

Biostatistical Analysis of DNA Methylation Profiling in Ovarian Cancer

A thesis presented for the degree of
Doctor of Philosophy of Imperial College
and the
Diploma of Imperial College
by

Wei Dai

Department of Surgery and Cancer
Imperial College
IRDB Building, Du Cane Road, London W12 0NN

JULY 19, 2011

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:



Copyright

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author and lodged in the doctorate thesis archive of the college central library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in Imperial College, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement. Further information on the conditions under which disclosures and exploitation may take place is available from the Imperial College registry.

Dedication

This thesis is dedicated to my family who gave me enormous support in the past four years, which is the source of my courage to start and to accomplish this PhD project.

Acknowledgements

I am extremely grateful to my supervisor, Prof. Robert Brown for his encouragement, guidance and support from the initial to the final stage of the study. His academic experience and patience is invaluable for me to develop an understanding of the project. I would also like to thank Mr. Jim Paul for offering advice in data analysis which has improved the quality of this work immeasurably, Dr. Sarah Blagden and Dr. Jenny Barrett for their thoughtful and constructive comments, and Dr. Natalie Shenker for proofreading the thesis.

Funding from Cancer Research UK (CRUK), has made this research possible. Ovarian Cancer Action (OCA) kindly provided the office space for this study. Scottish Gynaecological Clinical Trials (SGCT) group provided a high quality of ovarian tumour samples and clinical information. Expression data of the candidate genes in serous ovarian cancer and clinical information are provided by The Cancer Genome Atlas (TCGA) Pilot Project established by National Cancer Institute.

I am also very grateful that Dr. Jens M. Teodoridis and Dr. Constanze Zeller generated DMH data, and together with Dr. Janet Graham performed the bisulfite pyrosequencing and methylation specific PCR for the validation of methylation of candidate loci. Ms Jenny Hersey performed qRT-PCR to examine the expression of candidate genes identified in this study.

I wish to thank all my colleagues in Epigenetics Section at Department of Surgery and Cancer, as well as Tucker family for their kindness and support.

Abstract

Ovarian cancer is the most lethal gynaecological cancer. Although having good response to chemotherapy, the majority of the patients with advanced disease will eventually relapse. Aberrant DNA methylation in tumours has been proposed as biomarkers to predict patients' clinical outcome and response to chemotherapy.

An algorithm, Methylation Linear Discriminant Analysis (MLDA), was developed for large-scale methylation analysis using differential methylation hybridisation (DMH). MLDA identified loci differentially methylated between cisplatin sensitive and resistant derivatives of an ovarian tumour cell line with 89% accuracy and showed hypermethylation, rather than hypomethylation, predominantly occurred during the acquisition of cisplatin resistance.

Customised microarrays targeting promoter CpG islands in 10 key signaling pathways were designed for DMH analysis. Based on the power analysis epithelial ovarian tumours (screening study $n=120$, validation study $n=61$) prospectively collected through a cohort study, were firstly analysed by DMH at 302 loci spanning 189 promoter CGIs at 137 genes in the Wnt pathways for the association with progression free survival (PFS). Increased methylation of 6 loci, at *FZD4*, *FZD9*, *DVL1*, *NFATC3*, *ROCK1* and *NKDI* genes, were associated with shorter PFS independent from clinical parameters. A multivariate Cox model incorporates only *NKDI* and *DVL1*, identifying two groups differing in PFS (HR=2.72; permutation test $p = 4 \times 10^{-3}$). Consistent with epigenetic regulation, reduced expression of *FZD4* and *DVL1* is associated with poor relapse free survival in an independent

cohort ($p < 0.05$, $n = 321$). Analysis in further 9 pathways/families found 6 more independent biomarkers relevant to PFS at *PIK3R5*, *AKT1* and *VEGFB* from AKT/mTOR pathway, *PRDX2* and *TR2IT2* from Redox pathway and *MLH3* from MMR system.

The study shows DNA methylation changes are involved in acquired drug resistance, and demonstrates the importance of methylation at multiple promoter CGIs in key signaling pathways, especially in the Wnt pathway, for predicting clinical outcome in ovarian cancer and their potential as stratification biomarkers in future clinical studies for personalised treatment.

Table of contents

Dedication	4
Acknowledgements	5
Abstract	6
Contents	12
List of Tables	13
List of Figures	15
List of Abbreviations	16
List of Publications	18
1 Introduction	19
1.1 A Brief Introduction of Ovarian Cancer	20
1.1.1 Pathology and biology	21
1.1.2 Chemotherapy and acquired drug resistance	22
1.1.3 Prognostic and predictive factors of EOC	26
1.1.4 Gene expression profiling in EOC	30
1.2 Biological Functions of DNA Methylation	32
1.3 Aberrant DNA Methylation in Ovarian Cancer	35
1.3.1 Promoter CGI methylation and ovarian cancer	36
1.3.2 DNA methylation as a biomarker in ovarian cancer	36
1.4 High-throughput DNA Methylation Analysis	38
1.4.1 Restriction enzyme	41
1.4.2 Methylation antibody/methylation binding domain	42
1.4.3 Bisulfite conversion	43
1.4.4 Comparison of array-based DNA methylation profiling platforms	44
1.5 Bioinformatics Challenges in DNA Methylation Profiling Study	52
1.6 Aims of This Project and Outline of Thesis	53

2	Methylation Linear Discriminant Analysis (MLDA): An Algorithm Tailored for DNA Methylation Profiling Dataset	55
2.1	Introduction	56
2.2	Methods	60
2.2.1	Log-likelihood ratio transformation	62
2.2.2	Determination of log likelihood ratio cut-offs	64
2.2.3	Identification of robust regression outliers	66
2.2.4	Estimation of misclassification rate	68
2.2.5	SAM and PAM analysis	70
2.2.6	Making MLDA package under Windows	70
2.3	Results	71
2.3.1	DMH dataset on HCGI12K arrays	71
2.3.2	Validation of differential methylation	76
2.3.3	Implementation of MLDA on Agilent 15k custom-designed CpG island microarray	78
2.3.4	Modification of MLDA for within-array Normalisation	80
2.4	Discussion	82
2.4.1	Unmethylation controls and cross-hybridisation	84
2.4.2	Limitations of the study	85
2.4.3	Comparison of MLDA, SAM and PAM	86
3	Feasibility Study of DNA Methylation Profiling on Agilent Custom-designed Microarray in Ovarian Cancer Cell Lines	89
3.1	Introduction	90
3.2	Methods	94
3.2.1	Samples	94
3.2.2	Agilent custom-designed 8x15k microarray	95
3.2.3	Differential methylation hybridisation (DMH)	99
3.2.4	Statistical analysis	101
3.3	Results	104
3.3.1	Quality control in DMH	104
3.3.2	Reproducibility of Agilent custom-designed microarray	109
3.3.3	Dye bias correction	109
3.3.4	Methylation measurement of a CpG Island by the DMH assay	112
3.3.5	Sensitivity and specificity of DMH on Agilent custom-designed microarray	113
3.3.6	Correlation between DMH and bisulfite pyrosequencing	115
3.3.7	Determination of methylation status in cell line model	117
3.4	Discussion	119
3.4.1	Controls on the Agilent custom-designed microarray	119
3.4.2	Modification of DMH assay	122
3.4.3	Modification of MLDA	123

3.4.4	Identification of differential methylation within MseI fragment . . .	123
3.5	Conclusions	124
4	CpG Islands Methylation Profiling in the Wnt Pathway and Their Prognostic Significance in Epithelial Ovarian Cancer	127
4.1	Introduction	128
4.2	Methods	130
4.2.1	Patients	130
4.2.2	qRT-PCR	132
4.2.3	Target Selection on Agilent Custom-designed Microarray	133
4.2.4	Bisulfite Pyrosequencing	134
4.2.5	Statistical Analysis	135
4.2.6	The Cancer Genome Atlas (TCGA) Dataset	145
4.3	Results	146
4.3.1	Patients	146
4.3.2	Methylation of the Wnt pathway and PFS	150
4.3.3	Expression and progression free survival in TCGA dataset	157
4.3.4	Association of methylation at <i>DVL1</i> with response	157
4.3.5	Wnt pathway methylation index in late-stage EOC	158
4.3.6	Bisulfite pyrosequencing	160
4.4	Discussion	163
4.4.1	Systematic bias in DMH assay	164
4.4.2	Performance of methylation index in the Wnt pathway	164
4.4.3	Expression and methylation	165
4.4.4	Methylation of Wnt extracellular inhibitors in EOC	167
4.4.5	Methylation alterations in all Wnt pathways determine the patients with early disease progression	168
4.4.6	The biological role of <i>FZD4</i> and <i>DVL1</i> in the Wnt pathway	170
4.4.7	Sample size required for further validation study	174
4.5	Conclusions	175
5	CpG Islands Methylation Profiling in the Other Key Pathways and Their Prognostic Significance in Epithelial Ovarian Cancer	176
5.1	Introduction	177
5.2	Methods	179
5.3	Results	181
5.3.1	Methylation of other key pathways and progression free survival (PFS)	181
5.3.2	Expression of other key pathway and progression free survival (PFS)	184
5.3.3	Methylation and overall survival/response	187
5.4	Discussion	187

6	General Discussion and Future Perspectives	195
6.1	Association of Hypermethylation, DNA Damage and Cisplatin Treatment .	196
6.2	More Samples are Required for the Validation Study	197
6.3	Biological Mechanisms of Methylation at the Wnt Pathways Contributing to Ovarian Cancer Progression	197
6.4	Where Shall We Look at Besides Promoter CpG Islands?	198
6.5	Expansion of Methylation Model for Predicting Patients' Clinical Outcome	200
6.6	Methylation as Biomarkers for Surgery Decision	202
6.7	Identification of Aberrantly Methylated 'Hot Spots' in Ovarian Tumours . .	204
6.8	Methylation as Predictive Biomarkers for Treatment Decision	205
6.9	Methylation and Novel Treatments Targeting Molecular Pathways	207
	References	235

List of Tables

1.1	A summary of frequency, survival rate and molecular changes of malignant EOC	23
1.2	Response Evaluation Criteria in Solid Tumours (RECIST) criteria.	24
1.3	Prognostic factors in ovarian cancer at tumour presentation	28
1.4	REporting recommendations for tumour MARKer prognostic studies (RE-MARK)	31
1.5	Individual methylation prognostic/predictive biomarkers in ovarian cancer .	39
1.6	A panel of methylation prognostic biomarkers in ovarian cancer	40
1.7	Features of five types of microarrays combined with local techniques	48
2.1	Parameters of linear models in MLDA	75
2.2	Differentially methylated loci identified by MLDA, SAM and PAM on HCGI12K microarray	79
2.3	Differentially methylated CGIs identified by MLDA, SAM and PAM on Agilent custom-designed microarray	81
3.1	Controls on Agilent custom-designed microarray	96
3.2	Features for quality control in DMH assay	105
4.1	Estimated power of the analysis in the screening study	139
4.2	Estimated power of the analysis in the validation study	140
4.3	Patient characteristics	147
4.4	Univariate PFS analysis of clinical parameters in SGCTG cohort	148
4.5	Univariate overall survival analysis of clinical parameters in SGCTG cohort	149
4.6	Univariate PFS analysis of methylation in the Wnt pathway in the screening set	152
4.7	Univariate PFS analysis of methylation in the Wnt pathway in the validation set	153
4.8	Multivariate PFS analysis of methylation in the Wnt pathway in the screening set	154
4.9	Univariate overall survival analysis of methylation in the Wnt pathway in the screening set	156
4.10	Survival analysis of expression in the Wnt pathway in TCGA dataset	158

4.11	Correlation between expression and methylation	159
4.12	Response and methylation	160
4.13	Methylation frequency of extracellular inhibitors of the Wnt pathway	168
5.1	Disruption of cellular signaling pathways/families in human cancer by genetic and epigenetic alteration	180
5.2	Summary of univariate PFS analysis in 10 signaling pathways/families . . .	182
5.3	Univariate PFS analysis of methylation in 9 key pathways/families	185
5.4	Multivariate PFS analysis of validated methylation biomarkers in 9 key pathways/families	186
5.5	Univariate PFS analysis of expression of candidate genes in TCGA dataset .	187
5.6	Response and methylation of candidate genes in the key pathways	188
5.7	Methylation frequency of promoter CGI methylation in present and previous studies	190

List of Figures

1.1	Cytosine methylation	34
1.2	Coverage of DNA methylation microarrays at gene DVL1 and FZD4	50
1.3	An illustration of a CpG island related to two genes	52
2.1	Distribution of gene expression data and DMH data	58
2.2	An illustration of unmethylated and methylated model construction in MLDA in A2780 cell line	65
2.3	Determination of methylated and unmethylated cut-offs	67
2.4	Weighted scoring scheme in MLDA	69
2.5	Box plot of log ratios in 16 ovarian cancer cell lines	73
2.6	Consistent score and inconsistent score in 16 cell lines	74
2.7	Outliers identifications	77
2.8	Rotation of the coordinates using the origin as the centre	82
2.9	Within-array normalisation in MLDA to correct log DMH ratios	83
2.10	Venn diagram of loci identified by MLDA, SAM and PAM	87
2.11	Validation of differential methylation identified by MLDA and SAM on Illumina HumanMethylation27 BeadChip	88
3.1	Workflow of target selection on the Agilent custom-designed microarray . .	97
3.2	Annotation workflow of Agilent custom-designed microarray	99
3.3	Average correlation between replicates of filtered probe sets within arrays .	106
3.4	DMH dataset in pilot study	108
3.5	R^2 of linear regression model in duplicates	110
3.6	Dye bias correction in MLDA	111
3.7	Methylation of the loci within MseI fragment	114
3.8	Differential methylation between cisplatin sensitive and resistant cell lines .	116
3.9	Correlation between DMH and bisulfite pyrosequencing	118
3.10	Coverage of probes and significant level in Man-Whitney test	120
3.11	Methylation in DMH and MSP	121
3.12	Differential methylation hotspots at gene NR2E1 and WNT1	125
4.1	Schematic graph of the Wnt pathway	131
4.2	The principle of pyrosequencing	134

4.3	Genomic location of DVL1 tested by bisulfite pyrosequencing	136
4.4	Identification of prognostic biomarkers in the Wnt pathway	144
4.5	Unsupervised hierarchical clustering of the Wnt pathway	151
4.6	Kaplan-Meier plots of progression free survival	155
4.7	Kaplan-Meier plots of progression free survival in the patients with high/low methylation index	161
4.8	Kaplan-Meier plot of PFS in the risk groups predicted by leave-one-out cross validation	162
4.9	Correlation between DMH and bisulfite pyrosequencing at promoter CGI of DVL1	163
4.10	Correlation between DMH and bisulfite pyrosequencing at a locus on Chromosome 14	165
4.11	Receiver operator characteristic (ROC) curve	166
4.12	Methylation at Wnt β -catenin, PCP and/or Wnt/Ca ²⁺ pathways	171
4.13	Expression of FZD4 in IGROV1 SP and NSP cells	174
5.1	Unsupervised hierarchical clustering of methylation in 10 signaling pathways	183
5.2	Methylation of PRDX2 in two pairs of cell lines isolated from two patients before and after chemotherapy measured by DMH assay on Agilent custom-designed microarrays.	191
5.3	Methylation of 6 biomarkers in the patients received different treatments . .	193
6.1	Methylation interruption in the Wnt signaling pathways	199
6.2	Kaplan Meier plot of survival curves of late-stage patients with high/low methylation index estimated from the expanded model	203
6.3	Differential methylation observed in the patients received different platinum-based chemotherapy	206

List of Abbreviations

CGI	CpG island
CHARM	Comprehensive high-throughput arrays for relative methylation
CI	Confidence interval
CR	Complete response
DMH	Differential methylation hybridisation
DNMT	DNA methyltransferase
EOC	Epithelial ovarian cancer
FA	Fanconi anemia
FDR	False discovery rate
FIGO	International Federation of Gynaecology and Obstetrics
HDAC	Histone deacetylase
HMT	Histone methyltransferase
HR	Hazard ratio
HR	Homologous recombination
KEGG	Kyoto Encyclopedia of Genes and Genomes
KNN	k nearest neighbours algorithm
LOWESS	Locally weighted scatterplot smoothing
LR	Likelihood ratio
MI	Methylation Index

MLDA	Methylation Linear Discriminant Analysis
MMASS	Microarray based methylation assessment of single sample
MMR	Mismatch repair
MSP	Methylation specific polymerase chain reaction
NCRI	National Cancer Research Institute
NHEJ	Non-homologous end-joining
OR	Odd ratio
OS	Overall survival
PAM	Prediction Analysis of Microarrays
PBMC	Peripheral blood mononuclear cell
PCR	Polymerase chain reaction
PD	Progressive disease
PFS	Progression free survival
PR	Partial response
qRT-PCR	Quantitative real-time polymerase chain reaction
RECIST	Response evaluation criteria in solid tumours
REMARK	Reporting recommendations for tumour marker prognostic studies
SAM	Significant Analysis Microarrays
SD	Stable disease
SGCTG	Scottish Gynaecological Clinical Trials Group
TCGA	The Cancer Genome Atlas
TSS	Transcription start site

The full names of the genes identified in this study are in Appendix CD

List of Publications

Some of the research presented in this thesis can be found in the following publications:

- W. Dai, J. M. Teodoridis, J. Graham, C. Zeller, T. H. Huang, P. Yan, J. K. Vass, R. Brown, and J. Paul. Methylation linear discriminant analysis (MLDA) for identifying differentially methylated cpg islands. *BMC Bioinformatics*, 9:337, 2008.
- W. Dai, J. M. Teodoridis, C. Zeller, J. Graham, J. Hersey, J. M. Flanagan, E. Stronach, N. Siddiqui, J. Paul, and R. Brown. Systematic CpG islands methylation profiling of genes in the Wnt pathway in epithelial ovarian cancer identifies biomarkers of progression-free survival. *Clin Cancer Res* (*in press*)
- W. Dai, C. Zeller, J. M. Teodoridis, J. Hersey, J. M. Flanagan, N. Siddiqui, J. Paul, and R. Brown. Methylation biomarkers identified from systematic CpG islands methylation profiling in key signalling pathways predict progression-free survival in epithelial ovarian cancer (*in preparation*)

Chapter 1

Introduction

1.1 A Brief Introduction of Ovarian Cancer

Ovarian cancer is the most lethal and second most common of all the gynaecological cancers (Jemal et al., 2009). There are about 6850 new cases of ovarian cancer diagnosed each year in UK (<http://info.cancerresearchuk.org>). About 70% of ovarian cancers at diagnosis have spread beyond the pelvis into the abdominal cavity, near lymph nodes (stage III) or distant organs such as liver, lungs and brain (stage IV). Patients with advanced disease undergo surgical cytoreduction followed by platinum-based chemotherapy and most of them (>80%) are highly responsive to chemotherapy at the beginning of treatment. However, the majority of patients with advanced-stage disease eventually relapse and will acquire resistance to chemotherapy (Agarwal and Kaye, 2003). Therefore, the 5-year survival rate of ovarian cancer is low (40%)(<http://info.cancerresearchuk.org>).

Currently, the aetiology of ovarian cancer is still poorly understood. A strong family history of ovarian or breast cancer, often related to inherited mutation in one or two genes known as *BRCA1* and *BRCA2*, is the most important risk factor (Holschneider and Berek, 2000), however, only about 5-10% of women with ovarian cancer have *BRCA1* or *BRCA2* mutation (Holschneider and Berek, 2000; Sowter and Ashworth, 2005).

Ovarian cancer is a heterogeneous disease and can develop from epithelial cells, germ cells, stromal cells or sex-cord cells. Over 90% of ovarian cancers have been thought to arise from the epithelial surface of ovary. However, recently there is evidence that some epithelial ovarian cancers (EOC) arise from the fallopian tube epithelium (Crum et al., 2007; Kindelberger et al., 2007). As EOC accounts for the majority of ovarian cancers, the following sections are mainly focused on the pathology and biological properties, treatment

response, as well as the prognostic factors of EOC.

1.1.1 Pathology and biology

Epithelial ovarian cancers are classified as serous, endometrioid, mucinous, clear cell and undifferentiated (Rosen et al., 2009). The clinical behaviours of these subtypes of tumours vary, which is most likely to reflect variations in their biological properties. A number of studies have reviewed the clinicopathological features and molecular changes of these types of ovarian cancer (Cho and Shih Ie, 2009; Rosen et al., 2009; Bast et al., 2009). Table 1.1 summarises the frequency, 5-year survival rates and frequent molecular changes for these subtypes. In brief, the serous tumour is the most common subtype (about 70%) (Cho and Shih Ie, 2009). It is classified as high or low grade based on the extent of nuclear atypia and mitosis (Malpica et al., 2004). The morphologically defined grade subtypes are distinct diseases with different molecular alterations. *TP53* mutation is common in high grade serous ovarian cancer (HG-SOC). A recent comprehensive mapping study of *TP53* mutation shows a frequency as high as 96.7% and *TP53* dysfunction can be observed in 100% high grade serous carcinoma (Ahmed et al., 2010). *BRCA* dysfunction through germline and somatic mutation, down-regulation of gene expression by promoter methylation and amplification of trans-acting factors are observed in more than 50% of HG-SOC (Bowtell, 2010). A model that describes the genesis of high grade serous tumours proposed by Bowtell (2010), which suggests that *TP53* mutation followed by *BRCA* mutation is the primary event in the genesis of this type of tumours. The common precursor lesion of HG-SOC is not well recognised, although it may rise from ovarian surface inclusion cyst, peritoneal mesothelium, or the distal portion of the fallopian tube (Cho and Shih Ie, 2009). Low grade serous type is characterised by *KRAS* and *BRAF* mutations, while *TP53* mutation is rare. Endometrioid tumours comprise 10-20% of ovarian cancer cases. 5-year

survival rate of endometrioid tumours is 40-63%, which is higher than serous tumours (20-35%). *TP53* mutation is also common in high grade endometrioid tumours but is uncommon in the tumours with Wnt and/or PI3K/PTEN signalling defects (Cho and Shih Ie, 2009). However, mutations of β -catenin (*CTNNB1*) in the Wnt pathway and of *PTEN*, *PIK3CA* in the PI3K/PTEN pathway are frequently observed in low grade endometrioid tumours that are typically lack of *TP53* mutation. Mucinous and clear cell tumours account for 5-20% and 3-10% ovarian cancer cases, respectively. Molecular changes associated with these two tumour types have not been extensively studied yet. About 50% of mucinous tumours have *KRAS* mutation (Rosen et al., 2009). Mutations of *KRAS*, *BRAF* and *TP53* have been observed in clear cell tumours, but their frequencies are low (Rosen et al., 2009). *PIK3CA* mutations are reported in 20-25% clear cell tumours, and a small number of patients (about 8%) with clear cell cancers have *PTEN* mutations (Cho and Shih Ie, 2009). Recently, mutation of *ARID1A* was shown to be present in 46% of ovarian clear cell carcinomas and 30% of endometrioid carcinomas, thus has become a new molecular target for the treatment of these ovarian cancer subtypes (Wiegand et al., 2010).

1.1.2 Chemotherapy and acquired drug resistance

The recommended standard first line treatment for advanced ovarian cancer is platinum (cisplatin or carboplatin) combined with paclitaxel (du Bois et al., 2005). Platinum-based therapies can form intra- and inter-strand DNA adducts that inhibit replication and transcription. DNA damage caused by platinum can result in apoptosis (Agarwal and Kaye, 2003). Paclitaxel belongs to the taxane class of therapeutic agent and can bind to intracellular β -tubulin, which results in microtubule stabilization, growth arrest and apoptosis (Siddik, 2003). To monitor the response of treatment, the World Health Organisation (WHO), the National Cancer Institute and the European Organisation for Research and

Histology type	Frequency	5-year survival rate	Frequent molecular changes
Serous	~70%	20-35%	High grade: <i>TP53</i> Low grade: <i>KRAS</i> , <i>BRAF</i> mutation
Endometrioid	10-20%	40-63%	High grade: <i>TP53</i> mutation Low grade: <i>CTNNB1</i> , <i>PTEN</i> and <i>PIK3CA</i> mutation with microsatellite instability
Mucinous	5-20%	40-69%	<i>KRAS</i> mutation
Clear cell	3-10%	35-50%	<i>PTEN</i> , <i>PIK3CA</i> , <i>ARID1A</i> mutation
undifferentiated	1%	11-29%	unknown

Table 1.1: A summary of frequency, survival rate and molecular changes of malignant epithelial ovarian cancer adapt from Cho and Shih Ie (2009)

Treatment of Cancer have adopted a set of response criteria, Response Evaluation Criteria in Solid Tumours (RECIST) (Therasse et al., 2000). The RECIST criteria define response as complete response (CR), partial response (PR), stable disease (SD) and progressive disease (PD) based on the changes of tumour size measured by radiological modalities, such as computed tomography (CT) scanning and magnetic resonance imaging (MRI), at target tumour lesions of which the longest diameter was recorded (Table 1.2).

Although up to 80% of patients with ovarian cancer have a good response (complete/partial response) to initial chemotherapy, most of them will eventually relapse after a median period of 18 months (Greenlee et al., 2001). Although many tumours at the time of relapse are still chemo-responsive, they will eventually acquire resistance to treatment (Sood and Buller, 1998; Lister-Sharp et al., 2000; Greenlee et al., 2001). Therefore, an understanding the mechanisms of drug resistance, especially acquired drug resistance, have become very important for the successful treatment of ovarian cancer.

Response	Description
Complete response	the disappearance of all target lesions
Partial response	at least a 30% decrease of in the sum of the longest diameter of target lesions, taking as reference the baseline sum longest diameter
Progressive response	at least a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter recorded since the treatment started or the appearance of one or more new lesions
Stable disease	neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for progressive disease, taking as reference the smallest sum longest diameter since the treatment started

Table 1.2: Response Evaluation Criteria in Solid Tumours (RECIST) criteria.

Drug resistance can be intrinsic or acquired. Intrinsic resistance refers to tumours that do not respond to first-line chemotherapy and is conferred by the genotype of pretreatment clone, while acquired resistance is caused by mutations in progeny tumour cells after chemotherapy though the tumours initially respond (Balch et al., 2004). A number of studies have shown multiple mechanisms are involved in either intrinsic or acquired drug resistance in ovarian cancer. Generally, these mechanisms include decreased drug influx, increased drug efflux, altered drug target, drug metabolism, DNA damage repair, changes in cell cycle checkpoint mediators, and changes in downstream mediators of the apoptotic pathway (Hazlehurst et al., 2003). An inverse resistance relationship has been frequently observed between cisplatin and paclitaxel (Stordal and Davey, 2009), which adds another layer of complexity to drug resistance. As reviewed by Balch et al. (2004) altered gene expression levels have been found in resistant tumours compared to sensitive ones. Most

of those genes are involved in cell proliferation, DNA repair, angiogenesis, and cell migration. The changes of the gene expression are not only caused by mutation and copy number gain or loss, but aberrant epigenetic regulations, such as DNA methylation (see section 1.2 and 1.3) and histone deacetylation, also play a role. Several genes including tumour suppressors and genes involved in apoptosis have been shown epigenetically down-regulated in ovarian cancer (Balch et al., 2004). For example, the acquisition of methylation of the mismatch repair gene, *MLH1*, has been shown in 25% of plasma samples from the patients with ovarian cancer relapse (Gifford et al., 2004). The silencing of *MLH1* expression has previously been associated with resistance to platinum-based reagents (Brown et al., 1997). It is likely that the loss of *MLH1* expression results in defects in the mismatch repair system. This system can detect platinum-DNA adducts and activate apoptotic response. Therefore, the reduced levels of apoptosis in the cancer cells would contribute to the resistance to platinum-based drugs. Given epigenetic alterations have been shown to contribute to silencing of the genes associated with chemosensitivities, it has been proposed that the reversal of epigenetic silencing could resensitise ovarian tumours to chemotherapy (Balch et al., 2004). Our group has shown that the combination of DNA methyltransferase (DNMT) inhibitor and histone deacetylase (HDAC) inhibitor dramatically up-regulated a number of genes, including *MLH1*, in ovarian cancer cell line resistant to cisplatin (microarray data unpublished). Steele et al. (2009) also demonstrated that treatment with these two inhibitors resulted in a marked increase in the expression of epigenetically silenced *MLH1* both *in vitro* and *in vivo*. This further indicated the potential of using this combination to re-sensitise resistant ovarian tumours, as had previously been shown for a demethylating agent alone (Plumb et al., 2000).

1.1.3 Prognostic and predictive factors of EOC

As reviewed by Gasparini et al. (1993) in the context of breast cancer, a prognostic factor can be defined as "at a certain time point gives information about the subsequent clinical outcome". Accurate prognostic factors can identify the patients who are more likely to have an early-relapse and/or death, even though they have achieved complete response at the beginning of the treatment. Such patients are candidates for maintenance or consolidation therapy strategies. A predictive factor "gives information about selection of patients who will be benefit from treatment" (Gasparini et al., 1993). Accurate predictive biomarkers can identify the patients who are potential candidates for novel targeted therapy given that they have a high chance of exhibiting resistance to the conventional treatment, thus aiding personalised medicine strategies and potentially leading to more effective treatment. A prognostic factor is not necessarily predictive and vice versa, although certain factors can be both prognostic and predictive.

Currently, the established prognostic factors for EOC at clinical presentation before primary chemotherapy are residual disease, patient age at diagnosis, histology, stage, grade of tumour and performance status of the patient (Table 1.3) (Thigpen et al., 1993; Eisenhauer et al., 1999; Clark et al., 2001; Bristow et al., 2002; Cannistra, 2004). One of the most significant prognostic factors is stage of disease at diagnosis. Most cases already have advanced-stage tumours (Stage III or IV) at diagnosis. Patients with early-stage tumours have much better prognosis than the patients with advanced-stage tumours. It has been reported that five year survival rates for ovarian cancers are 90% in Stage I, 50% in Stage II, 20% in Stage III and 6% in Stage IV, respectively (<http://info.cancerresearchuk.org>). However, Clark et al. (2001) found the patients with Stage III and Stage IV have no difference in the risk of death, indicating in those patients stage

does not remain as an effective prognostic factor. Nevertheless, those clinical prognostic factors are imperfect predictors of outcome, and do not provide insights into the biological mechanisms that underlie the clinical behaviour of tumours. Therefore, the identification of reliable prognostic/predictive biomarkers would be of potential clinical value, provide the biological basis of tumour aggressiveness and resistance to chemotherapy, and identify drug targets for novel drug development strategies.

Although the potential prognostic value of individual biomarker in ovarian cancer have been investigated (Berchuck et al., 1990; Eisenhauer et al., 1999; Havrilesky et al., 2003; Verri et al., 2005; Bast et al., 2009; Lee and Park, 2009), to date, none of them have been clinically translated as independent prognostic biomarkers of sufficient value to direct clinical management of patients. There are many reasons that this may be the case, such as the lack of standardisation or reproducibility of the assays, poor study designs and inappropriate or misleading statistical analyses. Among these various reasons, study design has been a major issue, resulting in the study being underpowered or overly optimistic estimation of the effect size. Sample size is a major problem when complex data sets including thousands of biomarkers are examined simultaneously and where multiple testing caused a high false positive rate. As well as the small size of the samples, the following statistical problems are also common in these studies: (1) new biomarkers were examined individually without the simultaneous evaluation of prognostic value compared with conventional clinicopathological markers by performing a multivariate analysis; (2) the long-term clinical information were incomplete; (3) different chemotherapy treatment regimens were used, which would be likely to influence the prognostic value of the biomarkers, which was not accounted for in the analysis (Hall et al., 2004). Since, many prognostic biomarker studies have not been reported with sufficient information to allow for the adequate assessment of the quality of

Clinicalpathological factors	Description	Characteristics	Reference
Histology	Cell type of tumour	Serous, Mucinous, Endometrioid, Clear cell, others	(Teramukai et al., 2007; Clark et al., 2001)
Residual disease	Size of residual tumour after cytoreductive surgery	0cm, ≥ 1 cm or <2 cm, 2-5cm, >5 cm	(Teramukai et al., 2007; Clark et al., 2001)
Age	Age of patient at diagnosis	$\leq 69, \geq 70$ years	(Thigpen et al., 1993; Teramukai et al., 2007)
FIGO Stage	Extent of dissemination of disease	Stage I: Tumour is confined to the ovary Stage II: Tumour involves one or both ovaries and has extended into the pelvis Stage III: tumour involves one or both ovaries with microscopically confirmed peritoneal metastasis outside the pelvis and/or regional lymph node metastasis Stage IV: distant organ involvement, including pleural space or hepatic or splenic parenchyma	(Holschneider and Berek, 2000; Clark et al., 2001)
FIGO Grade	Degree of differentiation in tumour cells	I: well differentiated II: moderately differentiated III: poorly differentiated	(Clark et al., 2001)
Performance status	The physical status of the patients	0: normal activity 1: symptoms, but nearly ambulatory 2: some bed time, but needs to be in bed less than 50% of the normal daytime 3: needs to be in bed more than 50% of the normal daytime 4: unable to get out of bed	(Teramukai et al., 2007; Clark et al., 2001)

Table 1.3: Prognostic factors in ovarian cancer at tumour presentation. Abbreviation: FIGO International Federation of Gynaecology and Obstetrics

the study, REMARK guidelines (Table 1.4) (McShane et al., 2005) have been developed to encourage transparent and complete reporting in prognostic studies. These guidelines summarised key information on study design, patient characteristics, assay methods, data presentation and statistical analyses. In these guidelines, a clear objective of the study is required. The number of total patients and events (effective sample size) are the key items, and the rationale for the sample size should be given to determine the power of the study. Items 2, 3 and 6 of the guidelines recommend that the patient characteristics should be reported, including the treatment received and how the patients are selected in the study. This could help the reader understand how the study result can be applicable in a clinical situation, and to evaluate whether there was any selection bias in the study. The clinical endpoints examined in the study should be precisely defined as recommended in item 7. For overall survival, it should be defined whether the death referred to all deaths or only cancer-related deaths; with regards to progression free survival, the definition of disease progression should be clearly stated. In terms of statistical analysis, item 10 and 11 recommend that the statistical model methods and test of the assumption should be reported, for example if the Cox proportional hazard model is used, the assumption of the proportional hazards should be tested. Furthermore, the study should clarify how the marker values were handled, such as the use of a continuous variable or categorical variable. If a categorical variable was used, the paper should describe how the cut-point was determined should be stated. In data presentation (item 12 and 13), the flow of patients including the inclusion and exclusion criteria through the study should be described, and the distribution of clinical parameters should be reported to define the studied patients cohort further. In terms of data analysis (item 14 to 18), the relationship between the biomarkers and standard prognostic variables should be estimated. A univariate analysis is used to evaluate the correlation between biomarkers and outcome. The estimated effect of the biomarkers should

be adjusted in the multivariate analysis including standard prognostic factors, to assess if the biomarkers are independent or not. The hazard ratio, 95% confidence interval of the hazard ratio and p value should be reported for both biomarkers and standard prognostic factors. Internal validation or independent validation should be performed to evaluate how good the model is when fitting with new dataset. Finally, the limitations and implications of the study should be discussed (item 19 and 20).

Following REMARK guidelines, Mallett and his colleagues (2010) evaluated the prognostic studies published between January 2006 and April 2007 (the earliest publication of REMARK is in 2005, and very few studies referred REMARK criteria during the first two years). They found that the reporting of these key information was poor prior to the publication of the REMARK guidelines. A REMARK profile based on these guidelines was recommended by Mallett et al. (2010) to improve reporting and enhance prognostic research studies.

Compared to prognostic factors, the identification of predictive factors is more complicated and difficult. Although REMARK mentions the reporting of predictive markers, as claimed by the authors, the guidelines are not adequate for predictive studies. Therefore, a more specific set of guidelines for the investigation of predictive markers is required.

1.1.4 Gene expression profiling in EOC

With advances in DNA-microarray technology, it has become possible to analyse expression of thousands of genes simultaneously in ovarian tumour tissues, allowing the identification of gene expression profiles that are associated with clinical outcome and/or chemotherapy response in patients with EOC. This has provided insights into molecular changes

Introduction	
1	State the marker examined, the study objectives, and any prespecified hypotheses.
Materials and methods	
Patients	
2	Describe the characteristics (e.g. disease stage or comorbidities) of the study patients, including their source and inclusion and exclusion criteria.
3	Describe treatments received and how chosen (e.g. randomized or rule-based).
Specimen characteristics	
4	Describe type of biological material used (including control samples) and methods of preservation and storage.
Assay methods	
5	Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.
Study design	
6	State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g. by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.
7	Precisely define all clinical endpoints examined.
8	List all candidate variables initially examined or considered for inclusion in models.
9	Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.
Statistical analysis methods	
10	Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.
11	Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.
Results	
Data	
12	Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.
13	Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.
Analysis and presentation	
14	Show the relation of the marker to standard prognostic variables.
15	Present univariate analyses showing the relation between the marker and outcome, with the estimated effect (e.g. hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan–Meier plot is recommended.
16	For key multivariable analyses, report estimated effects (e.g. hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.
17	Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.
18	If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.
Discussion	
19	Interpret the results in the context of the prespecified hypotheses and other relevant studies; include a discussion of limitations of the study.
20	Discuss implications for future research and clinical value.

Table 1.4: REporting recommendations for tumour MARKer prognostic studies (RE-MARK) from McShane et al.(2005)

during cancer growth and chemotherapy resistance. Some of these studies have been reviewed by Konstantinopoulos et al in 2008 (Konstantinopoulos et al., 2008). Later that year, Tothill et al. (2008) reported on microarray gene expression profiling in 285 serous and endometrioid tumours of the ovary (GEO accession ID: GSE9899) and identified 6 molecular subtypes of ovarian cancer. Two subtypes predominantly represent the serous low malignant and low grade endometrioid subtypes. The remaining subtypes represent higher grade and advanced stage cancers of the serous and endometrioid subtypes. Crijns et al. (2009) (GEO accession ID: GSE13876) profiled the gene expression of 157 advanced stage serous ovarian cancers and identified a survival-related gene signature that involved 17 key signalling pathways as well as 13 transcription factors associated with overall survival. The majority of these gene expression signatures have been validated in an independent patient cohort, although most of validation sets are small. The Cancer Genome Atlas Project (<http://cancergenome.nih.gov/>) has accomplished the gene expression profiling of over 300 serous ovarian tumours, which has provided a valuable source for the further validation of these signatures. Meanwhile comprehensive catalogues of mutations, copy number variations and methylation profiling have also been developed in the same patient cohort. The integration of these "omics" datasets would be useful to identify patients who gain less benefit from conventional treatment strategies, and therefore are potential candidates for novel targeted treatments.

1.2 Biological Functions of DNA Methylation

In humans, DNA methylation usually occurs at CpG dinucleotides with the addition of a methyl-group to the 5' position of cytosine. This reaction is catalysed by members of DNA methyltransferase (DNMT) family (Figure 1.1). The DNMTs implicated in DNA methylation in humans include DNMT1, DNMT3a and DNMT3b. DNMT1 is involved in the

maintenance of methylation, primarily acting on hemi-methylated DNA during DNA replication, while DNMT3a and 3b are thought to be involved in de novo methylation facilitated by another protein DNMT3L, which is primarily expressed during development (Pradhan and Esteve, 2003; Jia et al., 2007). CpG dinucleotides are under-represented in the human genome (only 21% of the expected CpG sites)(Illingworth and Bird, 2009). This is mainly because the spontaneous deamination of cytosine residues gives rise to uracil residues that are recognised and repaired by the cell, but deaminated 5-methylcytosine resembles thymine and is left unrepaired (Lander et al., 2001). It was estimated that 70% of cytosines in CpG dinucleotides are heavily methylated, predominantly at repetitive elements to prevent chromosomal instability (Suzuki and Bird, 2008). However, some CpG-rich regions (on average 1 CpG per 10bp, which is much higher than the ratio of 1 per 100bp in the genome), accounting for <2% of CpG dinucleotides, are known as CpG island (CGI). CGIs are frequently located in the promoters and the first exons, generally remain unmethylated. Exception to this are methylated CGIs, which are found close to genes on the inactive X chromosome, imprinted genes, and tissue-specific genes (Baylin and Jones, 2006). To define a CGI a mathematical model was proposed by Gardiner-Garden: CG content is greater than 50%, the sequence length is over 200bp and the ratio of observed to expected CpGs is larger than 0.6 (Gardiner-Garden and Frommer, 1987). Using this model, genome-wide CGIs have been predicted in humans *in silico* and information about these CGIs is available on the UCSC Genome Bioinformatics website (<http://genome.ucsc.edu/>).

The mechanism by which CGI remain unmethylated compared to the global methylation is still unclear. It could be that CGIs are targeted by certain DNA demethylation mechanisms, which results in the removal of the methyl-group from cytosines (Frank et al., 1991). Another possible alternative is that since CGIs are often co-localised to the promoter

htb

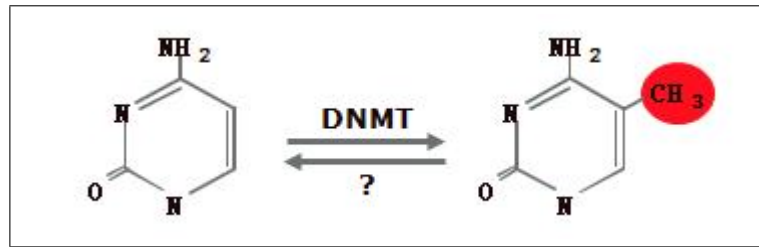


Figure 1.1: Cytosine methylation. DNA methylation is catalysed by DNA methyltransferase (DNMT)

region and are therefore bound by transcription factors, this precludes the association of DNMT at the CGIs (Illingworth and Bird, 2009).

Only about 3% of transcription start sites that are associated CGI are normally methylated in somatic tissue (Weber et al., 2007). Methylation of promoter CGIs is often associated with transcriptional silencing (Bird, 2002). Methylation could prevent transcription by the presence of the methyl-group in the major groove, which would directly prevent the binding of transcription factors. This possibility has been supported by the observation that several transcription factors cannot bind when the DNA recognition sequences of the transcription factors are methylated (Eden and Cedar, 1994). However, methylation is often capable of repressing transcription only after chromatin has been assembled, suggesting that this direct effect is not the main mechanism by which methylated DNA inhibits gene expression (Curradi et al., 2002). The other proposed mechanism involves the interaction between DNA methylation and histone modifications to determine gene expression status (Klose and Bird, 2006). It has been shown that methylated DNA can recruit methyl-CpG binding domain (MBD) proteins that associate with histone deacetylases (HDACs) and histone methyltransferases (HMTs). Histone deacetylation increases the positive charge of histones, which are then attracted by the negative charge of the DNA back bone, leading

to compression of the chromatin. Histone methylation, particularly H3K9 methylation and other repressive histone marker, binds with heterochromatin protein-1 (HP1) (Jenuwein and Allis, 2001). This establishes a repressive chromatin structure that represses gene expression. Meanwhile, DNMTs participate in multi-protein complexes that contain HDACs and HMTs causing the methylation of DNA. This further reinforces the repression of transcription (Feinberg and Tycko, 2004). Whether DNA methylation or histone modifications are the initiation event in the regulation of gene expression remains unclear.

1.3 Aberrant DNA Methylation in Ovarian Cancer

Two major types of aberrant DNA methylation changes occur in cancer compared to normal cells (Barton et al., 2008). One is global hypomethylation, which is mainly detected at repetitive elements. This could lead to chromosome instability. The other is hypermethylation at CGIs, which is associated with gene inactivation, and particularly affects tumour suppressor genes (Feinberg and Tycko, 2004). Since the methylation of CGI at the promoter region of the gene can cause gene silencing and aberrant methylation patterns have been observed in cancer cells, Peter A. Jones and Peter W. Laird (1999) suggested that promoter CGI methylation is the third type of hit that can cause the loss of function of tumour suppressor genes that are involved in tumorigenesis. Therefore, Knudson's two-hit hypothesis has been revised to include direct mutations in the coding regions, partial or complete loss of gene copies, and epigenetic silencing, which participate in different combinations to cause a complete loss of activity of a specific gene.

1.3.1 Promoter CGI methylation and ovarian cancer

As reviewed by Barton and his colleagues (2008), a number of genes have been found to be hypermethylated and become silenced in ovarian cancer compared to the normal ovarian surface tissues. The methylation frequencies of those genes varies widely in independent studies: the variation observed could be due to many reasons including difference in the composition of patients studied, different regions of interest examined and methods used to measure methylation. The genes with aberrant methylation in ovarian cancer are involved in multiple biological processes relevant to cancer development and progression including cell cycle, DNA repair, cell adhesion, microtubule stability, and cell proliferation amongst others. For instance, the tumour suppressor genes, *p16*, (*CDKN2A*), *BRCA1*, *RASSF1A* and *APC* have been found methylated in ovarian cancer tissues (Teodoridis et al., 2005). The methylation of tumour suppressor genes can serve as the second hit, thus leading to the initiation of cancer. As well as causing the direct inactivation of tumour suppressor genes, methylation can also silence transcription factors, such as *GAGT4* and *GAGT5* (Wakana et al., 2006), and cause the inactivation of downstream targets. In addition, the hypermethylation of genes in the DNA repair systems, such as, *BRCA1*, *GSTP1* and *MGMT*, is associated with a good response to chemotherapy in ovarian cancer (Teodoridis et al., 2005).

1.3.2 DNA methylation as a biomarker in ovarian cancer

There is an increasing amount of literature about the use of DNA methylation as a biomarker for the early detection of cancer, as well as the prediction of clinical outcome, and the patient response to chemotherapy. The use of DNA methylation as a biomarker has several advantages: (1) DNA is a more stable molecule than RNA or protein; (2) the detection of DNA methylation can focus on a small region, such as a CpG island, rather than screening

the whole gene; and (3) DNA methylation are detectable in body fluids such as serum or plasma (Cottrell and Laird, 2003; Gifford et al., 2004).

Table 1.5 lists the studies since 2004 reported on individual DNA methylation as prognostic or predictive biomarkers in malignant ovarian cancer at tumour presentation and response to first-line chemotherapy. Very few methylation prognostic biomarkers in ovarian cancer were reported before 2004, probably as a result of technological limitations. In spite of certain common problems in studies about prognostic biomarkers, such as small sample size, incomplete clinical information, no consideration of clinical parameters and treatment, and no further validation in an independent patient cohort (see section 1.1.3), other issues in many of these studies would confound the analysis regarding the biological features of DNA methylation, as outlined: (1) the promoter CGI is not well defined, and therefore, the regions that are methylated in the primary tumours are often unclear, with exception of the MCJ study (Strathdee et al., 2005); (2) expression of the candidate genes is not always evaluated in these studies, however, confirmation of the prognostic/predictive value at two molecular levels (e.g. methylation and expression) would increase the reliability of the biomarker and provide a better understanding of the function of methylation in the candidate genes; and (3) methylation is used as a categorical variable in the analysis rather than as a continuous measurement. This could reduce the statistical power of the analysis, especially when sample size is small. Furthermore, methylation cut-offs for a continuous variable can be arbitrary and are not always biologically relevant.

A panel of multiple biomarkers may be more informative given that cancer is a complex disease involving multiple biological mechanisms. Such a panel could improve the predictive accuracy since an individual biomarker is more likely to give a false negative

result than multiple biomarkers due to the limitations of the detection methods. Table 1-5 details some of the studies that have used the methylation levels of a panel of genes as the predictor of clinical outcome in ovarian cancer patients. Teodoridis et al. (2005) showed that methylation of at least one of three genes (*BRCA1*, *GSTP1* and *MGMT*) is correlated with good response to chemotherapy. Later, Su et al. (2010) showed the methylation of at least one of three genes (*SFRP1*, *SOX1* and *LMX1A*) is a predictor of overall survival, while methylation of at least one gene of *SFRP1*, *SFRP2* and *SOX1* is a predictor of disease recurrence in ovarian cancer patients. Since measurement of methylation at a panel of genes is very labour-intensive, array-based approaches have been utilized to identify the prognostic methylation signatures that can predict patients' clinical outcome. Wei et al. (2002) identified such a methylation signature including 956 loci that can stratify the late-stage patients into two distinct groups with different clinical outcome. Later, the signature was reduced to 112 loci to predict patient progression free survival with 95% accuracy using bioinformatics algorithms (Wei et al., 2006). Although all these studies still need further validation in randomised controlled trials, the identification of a panel of methylation biomarkers in ovarian cancer has become possible.

1.4 High-throughput DNA Methylation Analysis

Lots of efforts have been made to develop methods for large-scale DNA methylation analysis, since it has been realised that DNA methylation plays an important role in multiple processes of various diseases including cancer. Genome wide studies on DNA methylation primarily rely on three techniques based on restriction enzyme, sodium bisulfite conversion, or affinity purification with methylation binding domain proteins or a methyl-cytosine antibody. These methods can be combined with global approaches, such as DNA microar-

Genes	Genomic location	End point	Hazard ratio	95% CI	p value	number of age patients	Stage	Grade	Histology	Residual disease	Treatment	Method to measure methylation	Univariate analysis	Multivariate analysis	Method to measure expression	validation	Reference
<i>BRCA1</i>	promoter	OS			methylation vs. mutation 0.02; methylation vs. WT 0.08	11	median: 60	G1-G2 (100%)	S (55%); O (45%)	<1cm (80%)	Platinum-based	MSP	Y	Y	N	N	(Chiang et al. 2006)
<i>MCJ</i>	-400 - +700bp of TSS	response ¹ OS	2.9		0.027 0.023	41		III-IV (100%)			Platinum-based	Bisulfite sequencing	Y	N	N	N	(Strathdee et al. 2005)
<i>FANCF</i>	promoter	PFS response ¹	3.7* 1.43-8.54*		0.001* 0.0098	53	FANCF methylated, 64.1; FANCF unmethylated 59.3	G1-G2 (45.3%); III-IV (100%)	G1-G2 (45.3%); C3 (41.5%)	<1cm (26.4%); >=1cm (71.7%)	Platinum-based	MSP	Y	Y	qRT-PCR	N	(Lim et al., 2008)
		OS	1.162	0.343-3.113	0.77								Y	Y	N	N	
<i>CCND2</i>	Promoter	DFS OS			0.0212; multivariate analysis not significant 0.0625; multivariate analysis not significant	71	<50 (40.8%); >=50 (59.2%)	I-II (33.8%); G2 (49.3%); III-IV (31%); G3 (50.7%)	S (36.6%); E (21.1%); M (66.2%); C (9.9%); U (32.4%)	<2cm (66.2%); >=2cm (33.8%)		MSP	Y	Y	qRT-PCR	N	(Sakuma et al. 2006)
<i>p16</i>	Promoter	PFS OS response ²	1.54* 1.15*	1.01-2.34* 0.75-1.77*	<0.05* >0.05* 0.931	249		I-II (27.3%); G2 (16.1%); III-IV (61%)	S 34.5%; E (14.1%); M (38.6%); C (7.2%); U (16.1%); O (24.1%); G3 (15.3%); U (56.6%)	0cm (1.7%); 1-4cm (24.1%); >5cm (22.1%)	Platinum-based	MSP	Y	N	N	N	(Katsaros et al. 2004)
<i>Hsulf1</i>	CGI in Exon 1A	response ²			0.0146	60	Median: 61	III-IV (100%)	S (85%); O (15%)	optimal (1.7%); suboptimal (11.7%)	Platinum-based	Bisulfite sequencing ³	Y	N	TMA	N	(Staub, J et al. 2007)
<i>HOXA11</i>	Promoter	OS	3.4*	1.2-9.9*	0.03*	22	>60 (47.8%); <60 (52.2%)	I-II (32.6%); G1-G2 (66.3%); III (31.5%)	S (44.6%); E (15.2%); M (70.7%); C (35.9%); O (4.3%)	<2cm (73.8%); >2cm (23.3%)	Platinum-based	PMP	Y	Y	N	N	(Fiegl et al. 2008)
<i>FBXO32</i>	Promoter	PFS OS	1.003* 1.002*	1-1.006* 1-1.005*	0.028* 0.095	96	>=60 (32.3%); <60 (67.7%)	I-II (39.6%); G1 (67.4%); III-IV (60.4%)	S (65.9%); E (18.3%); M (6.3%); C (21.4%); O (1.6%)	optimal (73.8%); sub-optimal (26.2%)		qMSP	Y	Y	qRT-PCR	N	(Chou et al., 2010)
<i>SFRP5</i>	Promoter	response ²			p<0.0001	126	Median: 53	I-II (34.1%); G2 (21.4%); III-IV (65.9%)	S (65.9%); E (73.8%); M (6.3%); C (21.4%); O (1.6%)	optimal (73.8%); sub-optimal (26.2%)	Platinum-based	MSP	Y	N	qRT-PCR	N	(Su et al. 2010)

Table 1.5: Individual methylation prognostic/predictive biomarkers in ovarian cancer. ⁻¹Response was measured by tumour size. ⁻²Response was measured by both tumour size and CA125 levels. *Multivariate survival analysis result. Blank cell indicates the information is not given. S: serous; E: endometrioid; M: mucinous; C: clear cell; O: others. *BRCA1*(Chiang et al., 2006); *MCJ*(Strathdee et al., 2005); *FANCF*(Lim et al., 2008); *CCND2*(Sakuma et al., 2007); *p16*(Katsaros et al., 2004); *Hsulf1*(Staub et al., 2007); *HOXA11*(Fiegl et al., 2008); *FBXO32*(Chou et al., 2010); *SFRP5*(Su et al., 2010)

Genes	Genomic location	End point	Hazard ratio	95% CI	p value	number of patients	Stage	Grade	Histology	Residual disease	Treatment	Method to measure methylation	Method to measure expression	Validation	Reference
956 loci	CGI	PFS				19	III-IV (100%)				Platinum-based	DMH		N	(Wei et al. 2002)
BRCA1/GSTP1/ MGMT	Promoter	Response ¹			0.013	70	Median:6 0	III-IV (100%)	S(51.9%); NS (23.6%)		Platinum-based	MSP		N	(Teodoridis et al. 2005)
112 loci	CGI	PFS				40	III-IV (100%)					DMH		N	(Wei et al. 2006)
SOX1/SFRP1/SF RP2	Promoter	PFS	3.19	1.28-7.93		126	I-II (34.1%); Median:53 III-IV (65.9%)	G1 (18.3%); G2 (21.4%); G3 (59.5%)	S (65.9%); E (6.3%); M (26.2%); O (1.6%)	optimal (73.8%); sub-optimal (26.2%)	Platinum-based	MSP		N	(Su et al, 2008)
SOX1/LLMX1A/ SFRP1	Promoter	OS	2.71	1.06-6.90										N	

Table 1.6: A Panel of methylation prognostic biomarkers in ovarian cancer. ⁻¹Response was measured by tumour size. Blank cell indicates the information is not given. S: serous; NS: not serous; M: mucinous; E: endometrioid; O: others

rays, for methylation pattern study or high-throughput sequencing for single-base-pair resolution methylation analysis in the whole genome (Schones and Zhao, 2008). Although next-generation sequencing combined with these three local techniques is more powerful as it provides single-base resolution, and allele-specific readout and does not require the design of an appropriate microarray, array-based technology is still popular due to its low cost and the simplicity of down-stream data analysis. Therefore, only microarray-based technologies are discussed here, as they are more relevant to the content of this thesis.

1.4.1 Restriction enzyme

This approach uses a methylation-sensitive restriction enzyme, such as HpaII (CCCG), BtUI (GCGC), MspI (CCGG), Acil (CCGC and GCGG), HhaI (GCGC) and Tail (ACGT), which only recognizes a specific unmethylated DNA sequence. These sequences are then cleaved if unmethylated or remain uncleaved if methylated. One example of the use of this method is differential methylation hybridisation (DMH). In this method, the sample digested by methylation-sensitive restriction enzyme and the mock digested sample are labelled with different fluorescent dyes and competitively hybridised on the two-colour microarray. DNA methylation is measured by the relative fluorescent signal intensities. In DMH, genomic DNA is pre-digested by MseI, which recognises the site T/TAA. TTAA is rare in CGIs, so digestion by MseI could deplete the CpG-poor DNA. The advantage of this method is that it allows the methylation profiling of genome-wide regions by the identification of cleaved sites. The disadvantages of this method are as follows: (1) methylation analysis can only take place within the recognition sites, (2) it is prone to false negatives caused by incomplete digestion rather than DNA methylation, and (3) the size of MseI fragment introduces the bias in the detection of methylation.

In an adapted DMH method, the methylation-specific restriction enzyme, McrBC, is used to specifically recognise methylated DNA sequences. McrBC recognises two half-sites of the G/AmC, which are generally separated by about 55-103bps but can be separate by up to 3kb (Sutherland et al., 1992). This adaption can increase the sensitivity to densely methylated regions and reduce the false positives. This method was further modified in MethylScope using sheared DNA rather than MseI fragments to avoid the size bias in methylation analysis (Lippman et al., 2004). The McrBC-based method is further improved by comprehensive high-throughput arrays for relative methylation (CHARM) (Irizarry et al., 2008) by optimizing the workflow for microarray design and data processing. To overcome the regional limitations of using restriction enzyme and improve the sensitivity and specificity of methylation detection, a cocktail method, known as microarray-based methylation assessment of single samples (MMASS) (Ibrahim et al., 2006), was developed. In MMASS, one half of the DNA sample is digested by a number of methylation-sensitive enzymes and the other half is digested by McrBC, the two parts of sample are then competitively hybridized on the microarray.

1.4.2 Methylation antibody/methylation binding domain

Another popular method for achieving methylation enrichment is using methyl-binding proteins, such as MeCP2, or antibodies specific for 5-methylcytosine. The enrichment of methylated regions by the immunoprecipitation of genomic DNA with a 5-methylcytosine specific antibody, followed by hybridisation on the microarray is known as MeDIP (Weber et al., 2005). This method has been implemented by Agilent Technology for genome-wide methylation analysis by a CpG island tiling microarray. Recently, higher affinity methyl-binding proteins including multimerized MBD domains (Jorgensen et al., 2006) and a protein complex of MBD2b (short isoform of MBD2) and MBD3L (Rauch and Pfeifer,

2005) have been used to improve the affinity enrichment. However, this method requires the appropriate adjustment of the down-stream data analysis for variable CpG densities in different regions of the genome.

1.4.3 Bisulfite conversion

Treating with sodium bisulphite can convert cytosine to uracil, but leave the 5-methylcytosines intact. This method has been combined with bead arrays manufactured by Illumina (CA, USA) to obtain a DNA methylation profile in a large scale. The first Illumina assay for methylation studies was the GoldenGate Methylation assay (Bibikova et al., 2006), which covered 1505 CpG sites selected from cancer-related hot spots. Two primers that targeted the unmethylated and methylated sequences were labelled by different fluorescent dyes and the PCR products were hybridised on the microarray. Later, this assay was adapted to have a genome-wide resolution by using Infinium assay on HumanMethylation27 BeadChip, and targeted 27,578 CpG sites. In this assay, after bisulfite conversion, genomic DNA is amplified and enzymatically fragmented, and then hybridised on the BeadChip. DNA molecules are annealed to the oligos on the BeadChips that are linked to two different bead types- one links to the methylated sites and the other to the unmethylated CpG sites. After single-base extension using DNP- and Biotin-labeled dNTPs, both bead types incorporate the same type of labelled nucleotides, depending on the methylation status of the cytosine in the targeted CpG site. Both platforms achieve a single-base resolution read out and have been applied in methylation studies of various cancers (O’Riain et al., 2009; Christensen et al., 2010; Deneberg et al., 2010; Thirlwell et al., 2010; Wilhelm-Benartzi et al., 2010).

1.4.4 Comparison of array-based DNA methylation profiling platforms

To combine genome-wide technologies with local techniques, various microarrays have been designed. Many of them have competing strengths and weaknesses in human DNA methylation study. In this section, the advantages and disadvantages of five types of microarrays including HCGI12K, Illumina Infinium HumanMethylation27 BeadChip, Agilent 244K CpG island microarray as well as two custom-designed microarrays on the Agilent platform (see below) are discussed.

1. HCGI12K two-channel microarray: HCGI12k microarray (University Health Network, Toronto, Canada) includes about 12,000 probes from a library of CGIs purified by methyl-CpG binding domain (MBD) column (Cross et al., 1994).
2. Agilent 244K CpG island two-channel microarray: Agilent 244k CpG island microarray (product number: G4492A, array design ID: 014791, Agilent, UK) contains about 244,000 probes in or within 95bp of 27,800 CGIs fulfilled Gardiner-Garden criteria (Gardiner-Garden and Frommer, 1987) and were predicted by UCSC algorithm generated using a modification of a program developed by G. Miklem and L. Hillier (unpublished).
3. Illumina infinium HumanMethylation27 BeadChip: Infinium Human Methylation27 BeadChip (Illumina, CA, USA) simultaneously interrogates 27578 CpG sites within the promoter regions across more than 14000 genes. The methylation of each CpG site is represented by β value, a continuous variable ranging from 0 (unmethylated) to 1 (methylated).
4. Agilent custom-designed 15K two-channel microarray: this array was customised for this PhD project targeting about 800 promoter CGIs within 2kb of transcription

start sites of genes identified to be associated with acquired drug resistance and 10 key signalling pathways (see feasibility study in Chapter 3 and the screening study in Chapter 4). The 60-mer probes on the microarray were mainly selected from OligomeTM database (OGT, UK) and fabricated on an Agilent 8x15K platform (Agilent, UK).

5. Agilent custom-designed 60K two-channel microarray: this array was also designed in this PhD project targeting the regions from -2kb to +8kb of the transcription start site of the genes in 10 key signalling pathways (see validation study in Chapter 4 and 5). Similar to custom-designed 15K microarray, the 60-mer probes on the microarray were mainly selected from OligomeTM (OGT, UK) and fabricated on an Agilent 8x60K platform (Agilent, UK).

Among the five types of microarrays, the HumanMethylation27 BeadChip using an Infinium assay is based on bisulfite conversion, while the other 4 arrays have two-channel outputs and can be either combined with restriction enzyme or affinity enrichment approaches to achieve a methylation profile. The choice of techniques and microarrays are mainly influenced by the reproducibility and accuracy of the approaches, genomic coverage and read out resolution, input sample requirement, the number of samples that can be assayed on the microarray, the specificity of the probes, the quality of microarray annotation, and the purpose of the study. A summary of the features of these microarrays are shown in Table 1.7.

Sample requirements and throughput

Generally, restriction enzyme-based approaches require that the DNA is of high purity, quantity and integrity; while affinity-enrichment techniques are more tolerant to DNA impurities. However, the latter method requires substantial quantities of genomic input DNA

in order to achieve sufficient enriched output DNA. As low quantities of input DNA can be increased by PCR amplification after the DNA methylation-dependent step, Illumina Infinium assay only requires 0.5-1 μ g DNA. Furthermore, the Illumina Infinium platform has been shown to be compatible with degraded DNA extracted from formalin-fixed, paraffin-embedded samples (Thirlwell et al., 2010).

Restriction enzyme and affinity-based methods are very labour-intensive, which limits the use of these two types of method to small scale experiments only. Highly automated techniques have low labour costs but consume a large amount of reagents. Therefore, the semi-automated method of the Illumina Infinium assay is often reported in methylation studies that use a large number of samples.

Reproducibility, accuracy and quality controls

Variations can be introduced from any of the multiple steps of array-based approaches, including DNA preparation, hybridisation and image scanning. In particular, restriction enzyme-based methods, such as McrBC, involve extra steps of digestion, ligation and PCR amplification. Incomplete restriction is also one of the various factors that can influence the reproducibility and accuracy of the approach. On our custom-designed microarrays (Agilent custom-designed 15K and 60K microarrays), extra methylation controls were designed to monitor this bias. The introduction of unmethylated controls for normalisation purposes across the arrays in the data analysis could reduce systematic variations from various sources.

The HumanMethylation27 BeadChip is based on bisulfite conversion and tends to be more reproducible, although the result can be influenced by incomplete bisulfite conversion

and different PCR amplification efficiencies for the methylated and unmethylated versions of the same sequence. Successful bisulfite modification can be monitored by adding an extra quality control step by testing with calponin-specific primer sets prior to the Illumina Infinium assay (Teodoridis et al., 2005). In addition, as shown in Table 1.7, a variety of controls have been used on the HumanMethylation27 BeadChip to monitor multiple steps in the Infinium assay.

Influence of copy number changes

Three local techniques are influenced by copy number changes to various extents. MeDIP is particularly prone to measurement error due to the relative enrichment of gene-depleted genomic regions, because it is obvious in the unmethylated version of a sequence. Enzyme-based methods that rely on the balanced enrichment of methylated and unmethylated versions of a sequence, such as MMASS, are less prone to error than the methods that involve the unbalanced enrichment of methylated or unmethylated regions. Furthermore, bisulfite-based methods are less susceptible to copy number alterations.

Genomic coverage

In terms of genomic coverage and resolution, the comparison of three technologies with varying DNA fragment size and non-uniform distribution of the regions is complicated depending on the distribution of recognition sites of affinity targets/restriction enzyme, and composition of the assays.

Agilent CpG island 244k microarray targets 27,800 CGIs of about 20000 genes. In average, each gene has 13 probes covering promoter region, inside and/or downstream of the genes, while the Human methylation27 BeadChip targets only single CpG site within/out

Parameters	Microarray			
	HCG112K	Agilent CpG Island 244K	Illumina Methylation27	Agilent custom-designed 15k
Species	Human	Human	Human	Human
Method	Restriction enzyme/MeDIP	Restriction enzyme/MeDIP	Bisulfite conversion	Restriction enzyme/MeDIP
Platform	N/A(Non-commercial)	Agilent (commercial)	Illumina Infinium (commercial)	Agilent (commercial)
Format	1x12K	1x244K	12x27K	8x15k
Reproducibility	N/A	0.90	0.99	0.92
Internal quality controls	Mitochondrial controls, Arabadopsis controls and other controls ¹	Normalisation/unmethylation controls, negative controls	Staining, hybridisation, target removal, extension, bisulfite conversion, specificity, negative, non-polymorphic controls	Mitochondrial controls, unmethylated controls, negative controls and methylated controls
Content	Selected genome-wide	Genome-wide (CpG islands)	Genome-wide	Promoter regions (± 2 kb of TSS) of selected genes (about 620 genes)
Average coverage	1-120 CpG sites/probe, 1-5 probes/gene	1-12 CpG sites/probe, 13 probes/gene in average	single CpG site/probe ² , 2-8 probes/gene, 2 probes in average	2-13 CpG sites/probe, 35 probes/gene in average
Within CGI	Not necessary	Yes (with 95nt flanking region)	Not necessary	Yes (with 150nt flanking region)
Probe length	306mer ³	45-60mer	50mer	60mer
Unique sequence (%)	8525 (69.9%)	196921 (98.8%)	25762 (93.4%)	45174 (75.5%)
Repetitive elements (RepeatMask)	0	221 (0.1%)	3663 (13.3%) ⁴ , 2717 (9.9%) ⁵	807 (1.3%)
MseI recognition site	NA	16367 (8.2%)	N/A	1500 (2.5%)
Influence of CNA	Restriction enzyme (++) MeDIP (+++)	Restriction enzyme (++) MeDIP (+++)	+	Restriction enzyme (++) MeDIP (+++)
Sample throughput	1 sample/array	1 sample/array	12 samples/array	8 samples/array
Replicated probes	Various	None	25-30	None ⁶
Value	Log ratio	Log ratio	Beta value	Log ratio
Application	Discovery	Discovery	Discovery	Discovery/Validation
Annotation (Bioconductor package)	Poor (hcg12k.db)	Relatively good (N/A)	Relatively good (IlluminaHumanMethylation27.db)	Very good (N/A)
Cost	£20/sample		About £100/sample	About £180/sample
DNA amount		0.8-1µg	<1µg	

Table 1.7: Features of 5 types of microarrays combined with local techniques. ⁻¹Stratagene's SpotReport™ Alien™ cDNA Array Validation System. ⁻²50mer probes on the BeadChip cover the adjacent CpG sites, assuming the methylation status of the adjacent CpG sites is the same as that in the targeted CpG site. ⁻³Length of the probes are various on HCG112K microarray, ranging from 21 to over 1000bps. 306bp is the median size of the probe. ⁻⁴Probes is overlapped with the repetitive elements in at least 10bp. ⁻⁵Repeats occur at the targeted CpG site. ⁻⁶Majority of the probes have no replicates. ⁻⁷ β value represents the ratio of the intensity of the methylated bead type to the combined locus intensity: $\beta = \text{Max}(M, 0) / [\text{Max}(M, 0) + \text{Max}(U, 0) + 100]$ and reflects the methylation status of a specific CpG site. ⁻⁸The amount of DNA is required by restriction enzyme based method.

of CGI close to the transcription start site (TSS) of about 14,000 RefSeq genes. Clearly, the coverage of Methylation27 BeadChip is not satisfactory since each gene has only two CpG sites targeted on in average, while 8 CpG sites is maximum in the selected cancer genes. There is no external information about the design and selection of the beads on the BeadChip. Although studies have supported the hypothesis that the adjacent CpG sites have similar methylation status (Eckhardt et al., 2006), there is no concrete evidence that the measurement of methylation in two CpG sites within a CGI is sufficient to estimate the methylation level of the whole CGI sufficiently. To overcome the poor coverage of the microarrays currently available, our Agilent 15K custom microarray was designed to cover promoter CGIs in about 620 genes. In average, each gene is covered by 22 probes. Later, this coverage was extended from promoter CGIs to the regions from -2kb to +8kb to TSS of the genes on the 60K custom-designed microarray. The average number of the probes for each gene increased to 35. The coverage of five microarrays is illustrated in Figure 1.2 for two genes DVL1 and FZD4, which are both involved in Wnt signalling transduction. The BED files containing the custom tracks of microarrays that can be uploaded into UCSC genome browser to visualise the genomic locations of the probes are in Appendix 1.1 to 1.6.

Probe quality and repetitive elements

An appropriately designed microarray can improve the accuracy of the three local techniques. In principle, the selected probes should have a specific affinity to the targets with minimum bias introduced by polymorphisms and repeats. In our previous study, cross-hybridisation has been identified as a severe problem on the HCGI12K DNA methylation microarray (Dai et al., 2008). On this microarray, about 40% of probes target multiple sequences, in contrast, only about 2% of such probes are found on the Agilent 244K CGI

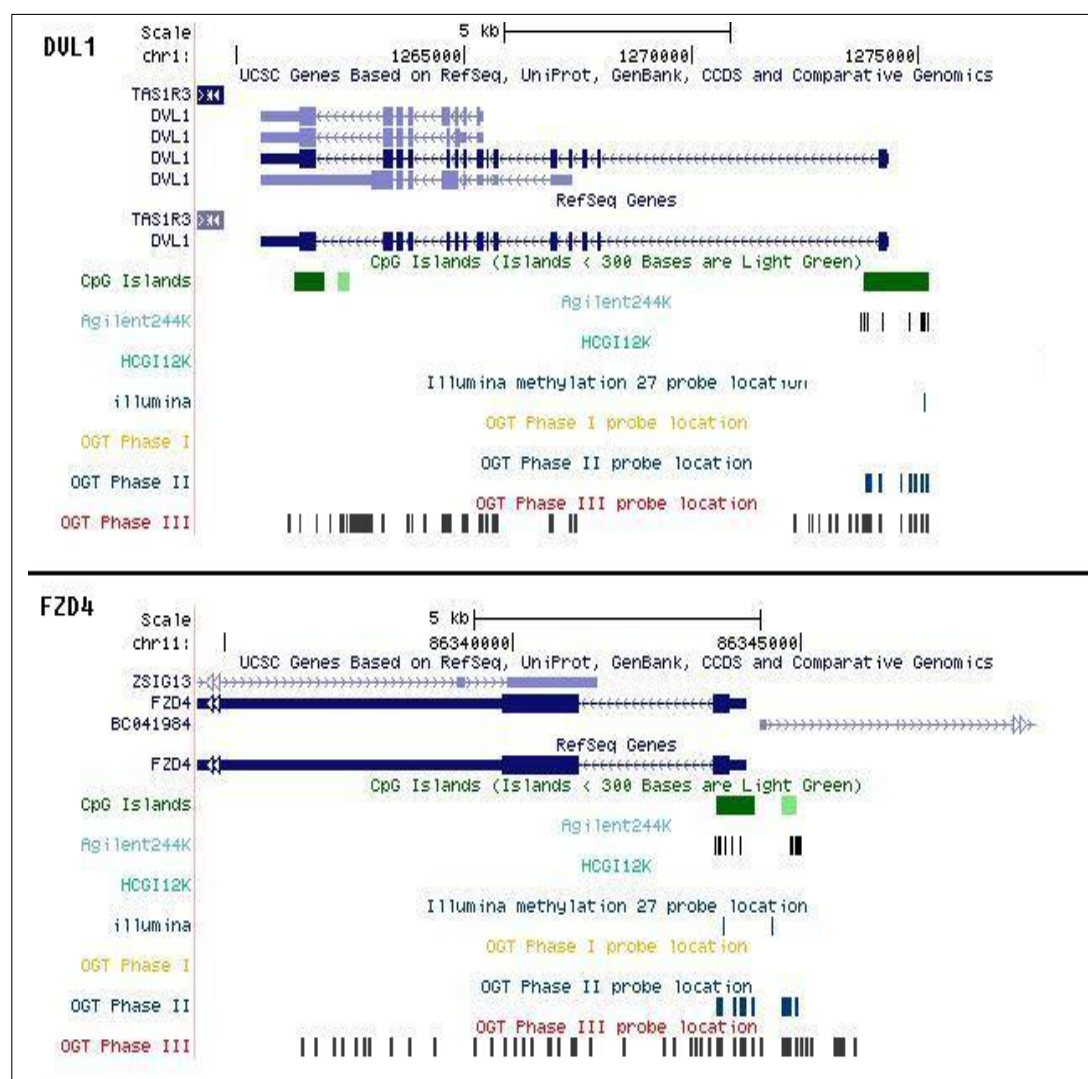


Figure 1.2: Coverage of five types of microarrays at gene DVL1 and FZD4. Green block: CGI. Dark blue line: Refseq gene. Light blue line: UCSC known gene. OGT phase I and II indicate the probes designed on Agilent 15K microarray associated with acquired drug resistance and key signalling pathways, respectively; OGT phase III indicates the probes designed on Agilent 60K microarray with expand the coverage of key signalling pathways. Agilent 244k and custom-designed 15K microarrays cover the CGIs only. Illumina Human-Methylation27 BeadChip has poor coverage and is not always targeting the CGI. No probes are found on HCGI12K arrays close to these two genes. Agilent custom-designed 60K microarray covers +2 to -8kb of TSS of the genes and the alternative transcripts without promoter CGI are also covered by this array. The figure was generated in UCSC genome browser by uploading the custom tracks of the microarrays.

microarray, 7% on the HumanMethylation27 BeadChip, and 4% on the Agilent custom-designed 15K microarray. The intensive coverage of the Agilent custom-designed 60K microarray might reduce the specificity of the array, as 25% of its probes are not unique. The sequences that contain repetitive elements, such as long interspersed nuclear elements (LINE) and short interspersed nuclear elements (SINE), could also influence the specificity of the probes leading to the low accuracy of methylation measurement. On the Illumina HumanMethylation27 BeadChip, about 10% of the probes map to the repetitive elements at the targeted CpG sites, while repetitive elements have only a very small influence on the rest of the arrays, especially in the Agilent 244K CpG island microarray which only has 0.1% such probes.

In addition, if the MseI restriction enzyme is used to fragmentise the genomic DNA, the existence of MseI recognition site (T/TAA) within the sequences targeted by the probes can also influence the affinity of the probes (see Chapter 3). The Agilent custom-designed 60K microarray has the lowest percentage of such probes (about 2%), while the Agilent 244K CpG island and custom-designed 15K arrays associated with acquired drug resistance or key signalling pathways have about 8% and 10% of such probes, respectively.

Annotation of the microarrays

The complete and accurate annotation of microarrays is of importance in order to enable the interpretation of the methylation analysis results. Unlike gene expression microarrays that only target unique genes (though sometimes with multiple transcripts), the CGIs/regions targeted by DNA methylation microarrays are sometimes associated with the regulation of more than one gene, normally two genes that are transcribed in two directions as shown in Figure 1.3. The single-gene annotation can be misleading in the down-stream data analysis

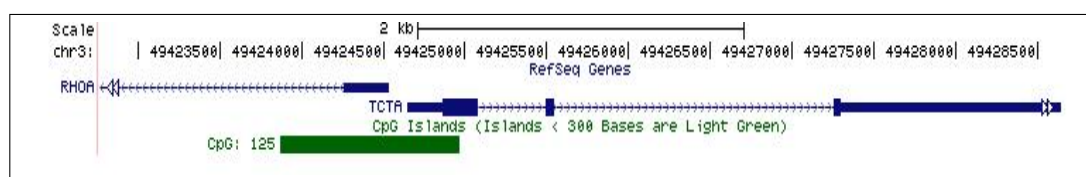


Figure 1.3: An illustration of a CpG island that is possibly regulated the expressions of two genes that are transcribed in two directions. Green block: CpG island. Arrow: direction of transcription. The figure was generated in UCSC genome browser.

such as the functional analysis since the annotation is incomplete. In addition, inconsistent definitions of CGIs used on different DNA methylation microarrays causes the difficulties in comparing the results across platforms. For example, the Agilent 244K CpG island microarray uses the definition for CGIs proposed by Gardiner Garden (1987) (see Section 1.2), while the Illumina HumanMethylation27 uses the definition published by Takai and Jones (2002), in which the following more stringent criteria were used: length over 500bps, CG content larger than 55% and the ratio of expected to observed CpG over 0.65. This definition could exclude the most of Alu-repetitive elements with an average CG frequency 52% and a CpG expected/observed ratio 0.62, as well as some small CGIs not located in promoter regions (Wang and Leung, 2004).

1.5 Bioinformatics Challenges in DNA Methylation Profiling Study

The application of bioinformatic approaches plays an important role in several aspects of DNA methylation profiling studies in cancer, as follows: (1) identifying DNA methylation profiling targets in silico, such as CGIs, depending on the purpose of DNA methylation analysis; (2) improving the quality of selected probes in microarray designs, such as the OGT Oligome databaseTM; (3) enhancing the DNA methylation analysis pipeline, including reduction of image and scanning artefacts, background correction, the exclusion of problematic probes, batch and array normalisation, and adjustment for GC content and

CpG density (especially in MeDIP); (4) improving the annotation of DNA methylation microarrays and simplifying the use of annotations, such as generating annotation packages that can be used in R (<http://www.r-project.org/>); and (5) integrating various sources of information with methylation data, such as expression profiling dataset, copy number changes and other epigenetic information like histone modifications and methylation patterns, to achieve a comprehensive view of the epigenetic and genetic changes in tumour initiation and progression.

As a result of the lack of the coverage of the Illumina BeadChip at potential regions of interest, and because MeDIP requires high quality DNA and introduces a CpG density bias, these two methods have not been used in this study. Instead, DMH was applied on a customized microarray that can measure the DNA methylation of targeted regions with multiple probes and only a moderate requirement for DNA quality. However, bioinformatic and statistical issues in terms of DNA methylation analysis remained while using this approach. The potential solutions have been evaluated and implemented in this study as described in the next section.

1.6 Aims of This Project and Outline of Thesis

Given the crucial roles of methylation at promoter CGIs in ovarian cancer development and progression, as well as the association of methylation with chemosensitivity in ovarian cancer, the major aim of this study was to identify the potential loci/CGIs that can be used as DNA methylation prognostic/predictive biomarkers that would be of clinical use in the management of ovarian cancer patients. Meanwhile, this project aimed to gain a further understanding of the biological mechanisms of ovarian tumour progression and the development of drug resistance. To achieve this goal, there were four primary sub-

aims of this PhD project that were relevant to bioinformatics and statistical issues in DNA methylation profiling studies of ovarian tumour samples using a DMH assay.

1. The development of an algorithm tailored for DNA methylation profiling studies based on DMH to identify differential methylation associated with acquired drug resistance: Methylation Linear Discriminant Analysis (MLDA). The principle and implementation of this algorithm is described and discussed in **Chapter 2**.
2. The design of custom-designed microarrays with intensive coverage of the targeted CGIs associated with acquired drug resistance and/or crucial signalling pathways that have been implemented in ovarian tumour development and progression. The target selection of the microarrays is detailed in **Chapter 3** and **Chapter 4**, respectively.
3. A feasibility and quality assurance study of using DMH with customised microarray to identify differential methylation associated with acquired drug resistance in ovarian cancer cell lines. **Chapter 3** gives the details about this feasibility study, and shows use of DMH on the custom-designed microarray with appropriate quality controls and data pre-processing method can identify differential methylation in ovarian cancer cell lines with high accuracy.
4. The evaluation of the prognostic values of 10 important signalling pathways/families in ovarian cancer by DMH on Agilent custom-designed 15K microarray and the cross-platform validation of those DNA methylation prognostic biomarkers by DMH on an Agilent custom-designed 60K microarray. The significant prognostic value of the Wnt pathway is addressed in **Chapter 4** and the prognostic values of the other nine key signalling pathways/families are discussed in **Chapter 5**.

A general discussion and future prospective for methylation analysis and the potential use of DNA methylation as biomarkers for ovarian cancer are given in **Chapter 6**.

Chapter 2

Methylation Linear Discriminant

Analysis (MLDA): An Algorithm

Tailored for DNA Methylation Profiling

Dataset

2.1 Introduction

DNA methylation frequently occurs in mammalian DNA at the 5 position of cytosine in CpG dinucleotides. It has been estimated that over 70% of cytosines of CpG dinucleotides are methylated in the human genome. CpG dinucleotides are under-represented in the genome and methylated CpG dinucleotides predominantly occur within repetitive elements (Cooper and Krawczak, 1989). However, there are CpG rich regions of the genome which generally remain unmethylated (Bird, 1986). These CpG rich regions are known as CpG islands and are frequently located in the promoter or the first exon regions of approximately 60% of all genes (Antequera and Bird, 1993). The unmethylated status of CpG islands is thought to be a prerequisite state to maintain the linked gene in an active transcribed and transcriptional permissive state (Illingworth and Bird 2009).

Differential Methylation Hybridisation (DMH) is one of several techniques for examining CpG island methylation at a genome-wide scale that has been applied to identify aberrantly methylated gene promoters in various cancers (Huang et al., 1999; Yan et al., 2001; Wei et al., 2002; Yan et al., 2002a,b,c; Rahmatpanah et al., 2006; Zighelboim et al., 2007) (see also Chapter 1, section 1.4). Nouzova et al. (2004) modified the original method by using digestion with a methylation-dependent enzyme, McrBC. This enzyme cleaves methylated CpG DNA sequences containing two half-sites of the G/A^mC generally separated by about 55-103bp but can be up to 3kb (Sutherland et al., 1992). Within-sample comparison is applied after competitive hybridisation with McrBC digested DNA and undigested (mock digested) DNA labelled with Cy3 (green) and Cy5 (red). If a locus is unmethylated the signal intensities of Cy3 and Cy5 are equivalent, while if methylated the Cy5/Cy3 (undigested/digested) ratio is greater than one. However, no common reference is generally used in the modified DMH method, and the unequal representation of methylated and

unmethylated sequences due to competitive hybridisation may reduce sensitivity and specificity to detect differential methylation.

Currently, Significance Analysis of Microarrays (SAM) (Tusher et al., 2001) and Prediction Analysis for Microarrays (PAM) (Tibshirani et al., 2002) are commonly applied in DNA methylation analysis. Based on the change in hybridisation relative to the standard deviation of repeated measurements, SAM assigns each gene a score that is an extension of the t-statistic. For significant genes with a score over a certain threshold, SAM uses permutations to estimate the false discovery rate (FDR). It has been implemented in many studies of gene expression data (Cheishvili et al., 2007; Goold et al., 2007; Lapointe et al., 2007; Lau et al., 2007; Flavell et al., 2008) as well as DMH data, e.g. Wei et al. (2006) applied SAM to detect differential methylation of CpG island loci between ovarian cancer patient groups with short and long progression-free survival (PFS). However, using a modified t statistic, SAM assumes that the microarray data conform to approximate normality and symmetry, leading to the loss of power in the analysis of DMH data that are inherently skewed (Figure 2.1). This is especially relevant to the analysis of methylation in cancer where many CpG islands that are generally unmethylated in normal cells become hypermethylated. This asymmetric change is different from the assumptions made in the analysis of changes of gene expression in cancer, where the numbers of genes up-regulated and down-regulated are similar.

In the modified DMH method, the ratios of raw signal intensities (undigested/digested) greater than 1 reflect the various methylation levels (Nouzova et al., 2004). A ratio cut-off is generally used to identify the hypermethylated loci (Wei et al., 2002). However, this is an arbitrary value and does not necessarily accurately reflect the various sources of varia-

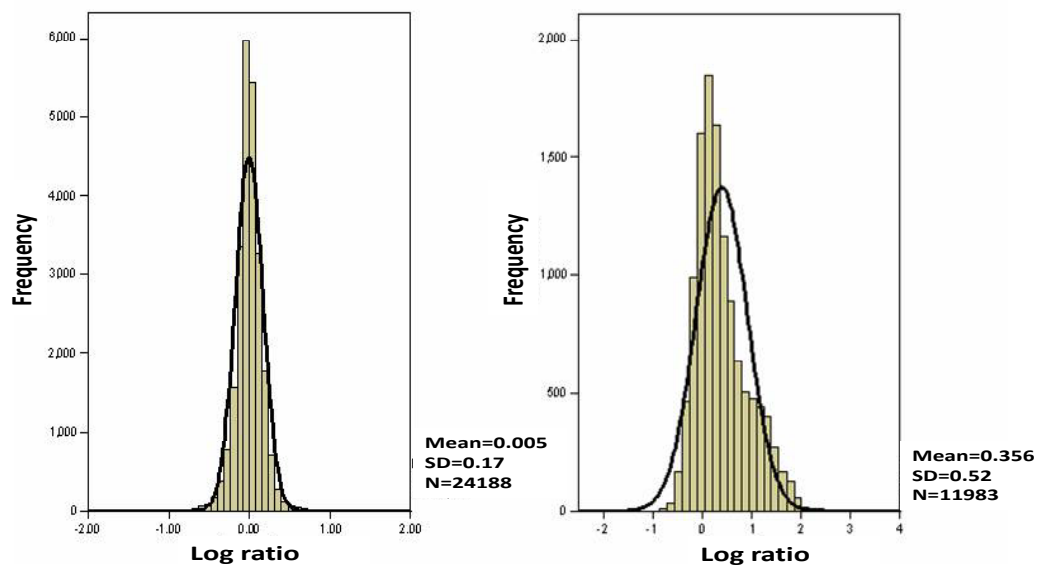


Figure 2.1: Distribution of log-transformed ratio of gene expression data in breast cancer and DMH data in A2780 cell line. The left histogram shows the distribution of log-transformed ratios (cy3/cy5) in gene expression profiling data from a previous study of breast cancer (van 't Veer et al., 2002) which is symmetric, while the right histogram shows the log-transformed ratios (undigested/digested) of DMH data from the present study which is skewed

tion in the experiment. It is therefore desirable to develop an algorithm to more objectively assess the methylation status of loci from DMH data.

PAM is a nearest centroid shrinkage method that identifies those genes that discriminate best between classes. This technique shrinks the class gene centroid towards the overall centroid by a "threshold" amount after standardising each gene by its within class standard deviation. The "threshold" is identified by cross-validation. This approach was applied in the study by Wei et al. (2006) and showed certain power in the identification of differentially methylated loci, but PAM is designed for class prediction rather than class comparison. Although the class predictor used in PAM can reflect the difference between classes, a large number of loci actually differentially methylated between the classes are excluded to improve the accuracy of prediction.

Although normalisation has become a standard procedure for the study of microarray data and is necessary for SAM and PAM analysis, unbalanced shifts in methylation status between class samples in DMH limit the use of between-class normalisation which assumes the changes are roughly symmetric. Thus, the differential methylation can be masked by the over-correction of normalisation and it would be preferable to use a method of analysis that does not require normalisation of the data.

Since PAM and SAM may have limitations for analysing DMH data, an alternative approach based on the specific features and known biological properties of the arrays used for DMH analysis is required. Therefore, the aims of the study in this chapter are (1) to identify the sequences that can be used as unmethylated controls to establish a methylation base line (unmethylated status) instead of using an arbitrary DMH ratio cutoff without con-

sidering the variations introduced in the DMH assay; (2) to develop an algorithm tailored for data from DMH assays to identify differential methylation between two groups of samples without between-group normalisation; (3) to develop an R package for this algorithm (named as Methylation Linear Discriminant Analysis, MLDA), that is publicly available at the Comprehensive R Archive Network (CRAN) (<http://cran.r-project.org/>); (4) to implement the MLDA package to identify a set of loci differentially methylated between ovarian cancer cell lines sensitive and resistant to cisplatin analysed by DMH using HCGI12K arrays (<http://data.microarrays.ca/cpg/>) and Agilent custom-designed 15K CpG island arrays (see Chapter 3).

2.2 Methods

Cell lines were grown in RPMI-1640 supplemented with glutamine (2mM) and 10% fetal bovine serum in humidified 5% CO₂ in air incubators at 37°C. Genomic DNA was extracted using Qiagen DNA extraction kits (Qiagen, UK) following manufacture's instruction.

All signal intensities from the McrBC digested and mock digested samples were extracted by GenePix Pro 6.0. A multiplicative background correction was applied to correct signal intensities for the background noise in each array. After background correction, the log-transformed signals from digested and mock digested samples show three approximately parallel linear patterns (Figure 2.2a). The first pattern (digested/undigested is close to 1) represents the unmethylated sequences. The second pattern is assumed to represent either hemi-methylated (mono-allelically methylated) sequences or the unmethylated sequences cross-hybridised with the methylated ones on the panel. This pattern could also be due to cell heterogeneity in the cell line population, but since the cell lines used in this study were clonal in origin, this is unlikely. The third pattern is assumed to represent bi-allelically

methyated sequences in the target DNA. Based on this observation an algorithm Methylation Linear Discriminant Analysis (MLDA) using three univariate linear regression models was developed in R (version 2.7.0).

MLDA is detailed in section 2.2.1 to 2.2.3. In brief, the first model is constructed by the log-transformed signal intensities of unmethylated features and used as the reference for unmethylation (Figure 2.2b). The second one is the intermediate model constructed through the point corresponding to the 97.5-quantile of the residuals below the first linear regression line (Figure 2.2c). The features with a standardised residual less than 2 from this intermediate model are used to generate the third model which is used as the reference for methylation (Figure 2.2d). The log likelihood ratio of a locus being methylated is then proportional to the difference between the squared standardised residual from the methylated line and that from the unmethylated line (see Section 2.2.1). The log likelihood threshold of zero then provides a more rational basis for distinguishing between methylated and unmethylated loci than a robust undigested/digested ratio of 1.5, as it takes into account the observed variability in the experiment. Then, the consistency and inconsistency scores, ranging from 0 (no consistency) to 2 (complete agreement), of log likelihood ratios on dye-swapped/duplicate arrays are used to determine methylation and unmethylation cut-offs, which keep the consistency score (CR) relatively high (about 1.4) and the inconsistency score (IR) low (about 0.01). Each locus is assigned a score based on the cut-offs using the weighted methylation scoring scheme (Figure 2.3). The feature consistently identified as methylated candidates on dye-swapped/duplicate arrays are scored as 1; similarly unmethylated features are scored as -1; the rest of the feature are assigned a weighted score corresponding to their location on the plot of log-likelihood ratios (Figure 2.4) (see Section 2.2.2). The averaged score for each locus is calculated in each sample class (e.g. resistant or sensitive) and plotted against each other. A robust regression model is then fitted

to these data. The standardised residuals from the robust regression model are assumed to follow a normal distribution. The outliers of the standardised residuals are identified as the differentially methylated loci between the class samples (see Section 2.2.3, an example in Figure 2.7).

2.2.1 Log-likelihood ratio transformation

1. An univariate linear regression model was constructed for the unmethylated probes (e.g. mitochondrial derived features) using formula 2.1 where α is the intercept, β is the slope of the model, and ζ is the error representing the unpredicted or unexplained variation in the model (Figure 2.2b). The parameters of regression line were estimated by the method of least squares (formula 2.2 and 2.3).

$$G_i = \alpha + \beta R_i + \zeta_i \quad i = 1, 2, 3 \dots k \quad (2.1)$$

$$\hat{\beta} = \frac{\sum (R_i - \bar{R}) (G_i - \bar{G})}{\sum (R_i - \bar{R})^2} \quad (2.2)$$

$$\hat{\alpha} = \hat{\beta} \bar{R} - \bar{G} \quad (2.3)$$

k is the number of unmethylated controls on DMH array. G_i and R_i are the logarithmic-transformed digested and undigested intensities of microarray probes for mitochondrial sequences, respectively. $\bar{G}$ and $\bar{R}$ are the averaged logarithmic-transformed undigested and digested intensities of the k unmethylated controls.

2. The scale estimate σ_{mito} associated with the error term in the linear regression model was estimated from the residuals from the observed points to the fitted line. The most extreme 10% of residuals was omitted from either end of the distribution to minimise the impact of extreme residuals on this estimate.

3. The standardised residuals (SR_{mito}) of all the microarray probes to the unmethylation regression line were calculated as formula 2.4.

$$SR_{mito} = \frac{residuals_{mito}}{\sigma_{mito}} \quad (2.4)$$

4. The point corresponding to the 97.5-quantiles residual below the unmethylation line is represented as X (R.975, G.975). The intermediate linear model (Figure 2.2c) was constructed through point X with a slope assumed to be 1 and the intercept estimated as formula 2.5.

$$\hat{\alpha} = G_{.975} - R_{.975} + 1.96\sigma_{mito} \quad (2.5)$$

5. The standardised residuals ($SR_{.975}$) of all the microarray probes to the line with slope 1 and intercept estimated from formula 2.5 were calculated as formula 2.6. The variance of the residuals to the intermediate model was assumed to be similar as that in the unmethylated model established in step (1).

$$SR_{.975} = \frac{residuals_{.975}}{\sigma_{mito}} \quad (2.6)$$

6. The microarray probes with standardised residuals less than 2 were included for later robust regression analysis. The line estimated from this regression analysis represents the methylation regression line (Figure 2.2d).
7. The scale estimate σ_{meth} of the methylation regression line was estimated using only those microarray probes below the line, with the most extreme 5% removed.
8. The standardised residuals (SR_{meth}) of all the microarray probes to the methylated

regression line were calculated as formula 2.7. The log likelihood ratio (LR) of all the microarray probes was estimated by formula 2.8 for further analysis.

$$SR_{meth} = \frac{residuals_{meth}}{\sigma_{meth}} \quad (2.7)$$

$$LR = SR_{mito}^2 - SR_{meth}^2 \quad (2.8)$$

2.2.2 Determination of log likelihood ratio cut-offs

Two inconsistency scores (IRmeth and IRunmeth) and two consistency scores (CRmeth and CRunmeth) between dye-swap/duplicate arrays were used to determine the log like likelihood ratio threshold. IRmeth (formula 2.9) represents the rate of the microarray probes identified as methylated in one array but as unmethylated in the other one, while IRunmeth (formula 2.10) is the rate of the microarray probes identified as unmethylated in one array but as methylated in the other one. CRmeth (formula 2.11) and CRunmeth (formula 2.12) are the rates for the spots identified as methylated (CRmeth) and unmethylated (CRunmeth) in both dye-swap/duplicate arrays (Figure 2.3). Therefore, the consistency/inconsistency score is within a range from 0 (no consistency) to 2 (complete agreement).

Results from 2 nd duplicate array	Results from 1 st duplicate array		
	unmethylated	uncertain	methylated
unmethylated	a	b	c
uncertain	d	e	f
methylated	g	h	i

$$IR_{meth} = \frac{c}{c + f + i} + \frac{g}{g + h + i} \quad (2.9)$$

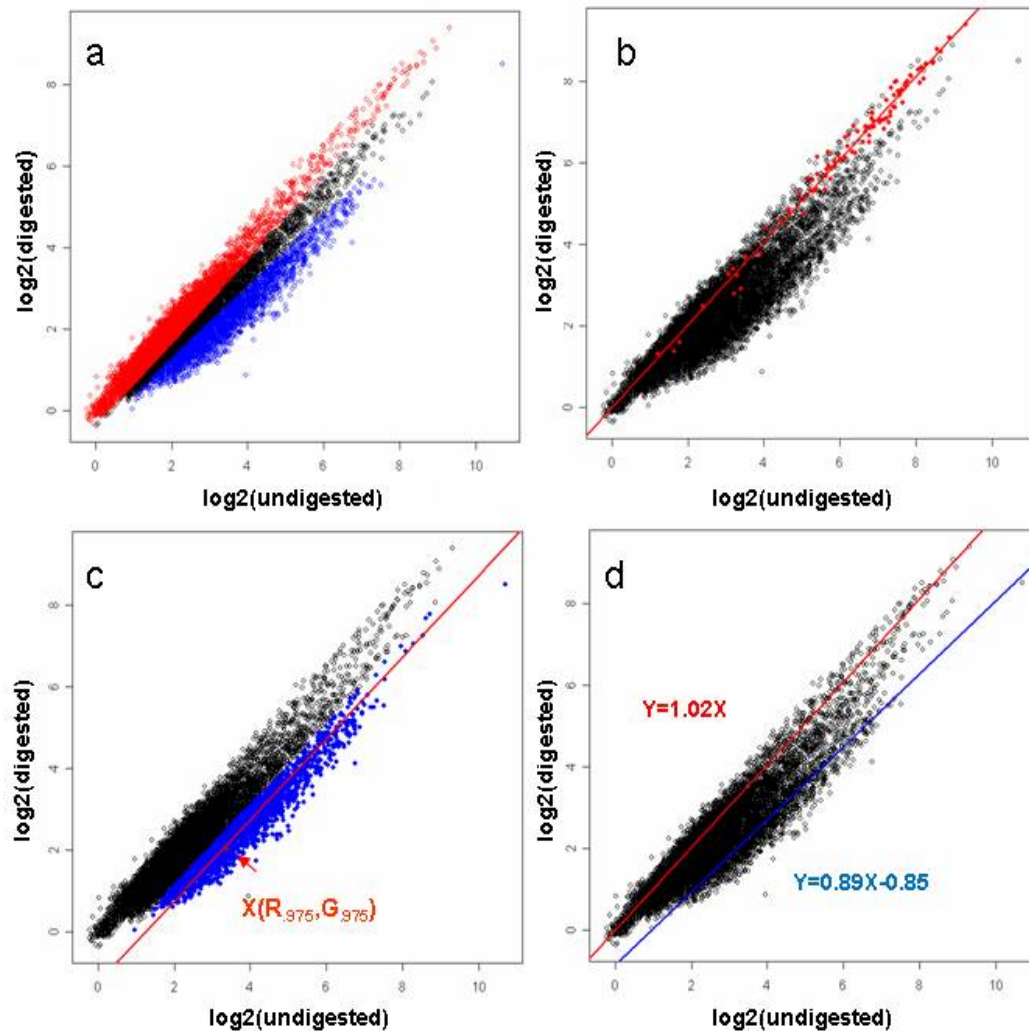


Figure 2.2: An illustration of unmethylated and methylated model construction in MLDA in A2780 cell line. a: Three patterns can be observed on the scatter plot of log-transformed Cy3 (undigested) against log-transformed Cy5 (digested) intensities. b: The unmethylated model constructed using 94 mitochondrial sequences as a unmethylation reference. c: The intermediate model constructed through the 97.5 quantile residual. The point X is the 97.5 quantile residual. The microarray probes in blue (standardised residual to the intermediate model is less than 2) are selected to construct the methylated model. d: Methylated (in blue) and unmethylated (in red) models in A2780 cell line.

$$IR_{unmeth} = \frac{g}{a + d + g} + \frac{c}{a + b + c} \quad (2.10)$$

$$CR_{meth} = \frac{i}{c + f + i} + \frac{i}{g + h + i} \quad (2.11)$$

$$CR_{unmeth} = \frac{a}{a + b + c} + \frac{a}{a + d + g} \quad (2.12)$$

The log likelihood ratio thresholds (LR_{meth} and LR_{unmeth}) for methylated and unmethylated microarray probes, which kept the inconsistency score low (at or close to 0.01) and the consistency score high (at or close to 1.4), were used as the cut-offs for methylated and unmethylated loci. IR tends to rise with the increase of CR slowly, but starts to increase dramatically when the CR goes above 1.4, at which point the inconsistency score is generally about 0.01. We have therefore used CR > 1.4 and IR < 0.01 as the criteria for determining the methylation cut-offs.

2.2.3 Identification of robust regression outliers

Each microarray probe was scored based on the cut-offs of likelihood ratios for methylation and unmethylation on dye-swap arrays using the weighted methylation scoring scheme shown in Figure 2.4. The microarray probes consistently identified as methylated candidates on dye-swap arrays were scored of 1; similarly unmethylated microarray probes were scored of -1. The rest of the microarray probes were assigned a weighted score based on their location on the plot.

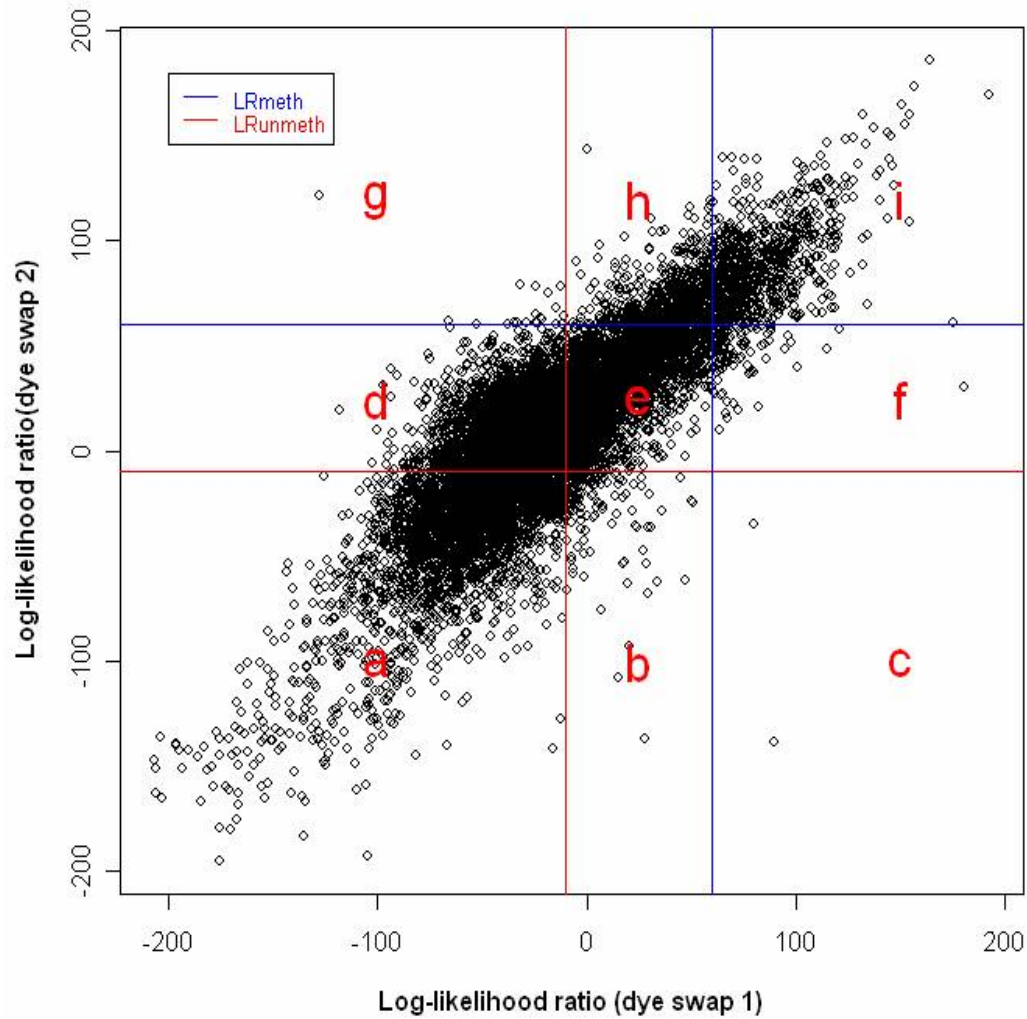


Figure 2.3: Determination of methylation and unmethylation cut-offs of likelihood ratios on dye-swapped arrays. LRmeth: log likelihood ratio cut-off for methylated spots; LRunmeth: log likelihood ratio cut-off for unmethylated spots.

A robust regression model (Tukey, 1977) was constructed with the averaged scores in one class of samples as the explanatory variable, and the corresponding scores in the other class of samples as the dependent variable. The degree of trimming was determined according to Barnett et al. (Wilcox, 2005a) when estimating the variance of residuals to the robust linear regression model. It was assumed that the standardised residuals (SRs) from the robust regression line followed a normal distribution $N(\mu, \sigma^2)$. μ and σ were estimated excluding outliers using the MAD-Median Rule (Wilcox, 2005b). The p value for each SR cut-off was calculated as described by Simon et al (2003). This p-value reflects the probability of observing a group of more extreme residuals from the fitted normal distribution. Microarray probes were identified as outliers if their SRs were larger than the cut-off for which the p-value was less than 0.01 (see an example in Figure 2.7).

2.2.4 Estimation of misclassification rate

The misclassification rate was estimated by drawing bootstrap samples 500 times with replacement from the two classes (sensitive and resistant) and carrying out hierarchical clustering based on the loci identified as differentially methylated using weighted scores for MLDA and log ratios without between-group normalisation for SAM and PAM, respectively. Clustering was carried out using Euclidean distance as the distance metric, and clusters were agglomerated using the average linkage criterion. The clustering tree was cut into two groups and the number of misclassified cell lines was counted. The misclassification rate was obtained from the averaged number of misclassified samples in 500 bootstraps divided by the total number of samples.

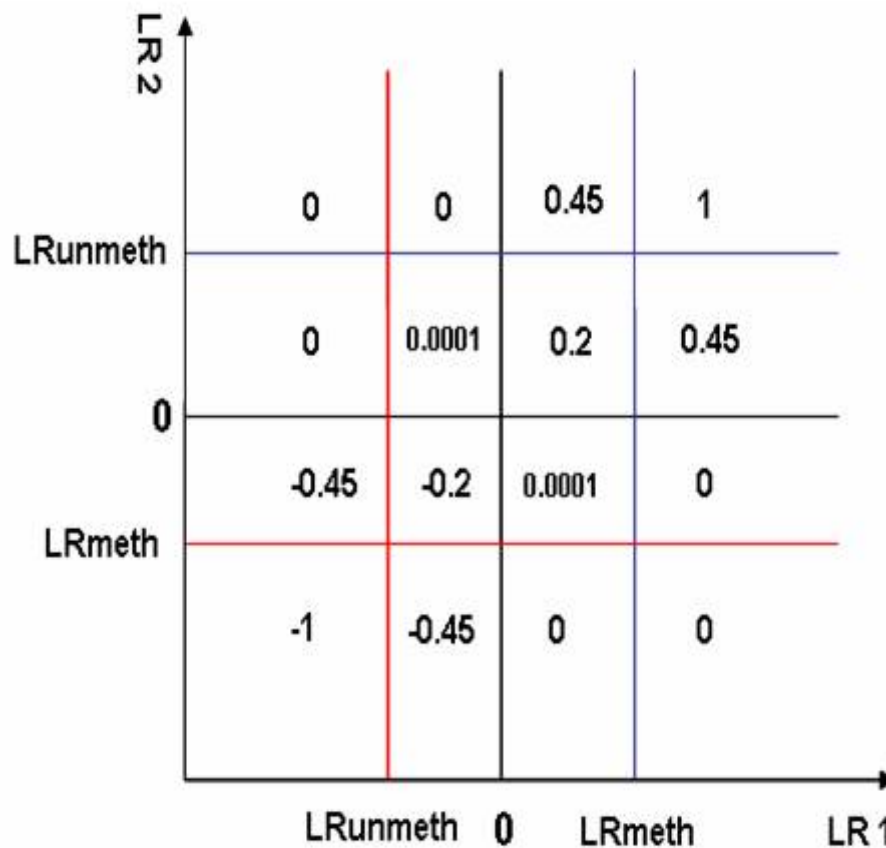


Figure 2.4: Weighted scoring scheme. The microarray probes consistently identified as methylated candidates on dye-swap arrays were scored 1; similarly unmethylated microarray probes were scored -1. The rest of the microarray probes were assigned a weighted score based on their location on the plot. LRmeth: log likelihood ratio cut-off for methylated loci; LRunmeth: log likelihood ratio cut-off for unmethylated loci. LR: log likelihood ratio on dye-swapped arrays.

2.2.5 SAM and PAM analysis

The raw signal intensities of each channel were subtracted by the median signal intensities of corresponding channel of controls on HCGI12K array. After this correction, SAM in samr package [samr(data, resp.type = "Two class unpaired", testStatistic = "standard", nperms = 100, center.arrays = FALSE)] and PAM in pamr package with default arguments were applied in the unpaired two groups comparison using logarithmic-transformed DMH ratios (digested/undigested) in R version 2.7.0. Between-group normalisation was not used in SAM and PAM to avoid masking the true differential methylation by over-correction. False discovery rate (FDR) less than 5% was used as the cut point to indentify the loci significantly differentially methylated between two groups.

2.2.6 Making MLDA package under Windows

The desktop used to construct the package has Microsoft Window XP (service pack 3) operating system with Intel Core Duo CPU E6859, and 2.99 GHz, 1.96 GB of memory. The components in Windows required to make MLDA package are (1) A minimal set of Unix utilities, so called "Rtools" (<http://www.murdoch-sutherland.com/Rtools/>), (2) ActivePerl 5.10.0 Building 1003 (<http://www.activestate.com/activeperl/>), (3) the Microsoft html help compiler (<http://msdn.microsoft.com/en-us/library/ms669985.aspx>) and (4) MikTex 2.7 (<http://www.miktex.org/>). The process of package construction followed the manual "Writing R Extension" (<http://www.r-project.org/>). Briefly, it has 7 steps: (1) create the directory tree using package.skeleton(); (2) modify the DESCRIPTION FILE and NAMESPACE file; (3) check R functions; (4) create and debug documentation files; (5) check the package using /textitR CMD check package test; (6) Build the package using the command *R CMD build -binary package*; (7) upload to CRAN (<ftp://cran.r-project.org/>)

incoming/).

2.3 Results

In this study, we have developed a novel approach, named MLDA, for analysing CpG island microarray hybridisation data that allows the identification of differentially methylated loci without normalisation. MLDA was programmed in R (version 2.7.0) and the package is available at CRAN (<http://cran.r-project.org/>).

2.3.1 DMH dataset on HCGI12K arrays

MLDA was applied to identify the CGIs differentially methylated from DMH data on HCGI12K arrays derived from sensitive A2780 derivatives (A2780, A2780p3, A2780p5, A2780p6, A2780p13, A2780p14) and isogenically matched, cisplatin resistant lines ((Anthony et al., 1996)) derived by multiple exposures to cytotoxic levels of cisplatin and which are 2-5 fold resistant to cisplatin in clonogenic assays (A2780cp70, A2780/MCP1, A2780/MCP2, A2780/MCP3, A2780/MCP4, A2780/MCP5, A2780/MCP6, A2780/MCP7, A2780/MCP8, A2780/MCP9). After background correction, the log-transformed digested and undigested intensities of the 13056 microarray probes show three approximately parallel linear patterns (Figure 2.2a). The first pattern (digested/undigested is close to 1) represents the unmethylated sequences. The second pattern represents either hemi-methylated sequences or the unmethylated sequences cross-hybridised with the methylated ones on the panel. The third pattern represents the methylated sequences in target DNA. The methylated and unmethylated loci in target DNA can be characterised by a linear regression model for each pattern. As previously mentioned, normalisation may not be appropriate for DMH data, so the log ratios of signal intensities in two classes of samples are not at the same level (Figure 2.5). Normalisation is not required for MLDA as the determination of the

methylation score is based on the data within each experiment.

Mitochondrial DNA is unmethylated (Maekawa et al., 2004), therefore, the signal intensities of both channels of microarray probes for mitochondrial sequences are expected to be equal. However, a bi-modal distribution is observed in the log-transformed fluorescence ratios (digested/undigested) of 121 mitochondrial sequences, representing the unmethylated mitochondrial sequences and the mitochondrial sequences assumed to be cross-hybridised with other methylated sequences on the panel. Thus, we selected 94 of 121 mitochondrial sequences that were consistently unmethylated through all the cell lines and used them as the unmethylation reference in target DNA.

The parameters of unmethylated and methylated models in all 16 cell lines were estimated (Table 2.1). The slope of the unmethylated regression line constructed by 94 mitochondrial sequences is indeed close to 1. After computing the log-likelihood ratios, the methylation and unmethylation cut-offs and associated IRs and CRs were determined from the dye-swapped array pairs (details in Method section). As shown in Figure 2.6, IR tends to rise with the increase of CR slowly, but starts to increase dramatically when the CR goes above 1.4, at which point IR is generally about 0.01. We have therefore used $CR > 1.4$ and $IR < 0.01$ as the criteria for determining the methylation and unmethylation cut-offs. Each locus was scored using the weighted scoring scheme based on those cut-offs. The averaged scores in 6 cisplatin-sensitive cell lines and 10 cisplatin-resistant cell lines were used to construct a robust regression model. Figure 2.7a shows that the standardised residuals ($\frac{residual}{\sigma}$) from the robust regression model roughly follow a normal distribution. The positive and negative outliers are determined as described in Method section.

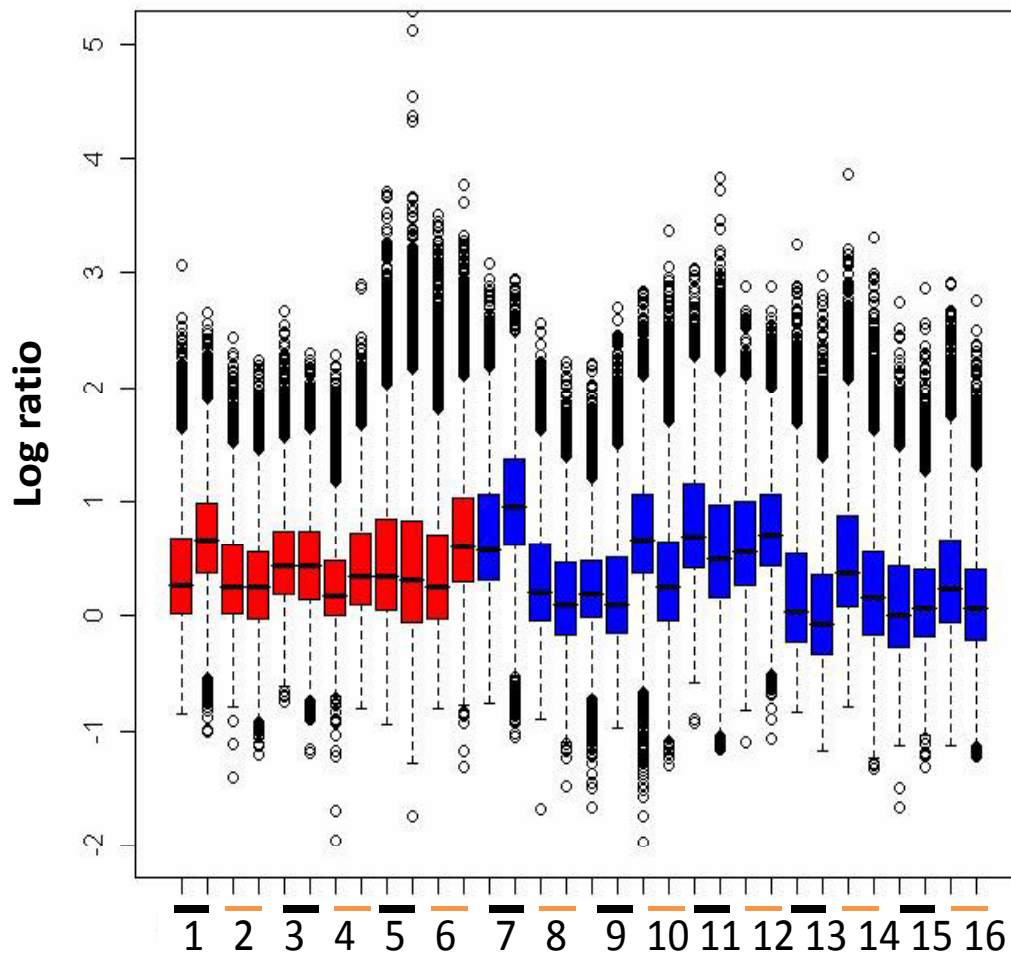


Figure 2.5: Box plot of log ratios of undigested signal intensities against digested signal intensities in 16 cell lines (dye-swapped arrays). The boxes in red are the A2780 sensitive cell lines; in blue are the A2780 resistant cell lines. Each cell line has duplicates. From 1 to 16, the cell lines are A2780 (1), A2780p3(2), A2780p5(3), A2780p6(4), A2780p13(5), A2780p14(6), cp70(7), MCP1(8), MCP2(9), MCP3(10), MCP4(11), MCP5(12), MCP6(13), MCP7(14), MCP8(15) and MCP9(16). As normalisation is not applied, the centre and scale of log ratios for the 16 cell lines are not at the similar level.

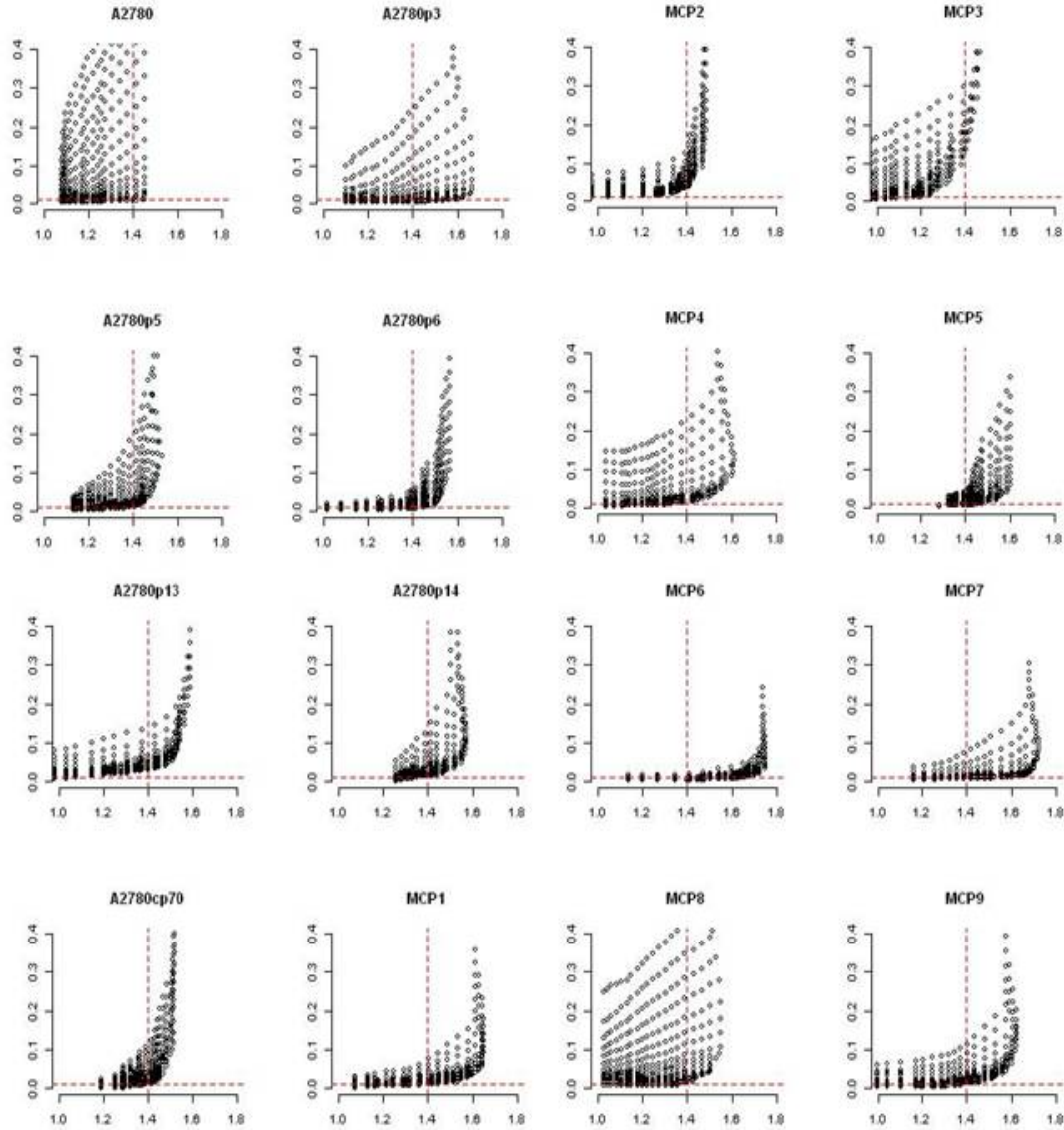


Figure 2.6: CR against IR in 16 cell lines. X axis is the consistency score (CR) and y axis is the inconsistency score (IR). IR tends to rise with the increase of CR slowly, but starts to increase dramatically when the CR goes above 1.4, at which point the inconsistency rate is generally about 0.01. Not all cell lines could reach this point e.g. MCP3

Unmethylated linear regression model								
cell line	Intercept (α)	slope (β)	σ	R^2	Interceptds (σ)	slopedds(β)	σ_{ds}	R^2_{ds}
A2780	-0.0003	1.0122	0.1727	0.9829	0.0005	1.0574	0.1897	0.9780
A2780p3	-0.0003	1.0343	0.1212	0.9897	0.0018	1.1065	0.1425	0.9882
A2780p5	-0.0003	1.0138	0.1684	0.9840	-0.0002	1.0728	0.1425	0.9883
A2780p6	0.0002	0.9914	0.1605	0.9778	-0.0001	1.0012	0.1638	0.9747
A2780p13	-0.0012	1.0240	0.1628	0.9786	-0.0005	1.0744	0.1436	0.9852
A2780p14	-0.0022	1.0499	0.1523	0.9809	-0.0009	1.0340	0.2069	0.9691
A2780/cp70	0.0013	0.9604	0.2532	0.9524	-0.0002	1.0119	0.2402	0.9479
MCP1	0.0002	0.9946	0.1450	0.9803	-0.0023	1.1120	0.1452	0.9836
MCP2	0.0000	0.9980	0.1370	0.9727	-0.0028	1.0475	0.1719	0.9653
MCP3	0.0004	0.9932	0.2253	0.9183	-0.0023	1.0517	0.2795	0.8978
MCP4	0.0006	0.9838	0.1739	0.9718	-0.0028	1.0770	0.1947	0.9751
MCP5	0.0009	0.9857	0.2464	0.9639	-0.0008	1.0170	0.2166	0.9692
MCP6	-0.0022	1.0352	0.1220	0.9751	-0.0068	1.1283	0.1540	0.9752
MCP7	-0.0005	1.0079	0.1379	0.9791	-0.0045	1.1529	0.1588	0.9764
MCP8	-0.0028	1.0578	0.1903	0.9431	-0.0068	1.1193	0.1885	0.9575
MCP9	-0.0017	1.0331	0.1834	0.9614	-0.0091	1.1538	0.1691	0.9674
Methylated linear regression model								
cell line	Intercept (α)	slope (β)	σ	R^2	Interceptds (σ)	slopedds(β)	σ_{ds}	R^2_{ds}
A2780	-0.8839	0.8917	0.1491	0.9438	-0.8055	0.9086	0.1706	0.9260
A2780p3	-1.1672	0.9797	0.1518	0.9553	-0.7414	0.9774	0.1790	0.9438
A2780p5	-0.8991	0.8978	0.1515	0.9476	-0.9246	0.9766	0.1673	0.9523
A2780p6	-0.9455	0.9562	0.1838	0.9378	-1.1995	0.9641	0.1788	0.9324
A2780p13	-1.8918	0.9807	0.2535	0.8962	-1.8049	0.9652	0.2936	0.8512
A2780p14	-1.5637	0.9142	0.2549	0.8857	-1.4468	0.9066	0.2128	0.8837
A2780/cp70	-1.0317	0.8501	0.1581	0.9115	-1.3074	0.8967	0.1541	0.9265
MCP1	-1.1990	0.9781	0.1692	0.9467	-1.0935	1.0384	0.1775	0.9525
MCP2	-0.8037	0.9292	0.1486	0.9557	-0.9739	0.9381	0.2176	0.8848
MCP3	-1.1244	0.9151	0.1755	0.9482	-0.9303	0.9205	0.2599	0.8098
MCP4	-1.4326	0.9171	0.1418	0.9660	-1.6205	0.9610	0.2348	0.8323
MCP5	-1.1187	0.9425	0.1839	0.9404	-1.2007	0.9546	0.1757	0.9295
MCP6	-1.2460	0.9250	0.1966	0.9294	-1.2182	0.9826	0.2248	0.8989
MCP7	-1.8972	0.9909	0.1977	0.9442	-1.4894	1.0139	0.2458	0.8886
MCP8	-1.0219	0.9905	0.1975	0.9468	-0.4735	0.9421	0.2280	0.8761
MCP9	-1.3399	0.9837	0.2073	0.9352	-1.1497	1.0078	0.1967	0.9115

Table 2.1: Parameters of linear models in MLDA for 16 cell lines in DMH dataset. ds:dye swap; σ : standard deviation; R^2 :coefficient of determination

Finally, 115 loci were identified as candidates differentially methylated between A2780 sensitive and these resistant cell lines (Appendix 2.1). Noticeably, 113 of 115 loci ($p = 8.8 \times 10^{-3}$, outlier detection test (Simon et al., 2003)) were hypermethylated, but only 2 loci ($p < 1.0 \times 10^{-3}$, outlier detection test) lost methylation in the resistant cell lines (Figure 2.7b). This is consistent with the unbalanced shift in DMH data and indicates cisplatin treatment of cells selects preferentially for hypermethylation of loci, rather than hypomethylation in these tumour cells.

To compare the results from MLDA, SAM and PAM, we analysed the DMH dataset by all three methods. MLDA identified 115 loci (113 hypermethylated and 2 hypomethylated loci, misclassification error < 0.001) (for details of 115 loci, see Appendix 2.1), SAM identified 152 loci (149 hypermethylated and 3 hypomethylated loci, misclassification error = 0.227, $FDR = 6.17 \times 10^{-3}$), and PAM found 24 hypermethylated loci (misclassification error = 0.084, $FDR < 0.001$) in the resistant cell lines. The lists of loci identified by MLDA, SAM and PAM are detailed in Appendix 2.2. Twenty-four loci identified by all three methods are listed in Table 2.2.

2.3.2 Validation of differential methylation

To confirm the differential methylation of loci identified in this study, we experimentally tested the methylation of 26 loci by methylation-specific PCR (MSP) and/or pyrosequencing of bisulphite modified DNA (Herman et al., 1996) in sensitive A2780 derivatives and cisplatin resistant derivatives. MSP and bisulfite pyrosequencing were performed by Dr. Janet Graham, Dr. Constanze Zeller and Dr. Jens M Teodoridis in Imperial College London. Twenty-three out of the 26 loci have been confirmed as differentially methylated. It should be noted that MSP and pyrosequencing only examine methylation at a limited num-

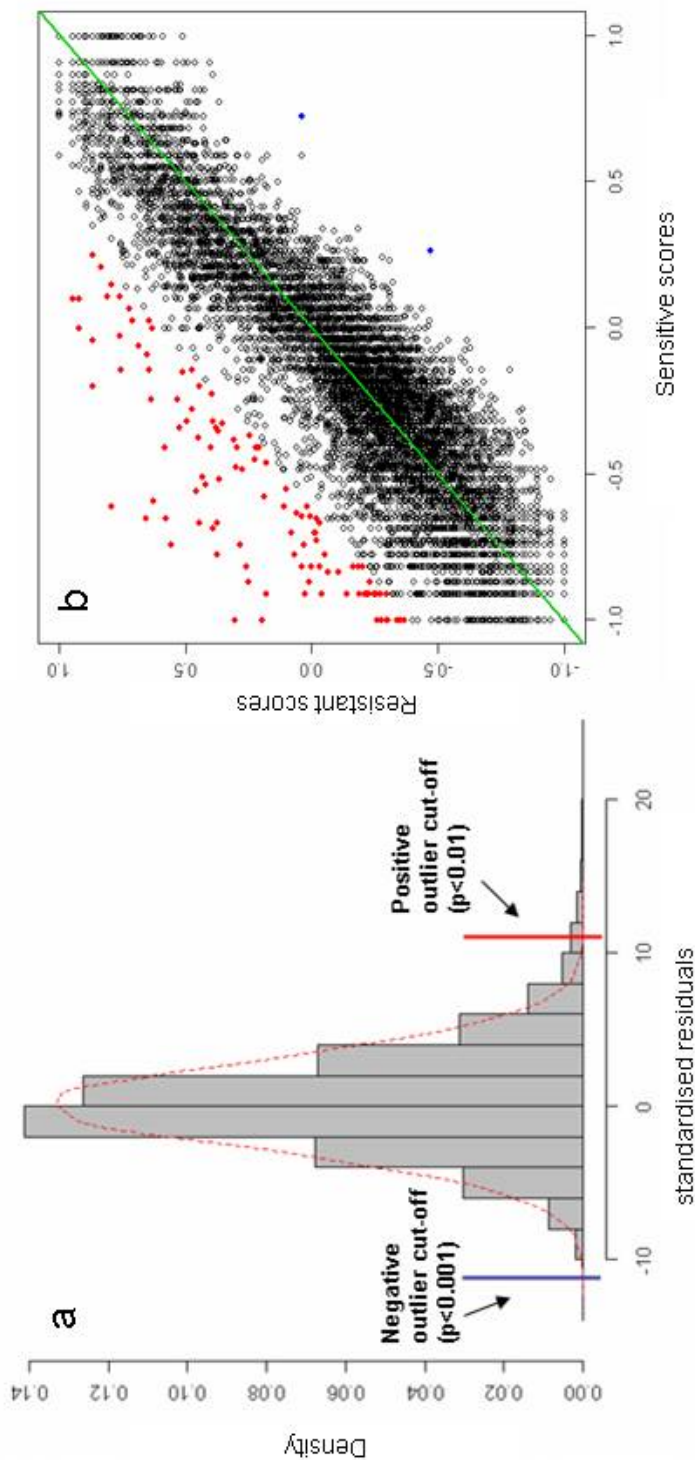


Figure 2.7: Outliers identifications. a: Distribution of the observed (histogram) standardised residuals and the theoretical distribution based on the fitted model (dashed smooth line in red). The red and blue solid line are the positive and negative cut-offs, respectively. b: Scatter plot of sensitive scores against resistant scores in A2780 series cell lines. The hypermethylated loci are colored in red and hypomethylated loci are in blue. The robust regression model is $Y = 0.9956X + 0.009$

ber of CpG sites of the sequence present on the DMH analysis. It is possible that the loci which were not confirmed as differentially methylated are methylated at other CpG sites which are detected by DMH but not targeted by MSP and/or pyrosequencing primers and so 23 out of 26 loci confirmed as differentially methylated is a minimum estimation.

2.3.3 Implementation of MLDA on Agilent 15k custom-designed CpG island microarray

MLDA was also implemented in DMH assay on another CpG island microarray, which was custom-designed for targeting the loci associated with acquired drug resistance in ovarian cancer (see Chapter 3), to identify the loci differentially methylated between A2780 cisplatin-sensitive derivatives (A2780p3, A2780p5 and A2780p6) and A2780 cisplatin-resistant derivatives (CP70, MCP1 and MCP6) (Chapter 3). Unmethylated controls selected from mitochondrial sequences on HCGI12K array (yellow spots on Figure 2.9a) and from chromosome 16 that are underrepresented for CG regions and contain very few M_{cr}BC recognition sites (ACG or GCG) (blue spots on Figure 2.9a) were used to construct the unmethylation linear model in MLDA. Out of 26 loci tested by bisulfite pyrosequencing/MSP, 9 loci that show consistent differential methylation between these three A2780 sensitive and resistant derivatives were selected as a test set, and 9/9 (100%) were identified by MLDA (see 'validation' in Table 2.3). This indicates the high sensitivity of MLDA using methylation scores of individual probes on the DMH arrays. Meanwhile, the performance of both SAM and PAM using the corrected DMH ratios by the method implemented in MLDA as described in section 2.3.4 are as good as MLDA at least at these 9 loci on the custom-designed microarray in terms of sensitivity (see 'validation' in Table 2-3), indicating the corrected DMH ratios could reflect the differential methylation between groups.

microarray ID	status	validation	MLDA rank	PAM rank	SAM rank	CGI	gene symbol	GenBank Accession	Chromosome
66_G_6	hypermethylated	Yes	1	1	1	Yes			
121_D_9	hypermethylated	Yes	2	5	6	Yes	CRABP1	NM_004378	15
39_E_1	hypermethylated	ND	3	2	2	No			
122_D_9	hypermethylated	No	4	11	11	Yes	SOX12	NM_006943	20
123_D_9	hypermethylated	No	5	10	10	Yes	SOX12	NM_006943	20
51_H_8	hypermethylated	Yes	6	18	19	No	FEZF2	NM_018008	3
58_A_1	hypermethylated	ND	7	22	22	Yes			
80_H_5	hypermethylated	ND	8	14	16	No			
21_A_11	hypermethylated	Yes	9	17	18	Yes	NTN4	NM_021229	12
38_D_7	hypermethylated	Yes	11	23	24	Yes	AGBL2	NM_024783	11
40_E_1	hypermethylated	ND	12	19	17	No			
18_A_7	hypermethylated	ND	13	20	20	No	EDIL3	NM_005711	5
55_F_8	hypermethylated	Yes	14	9	9	Yes	BC127881	BC127881	7
122_B_1	hypermethylated	ND	15	16	14	No			
109_A_6	hypermethylated	No	18	12	13	Yes			
41_D_9	hypermethylated	Yes	22	8	8	Yes	WNT1	NM_005430	12
42_D_9	hypermethylated	ND	23	4	4	No			
119_A_6	hypermethylated	Yes	24	6	5	Yes	NR2E1	NM_003269	6
63_A_8	hypermethylated	ND	26	15	15	No			
6_D_4	hypermethylated	Yes	31	3	3	Yes	LMX1A	NM_177398	1
17_H_9	hypermethylated	Yes	34	13	12	Yes	HRASLS3	NM_006290	6
5_D_4	hypermethylated	Yes	35	7	7	Yes	LMX1A	NM_177398	1
24_D_3	hypermethylated	Yes	75	24	23	Yes	SP5	NM_001003845	2
122_G_1	hypermethylated	ND	101	21	21	Yes			

Table 2.2: 24 loci identified by MLDA, PAM and SAM as differentially methylated candidates in the comparison between A2780 cisplatin sensitive and cisplatin multiple-selected resistant cell lines. ND: not done. Yes: validated. No: not validated. MLDA rank*: the rank of standardised residuals to the robust regression line constructed by the averaged sensitive scores against averaged resistant scores; Gene symbol**: only the gene of which transcription start site (TSS) is within 5 kb span of the loci; CGI***: CpG island defined by Gardiner-Garden and Frommer

The loci identified by all three methods are listed in Table 2.3. A full list of loci identified by three methods is detailed in Appendix 2.3.

2.3.4 Modification of MLDA for within-array Normalisation

Normalisation is not required in MLDA since it converts the continuous DMH ratios into ordinal variables using a scoring system. However, DMH ratio is generally used in the literature which requires within-array normalisation to reduce the influence of dye bias. MLDA was modified for this purpose using the univariate linear regression model constructed by unmethylation probes. The slope of the linear regression model is adjusted to 1 and intercept adjusted to 0 in this method, since the log transformed signal intensities of the unmethylated controls are assumed to be equal. The log transformed signal intensities of entire probes on the microarray are adjusted correspondingly as shown in formula 2.13 and 2.14.

$$\left\{ \begin{array}{l} \log_2 (G)' = \log_2 (G) - b \\ \log_2 (R)' = \log_2 (R) \end{array} \right\} \quad (2.13)$$

$$\left\{ \begin{array}{l} \log_2 (G)'' = \frac{\log_2 (G)' \sin \frac{\pi}{4}}{\sin(\tan^{-1} k)} \\ \log_2 (R)'' = \frac{\log_2 (R)' \cos \frac{\pi}{4}}{\cos(\tan^{-1} k)} \end{array} \right\} \quad (2.14)$$

G and R are the background subtracted signals from the Green and Red channels, respectively. In formula 2.13 and 2.14 above, b and k are the intercept and slope respectively of the linear regression model fitted to these original values (logged to the base 2). In 2.13 these values are initially adjusted to have 0 intercept and in 2.14 they are further adjusted to have a slope of 1 by rotating the coordinate using the origin as centre as shown in Figure 2.8. The signals of the probes were transformed from Figure 2.9a to Figure 2.9b. After this

Chapter 2. Methylation Linear Discriminant Analysis (MLDA): An Algorithm Tailored for DNA Methylation Profiling Dataset

81

CGI ID	status	chro	start	end	gene	GeneBank Accession	# MLDA hit ^c	# SAM hit ^c	# PAM hit ^c	# total ^b	validation
OGT_14	up	1	59020049	59023528	JUN	NM_002228	6	8	8	45	
OGT_19	up	1	94779425	94780639	F3	NM_001993	6	6	6	19	
OGT_23	up	1	111690661	111691062	C1orf88	NM_181643	6	13	10	16	
OGT_34	up	1	163590110	163590435	LMX1A	NM_177398	4	12	9	13	MSP
OGT_35	up	1	163590815	163592952	LMX1A	NM_177398	5	5	5	34	
OGT_38	up	1	210846774	210849709	ATF3	NM_001674	4	7	4	46	
OGT_39	up	1	219117071	219117487	HLX	NM_021958	3	8	7	16	
OGT_40	up	1	219118589	219120296	HLX	NM_021958	4	9	8	25	
OGT_53	up	2	30307070	30309159	LBH	NM_030915	5	8	5	31	
OGT_55	up	2	45014931	45016562	102_D_11*		1	6	2	14	
OGT_60	up	2	95054584	95056158	MAL	NM_022438	6	5	4	16	
OGT_70	up	2	171278123	171282150	SP5	NM_001003845	2	1	1	33	BP
OGT_72	up	2	174916661	174917322	2_A_6*		4	1	3	15	
OGT_88	up	3	27731216	27731426	66_A_7*		6	8	7	11	
OGT_89	up	3	37009232	37010360	EPH2AIP1	NM_014805	8	18	18	20	
OGT_89	up	3	37009232	37010360	MLH1	NM_000249	8	18	18	20	MSP
OGT_98	up	3	57179209	57179457	IL17RD	NM_017563	4	4	4	23	
OGT_125	up	4	30330302	30333940	PCDH7	NM_032456	1	1	1	51	
OGT_126	up	4	75183520	75184143	CXCL2	NM_002089	1	2	2	15	
OGT_129	up	4	84474750	84475423	HPSE	NM_006665	8	10	7	26	
OGT_131	up	4	85722570	85723917	CDS1	NM_001263	5	4	4	20	
OGT_154	up	5	139115773	139116160	47_H_6*		1	2	2	16	
OGT_161	up	5	172130088	172132212	DUSP1	NM_004417	5	6	6	22	
OGT_169	up	6	3172024	3174687	TUBB2B	NM_178012	10	11	8	62	
OGT_169	up	6	3172024	3174687	RDBP	NM_002904	10	11	8	62	
OGT_169	up	6	3172024	3174687	SKIV2L	NM_006929	10	11	8	62	
OGT_170	up	6	3174225	3174676	TUBB2B	NM_178012	6	7	4	30	
OGT_170	up	6	3174225	3174676	RDBP	NM_002904	6	7	4	30	
OGT_170	up	6	3174225	3174676	SKIV2L	NM_006929	6	7	4	30	
OGT_187	up	6	108592364	108597232	NR2E1	NM_003269	21	20	18	66	MSP
OGT_190	up	6	138229518	138230834	TNFAIP3	NM_006290	1	2	2	19	
OGT_202	up	7	27170440	27172987	HOXA9	NM_152739	4	2	1	167	
OGT_203	up	7	27241580	27242151	15_E_8*		2	3	2	18	
OGT_222	up	8	13034461	13035285	DLC1	NM_006094	7	3	3	16	
OGT_239	up	8	96029927	96031560	TP53INP1	NM_033285	7	6	6	14	
OGT_243	up	8	121205045	121205322	COL14A1	NM_021110	6	9	7	11	
OGT_249	up	9	106701	108177	FOXO4	NM_207305	1	6	3	85	
OGT_260	up	9	71848656	71849097	MAMDC2	NM_153267	11	5	4	16	
OGT_262	up	9	89302334	89303637	DAPK1	NM_004938	1	1	1	30	
OGT_265	up	9	95747937	95748308	66_G_6*		6	10	9	13	MSP
OGT_272	up	9	125813067	125820774	LHX2	NM_004789	15	9	9	81	
OGT_278	up	10	8131380	8138335	FLJ45983	NR_024255	4	4	2	86	
OGT_278	up	10	8131380	8138335	GATA3	NM_002051	4	4	2	86	
OGT_298	up	10	102974172	102980053	LBX1	NM_006562	24	26	23	78	
OGT_314	up	11	22170748	22171896	ANO5	NM_213599	1	4	2	20	
OGT_319	up	11	27698635	27698849	BDNF	NM_001143807	3	3	3	16	
OGT_322	up	11	44282233	44283093	33_C_10 40_C_10**		4	8	7	18	
OGT_324	up	11	47693315	47693682	AGBL2	NM_024783	3	1	1	18	
OGT_329	up	11	63137719	63138672	HRASLS3	NM_006290	6	6	6	26	MSP
OGT_358	up	12	47657957	47661817	WNT1	NM_005430	8	13	12	49	MSP
OGT_374	up	12	112393270	112395064	LHX5	NM_022363	4	4	3	22	
OGT_375	up	12	112513701	112514135	35_A_11*		6	6	5	20	
OGT_meth_587	up	13	19664208	19665779	GJB2	NM_004004	3	2	1	23	
OGT_388	up	13	100866117	100867259	NALCN	NM_052867	2	6	2	27	
OGT_393	up	14	37137197	37138958	32_F_11*		2	9	6	19	BP
OGT_408	up	15	43193759	43194174	DUOX2	NM_014080	8	9	8	11	
OGT_408	up	15	43193759	43194174	DUOX2	NM_207581	8	9	8	11	
OGT_409	up	15	60243307	60244940	C2CD4B	NM_001007595	8	8	8	40	
OGT_412	up	15	76419724	76420163	CRABP1	NM_004378	0	2	1	15	
OGT_413	up	15	76420377	76421144	CRABP1	NM_004378	10	15	15	19	
OGT_448	up	17	34111804	34115223	MLLT6	NM_005937	12	9	7	33	
OGT_473	up	18	45341073	45341444	LIPG	NM_006033	7	8	6	20	
OGT_511	up	20	1731824	1732552	115_H_11 116_H_11**		10	14	11	19	
OGT_513	up	20	10600573	10603611	JAG1	NM_000214	2	2	2	33	
OGT_520	up	20	54633686	54640196	TFAP2C	NM_003222	1	9	7	78	BP
OGT_547	up	X	1615879	1616123	P2RY8	NM_178129	10	12	8	14	
OGT_551	up	X	13866071	13866982	GPM6B	NM_001001994	1	4	2	12	
OGT_287	down	10	70747952	70749383	HK1	NM_033496	2	4	4	16	
OGT_431	down	16	65143769	65144220	TK2	NM_004614	7	12	13	20	
OGT_431	down	16	65143769	65144220	CKLF	NM_016326	7	12	13	20	
OGT_534	down	22	49310474	49317634	SCO2	NM_005138	1	1	3	61	
OGT_534	down	22	49310474	49317634	NCAPH2	NM_152299	1	1	3	61	

Table 2.3: CGI differentially methylated between A2780 cisplatin sensitive and resistant derivatives identified by MLDA, SAM and PAM on Agilent custom-designed 15K microarray. Abbreviation: MSP methylation specific PCR; BP bisulfite pyrosequencing.-¹up: hypermethylated locus; down: hypomethylated locus.-²number of probes identified by the method. -³number of probes targeting the whole CGI.*CGI identified on HCGI12K array but not within 2kb span of TSS of the genes.**Two loci related to the same CGI

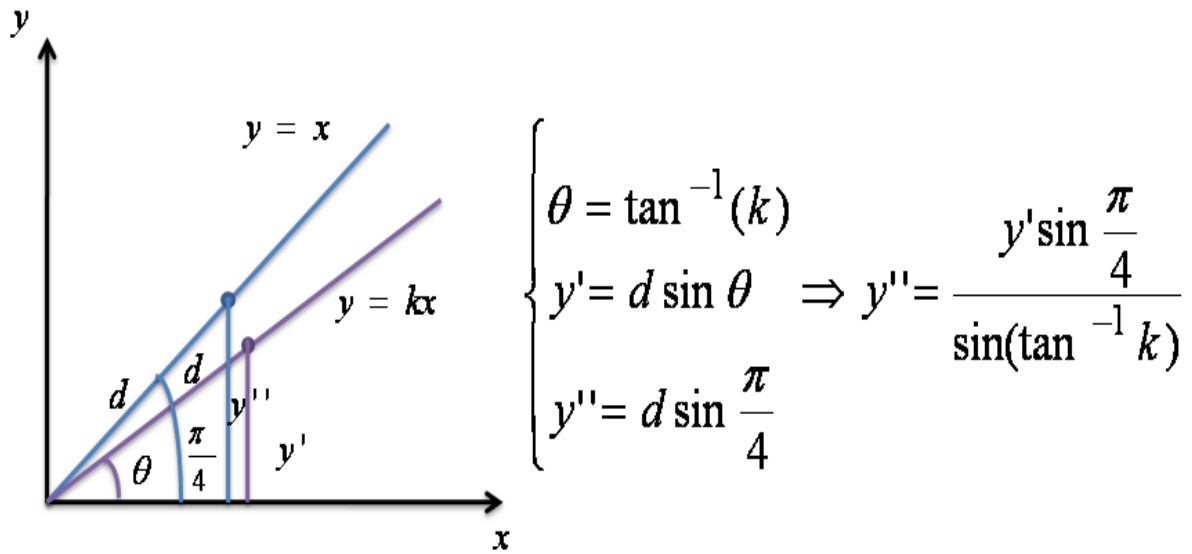


Figure 2.8: Rotation of the coordinates using the origin as the centre. After this adjustment, the log-transformed signals of the unmethylated controls from both channels were similar, and the distances of all the spots to the origin remain the same as before.

correction, the unmethylated model constructed using the corrected signals coincides with the diagonal. The median correlation of unmethylated controls in 20 dye-swap duplicates dramatically increased from 0.36 to 0.54 after this correction ($p = 2.09 \times 10^{-7}$, see Chapter 3), indicating the reduction of variations mainly introduced by dye-swap labeling. As an alternative of methylation scores, the log DMH ratios calculated from the corrected signals of McrBC digested and mock digested samples can be used in SAM and PAM analysis instead of the log ratios normalised across arrays using global method which could possibly mask the differential methylation between groups.

2.4 Discussion

Hypermethylation of promoter CpG islands is strongly correlated to transcriptional gene silencing and epigenetic maintenance of the silenced state and is a potential rich source

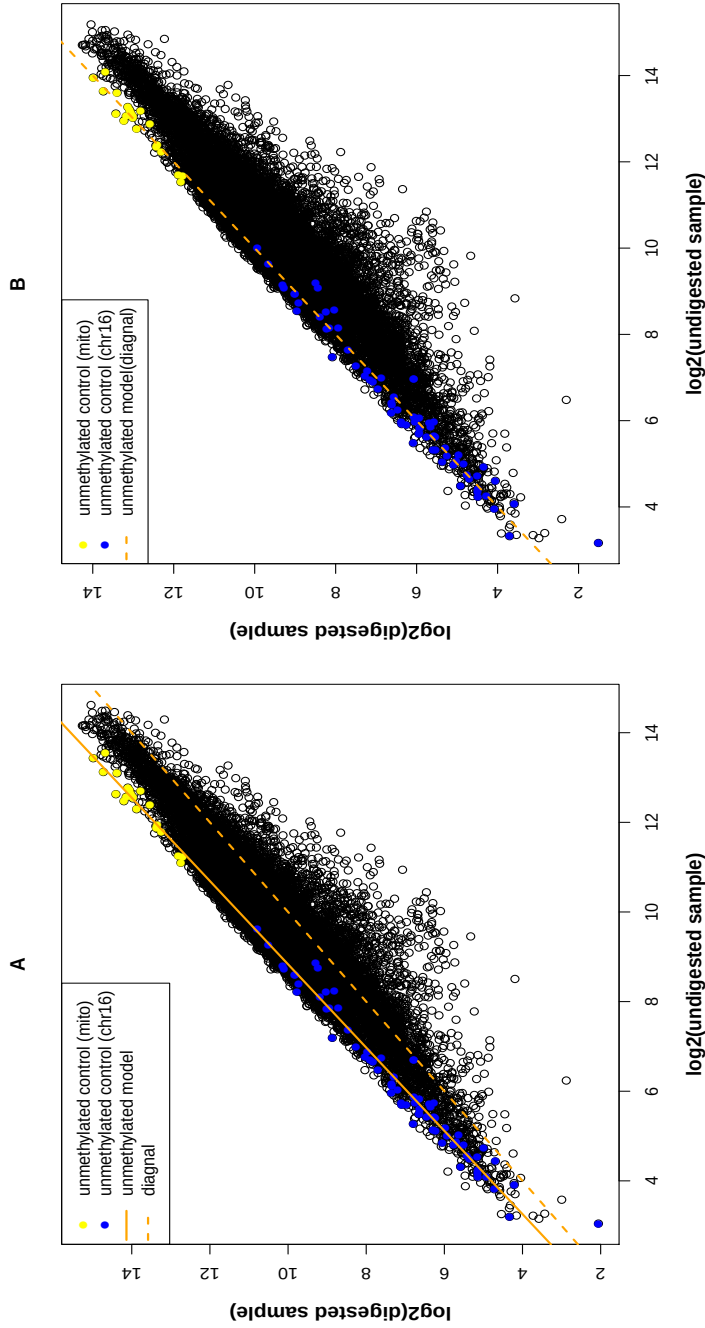


Figure 2.9: Within-array normalisation in MLDA to correct log DMH ratios. A: scatter plot of signals of McrBC digested and mock digested samples before normalisation. The unmethylated model does not overlap with the diagonal ($Y=X$) indicating the existence of dye bias. B: scatter plot of signals of McrBC digested and mock digested samples after normalisation. The unmethylated model coincided with the diagonal indicating the signals of unmethylated controls from both samples have been adjusted to be equal.

of biomarkers of cancer. Differential Methylation Hybridisation (DMH) is one technique used for genome-wide DNA methylation analysis. The study of such microarray data sets should ideally account for the specific biological features of DNA methylation and the non-symmetrical distribution of the ratios of unmethylated and methylated sequences hybridised on the array. We have therefore developed a novel algorithm tailored to this type of data, Methylation Linear Discriminant Analysis (MLDA). MLDA utilises log likelihood ratios representing the relative probability that loci are methylated instead of log ratios of signal intensities used in previous studies (Yan et al., 2001; Wei et al., 2002; Yan et al., 2002a,b,c; Wei et al., 2003). Validation of 23/26 identified loci using independent methods of methylation analysis shows that MLDA can robustly identify differentially methylated loci between ovarian cancer sensitive and resistant cell lines without requiring the data to be normalised.

2.4.1 Unmethylation controls and cross-hybridisation

Data on methylation status for 121 mitochondrial derived sequences were available in this study. Mitochondrial sequences would be expected to be unmethylated. We used 94 mitochondrial sequences to construct unmethylated linear model at the beginning of the study, and indeed, 93 of 121 mitochondrial loci were defined as unmethylated and 25 loci being of uncertain methylation status by MLDA. However, three mitochondrial loci were identified as hypermethylated candidates in the resistant ovarian carcinoma cell lines by both MLDA and SAM. One explanation of this discrepancy is that all these three loci have more than one BLAT hit indicating the existence of homology with nuclear DNA sequences, raising the possibility of hybridisation with these nuclear DNA sequences which may be differentially methylated. As shown in Figure 2.2a, the loci in the middle pattern represent either hemi-methylated sequences (only one allele is methylated) or the unmethylated

sequences cross-hybridised with the methylated ones on the panel. No specific allowance is made for these intermediate points in analysis by SAM and PAM, whereas MLDA attempts specifically to down-weight these points in the identification of the methylation regression line. By giving a lower weighted score (close to 0) (Figure 2.4) to those loci, MLDA reduces the influence of cross-hybridisation among this group of sequences. Of course cross-hybridisation may also occur in the loci in the other two patterns (methylated and unmethylated patterns), but it is not possible for any mathematical approach to identify this.

2.4.2 Limitations of the study

The selection of the weighted scores was arbitrary only based on the consistency of log likelihood ratios observed in duplicates. However, the validation result from bisulfite pyrosequencing shows 23 out of 26 loci (88%) identified using this weighted score scheme are truly differentially methylated between groups, indicating this score scheme is valid and useful, although ideally the loci identified using a series of weighted scores should be tested by bisulfite pyrosequencing and the scores that could identify differential methylation with highest sensitivity and specificity should be used in the algorithm.

Misclassification rate was estimated using weighted scores for MLDA and log ratios without between-group normalisation for SAM and PAM. Although it has been shown that MLDA has extremely small misclassification error (<0.001) compared to SAM ($=0.227$) and PAM ($=0.084$) using the training dataset described in this study, it is necessary to examine the misclassification error of these three methods in an independent DMH dataset to avoid the potential risk of over-fitting issue leading to the over-optimistic estimation of the performance of MLDA.

2.4.3 Comparison of MLDA, SAM and PAM

The misclassification error of MLDA (error<0.001) based on the methylation score is much lower than that for either SAM (error=0.227) or PAM (error=0.084) based on the log ratios, indicating the potential of MLDA methylation scores to be used as a reliable discriminator between classes of samples.

Validation by bisulfite pyrosequencing/MSP, showed that out of 26 loci randomly selected from 115 loci identified by MLDA, 23 loci have shown differential methylation between A2780 sensitive (A2780, A2780p3, A2780p5, A2780p6, A2780p13, A2780p14) and resistant derivatives (A2780cp70, A2780/MCP1, A2780/MCP2, A2780/MCP3, A2780/MCP4, A2780/MCP5, A2780/MCP6, A2780/MCP7, A2780/MCP8, A2780/MCP9) , indicating MLDA has high positive predictive value (88.5%) to identify differential methylation in DMH assay

Figures 2.10 shows the number of loci identified by MLDA, SAM and PAM, respectively as well as the number of loci that have been validated by bisulfite pyrosequencing/MSP. 35 loci were found by MLDA, but not by SAM and PAM. 2 loci randomly selected from these 35 loci have been validated by bisulfite pyrosequencing/MSP. The methylation status of 35 loci identified by MLDA only and 72 loci identified by SAM only were further examined in cisplatin sensitive cell line A2780p5, A2780p6, and cisplatin resistant cell line A2780cp70, MCP1 and MCP6 assayed by Illumina Infinium on HumanMethylation27K chips (more details about HumanMethylation27K are given in Chapter 1). Only 7 loci from MLDA and 6 loci from SAM were targeted by this array. Four out of seven (57.1%) loci from MLDA and only 2/6 (33.3%) loci from SAM showed increased methylation in at least two cisplatin resistant cell lines compared to sensitive cell lines (Figure 2.11).

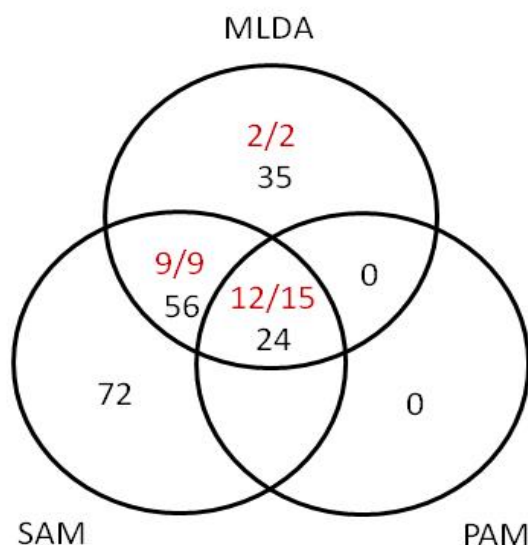


Figure 2.10: Venn diagram of the numbers of loci identified by MLDA, SAM and PAM. The numbers of loci identified by three methods are in black and the numbers of the loci that have been tested and validated by bisulfite pyrosequencing/MSP are in red

In conclusion, a novel method named MLDA for large scale DNA methylation study was developed in this study. This approach can transform the signal intensities to log-likelihood ratios through three linear regression models on two-channel DNA methylation microarrays, thus allowing to determine the methylation status of a locus based on dye-swapped/duplicate arrays. MLDA has been applied to identify the loci exhibiting differential methylation between A2780 sensitive and resistant cell lines and the high sensitivity of this method in DMH assay has been validated by bisulfite pyrosequencing and/or MSP, and showed hypermethylation, rather than hypomethylation, predominantly occurred during the acquisition of cisplatin resistance.

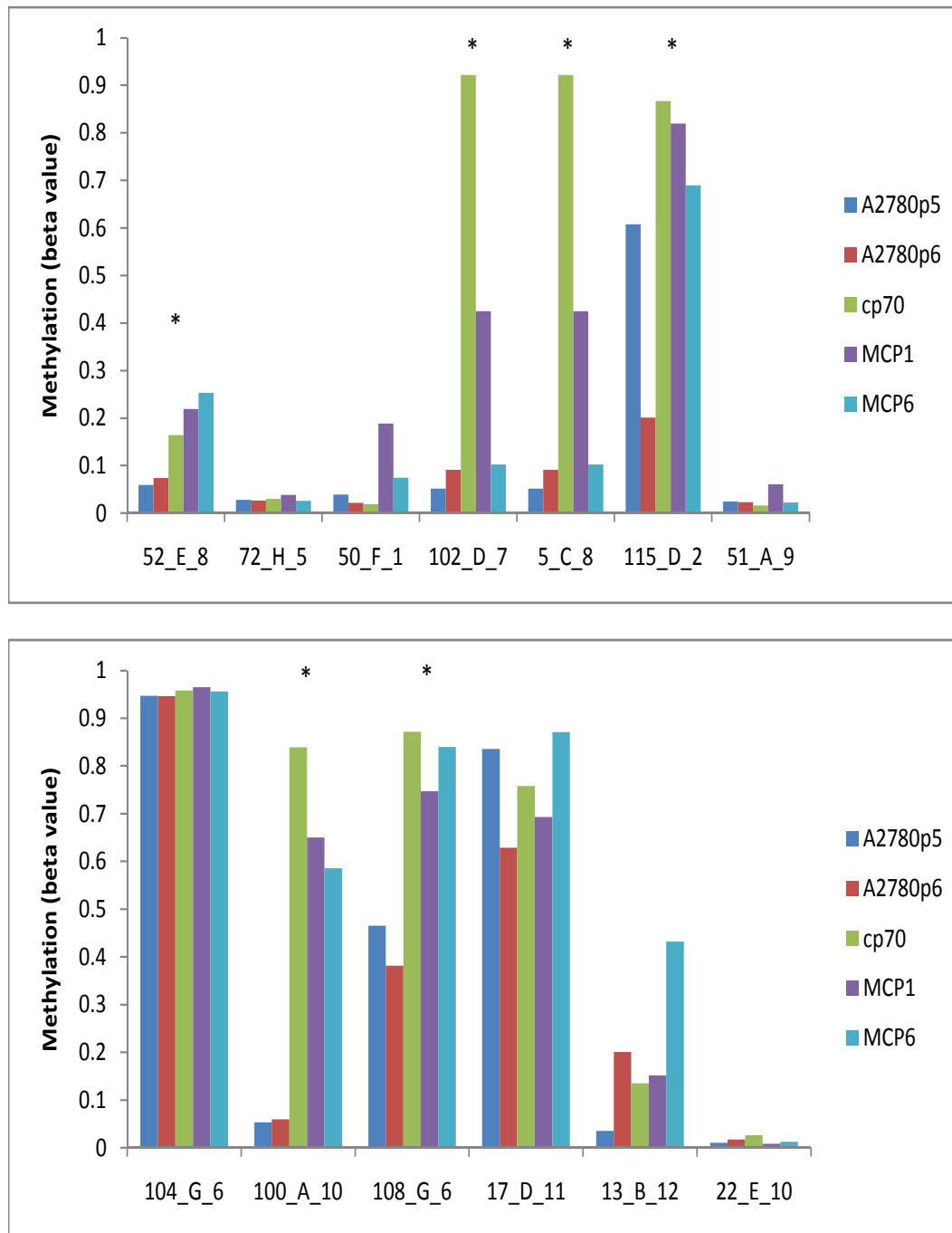


Figure 2.11: Methylation of loci identified by MLDA or SAM on Illumina HumanMethylation27 BeadChip. Methylation is represented as beta value ranging from 0 (unmethylated) to 1 (fully methylated). A: 7 loci identified by MLDA. B: 6 loci identified by SAM. *Methylation of loci increased in at least two resistant cell lines compared to sensitive ones. The significance of increased methylation (15%) was determined by the upper 95% CI estimated by bootstrapping from the null distribution of the difference among 4 biological replicates on the BeadChips.

Chapter 3

Feasibility Study of DNA Methylation Profiling on Agilent Custom-designed Microarray in Ovarian Cancer Cell Lines

3.1 Introduction

DNA methylation is an important epigenetic mechanism in the normal biological processes of X chromosome inactivation, imprinting genes and the inactivation of parasite DNA elements (Baylin and Jones, 2006). Although mitotically heritable and stable over time, changes of methylation have often been observed in cancers including ovarian cancer (Feinberg and Tycko, 2004; Barton et al., 2008; Sharma et al., 2010). Recent advance of technology have enabled quantitative assessment of DNA methylation at a genome wide level. For instance, Wei et al. (2002, 2006) obtained methylation profile on HCGI12K arrays (Cross et al., 1994) in advanced ovarian tumours to attempt to identify the methylation signatures that can distinguish patients with short and long progression-free survival. Similarly, we have obtained methylation profiling on the same array in 6 cisplatin sensitive derivatives and 10 cisplatin resistant derivatives of a human ovarian cell line and identified 115 loci associated with acquired drug resistance by Methylation Linear Regression Analysis (MLDA) (Dai et al., 2008). Teschendorff et al. (2009) analysed the methylation profile of blood cells from 148 healthy volunteers and 113 age-matched ovarian cancer patients prior to receiving chemotherapy in order to evaluate the diagnostic/risk value of methylation, Houshdaran et al. (2010) performed methylation profiling in 27 primary epithelial ovarian tumours and 15 ovarian cancer cell lines on the Illumina GoldenGate platform (Illumina, CA, USA) and showed the difference in methylation between ovarian cancer cell lines and tumours, as well as between different histological subtypes of ovarian cancer. Therefore, the methylation profile of tumours could be utilised to understand the epigenetic changes in tumourigenesis and progression better, and to discover candidate biomarkers for disease diagnosis/risk assessment, classification, prognosis and prediction of response to chemotherapy.

Currently, high-resolution methylation analysis relies on several existing technologies mainly based on (1) the use of methylation-sensitive and/or methylation-specific restriction enzyme that differentially recognise the methylated/unmethylated DNA, (2) the use of antibodies or other proteins that bind 5-methyl-cytosine (5mC) to enrich for methylated regions of DNA, or (3) the use of sodium bisulfite to convert the unmethylated cytosines to uracils but leave the methylated cytosines intact (Laird, 2010). High-throughput DNA methylation profiles can be obtained by combining these three methods either with microarrays, such as Illumina Infinium assay on HumanMethylation27k BeadChip (Illumina, CA, USA) and Agilent methylated DNA immunoprecipitation assay on Agilent 244k CpG island microarray (Agilent, UK) or with direct sequencing, for example bisulfite pyrosequencing and methylC-seq (Lister et al., 2009). Although sequencing, especially next generation sequencing, has the advantage of providing a single-base resolution DNA methylation map, microarray hybridisation has been widely used in DNA methylation profiling studies because it is more cost-effective and requires less down-stream data processing. The strengths and weaknesses of different types of methods for analysing DNA methylation combined with microarray hybridisation have been described in Chapter 1. In brief, the two major problems of the commercially available microarrays are the lack of sufficient coverage of loci of interest for the aims of a given study and low quality of annotation, leading to the wrong interpretation of the methylation analysis results. In addition, the reasons for the selection of regions targeted by commercial microarrays for large-scale DNA methylation studies are often unclear. Therefore, a custom-designed DNA methylation microarray, with well pre-defined targets based on known biological concepts or pathways, comprehensive annotation and good coverage of the probes is often the preferred option.

The objective of the work described in this Chapter is to test the feasibility of (1) cre-

ating a custom designed promoter CGI microarray targeting the genes involved in multiple cellular pathways and/or associated with acquired drug resistance in ovarian cancer, and (2) using differential methylation hybridisation (DMH)(Nouzova et al., 2004) to detect differential methylation of samples with this array. In this initial feasibility study, human ovarian cancer cell lines generated *in vitro* or *in vivo*, were used to address five major questions below.

1. What is the variability of DMH on the custom-designed microarray? As with any microarray experiment, variations can be introduced from any of the multiple steps of DMH: such as, DNA preparation, hybridisation and image scan. In particular, DMH analyses based on McrBC restriction enzyme, involves the extra steps of digestion, ligation and PCR amplification. Therefore, the estimation of DMH variability is necessary to understand a given DMH dataset and identify the loci statistically differentially methylated in cell lines and later in tumours.
2. Does a dye-swap design have any influence on the variability of DMH assays? As a result of the differential rates of incorporation of nucleotides labelled with the fluorescent dyes Cy3 (green) or Cy5 (red) into the labelling reaction and differences in relative fluorescence between the two dyes, labelling dye bias is often observed. Dye-swap design was recommended to remove the dye bias described by Ordway et al. (2006). However, Irizarry et al.(2008) showed the benefit of dye swap design did not exceed the extra cost. Therefore, the variation of dye-swap duplicates was estimated in this study to evaluate if the variation introduced by the dye swap step could possibly cover the true differential methylation between cisplatin sensitive and resistant cell lines.
3. Are the pre-selected unmethylated and methylated controls appropriate for the microarray? In previous DMH studies there has been an absence of control features

on the microarray leading to two main obstacles. The first one is that no criteria have been used to determine if the assay works appropriately to detect differences in methylation. Secondly, the performance of appropriate within-array and between-array normalisation was impossible to assess in DMH dataset pre-processing (see Chapter 2). The selection of control sequences as features on the microarray that are consistently unmethylated or methylated in samples is helpful to solve both problems. In this study, the selected sequences were evaluated by Methylation Linear Discriminant Analysis (MLDA, see Chapter 2) as unmethylated/methylated controls and were subsequently used in data pre-processing and down-stream data analysis.

4. How can the methylation level of a CpG island targeted by multiple probes to be measured? The DMH ratios of the probes targeting the same regions are not independent. A previous study has shown that the methylation levels of adjacent CpG sites are similar, especially within 50bps (Nautiyal et al., 2010), but the question of how the methylation level of multiple CpG sites within a CpG island/locus of >50bps can be summarised to reduce the number of data points has not been studied yet. An average value is often used to summarise multiple data points. Therefore, using the average DMH ratio, measured as a continuous variable, to measure the methylation level of a fragment was evaluated, as well the estimation of the optimal length of fragment for analysis in the proposed study.
5. Are the loci known to be differentially methylated in cell lines identified by DMH assay on the custom-designed microarrays? The methylation levels of a set of genes have been previously determined by bisulfite pyrosequencing (England and Pettersson, 2005) or methylation specific PCR (Herman et al., 1996) in cell lines. Using this set of genes as a test set, the DMH assay was evaluated in terms of sensitivity and specificity to detect differential methylation between cisplatin sensitive and resistant

cell lines after the four points mentioned above was taken into consideration.

3.2 Methods

3.2.1 Samples

In vitro generated cell lines

DNA was isolated from A2780 cisplatin sensitive derivatives: A2780p3, A2780p5 and A2780p6, and A2780 cisplatin resistant derivatives: A2780/CP70, A2780/MCP1 and A2780/MCP6 (Louie et al., 1985) using Qiagen DNA extraction kits (Qiagen, UK) following the manufacturer's instructions. The DNA from each cell line had dye-swap duplicates. In total, 12 samples underwent hybridisation (6 cell lines x 2=12).

In vivo generated cell line pairs

DNA was isolated from matched isogenic chemo-sensitive and resistant cell line pairs: PEO1 and. PEO4, PEO14 and PEO23, PEA1 and PEA2 (Langdon et al., 1988) using Qiagen DNA extraction kits (Qiagen, UK) following the manufacturer's instructions. Each of these cell lines had dye-swap duplicates hybridised to the arrays. To evaluate the variation of technical replicates of microarrays and DNA independent preparations in the DMH assay, independent DNA duplicates and hybridisation duplicates were made from PEO1 and PEO4. Therefore, the total number of samples for hybridisation was 20 (6 cell lines x 2 (dye-swap duplicates)+ 2 cell lines x 2 (DNA independent duplicates) + 2 cell lines x 2 (hybridisation duplicates) = 20).

3.2.2 Agilent custom-designed 8x15k microarray

Target selection

The workflow used to select the targets on the microarrays is shown in Figure 3.1. The loci targeted by the oligos on the microarray were mainly selected from our previous work and the published literature associated with acquired drug resistance: 1) CGIs identified from the DNA methylation profiling study by the DMH assay on a HCGI12K microarray between A2780 sensitive and resistant derivatives generated *in vitro* as well as between *in vivo* generated cell line pairs (PEO1 vs. PEO4, PEA1 vs. PEA2, PEO14 vs. PEO23) performed by Dr. Jens M. Teodoridis; 2) a prognostic DNA methylation signature associated with progression free survival in ovarian cancer (Wei et al., 2006); 3) and 4) CGIs identified from DNA methylation profiling study on an Orion MethylScope (Lippman et al., 2004) two-channel microarray based on McrBC restriction enzyme to detect differential methylation between A2780 sensitive and resistant cell lines (data not published); 5) candidate genes collected from the studies about acquired drug resistance in ovarian cancer through our collaborations (Professor Hani Gabra and Dr. Euan A. Stronach from Ovarian Cancer Action); and 6) other loci/genes differentially expressed before and after cisplatin treatment from gene expression profiling studies published from 2001 to 2008 (see Appendix 3.1 for the gene list and reference). Promoter CGIs within 2kb of the transcription start site (TSS) of the selected genes were defined by Gardiner-Garden (1987) by the following criteria: CG content greater than 50%, length of sequence over 200bp and the ratio of observed to expected CpGs is larger than 0.6. The genes without any promoter CGI were not taken into consideration. The CGIs that were not close to the TSS of the genes were also selected in the study if they showed differential methylation between the *in vitro* and *in vivo* generated cisplatin sensitive and resistant cell lines identified by MLDA on HCGI12K arrays (see Chapter 2) or on Orion MethylScope microarrays (data not published). In total, about 600

loci were covered by the custom-designed microarray.

Controls on the microarray

Three major types of controls were designed on the microarray (Table 3.1). Firstly, the sequences from the human mitochondrial genome which are generally unmethylated in human (see Chapter 2), and sequences from chromosome 16 that are underrepresented for CG regions and which contain very few McrBC recognition sites (ACG or GCG) were selected as unmethylated controls. Secondly, the sequences on HCG112K microarray (see Chapter 2) that showed high methylation levels in our previous DMH work were selected as positive methylation controls. Finally, sequences from non-human genome (see Table 3.1) were used as the negative controls which mainly reflected the background noise.

Controls	# probes	Source
Unmethylation	266	Mitochondrial genome
Unmethylation	400	Chr16: 59000000-61000000
Methylation	172	methylated probes on UHN12K microarray; promoter CGIs of GAGE family
Negative	577	Structural negatives (3xSLv1) ¹ Biological negatives ²

Table 3.1: Controls on the custom-designed microarray.¹This probe forms a hairpin and does not hybridize well with labelled samples of any species.²This probe has been shown not to have significant signal from any sample tested by Agilent.

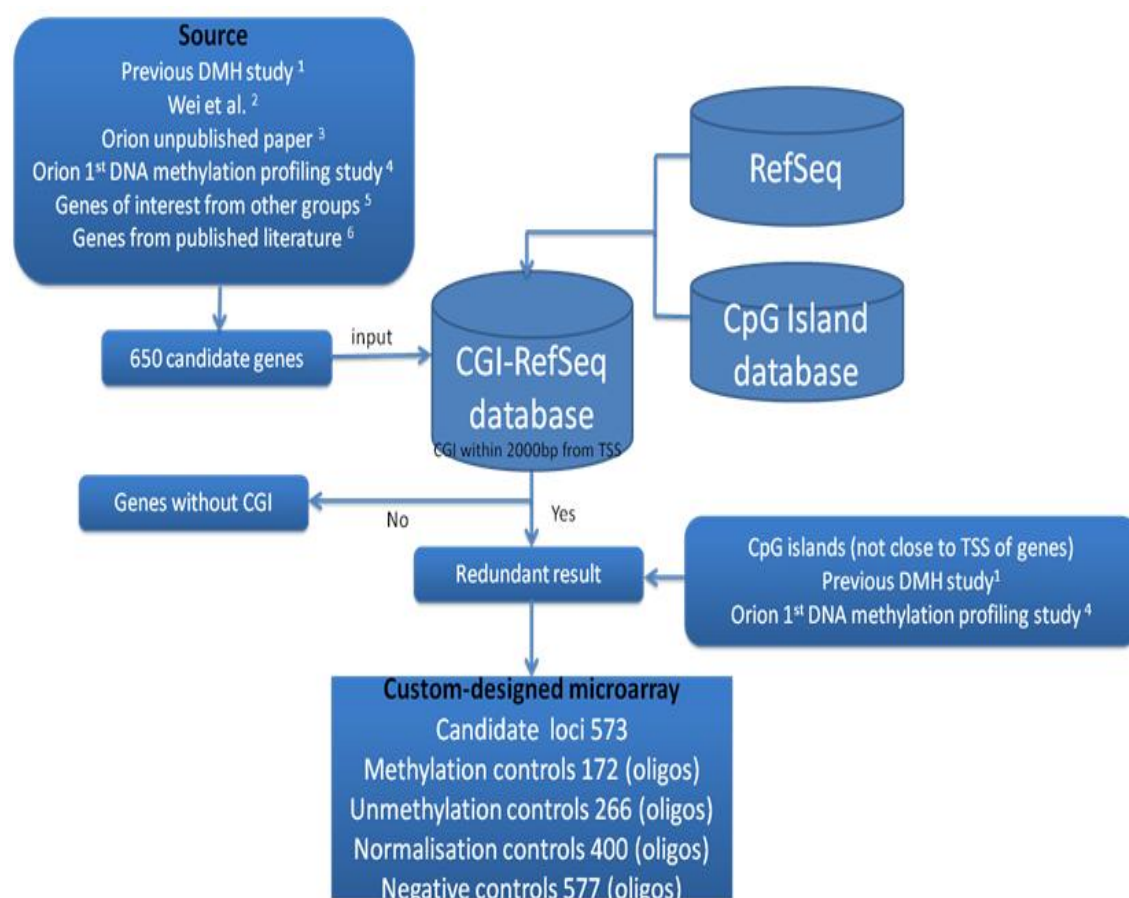


Figure 3.1: Workflow of target selection on Agilent custom-designed microarray in the feasibility study. ¹CGIs identified from DNA methylation profiling study by DMH assay on HCGI12K microarray between A2780 sensitive and resistant derivatives generated *in vitro* as well as between *in vivo* generated cell line pairs (PEO1 vs. PEO4, PEA1 vs. PEA2, PEO14 vs. PEO23) performed by Dr. Jens M. Teodoridis; ²the prognostic DNA methylation signature associated with progression free survival in ovarian cancer (Wei et al., 2006); ^{3,4}CGIs identified from DNA methylation profiling study on an Orion MethylScope (Lippman et al., 2004) two-channel microarray based on McrBC restriction enzyme to detect differential methylation between A2780 sensitive and resistant cell lines (data not published); ⁵candidate genes collected from the studies about acquired drug resistance in ovarian cancer through our collaborations (Professor Hani Gabra and Dr. Euan A. Stonach from Ovarian Cancer Action); ⁶other loci/genes differentially expressed before and after cisplatin treatment from gene expression profiling studies published from 2001 to 2008

Annotation of custom-designed microarray

The genomic annotation of human genes was obtained from the Refseq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) and the UCSC known gene database (<http://genome.ucsc.edu/>). Promoter CGIs within a 2kb of the transcription start site of the genes were obtained from UCSC database (Rhead et al., 2010) and from a genome-wide prediction of CGIs (Bock et al., 2007) as a supplementary set, both of which fulfil the mathematical model proposed by Gardiner-Garden to define a CGI (Gardiner-Garden and Frommer, 1987). MseI fragments were predicted by searching for the recognition sites (T/TAA) against the whole human genome sequence, masking the assembly gaps, intro-contig ambiguities and repeat regions. The prediction was performed in R using two packages: BSgenome and BSgenome.Hsapiens.UCSC.hg18. The genomic positions of targets were specified by Human Mar. 2006 (NCBI36/hg18) assembly. All the information was stored in an in-house MySQL database (<http://www.mysql.com/>). An R package, RODBC, provides an interface between R workspace and MySQL database.

The annotation pipeline is shown in Figure 3.2. The genomic locations of the probes were compared to the transcript start sites of the genes collected from Refseq database and UCSC known gene database, the genomic location of CGIs and MseI fragments. Any gene for which the TSS was within 2kb span of the probes was added into the annotation table. If multiple transcripts or different genes were related to the same probe, the annotation table included multiple official gene symbols, Entrez gene IDs, TSSs, and strands in the record for the corresponding probe. The genomic location and length of CGIs and MseI fragments targeted by the probes were also added in the annotation table. If the probes were in the 150nt flanking region of a CGI, it was annotated as a "flanking" probe. R functions [`mseI.anno()`, `cgi.anno()` and `gene.anno()`] for the microarray annotation are shown in the

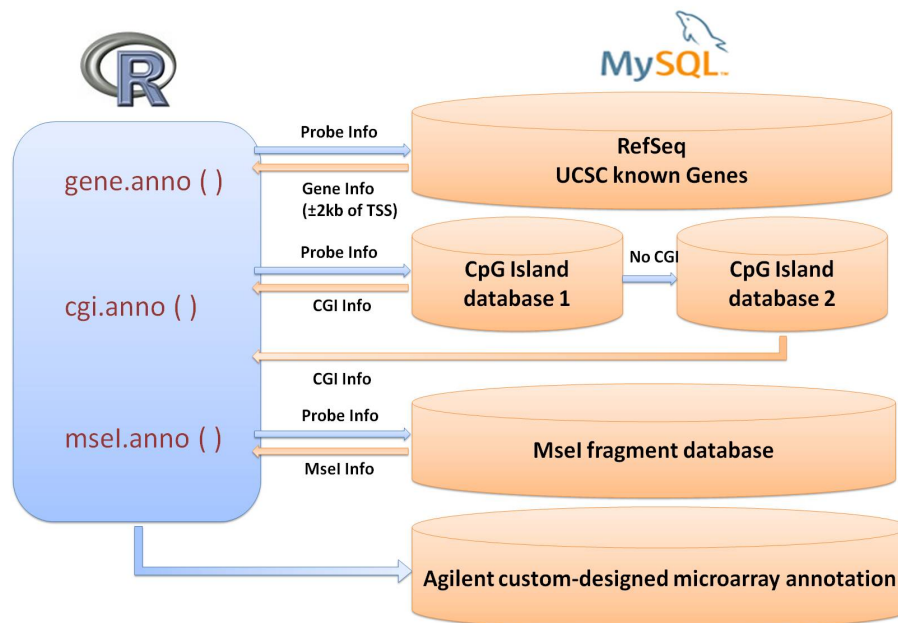


Figure 3.2: Annotation workflow of Agilent custom-designed microarray. *The probes are annotated as NA (not available) if they cover the MseI recognition site (T/TAA). CGI database 1 is from UCSC database and CGI database 2 is from Bock et al. (2007)

Appendix scripts and a complete annotation of this array is given in Appendix 3.2.

3.2.3 Differential methylation hybridisation (DMH)

Samples were assayed on Agilent custom-designed microarrays by DMH in duplicates (Nouzova et al., 2004). Briefly, DNAs were digested with MseI and fragments were ligated to annealed H12 (TAATCCCTCGGA) and H24 (AGGCAACTGTGCTATCCGAGGGAT) linkers. One fraction of linked fragments was digested with McrBC(5' $_{(A/G)C(N40-3000)}(A/G)$ 3') which cleaves methylated DNA sequences but leaves unmethylated sequences intact, the other fraction was mock digested. Both fractions were amplified using linker specific primers with only uncleaved fragments being amplified. A dye-swap design was used.

The amplicons from McrBC digested and mock digested samples are coupled either with Cy5(red) or Cy3(green), respectively, and then competitively hybridised on the microarray. Labelling of DNA, microarray fabrication, array hybridisation and image scanning was performed by Oxford Gene Technology (Oxford, UK).

The images of microarrays were scanned by Agilent C scanner at $5\mu\text{m}$ resolution and the raw signal intensities were extracted by Agilent Feature Extraction software (version 9.5.3.1) using CGH-v4_95_Feb07 protocol. The flagged probes were detected by Agilent Feature Extraction software. The probes flagged in more than 10% of the total samples were excluded from further analysis (Figure 3.4A). Otherwise, they were treated as missing data points (about 2%) and imputed by k-nearest neighbour (KNN) algorithm ($k=10$) (see section 3.2.4). The probes with signal intensities over 65000 were removed from the analysis due to the signal saturation. The probes with low intensities were also not taken into consideration. The MA-plot is a plot of the distribution of the signal intensity ratio (M) of the undigested (mock) sample to digested sample against the average signal intensity (A). M and A are defined as equations below.

$$\left\{ \begin{array}{l} M = \log_2(\text{undigest}) - \log_2(\text{digest}) \\ A = (\log_2(\text{undigest}) + \log_2(\text{digest}))/2 \end{array} \right\}$$

In a MA-plot, any probes with $A \leq \bar{A}_{control} + 1.96SD_{control}$ ($\bar{A}_{control}$ denotes the averaged A values of negative controls on the array, and $SD_{control}$ is the standard deviation of A values of negative controls) are considered as low intensity probes, and were thus excluded from the study (Figure 3.4C). Ultimately, about 90% of probes on the microarray were included in the subsequent analysis (Figure 3.4D).

Within-array normalisation was performed to reduce the influence of dye bias in MLDA

package (see Chapter 2 section 2.3.4), using a univariate linear regression model constructed using unmethylated probes selected from a poor CG region in chromosome 16, as well as from the mitochondrial genome. The slope of the linear regression model was adjusted to 1 and the intercept was adjusted to 0, since the log transformed signal intensities of the unmethylated controls were assumed to be equal. The log transformed signal intensities of all the probes on the microarray were adjusted correspondingly, as shown in formula 2.13 and 2.14 in Chapter 2.

3.2.4 Statistical analysis

The significance level was set at two-sided $p < 0.05$, unless it is stated as one-sided. All calculations and statistical analyses were performed in R (version 2.9.1) and SPSS (version 16.0).

Imputation of the missing data by KNN algorithm

The k-nearest neighbours ($k=10$) of the missing data point (a locus) used Euclidean distance (formula 3.1) to identify loci with data, and then the missing locus was imputed by averaging those non-missing loci of its neighbours.

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

Here, x_i and y_i denote the values of loci in two samples, respectively. n is the total number of loci on the array.

Coefficient of determination

The coefficient of determination, R^2 , is the proportion of variability of a set of data accounted for by the statistical model. In the linear regression model with single regulator, R^2 (formula 3.2) is the squared Pearson product-moment correlation coefficient (formula 3.3) between two samples. It ranges from 0 to 1 and an indicator of 'goodness of fit'. The higher R^2 , the better the fit is of the linear regression model. Therefore, it can be used as a measurement of consistency between duplicates in this study.

$$R^2 = \frac{l_{xy}^2}{l_{xx}l_{yy}} = r^2 \quad (3.2)$$

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} = \frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}} \quad (3.3)$$

Here, X is the DMH ratio (logged to the base 2) population in one sample and Y is the DMH ratio (logged to the base 2) population in the duplicate. $\bar{x}$ and $\bar{y}$ are the means of X and Y , respectively. l_{xy} is the sum of products of mean deviations of X and Y . l_{xx} and l_{yy} are similar to l_{xy} , which are the sum of products of mean deviations of X and X (or Y and Y).

Mann-Whitney U test

The Mann-Whitney U test is a non-parametric method to compare two independent random samples without the assumption of normal distribution. In two independent samples test, it is equivalent to the Wilcoxon rank sum test and the Kruskal-Wallis test for two groups. In the null hypothesis, the observations from two groups (X and Y populations) are combined and ranked, and the probability of an observation from one population X higher than that from the other population (Y) is equal to the probability of an observation from the pop-

ulation Y higher than that from the population X . Therefore, the probability of drawing a random larger sample is symmetrical in both populations. The alternative hypothesis is the probability of an observation from one population X exceeding the probability from the other population (Y) is not equal to 0.5. In two independent samples test, statistic U is calculated as described below.

1. The samples from two groups (n_1 and n_2) are ranked from the smallest to the largest.
2. The number of times that the rank from group 1 is larger than the rank from group 2, as well as the number of times that the rank from group 2 precedes the rank from group 1 is calculated.
3. The Mann-Whitney U statistic is the smaller time. In the case of small samples ($n_1 + n_2 \leq 30$), the distribution of U statistics is tabulated, but for sample sizes above 30 there is a good approximation with the use of normal distribution.

In this study, the methylation of loci were defined by DMH ratios of multiple probes that targeted the same locus being significantly greater than unmethylated regions selected from chromosome 16 and the mitochondrial genome. One-sided Mann-Whitney U test was used to test the difference and the significance level was set at $p < 0.0003$ based on Bonferroni correction ($0.05/161$, where 161 was the number of tumour samples that were assayed in DMH in the next stage of the study (see Chapter 4). To evaluate the prediction accuracy of methylation status and determine the minimum number of probes required to reach this significant level, seven loci that have been examined by MSP and are targeted by multiple probes on the microarray were used as a test set.

Locally Weighted Scatterplot Smoothing (LOWESS)

To estimate the optimal size of the MseI fragment, LOWESS was used to show (1) the relation between the distance of the probes to the MseI recognition site and signal intensities, where distance between the locus targeted by the probe and MseI fragment was calculated as the distance of the start site of the probe to the start site (recognition site) of the MseI fragment, and (2) the relationship between the DMH ratios and the length of the MseI fragments. The general idea of locally weighted scatterplot smoothing is a combination of local fitting of polynomials for smoothing and iterated weighted least square for robust estimation. Three parameters are crucial in LOWESS: f ($0 < f \leq 1$) is the proportion of the points used for smoothing. The increasing value of f increases the number of neighbourhood influential points, thus increases the smoothness; iter is the number of iterations for performing the fitting using the adjusted weights from the previous iteration; delta is used to speed up the computation, within which the points of the last computed point, are skipped and filled by the fitted value of linear interpolation. The default values for these three parameters were used in this study (f is 2/3, iter is 3 and delta is 1/100th of the range of x) to examine if the signals of the probes targeting different location of the MseI fragment are similar.

3.3 Results

3.3.1 Quality control in DMH

QC features in DNA methylation study

Since the Agilent custom-designed microarray was originally designed for the detection of copy number changes (CNA), Agilent Feature Extraction software does not provide quality

control (QC) reports that are tailored for DNA methylation analysis. Therefore the four features shown in Table 3.2 were used as indicators of DMH data quality, rather than using the QC report which is generated for comparative genomic hybridisation (CGH) microarrays by Feature Extraction. These four features indicate the outliers of non-uniformity and population at background and foreground in green and red channels, respectively. As shown in Figure 3.3, the highest consistency between duplicates was achieved using these four features to exclude probes (<5%) not fulfilling required data quality. The use of only two flags, *IsBGPpnOL* and *IsFeatNonUnifOL*, could also reach the same reproducibility of duplicates (Figure 3.3), and is therefore recommended when a large number of samples are involved to save the computing power.

Features (Green)	Features (Red)	Types	Description*
<i>gIsFeatNonUnifOL</i>	<i>rIsFeatNonUnifOL</i>	Boolean	Boolean flag indicating if a features is a NonUniformity Outlier or not. A feature is non-uniform if the pixel noise of feature exceeds a threshold established for a uniform feature
<i>gIsBGNonUniOL</i>	<i>rIsBGNonUniOL</i>	Boolean	The same concept as above but for background
<i>gIsFeatPopnOL</i>	<i>rIsFeatPopnOL</i>	Boolean	Boolean flag indicating if a feature is Population Outlier or not. Probes with replicate features on a microarray are examined using population statistics. A feature is a population outlier if its signal is less than a lower threshold or exceeds an upper threshold ($1.42 \times IQR$)
<i>gIsBGPpnOL</i>	<i>rIsBGPpnOL</i>	Boolean	The same concept as above but for background

Table 3.2: Features for quality control in DMH. * Reference Guide of Agilent Feature Extraction Software

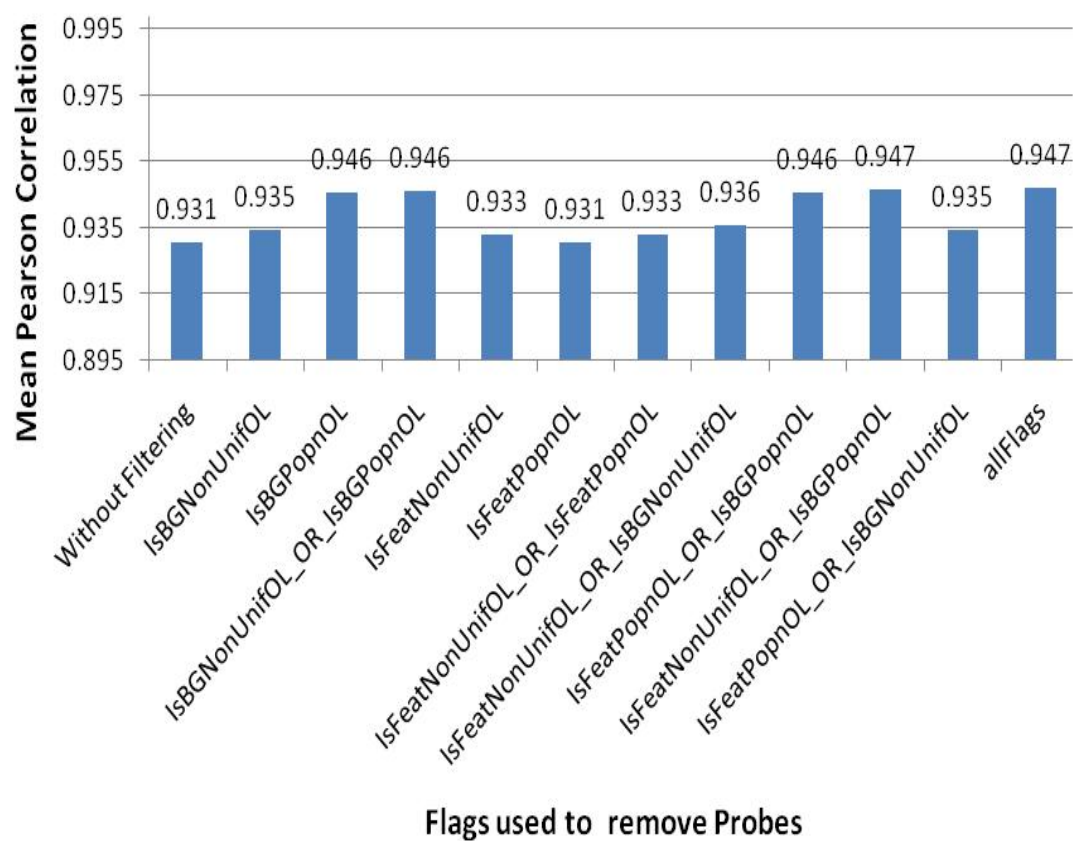


Figure 3.3: Average correlation between replicates of filtered probe sets within arrays. Various combinations of flags used to filter out the poor quality probes were tested. This analysis was performed by Dr. Volker Brenner from Oxford Gene Technology (OGT, UK)

Methylated and unmethylated controls

In general, unmethylated controls from mitochondrial genome and a poor CG region from chromosome 16 worked well and showed linear correlation between signals of McrBC digested and undigested samples ($p < 0.05$), although some probes targeting mitochondrial sequences reached signal saturation (signal > 65000) (Figure 3.4B). Two distinct groups of methylation controls were observed across all the samples in the pilot study (red spots in Figure 3.4B and D). One had a high level of methylation and the other had medium or even low level of methylation. Therefore, the former group would be appropriate controls to use in future studies, while the latter would not. The majority of the methylation controls were identified from the previous DMH study on UHN12K microarray (Chapter 2). However, as previously described (Chapter 1), over 10% probes on the UHN12K array targeted multiple sequences in the human genome according to the annotation of UHN12K microarray (<http://data.microarrays.ca/cpg/>) at the time we analysed the data. Thus, the high methylation level of those probes observed in a UHN12K study could be a confounding effect of cross-hybridisation. To identify appropriate methylation controls that would not be confounded by this effect, MLDA was applied as described in Chapter 2, in which methylation scores representing the methylation status of the loci mapped by the probes on duplicate arrays were calculated to each control probe based on the weighted scoring schema in MLDA (Figure 2.4 in Chapter 2). Only the methylation controls with an MLDA score > 0, which represented methylated on duplicate arrays, across all the cell lines were used in future study. The majority of these controls (66/89 probes) were from the promoter CGIs (49074918-49075532, 49103473-49103955, 49065375-49065985 and 49202750-49203235) of GAGE gene family on chromosome X.

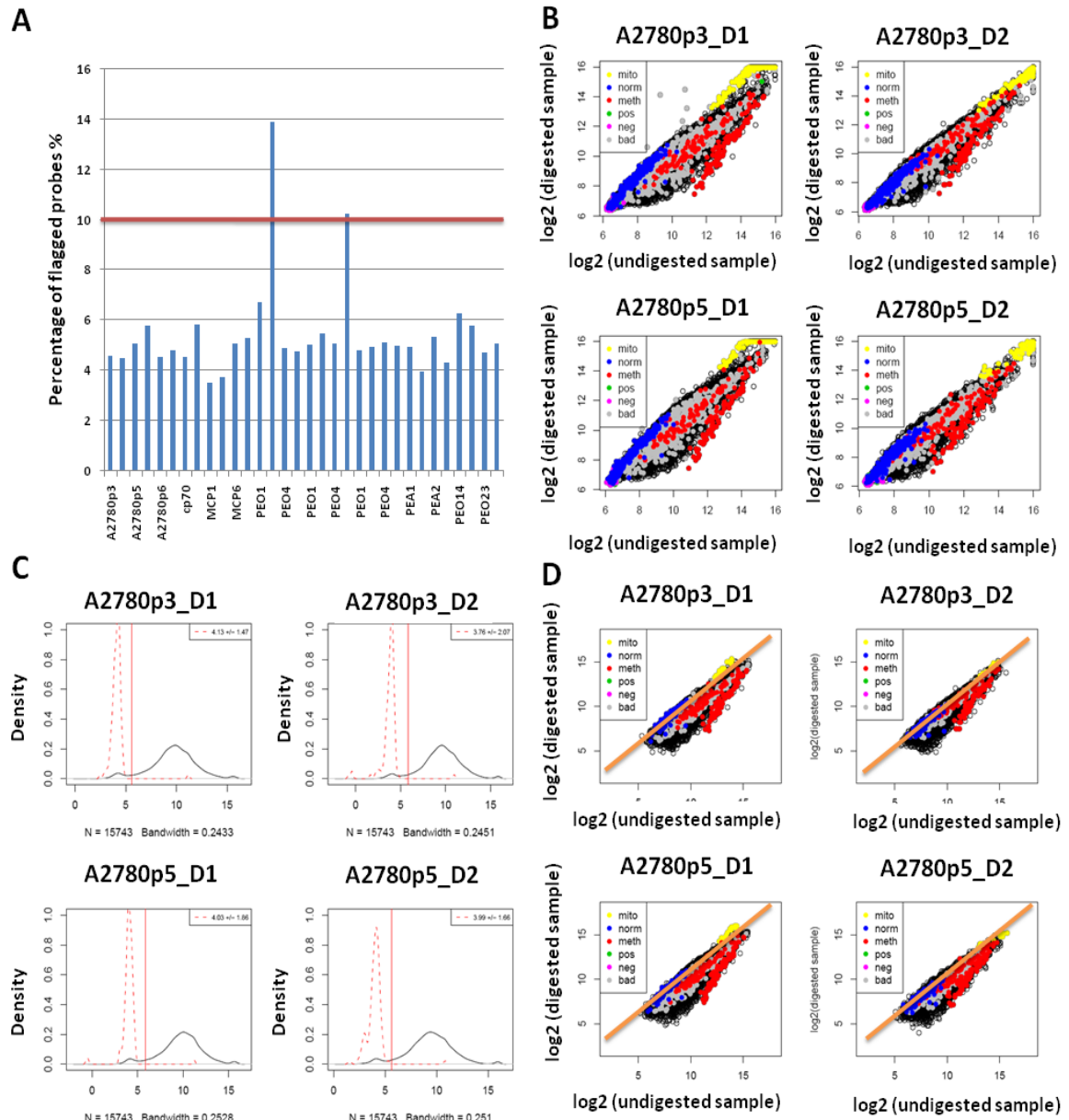


Figure 3.4: DMH data in pilot study. A: percentage of flagged probes. Each cell line has duplicates; B: scatter plots of raw signals of McrBC digested and mock samples in DMH; C: negative controls and low signals. Red broken line indicates the density of the negative controls. Red solid line indicates the cut off for low signals; D: scatter plots of pre-processed signals of McrBC digested and mock samples in DMH. Orange line: unmethylated regression line. Yellow spots: mitochondrial controls; Red spots: methylated controls; Blue spots: unmethylated controls selected from chromosome 16; Grey spots: the probes with quality issues.

3.3.2 Reproducibility of Agilent custom-designed microarray

To measure the reproducibility, R^2 , the coefficient of determination of the linear regression model constructed by duplicates was used. The independent preparation of DNA showed similar reproducibility as technical replicates ($mean \pm SD$: 0.86 ± 0.07 vs. 0.85 ± 0.06), suggesting that there was little variation from different DNA preparations. However, a large variation was introduced by dye bias in the dye-swap duplicates resulted in a low R^2 ($mean \pm SD$: 0.64 ± 0.09 , Figure 3.5A). The independently isolated DNA duplicates had a slightly higher R^2 when McrBC digested sample was labelled by Cy3 and mock sample labelled by Cy5 (dye labelling method 1) than those in dye-swap counterparts (dye labelling method 2) ($mean \pm SD$: 0.87 ± 0.07 vs. 0.83 ± 0.04 , Figure 3.5B). Therefore, technical replicates with McrBC digested sample labelled by Cy3 and mock sample labelled by Cy5 (dye labelling method 1) was used in the later study of tumours (Chapter 4 and 5) instead of using dye-swap duplicates, in order to minimise variability due to dye labelling.

3.3.3 Dye bias correction

The log transformed DMH ratios were corrected by MLDA (see Chapter 2 Section 2.3.4) to reduce dye bias and improve the reproducibility of DMH assay. As shown in Figure 3.6A, after correction the mean of log ratios of unmethylated controls in sample A2780p3 is around 0, therefore, the methylation baseline was adjusted to zero in each sample. The correlation of about 600 unmethylated controls in dye-swap duplicates was estimated before and after the correction. Figure 3.6B shows that the average correlation of unmethylated controls between 20 dye-swap duplicates increased from 0.36 to 0.54 after the correction ($p = 2.09 \times 10^{-7}$, $n = 20$, paired Student's t test) indicating the reduction of variation introduced by dye-swap labelling.

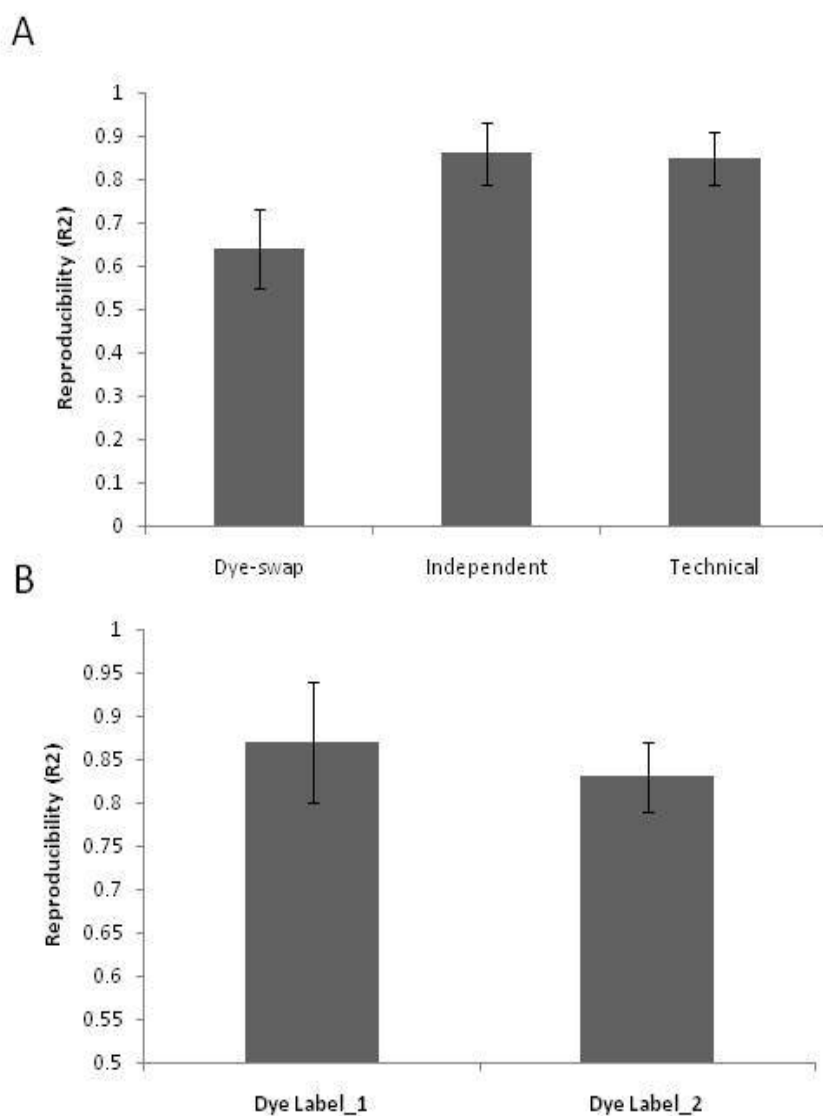


Figure 3.5: R square of linear regression model between replicates. A: reproducibility of duplicates. Dye-swap: dye-swap replicates; Independent: independently prepared DNA; Technical: hybridisation replicates. B: reproducibility of duplicates using two dye labelling methods. Dye label_1: McrBC digested sample labelled by Cy3 and mock sample labelled by Cy5. Dye label_2: dye-swap labelling method of Dye label_1. Error bar: Standard Error of Mean (SEM)

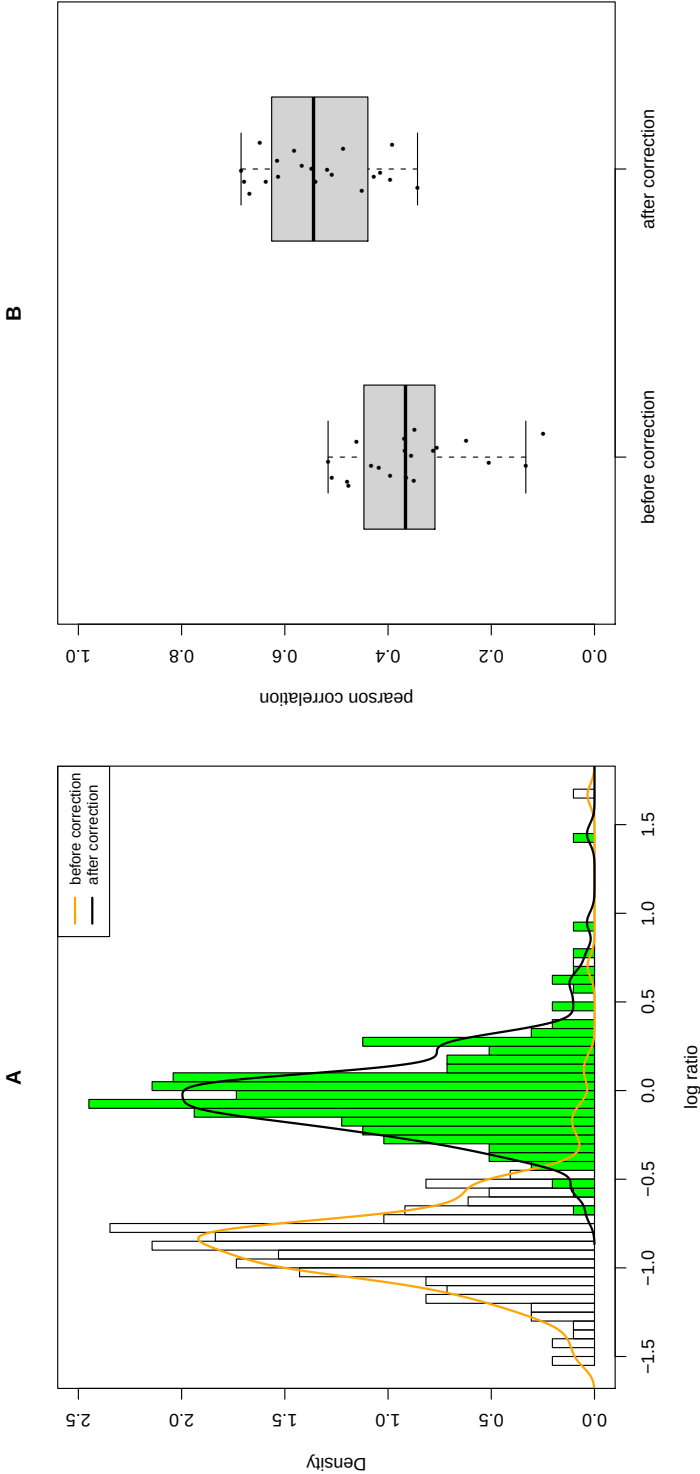


Figure 3.6: Dye bias correction in MLDA. A: distribution of log ratios of unmethylated controls before and after correction; B: Pearson correlation of unmethylated controls in duplicates before and after correction.

3.3.4 Methylation measurement of a CpG Island by the DMH assay

In DMH assay, PCR amplification is performed following MseI digestion on genomic DNA. In the early cycles of PCR, slight differences in amplification efficiency can produce different amount of MseI fragments, which results in changes in DMH ratios between samples. To understand how MseI digestion influences DMH ratios and how to summarise the methylation level of a CpG island, all of the MseI fragments were identified in the human genome using the MseI recognition site (T/TAA) *in silico*. Figure 3.7A shows the number of MseI fragments on 23 chromosomes. The average size of MseI fragment in humans is 158bp. Among 19,029,911 fragments predicted *in silico*, only 642 (0.30‰) is more than 6,000bp. The majority of fragments (92.5%) have a size of 0.1 to 2kbs, which is the size range of MseI fragment suggested to enhance restriction enzyme methylation analysis in recent study (Tran et al., 2009).

As illustrated in Figure 3.7B, the methylation levels of the loci targeted by the probes were indeed similar within a given MseI fragment. Therefore, individual MseI fragment instead of CGI was used as the unit by which the methylation level of a locus was estimated. About 50% of CpG Islands targeted by our custom-designed microarray had no MseI recognition site within the sequences, and were therefore treated as individual locus. Other CpG islands had various numbers of recognition sites ranging from 1 to 57, and were treated as multiple loci (MseI fragments) in the analysis.

Noticeably, the signals of the probes close to the MseI recognition size were not as high as those further away from the recognition size. To examine if the distance of the probes to the MseI recognition site had any influence on the signals and estimate an optimal MseI fragment size, the probes within ± 500 bp of MseI cut site were identified. Figure 3.7D shows the change of log transformed signal intensities in both green and red channels, as

well as DMH ratio, with the distance to MseI recognition site in two independent samples. The signals decrease at the region close to MseI cut site and reach their lowest value at around upstream 31bp. There are about 120bp (-90 to +30bp) distance between two inflections.

A minor shift of the signal intensities of two types of probes (forward and reverse), was observed in Figure 3.7C. The dip for forward and reverse probes is still at 31bp upstream of MseI cut site, except signal intensities in red channel (McrBC digest sample) of forward probes of which the dip is at -22bp. Therefore, the distance of probes towards and away from cut site is better to be -90 to +40bp for forward probes and -90 to +30bp for reverse complementary probes. So the minimal size of MseI fragment has to be 130bp. The log transformed signal intensities/DMH ratios of all the probes within the fragments were averaged. Average signal intensities increase with the size of MseI fragment and reach the peak at 440 bps, then decrease with the size slowly and reach the bottom at around 2500bp (Figure 3.7D). So the suggested size of MseI fragment that is recommended to be analysed by DMH is 130 to 2500bps.

3.3.5 Sensitivity and specificity of DMH on Agilent custom-designed microarray

To confirm that DMH assay on this custom-designed microarray could detect the loci known to be differentially methylated between cisplatin sensitive and resistant cell lines, nine loci (MseI fragments) were used as a test set. These loci have been examined by bisulfite pyrosequencing/methylation specific PCR (performed by Dr. Jens, M. Teodoridis, Dr. Janet Graham and Dr. Constanze Zeller in Imperial College London, UK) in both *in vivo* and *in vitro* generated cell lines previous, and were found differentially methylated between

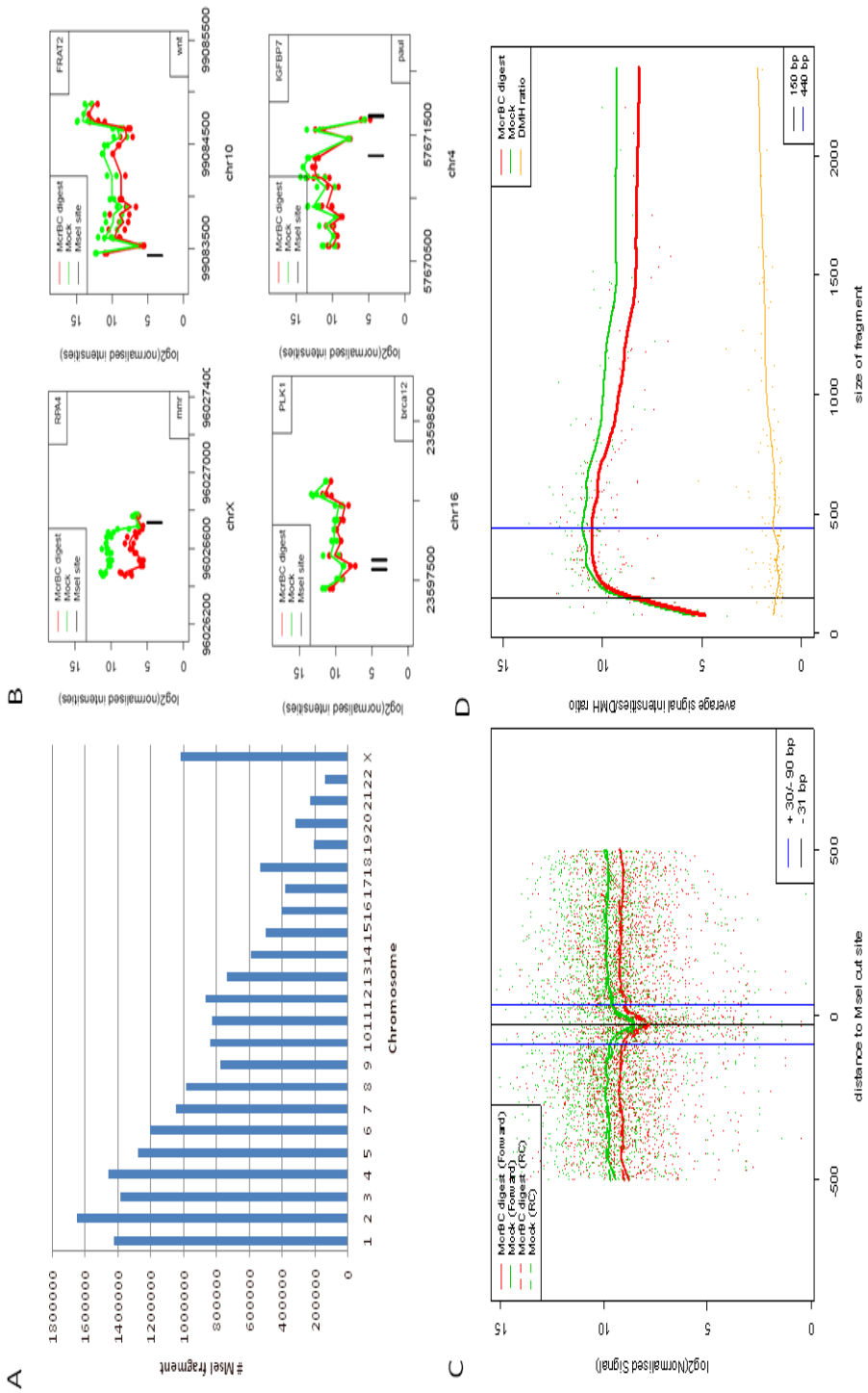


Figure 3.7: Methylation of the loci within MseI fragment. A: number of MseI fragments in Human genome. B: methylation of 4 loci containing MseI recognition site. Black solid line indicates the MseI recognition site. C: Reduction of signal intensities of probes close to MseI recognition site in both McrBC digested and mock-digested samples. Solid line: probes on forward strand. Broken line: probes on reverse complementary strand. D: signal intensities of probes and length of MseI fragments. Red: McrBC digested sample. Green: McrBC-mock digested sample. Orange: log DMH ratio

three A2780 sensitive and resistant derivatives, but 5 out of 9 had no change of methylation in *in vivo* cell line pairs (PEO1 vs. PEO4, PEA1 vs. PEA2 and PEO14 vs. PEO23).

The probes targeting these 9 loci were identified on the custom-designed microarray and methylation of the loci was measured by average DMH ratios of multiple probes. The result shows methylation of 9/9 (100%) loci is significantly different between A2780 sensitive and resistant cell lines if McrBC digested samples are labelled by Cy3 and reference samples are labelled by Cy5 ($p < 0.05$) (Figure 3.8A). This confirms high sensitivity of DMH assay to detect differential methylation on the custom-designed microarray. In the dye-swap replicate, 8/9 (89%) loci show significant differential methylation and methylation of one locus close to *SP5* was not identified as significant locus by Mann-Whitney test due to larger variation of methylation within the group (Figure 3.8B). This is consistent with the result in section 3.3.2 that showed the latter dye labelling method introduced larger variation which could possibly mask the true differential methylation between samples.

In addition, five loci that have no differential methylation in the sensitive and resistant cell line generated *in vivo* shown similar methylation level between the sensitive and resistant cell lines on the arrays (Figure 3.8C and D). Therefore, the DMH assay combining the custom-designed microarray not only has a high sensitivity but also has a high specificity for the detection of differential methylation in ovarian cancer cell lines.

3.3.6 Correlation between DMH and bisulfite pyrosequencing

To further explore the relationship between DMH and bisulfite pyrosequencing, methylation at three loci (*SP5*, *FOXA1* and *TFAP2C*) in A2780p3, A2780p5, A2780p6, cp70, MCP1 and MCP6 cell lines were tested by both methods. Average DMH ratios were linearly correlated with the result from bisulfite pyrosequencing (R^2 see Figure 3.9A and B),

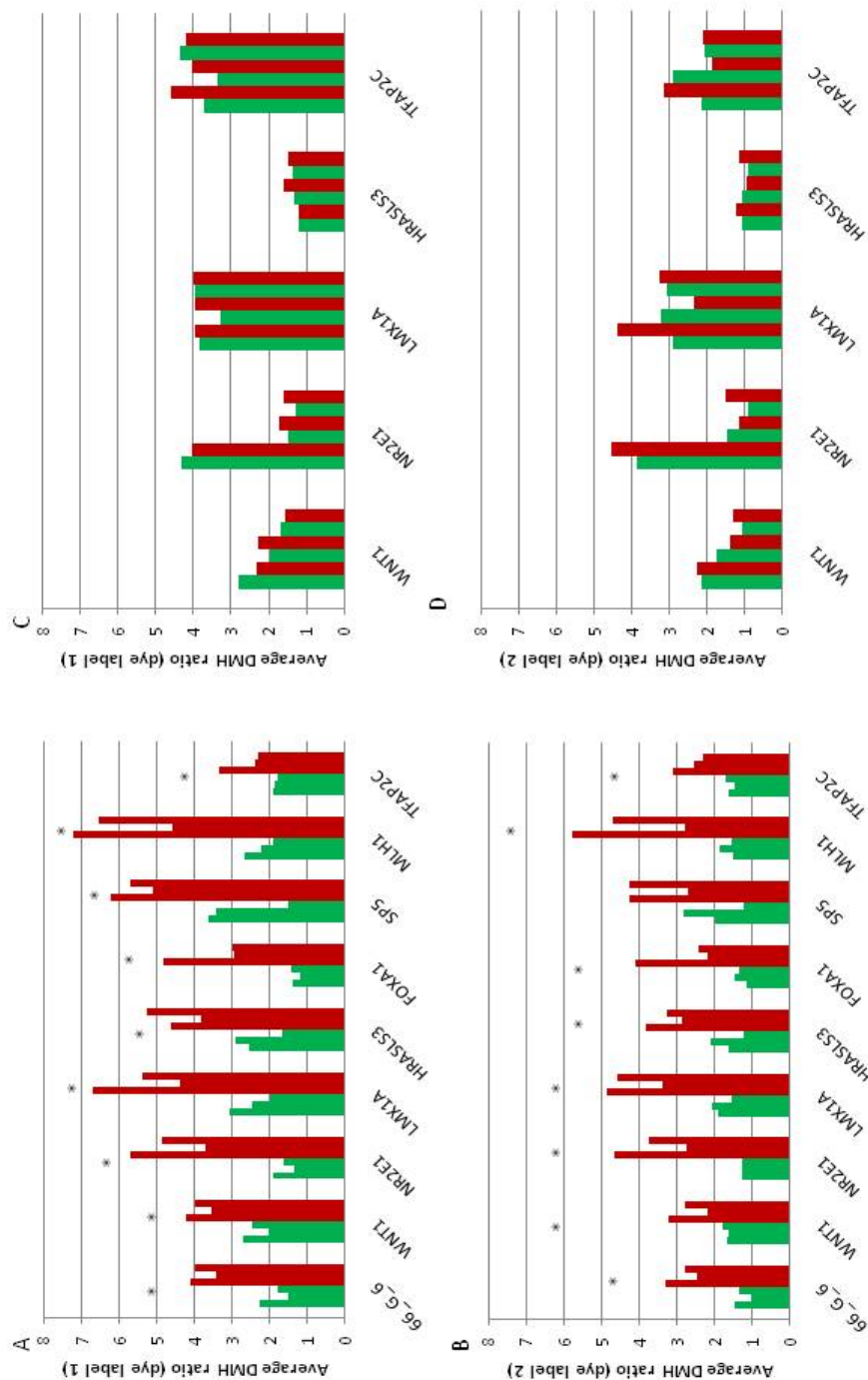


Figure 3.8: Differential methylation between cisplatin sensitive and resistant cell lines. A: Methylation of 9 loci measured by DMH using dye labelling method 1 in A2780 sensitive and resistant derivatives. B: Methylation of 9 loci measured by DMH using dye labelling method 2 in A2780 sensitive and resistant derivatives. C: Methylation of 5 loci measured by DMH using dye labelling method 1 in *in vivo* cisplatin sensitive and resistant cell line pairs. D: Methylation of 5 loci measured by DMH using dye labelling method 2 in *in vivo* cisplatin sensitive and resistant cell line pairs. Red: Cisplatin sensitive lines. Green: Cisplatin resistant lines. A2780 sensitive derivatives: A2780p3, A2780p5 and A2780p6. A2780 resistant derivatives: cp70, MCP1 and MCP6. *In vivo* cisplatin sensitive and resistant cell line pairs: PEO1 vs. PEO4, PEA1 vs. PEA2 and PEO14 vs. PEO23. * $p < 0.05$ (Mann-Whitney test).

especially when the McrBC digested sample was labelled by Cy3 (Figure 3.9A). This result further confirms that the use of average DMH ratios could represent the methylation level of the locus, and thus detect the differential methylation.

3.3.7 Determination of methylation status in cell line model

To quantify the methylation frequency of genes in ovarian tumours in later studies (see Chapter 4 and 5), an appropriate method of analysis is required to determine if the locus is methylated. Unfortunately, the use of continuous DMH ratios is not possible to achieve this goal unless a cutoff for methylation is determined. In previous DMH studies, an arbitrary cut-off has been used without considering the variations in DMH assays, leading to a potential increased in false positives or negatives. In the present study, a method based on Mann-Whitney U test was used to estimate the methylation status using DMH ratios of multiple probes (see Methods, section 3.2.4). The locus was identified as methylated if the p value is less than 0.0003, based on Bonferroni correction ($0.05/161$, 161 is the number of tumour samples that were assayed by DMH in the next stage of the study (see Chapter 4)). To evaluate how many probes are required to reach this significant level, 4 loci were used that were methylated in CP70 and targeted by multiple probes on the array. The probes were randomly selected within the loci and the number of probes used for the Mann-Whitney test was decreased from maximum to 3. As shown in Figure 3.10, at least 7 probes are required to reach this significant level.

Seven loci examined by MSP and known to be methylated/unmethylated in A2780 and cp70 cell lines were used to evaluate this method. The result shows 6 out of 7 loci were predicted correctly in A2780 and CP70 cell lines, while one locus (66_G_6, chr9: 95748103-95748629) was methylated in CP70 cell line but predicted as unmethylated locus

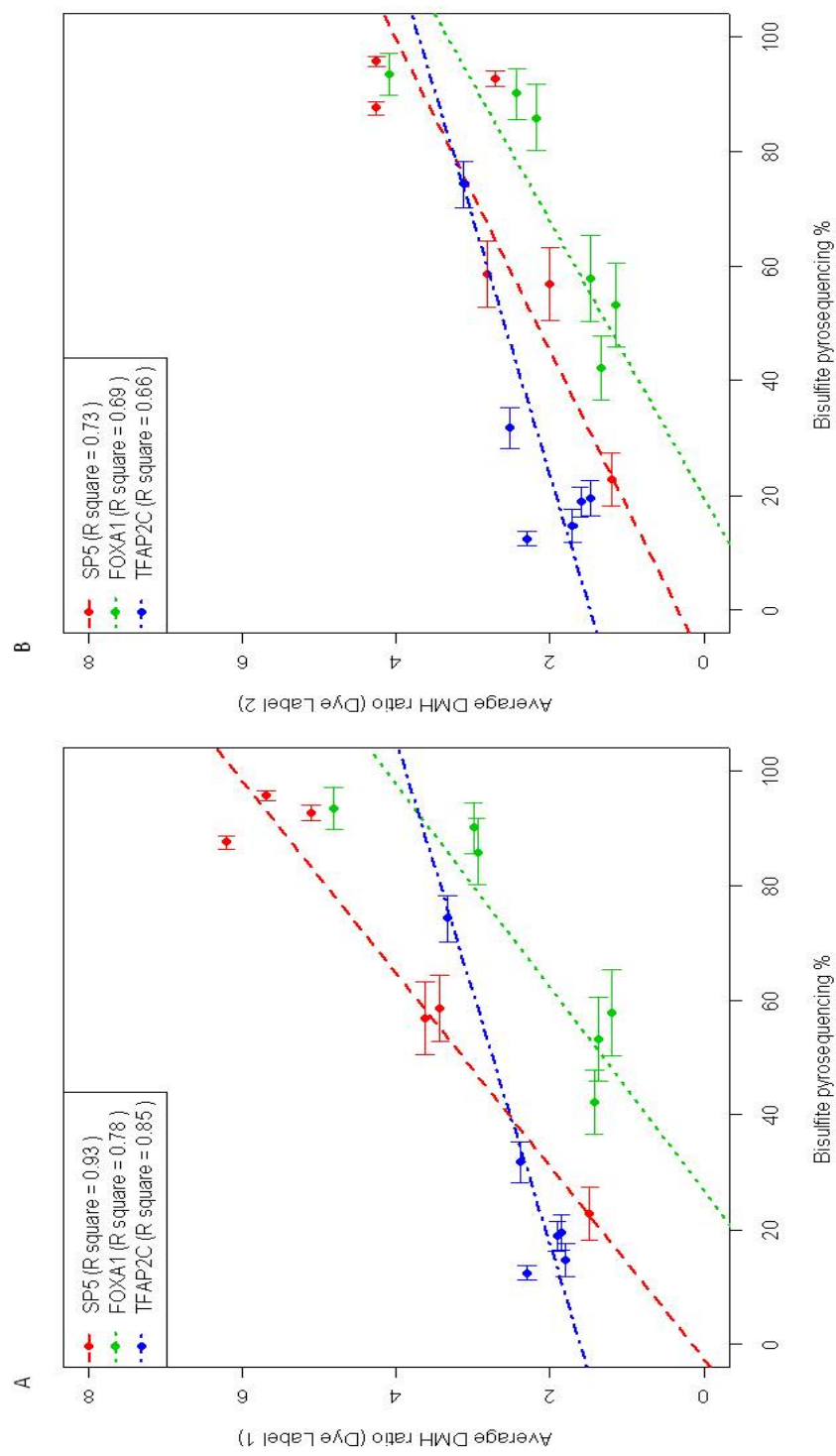


Figure 3.9: Correlation between DMH and bisulfite pyrosequencing. A: DMH using dye labelling method 1. B: DMH using dye labelling method 2. The linear regression models in three genes *SP5* (red), *FOXA1*(green) and *TFAP2C*(blue) are shown as broken lines. Error bar: one standard error of the methylation at CpG sites covering the same locus.

(Figure 3-10). This may be because only 4 probes target this locus, which is not sufficient to reach the level of $p < 0.0003$. Another locus in *AGBL2* was also targeted by only 4 probes, but as this locus was unmethylated in both cell lines, the unmethylated status determined by this method could be just as a result of the poor coverage of the probes. Therefore, further examination of the number of the probes is always necessary when the loci are predicted to be unmethylated.

3.4 Discussion

3.4.1 Controls on the Agilent custom-designed microarray

The use of mitochondrial sequences as the unmethylated controls have been described and discussed in detail in Chapter 2. In addition, 400 probes that mapped to a poor CG region of chromosome 16 were also selected as unmethylated controls (yellow spots in Figure 3.4B and D). In the DMH data pre-processing, log DMH ratios of those unmethylated controls were adjusted to zero by MLDA. Generally, LOWESS algorithm is used in two-channel microarrays to correct dye-bias. However, it requires the signals of the selected controls to cover the whole range of the signal intensities on the microarray. As shown in Figure 3.4B and D, the signal intensities of the unmethylated controls only cover the low and high signals; therefore, the linear regression model implemented in MLDA is more suitable to reduce the dye-bias in this case than LOWESS method. The methylated controls were selected and evaluated on the Agilent custom-designed microarray. CpG island methylation of testis specific genes, such as *MAGE*, *GAGE*, *XAGE* gene families, are the predominate mechanism of transcriptional repression of these genes in many tissues (Dunn et al., 2003; De Smet et al., 1999). The majority of selected methylated controls are from promoter CGIs of GAGE family (GAGE2E, GAGE2C, GAGE12J, GAGE12E, GAGE12G, GAGE4,

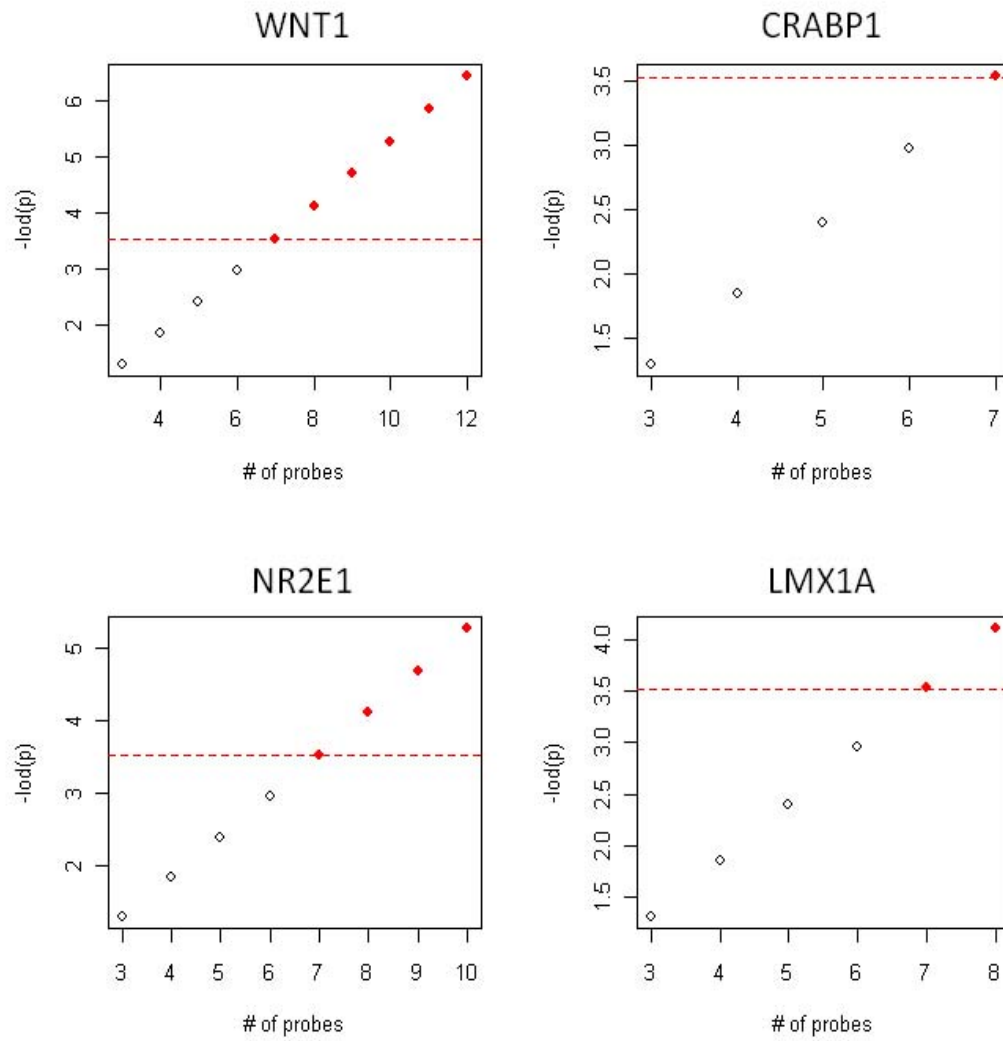


Figure 3.10: Coverage of probes and significant level in one side Man-Whitney test. The red broken line indicates the significant level $p = 0.0003$, i.e. $-\log(p) = 3.52$. The red spots show the numbers of probes required to reach the significant level ($p < 0.0003$) in one-side Mann-Whitney test.

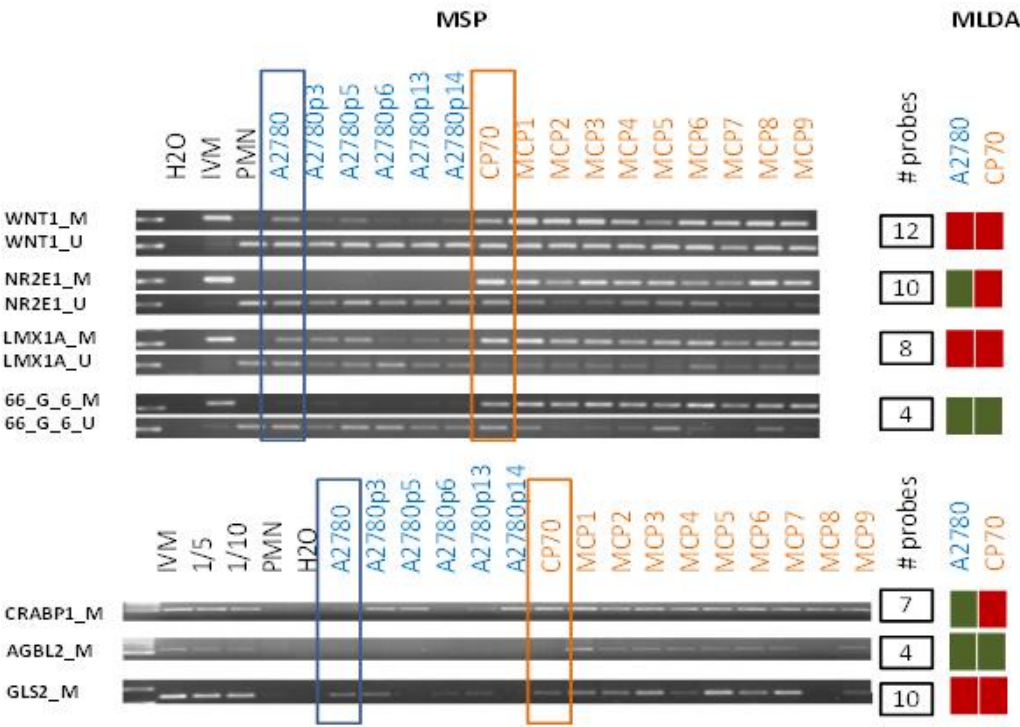


Figure 3.11: Methylation in DMH and MSP. Left: methylation of 7 loci tested by MSP in A2780 (blue frame) and CP70 (orange frame). Right: number of probes targeting the locus and methylation status determined on DMH. Red: methylated locus; Green: unmethylated locus.

GAGE7, GAGE8, and GAGE13) on chromosome X, but not from promoter CGI of GAGE1 that has been shown to express in ovarian cancer cell line A2780 as well as in 33.3% serious ovarian tumours and 16.7% of endometrioid ovarian tumours (Zhang et al., 2010). It will be necessary to further evaluate these selected methylated controls that were identified by the methylation analysis of ovarian cancer cell lines and tumours. Therefore, it will be important to show that they retain methylation (Chapter 4), since it has been reported that expression of these testis specific genes are increased in certain type of tumours, potentially indicating the loss of methylation at promoter regions of the genes.

3.4.2 Modification of DMH assay

A larger variation was introduced when McrBC digested sample was labelled by Cy5 (red) compared to Cy3 (green), which therefore influenced the detection of differential methylation between samples, particularly with the DMH assay in which distribution of data is asymmetric and a parametric test is not applicable. As shown in Figure 3.8B, 8 loci exhibit differential methylation using this labelling method tested by Mann-Whitney non-parametric statistics, however, the differential methylation of one locus in *SP5* was masked by the large dynamic range of DMH ratios within the group. Meanwhile, DMH assay labelling McrBC digested sample by Cy3 (green) performed better, as 9/9 (100%) loci showed differential methylation between two groups of cell lines. Therefore, given that the variation of methylation within tumour tissues is smaller than those in ovarian cancer cell lines, the latter labelling method is suggested to be used in duplicate DMH assays in ovarian tumour samples instead of dye-swap design. Indeed, this method appears to be more sensitive to detect difference in methylation with an appropriate correction for dye bias using the unmethylated controls designed on the arrays as described in the Method section of this chapter.

3.4.3 Modification of MLDA

Global between-array normalisation on custom-designed microarray is not recommended for two main reasons. Firstly, methylation changes are asymmetric (see Chapter 2). Secondly, custom-designed microarrays are focused on the regions that have been shown to be differentially methylated between ovarian cancer cell lines/tumour tissues either in our and collaborators previous DMH work or in the literature. Thus, there may be high frequency of methylated loci, which again biases the results away from a normal distribution of the data. MLDA were tailored for DMH data analysis without between-array normalisation, but this method is not applicable to use the average value to summarise the methylation level of the whole MseI fragment given the methylation scores from MLDA are ordinal. In this feasibility study, this algorithm has been modified (see section 2.3.4 in Chapter 2) to correct signals of McrBC digested and mock digested samples to reduce dye bias based on the linear regression, at the same time, the log ratios of the unmethylated controls were adjusted to zero. Correspondingly, the ratios of the rest of the probes were adjusted using the parameters estimated from the linear regression model. The adjusted log DMH ratios become more comparable between arrays as shown in Figure 3.6B. Consequently, the use of average DMH ratios estimated from the log ratios adjusted by MLDA could reflect the methylation changes of the regions targeted by multiple probes across the samples.

3.4.4 Identification of differential methylation within MseI fragment

We observed that differential methylation between cisplatin sensitive and resistant cell lines occurred in a small region, a so called "hot spot", close to the TSS of *NR2E1* and *WNT1* (Figure 3.12). Although the use of average DMH ratios can reflect the differential methylation of the whole MseI fragment between ovarian cancer cell lines in this study, bisulfite pyrosequencing, bisulfite sequencing or next generation sequencing of the whole MseI

fragment would be more powerful to identify such hot spots that are differentially methylated between samples, especially in tumours samples where differential methylation is not as dramatic as that in cancer cell lines.

3.5 Conclusions

This feasibility study shows that the use of differential methylation hybridisation (DMH) on our Agilent custom-designed 15K microarray could distinguish the regions of differential methylation between A2780 cisplatin sensitive and resistant derivatives with high sensitivity and specificity. Several recommendations can be made for the next stage of the study with regards to DNA methylation profiling in ovarian tumours: (1) the design of unmethylated and methylated controls on the arrays are necessary to monitor DMH assay, adjust for the dye bias and reduce the systematic variations in the experiments, (2) four features that indicated the outliers of non-uniformity and population in the background and foreground in both channels can be used to remove the poor quality probes, (3) simple duplicates instead of dye-swap duplicates are recommended to avoid the variations introduced by dye-swap labelling and the extra costs incurred, (4) to quantify the quality of DMH data, the percentage of the flagged probes, median signal intensities of McrBC mock digested samples, the reproducibility of duplicates as well as the difference of DMH ratios between methylated and unmethylated controls should be reported at the beginning of the analysis, (5) ideally, at least seven probes that target multiple regions for each MseI fragment should be designed on the microarray, (6) the suggested length of each MseI fragment included in the methylation analysis should be between 130 and 2500bp, and (7) the average DMH ratios of the MseI fragments estimated from the log ratios of multiple probes adjusted by MLDA should be used to reduce the number of data points, thus increasing the power of

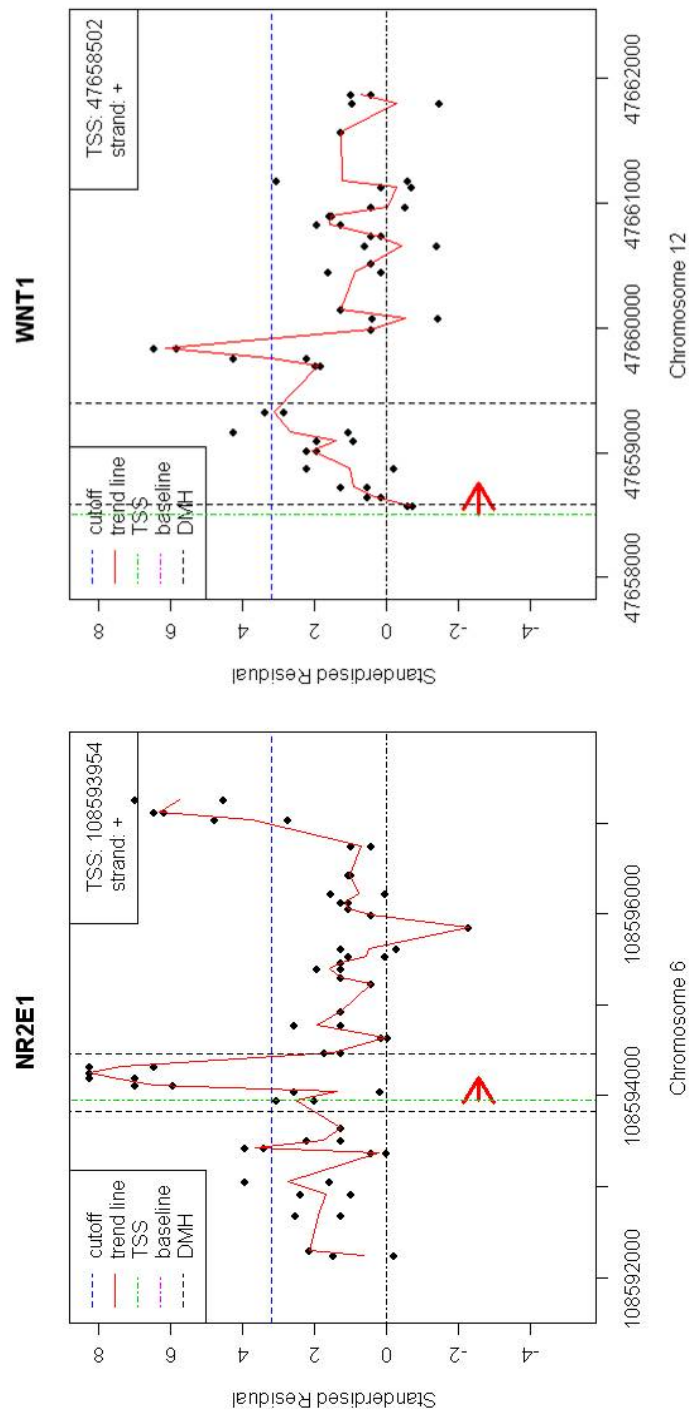


Figure 3.12: Differential methylation scores estimated from MLDA in *NR2E1* and *WNT1* between cisplatin sensitive ($n = 6$) and resistant ($n = 10$) cell lines. The black spots indicate the probes on the custom-designed microarray. Green line shows the transcription start site; Red arrow indicates the transcriptional direction of the gene; black horizontal line is the base line (no differential methylation); Blue line is the differential methylation cut-off determined by MLDA ($FDR < 0.05$); black vertical lines show the sequence exhibiting differential methylation between sensitive and resistant cell lines on HCG112K array

the analysis.

Chapter 4

CpG Islands Methylation Profiling in the Wnt Pathway and Their Prognostic Significance in Epithelial Ovarian Cancer

4.1 Introduction

Epithelial ovarian cancer (EOC) is the most lethal of all the gynaecological cancers accounting for 52% of all gynaecological cancer-related deaths (Jemal et al., 2009). Patients with advanced disease have surgical cytoreduction followed by platinum-based chemotherapy, but there is a low 5-year survival rate of 40% (<http://info.cancerresearchuk.org>). There is a need for prognostic biomarkers to assist our understanding of the molecular basis of ovarian cancer progression and of sufficient discriminatory prognostic power to be of clinical value. The established prognostic factors of EOC at clinical presentation before primary chemotherapy are residual disease and patient age, as well as histology, stage and grade of tumour (Cannistra, 2004; Konstantinopoulos et al., 2008; Kulasingam et al., 2010; Agarwal and Kaye, 2003). Although many individual biomarkers in EOC have been investigated (Berchuck et al., 1990; Havrilesky et al., 2003; Verri et al., 2005; Lee and Park, 2009), none of them have been translated as independent prognostic biomarkers of sufficient value to direct clinical management. Identifying robust prognostic markers following conventional treatment of EOC will provide a solid basis for characterising potential predictive biomarkers to novel treatment strategies and hence for stratification of patients for targeted care in clinical practice.

Aberrant DNA methylation frequently occurs in cancer, particularly at regions generally unmethylated in normal cells, known as CpG islands (CGIs) (Illingworth and Bird, 2009). CGIs often co-localise with the promoters of genes and promoter hypermethylation is associated with repression of gene transcription (Illingworth and Bird, 2009). Several studies have shown that CGI methylation has potential as a biomarker for monitoring tumour progression (Wei et al., 2006) and is associated with platinum-based chemoresistance in EOC (Li et al., 2009; Chou et al., 2010). However, many of these studies are either

limited by small sample size or lack of validation of the methylation biomarker as an independent marker.

The Wnt pathways control a variety of biological processes including cell proliferation, differentiation and migration, and have a crucial role in development and tissue homeostasis (Boyer et al., 2009). Wnt signalling is transduced through binding of Wnt molecules to the transmembrane Frizzled receptor proteins (Fz) and activation of downstream signalling pathways depending on the interaction between Wnt molecules and their specific receptors. Historically, the Wnt pathway is classified as "canonical" and "non-canonical" pathway (Figure 4.1). In the "canonical" pathway (McDonald and Silver 2009), binding of Wnt molecules to Frizzled/LEP transmembrane receptor complex results in dishevelled phosphorylation and inhibition of Axin, Gsk3b and APC complex, subsequently allowing β -catenin to accumulate in the nucleus and interact with transcription factors of the TCF/LEF family (Ricken et al., 2002; Gatcliffe et al., 2008; van 't Veer et al., 2002). Although mutations of APC and β -catenin are frequently observed in colorectal cancer (Vogelstein et al., 1988a; Sparks et al., 1998), they are rare in EOC (Ueda et al., 2001; Gatcliffe et al., 2008), with the exception of endometrioid subtypes (Wu et al., 2007). Among numerous non-canonical pathways, Wnt/planar cell polarity (PCP) signalling and Wnt/ Ca^{2+} signalling are involved in tumourigenesis (Miller et al., 1999; Kikuchi and Yamamoto, 2008; Wang, 2009).

In cancers where Wnt pathway gene mutations are rare, the Wnt pathway can become dysregulated through epigenetic mechanisms, such as promoter hypermethylation (Klarman et al., 2008; Valencia et al., 2009; Ying and Tao, 2009). The Wnt pathway extracellular inhibitors, such as SFRPs family and *WIF1*, are frequently silenced by promoter

hypermethylation in cancer including EOC (Takada et al., 2004; Veeck et al., 2006; Valencia et al., 2009; Yu et al., 2009; Bennett et al., 2010; Deng et al., 2010; Su et al., 2010). However, promoter methylation of the Wnt pathway has not been comprehensively studied in EOC. The aim of the study was to systematically profile DNA methylation at promoter CGIs of Wnt pathway genes and evaluate their role in tumour progression by assessing their prognostic value in EOC.

4.2 Methods

4.2.1 Patients

Fresh frozen tumour biopsies were prospectively collected from patients enrolled on an ongoing Scottish Gynaecology Clinical Trial Group (SGCTG)/National Cancer Research Institute (NCRI) study "DNA Methylation as a Predictor for Response and Progression-Free Survival in Patients with Ovarian Cancer". Tumours included in the present analysis were restricted to those from patients with confirmed FIGO stages Ic, III and IV epithelial ovarian cancer treated with cytoreductive surgery and platinum-based chemotherapy, excluding mucinous and clear cell tumours. These latter histological tumour types were not included due to their different clinical outcome from other more common serous and endometrioid ovarian tumours. Progression and survival status of the patient was examined 2 monthly for the first two years post-treatment, subsequently 6 monthly to 5 years and annually thereafter. The primary end point in this study is the association of DNA methylation with progression free survival (PFS), which is defined as the time after patients received first line chemotherapy to the first time that patients showed the progressive disease measured by RECIST 1.0 criteria, or had early death due to ovarian cancer or other causes. Information was collected, if available, on bulk of residual disease, stage, per-

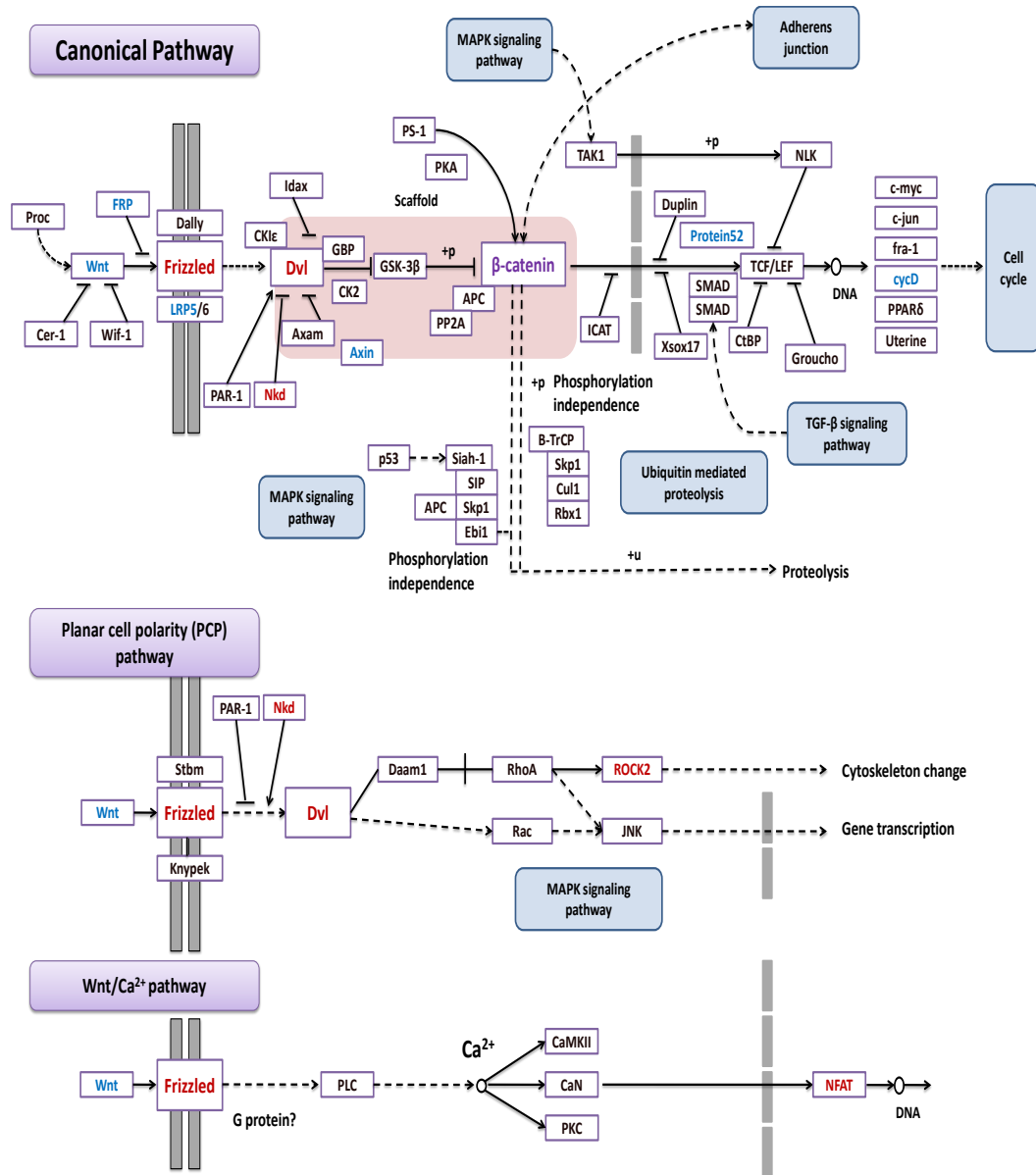


Figure 4.1: Schematic graph of the Wnt pathway. This graph is regenerated from Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway database (entry ID: hsa04310).

formance status, tumour grade, CA125 levels pre-treatment and after treatment, type of treatment received, response to treatment defined by RECIST 1.0 criteria, CA125 levels during treatment and overall survival, as defined as the time from start of chemotherapy to death due to ovarian cancer or other causes. This trial has been approved by the MREC for Scotland (reference number 01/165) and is carried out in accordance with the Declaration of Helsinki. Genomic DNA was extracted as previously described (Strathdee et al., 1999).

Differential Methylation Hybridisation (DMH)

Samples were assayed on Agilent custom-designed microarrays by DMH in duplicates as described in Method section in Chapter 3. Samples with more than 10% of the flagged probes in duplicates were excluded from further analysis. Otherwise, they were treated as missing data point (about 3%) and imputed by KNN algorithm ($k=10$). Probes with high (>65000) and low ($< mean + 1.96SD$ of negative controls) signal intensities were excluded in the analysis and intro-array normalisation was done using MLDA package tailored for DMH dataset (see Chapter 2). DMH ratio is the ratio of the signals from McrBC mock digested and digested samples. DMH dataset is available at GEO (accession ID: GSE23240).

4.2.2 qRT-PCR

qRT-PCR was done by Jenny Harsey at Institute for Cancer Research (Sutton, UK). Total cellular RNA was received from Gen-Probe Life Sciences and quantified by nanodrop at an optical density of $OD_{\lambda 260}$. Up to $1.2\mu g$ of RNA was reverse transcribed and cDNA synthesised in duplicate $20\mu l$ reactions using the Superscript II reverse transcriptase kit (Invitrogen) according to manufacturer's instructions. For qPCR individual TaqMan gene expression assays were obtained from Applied Biosystems, *FZD4* (assay no. Hs00201853.m1),

DVL1 (assay no. Hs00182896.m1), *NKDI* (assay no. Hs00263894.m1), *ROCK1* (assay no. Hs00178463.m1), *NFATC3* (assay no. Hs00190046.m1) and *GAPDH* (part no. 4326317E) were used. TaqMan Gene expression mastermix (Applied Biosystems) containing the enzyme AmpliTaq Gold was used for all reactions with the recommended cycling conditions. Relative expression analysis was performed according to the relative standard curve method.

4.2.3 Target Selection on Agilent Custom-designed Microarray

Genes involved in the Wnt signalling pathway were collected from Kyoto Encyclopaedia of Genes and Genomes (KEGG) (entry ID: hsa04310) (Kanehisa et al., 2006, 2010). Promoter CGIs within 2kb of the transcription start site of the genes were obtained from UCSC database (Rhead et al., 2010) and from a genome-wide prediction of CGIs (Bock et al., 2007) as a supplementary set, both of which fulfil the mathematical model proposed by Gardiner-Garden to define a CGI: CG content greater than 50%, length of sequence is over 200bp and ratio of observed to expected CpGs is larger than 0.6 (Gardiner-Garden and Frommer, 1987). The genome positions of targets are specified by Human Mar. 2006 (NCBI36/hg18) assembly. Out of 148 genes in the Wnt pathway, 137(92.6%) had a CpG island within 2Kb of the promoter and these are represented by 302 MseI fragments (130-6000bp) analysed by DMH.

Furthermore, the promoter CGIs of the genes involved in other signalling pathways/families (p53 AKT/mTOR, MMR, HR, NHEJ, BRCA1/2, FA family, igLon family and Redox) mainly collected from KEGG database were also targeted by the array. Evaluation and validation of prognostic values of the promoter CGIs in these pathways are detailed in Chapter 5.

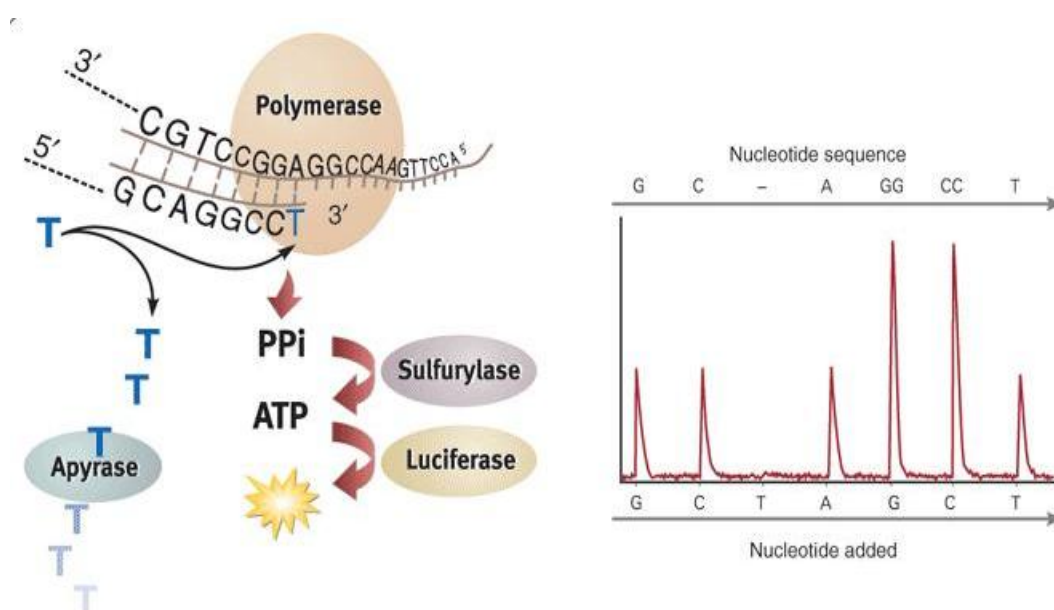


Figure 4.2: The principle of Pyrosequencing and the output ProgramTM. This figure is from England and Pettersson (2005).

4.2.4 Bisulfite Pyrosequencing

Bisulfite pyrosequencing is a quantitative measurement of methylation in CpG dinucleotides. The unmethylated cytosine (C) is converted into thymine (T) after bisulfite treatment, but the methylated cytosine (^mC) is intact. Pyrosequencing synthesizes the complementary strands of single-stranded DNA templates. In each incorporation of nucleotide, pyrophosphate (pp_i) is released and the light emission is proportional to the number of incorporations (Figure 4.2).

Methylation at promoter region of DVL1 (Chr1: 1275096-1275129, sequence: CGAGG CGTTTTTTGAGGTCGCGGGGCGTTCGCG, see Figure 4.3) was quantitated using bisulfite pyrosequencing. This sequence contains 7 CpG sites at chr1: 1275096, 1275102, 1275115, 1275117, 1275122, 1275126 and 1275128. Briefly, ovarian tumour genomic DNA was bisulfite treated by EZ DNA MethylationTM Kit (ZYMO research, CA, USA). Successful bisulfite modification was tested by CALPONIN-specific primer sets prior py-

pyrosequencing (Teodoridis et al., 2005). The bisulfite treated samples were amplified by PCR using forward primer GTAGGGTAGGGGAGATTT, reverse primer TCACTCCTC-CCATCCTAAAC and sequencing primer GGGTAGGGGAGATTTT. The PCR conditions were 95 °C 6 minutes, (95 °C 30 seconds, 60 °C 30 seconds and 72 °C 30 seconds) × 35 cycles, then 72 °C 5min. The PCR product is 151 base pair.

The Sepharose beads were used to purify the PCR product using a biotin-labelled primer. The sepharose beads containing PCR products were purified, denatured and washed using the Pyrosequencing Vacuum Prep Tool, then the sequencing primer was annealed to the PCR products. Pyrosequencing was done in PyroMarkTM MD system (Pyrosequencing, USA). The methylation was quantified as the percentage of methylated cytosine over the sum of methylated and unmethylated cytosines using Pyro Q-CpGTM software. The non-CpG cytosine was used as internal control to verify bisulfite conversion. Each sample had duplicates and the average of duplicates was used in the analysis.

4.2.5 Statistical Analysis

All the data analysis was done in R (2.10.1) and SPSS (version 16). The R scripts in this analysis are in Appendix scripts.

End point

The primary end point is progression free survival (PFS), which was measured by the time from treatment started to the progressive disease defined as either clinical evidence of progressive disease based on response evaluation criteria in solid tumours (RECIST 1.0) (Therasse et al., 2000) or elevated CA125 levels as defined previously (Rustin et al., 1996) and clinical or radiographic findings indicative of progression. PFS is considered as more objective and valuable end point to provide an estimate of biological effect of chemother-

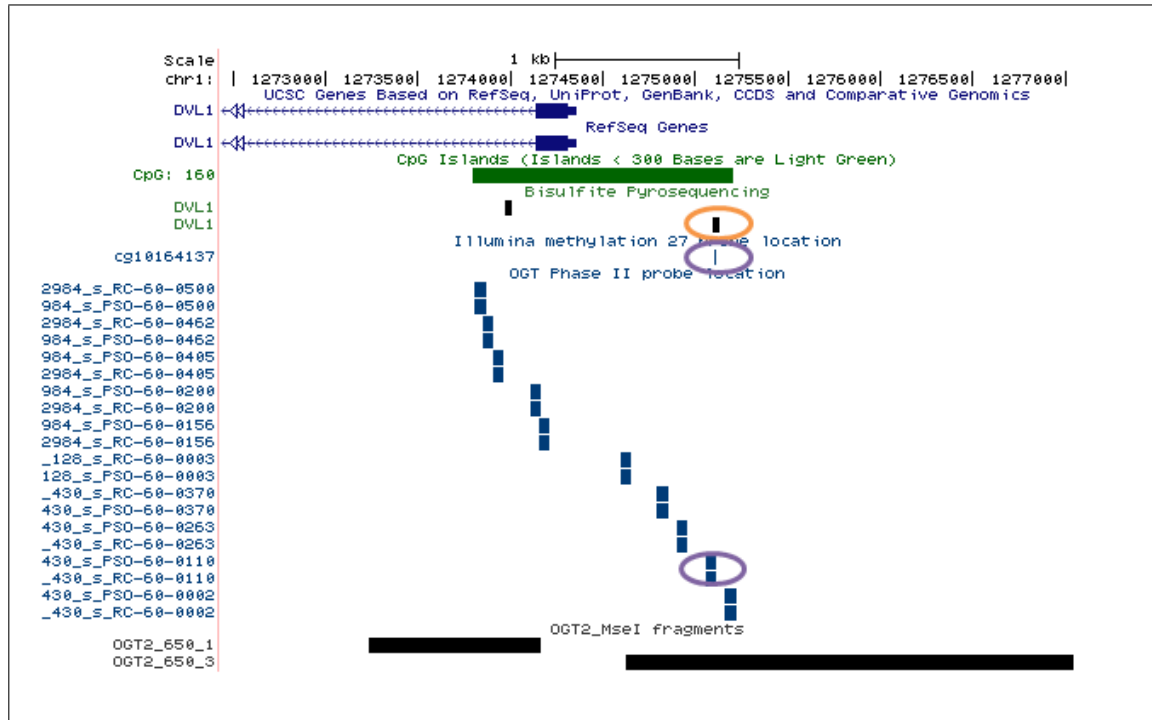


Figure 4.3: Genomic location of the region close to *DVL1* tested by bisulfite pyrosequencing, targeted by Agilent custom-designed microarray and Illumina HumanMethylation27. Purple circle: methylation of the region is correlated with PFS at $p < 0.05$ on custom-designed microarray (Chr1: 1275054-1275113) (see Result) and Illumina HumanMethylation27 BeadChip (Chr1: 1275054-1275103). Orange circle: CpG sites tested by bisulfite pyrosequencing (Chr1: 1275096-1275129).

apy, thus was used a primary end point in this analysis.

Cox proportional hazards model is used to examine the correlation between methylation (a continuous variable) and PFS, as shown in formula 4.1.

$$h(t) = h(t_0) \exp(\beta M) \quad (4.1)$$

Here, methylation is the explanatory variable (M). $h(t_0)$ is the baseline hazard function and corresponds to the probability of disease progression when the explanatory variable is zero. β is the regression coefficient giving the proportional changes that can be expected in the hazard, related to the changes of explanatory variable. Hazard ratio is $\frac{h(t)}{h(t_0)}$, indicating the relative hazard of disease progression in every unit increase of methylation.

The secondary end point is response measured by RECIST 1.0 criteria (Therasse et al., 2000) and categorised as complete response, partial response, progressive disease and stable disease (see Table 1.2 in Chapter 1).

Logistic regression model is used to test the relationship between methylation and response, as shown in formula 4.2.

$$\frac{P}{1-P} = \exp(\alpha + \beta M) \quad (4.2)$$

Methylation is the explanatory variable (M). α is the intercept indicating the baseline performance when explanatory variable is zero. β is the regression coefficient representing the change of probability of outcome (poor response in this case). $\exp(\beta)$ is the odd ratio corresponding to a one unit increase of methylation.

Power estimation in the screening study

With PFS as the end point, the Cox proportional hazard model was used to assess the correlation between methylation and survival time. Therefore, we estimated the required sample size for the Cox model using the formula 4.3 and 4.4 described before (Schoenfeld, 1983; Tong and Zhao, 2008).

$$N = \frac{(\mu_{1-\alpha/2} + \mu_{1-\beta})^2}{(\log\theta)^2 \varphi(1-p)p} \quad (4.3)$$

$$FDR = \frac{\pi\alpha}{\pi_0\alpha + (1 - \pi_0)(1 - \beta)} \quad (4.4)$$

Here, μ_γ indicates the γ -quantile of the standard normal distribution; θ is the hazard ratio; φ is the probability of an uncensored observation; p is the methylation frequency of loci in ovarian cancer; π_0 is the proportion of true null hypotheses; α is the significant level and $1 - \beta$ is the average power expected on the microarray.

The initial screening set consisted of DMH data from 120 tumours, with 111 subsequently shown to have suitable data quality and 102 patients had disease progression (see results). To estimate approximate statistical power of this screening set prior to analysis, we assumed 5% of the loci examined in the Wnt pathway were true positives and split patients into two groups based on the median methylation level at each locus (in the subsequent analysis methylation levels have been treated as a continuous variable, meaning we are underestimating the study power). With a hazard ratio at 1.75 and false discovery rate (FDR)(Benjamini and Hochberg, 1995) less than 50%, the estimated average power of the screening study is 0.8 (Table 4.1).

power ($1 - \beta$)	FDR	true positive ($1 - \pi_0$)	significant level (α)	uncensored (φ)	Hazard ratio (θ)	methylation frequency (p)	Sample size (N)
0.9	0.5	0.05	0.05	0.92	1.75	0.5	148
0.8	0.5	0.05	0.04	0.92	1.75	0.5	111
0.7	0.5	0.05	0.04	0.92	1.75	0.5	95

Table 4.1: Estimated power of the analysis in the screening study

Power estimation in the validation study

The subsequent validation study was independently analysed by DMH a year later and was based on the maximum feasible within existing financial constraints. The power of analysis for each biomarker was estimated using the formula 4.4 and 4.5 (Hsieh and Lavori, 2000).

$$Z_{1-\beta} = \sqrt{N\sigma^2(\log\theta)^2\varphi} - Z_{1-\alpha} \quad (4.5)$$

Here, σ^2 is the variance of DMH ratios of individual locus, θ is the hazard ratio of per unit increase of DMH ratio calculated from univariate Cox model using the locus as the covariate. Both parameters were estimated in the evaluation study. φ is the percentage of uncensored data and N is the number of samples. $Z_{1-\alpha}$ and $Z_{1-\beta}$ are standard normal deviates at the desired one-sided significance level α and power $1 - \beta$. Calculation of the statistical power for the sample size planned ($n = 50$) assumed the true positive loci rate was 40% (following on from the FDR rate observed in the screening set at $p < 0.05$) and using a FDR of 10%. Sample size calculations were carried out using the observed hazard ratios per unit change and observed variable variances from the screening data set. Testing was one-sided as associations of interest will be in the same direction as in the screening set. Table 4.2 gives the estimated power for individual locus. The validation power achieved for the largest observed effect size in the screening set was 88%, but only 30% for the median observed effect size. Among the 5 loci with FDR less than 10% (see

results) from the screening set the median power was 70%.

genes	Hazard ratio (Δ)	Variance (σ^2)	Squared standardised effect size ($\sigma^2 \log(\Delta)^2$)	Significant level (α)	power ($1 - \beta$)
FZD4	123.3	0.01	0.14	0.05	0.73
CCND1 [§]	3.9	0.11	0.21	0.07	0.88
NKD1	31.4	0.01	0.1	0.04	0.57
TAF8/CCND3	0.5	0.24	0.12	0.05	0.64
DVL1	12.1	0.02	0.13	0.05	0.69
FRAT1	0.3	0.05	0.08	0.03	0.39
ROCK1	2263.1	<0.01	0.09	0.04	0.49
PPP2R2B	34	<0.01	0.05	0.01	0.19
CTBP1 [§]	2.3	0.1	0.07	0.03	0.37
CCND1 [§]	48.6	<0.01	0.07	0.03	0.34
NKD2	10.6	0.01	0.07	0.03	0.35
FZD9	7.4	0.01	0.06	0.02	0.27
EEFSEC/RUVBL1	2.4	0.06	0.05	0.01	0.16
AXIN1	24.5	<0.01	0.05	0.01	0.17
C4orf42/CTBP1 [§]	1.9	0.13	0.05	0.02	0.22
WNT4	41.1	<0.01	0.05	0.01	0.17
LRP5	5.7	0.01	0.04	0.01	0.11
NFATC3	6.4	0.01	0.05	0.01	0.19
FRAT1	0.7	0.47	0.06	< 0.0005	< 0.007
SFRP5	51.2	<0.01	0.04	0.01	0.08

Table 4.2: Estimated power of the analysis in the validation study. [§]Different loci within the promoter region of *CCND1* and *CTBP1* were identified.

Determination of methylation clusters

Methylation of loci were defined by DMH ratios of the probes targeting the same locus being significantly greater than unmethylated regions selected from chromosome 16 and the mitochondrial genome. One-side Mann-Whitney U test was used to test the difference and the significance level was set at $p < 0.0003$ based on Bonferroni correction (see Mann-Whitney Test in section 3.2.4 of Chapter 3).

Unsupervised hierarchical clustering was performed using methylation status of MseI fragments in each pathway using Euclidean distance and average linkage method. The significance of the clusters was tested by bootstrap resampling method as previously described (Shimodaira, 2004). Pearson Chi-square test was used to examine the dependency between categorical clinical parameters (stage, grade, histology, response, and chemotherapy) and methylation clusters. One-way ANOVA was used to examine the difference of age in clusters. PFS and OS of the patients in the clusters were tested by log rank test. The significance level was set at two-sided $p \leq 0.05$.

Survival analysis

The workflow of the survival analysis is illustrated in Figure 4.4 (Survival Analysis). Cox proportional hazard model was used to examine for the association between methylation in each pathway and patient survival. The average DMH ratios were used as continuous variable in the Cox model. The variables applied for adjustment in multivariate analysis included histological type, FIGO stage, grade and age with potential confounding effects. The impact of DNA methylation on patients' progression free survival was determined by examining the relative hazard ratios with the increase of methylation level. The significance of estimated hazard ratios was tested by Score test.

In univariate survival analysis, the significant level was considered at the 5% level for clinical parameters/individual biomarker and at the 10% FDR for multiple test correction in the screening and validation study.

Kaplan-Meier curves were used to show the progression free survival in the patients with high/low methylation at promoter CGI, as well as those with high/low methylation

index (MI) (see "Construction of the multivariate Cox model"). Since average methylation frequency of the Wnt pathway genes observed is approximately 25%, the third quartile was used as the cut-point for high/low methylation. Kaplan-Meier survival curves of two groups were compared by the log-rank test (two-sided). The significance of log-rank test of two groups with high/low MI was assessed by a permutation test applied to the entire variable selection and model fitting process.

The correlation between response and promoter methylation in the Wnt pathway was tested by logistic regression. This is restricted to patients with measurable disease at baseline. Patients were classified as responders (complete or partial response) or non-responders (stable disease or progressive disease) according to RECIST 1.0 criteria.

Construction of the multivariate Cox model

The DMH data of late-stage tumours from the screening and validation study were standardised across the samples within the study as Z scores. The variance of the DMH data within the study was estimated by trimming 1% of the data at both sides, which is a robust estimation of the variance accommodating the outlier observations. A multivariate Cox model was constructed using Z score as a continuous variable. The covariates in the final model were selected by likelihood ratio forward stepwise algorithm from clinical parameters (stage, grade, histology and age), and those DNA methylation biomarkers identified from the univariate PFS analysis, which were independent from clinical parameters and have been validated in an independent patient cohort. The likelihood ratio forward stepwise method used maximum likelihood ratio test $p < 0.05$ for entry and $p > 0.1$ for removal of the variables from the model. Therefore, only the variables with the strongest association with PFS and independent contribution to the model were selected into the fi-

nal model, while those variables that did not significantly improve the model were left out. The variables included and excluded from the final model were shown in Appendix 4-1. Subsequently, a methylation index (MI) was calculated from the multivariate Cox model. The predictive value of MI and degree of over-fitting of the multivariate Cox model was estimated by LOOCV as described in the next section. The workflow of this analysis in the Wnt pathway was illustrated in Figure 4.4 (Model Construction).

The Receiver Operator Characteristic (ROC) curves considering the censored data were generated as described before (Heagerty et al., 2000) to evaluate the sensitivity and specificity of MI and the prognostic index (PI) estimated from the prognostic factors (histology type, stage, age and grade) that have been used in clinical practice.

Leave-one-out cross validation and permutation test

Leave-one-out cross validation (LOOCV) (Verweij and Van Houwelingen, 1993) was used to estimate the predictive value of our final multivariate Cox model. In each iteration of LOOCV, one sample was left out and the multivariate Cox model was constructed by the rest of the samples using exactly the same process as our modelling procedure for feature selection and model construction. Given both DNA methylation biomarkers selected into the final Cox model had $FDR < 10\%$, we simplified the feature selection process in LOOCV using $FDR < 10\%$ as the significant cut-point, therefore, in each iteration of model construction, only the biomarkers with $FDR < 10\%$ and independent from clinical parameters were considered to be included in the multivariate model construction and likelihood ratio forward stepwise algorithm was used to select the significant and independent covariates into the model.

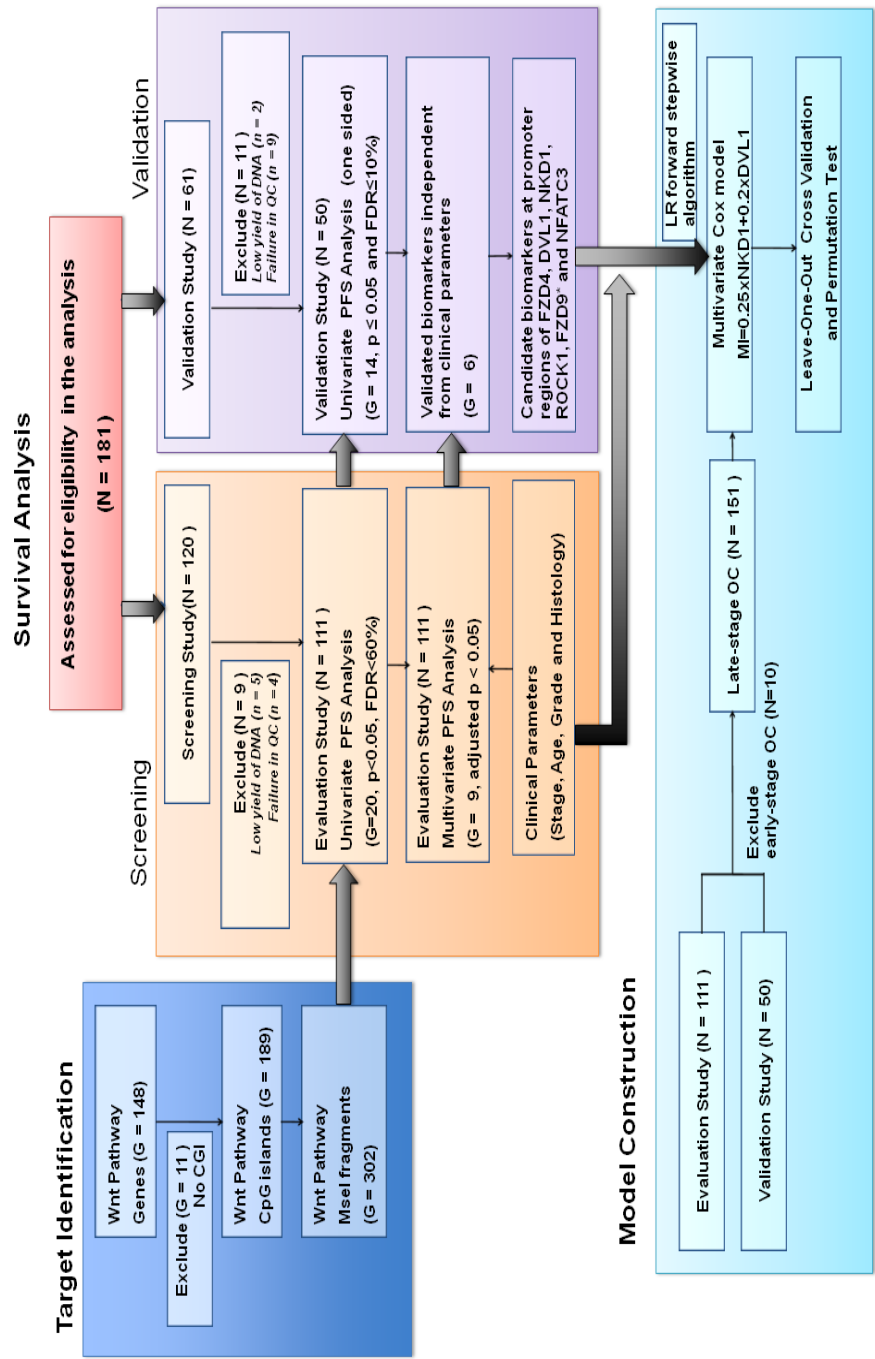


Figure 4.4: Workflow for identification of prognostics biomarkers in the Wnt pathway. N: number of tumour samples; G: number of biomarkers. OC: ovarian cancer. LR: likelihood ratio. *FZD9 was not considered in the model construction since in the combining analysis of screening and validation set, patients with high methylation at promoter region of this gene were not significantly distinct from those with low methylation ($p = 0.063$).

The coefficients in the multivariate Cox model were used to calculate a MI for each sample. In LOOCV, the omitted sample was classified as high risk of progressive disease or low risk based on its MI predicted from the multivariate Cox model constructed using the remaining samples above or below the third quartile of the methylation indices for the remaining cases in the training set. The entire procedure was repeated until each case had been omitted once and had been assigned into a risk group. Therefore, this cross-validated risk group for each case was determined based on a model that did not use that case in the model construction.

Kaplan-Meier survival curve was generated for the predicted risk groups for progressive disease and log-rank statistic was computed. To assess the significance of log-rank statistic and the degree of over-fitting, we performed a phenotype permutation test 500 times by shuffling among the cases and the entire cross-validation process described above was repeated. For each random shuffling, the predicted high/low risk groups were generated as described above and log-rank statistics was calculated between two risk groups. Permutation test provided a null-distribution of log rank test, so the probability that the log rank statistics in the null distribution are larger than the log rank statistic between the risk groups predicted by LOOCV without cases shuffling is the chance we obtain this log rank statistic randomly.

4.2.6 The Cancer Genome Atlas (TCGA) Dataset

The level 3 expression dataset on Affymetrix HGU133A microarrays of serous tumours were obtained from TCGA data portal (<http://cancergenome.nih.gov/dataportal>). The microarray data have been pre-processed and normalised across the samples (see http://tcga-data.nci.nih.gov/docs/TCGA_Data_Primer.pdf).

4.3 Results

4.3.1 Patients

Epithelial ovarian tumours prospectively collected through a cohort study, were analysed by differential methylation hybridisation (DMH) at 302 Wnt pathway associated loci. Mucinous and clear cell cancers were excluded due to their different clinical outcome from more common serous and endometrioid EOC (Sugiyama et al., 2000; Rosen et al., 2009). In an initial screening stage, 120 ovarian tumour samples, sequentially collected in a prospective cohort study were used. Four samples were excluded from the analysis due to poor quality of signal intensities in over 10% of probes in the duplicates and five samples were excluded as methylation controls did not reach acceptance criteria. Therefore, 111 tumours remained in the analysis. In a subsequent validation stage, 61 further tumours collected through the same protocol were examined. Eleven samples were excluded due to not reaching acceptance criteria, so 50 tumours remained in the analysis. Among the loci with $FDR < 10\%$ from the screening set the median power was 70%. In total, we examined DNA methylation at CGIs at the promoter regions of Wnt pathway genes, as defined by KEGG (entry ID: hsa04310)(Kanehisa et al., 2006, 2010), in 161 EOC. The genes in the Wnt pathway frequently methylated ($>5\%$) in ovarian cancer and unmethylated in PBMC are shown in Appendix 4.2. Following the REMARK recommendations (McShane et al., 2005), the full details of clinical parameters in Table 4.3 and their relationship to patient outcome are presented in Table 4.4 and 4.5.

Using unsupervised hierarchical clustering of methylation status, two groups of patients ($n_1 = 17$, $p = 0.05$ and $n_2 = 15$, $p < 0.001$) were identified, weakly correlated with stage ($p = 0.071$), but not associated with grade ($p = 0.156$), histology ($p = 0.106$)

Clinical Parameters		Screening N=111 (%)	Validation N=50 (%)
Age	Range	(32,83)	(44,86)
	mean (95% CI), Years	61 (59,63)	64 (61,66)
FIGO Stage	Ic	2 (1.8)	8 (16.0)
	III	85 (76.6)	35 (70.0)
	IV	24 (21.6)	7 (14.0)
Grade	G1	4 (3.6)	4 (8.0)
	G2	16 (14.4)	10 (20.0)
	G3	71 (64.0)	29 (58.0)
	unknown	20 (18.0)	7 (14.0)
Histology	Serous	42 (37.8)	19 (38.0)
	Endometrioid	8 (7.2)	3 (6.0)
	Other	35 (31.5)	28 (56.0)
	Unknown	26 (23.4)	0
Chemotherapy	Platinum	26 (23.4)	19 (38.0)
	Platinum + Taxane	67 (60.4)	24 (48.0)
	Other	18 (16.2)	3(6.0)
	missing	0	4 (8.0)
Response	CR	27 (24.3)	2 (4.0)
	PR	20 (18.0)	5 (10.0)
	SD	13 (11.7)	3 (6.0)
	PD	8 (7.2)	5 (10.0)
	NE	14 (12.6)	20 (40.0)
	missing	29 (26.1)	15 (30.0)
Overall survival (OS)	Alive/censored	37 (33.3)	27 (54.0)
	Dead	74 (66.7)	19 (38.0)
	missing	0	4 (8.0)
	Median (95% CI), Years	2.52 (1.67,3.37)	2.79 (2.17, 3.41)
Progression free survival (PFS)	Not progressed	9 (8.1)	14 (28.0)
	Progressed	102 (91.9)	32 (64.0)
	missing	0	4 (8.0)
	Median (95% CI), Years	0.92 (0.76,1.08)	1.11 (0.79, 1.43)

Table 4.3: Patient characteristics. Response was measured by RECIST 1.0; the patients with complete/partial response are the responders to chemotherapy, while the patients with progressive or stable disease are those who did not response to chemotherapy CR: complete response; PR: partial response; SD: stable disease; PD: progressive disease; NE: not measurable

Variable	No. of cases	Disease progression (%)	Median PFS (95% CI), years	1-Year PFS Rate (95% CI), %	HR ¹	95% CI ²	p value
Stage							0.1029 ³
Stage Ic	10	0	NR	100	NR	NR	
Stage III	120	103 (86)	0.92 (0.78, 1.06)	45 (35, 54)	1.00		
Stage IV	31	31 (100)	0.90 (0.67, 1.13)	39 (22, 56)	1.40	(0.93, 2.10)	0.1050
Histological type							0.0108** ³
serous	61	52 (85)	0.94 (0.72, 1.16)	48 (35, 61)	1.00		
endometrioid	11	7 (64)	1.58 (0.74, 2.42)	64 (35, 92)	0.51	(0.23, 1.12)	0.0929
Other	63	49 (78)	1.00 (0.80, 1.20)	49 (36, 62)	0.86	(0.58, 1.28)	0.4527
unknown	26	26 (100)	0.64 (0.47, 0.81)	31 (13, 49)	1.66	(1.03, 2.66)	0.0366*
Grade							0.0553 ³
G1	8	5 (63)	1.58 (0.45, 2.71)	63 (29, 96)	1.00		
G2	26	20 (77)	0.98 (0.40, 1.56)	47 (27, 66)	1.96	(0.73, 5.23)	0.1808
G3	100	84 (84)	1.00 (0.77, 1.23)	48 (38, 58)	1.77	(0.71, 4.37)	0.2185
unknown	27	25 (93)	0.92 (0.28, 1.56)	36 (17, 55)	2.97	(1.13, 7.78)	0.0271*
Age	161	134 (83)	0.98 (0.85, 1.11)	46 (46, 47)	1.01	(0.99, 1.03)	0.2185 ³
Chemotherapy							0.1337 ³
Platinum	45	40 (89)	0.88 (0.59, 1.17)	40 (26, 54)	1.00		
Platinum+Taxane	91	74 (81)	0.70 (0.93, 1.21)	52 (41, 63)	0.73	(0.49, 1.07)	0.1050
other	21	17 (81)	0.92 (0.48, 1.36)	47 (23, 71)	1.11	(0.63, 1.96)	0.7290
Response ⁴							
Responders (CR+PR)	54	51 (94)	1.08 (0.91, 1.25)	56 (43, 69)	1.00		
Non-responders (PD+SD)	29	27 (93)	0.58 (0.14, 1.02)	17 (3, 31)	2.55	(1.56, 4.17)	0.0001***

Table 4.4: Univariate PFS analysis of clinical parameters. Abbreviations: HR, hazard ratio; CI, confidence interval; NR, not reached; PFS, progression free survival.¹Hazard ratio estimated from Cox proportional hazard regression model.²Confidence interval of the estimated hazard ratio.³p value of Score test.⁴ Response was measured by RECIST 1.0; the patients with complete/partial response are the responders to chemotherapy, while the patients with progressive or stable disease are those who did not response to chemotherapy. ⁺ $p < 0.1$; ^{*} $p \leq 0.05$; ^{**} $p \leq 0.01$

Variable	No. of cases	Death, %	Median OS (95% CI), years	1-Year OS Rate (95% CI), %	HR	95% CI	p value
Stage							0.0566 ⁺³
Stage Ic	10	0	NR	100	NR	NR	
Stage III	120	67 (56)	2.81 (2.35, 3.27)	78 (67, 86)	1.00		
StageIV	31	26 (84)	1.53 (0.77, 2.29)	74 (58, 89)	1.56	(0.98, 2.48)	0.0586 ⁺
Histological type							0.1135
serous	61	34 (56)	2.75 (1.96, 3.54)	83 (73, 92)	1.00		
endometrioid	11	3 (27)	NR	91 (74, 100)	0.42	(0.13, 1.38)	0.1530
Other	63	34 (54)	2.81 (1.42, 4.20)	79 (68, 89)	1.01	(0.63, 1.63)	0.957
unknown	26	22 (85)	1.48 (0.44, 2.52)	64 (45, 83)	1.55	(0.89, 2.70)	0.126
Grade							0.0153 ^{*3}
G1	8	3 (38)	2.66 (1.92, 3.41)	88 (65, 100)	1.00		
G2	26	14 (54)	2.05 (0.28, 3.82)	66 (46, 85)	2.35	(0.67, 8.29)	0.1825
G3	100	57 (57)	3.00 (2.24, 3.76)	84 (77, 92)	1.62	(0.50, 5.21)	0.4218
unknown	27	19 (70)	1.31 (0.83, 1.79)	68 (49, 86)	3.84	(1.02, 11.92)	0.0466
Age	161	93 (58)	2.66 (2.02, 3.30)	77 (71, 83)	1.02	(1.00, 1.04)	0.0625 ⁺³
Chemotherapy							0.0306 ^{*3}
Platinum	45	29 (64)	1.53 (0.67, 2.40)	70 (56, 84)	1.00		
Platinum+Taxane	91	50 (56)	3.00 (2.71, 3.29)	87 (80, 94)	0.57	(0.36, 0.90)	0.0170 [*]
other	21	10 (48)	3.58 (1.08, 6.08)	74 (52, 96)	0.47	(0.22, 1.03)	0.0584 ⁺
Response ⁴							0.0462 ^{*3}
Responders (CR+PR)	54	33 (61)	3.05 (2.49, 3.61)	88 (78, 98)	1.00		
Non-responders (PD+SD)	29	18 (62)	1.51 (0.3, 3.45)	72 (56, 88)	1.83	(1.01, 3.33)	0.0464 [*]

Table 4.5: Univariate overall survival analysis of clinical parameters. Abbreviations: HR, hazard ratio; CI, confidence interval; NR, not reached; PFS, progression free survival.¹Hazard ratio estimated from Cox proportional hazard regression model.²Confidence interval of the estimated hazard ratio.³p value of Score test.⁴ Response was measured by RECIST 1.0; the patients with complete/partial response are the responders to chemotherapy, while the patients with progressive or stable disease are those who did not response to chemotherapy. ⁺ $p < 0.1$; $*p \leq 0.05$; $**p \leq 0.01$

and age ($p = 0.726$). However, two groups of patients have significant difference in PFS ($p = 0.007$, log rank test) and borderline difference in OS ($p = 0.048$, log rank test) (Figure 4.5).

4.3.2 Methylation of the Wnt pathway and PFS

To evaluate the association between methylation at promoter CGIs of the Wnt pathway and tumour progression, we used average DMH ratios as a continuous variable in the univariate Cox model with PFS as the endpoint (see methods). In the screening study we identified methylation of 20/302 loci at promoter CGIs of 17 genes that may be associated with PFS ($p < 0.05$), with five loci at genes *FZD4*, *DVL1*, *CCND1*, *CCND3* and *NKDI* significant at $FDR < 10\%$ after multiple test correction (see univariate survival analysis in Table 4.6). In an independent patient cohort ($n = 50$), 14 out of the 20 loci identified in the evaluation stage were still prognostic, $p \leq 0.05$ and $FDR \leq 10\%$ (Table 4.7).

The odds ratios of 14 loci identified as having $p \leq 0.05$ and $FDR \leq 10\%$ in univariate analysis from the validation study were adjusted by age, stage, grade and histological type and the patients were stratified into three groups who either received platinum alone, combination of platinum and taxane, or other platinum-based treatment. Hypermethylation at 6 loci was significantly associated with poor PFS independent from clinical parameters (adjusted $p < 0.05$): CGIs at *FZD4*, *DVL1*, *ROCK1*, *NFATC3*, *FZD9* and *NKDI* (see multivariate survival analysis in Table 4.8 and Figure 4.6A-F).

In addition to the association with PFS observed, the patients with higher methylation level at gene *DVL1*, *FZD4* and *NKDI* also had higher risk of death ($p < 0.05$, $FDR \leq 20\%$) (Table 4.9).

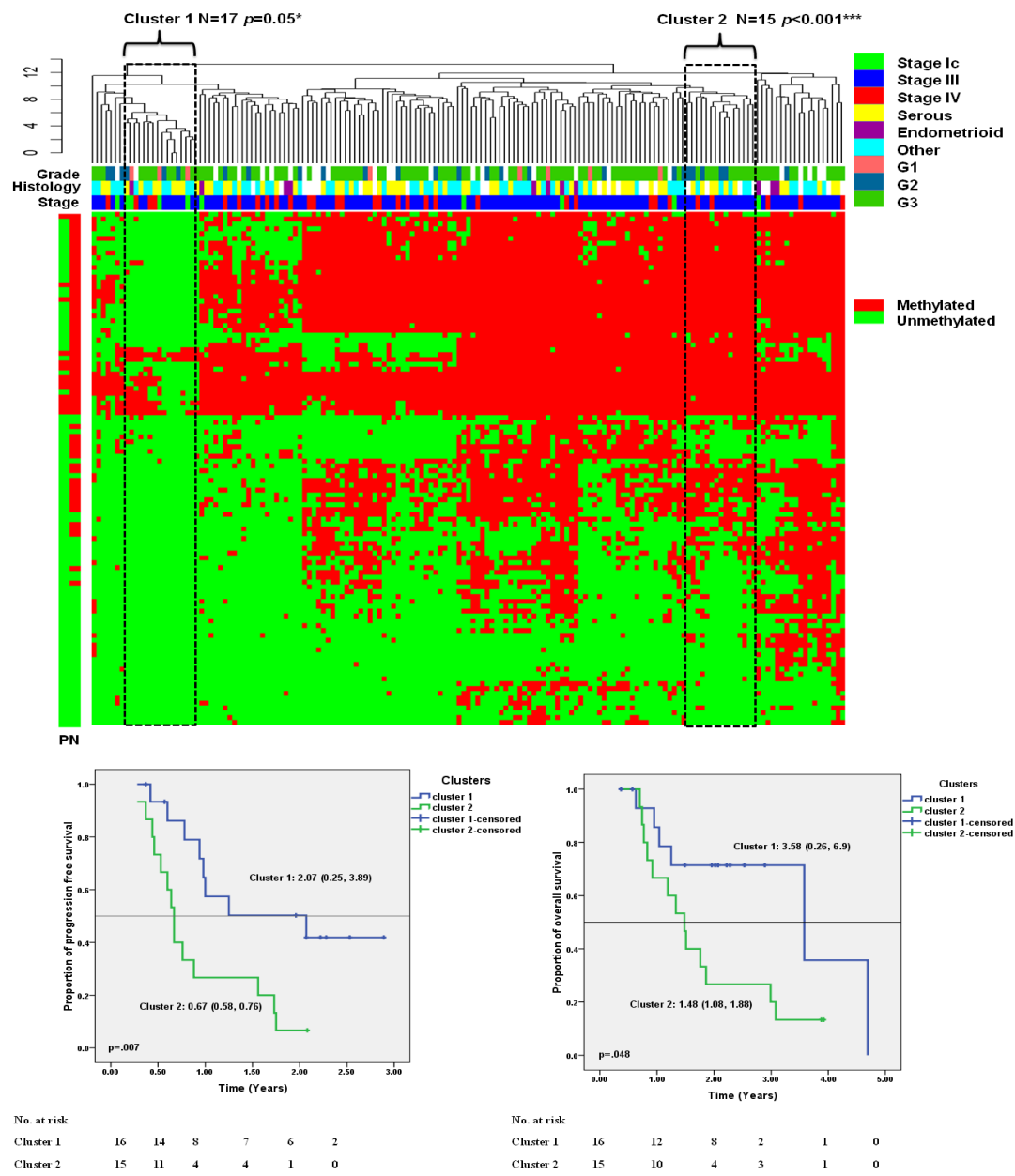


Figure 4.5: Unsupervised hierarchical clustering of binary DNA methylation data in 106 frequently methylated MseI fragments (row) in the Wnt pathway of 161 epithelial ovarian tumours (column). Red: methylated fragment; Green: unmethylated fragment. P: Peripheral Blood Mononuclear Cell (PBMC); N: adjacent samples of benign tumours. Distance: Euclidean distance; Linkage method: average linkage. Methylation frequency of MseI fragments are from 5% to 95% and size are from 130bp to 2500bp.

genes	Univariate PFS analysis (n=111)			
	HR ¹	95% CI ²	p value ³	FDR ⁴
FZD4	123.3	(8.9, 1701.30)	<0.001***	<0.1 ^{&}
CCND1 ^{\$}	3.9	(1.8, 8.3)	<0.001***	<0.1 ^{&}
NKD1	31.4	(3.7, 268.3)	0.002**	<0.1 ^{&}
TAF8 CCND3	0.5	(0.3, 0.8)	0.002**	<0.1 ^{&}
DVL1	12.1	(2.6, 55.5)	0.001**	<0.1 ^{&}
FRAT1	0.3	(0.1, 0.7)	0.005**	0.2
ROCK1	2263.1	(7.5, 681782.4)	0.008**	0.3
PPP2R2B	34.0	(2.4, 476.6)	0.009**	0.4
CTBP1 ^{\$}	2.3	(1.2, 4.3)	0.015*	0.4
CCND1 ^{\$}	48.6	(2.1, 1108.6)	0.015*	0.4
NKD2	10.6	(1.5, 74.1)	0.017*	0.5
FZD9	7.4	(1.4, 41.0)	0.021*	0.5
EEFSEC RUVBL1	2.4	(1.2, 5.2)	0.021*	0.5
AXIN1	24.5	(1.4, 423.3)	0.028*	0.6
C4orf42 CTBP1 ^{\$}	1.9	(1.1, 3.3)	0.030*	0.6
WNT4	41.1	(1.3, 1282.7)	0.034*	0.6
LRP5	5.7	(1.1, 30.2)	0.041*	0.6
NFATC3	6.4	(1.1, 37.6)	0.042*	0.6
FRAT1	0.7	(0.6, 1.0)	0.040*	0.6
SFRP5	51.2	(1.1, 2386.6)	0.045*	0.6

Table 4.6: Univariate progression free survival analysis of selected promoter methylation of Wnt pathway in the screening set ($n = 111$). Abbreviations: HR, hazard ratio; CI, confidence interval; FDR, false discovery rate. ¹Hazard ratio per unit increase in DMH ratio (continuous variable) estimated from Cox proportional hazard regression model. ²Confidence interval of the estimated hazard ratio. ³p value of Score test (two-sided). ⁴multiple test correction based on 302 loci. [&] $FDR \leq 0.1$; $*p \leq 0.05$; $**p \leq 0.01$; $***p \leq 0.001$. ^{\$}Different loci within the promoter region of CCND1/CTBP1 were identified

genes	Univariate PFS analysis (n=50)			
	HR ¹	95% CI ²	p value ³	FDR ⁴
FZD4	2.7	(1.0, 7.0)	0.02*	<0.1 ^{&}
CCND1 [§]	1.2	(0.5, 3.1)	0.335	0.3
NKD1	2.3	(1.0, 5.3)	0.025*	<0.1 ^{&}
TAF8 CCND3	1.3	(0.9, 1.8)	0.07	0.1 ^{&}
DVL1	4.3	(1.3, 13.8)	0.01*	<0.1 ^{&}
FRAT1	1.3	(0.7, 2.1)	0.2	0.2
ROCK1	4.8	(1.1, 20.4)	0.015*	<0.1 ^{&}
PPP2R2B	1.9	(0.1, 27.7)	0.315	0.3
CTBP1 [§]	1.4	(0.8, 2.4)	0.105	0.1 ^{&}
CCND1 [§]	311.9	(2.7, 36132.3)	0.01*	<0.1 ^{&}
NKD2	2.1	(0.9, 4.6)	0.04*	<.01 ^{&}
FZD9	1.8	(0.9, 3.4)	0.04*	<0.1 ^{&}
EEFSEC RUVBL1	1.7	(1.1, 2.7)	0.01*	<0.1 ^{&}
AXIN1	2.5	(1.0, 6.5)	0.025*	<0.1 ^{&}
C4orf42 CTBP1 [§]	1.6	(0.93,2.6)	0.045*	<0.1 ^{&}
WNT4	3.2	(1.23,8.55)	0.01*	<0.1 ^{&}
LRP5	2.7	(1.11,6.53)	0.02*	<0.1 ^{&}
NFATC3	6.6	(1.46,29.3)	0.005**	<0.1 ^{&}
FRAT1	1.2	(0.94,1.49)	0.08	0.1 ^{&}
SFRP5	2.4	(0.99,5.84)	0.025*	<0.1 ^{&}

Table 4.7: Univariate progression free survival analysis of selected promoter methylation of Wnt pathway in the validation set ($n = 50$). Abbreviations: HR, hazard ratio; CI, confidence interval; FDR, false discovery rate. ¹Hazard ratio per unit increase in DMH ratio (continuous variable) estimated from Cox proportional hazard regression model. -²Confidence interval of the estimated hazard ratio. -³p value of Score test (one-sided). -⁴multiple test correction based on 20 loci. [&] $FDR \leq 0.1$; * $p \leq 0.05$; ** $p \leq 0.01$.
[§]Different loci within the promoter region of CCND1/CTBP1 were identified

genes	Multivariate PFS analysis (n=111)		
	HR ¹	95% CI ²	adjusted p value ³
FZD4	37.6	(2.0, 694.6)	0.015*
NKD1	14.7	(1.3, 168.2)	0.031*
DVL1	7.5	(1.5, 37.2)	0.014*
ROCK1	3092.3	(5.8, 1650876.5)	0.012*
CCND1	5.4	(0.1, 219.4)	0.376
NKD2	5.6	(0.8, 39.3)	0.82
FZD9	6.4	(1.1, 36.4)	0.036*
EEFSEC RUVBL1	2.2	(0.9, 5.6)	0.105
AXIN1	3.4	(0.1, 104.7)	0.48
C4orf42 CTBP1	1.0	(0.5, 2.0)	0.901
WNT4	4.5	(0.1, 170.0)	0.414
LRP5	3.1	(0.6, 18.0)	0.198
NFATC3	12.6	(2.0, 78.6)	0.007***
SFRP5	9.7	(0.1, 964.1)	0.333

Table 4.8: Multivariate progression free survival analysis of selected promoter methylation of Wnt pathway in the screening set ($n = 111$). Abbreviations: HR, hazard ratio; CI, confidence interval. ¹Hazard ratio per unit increase in DMH ratio (continuous variable) estimated from Cox proportional hazard regression model. ²Confidence interval of the estimated hazard ratio. ³p value adjusted by histology, grade, stage and age. Age was used as a continuous variable and histology, grade and stage were used as categorical variables. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.005$.

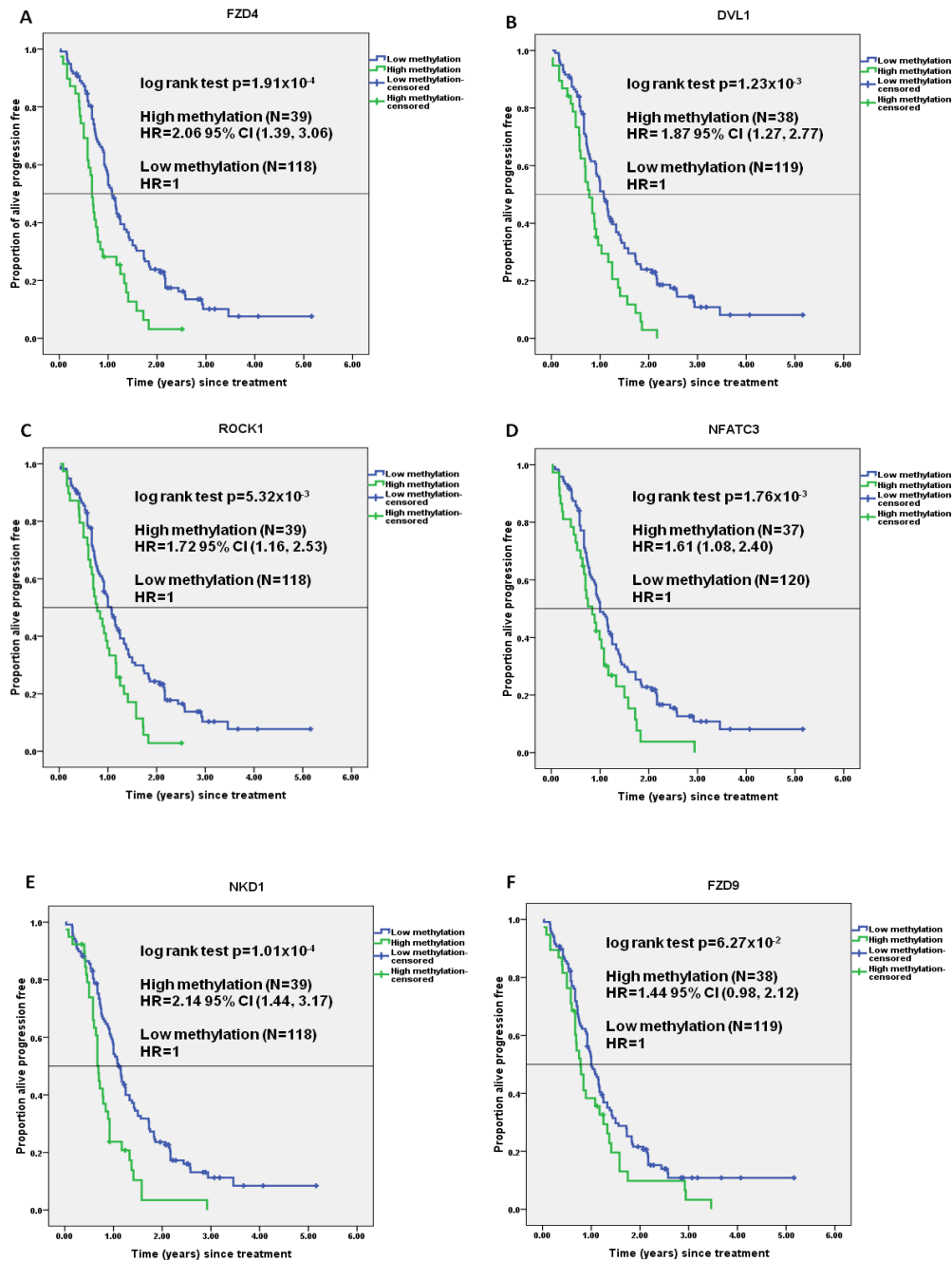


Figure 4.6: Kaplan-Meier plots of progression free survival. High/low methylation at A, *FZD4* CGI; B: *DVL1* CGI; C: *ROCK1* CGI; D: *NFATC3* CGI; E: *NKD1* CGI; F: *FZD9* CGI. The cut-off was determined by the third quartile in 157 patients ($n=161$, 4 patients have PFS missing). Time is from the patients received the first line of chemotherapy.

genes	Univariate Overall Survival (OS) (N=111)			
	HR ¹	95% CI ²	p value ³	FDR ⁴
FZD4	49.4	(2.5, 964.3)	0.01*	0.1 ^{&}
NKD1	21.3	(1.5, 299.2)	0.023*	0.2 ^{\$}
DVL1	14.0	(2.6, 75.0)	0.002***	<0.1 ^{&}
ROCK1	8.6	(0, 5642.9)	0.516	1
NKD2	1.3	(0.2, 11.2)	0.804	1
FZD9	3.3	(0.5, 23)	0.224	1
CCND1	33.4	(0.8, 1382.2)	0.065 ⁺	0.5
EEFSEC RUVBL1	1.4	(0.6, 3.4)	0.505	1
AXIN1	3.1	(0.1, 85.4)	0.506	1
C4orf42 CTBP1	1.4	(0.6, 3.4)	0.505	1
WNT4	5.9	(0.1, 335.6)	0.388	1
LRP5	4.6	(0.6, 33.8)	0.13	0.9
NFATC3	4.5	(0.6, 35.2)	0.148	0.9
SFRP5	1.1	(0, 110.1)	0.967	1

Table 4.9: Univariate overall survival analysis of selected promoter methylation of Wnt pathway in the screening set ($n = 111$). Abbreviations: HR, hazard ratio; CI, confidence interval; FDR, false discovery rate. ¹Hazard ratio per unit increase of DMH ratio (continuous variable) estimated from Cox proportional hazard regression model. ²Confidence interval of the estimated hazard ratio. ³p value of Score test. ⁴multiple test correction in the validation study based on 14 validated DNA methylation biomarkers. ^{\$} $FDR \leq 0.2$; [&] $FDR \leq 0.1$; ⁺ $p \leq 0.1$; ^{*} $p \leq 0.05$; ^{**} $p \leq 0.01$; ^{***} $p \leq 0.005$

4.3.3 Expression and progression free survival in TCGA dataset

Promoter hypermethylation is frequently associated with gene inactivation (Illingworth and Bird, 2009), therefore, we further examined the correlation between expression and PFS of the 6 linked candidate genes that were validated in the independent cohort and independent from clinical parameters. This was examined in a larger independent cohort of patients ($n = 321$) with serous ovarian cancer in stage III and IV generated by The Cancer Genome Atlas (TCGA) Pilot established by the NCI and NHGRI (unpublished). Among the 6 genes, *NKDI* is not presented on the microarray. For the remaining 5 genes, low expression of *FZD4* and *DVL1* indicated a higher risk of recurrent disease (*FZD4*: $HR = 0.8$; 95% CI (0.7, 1.0); adjusted $p = 0.009$; *DVL1*: $HR = 0.8$; 95% CI (0.6, 1.0); adjusted $p = 0.022$), independent from stage, grade, age and residual disease (Table 4.10). The reduction of expression of *FZD4* and *DVL1* correlating with poor PFS is consistent with promoter methylation leading to the down-regulation of the genes.

RNA levels of *FZD4* and *DVL1* were measured by qRT-PCR in a subset of tumours, with sufficient quality of RNA, from the SGCTG cohort. Low expression of *DVL1* was significantly associated with promoter hypermethylation ($n = 27$, $p = 0.013$) and we found weak dependence between expression and methylation in gene *FZD4* ($n = 25$, $p = 0.066$) (Table 4.11). However, no evidence of association between methylation and expression for *NKDI*, *ROCK1*, *NFATC3* was observed in this small cohort of tumours.

4.3.4 Association of methylation at *DVL1* with response

To examine the role of the 6 DNA methylation biomarkers described above in chemoresponsiveness of ovarian cancer, we assessed the relationship between methylation and patients' response to first line platinum-based chemotherapy and found an increased methylation of *DVL1* was weakly correlated with poor response (Table 4.12). The odd ratio of the

Gene	Univariate PFS analysis				Multivariate PFS analysis		
	HR ¹	95%CI ²	p value ³	FDR ⁴	HR ¹	95%CI ²	p value ⁵
FZD4	0.8	(0.7, 0.9)	0.004**	<0.1 ^{&}	0.8	(0.7, 1.0)	0.009**
NKD1	-	-	-	-	-	-	-
DVL1	0.8	(0.7, 1.0)	0.041*	0.1 ^{&}	0.8	(0.6, 1.0)	0.022*
ROCK1	0.8	(0.4, 1.5)	0.444	0.8	0.9	(0.4, 1.8)	0.709
FZD9	0.8	(0.4, 1.7)	0.560	0.8	0.7	(0.3, 1.6)	0.345
NFATC3	1.0	(0.6, 1.6)	0.978	1	1.2	(0.7, 2.1)	0.441

Table 4.10: Survival analysis of expression of 6 candidate genes in the Wnt pathway in TCGA dataset ($n = 321$). Abbreviations: PFS, progression free survival; HR, hazard ratio; CI, confidence interval; FDR, false discovery rate.¹Hazard ratio per unit increase in expression estimated from Cox proportional hazard regression model.²Confidence interval of the estimated hazard ratio.³p value of Score test.⁴Multiple test correction based on 6 loci.⁵p value adjusted by stage, grade, residual disease and age. Age was used as a continuous variable; stage, grade and residual disease were used as categorical variables. [&] $FDR \leq 0.1$; * $p \leq 0.05$; ** $p \leq 0.01$

patients with progressive or stable disease ($n = 29$) to the patients with partial or complete response ($n = 54$) in every unit increase of methylation Z score is 1.6 (95% CI: 1.0-2.6, $p = 0.043$, $FDR = 15\%$). Again consistent with the methylation data is that decreased expression of DVL1 was related to poor response to platinum-based chemotherapy in a large patient cohort collected by TCGA (PD+SD ($n = 36$) vs. PR+CR ($n = 217$), $OR = 0.5$, 95% CI 0.3-0.9, $p = 0.035$).

4.3.5 Wnt pathway methylation index in late-stage EOC

To identify the best methylation predictors of PFS in the ovarian tumours from both screening and validation sets, the average DMH ratio of each locus was transformed to a Z score.

Gene	Expression	Methylation		P value ¹
		Hypermethylation	Hypomethylation	
FZD4	High expression	0	7	0.066 ⁺
	Low expression	7	11	
DVL1	High expression	0	7	0.013*
	Low expression	10	10	
NFATC3	High expression	0	7	0.276
	Low expression	4	16	
ROCK1	High expression	0	7	0.5413
	Low expression	2	18	
NKD1	High expression	3	3	0.903
	Low expression	6	13	

Table 4.11: Correlation between expression and methylation. ¹P value: Fisher's exact test (one sided) in cases samples in the group are smaller than 5; ⁺ $P \leq 0.1$; * $p \leq 0.05$

We excluded 10 Stage I tumours from 161 tumours since these may be biologically distinct entities and patients with early-stage ovarian cancer have much better clinical outcome than patients with late-stage cancer (Trope and Kaern, 2007). From the Z scores of 151 patients with late-stage EOC (stage III and IV), we constructed a multivariate Cox model, of which the covariates were selected by likelihood ratio (LR) forward stepwise algorithm from clinical parameters (age, stage, grade and histological type) and promoter methylation at *FZD4*, *DVL1*, *NKD1*, *ROCK1*, *FZD9* and *NFATC3* (see Methods). The final model includes two best predictors at *DVL1* ($HR = 1.21$; 95% CI 1.03 to 1.43; $p = 0.02$) and *NKD1* ($HR = 1.28$; 95% CI 1.05 to 1.57; $p = 0.02$). The hazard ratio (HR) represents the relative risk per unit increase in Z score. The remaining variables including clinical parameters and 4 methylation biomarkers were not selected into the model due to not meeting the entry criteria in forward stepwise method (see Methods). A methylation index (MI) calculated from this model ($MI = 0.25Z_{NKD1} + 0.2Z_{DVL1}$, Z denotes Z score) can identify two distinct prognostic groups using the third quartile of the index as the cut-off ($HR = 2.72$;

genes	Logistic regression (Response ¹ : PR+CR n=54; PD+SD n=29)			
	OR ²	95% CI ³	p value ⁴	FDR ⁵
FZD4	1.3	(0.8, 1.9)	0.270	0.32
NKD1	1.4	(0.9, 2.1)	0.115	0.23
DVL1	1.6	(1.0, 2.6)	0.043*	0.15 ^{\$}
ROCK1	1.0	(0.6, 1.4)	0.875	0.88
FZD9	1.3	(0.8, 2.0)	0.239	0.34
NFATC3	1.5	(1.0, 2.0)	0.051 ⁺	0.15 ^{\$}

Table 4.12: Response and methylation of 6 candidate genes in the Wnt pathway in the combined analysis of screening and validation set. Abbreviations: OR, odd ratio; CI, confidence interval; FDR, false discovery rate. ¹Response to first line chemotherapy was measured by RESIST 1.0; the patients with complete/partial response are the responders to chemotherapy, while the patients with progressive or stable disease are those who did not response to chemotherapy. ²Odd ratio of the patients had poor response to chemotherapy (PD+SD, $n = 29$) to the patient with good response to chemotherapy (CR+PR, $n = 54$) per unit increase of Z scores. ³Confidence interval of the estimated odd ratio. ⁴p value of score test of logistic regression model. ⁵Multiple test correction. ^{\$} $FDR \leq 0.2$; ⁺ $p \leq 0.1$; * $p \leq 0.05$

95%CI (1.74, 4.25); log rank test $p = 4.1 \times 10^{-6}$; permutation test $p = 4 \times 10^{-3}$) (Figure 4.7). Leave-One-Out Cross Validation (LOOCV) was used to assess the predictive value of the multivariate Cox model in 151 late-stage EOCs. As shown in Figure 4.8, the patients with high/low predicted MI in LOOCV have distinct PFS ($p = 5.75 \times 10^{-4}$, log rank test; $p < 0.01$, permutation test).

4.3.6 Bisulfite pyrosequencing

Bisulfite pyrosequencing has been used as a 'gold standard' in the measurement of methylation, therefore, methylation at promoter region of DVL1 was examined by bisulfite pyrosequencing in the tumours used in the screening set. Using the average methylation at 7 CpG sites targeted by bisulfite pyrosequencing, we found the correlation between DMH ratios (chr1: 1275054-1275112) and bisulfite pyrosequencing (chr1: 1275096-1275129)

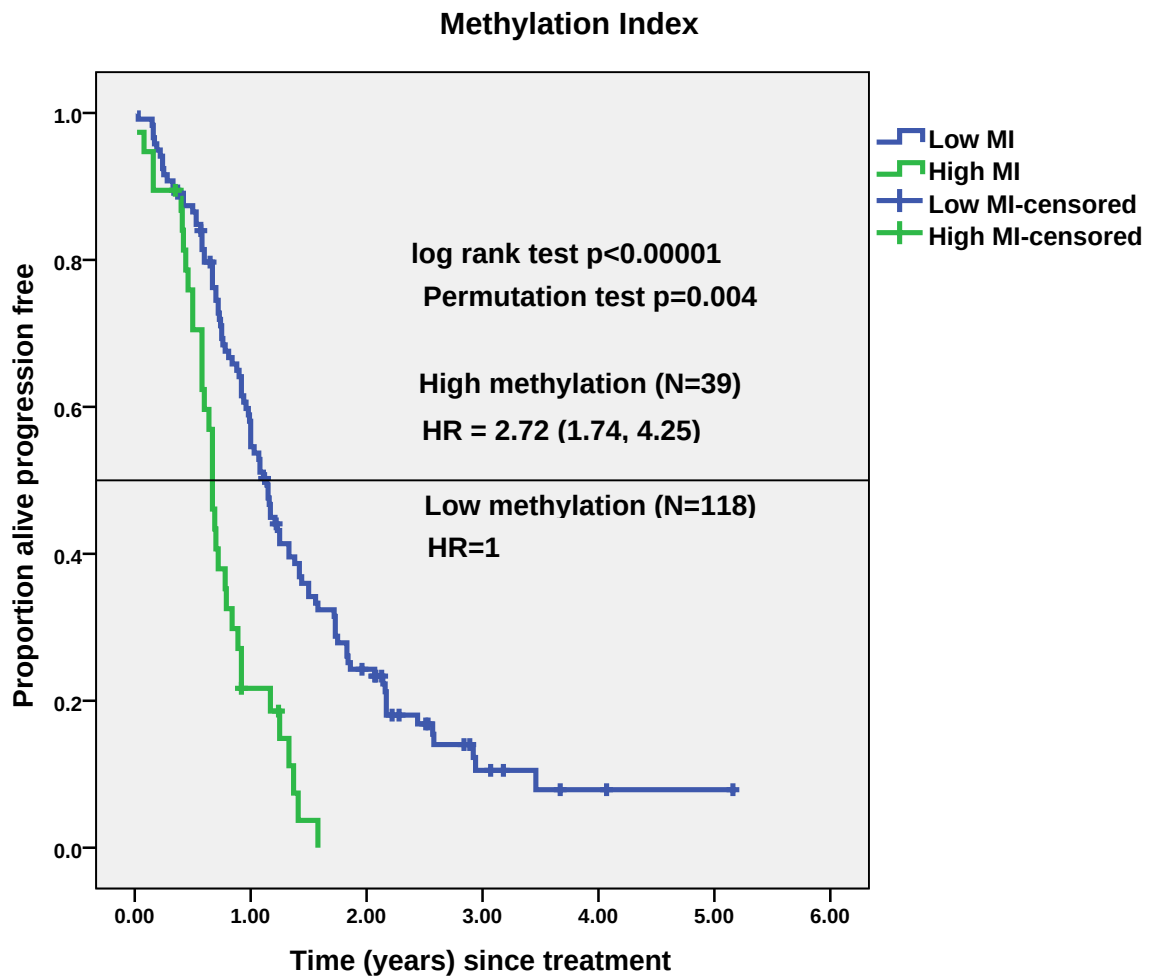


Figure 4.7: Kaplan-Meier plots of progression free survival in the patients with high/low methylation index (*MI*) estimated from the multivariate Cox model in late-stage ovarian cancer. The cut-off was determined by the third quartile in 148 patients ($n = 151$, 3 patients have PFS missing). Time is from the patients received the first line of chemotherapy.

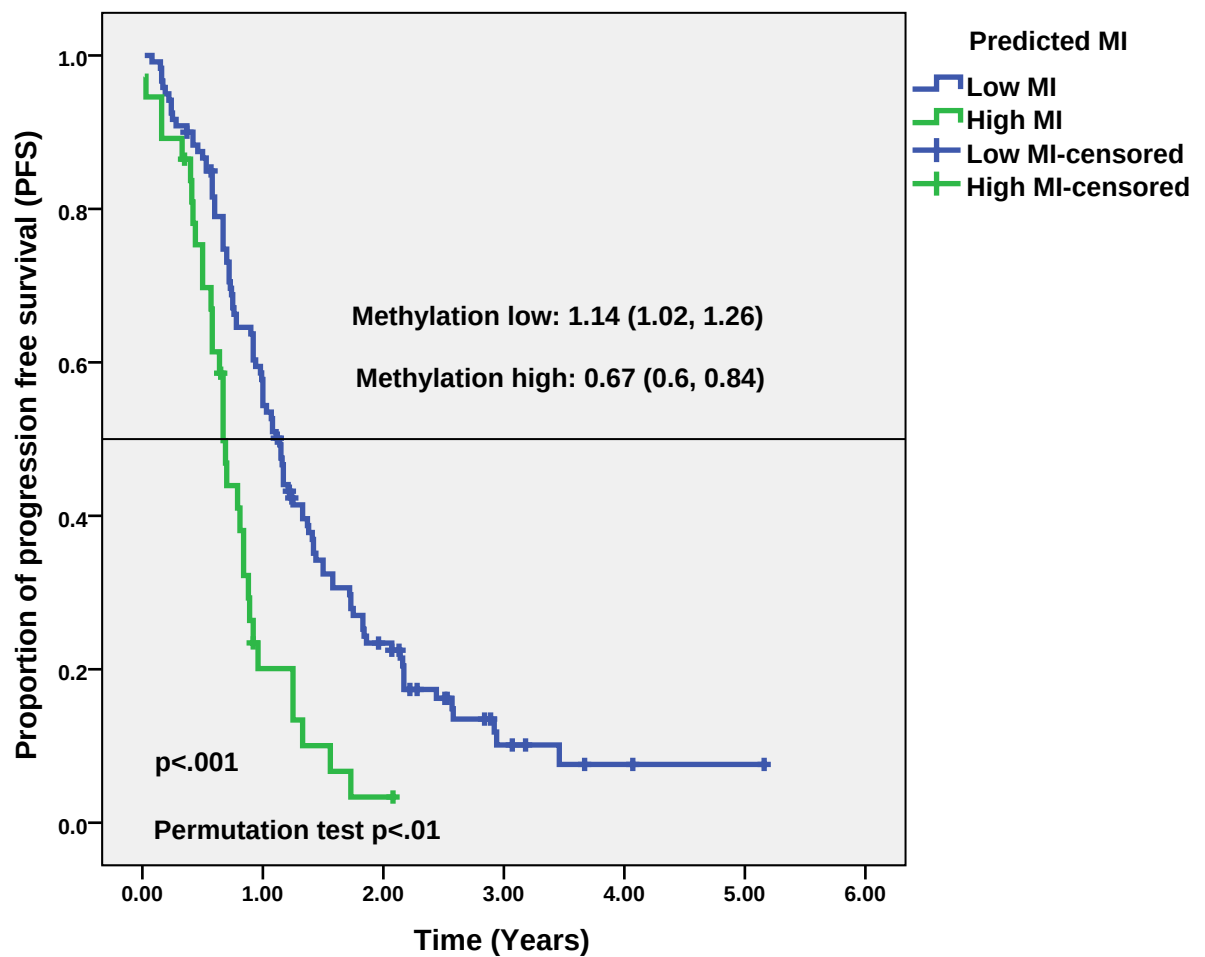


Figure 4.8: Kaplan-Meier plot of PFS in the risk groups predicted by leave-one-out cross validation. *MI*: predicted methylation index. Blue: patients with low predicted *MI*. Green: patients with high predicted *MI*. The third quartile of *MI* of the training set was used to determine the high/low *MI* of the test observation.

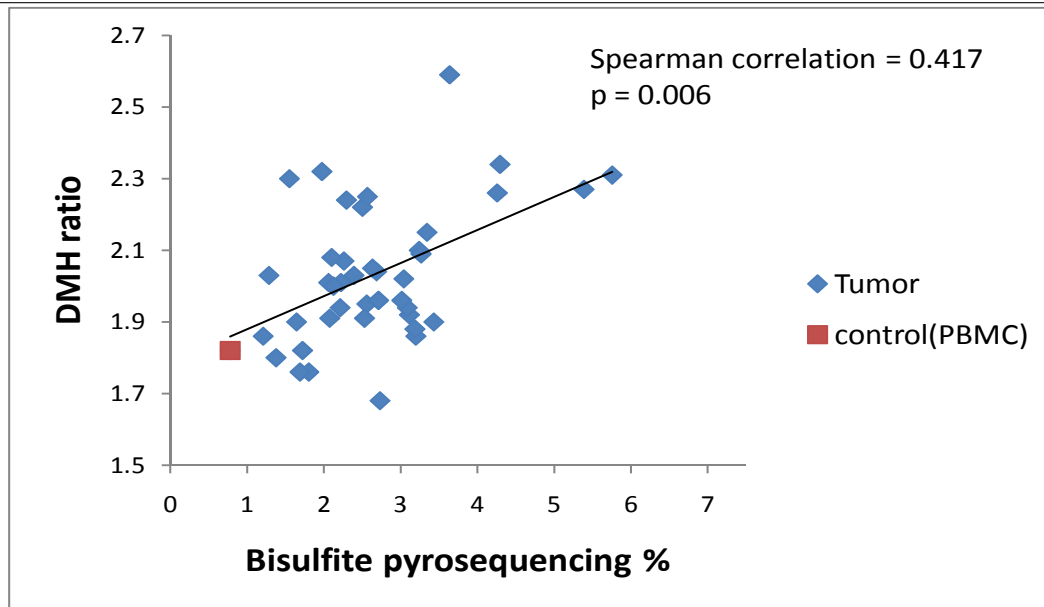


Figure 4.9: Correlation between DMH ratios at Chr1: 1275054-1275113 and bisulfite pyrosequencing at the locus Chr1:1275996 at promoter region of *DVLI*.

($n = 41$, spearman correlation $\rho = 0.417$, $p = 0.006$), though methylation level at this region is generally low ($<10\%$)(Figure 4.9).

4.4 Discussion

We examined DNA methylation at CGIs at the promoter regions of Wnt pathway genes, as defined by KEGG (entry ID: hsa04310)(Kanehisa et al., 2006, 2010), in 161 EOC. Using average DMH ratio as a continuous variable we identified 6 loci (*FZD4*, *DVLI*, *ROCK1*, *NFATC3*, *FZD9* and *NKDI*) of which methylation was associated with PFS in screening and validation tumour cohorts and are independent from known clinical prognostic features.

4.4.1 Systematic bias in DMH assay

As with any microarray experiment, systematic bias can be introduced from any of the multiple steps of DMH: such as, DNA preparation, hybridisation and image scan. In particular, DMH based on McrBC restriction enzyme, involves extra steps of digestion, ligation and PCR amplification. Therefore, the reproducibility of technical duplicates was evaluated using R^2 (coefficient of determination, ranging from 0 to 1) to estimate the variations introduced in DMH assays. After background correction, with-array normalisation and feature selection, the average R^2 of duplicate arrays in the screening set ($n = 111$) and in the following validation set ($n = 50$) was high, 0.92 ($n = 111$; $mean \pm SD$: 0.92 ± 0.04) and 0.96 ($n = 50$; $mean \pm SD$: 0.96 ± 0.02), respectively. This indicates the bias in our DMH assays was small, especially after appropriate data pre-processing (see Chapter 3). In addition, prior to conducting the survival analysis of the Wnt pathway, we observed a good correlation between existing bisulphite pyrosequencing methylation data at locus (chr14: 60174197-60174329) performed by Dr. Janet Graham from Imperial College London and average DMH ratio observed at this locus (Spearman correlation $\rho = 0.87$, $p < 0.01$) (Figure 4.10), supporting the use of average DMH ratio as a continuous variable in the methylation analysis.

4.4.2 Performance of methylation index in the Wnt pathway

We also evaluated methylation indices in the Wnt pathway based on these loci for sensitivity and specificity to identify patients who progressed within one year after receiving first line chemotherapy. This group of patients benefit less from platinum-based chemotherapy at second-line treatment than patients progressing after 12 months (Agarwal and Kaye, 2003) and may be an important patient population to treat with novel therapeutic approaches at presentation or with maintenance therapies during remission. As shown in

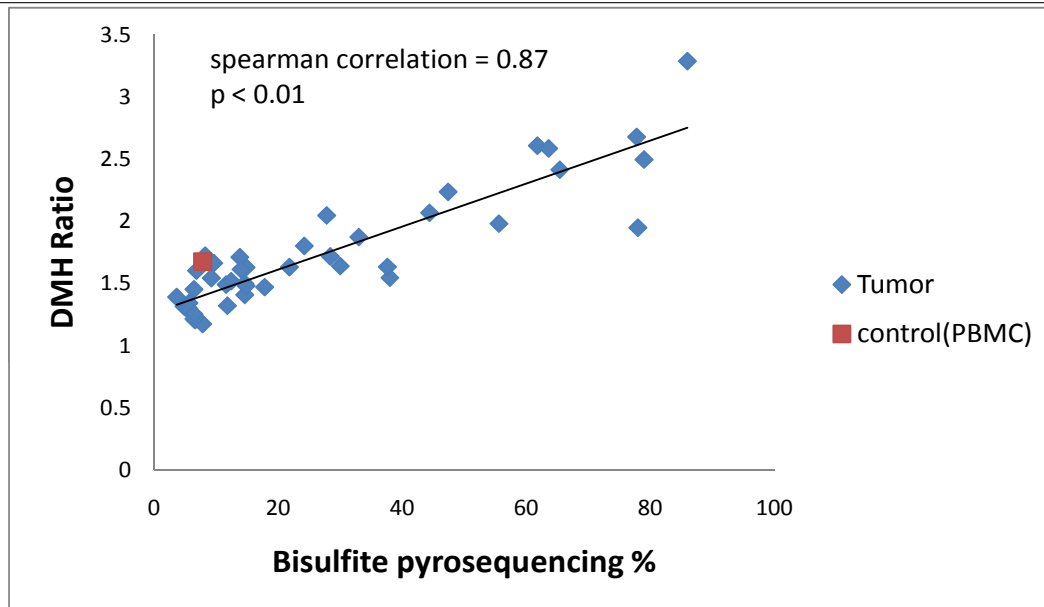


Figure 4.10: Correlation between DMH and bisulfite pyrosequencing at the locus (chr14: 60174197-60174329). PBMC: Peripheral Blood Mononuclear Cell

Figure 4.11, this methylation index appears slightly better than a prognostic index estimated from clinical parameters (age, stage, grade and histology type) to identify this group of patients, though neither of them reach the satisfying specificity and sensitivity to be of use in clinics.

Meanwhile, the regression coefficients ($B = 0.68$) of the predicted methylation indices by LOOCV can be used as the shrinkage factor (Verweij and Van Houwelingen, 1993) to estimate required sample size for validating this multivariate Cox model in the future study.

4.4.3 Expression and methylation

Two out of 5 genes examined by qRT-PCR show a correlation between CGI methylation and expression. Further investigation is required for the genes that did not show the correlation between expression and methylation because (1) tumour samples were not microdis-

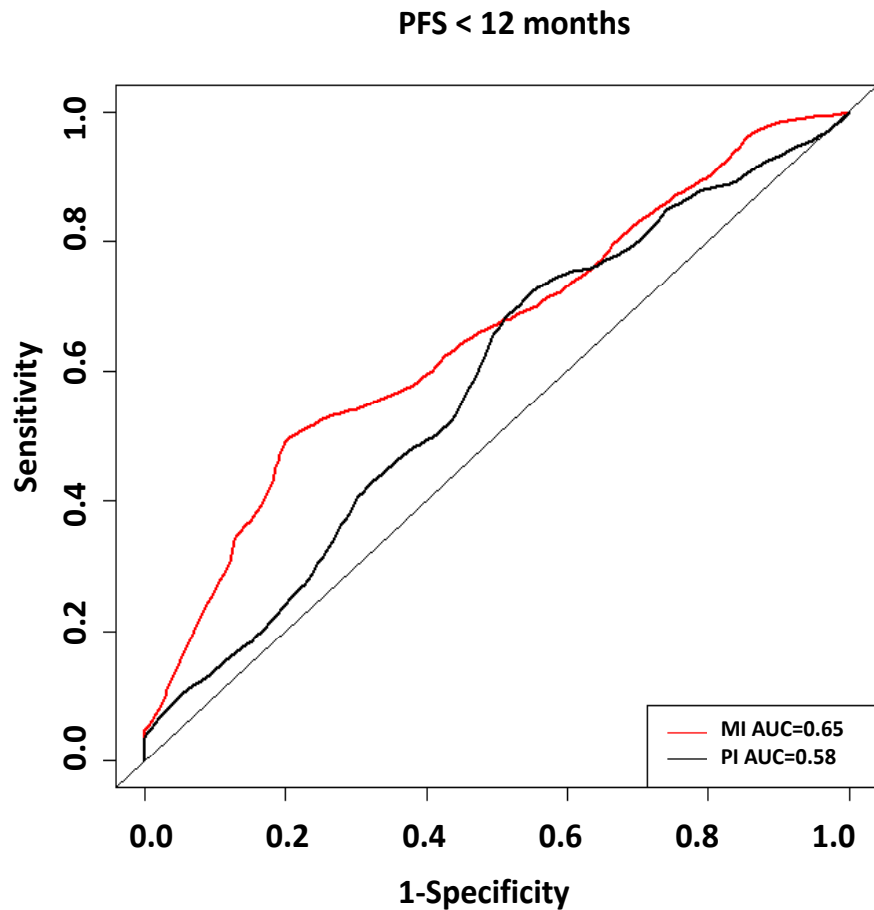


Figure 4.11: Receiver Operator Characteristic (ROC) curve. Red line: methylation index estimated from the multivariate Cox model including two independent biomarkers *NKDI* and *DVLI*; Black line: prognostic index estimated from the multivariate Cox model including Grade, Stage, age and Histology type of tumours. AUC: area under curve.

sected, which could confound the RT-PCR data or TCGA expression data due to expression of the genes in normal infiltrating tissues and (2) it is also possible that the biomarkers identified are heterogeneously methylated in the tumour tissue, i.e. differential methylation of those biomarkers only exist in a subset of tumour (or normal infiltrating) cells, therefore, the changes of expression of those genes in a subpopulation are masked by the general expression levels in the entire tumour.

4.4.4 Methylation of Wnt extracellular inhibitors in EOC

Though Wnt extracellular inhibitors are frequently methylated in EOC (Table 4.13), methylation of those inhibitors have no correlation with PFS in this study except *SFRP5*. *SFRP5* has been previously reported to correlate with poor response to platinum-based chemotherapy (Su et al., 2010). We found methylation at *SFRP5* was weakly associated with poor PFS, but was not independent from clinical parameters (adjusted $p = 0.094$, $n = 161$), and had no correlation with response to chemotherapy ($OR = 1.1$, 95% CI (0.7, 1.6), $p = 0.791$). Since it has been shown that silence of *SFRP5* is associated with promoter methylation (Su et al., 2010), we further examined if expression of *SFRP5* was correlated with response to first line platinum-chemotherapy in TCGA dataset, and no association was observed. However, we noted response was measured by both the reduction of tumour size and CA125 concentration in the previous study, while in our study response was measured by RECIST 1.0 criteria based on the shrinkage of tumour size. Furthermore, different composition of tumour stage and histology type in the cohort studies might also cause the differences.

genes	methylation frequency	methylation status in PBMC
DKK1	13.04%	U
DKK2	3.11%	U
DKK4	0%	U
SFRP1	9.32%	U
SFRP2*	50.31%	U
SFRP2*	27.33%	U
SFRP4	42.86%	U
SFRP5	67.70%	U
WIF1	78.88%	M

Table 4.13: Methylation frequency of extracellular inhibitors of the Wnt pathway. *Different promoter CGIs of *SFRP2*. U: unmethylated loci; M: methylated loci

4.4.5 Methylation alterations in all Wnt pathways determine the patients with early disease progression

The methylation frequencies of 6 independent biomarkers identified in this study are varied, ranging from 16% to 93% (Figure 4.12A). Given at least 7 probes are required to determine the methylation status (methylated/unmethylated) of each locus (see more details in section 3.3.7 of Chapter 3), methylation frequency of *DVL1* could not be estimated since there is only 6 probes targeting the promoter region we identified. Three major pathways are implicated in Wnt signalling transduction, Wnt/ β -catenin (canonical), PCP and Wnt/Ca²⁺ pathway (Boyer et al., 2009). Using NCI protein and interaction database (<http://pid.nci.nih.gov/>), we identified the genes in the canonical, PCP and Wnt/Ca²⁺ pathways. To identify methylation alterations of which Wnt pathway have major impact on disease progression, we categorised the patients into 5 groups based on the methylation alterations in different Wnt pathways, as shown in Figure 4.12A. Ten patients have no methylation at all 5 loci and 39 patients ("mild Wnt/ β -catenin methylation" in Fig-

ure 4.12A) have methylation only at *FZD9*. This gene has been shown to be involved in Wnt signaling transduction through Wnt/ β -catenin pathway in NCI curated protein/interaction pathway. Eighty-five patients ("Wnt/ β -catenin or PCP methylation" in Figure 4.12A) have methylation of at least two out of four loci (*FZD4*, *NKD1*, *ROCK1* and *FZD9*) that involves Wnt/ β -catenin pathway and/or PCP pathway. Sixteen patients ("Wnt/ Ca^{2+} methylation" in Figure 4.12A) have methylated promoters of *NFATC3* from Wnt/ Ca^{2+} pathway, but not all other 4 loci are methylated in those patients. Only 6 patients have severe methylation at all 5 loci indicating methylation alterations in all three Wnt signalling pathways. Methylation of *NFATC3* from Wnt/ Ca^{2+} pathway is always accompanied by methylation either in canonical and/or PCP pathways. "Wnt/ Ca^{2+} methylation" group have slightly shorter PFS compared to the patients with methylation in canonical or PCP pathways ("mild Wnt/ catenin methylation" and "Wnt/ β -catenin or PCP methylation") and without promoter methylation of *NFATC3* (Figure 4.12B and D), indicating methylation alterations in Wnt/ Ca^{2+} pathway probably have more impact to disease progression than those in the other two pathways in ovarian cancer. The patients with severe methylation in three Wnt signalling pathways have much shorter progression free survival compared to the rest four groups, especially to the patients without any methylation (HR=10.68, 95% CI 2.63 to 43.44, p=0.001) (Figure 4.12B and D). Since all the patients in this group have disease progression within about 12 months, most of them are potentially resistant to the second line chemotherapy as discussed in previous section, while all the patients without methylation at 5 loci in the Wnt pathway have no disease relapse in the first 6 months of treatment, so they are unlikely to be platinum-refractory (intrinsic resistance) (Agarwal and Kaye, 2003). This shows the importance of methylation at three major pathways for early disease relapse. Four of those loci are excluded from the multivariate methylation model that failed to stratify the patients with disease progression within 12 months, suggesting methylation pattern of multiple loci

is probably more useful than the mathematical model incorporating multiple loci to identify the patients with early disease progression. However, only the patients with no methylation changes and severe methylation in Wnt pathways have significant difference in overall survival, and there is no significant difference in overall survival of the remaining groups (Figure 4.12C).

4.4.6 The biological role of *FZD4* and *DVL1* in the Wnt pathway

We observe that the prognostic value of *FZD4* and *DVL1* at two molecular levels, i.e. both increased methylation and decreased expression of *FZD4* and *DVL1* are correlated with poor PFS in EOC. This indicates expression of these two genes might be tightly regulated by promoter methylation. Both Fzd4 and Dvl1 function at the upstream of the Wnt signalling and are of importance in the Wnt signalling transduction.

Frizzled homology 4 (*FZD4*)

FZD4 is frizzled homology 4, one member of frizzled receptor family. Recently, *FZD4* has been shown to express in granulosa cells in rat ovary (Ricken et al., 2002) and play a crucial role in corpus luteum formation and function in mouse ovary (Boyer et al., 2009), indicating the importance of *FZD4* in the normal function of ovary.

In late-stage serous EOC (n=321, TCGA dataset), among 9 frizzled receptors (*FZD1*, 3 to 10) presented on the microarray, only *FZD4* is co-expressed with *LRP5* ($\rho=0.27$, $p=1.23 \times 10^{-6}$, Pearson correlation). Methylation of *LRP5* is one of 14 validated biomarkers associated with PFS (hypermethylation is correlated with high risk, Table 4.6 and 4.7) in this study, and similar to *FZD4*, expression of *LRP5* is correlated with PFS in this patient cohort as well (n=321, HR=0.8, 95% CI (0.6,0.9), $p=0.008$, TCGA dataset) (low expres-

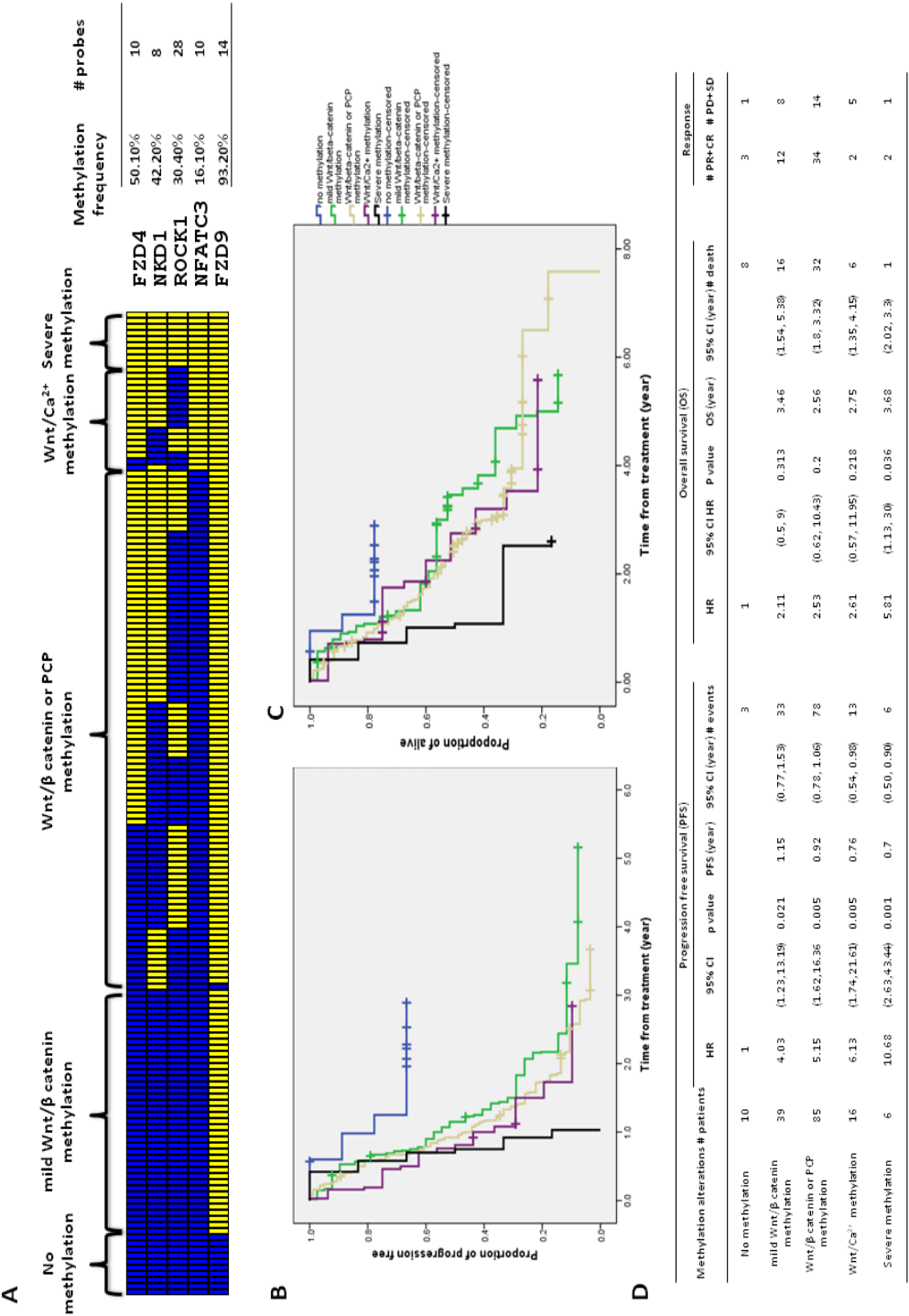


Figure 4.12: Methylation at Wnt/ β -catenin, PCP and/or Wnt/ Ca^{2+} pathways. A: Methylation of 5 independent biomarkers in Wnt/ β catenin, PCP and Wnt/ Ca^{2+} pathways. Yellow: patients with methylated locus; Blue: patients with unmethylated locus; B: Kaplan Meier curves of progression free survival in 5 groups of patients interrupt by methylation; C: Kaplan Meier curves of overall survival in 5 groups of patients with different Wnt pathways interrupt by methylation; D: hazard ratios of disease progression and death in 5 groups of patients.

sion is correlated with high risk). Therefore, *FZD4* and *LRP5* may function together in the signalling transduction of the Wnt pathway in ovarian cancer.

FZD4 receptor does not mediate Wnt/ β -catenin signalling, but is involved in Wnt/ Ca^{2+} pathway, which is stimulated by Wnt5a protein (McDonald and Silver, 2009). It has been reported that Wnt5a protein activates downstream calcium effector molecules and may inhibit uncontrolled the Wnt/ β -catenin signalling through the activation of receptor Ror2, processing an extracellular cysteine-rich Wnt binding domain. However, with the absence of Ror2, Wnt5a has been shown to activate TCF/LEF transcription in the presence of Lrp5 and Fzd4, thus promoting canonical Wnt signalling (Mikels and Nusse, 2006). Therefore, the expression of *ROR2* in EOC using GeneCards (<http://www.genecards.org/>) was examined and observed that this gene is under-expressed in normal ovary and ovarian tumours. Interestingly, expression of β -catenin alone is not correlated with progression free survival in TCGA dataset (n=321, HR=0.76, p=0.11), but the prognostic index from the multivariate Cox model including *ROR2*, *FZD4*, *LRP5* and β -catenin is a significant predictor of disease relapse and each gene has independent contribution to this association (m=321; *ROR2*, HR=0.62, p=0.025; *LRP5*, HR=0.77, p=0.009; *FZD4*, HR=0.80, p=0.006; β -catenin, HR=0.68, p=0.033; multivariate Cox model, p= 3×10^{-4} , score test), indicating their co-function with β -catenin in Wnt/ β -catenin. Further investigation of this pathway in EOC will be necessary to further understand how the aberrant promoter methylation affects the Wnt signalling pathway and contributes to ovarian cancer progression.

In addition, ovarian cancer is a highly heterogeneous disease. It has been found that a small subset of tumour cells, known as side population, enriched for tumour initiating capacity, expressing stem-like genes and resistant to chemotherapeutic drugs are present in

ascites of patients with ovarian tumours. Patients with higher proportion of such population of cells are subject to have unfavourable survival (Hu et al., 2010; Rizzo et al., 2011). Given the variation of methylation observed in this study is generally small, especially for the loci that increased methylation relevant to short PFS, methylation possibly occurs in such a small population of tumour cells. Consistent with this hypothesis, *FZD4*, also known as cell surface marker CD344, has been found to be significantly down-regulated in side population compared to non-side population in IGROV1 cell line in an independent analysis in our previous study ($FDR = 0.026$, rank products; fold change=0.34) (Figure 4.13). Side population cells in IGROV1 cell line are a highly tumourigenic subpopulation and have been characterised with an embryonic stem-cell-associated "NOS" signature (Notch/Oct4/Sox2 regulated genes) ($p < 10^{-6}$ and $FDR < 10^{-4}$, Gene Set Analysis). This suggests the dysregulation of the Wnt pathway by promoter methylation could possibly only occur in a subset of tumour cells which have more growth advantages in tumour progression and are subject to be resistant to the chemotherapy (Hu et al., 2010).

Dishevelled, dsh homology 1 (*DVL1*)

Dishevelled protein is a key component of Wnt signalling and plays a central role in propagating Wnt signalling to canonical and non-canonical pathways. Three dishevelled genes, *DVL1*, *DVL2* and *DVL3* have been identified in human. Wnt3a-sensitive canonical pathway was suppressed by knock-down of *Dvl1*, *Dvl2* and *Dvl3*, but most sensitive to the changes of either *Dvl1* or *Dvl3* (Lee et al., 2008). Down-regulation of *DVL1* by promoter CGI methylation can possibly interrupt the balance of Wnt signalling transduction between canonical and non-canonical pathways, leading to the distortion of Wnt signalling transduction.

In addition, *Dvl* is an indirect negative regulator of Notch pathway (Lu et al., 2007),

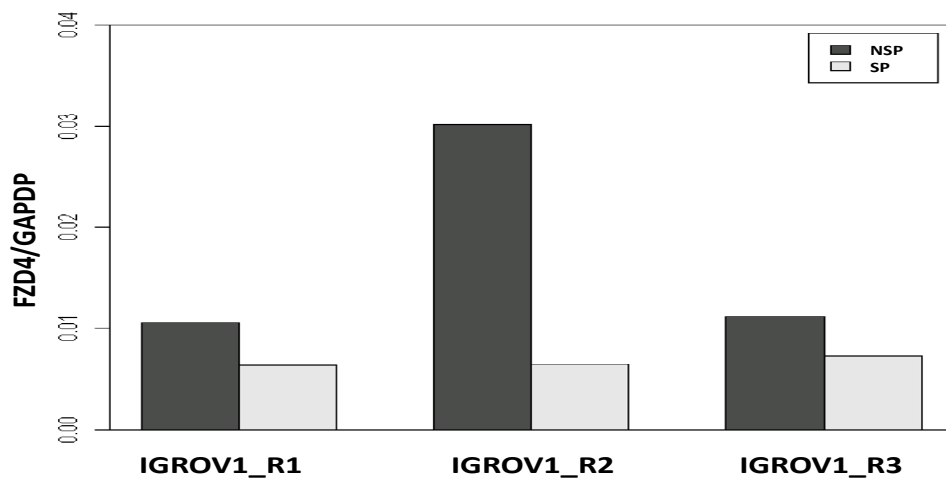


Figure 4.13: Expression of *FZD4* in IGROV1 SP and NSP cells. The average expression of *FZD4* is calculated by two probes (218665_at and 224337_s_at) and relative to the expression of *GAPDH* on Affymetrix HGU133plus2 microarray.

which has been shown activated in ovarian cancer (Hopfer et al., 2005; Rose, 2009). Further investigation about the effect of down-regulation of *DVL1* by promoter methylation on Notch pathway is also necessary to understand the cross talk between the Wnt and Notch pathways potentially mediated by Dvl in ovarian cancer progression.

4.4.7 Sample size required for further validation study

A limitation of the study is the low average power of the analysis in the validation set (see Materials and Methods). The current validation study while achieving 88% power for the loci with the greatest effect had only 30% power for the median observed effect in the initial evaluation. However, 14 out of 20 independent biomarkers identified in the first evaluation set have been validated in the second patient cohort. Nevertheless, those loci that have not been validated need further investigation in a larger sample set. A further 71 tumours are

required to reach the power of 90%, which is the power generally used in the validations study.

4.5 Conclusions

We have shown that methylation of multiple CGIs in the Wnt pathway is frequently observed in epithelial ovarian cancer and identified methylation of key loci as significantly associated with PFS (CGIs at *FZD4*, *DVL1*, *NKD1*, *ROCK1*, *FZD9* and *NFATC3*) that are independent from clinical parameters and constructed a multivariate Cox model that incorporates two independent CGIs at *NKD1* and *DVL1*, which can identify two groups of patients with distinct PFS. These data support the importance of epigenetic regulation of the Wnt pathway in ovarian cancer progression and provide novel prognostic biomarkers that have advantages over the use of expression in ovarian tumours that are highly heterogeneous as a biomarker. This will be relevant for patient stratification in future clinical trials of ovarian cancer, particularly for novel drugs targeting the Wnt pathway. Furthermore, it should also be pointed out the samples used in this analysis was obtained before treatment, and methylation status of these genes may changes in some of patients after treatment. Therefore, repeat measurement of methylation at relapse or disease progression would be necessary to further understand the association between dynamic methylation changes in the Wnt pathway, response and clinical outcome.

Chapter 5

CpG Islands Methylation Profiling in the Other Key Pathways and Their Prognostic Significance in Epithelial Ovarian Cancer

5.1 Introduction

Cancer is a disease of gene expression in which the multiple complex networks governing homeostasis become deranged. Key alterations in a variety of processes have been implicated in tumour development, including self-sufficiency in growth signals, insensitivity to growth-inhibitory signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis (Hanahan and Weinberg, 2000). A limited number of cellular pathways are altered and heritably disabled in almost all cancers (Jones, 2005; Baylin and Jones, 2006). Molecular alterations in these pathways, such as mutation and copy number changes have been extensively studied in cancers. However, a growing number of data suggest that heritable changes, regulated by epigenetic alterations, may also be critical for all human cancer initiation and progression (Sharma et al., 2010). Among the three major types of epigenetic changes include DNA methylation, histone modification and microRNA, DNA methylation is perhaps the most extensively studied in mammals (Zeller and Brown, 2010). DNA methylation provides a stable gene silencing mechanism that is of importance in regulating gene expression in association with histone modifications and transcriptional repression.

Table 5.1 gives some examples in 10 key cellular pathways studied in this thesis, including AKT/mTOR, BRCA1/2, Redox, p53, Wnt, Fanconi Anemia (FA) family, igLON family, genes involved in homologous recombination (HR), non-homologous end joining (NHER) and mismatch repair (MMR), disrupted in human cancers by genetic and epigenetic mechanisms (Altomare and Testa, 2005; Majumder and Sellers, 2005; Deng et al., 2010; Senturk et al., 2010; Wang and Yi, 2008; Hayes and McMahon, 2009; Teodoridis et al., 2005; Levine et al., 2006; He et al., 2010; Vogelstein et al., 1988b; Sparks et al., 1998; Boyer et al., 2009; Ying and Tao, 2009; Lyakhovich and Surralles, 2006; Lim et al.,

2008; Sellar et al., 2003; Ntougkos et al., 2005; Pasic et al., 2010; Helleday, 2003; Evers et al., 2010; Meindl et al., 2010; van Gent and van der Burg, 2007; Curtin et al., 2009; Sellar et al., 2003; Ntougkos et al., 2005; Pasic et al., 2010). The p53 pathway is undoubtedly among the most extensively studied pathways, which promote cell cycle arrest, apoptosis and senescence. TP53 mutation are frequently observed in most of the common types of human tumours (Levine et al., 2006), including almost all high grade serous ovarian cancer (Bowtell, 2010; Ahmed et al., 2010). It has been reported that methylation at promoter region of p14^{ARF}, upstream of p53 pathway, were observed in gliomas, breast cancer and ovarian cancer (Teodoridis et al., 2005; Barekati et al., 2010; He et al., 2010), and deregulation of p14^{ARF} by promoter methylation is critical in the p53 pathway since it indirectly regulates the level of p53 protein. The other example is BRCA1/2 signalling pathway. This pathway participates in multiple biological functions including cell cycle checkpoint and double strand break repair. BRCA1 mutation has also been observed in high grade serous cancer (Kobel et al., 2008), while silencing of BRCA1 by promoter methylation have been reported in ovarian cancer and is correlated with poor survival in patients compared to those with BRCA1 mutation or wild-type BRCA1 (Senturk et al., 2010). Aberrant promoter methylation has been found in some of these pathways (Table 5.1), but have not been comprehensively studied yet. Systematic methylation profiling of the key pathways in ovarian cancer could provide an insight into the aberrant methylation changes of these pathways in primary tumours. In Chapter 4, the promoter methylation of the genes in the Wnt pathway has been systematically examined and their prognostic value evaluated in ovarian cancer. Subsequently, 9 key pathways/families (AKT/mTOR, BRCA1/2, Redox, p53, FA family, igLON family, HR, NHER and MMR) were further examined for prognostic value of methylation at promoter CGIs in ovarian cancer. Therefore, this chapter focuses on the investigation of DNA methylation of multiple promoter CGIs of the genes in these 9 key

signalling pathways for predicting clinical outcome in ovarian cancer.

5.2 Methods

The patients selected into this analysis are the same patients cohort used to identify the prognostic DNA methylation biomarkers in the Wnt pathway (see Chapter 4) and analysed using the methods described in Chapter 4, except that no multivariate Cox model including multiple DNA methylation biomarkers was constructed. In brief, fresh frozen tumours were taken before chemotherapy. Genomic DNA was extracted as previously described (Strathdee et al., 1999), then were assayed by DMH in duplicates on Agilent custom-designed 60-microarray targeting the promoter CGIs of 10 key signaling pathways/families (p53, Wnt, AKT/mTOR, MMR, HR, NHEJ, BRCA1/2, FA family, IgLON family and Redox). The genes were mainly collected from Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Table 5.2). Unsupervised hierarchical clustering was performed using methylation status (methylated/unmethylated) of MseI fragments defined by Mann-Whitney U test using Euclidean distance and average linkage method. The significance of the clusters was tested by bootstrap resampling method as previously described (Shimodaira, 2004). The association of DNA methylation of individual locus (MseI fragment) with progression free survival (PFS) was examined by univariate and multivariate Cox model adjusted by clinical parameters (grade, stage, histology and age) using DMH ratios as a continuous variable in each pathway/family, separately. The difference of PFS of the patients with high/low methylation was tested by log rank test. The significance level was set at two-sided $p \leq 0.05$. All the analysis was done in R (version 2.10.1) and SPSS (version 16.0).

Pathway/ family	Function	Example of genetic alteration	Example of epigenetic alteration	reference
p53	The p53 pathway respond to a wide variety of stress signals, leading to apoptosis, cellular senescence or cell cycle arrest	p53 mutation	<i>P14^{ARF}</i> , <i>APAF1</i> , <i>FAS</i> methylation	(Teodoridis, Hall et al. 2005; Levine and Oren 2009; He, Qiao et al. 2010)
BRCA1/2	The BRCA1/2 pathway participates multiple biological functions including cell cycle checkpoint and double strand break repair	Mutation of <i>BRCA1</i> and <i>BRCA2</i>	<i>BRCA1</i> methylation	(Deng 2006; Senturk, Cohen et al. 2010)
FA family	Fanconi anemia proteins are a family has cross-talk with BRCA1/2 pathway and involves in double strand break repair	<i>FANCA</i> , <i>FANCC</i> and <i>FANCG</i> mutation	<i>FANCF</i> methylation	(Lyakhovich and Surrallès 2006; Lim, Smith et al. 2008)
Wnt	The Wnt pathways control a variety of biological processes including cell proliferation, differentiation and migration, and have a crucial role in development and tissue homeostasis	Mutation of <i>APC</i> , <i>β-catenin</i>	<i>E-cadherin</i> , <i>SFRPs</i> , <i>APC</i> methylation	(Vogelstein, Fearon et al. 1988; Sparks, Morin et al. 1998; Boyer, Goff et al. 2009; Ying and Tao 2009)
Akt/mTOR	AKT/mTOR pathway regulates multiple cellular processes including cell proliferation and survival, cell size, tissue invasion and angiogenesis	<i>PI3K</i> , <i>PTEN</i> mutation, <i>PTEN</i> deletion	PTEN methylation	(Altomare and Testa 2005; Majumder and Sellers 2005; Carnero, Blanco-Aparicio et al. 2008)
Redox	Redox signaling pathway is the process where free radicals, reactive oxygen species (ROC) and other electronically-activated species act as messengers for various biological processes	<i>NRF2</i> and <i>KEAP1</i> mutation	<i>PRDX2</i> , <i>GSTP1</i> methylation	(Wang and Yi 2008; Hayes and McMahon 2009)
MMR	The mismatch repair (MMR) pathway is involved in the removal of DNA base mismatch arising either during DNA replication or DNA damage	Mutation of <i>MLH1</i> and <i>MSH2</i>	<i>MLH1</i> methylation	(Pedroni, Sala et al. 2001; Hitchins, Lin et al. 2007; Martin, Lord et al. 2010)
HR	Homologous recombination (HR) repairs DNA double strand breaks using a homologous DNA sequence as a template for error-free repair	<i>RAD51C</i> mutation	NA	(Helleday 2003; Evers, Helleday et al. 2010; Meindl, Hellebrand et al. 2010)
NHEJ	Non-homologous end joining repairs double-strand breaks without requirement of a homologous DNA template	<i>XRCC2</i> single nucleotide polymorphism	NA	(van Gent and van der Burg 2007; Curtin, Lin et al. 2009)
Ig/LON Family	The Ig/LONs are an immunoglobulin subfamily of glycosylphosphatidylinositol-anchored cell adhesion molecules	<i>LASMP</i> copy number alteration and loss of heterozygosity (LOH)	<i>OPCML</i> methylation	(Sellar, Watt et al. 2003; Ntougkos, Rush et al. 2005; Pasic, Shlien et al. 2010)

Table 5.1: Disruption of cellular signaling pathways/families in human cancer by genetic and epigenetic alteration. NA: no changes available.

5.3 Results

We obtained the methylation profile of 10 key signaling pathways in 161 epithelial ovarian cancers. Among 10 signaling pathways, 69% of loci in the Wnt pathway are methylated in EOC, significantly higher than 60% estimated from all the loci presented on the microarray ($p=0.0447$). The genes frequently methylated ($>5\%$) in ovarian tumours and unmethylated in peripheral blood mononuclear cell (PBMC) are listed in Appendix 5.1. Using unsupervised hierarchical clustering (Euclidean distance and average linkage method), only two groups of patients were identified consistently clustered together, and confirmed by 1000 fold bootstrap resamplings ($n_1=10$, $p=0.02$ and $n_2=26$, $p<0.001$, respectively). The groups identified have no difference in Stage, Grade, histology and age (Figure 5.1A). These two groups of patients have distinct progression free survival (cluster 1: median PFS 2.05, 95% CI (1.42, 2.67); cluster 2: median PFS 1.34, 95% CI (1.00, 1.67); $p=0.017$ log rank test) but no difference in overall survival (Figure 5.1B). 9 out of 10 patients in cluster 1 are consistent with one cluster of patients (see cluster 1 in Figure 4.5, $n=17$) identified using the frequently methylated loci in the Wnt pathway on the microarray ($p<0.0001$), but the patients in cluster 2 are not overlapped with any group of patients identified in the Wnt pathway.

5.3.1 Methylation of other key pathways and progression free survival (PFS)

Association between methylation of individual locus and PFS was tested by univariate and multivariate Cox model adjusted by age, stage, grade and histology type. Table 5.2 lists the number of DNA methylation biomarkers identified in the key pathways in the screening set ($n=111$), as well as the numbers of loci had been validated in the following

Pathway/ Family	Source	Total # Genes	Total # CGIs	# Msel fragments (G) (130-6000bp)	Screening (N=111)			Validation (N=50)	
					# Msel fragments (G1) P<0.05 (G1/G %)	# Msel fragments (G2) FDR<0.1 (G2/G %)	# Msel fragments (G3) P<0.05 ⁻³ (G3/G1 %)		
p53	hsa04115 ¹	68	87	139	13 (9.4)	2 (1.4)*	4 (31.8)		
BRCA1/2	Biocompare ²	64	72	99	6 (6.1)	0	2 (33.3)		
FA family	Biocompare ²	14	15	16	0	0	NE		
Wnt	hsa04310 ¹	148	189	302	20 (6.6)	5 (1.7)	14 (70.0)		
Akt/mTOR	hsa04150 ¹	51	51	78	6 (7.7)	2 (2.56%)	4 (66.7)		
Redox	Collaboration	48	44	63	3 (4.8)	0	2 (66.7)		
MMR	ko03430 ¹	26	29	39	2 (5.1)	0	2 (100.0)		
HR	ko03440 ¹	35	31	45	1 (2.2)	0	1 (100.0)		
NHEJ	ko03450 ¹	17	14	17	0	0	NE		
igLon Family	Collaboration	6	10	13	1 (7.7)	0	0		

Table 5.2: Summary of univariate progression-free survival analysis in 10 signaling pathways/families. * The two biomarkers at the promoter regions of two genes with function in the p53 pathway, and also downstream targets of the Wnt pathway. +Two biomarkers have trend to correlate with PFS ($p < 0.1$).⁻¹pathway database from Kyoto Encyclopedia of Genes and Genomes (KEGG).⁻²Biocompare: <http://gspdb.biocompare.com/Gene/Pathway.aspx?gp=55>.⁻³p value of score test (one sided). NE: not evaluable since there is no significant biomarker at $FDR \leq 0.1$ in the evaluation study

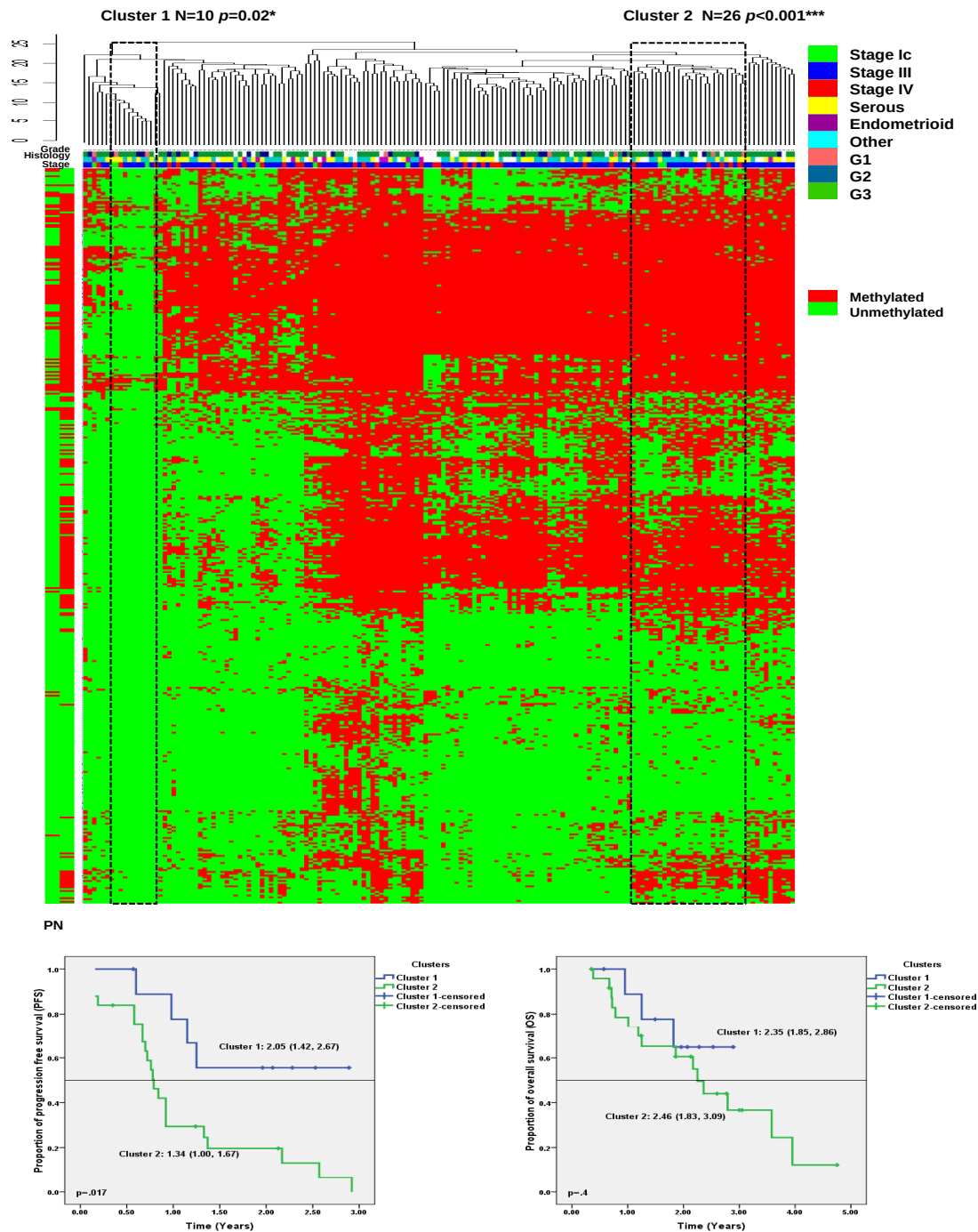


Figure 5.1: Unsupervised hierarchical clustering of binary DNA methylation data in 411 frequently methylated MseI fragments (row) on microarray in 161 epithelial ovarian tumours (column). Red: methylated fragment; Green: unmethylated fragment. P: Peripheral Blood Mononuclear Cell (PBMC); N: adjacent samples of benign tumours. Distance: Euclidean distance; Linkage method: average linkage. Methylation frequency of MseI fragments are from 5% to 95% and size are from 130bp to 2500bp.

patient cohort (n=50). The loci identified and validated in the Wnt pathway have been described in details in Chapter 4. In the other key pathways, methylation at 13/139 in the p53 pathway, 5/99 in the BRCA1/2 pathway, 6/78 in the AKT/mTOR pathway, 3/63 in the Redox pathway, 2/39 in MMR and 1/45 in HR were correlated with PFS ($p<0.05$) in the screening study (Table 5.3). In the subsequent validation study, the correlation between methylation and PFS remained in 15 loci ($p<0.05$ and $FDR<10\%$), of which 4 loci are from the p53 pathway (*BAI2*, *BAX*, *LRDD*, and *CCND1*), 2 loci from the BRCA1/2 pathway (*HDAC4* and *HDAC11*), 4 loci from the AKT/mTOR pathway (*VEGFA*, *PIK3R5*, *AKT1* and *VEGFB*), 2 loci from the Redox pathway (*PRDX2* and *TR2IT2*), 1 locus from HR (*RECQL4*) and 2 loci from MMR (*LIG1* and *MLH3*), respectively (Table 5.3). Multivariate survival analysis shows the association between methylation and PFS at *PIK3R5*, *AKT1* and *VEGFB* from AKT/mTOR pathway, *PRDX2* and *TR2IT2* from redox pathway, and *MLH3* from MMR are independent from grade, histology, stage and age (adjusted $p<0.05$) (Table 5.4).

5.3.2 Expression of other key pathway and progression free survival (PFS)

Further examination of the expression of the candidate genes associated with methylation biomarkers independent from clinical parameters (*PI3R5*, *AKT1*, *VEGFB*, *PRDX2*, *TR2IT2* and *MLH3*) was obtained from gene expression profiling study in ovarian cancer on Affymatrix HGU133A microarray (TCGA dataset, unpublished). This analysis found although none of those genes have significant correlation in expression and PFS, expression of *PRDX2* and *MLH3* show a trend of correlation with PFS in this patient cohort (low expression is correlated with high risk, $n=321$, $p<0.1$) (Table 5.5).

Pathway/ family	genes	Univariate PFS analysis (Screening N=111)				Univariate PFS analysis (Validation N=50)			
		HR ¹	95% CI ²	p value ³	FDR ⁴	HR ¹	95% CI ²	p value ⁵	FDR ⁶
AKT/mTOR	RPS6KA2	172.7	(6.2, 4803.5)	2.97E-03**	0.1 ^{&}	16.5	(0.5, 554.3)	5.80E-02 ⁺	<0.1 ^{&}
	PIK3CG	0.7	(0.5, 0.9)	7.50E-03**	0.2	1.2	(1, 1.3)	9.91E-01	1
	VEGFA	13.7	(1.1, 173.7)	4.29E-02*	0.6	3.3	(1.1, 10.2)	1.77E-02*	<0.1 ^{&}
	PIK3R5	22.6	(1.9, 266.9)	1.38E-02*	0.3	1.8	(0.9, 3.5)	4.09E-02*	<0.1 ^{&}
	AKT1	30.3	(3, 307.3)	3.80E-03**	0.1 ^{&}	4.3	(1, 17.9)	2.03E-02*	<0.1 ^{&}
p53	VEGFB DNAJC4	14.7	(1.6, 133.1)	1.64E-02*	0.3	22.7	(2.1, 24592.9)	1.08E-02*	<0.1 ^{&}
	BAI1	108.1	(4.6, 2515.4)	3.73E-03**	0.2	2.7	(1.1, 6.3)	1.06E-02*	<0.1 ^{&}
	CDK6	926	(1, 852982.5)	4.95E-02*	0.5	1.4	(0, 1067.2)	4.64E-01	0.5
	GTSE1	2.4	(1.2, 4.8)	1.49E-02*	0.3	1.2	(0.9, 1.5)	1.17E-01	0.3
	BAX	2.4	(1, 5.7)	4.47E-02*	0.5	2.5	(1.3, 5)	3.92E-03**	<0.1 ^{&}
BRCA1/2	SERPINB5	1.3	(1, 1.6)	4.29E-02*	0.5	1	(0.8, 1.3)	3.78E-01	0.1
	LRDD	17.6	(1.9, 165.3)	1.20E-02*	0.3	3.1	(1.1, 8.6)	1.46E-02*	<0.1 ^{&}
	CDKN2BAS CDKN2A	0.4	(0.2, 1)	3.84E-02*	0.5	1.2	(0.7, 2.2)	2.64E-01	0.4
	TAF8 CCND3	0.5	(0.3, 0.8)	1.62E-03**	0.1 ^{&}	1.3	(0.9, 1.8)	6.90E-02 ⁺	0.2
	CCND1	3.8	(1.8, 8.3)	5.20E-04***	0.1 ^{&}	1.4	(0.6, 3.4)	2.30E-01	0.4
Redox	RPRM	0.5	(0.3, 1)	3.88E-02*	0.5	1.5	(0.7, 3.2)	1.46E-01	0.3
	SERPINE1	6.4	(1.1, 38.3)	4.16E-02*	0.5	1.8	(0.1, 26.7)	3.28E-01	0.4
	CCNG2	44.4	(1.4, 1440.3)	3.28E-02*	0.5	0.6	(0.1, 6.7)	6.49E-01	0.2
	CCND1	48.6	(2.1, 1108.6)	1.45E-02*	0.3	2353.7	(12.7, 435569.4)	1.58E-03**	<0.1 ^{&}
	HLTF	0.8	(0.6, 1)	4.01E-02*	0.6	1.2	(0.9, 1.6)	9.22E-01	1
MMR	RBBP8 RBBP8	22.6	(1.9, 274.9)	1.55E-02*	0.4	0.6	(0.1, 5.8)	6.74E-01	0.4
	HDAC4	8.9	(1.5, 51)	1.44E-02*	0.4	2.3	(1, 5.4)	3.09E-02*	<0.1 ^{&}
	HDAC11 HDAC11	10.6	(1.8, 62)	9.03E-03**	0.4	2.3	(1.1, 4.7)	1.23E-02*	<0.1 ^{&}
	RBBP8 RBBP8	39.6	(4.2, 372.2)	1.71E-03**	0.2	0.8	(0.2, 4.6)	5.76E-01	0.4
	TXNRD1	0.4	(0.2, 0.9)	2.50E-02*	0.5	1.6	(0.6, 4.8)	8.16E-01	1
HR	PRDX2	2.5	(1.4, 4.6)	3.48E-03**	0.2	2.1	(1.1, 4.2)	1.39E-02*	<0.1 ^{&}
	TR2IT1	21.1	(1.8, 249.7)	1.62E-02*	0.5	5.6	(1.5, 21.1)	4.82E-03**	<0.1 ^{&}
	LIG1	2	(1.1, 3.8)	2.63E-02*	0.5	1.4	(1.0, 1.9)	2.95E-02*	<0.1 ^{&}
	MLH3	43.6	(1.7, 1102.7)	2.19E-02*	0.5	1.6	(0.9, 267.5)	2.90E-02*	<0.1 ^{&}
	LRRCL14 RECQL4	4.7	(2.0, 602.4)	2.09E-02*	0.8	3.5	(1, 11.8)	2.32E-02*	NE

Table 5.3: Survival analysis (PFS) of DNA methylation biomarkers in 9 key pathways. Abbreviations: HR, hazard ratio; CI, confidence interval; FDR, false discovery rate.¹Hazard ratio per unit increase in DMH ratio (continuous variable) estimated from Cox proportional hazard regression model.²Confidence interval of the estimated hazard ratio. ³p value of Score test (two-sided). ⁴multiple test correction based on the loci in each pathway/family.⁵ p value of Score test (one sided).⁶multiple test correction in the validation study based on the loci identified in the screening study.[&]FDR ≤ 0.1 ; ⁺p < 0.1; *p ≤ 0.05 ; **p ≤ 0.01 ; ***p ≤ 0.001

Pathway/ family	genes	Multivariate PFS analysis (N=111)		
		HR ¹	95% CI ²	adjusted p value ³
AKT/mTOR	VEGFA	4.1	(0.3, 57.3)	2.94E-01
	PIK3R5	15	(1.2, 184.6)	3.44E-02*
	AKT1	53.7	(4.3, 674.9)	2.04E-03**
	VEGFB DNAJC4	13.4	(1.3, 141.6)	3.07E-02*
p53	BAI1	20.4	(0.7, 592.5)	7.94E-02 ⁺
	BAX	1.6	(0.6, 4.1)	3.28E-01
	LRDD	8.2	(0.8, 86.6)	7.90E-02 ⁺
	CCND1	1.9	(0.8, 4.8)	1.69E-01
BRCA1/2	HDAC4	3	(0.4, 21.4)	2.64E-01
	HDAC11 HDAC11	4.6	(0.7, 29.7)	1.08E-01
Redox	PRDX2	3.3	(1.6, 6.6)	7.79E-04***
	TR2IT1	22.4	(1.7, 298.4)	1.85E-02*
MMR	LIG1	1.7	(0.8, 3.4)	1.63E-01
	MLH3	58	(1.9, 1782.3)	2.02E-02*

Table 5.4: Multivariate survival analysis (PFS) of validated DNA methylation biomarkers in 9 key pathways. Abbreviations: HR, hazard ratio; CI, confidence interval; FDR, false discovery rate.⁻¹Hazard ratio per unit increase in DMH ratio (continuous variable) estimated from Cox proportional hazard regression model.⁻²Confidence interval of the estimated hazard ratio. ⁻³p value adjusted by histology, grade, stage and age. Age was used as a continuous variable and histology, grade and stage were used as categorical variables. ⁺ $p < 0.1$; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$

Pathways/ Families	Gene	Univariate PFS analysis (TCGA dataset, N=321)		
		HR ¹	95%CI ²	p value ³
AKT/mTOR	VEGFB	1.15	(0.83, 1.60)	0.4
	PIK3R5	0.77	(0.35,1.67)	0.506
	AKT1	0.88	(0.72, 1.08)	0.231
Redox	PRDX2	0.69	(0.47, 1.02)	0.063 ⁺
	TR2IT1	-	-	-
MMR	MLH3	0.73	(0.53, 1.01)	0.06 ⁺

Table 5.5: Univariate survival analysis of expression of candidate genes in the other pathways/families in TCGA dataset. Abbreviations: B, coefficient of Cox proportional hazard regression model; HR, hazard ratio; CI, confidence interval; FDR, false discovery rate. -¹Hazard ratio estimated from Cox proportional hazard regression model.-²Confidence interval of the estimated hazard ratio. -³p value of Score test. +p<0.1; *p≤0.05; **p≤0.01;***p≤0.001

5.3.3 Methylation and overall survival/response

Among all 6 independent and validated biomarkers, only *PRDX2* shows an association of methylation to overall survival (screening set: HR=1.90, 95% CI (1.21, 2.99), p=0.006; validation set: HR=3.22, 95% CI (1.18,9.79), one sided p=0.01). But no correlation between methylation of *PRDX2* and response was observed (p=0.553). Out of 6 biomarkers, only increased methylation of a locus close to *VEGFB* or *DNAJC4* is significantly related to poor response to chemotherapy in the combined analysis of screening and validation set (Table 5.6).

5.4 Discussion

Multiple cellular signaling pathways are disordered in cancers. Though genetic alterations of these pathways have been extensively investigated, systematic study of promoter methylation of the genes in the pathways has not been done in cancers yet. In this study, methy-

Pathway/family	genes	Logistic regression (Response ¹ : PR+CR n=54; PD+SD n=29)			
		OR ²	95% CI ³	P value ⁴	FDR ⁵
AKT/mTOR	PIK3R5	0.78	(0.48, 1.26)	0.305	0.672
	AKT1	0.94	(0.57, 1.56)	0.812	0.812
	VEGFB DNAJC4	1.94	(1.11, 3.40)	0.021*	0.126 ^{\$}
Redox	PRDX2	0.87	(0.56, 1.36)	0.553	0.672
	TR2IT1	1.24	(0.77, 1.98)	0.372	0.672
MMR	MLH3	1.17	(0.69, 2.01)	0.56	0.672

Table 5.6: Response and methylation of 6 candidate genes in the key pathways in the combined analysis of screening and validation set. Abbreviations: OR, odd ratio; CI, confidence interval; FDR, false discovery rate. ¹Response to first line chemotherapy was measured by RESIST 1.0; the patients with complete/partial response are the responders to chemotherapy, while the patients with progressive or stable disease are those who did not response to chemotherapy. ²Odd ratio of the patients had poor response to chemotherapy (PD+SD, $n = 29$) to the patient with good response to chemotherapy (CR+PR, $n = 54$) per unit increase of Z scores. ³Confidence interval of the estimated odd ratio. ⁴p value of score test of logistic regression model. ⁵Multiple test correction. ^{\$} $FDR \leq 0.2$; ⁺ $p \leq 0.1$; ^{*} $p \leq 0.05$

lation profile of 9 signaling pathways/families was obtained in 161 ovarian tumours and association between methylation and clinical outcome has been evaluated.

In a previous study, methylation frequency of 24 CpG islands of genes associated with DNA damage responses have been examined by methylation specific PCR (MSP) in late-stage ovarian cancer (Teodoridis et al., 2005). Eleven out of 24 CGIs are presented on the custom-designed microarray in present study (Table 5.7). The estimated methylation frequency in this study is generally higher compared to previous study. For example, promoter CGI methylation of *FANCF* was found in 14.9% of patients in present study of which majority (96%) have late-stage tumours, while promoter methylation was not detected in late-stage ovarian cancer by Teodoridis et al., probably because the custom-designed microarray is targeting the whole CpG island by multiple probes, while MSP primers only

cover a few CpG sites within CGI and possibly the informative loci are missed due to the restriction of MSP primer design in Teodoridis et al. study. In addition, the tumour tissues that is a mixture of cells, with a number of patterns of methylation could influence the sensitivity of MSP (Smith et al., 2003). Our result is comparable to the study performed by Lim et al. (2008), in which promoter methylation of *FANCF* was detected in 13.2% of advanced ovarian cancer. Promoter region of *BRCA1* was methylated in only 1.2% of patients, which is much lower than the estimation from previous study. Although the region was targeted by 8 probes fulfilling the requirement that at least 7 probes should be used to determine the methylation status estimated in feasibility study (Chapter3), the promoter CGI of *BRCA1* is within a large MseI fragment (2274bps). The larger the MseI fragment is, the more likely to observe the larger variation depending on the distance between the probes. This can potentially influence the Mann-Whitney test, and partially explain the reason that much less methylation frequency was found at *BRCA1* in our study, indicating that fragment size bias exist in DMH assay using MseI restriction enzyme to fragmentize the genomic DNA.

Expression and methylation at two genes (*MLH3* and *PRDX2*) are weakly correlated with PFS in this study, indicating promoter methylation at these two genes probably regulates gene expression, though direct relation between methylation and expression of these two genes has not been validated yet. *MLH3* is a member of the MutL-homology (MLH) family of DNA mismatch repair (MMR) genes, interacting with MLH1. Cells in culture with dominant negative MLH3 protein have been shown exhibiting microsatellite instability (Lipkin et al., 2000). There is no report about methylation at *MLH3* so far, but it has been shown that *PRDX2* silencing caused by putative promoter region methylation can be restored by a demethylating agent in human malignant melanomas (Furuta et al., 2006).

Gene	Pathway/Family	Methylation Frequency	
		Teodoridis et al.	Present study
APAF1	p53	0%	11.2%
BRCA1	BRCA1/2	12.3%	1.2%
CASP8	p53	1.2%	1.9%
FANCF	FA family	0%	14.9%
FAS	p53	0%	1.9%
GSTP1	redox	1.2%	13.7%
MLH1	MMR	5.7%	3.7%
OPCML	IgLON family	33.3%	85.7% ¹ or 40.3% ²
CDKN2A	p53	0%	0%
CDKN1A	p53	0%	67.7%
TP73	p53	0%	83.9%
SFRP1	Wnt	5.3%	9.9%

Table 5.7: Methylation frequency of promoter CGIs of 12 genes in ovarian cancer estimated from Teodoridis et al study and present study.⁻¹Chr11: 132317872-132319158. ⁻²Chr11: 132319158-132319639

This proves the dependency between expression and promoter methylation at this gene. However, *PRDX2* is observed to be up-regulated after chemotherapy and increased expression of *PRDX2* is associated with cisplatin resistance in late-stage and high grade ovarian cancer (L'Esperance et al., 2006). In contrast, we observe no association between response to chemotherapy and methylation. Thus, methylation of *PRDX2* alone has no major impact on response to the first line platinum-based chemotherapy in ovarian cancer. Furthermore, it should also be pointed out the samples used in this analysis was obtained before treatment, and methylation status of *PRDX2* may changes in some of patients after exposure. Among two pairs of isogenic cell lines isolated from two patients before and after chemotherapy (PEA1 vs. PEA2 and PEO14 and PEO23), *PRDX2* is methylated in PEA1, PEA2, PEO14, but become unmethylated in PEO23 cell line (Figure 5.2). Therefore, repeat measurement of methylation at relapse or disease progression would be necessary to further understand the association between dynamic methylation changes, response and clinical outcome.

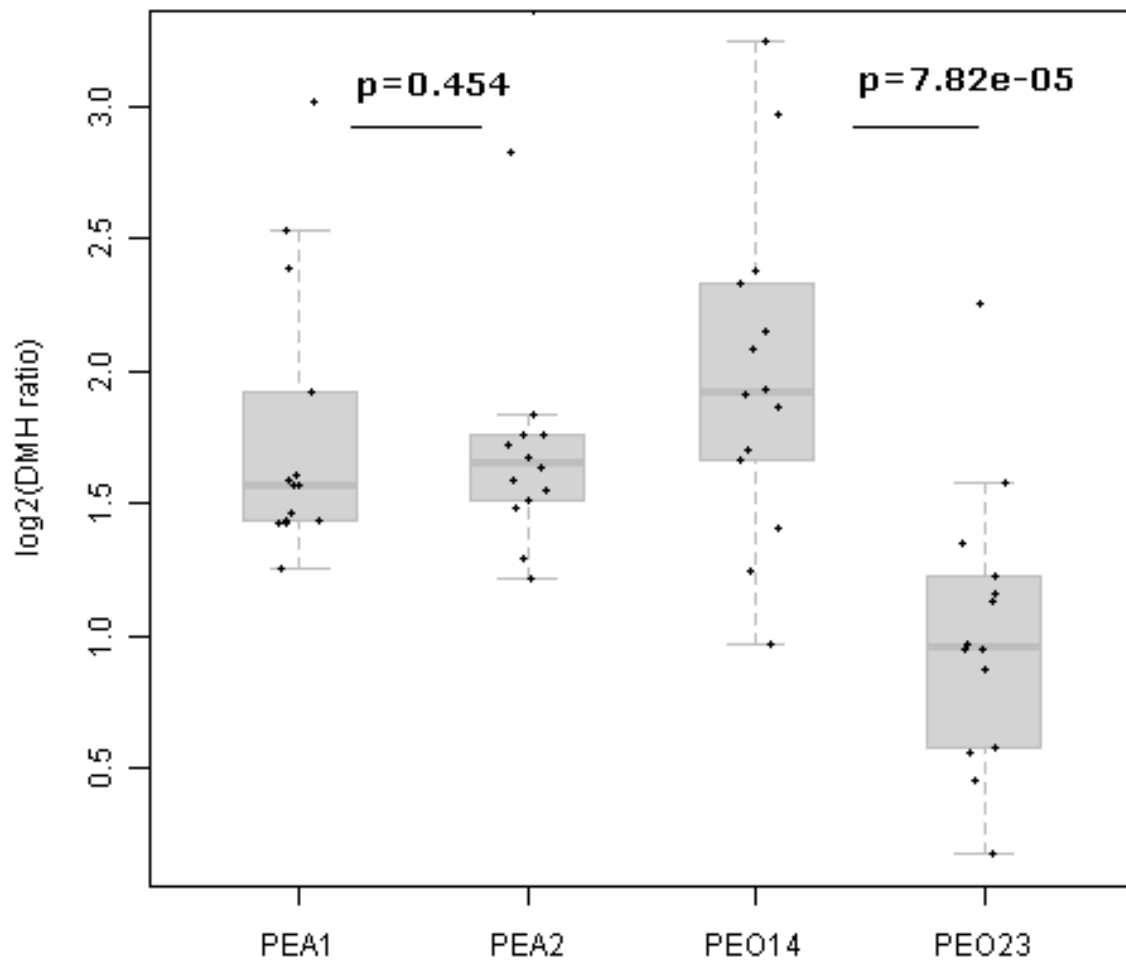
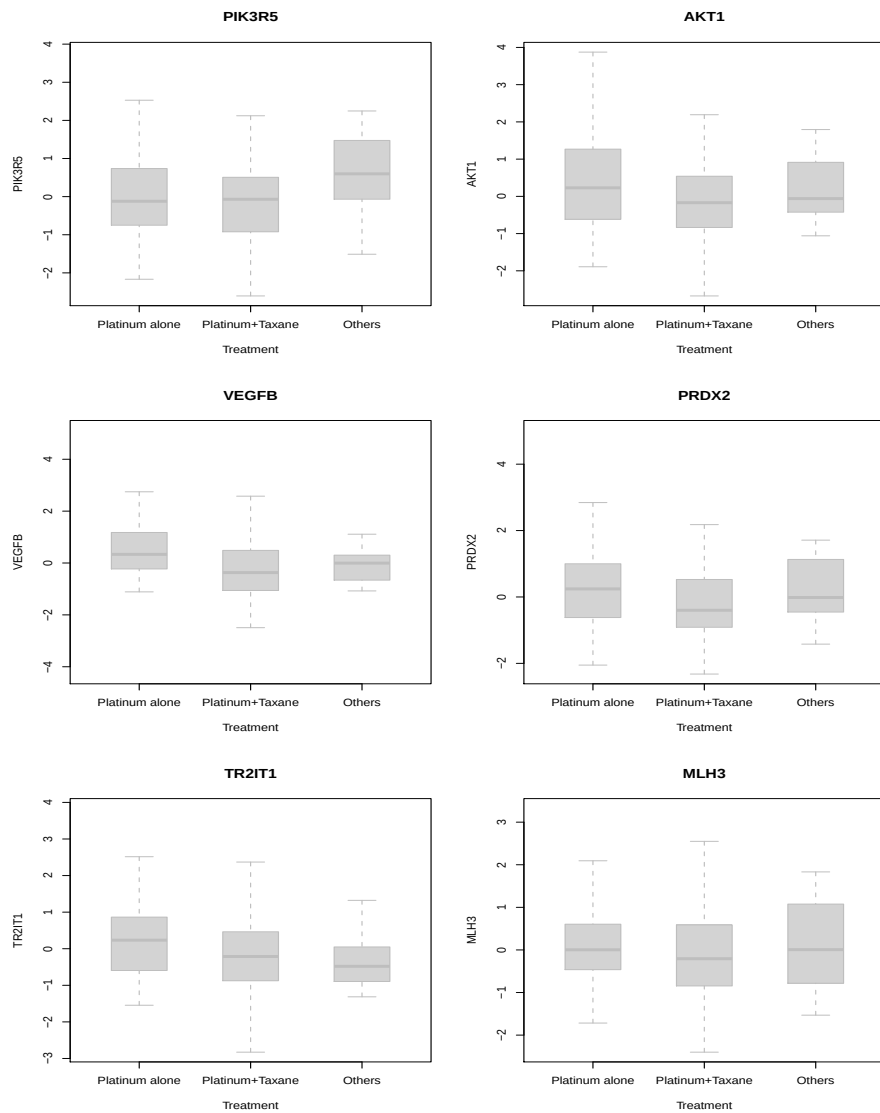


Figure 5.2: Methylation of *PRDX2* in two pairs of cell lines isolated from two patients before and after chemotherapy measured by DMH assay on Agilent custom-designed microarrays. *PRDX2* are methylated in PEA1, PEA2 and PEO14, and unmethylated PEO23 determined using 14 probes targeting the promoter CGI by the method described in section 3.2.4. Methylation was dramatically decreased in PEO23 compared to PEO14 ($p = 7.82 \times 10^{-5}$, Man-Whitney test).

In SGCTG patient cohort, the response rates of the patients received platinum alone and platinum combined with taxane treatment, platinum combined with other agents as advised by oncologists are different (PR+CR: 40% vs. 71% vs. 80%, $p=0.011$). As shown in Figure 5.3, methylation at *PIK3R5*, *AKT1*, *VEGFB*, *TR2IT1* and *PRDX2* are varied in the patients received different treatments, especially between the patients treated by platinum alone and combination of Platinum and Taxane. Therefore, different treatments in non-randomised clinical trial might be a confounding factor in the response analysis, suggesting the correlation analysis between methylation and response should be performed in the patients stratified by chemotherapy in a larger patient cohort.

The genes involved in the key signaling pathways were obtained mainly from KEGG database including the classical representation of the established pathways collected from literatures. Although KEGG provides valuable resources for selecting the targets in methylation profiling study, in many cases, the boundary of the pathway is vague because the components included in the pathway are based on the curator's knowledge and experience (Lu et al., 2007). Even though the core components of the pathways in different signalling pathway database are generally similar, the additional components involved in the pathway can be very different across different databases. For example, 148 genes are presented in the Wnt pathway in KEGG database (hsa04310, <http://www.genome.jp/kegg/>), while only 25 genes are involved in the same pathway in BioCarta (ID: H_wntPathway, <http://www.biocarta.com/>). Expansion of 10 signalling pathways with the additional components that are potentially involved in the signaling transduction as well as further classification of the components based on their specific function in a pathway are necessary to uncover how the methylation changes of multiple components influence the signalling transduction in the pathways.



Biomarkers	Platinum vs. combination	Platinum vs. others	combination vs. others
PIK3R5	0.5294	0.0073**	0.0394*
AKT1	0.0255*	0.2583	0.5734
VEGFB	0.0006**	0.2917	0.0771
PRDX2	0.0117*	0.1691	0.5734
TR2IT1	0.0374*	0.7485	0.0826
MLH3	0.0870	0.3712	0.8026

Figure 5.3: Methylation of 6 biomarkers in the patients received different treatments. * $p < 0.05$ (Man-Whitney test); ** $p < 0.01$ (Mann-Whitney test).

In conclusion, methylation profiles of promoter CGIs in 9 key signalling pathways were achieved and their prognostic value was evaluated in terms of association with progression free survival in ovarian cancers. Methylation of promoter CGIs at *PI3R5*, *AKT1*, *VEGFB*, *PRDX2*, *TR2IT2* and *MLH3* is correlated with poor PFS and independent from clinical parameters. Meanwhile, reduced expression of *PRDX2* and *MLH3* showed weak association with unfavourable PFS in another patient cohort.

Chapter 6

General Discussion and Future Perspectives

Ovarian cancer is the leading cause of all the gynaecological cancers. Intrinsic or acquired drug resistance to conventional treatment is one of major obstacles in ovarian cancer treatment (Agarwal and Kaye, 2003). Advance of high-throughput methylation profiling makes evaluation of the association between methylation of CGIs and acquired drug resistance and identification of methylation prognostic/predictive biomarkers in large-scale methylation studies possible (Laird, 2010). Using biostatistical analysis enhanced by bioinformatics approaches to identify such biomarkers in ovarian cancer has been explored in this thesis using restriction enzyme-based methylation profiling technologies.

6.1 Association of Hypermethylation, DNA Damage and Cisplatin Treatment

Using a novel method, Methylation Linear Discriminant Analysis (MLDA) (see Chapter 2), we found more aberrant hypermethylation than hypomethylation of CpG islands arose during the exposure of ovarian cancer cell lines to cisplatin, indicating mechanisms driving the occurrence of hypermethylation, rather than hypomethylation, in the acquisition of cisplatin resistance. DNA is the primary target of cisplatin. Formation of cisplatin-DNA adducts interferes with DNA replication and transcription. The interstrand and intrastrand crosslinks caused by cisplatin disrupt the DNA structure, which can be recognised by cellular proteins to repair the damage (Balch et al., 2004). It is possible that the methylation changes occur in the cancer cell lines is a consequence of DNA damage caused by cisplatin and repair of DNA, especially double-strand repair, in the genome. Consistent with this hypothesis, recent data has shown that DNA methylation at exogenous introduced sequences can be influenced by DNA double strand breaks (DSB) in cells and DNMT1 is present at sites of DNA repair (Mortusewicz et al., 2005; Cuzzo et al., 2007; O'Hagan et al., 2008; Palii et al., 2008). Although it is still unclear if this occurs at endogenous genes, this could

potentially explain why many loci become hypermethylated during acquired cisplatin resistance.

6.2 More Samples are Required for the Validation Study

As shown in Chapter 4 and 5, epithelial ovarian tumours prospectively collected from a SGCTG/NCRI cohort study have been systematically examined by DMH assay at promoter CGIs of 10 key signalling pathways and we have shown that levels of methylation of multiple CGIs have the potential to be prognostic biomarkers for progression-free survival. Although the validation study was adequately powered for the majority of hypermethylated loci, this was not the case for hypomethylated loci which generally had lower hazard ratios. At least a further 70 tumours with progression events are required to reach 90% power for all loci.

6.3 Biological Mechanisms of Methylation at the Wnt Pathways Contributing to Ovarian Cancer Progression

DNA methylation at the loci that are associated with patients' clinical outcome helps to define key genes and pathways important for ovarian cancer progression. The Wnt pathways are especially important, given 6 independent biomarkers (*FZD4*, *NKD1*, *DVL1*, *ROCK1*, *NFATC3*, *FZD9*) from this pathway showed the correlation with progression free survival (Chapter 4), accounting for 50% of the biomarkers that have been identified in this project. Three major pathways are implicated in Wnt signalling transductions, Wnt/ β -catenin (canonical), PCP and Wnt/ Ca^{2+} pathway (Boyer et al., 2009). However, the genes relevant to the 6 biomarkers identified are involved in different Wnt pathways as shown in Figure 6.1. As discussed in Chapter 4, based on the methylation status of 5 biomarkers

(FZD4, NKD1, NFATC3, ROCK1, and FZD9) (note methylation status of *DVL1* could not be determined due to the limitation of the method used, see Discussion in Chapter 4), we defined 5 groups of patients (1) "no methylation", (2) "mild Wnt/ β -catenin methylation", (3) "Wnt/ β -catenin or PCP methylation", (4) "Wnt/Ca²⁺ methylation" and (5) "severe methylation" in all three Wnt pathways. "Wnt/Ca²⁺ methylation" group have slightly shorter PFS compared to the patients with "mild Wnt/ β -catenin methylation" and "Wnt/ β -catenin or PCP methylation", indicating methylation alterations in Wnt/Ca²⁺ pathway probably have more impact on disease progression than those in the other two pathways in ovarian cancer. However, this systematic approach does not provide a functional read-out of Wnt pathways, so experimental investigation of specific candidate Wnt genes for their potential role in cell transformation and response to treatment in EOC is necessary, especially for *DVL1* and *FZD4* that have been shown 1) methylation associates with PFS and overall survival in a manner that is independent from other clinical/pathological criteria, 2) an association between methylation and expression and 3) an association between expression and survival in independent tumour cohorts (TCGA pilot study).

6.4 Where Shall We Look at Besides Promoter CpG Islands?

Although methylation at promoter CGI is of importance in regulating gene expression, an increasing number of studies show several regions are potentially interesting for methylation analysis. First, certain relationship between expression of the genes and methylation within gene body but not in CGI have been observed (Zilberman et al., 2007; Maunakea et al., 2010), indicating intragenic methylation plays a role in regulating transcription. Second, it is estimated that about 30-40% of the genes have relatively low CpG density at the promoter regions (non-CGI promoters), and those regions are observed to be generally

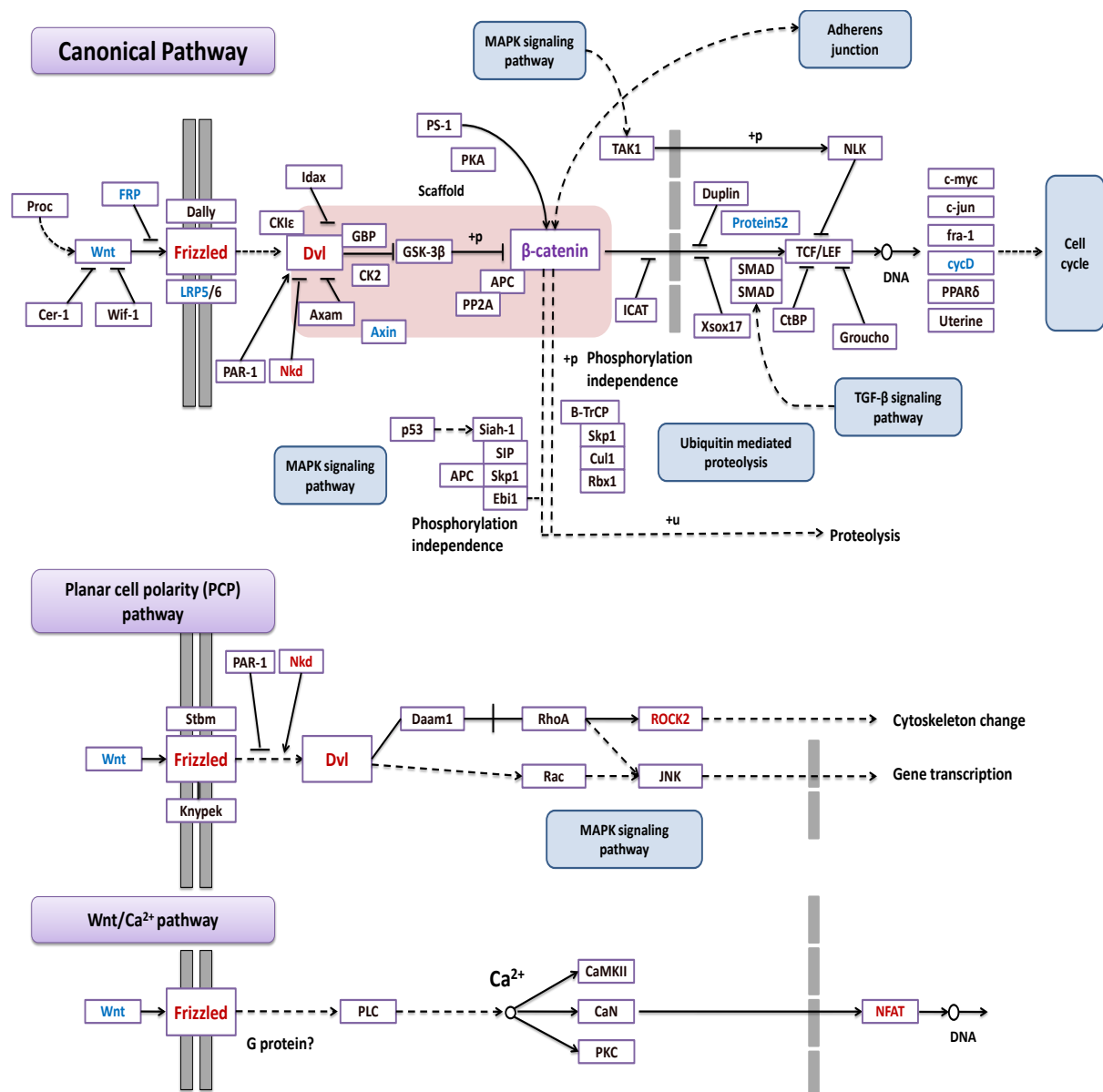


Figure 6.1: Methylation interruption in the Wnt signaling pathways. The genes highlighted in red are relevant to the biomarkers have been identified in the screening set and validated in an independent cohort, while the genes highlighted in blue are relevant to the biomarkers that have been identified in the screening set, but not validated in the following study.

methyated in somatic tissues, not associated with activities of the genes (Shen et al., 2007; Weber et al., 2007). However, a large number of these regions are located at alternative promoters that are within the gene body of dominantly expressed transcript which are generally methylated in normal cells. Identification of 'true' non-CGI promoter irrelevant to intragenic methylation would be helpful to evaluate the relation between methylation at such regions and expression of the relevant genes. A subset of non-CGI promoters have been found unmethylated in the somatic tissues when the genes become active (Weber et al., 2007), making methylation changes of these regions potentially interesting. Third, although CGIs at alternative promoters that are annotated in Refseq database and/or UCSC known gene database have been analysed in this study, some intragenic CGIs not associated with known genes have not been included in the analysis. Since these CGIs are potentially related to the alternative promoters that have not been identified yet and methylation at these regions can influence the elongation of transcription (Illingworth and Bird, 2009), evaluation of methylation at these regions might be also of interest. Therefore, the probe coverage on Agilent custom-designed 60K microarray used in the validation study of candidate methylation biomarkers has been expanded from promoter CGI to +2k to -8kb distance of the transcription start site of genes involved in the key signalling pathways, allowing evaluation of methylation at gene body and non-CGI promoter regions in ovarian cancer in the future studies.

6.5 Expansion of Methylation Model for Predicting Patients' Clinical Outcome

The prognostic values of multiple promoter CGIs were evaluated and validated individually in this study, and then a multivariate Cox model combining multiple methylation biomarkers from the Wnt pathway was constructed. Only two methylation biomarkers

at *DVL1* and *NKDI* were incorporated into the model and identified two groups of patients with favourable and unfavourable progression free survival (see Chapter 4). Given single pathway analysis is probably too simple to accurately reflect tumour progression and chemosensitivity which involve multiple biological mechanisms. We attempted to explore the possibility to expand the multivariate methylation model using all independent biomarkers identified in this project including promoter methylation at *FZD4*, *NKDI*, *DVL1*, *ROCK1*, *NFATC3*, *FZD9* from the Wnt pathways, *PIK3R5*, *AKT1*, *VEGFB* (or *DNAJC4*) from AKT/mTOR pathways, *PRDX2*, *TR2IT1* from Redox pathway, and *MLH3* from MMR system. Among 12 biomarkers, 4 methylation biomarkers at *FZD4*, *NFATC3*, *VEGFB* (or *DNAJC4*) and *PRDX2* were incorporated into the model using forward likelihood ratio stepwise algorithm. The methylation index (MI) calculated from this model can identify two groups of patients with distinguished PFS and OS (PFS: $p = 5.39 \times 10^{-8}$, events=134, Figure 6.2A; OS: $p = 9 \times 10^{-3}$, death=93, Figure 6.2B) and increased MI is significantly correlated to the patients with poor response to platinum-based chemotherapy ($p=0.02$, OR=3.32, 95% CI 1.21 to 9.17, logistic regression), suggesting the association between MI and PFS may be partly related to the likelihood of achieving good response with first-line chemotherapy. However, excluding the patients without response to chemotherapy, this MI is still associated with PFS in the patients who achieved complete response or partial response to chemotherapy ($n=54$, HR=3.32, 95% CI 1.82 to 5.95, $p = 8.08 \times 10^{-5}$), indicating the prognostic value of this MI is partly independent from response to first-line treatment. Consistent with the prognostic value of multivariate methylation model, the multivariate expression model constructed by *FZD4*, *NFATC3*, *DNAJC4* and *PRDX2* is also significantly relevant to disease free survival in patient cohort with late-stage disease generated by The Cancer Genome Atlas (TCGA) pilot study ($n=321$, multivariate Cox model $p=0.003$; *FZD4*: HR=0.8, 95% CI 0.69 to 0.93, $p=0.003$; *NFATC3*: HR=0.96, 95%

CI 0.76 to 1.0, $p=0.707$; *DNAJC4*: HR=0.62, 95% CI 0.37 to 1.05; *PRDX2*: HR=0.69, 95% CI 0.48 to 1.0, $p=0.052$). Though the predictive values of both expression and methylation models are still awaiting for further assessment either by cross-validation or ideally by other patient cohort with adequate power, this shows incorporating multiple biological mechanisms are potentially more valuable in prediction of patients' outcome and response. However, based on the "rule of thumb" we cannot include more variables than one-tenth of the number of events available for the analysis (Van Belle, 2008). In our case 151 patients with late-stage disease are included in the multivariate analysis with 134 events (disease progression), suggesting at maximum only 13 loci can be incorporated into the multivariate Cox model. Therefore, a much larger patient cohort is required to identify a methylation signature probably including hundreds of loci using more robust method, such as supervised gene shaving (Hastie et al., 2000), which can extract coherent methylation at multiple promoters and gene-body regions that vary across the samples in such a way that the methylation is correlated with clinical outcome.

6.6 Methylation as Biomarkers for Surgery Decision

Volume of residual disease following surgery is an important prognostic factor in ovarian cancer (Bristow et al., 2002). Meta-analysis shows a positive correlation between surgical debulk status and survival in advanced disease (Hunter et al., 1992). However, in a number of patients who are thought to be suitable for the debulking surgery based on the preoperative assessment, an operation is not undertaken because disease resection will result in long-term morbidity or patient demise. Chemotherapy is allowed to start earlier in those patients, therefore, identification of biomarkers which can predict the resectability of tumours would aid clinical decision making in terms of which patients should have primary surgery or neoadjuvant treatment to achieve a complete macroscopic debulk. Unfortun-

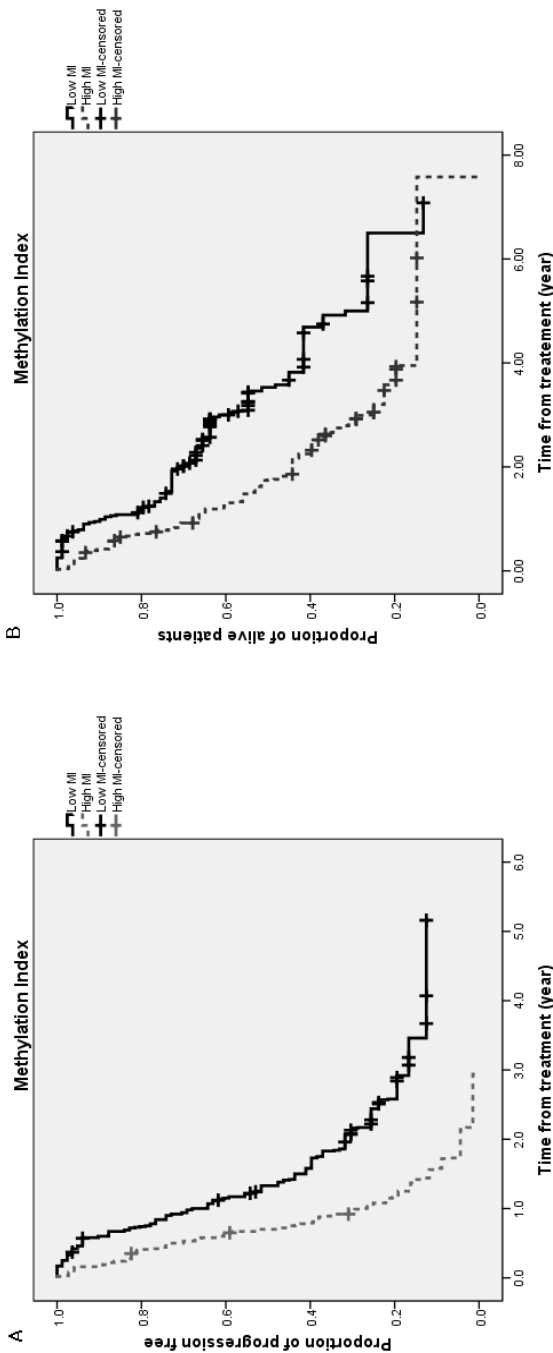


Figure 6.2: Kaplan Meier plot of survival curves of late-stage patients with high (n=74) and low (n=74) methylation index estimated from promoter of FZD4, NFATC3, VEGFB (or DNAJC4) and PRDX2. A: progression free survival curves; B: overall survival curves.

nately, only a few samples from this study have surgical annotation, though methylation of ROCK1 did show the association with the level of residual disease following surgery ($p=0.032$). We have proposed to further examine the correlation between the loci that are shown to be associated with PFS in SGCTG cohort study, and a discovery analysis for new candidates associated with success of surgical resection from the data generated by DMH on Agilent 8x60K array using the samples collected in Hammersmith Hospital, UK and by Illumina Infinium on HumanMethylation27 chips generated by TCGA pilot study. Meanwhile, development of a non-invasive method to detect methylation of multiple selected loci in the serum or plasma of ovarian cancer patients in a robust manner with rapid assay times will allow us to apply these findings in the treatment of ovarian cancer, effectively and efficiently.

6.7 Identification of Aberrantly Methylated 'Hot Spots' in Ovarian Tumours

Methylation of a small region (about 150bps) at promoter CGI of *DVL1*, one biomarker identified from the Wnt pathway, was examined by bisulfite sequencing and the result shows that methylation at this locus is significantly correlated with DMH ratios ($p=0.006$, see results in Chapter 4) in tumours, but methylation level is generally low ($<10\%$) and differential methylation at *DVL1* among the tumours is subtle which is different from what we observed in ovarian cancer cell line model where dramatic increase of methylation ($>15\%$) at the examined loci were observed in the cell lines resistant to cisplatin compared to the sensitive ones. There are two possibilities to explain this: (1) the examined locus generally has low methylation in all tumour cells and (2) the locus has high methylation only in a small subset of tumour cells due to the heterogeneity of disease. If the former possibility is the case, an extensive bisulphite sequencing of the chromosomal region targeted by DMH

is required. Therefore, we have proposed to validate differential methylation using Illumina GAIIX high throughput DNA sequencing of bisulfite modified DNA and PCR amplification of target loci to define the exact location. If the latter possibility is the case, identification of the subpopulation of tumour cells, such as drug resistant tumour stem cells, and characterisation of this subpopulation is of importance to understand how this subset of tumour cells contributes to the tumour progression and have impact on patient survival.

6.8 Methylation as Predictive Biomarkers for Treatment Decision

Combination of platinum and taxane are recommended in ovarian cancer treatment (du Bois et al. 2005). However, some patients can not receive the combined treatment due to poor performance status. We explored the potential of using the prognostic biomarkers identified in this study as predictive biomarkers for treatment decision making (single or combined treatment regimen). Of 12 independent biomarkers, methylation at *DVL1*, *ROCK1*, *PIK3R5*, *AKT1*, *VEGFB* (or *DNAJC4*), *TR2IT1* and *PRDX2* show a trend of differential methylation between patients receiving platinum alone (n=45), combined treatment (n=91) and other platinum-based treatment (n=21). In the patients received the combined treatment, lower methylation of *DVL1* and *VEGFB* (or *DNAJC4*) are found in the responders (PR+CR, n=34) than non-responders (PD+SD, n=14) (Figure 6.3) who have significantly higher risk of disease progression than the former group (HR=2.48, p=0.01). The number of samples available for this exploratory analysis is small. Further evaluation and validation analyses of these biomarkers in a randomised clinical trial following a more specific analysis plan for predictive biomarkers of platinum alone or combined treatment are clearly essential.

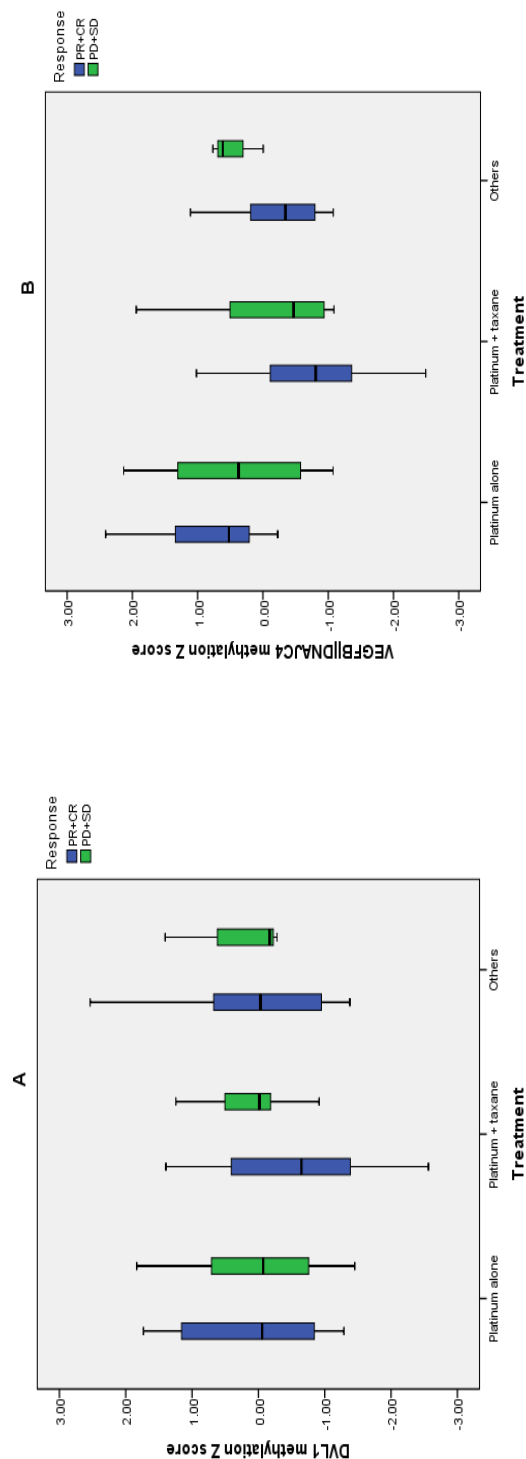


Figure 6.3: Differential methylation observed in the patients received different platinum-based chemotherapy (platinum alone n=45, platinum and taxane n=91, other platinum-based treatment n=21). A: methylation at DVL1; B: methylation at VEGFB or DNAJC4.

6.9 Methylation and Novel Treatments Targeting Molecular Pathways

Several molecular therapies targeting key pathways, such as poly(adenosine diphosphate-ribose) polymerase (PARP) inhibitor (Bast and Mills, 2010; Yap et al., 2009) (ClinicalTrials.gov ID: NCT00494442) targeting base excision repair in the patients with BRCA mutation, Wee1-inhibitor targeting G2 checkpoint in the patients with p53 mutation (ClinicalTrials.gov ID: NCT01164995), mTOR inhibitors targeting PI3K-AKT pathway (ClinicalTrials.gov ID: NCT00429793), have been approved for Clinical trials in ovarian cancer. Methylation changes identified in this study at the key pathways in the patients received platinum-based chemotherapy will be relevant for patient stratification, in future clinical trials of ovarian cancer, particularly for novel drugs targeting the Wnt pathway (Luu et al., 2004), redox pathway, AKT/mTOR pathway (Yap et al., 2009) as well as MMR system.

References

- R. Agarwal and S. B. Kaye. Ovarian cancer: strategies for overcoming resistance to chemotherapy. *Nat Rev Cancer*, 3(7):502–16, 2003.
- A. A. Ahmed, D. Etemadmoghadam, J. Temple, A. G. Lynch, M. Riad, R. Sharma, C. Stewart, S. Fereday, C. Caldas, A. Defazio, D. Bowtell, and J. D. Brenton. Driver mutations in tp53 are ubiquitous in high grade serous carcinoma of the ovary. *J Pathol*, 221(1): 49–56, 2010.
- D. A. Altomare and J. R. Testa. Perturbations of the akt signaling pathway in human cancer. *Oncogene*, 24(50):7455–64, 2005.
- F. Antequera and A. Bird. Number of cpg islands and genes in human and mouse. *Proc Natl Acad Sci U S A*, 90(24):11995–9, 1993.
- D. A. Anthoney, A. J. McIlwrath, W. M. Gallagher, A. R. Edlin, and R. Brown. Microsatellite instability, apoptosis, and loss of p53 function in drug-resistant tumor cells. *Cancer Res*, 56(6):1374–81, 1996.
- C. Balch, T. H. Huang, R. Brown, and K. P. Nephew. The epigenetics of ovarian cancer drug resistance and resensitization. *Am J Obstet Gynecol*, 191(5):1552–72, 2004.
- Z. Barekati, R. Radpour, C. Kohler, B. Zhang, P. Toniolo, P. Lenner, Q. Lv, H. Zheng, and

- X. Y. Zhong. Methylation profile of tp53 regulatory pathway and mtdna alterations in breast cancer patients lacking tp53 mutations. *Hum Mol Genet*, 19(15):2936–46, 2010.
- C. A. Barton, N. F. Hacker, S. J. Clark, and P. M. O’Brien. Dna methylation changes in ovarian cancer: implications for early diagnosis, prognosis and treatment. *Gynecol Oncol*, 109(1):129–39, 2008.
- Jr. Bast, R. C. and G. B. Mills. Personalizing therapy for ovarian cancer: Brcaness and beyond. *J Clin Oncol*, 28(22):3545–8, 2010.
- Jr. Bast, R. C., B. Hennessy, and G. B. Mills. The biology of ovarian cancer: new opportunities for translation. *Nat Rev Cancer*, 9(6):415–28, 2009.
- S.B. Baylin and P.A. Jones. Epigenetics. In CD Allis, T Jenuwein, and D Reinberg, editors, *Epigenetic determinants of cancer*. John Inglis, NewYork, 2006.
- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- L. B. Bennett, K. H. Taylor, G. L. Arthur, F. B. Rahmatpanah, S. I. Hooshmand, and C. W. Caldwell. Epigenetic regulation of wnt signaling in chronic lymphocytic leukemia. *Epigenomics*, 2(1):53–70, 2010.
- A. Berchuck, A. Kamel, R. Whitaker, B. Kerns, G. Olt, R. Kinney, J. T. Soper, R. Dodge, D. L. Clarke-Pearson, P. Marks, and et al. Overexpression of her-2/neu is associated with poor survival in advanced epithelial ovarian cancer. *Cancer Res*, 50(13):4087–91, 1990.
- M. Bibikova, Z. Lin, L. Zhou, E. Chudin, E. W. Garcia, B. Wu, D. Doucet, N. J. Thomas, Y. Wang, E. Vollmer, T. Goldmann, C. Seifart, W. Jiang, D. L. Barker, M. S. Chee,

- J. Floros, and J. B. Fan. High-throughput dna methylation profiling using universal bead arrays. *Genome Res*, 16(3):383–93, 2006.
- A. Bird. Dna methylation patterns and epigenetic memory. *Genes Dev*, 16(1):6–21, 2002.
- A. P. Bird. CpG-rich islands and the function of dna methylation. *Nature*, 321(6067):209–13, 1986.
- C. Bock, J. Walter, M. Paulsen, and T. Lengauer. CpG island mapping by epigenome prediction. *PLoS Comput Biol*, 3(6):e110, 2007.
- D. D. Bowtell. The genesis and evolution of high-grade serous ovarian cancer. *Nat Rev Cancer*, 10(11):803–8, 2010.
- A. Boyer, A. K. Goff, and D. Boerboom. Wnt signaling in ovarian follicle biology and tumorigenesis. *Trends Endocrinol Metab*, 21(1):25–32, 2009.
- R. E. Bristow, R. S. Tomacruz, D. K. Armstrong, E. L. Trimble, and F. J. Montz. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol*, 20(5):1248–59, 2002.
- R. Brown, G. L. Hirst, W. M. Gallagher, A. J. McIlwrath, G. P. Margison, A. G. van der Zee, and D. A. Anthoney. hmlh1 expression and cellular responses of ovarian tumour cells to treatment with cytotoxic anticancer agents. *Oncogene*, 15(1):45–52, 1997.
- S. A. Cannistra. Cancer of the ovary. *N Engl J Med*, 351(24):2519–29, 2004.
- D. Cheishvili, C. Maayan, Y. Smith, G. Ast, and A. Razin. Ikap/help1 deficiency in the cerebrum of familial dysautonomia patients results in down regulation of genes involved in oligodendrocyte differentiation and in myelination. *Hum Mol Genet*, 16(17):2097–104, 2007.

- J. W. Chiang, B. Y. Karlan, L. Cass, and R. L. Baldwin. Brca1 promoter methylation predicts adverse ovarian cancer prognosis. *Gynecol Oncol*, 101(3):403–10, 2006.
- K. R. Cho and M. Shih Ie. Ovarian cancer. *Annu Rev Pathol*, 4:287–313, 2009.
- J. L. Chou, H. Y. Su, L. Y. Chen, Y. P. Liao, C. Hartman-Frey, Y. H. Lai, H. W. Yang, D. E. Deatherage, C. T. Kuo, Y. W. Huang, P. S. Yan, S. H. Hsiao, C. K. Tai, H. J. Lin, R. V. Davuluri, T. K. Chao, K. P. Nephew, T. H. Huang, H. C. Lai, and M. W. Chan. Promoter hypermethylation of fbxo32, a novel tgf-beta/smad4 target gene and tumor suppressor, is associated with poor prognosis in human ovarian cancer. *Lab Invest*, 90(3):414–25, 2010.
- B. C. Christensen, K. T. Kelsey, S. Zheng, E. A. Houseman, C. J. Marsit, M. R. Wrensch, J. L. Wiemels, H. H. Nelson, M. R. Karagas, L. H. Kushi, M. L. Kwan, and J. K. Wiencke. Breast cancer dna methylation profiles are associated with tumor size and alcohol and folate intake. *PLoS Genet*, 6(7):e1001043, 2010.
- T. G. Clark, M. E. Stewart, D. G. Altman, H. Gabra, and J. F. Smyth. A prognostic model for ovarian cancer. *Br J Cancer*, 85(7):944–52, 2001.
- D. N. Cooper and M. Krawczak. Cytosine methylation and the fate of cpg dinucleotides in vertebrate genomes. *Hum Genet*, 83(2):181–8, 1989.
- S. E. Cottrell and P. W. Laird. Sensitive detection of dna methylation. *Ann N Y Acad Sci*, 983:120–30, 2003.
- A. P. Crijns, R. S. Fehrmann, S. de Jong, F. Gerbens, G. J. Meersma, H. G. Klip, H. Hollema, R. M. Hofstra, G. J. te Meerman, E. G. de Vries, and A. G. van der Zee. Survival-related profile, pathways, and transcription factors in ovarian cancer. *PLoS Med*, 6(2):e24, 2009.

- S. H. Cross, J. A. Charlton, X. Nan, and A. P. Bird. Purification of cpg islands using a methylated dna binding column. *Nat Genet*, 6(3):236–44, 1994.
- C. P. Crum, R. Drapkin, D. Kindelberger, F. Medeiros, A. Miron, and Y. Lee. Lessons from brca: the tubal fimbria emerges as an origin for pelvic serous cancer. *Clin Med Res*, 5(1):35–44, 2007.
- C. Cuozzo, A. Porcellini, T. Angrisano, A. Morano, B. Lee, A. Di Pardo, S. Messina, R. Iuliano, A. Fusco, M. R. Santillo, M. T. Muller, L. Chiariotti, M. E. Gottesman, and E. V. Avvedimento. Dna damage, homology-directed repair, and dna methylation. *PLoS Genet*, 3(7):e110, 2007.
- M. Curradi, A. Izzo, G. Badaracco, and N. Landsberger. Molecular mechanisms of gene silencing mediated by dna methylation. *Mol Cell Biol*, 22(9):3157–73, 2002.
- K. Curtin, W. Y. Lin, R. George, M. Katory, J. Shorto, L. A. Cannon-Albright, G. Smith, D. T. Bishop, A. Cox, and N. J. Camp. Genetic variants in xrrc2: new insights into colorectal cancer tumorigenesis. *Cancer Epidemiol Biomarkers Prev*, 18(9):2476–84, 2009.
- W. Dai, J. M. Teodoridis, J. Graham, C. Zeller, T. H. Huang, P. Yan, J. K. Vass, R. Brown, and J. Paul. Methylation linear discriminant analysis (mlda) for identifying differentially methylated cpg islands. *BMC Bioinformatics*, 9:337, 2008.
- C. De Smet, C. Lurquin, B. Lethe, V. Martelange, and T. Boon. Dna methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a cpg-rich promoter. *Mol Cell Biol*, 19(11):7327–35, 1999.
- S. Deneberg, M. Grovdal, M. Karimi, M. Jansson, H. Nahi, A. Corbacioglu, V. Gaidzik, K. Dohner, C. Paul, T. J. Ekstrom, E. Hellstrom-Lindberg, and S. Lehmann. Gene-

- specific and global methylation patterns predict outcome in patients with acute myeloid leukemia. *Leukemia*, 24(5):932–41, 2010.
- Y. Deng, B. Yu, Q. Cheng, J. Jin, H. You, R. Ke, N. Tang, Q. Shen, H. Shu, G. Yao, Z. Zhang, and W. Qin. Epigenetic silencing of wif-1 in hepatocellular carcinomas. *J Cancer Res Clin Oncol*, 2010.
- A. du Bois, M. Quinn, T. Thigpen, J. Vermorken, E. Avall-Lundqvist, M. Bookman, D. Bowtell, M. Brady, A. Casado, A. Cervantes, E. Eisenhauer, M. Friedlaender, K. Fujiwara, S. Grenman, J. P. Guastalla, P. Harper, T. Hogberg, S. Kaye, H. Kitchener, G. Kristensen, et al. 2004 consensus statements on the management of ovarian cancer: final document of the 3rd international gynecologic cancer intergroup ovarian cancer consensus conference (gcig occc 2004). *Ann Oncol*, 16 Suppl 8:viii7–viii12, 2005.
- B. K. Dunn, M. Verma, and A. Umar. Epigenetics in cancer prevention: Early detection and risk assessment. *Annals of the New York Academy of Sciences*, 983(1):1–4, 2003.
- F. Eckhardt, J. Lewin, R. Cortese, V. K. Rakyen, J. Attwood, M. Burger, J. Burton, T. V. Cox, R. Davies, T. A. Down, C. Haefliger, R. Horton, K. Howe, D. K. Jackson, J. Kunde, C. Koenig, J. Liddle, D. Niblett, T. Otto, R. Pettett, S. Seemann, C. Thompson, T. West, J. Rogers, A. Olek, K. Berlin, and S. Beck. Dna methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*, 38(12):1378–85, 2006.
- S. Eden and H. Cedar. Role of dna methylation in the regulation of transcription. *Curr Opin Genet Dev*, 4(2):255–9, 1994.
- E. A. Eisenhauer, M. Gore, and J. P. Neijt. Ovarian cancer: should we be managing patients with good and bad prognostic factors in the same manner? *Ann Oncol*, 10 Suppl 1:9–15, 1999.

- R England and M Pettersson. Pyro q-cpgrtm: quantitative analysis of methylation in multiple cpg sites by pyrosequencing. *Nature Methods*, 2:doi:10.1038/nmeth800, 2005.
- B. Evers, T. Helleday, and J. Jonkers. Targeting homologous recombination repair defects in cancer. *Trends in Pharmacological Sciences*, 31(8):372–380, 2010.
- A. P. Feinberg and B. Tycko. The history of cancer epigenetics. *Nat Rev Cancer*, 4(2):143–53, 2004.
- H. Fiegl, A. Jones, C. Hauser-Kronberger, G. Hutarew, R. Reitsamer, R. L. Jones, M. Dowsett, E. Mueller-Holzner, G. Windbichler, G. Daxenbichler, G. Goebel, C. Enssinger, I. Jacobs, and M. Widschwendter. Methylated neurod1 promoter is a marker for chemosensitivity in breast cancer. *Clin Cancer Res*, 14(11):3494–502, 2008.
- J. R. Flavell, K. R. Baumforth, V. H. Wood, G. L. Davies, W. Wei, G. M. Reynolds, S. Morgan, A. Boyce, G. L. Kelly, L. S. Young, and P. G. Murray. Down-regulation of the tgf-beta target gene, ptpkr, by the epstein-barr virus encoded ebna1 contributes to the growth and survival of hodgkin lymphoma cells. *Blood*, 111(1):292–301, 2008.
- D. Frank, I. Keshet, M. Shani, A. Levine, A. Razin, and H. Cedar. Demethylation of cpg islands in embryonic cells. *Nature*, 351(6323):239–41, 1991.
- J. Furuta, Y. Nobeyama, Y. Umebayashi, F. Otsuka, K. Kikuchi, and T. Ushijima. Silencing of peroxiredoxin 2 and aberrant methylation of 33 cpg islands in putative promoter regions in human malignant melanomas. *Cancer Res*, 66(12):6080–6, 2006.
- M. Gardiner-Garden and M. Frommer. Cpg islands in vertebrate genomes. *J Mol Biol*, 196(2):261–82, 1987.

- G. Gasparini, F. Pozza, and A. L. Harris. Evaluating the potential usefulness of new prognostic and predictive indicators in node-negative breast cancer patients. *J Natl Cancer Inst*, 85(15):1206–19, 1993.
- T. A. Gatliffe, B. J. Monk, K. Planutis, and R. F. Holcombe. Wnt signaling in ovarian tumorigenesis. *Int J Gynecol Cancer*, 18(5):954–62, 2008.
- G. Gifford, J. Paul, P. A. Vasey, S. B. Kaye, and R. Brown. The acquisition of hmlh1 methylation in plasma dna after chemotherapy predicts poor survival for ovarian cancer patients. *Clin Cancer Res*, 10(13):4420–6, 2004.
- R. Goold, M. Hubank, A. Hunt, J. Holton, R. P. Menon, T. Revesz, M. Pandolfo, and A. Matilla-Duenas. Down-regulation of the dopamine receptor d2 in mice lacking ataxin 1. *Hum Mol Genet*, 16(17):2122–34, 2007.
- R. T. Greenlee, M. B. Hill-Harmon, T. Murray, and M. Thun. Cancer statistics, 2001. *CA Cancer J Clin*, 51(1):15–36, 2001.
- J. Hall, J. Paul, and R. Brown. Critical evaluation of p53 as a prognostic marker in ovarian cancer. *Expert Rev Mol Med*, 6(12):1–20, 2004.
- D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100(1):57–70, 2000.
- T. Hastie, R. Tibshirani, M. B. Eisen, A. Alizadeh, R. Levy, L. Staudt, W. C. Chan, D. Botstein, and P. Brown. 'gene shaving' as a method for identifying distinct sets of genes with similar expression patterns. *Genome Biol*, 1(2):RESEARCH0003, 2000.
- L. Havrilesky, M. Darcy, H. Hamdan, R. L. Priore, J. Leon, J. Bell, and A. Berchuck. Prognostic significance of p53 mutation and p53 overexpression in advanced epithelial ovarian cancer: a gynecologic oncology group study. *J Clin Oncol*, 21(20):3814–25, 2003.

- J. D. Hayes and M. McMahon. Nrf2 and keap1 mutations: permanent activation of an adaptive response in cancer. *Trends Biochem Sci*, 34(4):176–88, 2009.
- L. A. Hazlehurst, T. H. Landowski, and W. S. Dalton. Role of the tumor microenvironment in mediating de novo resistance to drugs and physiological mediators of cell death. *Oncogene*, 22(47):7396–7402, 2003.
- J. He, J. B. Qiao, and H. Zhu. p14(arf) promoter region methylation as a marker for gliomas diagnosis. *Med Oncol*, 2010.
- P. J. Heagerty, T. Lumley, and M. S. Pepe. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics*, 56(2):337–44, 2000.
- T. Helleday. Pathways for mitotic homologous recombination in mammalian cells. *Mutat Res*, 532(1-2):103–15, 2003.
- J. G. Herman, J. R. Graff, S. Myohanen, B. D. Nelkin, and S. B. Baylin. Methylation-specific pcr: a novel pcr assay for methylation status of cpg islands. *Proc Natl Acad Sci U S A*, 93(18):9821–6, 1996.
- C. H. Holschneider and J. S. Berek. Ovarian cancer: epidemiology, biology, and prognostic factors. *Semin Surg Oncol*, 19(1):3–10, 2000.
- O. Hopfer, D. Zwahlen, M. F. Fey, and S. Aepli. The notch pathway in ovarian carcinomas and adenomas. *Br J Cancer*, 93(6):709–18, 2005.
- S. Houshdaran, S. Hawley, C. Palmer, M. Campan, M. N. Olsen, A. P. Ventura, B. S. Knudsen, C. W. Drescher, N. D. Urban, P. O. Brown, and P. W. Laird. Dna methylation profiles of ovarian epithelial carcinoma tumors and cell lines. *PLoS One*, 5(2):e9359, 2010.

- F. Y. Hsieh and P. W. Lavori. Sample-size calculations for the cox proportional hazards regression model with nonbinary covariates. *Control Clin Trials*, 21(6):552–60, 2000.
- L. Hu, C. McArthur, and R. B. Jaffe. Ovarian cancer stem-like side-population cells are tumourigenic and chemoresistant. *Br J Cancer*, 102(8):1276–83, 2010.
- T. H. Huang, M. R. Perry, and D. E. Laux. Methylation profiling of cpg islands in human breast cancer cells. *Hum Mol Genet*, 8(3):459–70, 1999.
- R. W. Hunter, N. D. Alexander, and W. P. Soutter. Meta-analysis of surgery in advanced ovarian carcinoma: is maximum cytoreductive surgery an independent determinant of prognosis? *Am J Obstet Gynecol*, 166(2):504–11, 1992.
- A. E. Ibrahim, N. P. Thorne, K. Baird, N. L. Barbosa-Morais, S. Tavare, V. P. Collins, A. H. Wyllie, M. J. Arends, and J. D. Brenton. Mmass: an optimized array-based method for assessing cpg island methylation. *Nucleic Acids Res*, 34(20):e136, 2006.
- R. S. Illingworth and A. P. Bird. CpG islands—'a rough guide'. *FEBS Lett*, 583(11):1713–20, 2009.
- R. A. Irizarry, C. Ladd-Acosta, B. Carvalho, H. Wu, S. A. Brandenburg, J. A. Jeddloh, B. Wen, and A. P. Feinberg. Comprehensive high-throughput arrays for relative methylation (charm). *Genome Res*, 18(5):780–90, 2008.
- A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun. Cancer statistics, 2009. *CA Cancer J Clin*, 59(4):225–49, 2009.
- T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–80, 2001.

- D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, and X. Cheng. Structure of dnmt3a bound to dnmt3l suggests a model for de novo dna methylation. *Nature*, 449(7159):248–51, 2007.
- P. A. Jones. Overview of cancer epigenetics. *Semin Hematol*, 42(3 Suppl 2):S3–8, 2005.
- P. A. Jones and P. W. Laird. Cancer epigenetics comes of age. *Nat Genet*, 21(2):163–7, 1999.
- H. F. Jorgensen, K. Adie, P. Chaubert, and A. P. Bird. Engineering a high-affinity methyl-cpg-binding protein. *Nucleic Acids Res*, 34(13):e96, 2006.
- M. Kanehisa, S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. From genomics to chemical genomics: new developments in kegg. *Nucleic Acids Res*, 34(Database issue):D354–7, 2006.
- M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa. Kegg for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*, 38(Database issue):D355–60, 2010.
- D. Katsaros, W. Cho, R. Singal, S. Fracchioli, I. A. Rigault De La Longrais, R. Arisio, M. Massobrio, M. Smith, W. Zheng, J. Glass, and H. Yu. Methylation of tumor suppressor gene p16 and prognosis of epithelial ovarian cancer. *Gynecol Oncol*, 94(3):685–92, 2004.
- A. Kikuchi and H. Yamamoto. Tumor formation due to abnormalities in the beta-catenin-independent pathway of wnt signaling. *Cancer Sci*, 99(2):202–8, 2008.
- D. W. Kindelberger, Y. Lee, A. Miron, M. S. Hirsch, C. Feltmate, F. Medeiros, M. J. Callahan, E. O. Garner, R. W. Gordon, C. Birch, R. S. Berkowitz, M. G. Muto, and C. P. Crum. Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: Evidence for a causal relationship. *Am J Surg Pathol*, 31(2):161–9, 2007.

- G. J. Klarmann, A. Decker, and W. L. Farrar. Epigenetic gene silencing in the wnt pathway in breast cancer. *Epigenetics*, 3(2):59–63, 2008.
- R. J. Klose and A. P. Bird. Genomic dna methylation: the mark and its mediators. *Trends Biochem Sci*, 31(2):89–97, 2006.
- M. Kobel, D. Huntsman, and C. B. Gilks. Critical molecular abnormalities in high-grade serous carcinoma of the ovary. *Expert Rev Mol Med*, 10:e22, 2008.
- P. A. Konstantinopoulos, D. Spentzos, and S. A. Cannistra. Gene-expression profiling in epithelial ovarian cancer. *Nat Clin Pract Oncol*, 5(10):577–87, 2008.
- V. Kulasingam, M. P. Pavlou, and E. P. Diamandis. Integrating high-throughput technologies in the quest for effective biomarkers for ovarian cancer. *Nat Rev Cancer*, 10(5):371–8, 2010.
- P. W. Laird. Principles and challenges of genome-wide dna methylation analysis. *Nat Rev Genet*, 11(3):191–203, 2010. Journal article.
- E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.
- S. P. Langdon, S. S. Lawrie, F. G. Hay, M. M. Hawkes, A. McDonald, I. P. Hayward, D. J. Schol, J. Hilgers, R. C. Leonard, and J. F. Smyth. Characterization and properties of nine human ovarian adenocarcinoma cell lines. *Cancer Res*, 48(21):6166–72, 1988.
- J. Lapointe, C. Li, C. P. Giacomini, K. Salari, S. Huang, P. Wang, M. Ferrari, T. Hernandez-Boussard, J. D. Brooks, and J. R. Pollack. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res*, 67(18):8504–10, 2007.

- S. K. Lau, P. C. Boutros, M. Pintilie, F. H. Blackhall, C. Q. Zhu, D. Strumpf, M. R. Johnston, G. Darling, S. Keshavjee, T. K. Waddell, N. Liu, D. Lau, L. Z. Penn, F. A. Shepherd, I. Jurisica, S. D. Der, and M. S. Tsao. Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol*, 25(35):5562–9, 2007.
- Y. K. Lee and N. H. Park. Prognostic value and clinicopathological significance of p53 and pten in epithelial ovarian cancers. *Gynecol Oncol*, 112(3):475–80, 2009.
- Y. N. Lee, Y. Gao, and H. Y. Wang. Differential mediation of the wnt canonical pathway by mammalian dishevelleds-1, -2, and -3. *Cell Signal*, 20(2):443–52, 2008.
- S. L’Esperance, I. Popa, M. Bachvarova, M. Plante, N. Patten, L. Wu, B. Tetu, and D. Bachvarov. Gene expression profiling of paired ovarian tumors obtained prior to and following adjuvant chemotherapy: molecular signatures of chemoresistant tumors. *Int J Oncol*, 29(1):5–24, 2006.
- A. J. Levine, W. Hu, and Z. Feng. The p53 pathway: what questions remain to be explored? *Cell Death Differ*, 13(6):1027–36, 2006.
- M. Li, C. Balch, J. S. Montgomery, M. Jeong, J. H. Chung, P. Yan, T. H. Huang, S. Kim, and K. P. Nephew. Integrated analysis of dna methylation and gene expression reveals specific signaling pathways associated with platinum resistance in ovarian cancer. *BMC Med Genomics*, 2:34, 2009.
- S. L. Lim, P. Smith, N. Syed, C. Coens, H. Wong, M. van der Burg, P. Szlosarek, T. Crook, and J. A. Green. Promoter hypermethylation of fancf and outcome in advanced ovarian cancer. *Br J Cancer*, 98(8):1452–6, 2008.
- S. M. Lipkin, V. Wang, R. Jacoby, S. Banerjee-Basu, A. D. Baxevanis, H. T. Lynch, R. M.

- Elliott, and F. S. Collins. Mlh3: a dna mismatch repair gene associated with mammalian microsatellite instability. *Nat Genet*, 24(1):27–35, 2000.
- Z. Lippman, A. V. Gendrel, M. Black, M. W. Vaughn, N. Dedhia, W. R. McCombie, K. Lavine, V. Mittal, B. May, K. D. Kasschau, J. C. Carrington, R. W. Doerge, V. Colot, and R. Martienssen. Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998):471–6, 2004.
- R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. Human dna methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271):315–22, 2009.
- D. Lister-Sharp, M. S. McDonagh, K. S. Khan, and J. Kleijnen. A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer. *Health Technol Assess*, 4(17):1–113, 2000.
- K. G. Louie, B. C. Behrens, T. J. Kinsella, T. C. Hamilton, K. R. Grotzinger, W. M. McKoy, M. A. Winker, and R. F. Ozols. Radiation survival parameters of antineoplastic drug-sensitive and -resistant human ovarian cancer cell lines and their modification by buthionine sulfoximine. *Cancer Res*, 45(5):2110–5, 1985.
- L. J. Lu, A. Sboner, Y. J. Huang, H. X. Lu, T. A. Gianoulis, K. Y. Yip, P. M. Kim, G. T. Montelione, and M. B. Gerstein. Comparing classical pathways and modern networks: towards the development of an edge ontology. *Trends Biochem Sci*, 32(7):320–31, 2007.
- H. H. Luu, R. Zhang, R. C. Haydon, E. Rayburn, Q. Kang, W. Si, J. K. Park, H. Wang, Y. Peng, W. Jiang, and T. C. He. Wnt/beta-catenin signaling pathway as a novel cancer drug target. *Curr Cancer Drug Targets*, 4(8):653–71, 2004.

- A. Lyakhovich and J. Surralles. Disruption of the fanconi anemia/brca pathway in sporadic cancer. *Cancer Lett*, 232(1):99–106, 2006.
- M. Maekawa, T. Taniguchi, H. Higashi, H. Sugimura, K. Sugano, and T. Kanno. Methylation of mitochondrial dna is not a useful marker for cancer detection. *Clin Chem*, 50(8):1480–1, 2004.
- P. K. Majumder and W. R. Sellers. Akt-regulated pathways in prostate cancer. *Oncogene*, 24(50):7465–74, 2005.
- S. Mallett, A. Timmer, W. Sauerbrei, and D. G. Altman. Reporting of prognostic studies of tumour markers: a review of published articles in relation to remark guidelines. *Br J Cancer*, 102(1):173–80, 2010.
- A. Malpica, M. T. Deavers, K. Lu, D. C. Bodurka, E. N. Atkinson, D. M. Gershenson, and E. G. Silva. Grading ovarian serous carcinoma using a two-tier system. *Am J Surg Pathol*, 28(4):496–504, 2004.
- A. K. Maunakea, R. P. Nagarajan, M. Bilenky, T. J. Ballinger, C. D’Souza, S. D. Fouse, B. E. Johnson, C. Hong, C. Nielsen, Y. Zhao, G. Turecki, A. Delaney, R. Varhol, N. Thiessen, K. Shchors, V. M. Heine, D. H. Rowitch, X. Xing, C. Fiore, M. Schillebeeckx, S. J. Jones, D. Haussler, M. A. Marra, M. Hirst, T. Wang, and J. F. Costello. Conserved role of intragenic dna methylation in regulating alternative promoters. *Nature*, 466(7303):253–7, 2010.
- S. L. McDonald and A. Silver. The opposing roles of wnt-5a in cancer. *Br J Cancer*, 101(2):209–14, 2009.
- L. M. McShane, D. G. Altman, W. Sauerbrei, S. E. Taube, M. Gion, and G. M. Clark.

- Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol*, 23(36): 9067–72, 2005.
- A. Meindl, H. Hellebrand, C. Wiek, V. Erven, B. Wappenschmidt, D. Niederacher, M. Freund, P. Lichtner, L. Hartmann, H. Schaal, J. Ramser, E. Honisch, C. Kubisch, H. E. Wichmann, K. Kast, H. Deissler, C. Engel, B. Muller-Myhsok, K. Neveling, M. Kiechle, C. G. Mathew, D. Schindler, R. K. Schmutzler, and H. Hanenberg. Germline mutations in breast and ovarian cancer pedigrees establish *rad51c* as a human cancer susceptibility gene. *Nat Genet*, 42(5):410–4, 2010.
- A. J. Mikels and R. Nusse. Purified *wnt5a* protein activates or inhibits *beta*-catenin-*tcf* signaling depending on receptor context. *PLoS Biol*, 4(4):e115, 2006.
- J. R. Miller, A. M. Hocking, J. D. Brown, and R. T. Moon. Mechanism and function of signal transduction by the *wnt*/*beta*-catenin and *wnt*/*ca2+* pathways. *Oncogene*, 18(55): 7860–72, 1999.
- O. Mortusewicz, L. Schermelleh, J. Walter, M. C. Cardoso, and H. Leonhardt. Recruitment of *dna* methyltransferase *i* to *dna* repair sites. *Proc Natl Acad Sci U S A*, 102(25):8905–9, 2005.
- S. Nautiyal, V. E. Carlton, Y. Lu, J. S. Ireland, D. Flaucher, M. Moorhead, J. W. Gray, P. Spellman, M. Mindrinos, P. Berg, and M. Faham. High-throughput method for analyzing methylation of *cpgs* in targeted genomic regions. *Proc Natl Acad Sci U S A*, 107(28):12587–92, 2010.
- M. Nouzova, N. Holtan, M. M. Oshiro, R. B. Isett, J. L. Munoz-Rodriguez, A. F. List, M. L. Narro, S. J. Miller, N. C. Merchant, and B. W. Futscher. Epigenomic changes during

- leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using cpg island microarrays. *J Pharmacol Exp Ther*, 311(3):968–81, 2004.
- E. Ntougkos, R. Rush, D. Scott, T. Frankenberg, H. Gabra, J. F. Smyth, and G. C. Sellar. The iglon family in epithelial ovarian cancer: expression profiles and clinicopathologic correlates. *Clin Cancer Res*, 11(16):5764–8, 2005.
- H. M. O’Hagan, H. P. Mohammad, and S. B. Baylin. Double strand breaks can initiate gene silencing and sirt1-dependent onset of dna methylation in an exogenous promoter cpg island. *PLoS Genet*, 4(8):e1000155, 2008.
- J. M. Ordway, J. A. Bedell, R. W. Citek, A. Nunberg, A. Garrido, R. Kendall, J. R. Stevens, D. Cao, R. W. Doerge, Y. Korshunova, H. Holemon, J. D. McPherson, N. Lakey, J. Leon, R. A. Martienssen, and J. A. Jeddloh. Comprehensive dna methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis*, 27(12):2409–23, 2006.
- C. O’Riain, D. M. O’Shea, Y. Yang, R. Le Dieu, J. G. Gribben, K. Summers, J. Yeboah-Afari, L. Bhaw-Rosun, C. Fleischmann, C. A. Mein, T. Crook, P. Smith, G. Kelly, A. Rosenwald, G. Ott, E. Campo, L. M. Rimsza, E. B. Smeland, W. C. Chan, N. Johnson, R. D. Gascoyne, S. Reimer, R. M. Braziel, G. W. Wright, L. M. Staudt, T. A. Lister, and J. Fitzgibbon. Array-based dna methylation profiling in follicular lymphoma. *Leukemia*, 23(10):1858–66, 2009.
- S. S. Palii, B. O. Van Emburgh, U. T. Sankpal, K. D. Brown, and K. D. Robertson. Dna methylation inhibitor 5-aza-2’-deoxycytidine induces reversible genome-wide dna damage that is distinctly influenced by dna methyltransferases 1 and 3b. *Mol Cell Biol*, 28(2):752–71, 2008.

- I. Pasic, A. Shlien, A. D. Durbin, D. J. Stavropoulos, B. Baskin, P. N. Ray, A. Novokmet, and D. Malkin. Recurrent focal copy-number changes and loss of heterozygosity implicate two noncoding rnas and one tumor suppressor gene at chromosome 3q13.31 in osteosarcoma. *Cancer Res*, 70(1):160–71, 2010.
- J. A. Plumb, G. Strathdee, J. Sludden, S. B. Kaye, and R. Brown. Reversal of drug resistance in human tumor xenografts by 2'-deoxy-5-azacytidine-induced demethylation of the hmlh1 gene promoter. *Cancer Res*, 60(21):6039–44, 2000.
- S. Pradhan and P. O. Esteve. Mammalian dna (cytosine-5) methyltransferases and their expression. *Clin Immunol*, 109(1):6–16, 2003.
- F. B. Rahmatpanah, S. Carstens, J. Guo, O. Sjahputera, K. H. Taylor, D. Duff, H. Shi, J. W. Davis, S. I. Hooshmand, R. Chitma-Matsiga, and C. W. Caldwell. Differential dna methylation patterns of small b-cell lymphoma subclasses with different clinical behavior. *Leukemia*, 20(10):1855–62, 2006.
- T. Rauch and G. P. Pfeifer. Methylated-cpg island recovery assay: a new technique for the rapid detection of methylated-cpg islands in cancer. *Lab Invest*, 85(9):1172–80, 2005.
- B. Rhead, D. Karolchik, R. M. Kuhn, A. S. Hinrichs, A. S. Zweig, P. A. Fujita, M. Diekhans, K. E. Smith, K. R. Rosenbloom, B. J. Raney, A. Pohl, M. Pheasant, L. R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R. A. Harte, B. Giardine, T. R. Dreszer, H. Clawson, G. P. Barber, D. Haussler, and W. J. Kent. The ucsc genome browser database: update 2010. *Nucleic Acids Res*, 38(Database issue):D613–9, 2010.
- A. Ricken, P. Lochhead, M. Kontogiannia, and R. Farookhi. Wnt signaling in the ovary: identification and compartmentalized expression of wnt-2, wnt-2b, and frizzled-4 mrnas. *Endocrinology*, 143(7):2741–9, 2002.

- S. Rizzo, J. Hersey, P. Mellor, W. Dai, A. Santos-Silva, D. Liber, L. Luk, I. Titley, C. Carden, G. Box, D. Hudson, S. Kaye, and B. Robert. Ovarian cancer stem cell like side populations are enriched following chemotherapy and overexpress ezh2. *Mol Cancer Ther*, In press, 2011.
- S. L. Rose. Notch signaling pathway in ovarian cancer. *Int J Gynecol Cancer*, 19(4):564–6, 2009.
- D. G. Rosen, G. Yang, G. Liu, I. Mercado-Urbe, B. Chang, X. S. Xiao, J. Zheng, F. X. Xue, and J. Liu. Ovarian cancer: pathology, biology, and disease models. *Front Biosci*, 14:2089–102, 2009.
- G. J. Rustin, A. E. Nelstrop, P. McClean, M. F. Brady, W. P. McGuire, W. J. Hoskins, H. Mitchell, and H. E. Lambert. Defining response of ovarian carcinoma to initial chemotherapy according to serum ca 125. *J Clin Oncol*, 14(5):1545–51, 1996.
- M. Sakuma, J. Akahira, K. Ito, H. Niikura, T. Moriya, K. Okamura, H. Sasano, and N. Yae-gashi. Promoter methylation status of the cyclin d2 gene is associated with poor prognosis in human epithelial ovarian cancer. *Cancer Sci*, 98(3):380–6, 2007.
- D. A. Schoenfeld. Sample-size formula for the proportional-hazards regression model. *Biometrics*, 39(2):499–503, 1983.
- D. E. Schones and K. Zhao. Genome-wide approaches to studying chromatin modifications. *Nat Rev Genet*, 9(3):179–191, 2008.
- G. C. Sellar, K. P. Watt, G. J. Rabiasz, E. A. Stronach, L. Li, E. P. Miller, C. E. Massie, J. Miller, B. Contreras-Moreira, D. Scott, I. Brown, A. R. Williams, P. A. Bates, J. F. Smyth, and H. Gabra. Opcml at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. *Nat Genet*, 34(3):337–43, 2003.

- E. Senturk, S. Cohen, P. R. Dottino, and J. A. Martignetti. A critical re-appraisal of brca1 methylation studies in ovarian cancer. *Gynecol Oncol*, 2010.
- S. Sharma, T. K. Kelly, and P. A. Jones. Epigenetics in cancer. *Carcinogenesis*, 31(1): 27–36, 2010.
- L. Shen, Y. Kondo, Y. Guo, J. Zhang, L. Zhang, S. Ahmed, J. Shu, X. Chen, R. A. Waterland, and J. J. Issa. Genome-wide profiling of dna methylation reveals a class of normally methylated cpg island promoters. *PLoS Genet*, 3(10):e181, 2007.
- H. Shimodaira. Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *The Annals of Statistics*, 32(6):2616–2641, 2004.
- Z. H. Siddik. Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene*, 22(47):7265–79, 2003.
- R. M. Simon, E. L. Korn, L. M. McShane, M. D. Radmacher, G. W. Wright, and Y. Zhao. *Design and Analysis of DNA Microarray Investigations*. Statistics for Biology and Health. New York, New York, 2003.
- E. Smith, T. Bianco-Miotto, P. Drew, and D. Watson. Method for optimizing methylation-specific pcr. *Biotechniques*, 35(1):32–3, 2003.
- A. K. Sood and R. E. Buller. Drug resistance in ovarian cancer: from the laboratory to the clinic. *Obstet Gynecol*, 92(2):312–9, 1998.
- H. M. Sowter and A. Ashworth. Brca1 and brca2 as ovarian cancer susceptibility genes. *Carcinogenesis*, 26(10):1651–6, 2005.
- A. B. Sparks, P. J. Morin, B. Vogelstein, and K. W. Kinzler. Mutational analysis of the apc/beta-catenin/tcf pathway in colorectal cancer. *Cancer Res*, 58(6):1130–4, 1998.

- J. Staub, J. Chien, Y. Pan, X. Qian, K. Narita, G. Aletti, M. Scheerer, L. R. Roberts, J. Molina, and V. Shridhar. Epigenetic silencing of hsnf-1 in ovarian cancer: implications in chemoresistance. *Oncogene*, 26(34):4969–78, 2007.
- N. Steele, P. Finn, R. Brown, and J. A. Plumb. Combined inhibition of dna methylation and histone acetylation enhances gene re-expression and drug sensitivity in vivo. *Br J Cancer*, 100(5):758–63, 2009.
- B. Stordal and R. Davey. A systematic review of genes involved in the inverse resistance relationship between cisplatin and paclitaxel chemotherapy: role of brca1. *Curr Cancer Drug Targets*, 9(3):354–65, 2009.
- G. Strathdee, M. J. MacKean, M. Illand, and R. Brown. A role for methylation of the hmlh1 promoter in loss of hmlh1 expression and drug resistance in ovarian cancer. *Oncogene*, 18(14):2335–41, 1999.
- G. Strathdee, J. K. Vass, K. A. Oien, N. Siddiqui, J. Curto-Garcia, and R. Brown. Demethylation of the mcj gene in stage iii/iv epithelial ovarian cancer and response to chemotherapy. *Gynecol Oncol*, 97(3):898–903, 2005.
- H. Y. Su, H. C. Lai, Y. W. Lin, C. Y. Liu, C. K. Chen, Y. C. Chou, S. P. Lin, W. C. Lin, H. Y. Lee, and M. H. Yu. Epigenetic silencing of sfrp5 is related to malignant phenotype and chemoresistance of ovarian cancer through wnt signaling pathway. *Int J Cancer*, 127(3):555–67, 2010.
- T. Sugiyama, T. Kamura, J. Kigawa, N. Terakawa, Y. Kikuchi, T. Kita, M. Suzuki, I. Sato, and K. Taguchi. Clinical characteristics of clear cell carcinoma of the ovary: a distinct histologic type with poor prognosis and resistance to platinum-based chemotherapy. *Cancer*, 88(11):2584–9, 2000.

- E. Sutherland, L. Coe, and E. A. Raleigh. Mcrbc: a multisubunit gtp-dependent restriction endonuclease. *J Mol Biol*, 225(2):327–48, 1992.
- M. M. Suzuki and A. Bird. Dna methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 9(6):465–76, 2008.
- T. Takada, Y. Yagi, T. Maekita, M. Imura, S. Nakagawa, S. W. Tsao, K. Miyamoto, O. Yoshino, T. Yasugi, Y. Taketani, and T. Ushijima. Methylation-associated silencing of the wnt antagonist sfrp1 gene in human ovarian cancers. *Cancer Sci*, 95(9):741–4, 2004.
- D. Takai and P. A. Jones. Comprehensive analysis of cpg islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*, 99(6):3740–5, 2002.
- J. M. Teodoridis, J. Hall, S. Marsh, H. D. Kannall, C. Smyth, J. Curto, N. Siddiqui, H. Gabra, H. L. McLeod, G. Strathdee, and R. Brown. Cpg island methylation of dna damage response genes in advanced ovarian cancer. *Cancer Res*, 65(19):8961–7, 2005.
- S. Teramukai, K. Ochiai, H. Tada, and M. Fukushima. Piepoc: a new prognostic index for advanced epithelial ovarian cancer–japan multinational trial organization oc01-01. *J Clin Oncol*, 25(22):3302–6, 2007.
- A. E. Teschendorff, U. Menon, A. Gentry-Maharaj, S. J. Ramus, S. A. Gayther, S. Apostolidou, A. Jones, M. Lechner, S. Beck, I. J. Jacobs, and M. Widschwendter. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*, 4(12):e8274, 2009.
- P. Therasse, S. G. Arbuck, E. A. Eisenhauer, J. Wanders, R. S. Kaplan, L. Rubinstein, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, M. C. Christian, and S. G. Gwyther.

- New guidelines to evaluate the response to treatment in solid tumors. european organization for research and treatment of cancer, national cancer institute of the united states, national cancer institute of canada. *J Natl Cancer Inst*, 92(3):205–16, 2000.
- T. Thigpen, M. F. Brady, G. A. Omura, W. T. Creasman, W. P. McGuire, W. J. Hoskins, and S. Williams. Age as a prognostic factor in ovarian carcinoma. the gynecologic oncology group experience. *Cancer*, 71(2 Suppl):606–14, 1993.
- C. Thirlwell, M. Eymard, A. Feber, A. Teschendorff, K. Pearce, M. Lechner, M. Widschwendter, and S. Beck. Genome-wide dna methylation analysis of archival formalin-fixed paraffin-embedded tissue using the illumina infinium humanmethylation27 beadchip. *Methods*, 2010.
- R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc Natl Acad Sci U S A*, 99(10):6567–72, 2002.
- T. Tong and H. Zhao. Practical guidelines for assessing power and false discovery rate for a fixed sample size in microarray experiments. *Stat Med*, 27(11):1960–72, 2008.
- R. W. Tothill, A. V. Tinker, J. George, R. Brown, S. B. Fox, S. Lade, D. S. Johnson, M. K. Trivett, D. Etemadmoghadam, B. Locandro, N. Traficante, S. Fereday, J. A. Hung, Y. E. Chiew, I. Haviv, D. Gertig, A. DeFazio, and D. D. Bowtell. Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res*, 14(16):5198–208, 2008.
- A. Tran, C. Escovedo, J. Migdall-Wilson, A. P. Chou, W. Chen, T. Cloughesy, S. Nelson, and A. Lai. In silico enhanced restriction enzyme based methylation analysis of the

- human glioblastoma genome using agilent 244k cpg island microarrays. *Front Neurosci*, 3:57, 2009.
- C. Trope and J. Kaern. Adjuvant chemotherapy for early-stage ovarian cancer: review of the literature. *J Clin Oncol*, 25(20):2909–20, 2007.
- J.W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, Reading Massachusetts, 1977.
- V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A*, 98(9):5116–21, 2001.
- M. Ueda, R. M. Gemmill, J. West, R. Winn, M. Sugita, N. Tanaka, M. Ueki, and H. A. Drabkin. Mutations of the beta- and gamma-catenin genes are uncommon in human lung, breast, kidney, cervical and ovarian carcinomas. *Br J Cancer*, 85(1):64–8, 2001.
- A. Valencia, J. Roman-Gomez, J. Cervera, E. Such, E. Barragan, P. Bolufer, F. Moscardo, G. F. Sanz, and M. A. Sanz. Wnt signaling pathway is epigenetically regulated by methylation of wnt antagonists in acute myeloid leukemia. *Leukemia*, 23(9):1658–66, 2009.
- G Van Belle. *Statistical Rules of Thumb*. Wiley, New York, 2008.
- D. C. van Gent and M. van der Burg. Non-homologous end-joining, a sticky affair. *Oncogene*, 26(56):7731–40, 2007.
- L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415(6871):530–6, 2002.
- J. Veeck, D. Niederacher, H. An, E. Klopocki, F. Wiesmann, B. Betz, O. Galm, O. Camara, M. Durst, G. Kristiansen, C. Huszka, R. Knuchel, and E. Dahl. Aberrant methylation

- of the wnt antagonist sfrp1 in breast cancer is associated with unfavourable prognosis. *Oncogene*, 25(24):3479–88, 2006.
- E. Verri, P. Guglielmini, M. Puntoni, L. Perdelli, A. Papadia, P. Lorenzi, A. Rubagotti, N. Ragni, and F. Boccardo. Her2/neu oncoprotein overexpression in epithelial ovarian cancer: evaluation of its prevalence and prognostic significance. clinical study. *Oncology*, 68(2-3):154–61, 2005.
- P. J. Verweij and H. C. Van Houwelingen. Cross-validation in survival analysis. *Stat Med*, 12(24):2305–14, 1993.
- B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits, and J. L. Bos. Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319(9):525–32, 1988a.
- B. Vogelstein, E. R. Fearon, S. R. Hamilton, S. E. Kern, A. C. Preisinger, M. Leppert, Y. Nakamura, R. White, A. M. Smits, and J. L. Bos. Genetic alterations during colorectal-tumor development. *N Engl J Med*, 319(9):525–32, 1988b.
- K. Wakana, Y. Akiyama, T. Aso, and Y. Yuasa. Involvement of gata-4/-5 transcription factors in ovarian carcinogenesis. *Cancer Lett*, 241(2):281–8, 2006.
- J. Wang and J. Yi. Cancer cell killing via ros: to increase or decrease, that is the question. *Cancer Biol Ther*, 7(12):1875–84, 2008.
- Y. Wang. Wnt/planar cell polarity signaling: a new paradigm for cancer therapy. *Mol Cancer Ther*, 8(8):2103–9, 2009.
- Y. Wang and F. C. Leung. An evaluation of new criteria for cpg islands in the human genome as gene markers. *Bioinformatics*, 20(7):1170–7, 2004.

- M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam, and D. Schubeler. Chromosome-wide and promoter-specific analyses identify sites of differential dna methylation in normal and transformed human cells. *Nat Genet*, 37(8):853–862, 2005.
- M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, and D. Schubeler. Distribution, silencing potential and evolutionary impact of promoter dna methylation in the human genome. *Nat Genet*, 39(4):457–66, 2007.
- S. H. Wei, C. M. Chen, G. Strathdee, J. Harnsomburana, C. R. Shyu, F. Rahmatpanah, H. Shi, S. W. Ng, P. S. Yan, K. P. Nephew, R. Brown, and T. H. Huang. Methylation microarray analysis of late-stage ovarian carcinomas distinguishes progression-free survival in patients and identifies candidate epigenetic markers. *Clin Cancer Res*, 8(7):2246–52, 2002.
- S. H. Wei, R. Brown, and T. H. Huang. Aberrant dna methylation in ovarian cancer: is there an epigenetic predisposition to drug response? *Ann N Y Acad Sci*, 983:243–50, 2003.
- S. H. Wei, C. Balch, H. H. Paik, Y. S. Kim, R. L. Baldwin, S. Liyanarachchi, L. Li, Z. Wang, J. C. Wan, R. V. Davuluri, B. Y. Karlan, G. Gifford, R. Brown, S. Kim, T. H. Huang, and K. P. Nephew. Prognostic dna methylation biomarkers in ovarian cancer. *Clin Cancer Res*, 12(9):2788–94, 2006.
- K. C. Wiegand, S. P. Shah, O. M. Al-Agha, Y. Zhao, K. Tse, T. Zeng, J. Senz, M. K. McConechy, M. S. Anglesio, S. E. Kalloger, W. Yang, A. Heravi-Moussavi, R. Giuliany, C. Chow, J. Fee, A. Zayed, L. Prentice, N. Melnyk, G. Turashvili, A. D. Delaney, J. Madore, S. Yip, A. W. McPherson, G. Ha, L. Bell, S. Fereday, A. Tam, L. Galletta, P. N. Tonin, D. Provencher, D. Miller, S. J. Jones, R. A. Moore, G. B. Morin, A. Oloumi, N. Boyd, S. A. Aparicio, M. Shih Ie, A. M. Mes-Masson, D. D. Bowtell, M. Hirst,

- B. Gilks, M. A. Marra, and D. G. Huntsman. Arid1a mutations in endometriosis-associated ovarian carcinomas. *N Engl J Med*, 363(16):1532–43, 2010.
- R.R. Wilcox. Robust regression. In B.A. Holland, editor, *Robust Estimation and Hypothesis Testing*. Elsevier Academic Press, London, 2005a.
- R.R. Wilcox. Estimating measures of location and scale. In B.A. Holland, editor, *Robust Estimation and Hypothesis Testing*. Elsevier Academic Press, London, 2005b.
- C. S. Wilhelm-Benartzi, D. Koestler, E. A. Houseman, B. C. Christensen, J. K. Wiencke, A. R. Schned, M. R. Karagas, K. T. Kelsey, and C. J. Marsit. Dna methylation profiles delineate etiologic heterogeneity and clinically important subgroups of bladder cancer. *Carcinogenesis*, 2010.
- R. Wu, N. Hendrix-Lucas, R. Kuick, Y. Zhai, D. R. Schwartz, A. Akyol, S. Hanash, D. E. Misek, H. Katabuchi, B. O. Williams, E. R. Fearon, and K. R. Cho. Mouse model of human ovarian endometrioid adenocarcinoma based on somatic defects in the wnt/beta-catenin and pi3k/pten signaling pathways. *Cancer Cell*, 11(4):321–33, 2007.
- P. S. Yan, C. M. Chen, H. Shi, F. Rahmatpanah, S. H. Wei, C. W. Caldwell, and T. H. Huang. Dissecting complex epigenetic alterations in breast cancer using cpg island microarrays. *Cancer Res*, 61(23):8375–80, 2001.
- P. S. Yan, C. M. Chen, H. Shi, F. Rahmatpanah, S. H. Wei, and T. H. Huang. Applications of cpg island microarrays for high-throughput analysis of dna methylation. *J Nutr*, 132(8 Suppl):2430S–2434S, 2002a.
- P. S. Yan, T. Efferth, H. L. Chen, J. Lin, F. Rodel, L. Fuzesi, and T. H. Huang. Use of cpg island microarrays to identify colorectal tumors with a high degree of concurrent methylation. *Methods*, 27(2):162–9, 2002b.

- P. S. Yan, S. H. Wei, and T. H. Huang. Differential methylation hybridization using cpg island arrays. *Methods Mol Biol*, 200:87–100, 2002c.
- T. A. Yap, C. P. Carden, and S. B. Kaye. Beyond chemotherapy: targeted therapies in ovarian cancer. *Nat Rev Cancer*, 9(3):167–81, 2009.
- Y. Ying and Q. Tao. Epigenetic disruption of the wnt/beta-catenin signaling pathway in human cancers. *Epigenetics*, 4(5):307–12, 2009.
- J. Yu, Q. Tao, Y. Y. Cheng, K. Y. Lee, S. S. Ng, K. F. Cheung, L. Tian, S. Y. Rha, U. Neumann, C. Rocken, M. P. Ebert, F. K. Chan, and J. J. Sung. Promoter methylation of the wnt/beta-catenin signaling antagonist dkk-3 is associated with poor survival in gastric cancer. *Cancer*, 115(1):49–60, 2009.
- C. Zeller and R. Brown. Therapeutic modulation of epigenetic drivers of drug resistance in ovarian cancer, 2010.
- S. Zhang, X. Zhou, H. Yu, and Y. Yu. Expression of tumor-specific antigen mage, gage and bage in ovarian cancer tissues and cell lines. *BMC Cancer*, 10:163, 2010.
- I. Zigelboim, P. J. Goodfellow, A. P. Schmidt, K. C. Walls, M. A. Mallon, D. G. Mutch, P. S. Yan, T. H. Huang, and M. A. Powell. Differential methylation hybridization array of endometrial cancers reveals two novel cancer-specific methylation markers. *Clin Cancer Res*, 13(10):2882–9, 2007.
- D. Zilberman, M. Gehring, R. K. Tran, T. Ballinger, and S. Henikoff. Genome-wide analysis of arabidopsis thaliana dna methylation uncovers an interdependence between methylation and transcription. *Nat Genet*, 39(1):61–9, 2007.