

**Identification and Characterisation of
Differentially Methylated Regions
within the human
Major Histocompatibility Complex**

Eleni M. Tomazou

The Wellcome Trust Sanger Institute
University of Cambridge
Clare College

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This thesis describes my work undertaken in the laboratory of Prof. Stephan Beck at the Wellcome Trust Sanger Institute while member of Clare College, University of Cambridge. It is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy. The work described here has not been submitted for any degree, diploma, or any other qualification. This thesis does not exceed 300, single-sided pages of double spaced text, not including the bibliography and appendices.

This dissertation is the result of my own work and includes nothing that it is the outcome done in collaboration except as detailed in the text below.

Microarray data analysis was done with the help of Gregory Lefebvre (Wellcome Trust Sanger Institute).

Bioinformatics analysis for identification of genomic features of tDMRs was performed with the help of Stephan Rice (Wellcome Trust Sanger Institute).

MHC tile-path array was printed by the Wellcome Trust Sanger Institute Microarray facility.

Eleni M. Tomazou
Cambridge, 21 July 2008

Abstract

DNA methylation is one of several epigenetic marks capable of modulating genome function. Alterations to the temporal or spatial patterns of DNA methylation give rise to differentially methylated regions (DMRs). DMRs can arise during normal development and can be associated with specific tissues (tissue-specific DMRs, tDMRs) as well as during the development of aberrant phenotypes (phenotype specific DMRs, pDMRs) and in many cases can be implicated in the aetiology of complex diseases.

This dissertation describes an array-based assay for the unbiased identification and characterisation of DMRs (both tDMRs and pDMRs) within the human Major Histocompatibility Complex (MHC). The MHC, a 4Mb region on chromosome 6, is an ideal model system for studying DMRs as it is gene dense and associated with many complex diseases including immune-linked diseases as well as cancer.

I identified and characterised 55 MHC loci as tDMRs of which about 27% could be correlated with tissue specific gene expression. This implicates DNA methylation as an additional regulatory layer in the control of MHC loci. DNA methylation was also found to be associated with the regulation of genes involved in the MHC class I antigen processing and presentation pathway. Cell lines that displayed the MHC class I⁻ phenotype, which is a common disease phenotype, were tested for the presence of pDMRs. I identified two pDMRs that were correlated with the down-regulation of the *HLA-A*, *HLA-B*, *TAP1* and *PSMB8* genes and 14 pDMRs associated with *PSMB9* up-regulation. Three DMRs were identified within the TNF gene cluster which may contribute to the development of the MHC class I⁻ phenotype. Finally, two DMRs within the promoter regions of the *PSMB8* and *B2M* genes showed strong correlation with low expression levels. These findings are consistent with previous studies supporting the notion that transcriptional gene silencing promotes DNA hypermethylation or vice versa. The former implies that, in some cases, DNA hypermethylation may be the consequence rather than the cause of gene silencing.

The genomic features and functional aspects of some of the identified DMRs were tested and it was shown that DNA methylation inhibitors can restore parts of the MHC class I pathway that were silenced by hypermethylation.

The results presented in this thesis support the role of DNA methylation in phenotypic plasticity. They complement the extensive amount of genetic data available for the MHC and open the way for the development of integrated (epi)genetic approaches to complex phenotypes and common diseases.

Publications

Publication list arising from the work described in this thesis at the time of submission:

1. **Tomazou EM** and Powell GT. Look who's talking too: graduates developing skills through communication. *Nat Rev Genet*. 2007 Sep;8(9):724-6
2. **Tomazou EM**, Rakyan VK, Lefebvre G, Andrews R, Ellis P, Jackson DK, Langford C, Francis MD, Bäckdahl L, Miretti M, Coggill P, Ottaviani D, Sheer D, Murrell A, Beck S. Generation of a genomic tiling array of the human Major Histocompatibility Complex (MHC) and its application for DNA methylation analysis. *BMC Med Genomics*. 2008 May 30;1:19.
3. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Thorne NP, Kulesha E, Gräf S, **Tomazou EM**, Bäckdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJP, Durbin R, Tavare S, Beck S. A Bayesian de-convolution strategy for immunoprecipitation-based DNA methylation analysis. *Nat Biotechnol*. 2008 Jul 8;26(7):779-785.
4. Rakyan, VK, Down TA, Thorne NP, Flicek P, Kulesha E, Gräf S, **Tomazou EM**, Bäckdahl L, Johnson N, Herberth M, Howe KL, Jackson DK, Miretti MM, Fiegler H, Marioni JC, Birney E, Hubbard TJP, Carter NP, Tavare S, Beck S. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res*. 2008 Jun 24 (online)
5. Ottaviani D, Lever E, Mitter R, Jones T, Forshew T, **Tomazou EM**, Beck S, Krawetz SA, Platts AE, Segarane B, Sheer D. Recruitment of Genome Anchors to the Nuclear Matrix: a Novel Mechanism for Regulating Expression of the Human MHC? *Genome Res*. accepted

Acknowledgments

Stephan Beck, my supervisor. Thank you for your support, advice, for keeping up my enthusiasm, and for always remembering my birthday. I could not have asked for a better mentor.

Mike Stratton for his constructive comments, as a member of my thesis committee, and for his support during the last year. It has been really helpful. Thank you very much.

Vardhman Rakyan my lab-mate, bench-mate, office-mate and a good friend. Thank you for introducing me to MeDIP and bisulphite sequencing and for the interesting discussions. Keep in touch!

Marcos Miretti, my favourite post-doc. Thank you for being there when the others had to go...Hope we meet in America – North and/or South...

Adele Murrell, my second supervisor. It was a pleasure meeting you and working with you.

ex-Team-50 members: Roger Horton, Penny Coggill, Liselotte Backdahl and Jen Sambrook. Thank you for all the good times in the office and for all the chocolate we have shared. Good luck to all of you and hope we meet again.

The Cancer Genome Project team, and especially Andy Futreal and Mike Stratton; thank you for letting me use your space, equipment and money during the last year.

The people in the analysis office; although we have not interacted a lot it was a great pleasure to meet you all!

Gregory Lefebvre, my statistician. Thank you very much for analysing my data. It was nice working with you and good luck in Lausanne.

The Wellcome Trust Sanger Institute Microarray Facilities and especially Cordelia Langford, Rob Andrews and Peter Ellis; thanks for your support during the last three years.

The Wellcome Trust Sanger Institute PhD program: for the generous financial support and training. A special thank you to Christina Hedberg-Delouka and Alex Bateman for always being helpful.

Antigone and Samrah, my 'Cambridge family'. Thank you for proof-reading my thesis and for all the great times at the "spitaki" residence. I will miss you...

Lampros and Loukia, my lovely parents; thank you for letting me go... I miss you too...

Table of contents

Chapter 1	General Introduction	1
1.1	Introduction	2
1.2	The Major Histocompatibility Complex – MHC	3
1.2.1	MHC encoded genes	3
1.2.2	MHC polymorphism	5
1.2.3	MHC-linked diseases	6
	1.2.3.1 Future challenges in studying MHC-linked diseases	8
	1.2.4 MHC and Epigenetics – What is known so far	8
1.3	Epigenetics	12
1.3.1	Definition	12
1.3.2	Epigenetic Modification in Mammalian Genomes	12
1.3.3	DNA methylation in mammals	14
1.3.4	Function of DNA methylation	16
1.3.5	DNA methylation inhibition assay	19
1.3.6	Methodologies for detection of DNA methylation	20
	1.3.6.1 Bisulphite Sequencing	21
	1.3.6.2 Methylated DNA Immunoprecipitation – MeDIP	23
1.3.7	Epigenetic variation in humans	24
	1.3.7.1 Differentially Methylated Regions – DMRs	25
	1.3.7.2 DMR identification goes global in the human genome	29
1.4	Rationale of my thesis	31
Chapter 2	Materials & Methods	33
2.1	Materials	34

2.1.1	Reagents	34
2.1.2	Commercial kits	35
2.1.3	Solutions & Buffers	35
2.1.4	DNA used for tDMR screen – chapter 4	38
2.1.5	Cell Lines Used for pDMR screen – chapter 5	38
2.1.6	Cell Culture Media and Reagents	39
2.1.7	Bacterial Clones	39
2.1.8	Key Web Addresses	40
2.2	Methods	41
2.2.1	Tissue Culture	41
2.2.1.1	Culturing of Cell Lines	41
2.2.1.2	Cell cryo-preservation	42
2.2.1.3	5-Aza-2'-Deoxycytidine Treatment	42
2.2.1.4	DNA extraction and manipulation	43
2.2.1.5	RNA manipulation	43
	2.2.1.5.1 RNA extraction	44
	2.2.1.5.2 cDNA synthesis	44
2.2.2	Methylated DNA Immunoprecipitation (MeDIP)	44
2.2.2.1	Sonication of genomic DNA	44
2.2.2.2	Immunoprecipitation - pDMR screen	45
2.2.2.3	Immunoprecipitation - tDMR screen	46
2.2.3	Bisulphite Sequencing	47
2.2.3.1	Primer Design	47
2.2.3.2	Bisulphite treatment	48
2.2.3.3	PCR amplification	48
2.2.3.4	Sequencing	49

2.2.3.5	ESME analysis	49
2.2.4	Quantitative real-time PCR	50
2.2.4.1	Primer Design	50
2.2.4.2	qRT-PCR amplification	50
2.2.4.3	qRT-PCR assay data analysis	51
2.2.5	Bacterial Cloning	53
2.2.6	Mini-preps of plasmid DNA	53
2.2.7	Restriction Digests	54
2.2.8	Colony PCR	54
2.2.9	MHC tile path array	55
2.2.9.1	Generation of amino-linked probes	55
2.2.9.2	Gap closure and control clones	55
2.2.9.3	Array printing and processing	56
2.2.10	Microarray hybridization	57
2.2.11	Microarray Scanning	57
2.2.12	Microarray Data Analysis	58
2.2.13	Identification of genomic features of DMRs	59
Chapter 3	Development and validation of an array-based assay for the identification of DMRs	60
3.1	Introduction	61
3.2	MHC tiling array	62
3.2.1	Chemistry of the MHC tiling array	62
3.2.2	Generation and quality control of MHC array probes	63
3.2.3	Validation of the MHC tiling array	65
3.2.4	Repetitive elements	66
3.3	MeDIP optimization and validation	68

3.3.1	Genomic DNA fragmentation	68
3.3.2	Validation of MeDIP	69
3.4	Application of the MHC tiling array for methylation analysis	71
3.4.1	Normalization of MHC tiling array data	71
3.4.2	MeDIP-MHC tiling array hybridization quality control	72
3.4.3	DMR identification and validation	74
3.5	Discussion	75
3.6	Conclusion	76
Chapter 4	tDMR screen	77
4.1	Introduction	78
4.2	Samples used for the tDMR screen	80
4.3	Tissue-specific DNA methylation profiles of the MHC	80
4.4	tDMR identification	82
4.5	Validation of tDMRs	83
4.5.1	Validation of tDMRs by bisulphite sequencing	84
4.5.2	Correlation with HEP data	86
4.6	Correlation of tDMRs with expression data	88
4.7	Non-redundant tDMRs	88
4.8	Genomic features of non-redundant tDMRs	91
4.9	Discussion	93
4.10	Conclusion	95
Chapter 5	pDMR screen	96
5.1	Introduction	97
5.2	Samples used for pDMR screen	101
5.3	Relative expression of MHC class I pathway genes encoded within the MHC	102

5.3.1	Expression analysis	104
5.4	Effect of DNA methylation inhibition on MHC encoded class I pathway genes	109
5.5	MHC DMRs associated with the MHC class I ⁺ phenotype	111
5.5.1	Generation of DNA methylation profiles within the MHC region	112
5.5.2	DMRs between the cancer cell lines and shared controls	113
5.5.3	pDMR identification	116
5.5.3.1	pDMRs associated with HLA-A, HLA-B, TAP1 and PSMB8 expression	116
5.5.3.2	pDMRs associated with TAP2, TAPBP, HLA-C and PSMB9 expression	120
5.6	DNA methylation and levels of transcriptional activity	123
5.6.1	DMRs overlapping with the PSMB8 promoter region	124
5.7	Prominent DMRs within the MHC region	126
5.7.1	The tumour necrosis factor cluster	127
5.7.2	DMRs within the TNF cluster	127
5.7.3	Expression of TNF cluster genes and correlation with DMRs	130
5.8	Discussion	133
5.9	Conclusion	136
Chapter 6	MHC class I pathway genes not encoded within the MHC region	137
6.1	Introduction	138
6.2	Non-MHC encoded MHC class I pathway components	138
6.3	Expression analysis of B2M, ERp57, CRT and CANX	139
6.4	Methylation analysis of the B2M gene	141

6.5	Discussion	142
6.6	Conclusion	143
Chapter 7	General Discussion	144
7.1	Introduction	145
7.2	Array-based assay for DMR identification	145
	7.2.1 Future directions	146
7.3	tDMR screen	147
	7.3.1 tDMRs within the MHC	147
	7.3.2 Genomic Features of tDMRs	147
	7.3.3 Copy number variation and DNA methylation	148
	7.3.4 Future directions	149
7.4	pDMR screen	151
	7.4.1 pDMRs within the MHC	151
	7.4.2 DMRs within the TNF cluster	152
	7.4.3 Transcriptional silencing and DNA hypermethylation	155
	7.4.4 Future directions	155
7.5	Long Range Epigenetic Silencing	156
7.6	Recombination hotspots and epigenetic events	157
7.7	Conclusion	157
	Bibliography	159
	Appendix	179

List of figures

Chapter 1 General Introduction

1.1	Gene map of the human MHC	7
1.2	Autoimmune diseases caused by complex traits	9

1.3	DNA methylation and histone modifications	13
1.4	Mechanism of DNA methylation	14
1.5	Transcriptional repression by DNA methylation	17
1.6	DNA methylation inhibitors	20
1.7	Bisulphite conversion	22
1.8	Methylated DNA immunoprecipitation (MeDIP)	24
1.9	DNA methylation heterogeneity among individuals and cell types	26
1.10	Selected landmarks in large-scale DNA methylation studies and DMR identification in the human genome	30
Chapter 2	Materials and Methods	
2.1	Validation of UBC primers	52
Chapter 3	Development and validation of an array-based assay for the identification of DMRs	
3.1	Diagrammatic representation of processing of single-stranded array probes	62
3.2	Quality control of PCR-amplified probes	64
3.3	Hybridization variation	66
3.4	Distribution and suppression of repeat sequences	67
3.5	Relationship between target fragments and array probes in methylation analysis	69
3.6	Fragmentation of genomic DNA	69
3.7	Correlation between enrichment after MeDIP and CpG density	71
3.8	Normalization within arrays	72
3.9	Comparisons of MHC tiling array hybridizations	73
3.10	Design of approach for calling DMRs	74
Chapter 4	tDMR screen	
4.1	Regulation of gene expression	78

4.2	DNA methylation profiles of the MHC	82
4.3	tDMRs within the MHC region	83
4.4	tDMR validation	85
4.5	Example of a tDMR identified by both HEP and MeDIP-MHC tiling array studies	87
4.6	Example of tDMRs correlating with tissue-specific gene expression	89
4.7	Non-redundant tDMRs within the MHC region	90
4.8	Genomic features of putative tDMRs	93
Chapter 5 pDMR screen		
5.1	MHC class I molecule	98
5.2	MHC class I antigen presentation pathway	99
5.3	Relative expression of MHC encoded MHC class I pathway genes	106
5.4	Restoration of gene expression after DNA methylation inhibition in two cancer cell lines	111
5.5	DNA methylation profiles of the MHC	113
5.6	DMRs identified between the cancer cell lines and shared controls	115
5.7	A pDMR within the TAP1/PSMB9 bidirectional promoter	118
5.8	Sequence of the DMR overlapping with the TAP1/PSMB9 bidirectional promoter	119
5.9	pDMRs associated with HLA-A, HLA-B, TAP1 and PSMB8 expression	120
5.10	pDMRs associated with PSMB9 up-regulation	122
5.11	A DMR within the PSMB8 promoter	125
5.12	DMRs within the TNF cluster	129
5.13	Relative expression of TNF cluster genes	131
5.14	TNF-cluster gene expression after DNA methylation inhibition	132

in two cancer cell lines

Chapter 6	MHC class I pathway genes not encoded within the MHC region	
6.1	Relative expression of non-MHC encoded MHC class I pathway genes	140
6.2	Gene expression after DNA methylation inhibition in two cancer cell lines	140
6.3	Methylation analysis of the B2M gene	142

List of Tables

Chapter 1	General Introduction	
1.1	Genes in the MHC in which variation has a relationship to disease	10
Chapter 2	Materials & Methods	
2.1	Tissues and cell types used in this study	38
2.2	List of all cell lines used	41
Chapter 3	Development and validation of an array-based assay for the identification of DMRs	
3.1	Summary of MHC tiling array probes	65
Chapter 4	tDMR screen	
4.1	Genomic features of non-redundant tDMRs	92
Chapter 5	pDMR screen	
5.1	Characteristics of cell lines used in the pDMR screen	102
5.2	Summary of MHC encoded MHC class I pathway gene expression	105
Appendix		
2.1	Primer sets used within this thesis	180
2.2	MHC tiling array clones	187
4.1	tDMRs within the MHC region	205
5.1	DMRs common in both tDMR and pDMR screens	206

Abbreviations

5-aza-CR	5-Azacytidine
5-aza-CdR	5-Aza-2'-deoxycytidine
5m-CpG	methyated CpG at 5-carbon position of cytosine
aCGH	array comparative genomic hybridization
ASM	allele specific DNA methylation
bp	base pair
BAC	bacterial artificial chromosome
B2M	β 2-microglobulin
BSA	bovine serum albumin
°C	degrees Celcius
CANX	calnexin
CALR	calreticulin
CGI	CpG island
ChIP	chromatin immunoprecipitation
CNV	Copy Number Variation
CpG	cytidine-guanosine dinucleotide
CIITA	MHC class II transactivator
Cy3	Cyanine 3-dCTP
Cy5	Cyanine 5-dCTP
DMR (tDMR, pDMR)	Differentially Methylated Region (tissue-specific-, phenotype-specific)
DMSO	dimethyl sulphoxide
DNMT	DNA methyltransferase
dNTP	2'-deoxyribonucleoside 5'-triphosphate
ds	double stranded
EBV	Epstein-Barr Virus
ECR	Evolutionary Conserved Region
EDTA	ethylenediamine tetra-acetic acid
ER	Endoplasmatic Reticulum

FBS	Foetal Bovine Serum
GA	Genetic Analyser
HCMV	human cytomegalovirus
HERV	human endogenous retrovirus
HEP	Human Epigenome Project
HLA	human leukocyte antigen
HSP	heat shock protein
ICF	immunodeficiency syndrome
IFN	interferon
kb	kilobase pairs
LB	Luria-Bertani broth
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LITAF	lipopolysaccharide-induced TNF- α factor
LM-PCR	Ligation Mediated PCR
LPS	lipopolysaccharide
LRES	long range epigenetic silencing
LTR	long terminal repeat
MeDIP	Methylated DNA Immunoprecipitation
MBD	methyl binding domain
μ g	microgram
MHC	Major Histocompatibility Complex
min	minute
ml	millilitre
μ l	microlitre
μ M	micromolar
mM	millimolar
mm	millimetre
NAHR	non-allelic homologous recombination
NRM	nurim
MVP	Methylation Variable Position
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
ng	nanogram

PAC	P1 artificial chromosome
PBS	phosphate buffer saline
PCR	Polymerase Chain Reaction
RFX	regulatory factor X
rpm	revolutions per minute
RRBS	Reduced Representation Bisulphite Sequencing
RT-PCR	Real Time PCR
SAM	S-adenosyl-methionine
SDS	sodium dodecyl sulphate
sec	second
SNP	Single Nucleotide Polymorphism
ss	single stranded
SSC	saline sodium citrate
TNF	Tumour Necrosis Factor
TSS	Transcription Start Site
Tris	tris(hydroxymethyl)aminomethane
U	unit
UCSC	University of California Santa Cruz
UTR	untranslated region
WTCCC	Wellcome Trust Case Control Consortium
WGA	whole genome association