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

List of figures

List of Tables

Abbreviations

5-aza-CR	5-Azacytidine
5-aza-CdR	5-Aza-2'-deoxycytidine
5m-CpG	methylated 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	β2-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'-triphophate
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
μg	microgram
MHC	Major Histocompatibility Complex
min	minute
ml	millilitre
μl	microlitre
μM	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