Folic Acid: Investigation of its relevance for DNA methylation and the enzymes that reduce it.

Alan Harrison, BSc.

A thesis presented to Dublin City University for the Degree of Doctor of Philosophy

Supervisor: Dr. Anne Parle-McDermott

School of Biotechnology

Dublin City University

December 2014

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of ..... is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date:_____

Acknowledgements

First off, I'd like to thank my supervisor, Dr. Anne Parle-McDermott, who provided excellent help and support from the beginning to the end of my time on this project. Anne was a great mentor, who was always encouraging me to bring my own ideas to the table, which I realise now was what made me grow so much as a scientist. I'd also like to thank Dr. Linda Hughes, whose expertise was invaluable during the cell culture and enzyme kinetics sections of this thesis. In fact, all members of the Nutritional Genomics research group, past and present, deserve a huge thanks for their support during this this, through both the good days and the bad months that comes with research.

A big thank you to the Biological Research Society of DCU, who provided enough well-earned distractions and 'refreshments' to keep us all motivated throughout our post-graduate studies. The DCU Weightlifting Club also deserves similar thanks, as they played a pivotal role in the writing of Chapter 3 of thesis… by allowing me to go throw heavy things around a room in a during the days of MMSDK. The amount of support a DCU student has between the clubs and societies here is incredible, and I'd urge anyone at the beginning of their studies to take advantage of these resources.

A massive thank you to my family for their perpetual support. Mam, Dad, Maria, Tara, and Ronan deserve almost as much credit as I do for producing this thesis (disclaimer: all work in this thesis is my own unless cited and acknowledged otherwise). I'd also like to thank my friends who stood by me over the past few years, especially the lads from the Generator WhatsApp group. Another big thank you to housemates Gareth and Kevin for putting up with me even as I went completely insane during the last stages of the project.

Speaking of which, I'd finally like to thank my imaginary friends, Jad Abumrad and Robert Krulwich, for keeping me company during all those late nights in the lab.

Table of Contents

Abstract 11
Abbreviations 12
Units 13
Publications and Presentations 14
Chapter 1: Introduction 15
1.1: Overview 16
1.2: Folate Metabolism 16
1.2.1: Folic Acid Supplementation 16
1.2.2: The FASSTT Study 17
1.2.3: Folate-mediated One-Carbon Metabolism 18
1.3: Epigenetics 20
1.4: DNA Methylation 22
1.5: The Foetal Origin of Human Disease 23
1.6: DNA methylation and Nutrition 24
1.7: Methods in DNA methylation analysis 26
1.7.1: Genome-wide DNA methylation analysis 26
1.7.2: Modified Methylation Specific Digital Karyotyping 27
1.7.3: Methylated DNA Immunoprecipitation 29
1.7.4: Gene-specific DNA Methylation analysis 29
1.8: Aims and Objectives 31
Figure 1.1: FASSTT Study Overview 33
Figure 1.2: Folate mediated one-carbon metabolism 34
Figure 1.3: Stages of Neural Tube Defect Development 35
Figure 1.4: Passive and Active DNA Demethylation 36
Figure 1.5: Maternal methyl donor supplementation impacts on IAP methylation 37
Figure 1.6: Modified Methylation-Specific Digital Karyotyping (MMSDK) 39
Figure 1.7: Methylated DNA Immunoprecipitation 40
Table 1.1: Clinical and Biochemical data from FASSTT cohort, taken from McNulty *et al.*, 2013[4] 41
Table 1.2: Overview of whole-genome methods of DNA methylation analysis 42
Chapter 2: Materials and Methods 43
2.1 Materials 44
2.1.1 Biological Materials 44
2.1.2: Stock Solutions and Reagents 45

2.1.3 Oligonucleotide and primer sequence list 47
2.2: Cell Culture Methods 49
2.2.1: HEK293 Cell Culture 49
2.2.2: HEK293 Cell Counts 49
2.2.3: HEK293 Cell Growth Curve 49
2.2.4: Cytotoxicity assay of 5'Azacytidine for HEK293 Cells 49
2.3: Molecular Biological Laboratory Techniques 50
2.3.1: Nucleic Acid Extraction from HEK293 Cells 50
2.3.2 Non-denaturing Polyacrylamide Gel Electrophoresis 50
2.3.3: Agarose Gel Electrophoresis 51
2.3.4: SDS PAGE and Western Blot 51
2.3.5: Estimation of nucleic acid concentration 52
2.3.6: Phenol Extraction and Precipitation of DNA 52
2.3.7: TA Cloning and Transformation of Competent Cells 52
2.3.8: Recombinant Protein Production 53
2.3.9: Recombinant Protein Fractionation 53
2.3.10: Concentration of DHFR soluble fraction 53
2.3.11: Resolubilisation of DHFR insoluble fraction via Folding and Refolding 53
2.3.12: Resolubilisation of DHFR insoluble fraction via Resuspension with N-lauroyl-sarcosine and triethanolamine 54
2.3.13: Gravity-Flow Column Purification of GST-tagged Protein 54
2.3.14: Batch Purification of GST-tagged Protein 54
2.3.15: Bradford Protein Assay 55
2.3.16: DHFR Kit assay for enzyme activity 55
2.3.17: *In vitro* SUMOylation assay 55
2.3.18: Plasmid DNA Preparation from Bacterial Cells 56
2.3.19: Sodium Bisulfite Treatment of DNA 56
2.3.20: Sensitive Melting After Real Time Methylation Specific PCR 56
2.3.21: cDNA Synthesis of HEK293 extracted RNA 57
2.3.22: Genomic DNA Contamination Test 57
2.3.23: Reverse Transcriptase-Quantitative PCR 57
2.3.24 Data Analysis and Statistics 57
Chapter 3: DNA Methylation Method Development and Identification of Folate Sensitive Differential Methylation Regions (FS-DMRs) 59
3.1 Introduction 60
3.1.1: The Human DNA Methylome 60

3.1.2: Selection of a genome-wide method of DNA methylation analysis 61
3.1.3: Aims and Objectives 62
3.2: Methods 63
3.2.1: The FASSTT Study 63
3.2.2: Annealing of MMSDK Ligation linkers 63
3.2.3: Modified methylation specific digital karyotyping 63
3.2.4: PCR amplification of MMSDK products 64
3.2.5: DNA extraction form PAGE gel 65
3.2.6: Methylated DNA Immunoprecipitation 65
3.2.7: Whole genome-wide amplification of MeDIP DNA 66
3.2.8: Quality Control Analysis of Immunoprecipitated DNA 66
3.2.9: Nimblegen Microarray Analysis 67
3.2.10: Selection of samples for genome-wide analysis 67
3.3: DNA Microarray Data Analysis 69
3.3.1: Introduction 69
3.3.2: Comparison of Pre- versus Post-intervention for each Individual 70
3.3.3: FS-DMR Comparison between folic acid intervention and placebo groups 70
3.3.4: Obtaining lists for complete DNA methylation responses to folic acid supplementation 71
3.3.5: Pregnancy related Differential Methylated Regions 71
3.3.6: In-depth analysis of FS-DMRs: Individual Enrichment Levels across the Subset 71
3.3.7: In depth analysis of FS-DMRs: Probe distribution across remaining peaks 72
3.4: Results 72
3.4.1: MMSDK Pilot Study 1, Lambda Phage DNA 72
3.4.2: MMSDK Pilot Study 2, Human Genomic DNA 73
3.4.3: Methylated DNA Immunoprecipitation, another method for genome-wide methylation Analysis 74
3.4.4: Quality Control of MeDIP DNA 74
3.4.5: Novel FS-DMRs 75
3.5: Discussion 76
Figure 3.1: Project Overview 81
Figure 3.2: Analysis overview of Microarray Data: Quantitative Methylation Changes 82
Figure 3.3: Analysis overview of Microarray Data: Complete Methylation Changes 83
Figure 3.4: Sample dataset for pre- and post-intervention comparison 84
Figure 3.5: Sample dataset for comparison between three participants 85
Figure 3.6: Sample illustration between probe and peak enrichment scores 86

Figure 3.7: MMSDK Pilot Test on Lambda DNA: PCR amplification step 87
Figure 3.8: Restriction sequences of the Lambda Genome. 88
Figure 3.9: MMSDK PCR Products on Human Genomic DNA 89
Figure 3.10: MMSDK PCR Products on Human Genomic DNA 90
Figure 3.11: Quality Control analysis of MeDIP DNA, gel electrophoresis 91
Figure 3.12: Quality Control analysis of Immunoprecipitated DNA by qPCR 92
Figure 3.13: IP6K1 MeDIP probe positions 93
Figure 3.14: Chr9ORF44 MeDIP probe positions 94
Figure 3.15: RASA4 MeDIP Probe positions 95
Figure 3.16: SGPL1 MeDIP probe positions 96
Figure 3.17: Chr19ORF75 MeDIP probe positions 97
Table 3.1: Criteria for selecting FASSTT samples for MeDIP analysis. 98
Table 3.2: Sequences of Sanger reads obtained for MMSDK. 99
Table 3.3: Genes that exhibited changes in methylation at their promoter regions in all individuals tested 100
Table 3.4: Genes that changed their methylation status from ON to OFF across all three individuals 101
Table 3.5: Genes that changed their methylation status from OFF to ON across all three individuals 102
Table 3.6: Narrowing down of FS-DMRs based on individual peak scores 103
Table 3.7: Final List of FS-DMRs 104
Table 3.8: Probe positions of FS-DMR Peaks 105
Table 3.8: Probe positions of FS-DMR Peaks (continued) 106
Chapter 4: Confirmation of candidate FS-DMR sites and assessment of their impact on gene expression 107
4.1: Introduction 108
4.1.1: Overview 108
4.1.2: The impact of DNA methylation changes on gene expression 108
4.1.3 Aims and Objectives 109
4.2: Methods 110
4.2.1: SMART MSP assay design 110
4.2.2: Rationale for FS-DMR selection for Confirmation in the FASSTT cohort and gene expression analysis 111
4.2.3: SMART MSP Optimisation 111
4.2.4: SMART MSP FASSTT Screening 112
4.2.5: HEK293 5’Azacytidine cytotoxicity assay 112

4.2.6: Reverse-Transcription Quantitative-PCR 113
4.2.7: DNA Extraction from HEK293 cells 113
4.2.8: Data analysis and statistics 113
4.3: Results 114
4.3.1: Design and Optimisation of SMART-MSP assays 114
4.3.2: Assessment of the control Col2A1 assay in the FASSTT cohort 115
4.3.3: Methylation analysis of the FASSTT cohort 115
4.3.4: Assessment of HEK293 Growth Curve in the presence of varying concentrations of 5'Azacytidine 117
4.3.5: Nucleic Acid Extraction and Integrity 117
4.3.6: Gene Expression Analysis of HEK293 Cells treated with 5'Azacytidine by RT-qPCR 117
4.3.7: Gene specific DNA methylation analysis of 5'Azacytidine treated HEK293 cells 117
4.4: Discussion 118
Figure 4.1: Melting profile of IP6K1 SMART MSP Assay 121
Figure 4.2: Melting profile of Chr9ORF44 SMART MSP Assay 122
Figure 4.3: Melting profile of RASA4 SMART MSP Assay 123
Figure 4.4: Melting profile of GPS2 SMART MSP Assay 124
Figure 4.5: Scatterplot of IP6K1 methylation changes for all FASSTT samples 125
Figure 4.6: FASSTT cohort methylation changes for IP6K1 126
Figure 4.7: Scatterplot of Chr9ORF44 methylation changes for all FASSTT samples 127
Figure 4.8: FASSTT cohort methylation changes for Chr9ORF44 128
Figure 4.9: Scatterplot of RASA4 methylation changes for all FASSTT samples 129
Figure 4.10: FASSTT cohort methylation changes for RASA4 130
Figure 4.11: HEK293 Cytotoxcity Assay with 5'Azacytidine Treatment 131
Figure 4.12: HEK293 Percentage Cell Death with 5'Azacytidine 132
Figure 4.13: RNA extracted from 5aC-treated HEK293 Cells 133
Figure 4.14: cDNA Genomic Contamination test 134
Figure 4.15: IP6K1 Gene Expression in response to 5aC Treatment 135
Figure 4.16: RASA4 Gene Expression in response to 5aC Treatment 136
Figure 4.17: GPS2 Gene Expression in response to 5aC Treatment 137
Figure 4.18: IP6K1 DNA Methylation in response to 5aC Treatment 138
Figure 4.19: RASA4 DNA Methylation in response to 5aC Treatment 139
Figure 4.20: GPS2 DNA Methylation in response to 5aC Treatment 140
Table 4.1: SMART MSP Assay Conditions 141

Table 4.2: COL2A1 Consistency check across the FASSTT cohort 142
Chapter 5: Investigation of Dihydrofolate Reductase Post-Translational Modifications 143
5.1: Introduction 144
5.1.1: The Epigenome beyond DNA methylation 144
5.1.2: Post translational modifications of Dihydrofolate Reductase 145
5.1.3: Aims and Objectives 146
5.2: Methods 146
5.2.1: *In Silico* Post-translational modification analysis 146
5.2.2: Recombinant protein production 146
5.2.3: Recombinant protein analysis 147
5.2.4: Solubilisation of DHFR-GST insoluble fractions 147
5.2.5: Purification of GST-bound DHFR 147
5.2.6: In vitro SUMOylation 148
5.2.7: Dihydrofolate Reductase Activity Assay 148
5.3: Results 149
5.3.1 *In silico* analysis of DHFR post-translational modifications 149
5.3.2: DHFR is primarily found in the insoluble fraction when the human protein is expressed by BL21-DE3 cells 149
5.3.3: Glutathione agarose purification of soluble DHFR and Western blot analysis 150
5.3.4: Resolubilisation of DHFR insoluble fraction with Guanidine-HCl and Glycerol 150
5.3.5: Resolubilisation of DHFR with N-lauroyl-sarcosine and triethanolamine 151
5.3.6: *In vitro* SUMOylation of DHFR 152
5.3.7: *In vitro* Acetylation of DHFR 153
5.4: Discussion 154
Figure 5.1: *In Silico* Post Translation Modification Analysis of DHFR 157
Figure 5.2: Peptide sequence alignment and conserved SUMOylation sites in mammals 158
Figure 5.3: Soluble and insoluble fractions of BL21-DE3 cell lysates 159
Figure 5.4: Eluted soluble DHFR 160
Figure 5.5: Standard Curve for Bradford Assay 161
Figure 5.6: Resolubilisation of DHFR using Guanidine-HCl and Glycerol 162
Figure 5.7: Purification of resolubilised DHFR-GST using Glutathione Agarose Beads 163
Figure 5.8: Glutathione purification of DHFR-GST resolubilised with N-lauroyl-sarcosine and triethanolamine 164
Figure 5.9: Glutathione column-purification of combined DHFR-GST pellets resolubilised with N-lauroyl-sarcosine and triethanolamine 165

Figure 5.10: Concentration DHFR-GST wash fraction from protein resolubilised with N-lauroyl-sarcosine and triethanolamine 166
Figure 5.11: DHFR Enzyme Activity Assay on DHFR-GST Partially Purified from Resolubilised Protein Wash Fraction 167
Figure 5.12: Products of *In vitro* SUMOylation assay of DHFR probed with anti-DHFR and anti-SUMO 168
Figure 5.13: Coomassie stain of gel used in SUMOylation experiments highlights the range of proteins present in the reaction 169
Figure 5.14: *In vitro* SUMOylation reactions on recombinant DHFR and P53 probed with anti-p53 and anti-SUMO 170
Figure 5.15: DHFR Enzyme Activity Assay on dried down recombinant DHFR from Sigma-Aldrich 171
Figure 5.16: In Silico SUMOylation analysis on UBC9 172
Chapter 6: Discussion and Future Work 173
6.1: The Impact of Folic Acid Supplementation on DNA methylation 174
6.2: Investigation of post-translational modifications affecting dihydrofolate reductase 177
6.3: Future Work 178
References 180
Appendix A: Visual Basic 6 Macro Scripts 198
A.1: Sub Macros used for removing and pasting mismatched entries 199
A.2 Looping Macro for Step 1 of data analysis 199
A.3 Sub Macros for Step 2 of data analysis 200
A.4 Looping macro for Step 2 of data analysis 201
A.5: Sub Macros for comparing Intervention DMRs against Placebo DMRs 202
A.6: Looping Macro to Remove DMRs Common in Placebo and Intervention Groups 202
Appendix B: Raw data from MeDIP qPCR QC analysis. 204
Appendix C: Pregnancy Related Differentially Methylated Regions 209
Appendix C.1: Genes that exhibited a decrease in methylation at their promoter regions in response to pregnancy 210
Appendix D: SMART MSP Assay Data 211
Figure D.1: IP6K1 SMART-MSP Assay 212
Figure D.2: Chr9ORF44 SMART MSP Assay 213
Figure D.3: RASA4 SMART MSP Assay 214
Figure D.4: GPS2 SMART MSP Assay 215
Figure D.5: COL2A1 SMART MSP Assay Standard Curve 216
Figure D.6: IP6K1 SMART MSP Assay Standard Curve 217
Figure D.7: Chr9ORF44 SMART MSP Assay Standard Curve 218

Figure D.8: RASA4 SMART MSP Assay Standard Curve 219
Figure D.9: GPS2 SMART MSP Assay Standard Curve 220
Appendix E: FASSTT Cohort Data 221
Appendix E.1: COL2A1 Screening Experiments 222
Appendix E.2: IP6K1 SMART-MSP Raw Data 232
Appendix E.3: Chr9ORF44 SMART-MSP Raw Data 240
Appendix E.4: RASA4 SMART-MSP Raw Data 250
Appendix F: Cell culture raw data 261
Appendix F.1: HEK293 5'azacytidine Cytotoxicity Assay Raw Data 262
Appendix G: RT qPCR Raw Data 263
Appendix G.1 GUS RT qPCR data 264
Appendix G.2 IP6K1 rt qPCR assay standard Curve 265
Appendix G.3 RASA4 RTqPCR assay standard Curve 265
Appendix G.4 GPS2 RTqPCR assay standard Curve 266
Appendix G.5: IP6K1 RT qPCR data 267
Appendix G.6: RASA4 RT qPCR data 267
Appendix G.7 GPS2 RT qPCR data 267
Appendix H: "DNA Methylation: A Timeline of Methods and Applications" 268

Abstract

DNA methylation is a biochemical process where a DNA base, usually cytosine, is enzymatically methylated at the 5-carbon position. An epigenetic modification associated with gene regulation, DNA methylation plays a key role in a wide variety of biological processes. Changes in DNA methylation patterns have also been implicated in the pathogenesis of many human diseases. Following up on a randomised control trial (RCT) studying the effects of prolonged folic acid supplementation during pregnancy, we sought to identify regions of the genome that altered their methylation status in response to intervention, considering the methyl-donor for DNA methyltransferase reactions is a product of folate metabolism. Using a genome-wide, array-based method of analysis on a subset of the cohort (n=6) from the RCT, we identified 5 novel folate-sensitive differentially methylated regions (FS-DMRs): Inosital hexakisphosphate kinase 1 (IP6K1), Chromosome 9 Open Reading Frame 44 (Chr9ORF44), RAS p21 protein activator 4 (RASA4), Sphingosine-1-phosphate lyase 1 (SPGL1) and Chromosome 19 Open Reading Frame 75 (Chr19ORF75). A confirmation analysis was carried out using a loci-specific method of analysis across the rest of the cohort (n = 238) on three of these FS-DMRs (IP6K1, RASA4, and Chr9ORF44), but they did not reach statistical significance. In a cell culture model, changes in methylation across selected regions were confirmed to have a significant impact on gene expression (IP6K1, RASA4 and GPS2). A final analysis of dihydrofolate reductase (DHFR), responsible for reducing folic acid to its active form, investigated whether specific post-translational modifications was a feature of this enzyme. Following up on a SUMOylation analysis of DHFR previously published in the literature, our own analysis did not yield the same results, exposing a potential source of false positive results from a commercially available kit wildly used in this field.

Abbreviations

5cC	5’carboxylcytosine
5fC	5’formylcytosine
5hmC	5’hydroxymethylcytosine
5mC	5’methylcytosine
AFPPS	Aspirin and Folic Acid Polyp Prevention Study
APS	Ammonium Persulfate
BSA	Bovine serum albumin
cDNA	Complementary Deoxyribonucleic Acid
Chr19ORF75	Chromosome 19 Open Reading Frame 75
Chr9ORF44	Chromosome 9 Open Reading Frame 44
DHFR	Dihyrdofolate Reductase
DMEM	Dublecco’s Modified Eagle’s Medium
DMR	Differentially Methylated Region
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
DNMT	DNA Methyltransferase
dNTP	Deoxyribonucleotide Triphosphate
EDTA	Ethylenediamine Tetracetic Acid
EtOH	Ethanol
FASSTT	Folic Acid Supplementation during the Second and Third Trimester
FBS	Foetal Bovine Serum
FS-DMR	Folate Sensitive Differentially Methylated Region
GPS2	G protein pathway suppressor 2
HEK	Human Embryonic Kidney
IAP	Intracisternal A Particle
IP6K1	Inosital hexakisphosphate kinase 1
IPTG	Isopropyl β- D –thiogalactoside
LoTE	Low Tris Buffer EDTA
MeDIP	Methylated DNA Immunoprecipitation
MMSDK	Modified Methylation Specific Digital Karyotyping
mRNA	Messenger RNA
MTHFD1	Methylenetetrahydrofolate Dehydrogenase
MTHFR	Methylenetetrahydrofolate Reductase
MTX	Methotrexate
NGS	Next Generation Sequencing
NTD	Neural Tube Defect
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase Chain Reaction
qPCR	Quantitative PCR
RASA4	RAS p21 protein activator 4
RLGS	Restriction Landmark Genomic Scanning
RNA	Ribonucleic Acid
RNase	Ribonuclease
RT-qPCR	Reverse Transcriptase Quantitative PCR
SAM	S-adenosylmethionine
SDS	Sodium Dodecyl Sulphate
SHMT	Serine Hydroxymethyltransferase

SMART-MSP	Sensitive Melting After Real Time Methylation Specific PCR
SNP	Single Nucleotide Polymorphism
lncRNA	Long non-coding RNA
SPGL1	Sphingosine-1-phosphate lyase 1
SUMO	Small Ubiquitin-like Modifier Protein
TBE	Tris Borate EDTA
TBS	Tris Buffered Saline
TBST	Tris Buffered Saline plus Tween
TEMED	Tetramethylethylenediamine
TET1	Ten-eleven translocation methylcytosine dioxygenase 1
THF	Tetrahydrofolate

Units

bp	Base Pair
cm^2	Centimetre squared
C_P	Crossing point
g	Gram or gravities
hr	Hour(s)
kDa	Kilodalton
M	Molar
mA	Milliamps
mg	Milligram
min	Minute(s)
ml	Millilitre
mM	Millimolar
nm	Nanometre
°C	Degrees Celsius
rpm	Revolutions Per Minute
sec	Second(s)
V	Volts
W	Watts
U	Units
µg	Microgram
µl	Microlitre
µm	Micrometre
µM	Micromolar

Publications and Presentations

Harrison, A. & Parle-McDermott, A. DNA methylation: a timeline of methods and applications. *Front. Genet.* **2,** 74 (2011).

Harrison, A. Pentieva, K, McNulty, H, & Parle-McDermott, A. The effect of a folic acid intervention trial on DNA methylation patterns in pregnant women. Poster Presentation, 9th International Homocysteine and One-Carbon Metabolism Conference, 2013.

Harrison, A. Modified Methylation Specific Digital Karyotyping (MMSDK): A novel sequencing method for the analysis of DNA methylation patterns on a genome-wide scale. Poster Presentaiton, School of Biotechnology Research Day, 2012.

Harrison, A. Pentieva, K, McNulty, H, & Parle-McDermott, A. The effect of a folic acid intervention trial on DNA methylation patterns in pregnant women. Poster Presentation, 2014 Keystone Symposia Conference, D2: Epigenetic Programming and Inheritance.

Publications in preparation:

Harrison, A. Pentieva, K. Ozaki, M. McNulty, H. & Parle-McDermott, A. The impact of a folic acid intervention trial on DNA methylation patterns in pregnant women: a genome-wide and gene-specific approach (journal to be confirmed).

Harrison A, Hughes L, Parle-McDermott A. Potential false positive results from Active Motif's SUMOlink kit. Bio-techniques.

Chapter 1:

Introduction

1.1: Overview

The aim of this project is to assess the effects of folic acid supplementation on genome-wide DNA methylation patterns i.e., to identify human 'Folate-Sensitive DNA methylation sites." As an epigenetic modification known to influence gene expression, DNA methylation has been the focus of extensive research over the past decade due to its potential to be a mechanism for the "foetal origins of adult disease" hypothesis, i.e. the effect of nutritional and environmental factors during development on health later in life[1–3].

In an intervention study recently carried out by the University of Ulster[4], 119 pregnant women were recruited and given either a 400μg folic acid supplement every day during their second and third trimesters, or a placebo. Blood samples were taken before intervention, after intervention, and from the umbilical cord upon delivery (Figure 1.1). DNA from this study, called FASSTT (Folic Acid Supplementation during the Second and Third Trimester), was the focus of this project, aiming to examine the impact of folic acid supplementation on DNA methylation. In order to assess this, a literature review examining and comparing DNA methylation methods of analysis was written and published in 2011[5]. Following this, an analysis of post-translational modifications affecting dihydrofolate reductase – an enzyme involved in producing a functional form of folate from folic acid – was examined *in vitro* using protein recombination technology to provide insight into the regulation of this enzyme.

1.2: Folate Metabolism

1.2.1: Folic Acid Supplementation and Neural Tube Defects

Folic acid is the synthetic form of folate used as dietary supplements and in the fortification of food. While biologically inactive itself, folic acid is processed by the cell into its active form, tetrahydrofolate (THF), by the enzyme dihydrofolate reductase (DHFR). From here, tetrahydrofolate and its derivatives play a key role in one-carbon metabolism, acting as co-enzymes for many of the reactions involved (Figure 1.2)[6,7]. As an essential B_9 vitamin, folate and its derivatives are essential components of a healthy diet; the recommended daily allowance is 400μg for adults and 600μg for pregnant women[8].

Neural tube defects (NTDs) are complex multifactorial disorders caused by the failure of the neural tube to close during development. This process, known as neurulation, occurs between 21 and 28 days after conception, usually before the mother knows that she is pregnant[9]. There are two main types of NTDs – cranial and caudal defects – each arising from the position along the neural tube the malformation occurs as depicted in Figure 1.3[10]. Anencephaly in an

invariably lethal NTD, characterised by severe cerebral and cranial defects[11]. On the other hand, spina bifida is a caudal defect that makes up about two-thirds of NTD cases, and causes paraplegia, though is not usually fatal on its own[10,11]. In 2001, the occurrence of NTDs in Ireland was recorded at 0.8 to 1.5 per 1,000 births[12].

The link between micronutrient availability and neural tube defects was first suggested in 1972, when Smithells *et al.*, examined serum folate, red cell folate, white blood cell vitamin C, and riboflavin levels in 900 cases of pregnancy[13]. Of these pregnancies, 6 resulted in neural tube defects. Despite the small numbers, red cell folate levels were found to be significantly lower in these individuals when compared to the rest of the sample set. Following this finding, the same research group examined this link further, carrying out a nonrandomised control trial on 243 women with high-risk pregnancies (i.e. where NTDs had occurred before). An 86% lower risk was reported as a result of this intervention[14]. However, no public health action was taken following this, due to the lack of a randomised element to the study[10].

In 1992, a study published in the *Lancet* by the British Medical Research Council found that folic acid supplementation alone has a strong protective effect against the occurrence of NTDs[15]. These results have now been replicated by a large number of studies carried out all over the world[16–19].

However, the mechanism of action and the metabolic basis from which folic acid supplementation prevents these birth defects is not well understood, and there have been questions raised over the possible adverse effects of mandatory fortification[20–23]. In 2007, data from the Aspirin and Folic Acid Polyp Prevention Study (AFPPS) suggested that treatment with folic acid leads to an increased risk of colorectal adenomas and prostate cancer, adding to the growing concern[24]. More recently, a meta-analysis reported in The Lancet demonstrated that, in aggregate, these epidemiological trials provide no significant evidence for the effects of folic acid supplementation on cancer incidence[25].

There will always be controversy in the field as long as the molecular mechanisms and evidence for the benefits of folic acid supplementation remain elusive. For this reason, it is of paramount importance that we learn as much as we can about how folic acid supplementation works to cause or prevent disease.

1.2.2: The FASSTT Study

Today, clear recommendations exist regarding supplementation during pregnancy. Women planning on having children are advised to take 600µg of folic acid a day until the end of the first trimester. However, there are no guidelines for folic acid supplementation during the latter

stages of pregnancy, although a decline in serum folate levels have been observed during the second and third trimester in multiple studies[26,27].

Recently, a randomised control trial carried out in the University of Ulster, Coleraine sought to examine the effect of continued folic acid supplementation past the first trimester[4]. The study, named FASSTT (Folic Acid Supplementation during the Second and Third Trimester), recruited 119 pregnant women attending an antenatal clinic in Northern Ireland. They were randomly assigned into two groups, taking either 400µg folic acid daily, or a placebo (Figure 1.1). All participants had been taking a dose of 400µg/d folic acid prior to recruitment, and all had singleton uncomplicated pregnancies. Participants were excluded if they were aged over 35, or suffered from gastrointestinal, hepatic, renal, or vascular disease. Women who previously had pregnancies with NTDs, had first-degree relatives with NTDs, or suffered from NTDs themselves were also excluded from the trial. Non-fasting blood samples were taken at recruitment, (at approximately week 14 of pregnancy), and at week 36. Food frequency questionnaires were taken during the second trimester, focusing on B-vitamin intake. From blood samples, plasma homocysteine, serum folate, red cell folate, and serum vitamin B12 levels were measured using microbial and immunological assays. Genotypes of the *MTHFR* 677 C>T polymorphism, known to affect folate metabolism, were also measured using PCR coupled with HinF1 restriction digestion[4]. Clinical and biological data from the FASSTT is presented in Table 1.1.

The FASSTT study found that continued folic acid supplementation can prevent the typical decline in folate status typically observed during the latter stages of pregnancy[4]. Using DNA extracted from the buffy coats of these samples, we aim to examine the effect of intervention on DNA methylation.

1.2.3: Folate-mediated One-Carbon Metabolism

Folate metabolism is compartmentalised between the mitochondria, nucleus, and cytoplasm of mammalian cells. Cytoplasmic one-carbon metabolism can be broadly divided into two sections: the DNA cycle and the methylation cycle[7].

This metabolic network is outlined in Figure 1.2. First, folic acid and dietary folate are reduced by DHFR to their biologically active form, tetrahydrofolate. The one-carbon unit is introduced to THF by serine and the enzyme serine hydroxymethyltransferase (SHMT), producing glycine and 5,10-methylenetetrahydrofolate, the form of folate required for *de novo* thymidylate synthesis. In rapidly growing cells, the trifunctional enzyme methylenetetrahydrofolate dehydrogenase (MTHFD1) may also utilise 5,10-methylenetetrahydrofolate for purine biosynthesis. It is the significance of the DNA cycle to

cell survival that makes methotrexate – an anti-folate drug that competitively inhibits DHFR – such a potent chemotherapeutic agent[28].

The methylation cycle is comprised of four reactions. One product of this cycle is S-adenosylmethionine (SAM), the methyl donor for cellular methylation reactions including DNA methylation[29–31]. The DNA and methylation cycles are linked together by methylenetetrahydrofolate reductase (MTHFR), which reduces 5,10-methylenetetrahydrofolate to 5-methylenetetrahydrofolate. This reaction is vital for feeding methyl groups into the methylation cycle. The *MTHFR* polymorphism 677C>T ($Ala^{222}Val$) reduces MTHFR enzyme activity, possibly steering one-carbon units away from the methylation cycle and compromising methylation reactions – including DNA methylation – by reducing SAM availability[32–34]. Elevated risk of NTDs have been associated with the homozygous TT variant of the polymorphism[35].

Variations of *DHFR*, the gene encoding dihydrofolate reductase, have been associated with disease risk and the efficacy of methotrexate treatment in patients with cancer or arthritis[36]. Tumour cells have been found to resist this chemotherapeutic effect either by upregulating DHFR transcription[37], or through mutation of the *DHFR* gene itself[28,38].

Due to its role in health and disease and its vital function in one-carbon metabolism, *DHFR* has been the subject of extensive investigation since its discovery in 1983[39]. Recently, our research group found that *DHFRL1*, once annotated as a non-functional pseudogene of *DHFR*, is expressed and functional, capable of reducing tetrahydrofolate in *DHFR*-null mammalian cells[40]. With a lower affinity for DHF than DHFR, DHFRL1 was also found to localise to the mitochondria, possibly facilitating on-site synthesis of mitochondrial DNA. McEntee *et al*., (2011)[40] also hypothesised that the second reductase enzyme may have a differential response to methotrexate treatment, possibly providing an insight to the mechanisms of resistance, or new potential therapeutic treatment strategies.

As aforementioned, one-carbon metabolism is compartmentalised between the nucleus, the mitochondria, and the cytoplasm. Whereas McEntee and colleagues found evidence for mitochondrial DHFRL1, DHFR has been well established to localise to the nucleus during the S phase of the cell cycle[41,42]. Anderson *et al.*, (2007)[43] have found evidence that this nuclear translocation is facilitated by SUMOylation of DHFR, a post-translational modification associated with localisation to the nucleus. On the other hand, DHFRL1 does not have a mitochondrial targeting sequence in its primary structure, although it does have a SUMOylation site. Indeed, all enzymes involved in nuclear one-carbon metabolism have consensus sequences for SUMOylation, and have been demonstrated to be substrates for *in*

vitro SUMOylation reactions[43,44]. No evidence for nuclear DHFRL1 has been uncovered as of yet, but this possibility cannot yet be ruled out[40].

As demonstrated by McEntee *et al*., (2011)[40], research into the enzymes of one-carbon metabolism is at the very frontier of the field, with many more questions remaining to be answered. There is still much to be learned about the compartmentalisation of folate metabolism, and the factors that influence localisation of its various components.

1.3: Epigenetics

In the 19th century, embryologists trying to discern the 'developmental plan' for organisms were divided into two groups: those endorsing preformationism – the concept that each cell with preformed elements simply grows larger during development, and those who believed that development was the result of a complex series of chemical reactions. The latter, now known to be most correct, named this theory 'epigenesis[45].'

The term 'epigenetics' was coined by CH Waddington in 1942 by combining the words 'genetics' and 'epigenesis[46,47].' Originally used to simply describe the role of genetics during development, Waddingtonian epigenetics is quite different from the epigenetics we know today. However, by examining the scientific advances made following Waddington's work, the link between these two schools of thought becomes clear.

As the understanding of embryology improved into the 20th century, it became clear that development adheres to the model of epigenesis. One piece of evidence that seemed to support this was the fact that the complement of DNA in a fertilised egg is identical to that of every somatic cell arising from it. Watson and Crick's model of semi-conservative replication provides the mechanism of how this level of genetic homology can arise during development[48,49]. However, this observation raised another challenge in the field: How could identical genetic material across cells produce such a wide diversity of cell types and tissues?[45] Stedman and Stedman provided the first explanation to this phenomena, understanding that the proteins bound to DNA in the nucleus could act for repressors for gene expression[50]. At the time, it was assumed that different cell types must have a different set of DNA binding proteins. It wasn't until 1964 when Allfrey and Mirsky began describing how post-translational modification of these proteins can impact on gene expression that epigenetics took the form we know of today[51].

The archaic form of term epigenetics was coined to further understand the role of genetics during development, aiming to answer one of the oldest questions in biology: "Does the

embryo contain all its parts in little from the beginning… or is there a true formation of new structures as it develops?" (Aristotle, 350BC)[52,53]. By examining the evidence supporting this hypothesis, scientists throughout the 20th century discovered that DNA can be altered without affecting the sequence itself, ushering in the new age of epigenetics.

The core histones within a eukaryotic nucleus are capable of acquiring more than 100 different post-translational modifications: including acetylation, ubiquitination, acetylation, SUMOylation, and phosphorylation[54]. By altering how histones interact with DNA, these modifications make up a complex mosaic of changes commonly referred to as the "histone code[55]." For example, methylation of histone lysines H3K4 is associated with transcriptional activation, while methylation of histone lysines H3K9 and H3K27 are both associated with transcriptional repression. In these cases, histone methylation acts to recruit activating enzyme complexes to chromatin[56,57]. On the other hand, acetylation of lysine residues impacts on gene expression by neutralising the ionic interaction between DNA and histones[54]. Histone deacetylation, a separate enzymatic process, complements this interaction, alluding to a complex system of biochemical switches that alter gene expression while leaving the DNA sequence intact[58].

Long non-coding RNAs (lncRNAs) are also considered to be an element of epigenetic regulation[59]. Once believed to be non-functional, now it is understood that lncRNAs regulate transcription by base-pairing with mRNAs in the cytoplasm. For example, a mouse lncRNA complementary to the gene *Uchl1* can control UCHL1 synthesis at a post-transcriptional level[60]. Although a very promising area of research, the sheer abundance of lncRNAs is proving to be a significant challenge to overcome[59]. However, efforts are underway to categorise the structure-function dynamics of lncRNAs that regulate protein expression – including the identification of protein and RNA binding domains[61]. With such a wide diversity of individual transcripts yet to discerned, uncovering the "global landscape of elements in lcnRNAs" may bring the field closer to unravelling this complex epigenetic system[59].

There is a third type of epigenetic mechanism more extensively studied than those mentioned here: DNA methylation[62]. However, despite the huge volume of research being carried out to investigate the nature of this modification, many of its mechanisms and characteristics are still unknown. Nonetheless, with a wide range of health and biological implications, the study of DNA methylation is an incredibly promising field[63].

1.4: DNA Methylation

DNA methylation is a biochemical process where a DNA base, usually cytosine, is enzymatically methylated at the 5-carbon position[64]. An epigenetic modification associated with gene regulation, DNA methylation plays a key role in biological health and disease, acting through processes such as genomic imprinting, X-inactivation, and tissue specific gene expression with possible trans-generational effects[65–67]. Changes in DNA methylation patterns have also been implicated in the pathogenesis of human diseases such as cancer[68]. Although DNA methylation also occurs in prokaryotic cells, the purpose for which is fundamentally different to eukaryotic cells: bacteria methylate their own DNA in order to differentiate it from that of invading bacteriophages. Utilized as a restriction/modification mechanism, DNA methylation allows bacterial cells to protect their own DNA from the restriction enzymes used to digest incoming phage DNA[69].

Relatively little is known about the regulatory systems in place that influence DNA methylation patterns *in vivo*. The raw materials for all DNA methyltransferase reactions are derived from the "methylation cycle" of one-carbon metabolism in the form of S-adenosylmethionine. The DNA methyltransferase family of enzymes catalyse the transfer of methyl moieties from S-adenosylmethionine, and are broadly categorised into two groups; maintenance and *de novo* methyltransferases. The maintenance methyltransferase, DNMT1, has a substrate preference for hemi-methylated DNA; methylated DNA that has undergone semi-conservative replication with one parent retaining its methylation mark[70,71]. The enzymes DNMT3a and DNMT3b are capable of methylating unmethylated DNA, and constitute *de novo* methylation[70]. DNMT2 was once believed to not have any DNA methyltransferase activity, localising to the cytoplasm in transfected mouse fibroblasts[72]. Now, however, evidence has emerged that this elusive enzyme targets RNA molecules for methylation[73].

It was previously believed that DNA demethylation only occurs passively; DNMT1 would switch off during DNA replication, eventually decreasing the number of hemi-methylated nascent strands after subsequent cycles[74]. Although this has been demonstrated to be the case in embryonic stems cells during development, evidence is emerging for an active DNA demethylation mechanism[75,76]. The enzyme TET1 has been demonstrated to oxidate 5-methylcytosine to 5-hydroxymethylcytosine[77]. Although there were some proposals that 5-hydroxymethylcytosine acts as a unique epigenetic mark involved in transcriptional regulation[78], there is more coherent data in the literature to suggest that 5-hydroxymethylcytosine acts as an intermediate between 5-methylcytosine to unmethylated cytosine. Two possible enzymatic processes have been proposed to explain this, either through

further oxidation to 5-formylcytosine and 5-carboxycytosine, or by deamination of the 5-hydroxymethylcytosine molecule to 5-hydroxymethyluracil. In both the oxidation and the deamination models, possible T:G mismatches are introduced which, when corrected by base-excision repair machinery, produce unmethylated cytosines (Figure 1.4)[79–82].

However, the role of 5-hydroxycytosine may not be limited to passive demethylation. Indeed, 5-methylcytosine is thought to have originated as an ancestral, protective agent in mammalian cells against invading retroviral elements, based on the high levels of methylation found in repeated and transposable regions of chromosome centromeres[83]. Although the DNMT enzymes are highly conserved across various organisms, other complex processes associated with DNA methylation have evolved beyond its ancient role, and are not common across all species[84]. For example, imprinted genes in mammals and plants and have been found to be methylated monoallelically based on their parent of origin – something now known to have emerged from convergent evolution[85]. Thus, it is possible that hydroxymethylation could have evolved beyond its role in active demethylation over the course of evolutionary history. For example, 5'hydroxymethylation is also known to inhibit methylation via DNMT1, possibly facilitating passive demethylation too[86].

In a recently published review, Hahn *et al.,* (2014)[87] describe how 5' hydroxymethylation has been demonstrated to be stable in various cell and tissue types, alluding to a function beyond demethylation. Even more recently, Bachman *et al.,* (2014)[88] found 5'hydroxycytosine to be stable in mouse tissues *in vivo* using isotopic labelling.

These studies show that the paradigm dictating the role of 5'hydroxymethylation is constantly changing as new data emerges. There is much more to learn about the process of DNA methylation and demethylation. This thesis aims to add to the collective knowledge of the influences on these processes by examining the impact of maternal nutrition on health and disease during pregnancy.

1.5: The Foetal Origin of Human Disease

In the winter of 1944-1945, a devastating famine in The Netherlands caused by a German imposed food embargo claimed the lives of over 200,000 people. Due to the fact that this was a manmade famine enforced upon a developed country for a short period of time with excellent medical records kept throughout, the 'Dutch Hunger Winter' has gained considerable attention from the scientific community. Many studies have been carried out on birth cohorts recorded during the famine in order to learn about the effects of malnutrition on the *in utero*

environment[89–92]. In 1976, it was found that children exposed to the famine during early gestation were more susceptible to obesity later in life, compared those exposed to the famine during the latter stages of development[93]. In another study, it was found that although reduced glucose tolerance was associated with exposure to the famine during any point during gestation, those exposed to the famine during late gestation were more susceptible to coronary heart disease, atherogenic plasma lipid profiles, and increased risk of breast cancer[94].

The Dutch Hunger Winter provided data that adhered to the now well established 'Barker Hypothesis', which holds that various adult diseases originate from the *in utero* environment from which the foetus developed[1]. It has been hypothesised that these disease states arise due to the predicted post-natal environment being radically different than the pre-natal environment[94]. For example, epidemiological associations between poor foetal growth and development of type 2 diabetes later in life indicate that a developing child is capable of adapting a 'thrifty phenotype' in order to prepare for a possibly adverse environment upon birth. It is when a baby is born into an environment that does not reflect it gestational conditions – such as poor foetal nutrition during a short famine – that susceptibility to diseases like diabetes increases[95,96].

There seems to be an obvious, intricate, and complicated relationship between the *in utero* environment and adult health and disease. A rather elaborate and elusive molecular system, like the plasticity of the epigenome, may hold the biological explanation to what has been observed in these studies[97].

During the foetal development in mice, the parental genomes both undergo significant epigenetic reprogramming. At the very early stages, before blastocyst formation, genome-wide DNA demethylation takes place[74]. After implantation, genome-wide levels of methylation increase, along with activity of the *de novo* methyltransferases, DNMT3a and DNMT3b[98]. Very little is known about the mechanisms that control this reprogramming, but researchers are beginning to speculate that Barker's 'thrifty phenotype' may in fact be a 'thrifty epigenotype'[99,100].

1.6: DNA methylation and Nutrition

Although it has been known for a long time that diseases like asthma, neurological disorders and some forms of cancer originate through an alteration in the DNA methylation status, evidence is now mounting that these changes occur through aberrant environmental conditions[101–103]. The sites that are subject to change are called differentially methylated regions (DMRs), and the variations themselves are called metastable epialleles[104,105]. A variety

of environmental influences have been observed that can bring about changes to DMRs, including diet, malnutrition, and exposure to chemicals such as tobacco smoke, alcohol, metals, and other sources of oxidative stress[106–108].

As a precursor to the methyl-donor S-adenosylmethionine, interest is growing around folate's role in DNA methylation. In particular, a study on mice has yielded proof that genetically identical animals can have drastically different phenotypes due to observable changes in a particular DMR brought on by the diet of the mother[109]. In this case, a pregnant mouse with a diet supplemented with methyl donating vitamins, such as folic acid itself, exhibited a change in the DMR of the *agouti* gene, generating metastable epialleles with drastically differing phenotypes. When the retroelement Intracisternal A Particle (IAP) (a type of transposable element or 'genetic parasite' capable of reverse-transcribing to RNA and changing position in the genome[110]) upstream of the *agouti* gene is methylated, expression of the agouti signalling protein increases and produces the phenotype of yellow fur, obesity, Type II diabetes, and a predisposition to tumours. When IAP is silenced through DNA methylation, the *agouti* gene is expressed in its healthy, wild type form. Due to the variable levels of activity from IAP, a spectrum of colours from yellow to brown was observed in correlation with levels of upstream methylation. These changes persisted to the following generation, even without exposure to the same *in utero* environment (Figure 1.5)[110].

In humans, the effects of nutrition on particular DMRs *in utero* have been observed from the Dutch Hunger Winter. In a study tracing those that had been exposed to the famine prenatally, Heijmans *et al.* (2008) found a significant difference in the DNA methylation statuses of *IGF2* between same sex siblings; those prenatally exposed to the famine showed a significant decrease in DNA methylation at the imprinted locus compared to those who did not[97]. At the same time, other imprinted loci such as *IL10*, *LEP*, *ABCA1* and *MEG3* showed increased methylation in the same exposed individuals[107].

With respect to folate status, several studies have shown that folic acid supplementation has an effect on the methylation status of particular genes. For example, the *IGF2* DMR of children whose mothers consumed 400μg of folic acid periconceptionally was shown to be highly methylated in comparison to those who had not[108]. In another study by Fryer *et al.* (2011)[111], a genome-wide analysis correlating DNA methylation levels with homocysteine levels (a product of folate metabolism) found many more DMRs that can be affected by folate status[111]. This study was carried out on cord blood taken at the end of term, demonstrating that the effect of folate status on DNA methylation continues throughout the entire pregnancy.

More recently, Dominguez-Salas *et al.*, (2014) demonstrated that variations in folate status, caused by seasonal influences, impacted on the methylation status of six metastable epialleles of children born in rural Gambia[112]. In another recent study, Amarasekera *et al.*, (2014) examined the impact of folate status during the third trimester of pregnancy on neonatal DNA methylation. Seven FS-DMRs emerged from the analysis with one, the promoter region of *ZFP57*, significantly hypomethylated in response to high folate levels. With *ZFP57* expression levels found to increase as a result of hypomethylation, and its product known to be a regulator of imprinting-associated DNA methylation during development, Amarasekera and colleagues tentatively speculate that a compensatory mechanism may exist to regulate ZFP57 production in response to folate levels[113].

The studies listed above suggest that similar changes to DNA methylation status of these DMRs may also be observed in the FASSTT samples; the Steegers-Theunissen paper describes a study with individuals taking the same amount of folic acid, 400μg/day. The imprinted loci of *IGF2* was found to be methylated in the lymphocytes of test subjects[108]. At the same time, data from the Fryer *et al*. (2011) study on homocysteine suggest that, although the beneficial effects of folic acid supplementation with respect to the prevention of neural tube defects is limited to the first few weeks of gestation, the relationship between folic acid supplementation and DNA methylation changes may be relevant throughout the second and third trimester[111].

In a review published in 2011, Parle-McDermott & Ozaki examined studies in the literature that identified nutrient sensitive DNA methylation sites in humans[3]. There is a wealth of information in the literature to suggest that factors affecting one-carbon metabolism can have a significant impact on DNA methylation throughout the genome. By examining DNA obtained from the FASSTT study, we aim to add to the growing list of FS-DMRs already established in this incredibly fast-moving field.

1.7: Methods in DNA methylation analysis

1.7.1: Genome-wide DNA methylation analysis

Upon the completion of the Human Genome Project in 2001, it was clear that the full 'Book of Man' had yet to be elucidated. The inexplicable complexity of human biology could not be explained by genetics alone, as with only 30,000 protein-coding genes we possess just 33% more than the nematode, *Caenorhabditis elegans*, and approximately 6,000 less than the humble banana plant, *Musa acuminate*[114–117]. However, mounting evidence from the past few decades is pointing to a new set of variables that contribute to our individuality. The Human

Genome Project has already unveiled the genetic hardware needed to create a person, but the search for the biochemical software is still underway[5].

Methylated cytosine itself is very difficult to measure without further modification. Early genome-wide methylation analysis based on Reversed-Phase High Performance Liquid Chromatography relied on the separation of methylated cytosine from unmethylated cytosine, but was limited to only displaying the total level of methylation in a given sample, independent of the associated loci[118]. Methods of this nature elucidate very little about the epigenome on a practical scale, especially when aiming to eventually link specific DNA methylation alterations to changes in gene expression.

Recent advances in Next Generation Sequencing (NGS) technology have yielded a variety of sophisticated and high-throughput, high-resolution methods for whole genome-wide epigenetic analysis[119]. At the moment, the gold-standard profiling method with the highest resolution is bisulfite sequencing coupled with NGS. Although this is a highly expensive approach[120], cost is becoming less of an issue in recent years[5]. For example, Illumina unveiled its Infinium HumanMethylation450 BeadChip: a microarray capable of processing 480,000 of the 28 million CpG dinucleotides across the human genome without the use of NGS technology[121,122]. As comprehensive as this seems, the use of a DNA microarray to obtain data immediately introduces bias; only regions of the genome represented by the array itself can be probed for changes in DNA methylation.

As part of a review in *Frontiers in Genetics*[5], the available methods for DNA methylation analysis were examined. Table 1.2 outlines the features and differences between a selection of genome-wide methods. Of those that are based on NGS, whole bisulfite sequencing as described by Cokus *et al*., (2008) is the most comprehensive[123]. MeDIP-seq (Methylated DNA Immunoprecipitation Sequencing) and anti-5mC MBDiGS (Methyl-Binding Domain isolated Genome Sequencing) follow similar premises, with shearing of the whole genome followed by immunoprecipitation of methylated DNA fragments for the former, and immunoprecipitation of methyl-binding proteins for the latter[124,125]. Unlike sodium bisulfite sequencing, however, these methods do not produce data at a single-base pair resolution, as NGS is utilised to identify the tags enriched for methylation rather than the individually methylated cytosines themselves.

1.7.2: Modified Methylation Specific Digital Karyotyping

It has been mentioned in Section 1.4 that DNA methylation exists in bacteria as a restriction modification system to differentiate between the organism's own DNA and that of invading

phage[69]. In the context of mammalian DNA methylation, this characteristic is still of interest to researchers from a technical standpoint. Of the restriction enzymes used by bacteria to fend off phage DNA, there are some that can differentiate between methylated and unmethylated DNA. For example, the enzyme HpaII will only digest DNA at the site CCGG when unmethylated, but not when methylated. Exploiting the redundancy of the many palindromic motifs of restriction sites throughout the genome, Cedar *et al.*, (1979) used a pair of 'isoschizomers' to map the methylation patterns of calf thymus DNA[126]. Since both MspI and HpaII cut at the same site, and the latter will only cut unmethylated DNA, digesting DNA with both in tandem can be used to discriminate between methylated and unmethylated loci[127].

Since the publication of this MspI/HpaII assay in 1990, other groups began to experiment with similar enzymes to obtain DNA methylation data. Restriction landmark genomic scanning (RLGS) is a genomic scanning method that takes advantage of the specificity of restriction endonucleases and allows a low resolution comparison of genome-wide differences between individuals[128]. Radiolabelled DNA is digested with two restriction enzymes and separated in two dimensions. This produces a profile of thousands of spots spread through the gel, each representing a restriction site. This method was adapted for DNA methylation analysis (RLGS-M) by employing methylation-sensitive restriction enzymes to differentiate methylation differences between individuals[129,130]. Later, simpler and less expensive genome-wide screening strategies came into practice. Using methylation-sensitive and insensitive restriction enzymes with two low-stringency annealing steps, Liang *et al.*, (2002) found that methylation profiles could be obtained by digesting DNA with methylation-specific endonucleases followed by a PCR reaction with random primers[131]. This process is known as arbitrarily primed PCR (AP-PCR), and is based on a method developed by Welsh and McClelland in 1990, initially used to identify bacterial species[132]. AP-PCR was adapted in order to scour tumour genomes for new differential methylation sites, comparing the difference in product sizes and identifying the sites by cutting, cloning and sequencing the bands[104,131].

Today, there are many techniques in the literature that apply the same principles of the Singer-Sam paper to Next-Generation Sequencing[124,133,134]. One is Methyl-sensitive restriction enzyme sequencing (MRE-seq), where genomic DNA samples are digested with restriction enzymes and the subsequent DNA fragments are size selected and sequenced. Differential DNA methylation may be identified by comparison of the fragments. In a similar manner, Li *et al.*, (2009) developed a method called Modified Methylation Specific Digital Karyotyping (MMSDK)[133,134]. Here, methylated DNA is mapped with a methylation specific enzyme, and

oligonucleotide linkers are ligated to either side of the fragments, incorporating Illumina NGS adaptors into their sequences. This method allows high-throughput and low-cost genome-wide DNA methylation mapping with multiple samples.

The MMSDK process, outlined in Figure 1.6, involves the tandem digestion of genomic DNA with a methylation sensitive restriction enzyme (AscI) and a frequent cutting fragmenting enzyme (NlaIII)[133]. The ligation of a biotinylated oligonucleotide linker to the overhangs generated by AscI allows for the separation of fragments flanked by two NlaIII sites from those with one AscI site and one NlaIII site. The latter are retained via streptavidin-bound magnetic beads, and ligated to a second oligonucleotide which contains a restriction site for the enzyme MmeI at its 5' end. The enzyme MmeI recognises this sequence, and cuts 16-18bp into the downstream unknown genomic DNA. A final linker is ligated to the overhang generated by MmeI, resulting in the flanking of an unknown 16-18bp genomic sequence by two Illumina NGS adapters. Since AscI does not cut methylated fragments, the library obtained from these reactions represents areas adjacent to unmethylated AscI sites throughout the genome. By aligning these sequences back to a reference genome, an accurate map of CpG methylation can be obtained.

1.7.3: Methylated DNA Immunoprecipitation

The use of antibodies raised against 5'methylcytosine as an agent to discriminate between methylated and unmethylated DNA was first discussed in 1985[135]. Building on this work, Oakeley *et al.* (1997)[136] devised an efficient method to study global changes in DNA methylation during tobacco pollen maturation with confocal microscopy and secondary antibodies bound to fluorescent isothiocyanate. Later, the dynamic epigenetic reprogramming associated with mammalian embryonic development was uncovered using a similar immunofluorescence-based technique[137]. However it would take several more years until DNA microarray and next generation sequencing technology would advance to a stage where methylated DNA could be immunoprecipitated and scrutinised at a single base-pair resolution[14].

Methylated DNA Immunoprecipitation (MeDIP) is a protocol that requires genomic DNA to be sheared and denatured prior to enrichment with anti-methylcytosine (Figure 1.7)[138]. This can be achieved through sonication or by digestion with a restriction endonuclease.

1.7.4: Gene-specific DNA Methylation analysis

As described previously, there are three main strategies for analysing DNA methylation at a gene-specific and genome-wide context: digestion with methylation-specific restriction

endonucleases, immunoprecipitation with anti-5'methylcytosine, and treatment with sodium bisulfite[5]. The use of methylation-specific restriction endonucleases and anti-5'methylcytosine allow indirect analysis of DNA methylation patterns, though both lack the capacity to generate a significant amount of data at single-base resolution. The third approach, adapted here for high-throughput gene-specific analysis, achieves just this.

It was first found in 1970 that treating DNA with sodium bisulfite deaminates cytosine residues to uracil[129]. Later, it was discovered that the same reaction, when applied to methylated cytosine, takes place at a much slower rate[139]. Frommer *et al*., (1992)[140] exploited this property in a classic paper to demonstrate that sodium bisulfite treatment can be used in conjunction with PCR to change all unmethylated cytosines to thymidine, while all methylated cytosines deaminate to unmethylated cytosines. Although this method was used to establish a genome-sequencing based protocol at the time, sodium bisulfite treatment of DNA has now been applied to conventional genotyping techniques for gene-specific DNA methylation analysis[5].

For example, in Methylation-Sensitive Single Nucleotide Primer Extension (MS-SNuPE)[141], a target sequence is treated with sodium bisulfite and amplified with conventional PCR. The product is separated and isolated via electrophoresis, and two SNuPE primers are annealed to the amplified fragment. These primers are designed to lie immediately upstream of the targeted nucleotide. From here, ^{32}P-dCTP or ^{32}P-TTP are incorporated into an amplified fragment of sodium bisulfite treated DNA, and the particular nucleotide incorporated can be visualised using autoradiographic film or phosphorimage quantitation. In a methylated target cytosine, the nucleotide would not undergo bisulfite conversion, and ^{32}P-dCTP would be incorporated. In the case that the target cytosine is unmethylated, it would be converted to thymidine, and the incorporation of ^{32}P-TTP would be observed at that locus.

Based on a well-known PCR method for resolving single-base restriction fragment length polymorphisms[142], Methylation-Sensitive Single-Strand Conformation Analysis (MS-SSCA) is another a technique used to screen and analyse DNA methylation in a gene-specific manner[143]. Genomic DNA is bisulfite treated and the gene of interest is amplified with PCR. The methylation differences, manifesting as a sequence change within the PCR product, can be discerned by digestion with a restriction enzyme. The digestion patterns of samples are compared to a methylation standard and variations in pattern imply changes in DNA methylation. Methylation differences are characterized using a gel stabbing technique and sequencing[144].

Similarly to MS-SSCA, Methylation Specific PCR utilises PCR primers specific to a methylated, sodium bisulfite template[145]. Although incredibly sensitive, this method leaves no means to identify false positives following amplification – a common occurrence due to the decreased sequence complexity of sodium bisulfite treated DNA. In order to combat these pitfalls, MSP has been adapted for Sensitive Melting after Real Time Methylation Specific PCR (SMART-MSP)[146]. By carrying out a high-resolution melting analysis on the amplified PCR product, false positives obtained from non-specific amplification or incomplete bisulfite conversion can be eliminated based on the temperature of the melted PCR product.

SMART-MSP was selected as our method for gene-specific DNA methylation analysis. The FASSTT samples require a reliable and high-throughput DNA methylation analysis; something that only SMART-MSP can provide with the resources accessible to us. This method can also be adapted to use on the Roche Lightcycler 480™ platform available in our laboratory.

Although SMART-MSP does not yield the specific methylation tags on a particular locus at a single-base pair resolution like MS-SSCA and MSNuPE, it can be scaled easily for multiple samples on 96-well plates once the PCR assay has been optimised, making it a cheaper, efficient alternative to the other methods described here. The current 'gold-standard' method for DNA methylation analysis is Pyrosequencing following sodium bisulfite treatment[5,147], but this platform was not available to us.

1.8: Aims and Objectives

Hypothesis

Folic acid supplementation during pregnancy affects the methylation status of specific regions of the genome known as Folate Sensitive Differentially Methylated Regions (FS-DMRs). In folate-mediated one carbon metabolism, DHFR is controlled by competing post-translational modifications.

Aim

To identify differentially methylated regions of the genome which change their methylation status in response to folic acid supplementation, and to examine the potential of DHFR to acquire post-translational modifications.

Objectives

- To conduct a review of the literature and determine what the most suitable methods of gene-specific and genome-wide DNA methylation analysis are, taking into consideration cost, efficiency, coverage, bias, specificity, and scalability.
- To carry out genome-wide DNA methylation analysis on a subset of FASSTT samples to generate a short-list of candidate FS-DMRs.
- To validate these FS-DMRs in the entire cohort using gene-specific DNA methylation analysis.
- To examine the effect of altered DNA methylation patterns of these DMRs on gene expression in a cell culture model.
- To produce recombinant GST-tagged DHFR and examine its susceptibility for post-translational modifications *in vitro*.

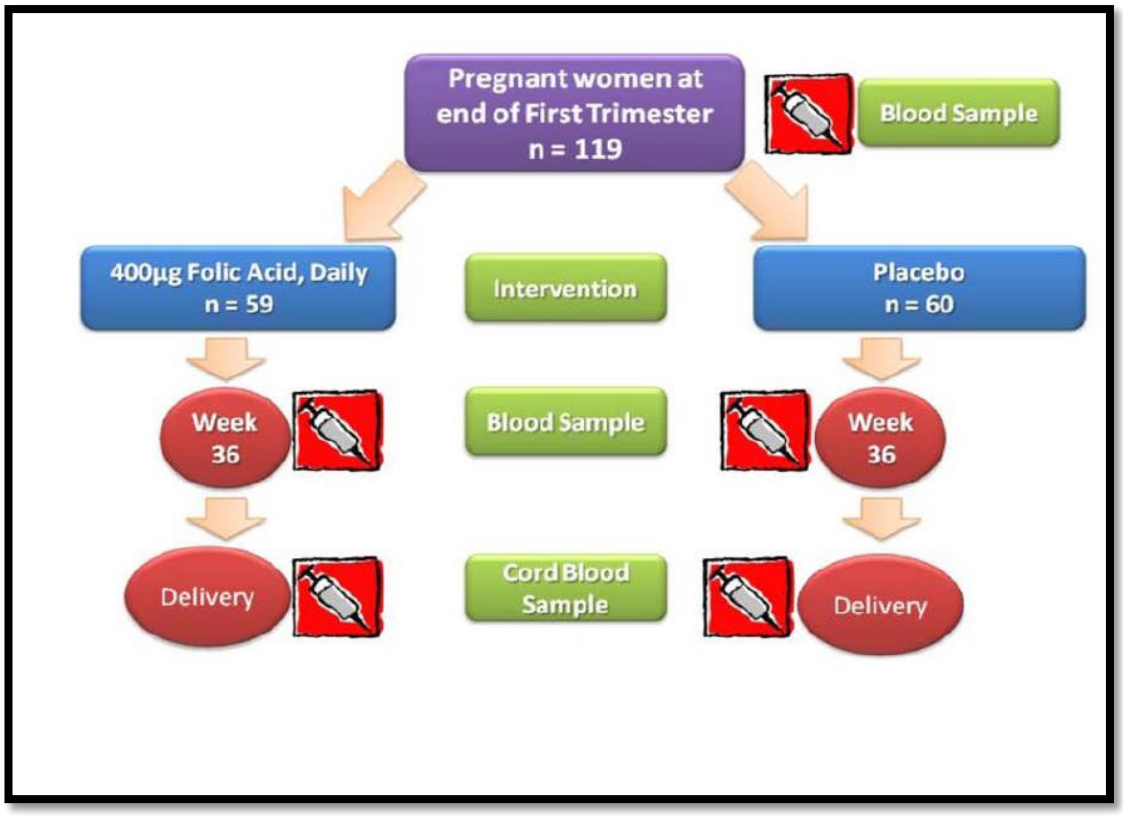


Figure 1.1: FASSTT Study Overview

All women were recruited at the beginning of their second trimester, after the known beneficial effects of folic acid have taken place. Those in the intervention group continued to take 400µg of folic acid daily, while those in the placebo group received a placebo. A total of 3 blood samples from each woman were taken during the FASSTT study. One before intervention, one after 36 weeks of pregnancy, and one more from the umbilical cord upon delivery. Out of 119 total participants, 59 were recruited for the intervention group and 60 for the placebo.

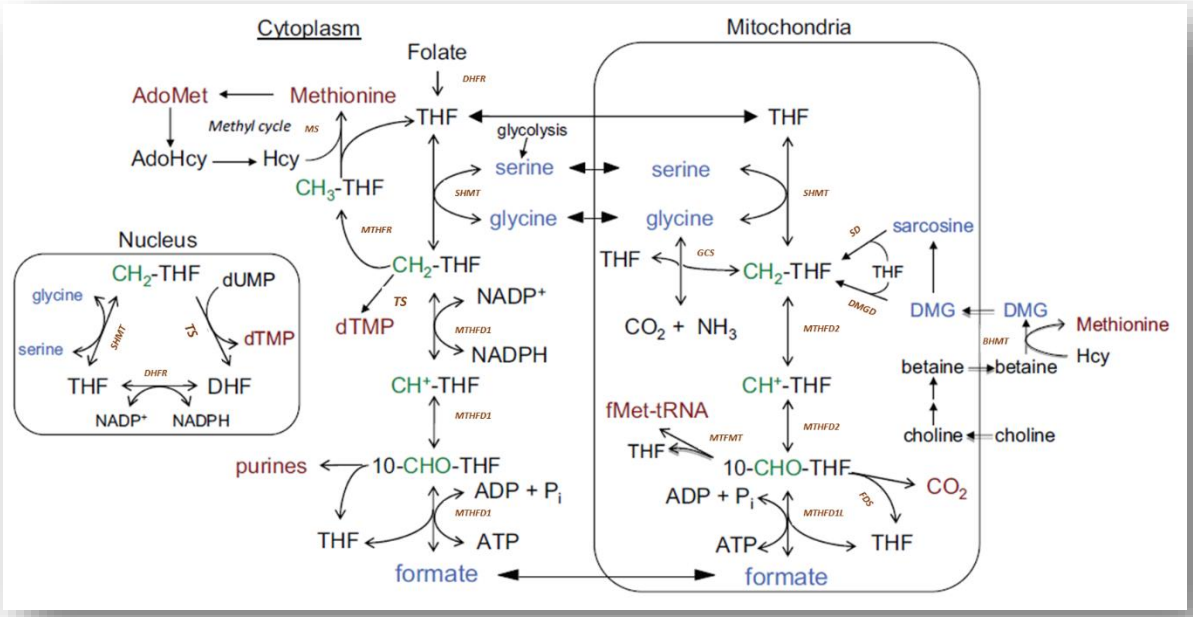


Figure 1.2: Folate mediated one-carbon metabolism

Taken from Tibbets and Appling, (2010)[7]. Folate mediated one carbon metabolism is compartmentalised between the nucleus, cytoplasm, and mitochondria. Dietary folate is reduced by DHFR and incorporated into either the DNA cycle, producing purines and thymidylate (dTMP), or the methylation cycle, producing S-adenosylmethionine (AdoMet). Figure legend: Red, end products; Blue, one-carbon donors; Green, activated one-carbon units; MTHFD1, methylenetetrahydrofolate dehydrogenase 1; MTHFD2, methylenetetrahydrofolate dehydrogenase 2; MTHFDL1, methylenetetrahydrofolate dehydrogenase 1-like; SHMT, serine hydroxymethyltransferase; GSC, glycine cleavage system; MTHFR, 5,10-methylene-THF reductase; MS, methionine synthase; DMGD, dimethylglycine dehydrogenase; SD, sarcosine dehydrogenase; TS, thymidylate synthase; FDS, 10-formyl-THF. MTFMT, methionyl-tRNA formyltransferase; DHFR, dihydrofolate reductase; BHMT, betaine-homocysteine methyltransferase.

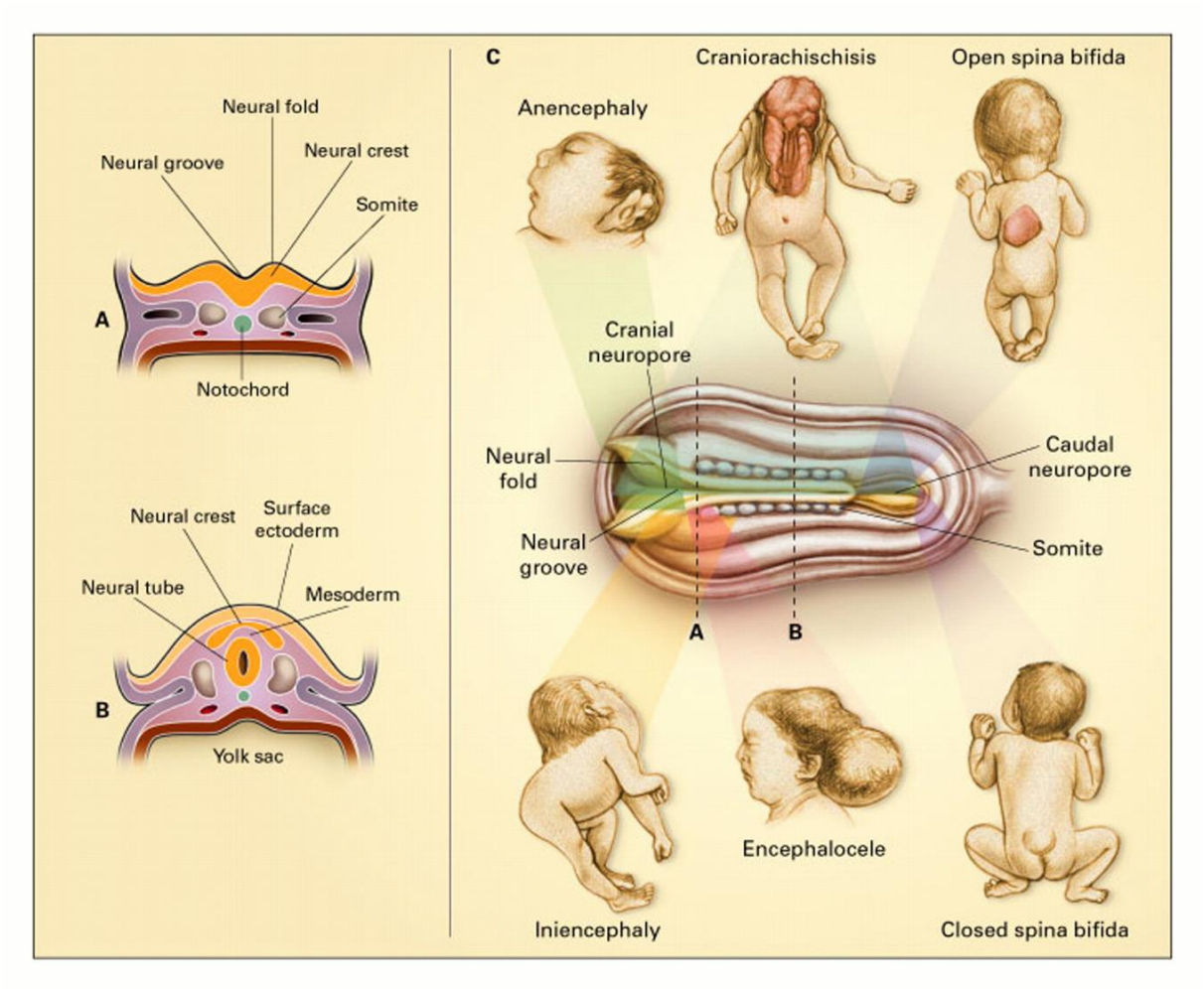


Figure 1.3: Stages of Neural Tube Defect Development

Cross-sections of the rostal end of the embryo before the neural tube has closed (A), and at the centre of the embryo after neurulation is complete (B). Two types of NTDs are depicted relative to where the neural tube fails to close along a dorsal view of the embryo (C). Anencephaly, inencephaly, and encephalocele are cranial defects, while both open and closed spina bifida are caudal defects. Cranioarchischisis is the most severe NTD, characterised by both cranial and caudal defects in neural tube formation. Image taken from Botto *et al.*, (1999)[10].

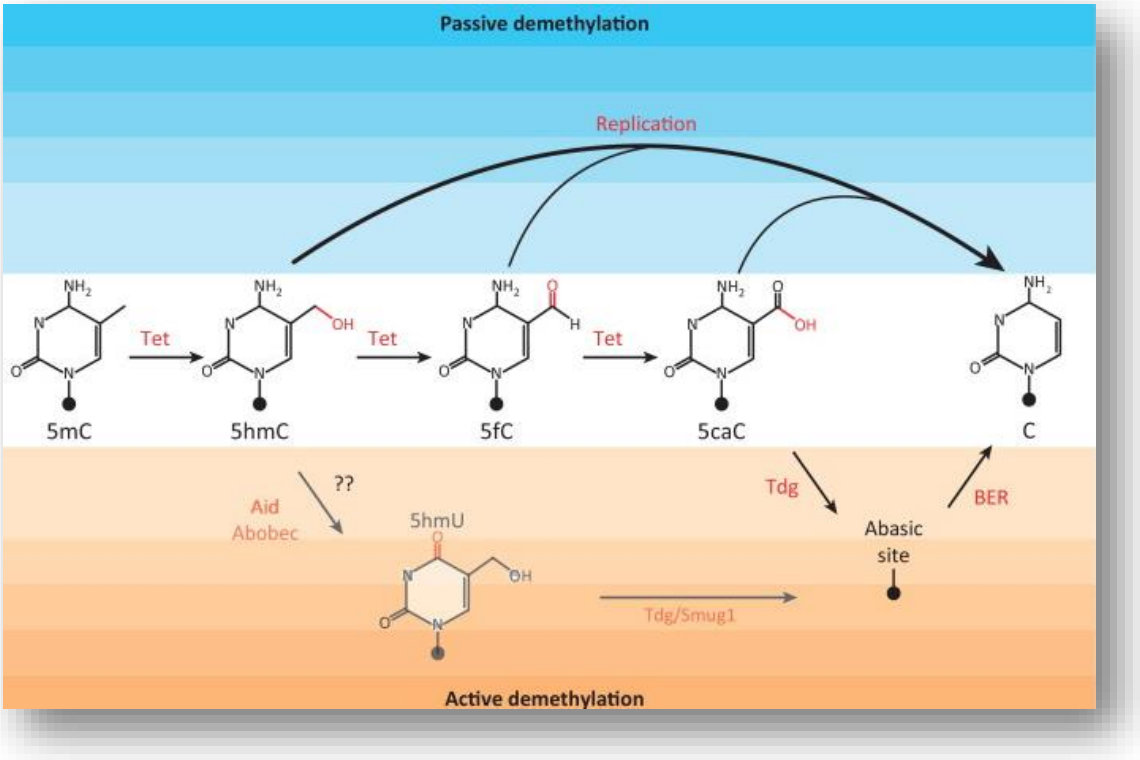


Figure 1.4: Passive and Active DNA Demethylation

The enzyme Tet1 mediates the oxidation of 5'methylcytosine (5mC) to 5'hydroxymethylcytosine (5hmC). With maintenance DNA methyltransferase unable to bind 5hmC, successive replication cycles may dilute this epigenetic mark, leading to the incorporation of unmethylated cytosine (C) in its place (passive demethylation). 5hmC may also be further oxidated to 5'formylcytosine (5fC) and 5'carboxycytosine (5caC). Although passive methylation through replication may occur with these molecules too, they can also be removed by the DNA glycosylase Tdg. The abasic site left by this reaction can be replaced by cytosine through base-excision repair (BER). Another model has proposed that 5hmC may be modified by the deaminase enzymes Aid/Abobec to generate 5'hydroxymethyluracil. This in turn can also be excised by the DNA glycosylases Tgd or Smug1. Both pathways involving BER constitute active demethylation. Image taken from Piccolo *et al.*, 2014[84].

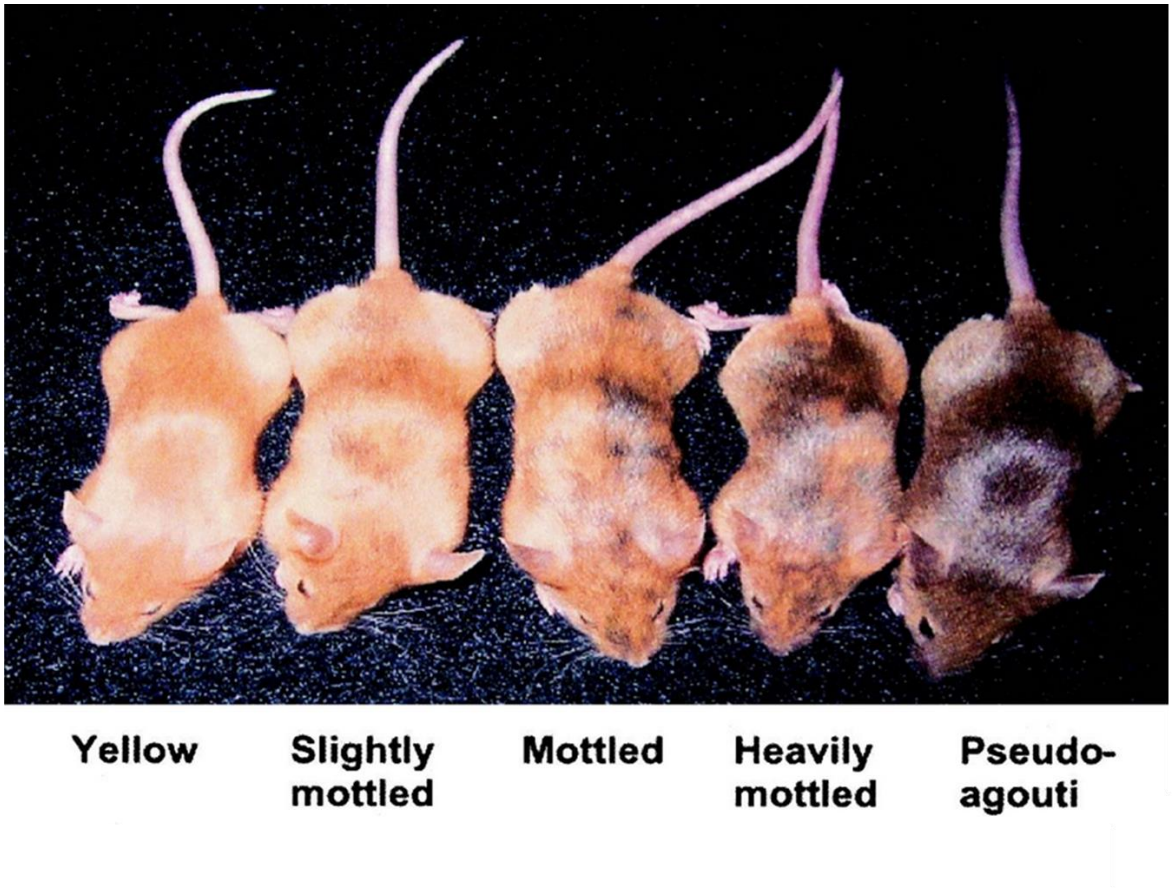


Figure 1.5: Maternal methyl donor supplementation impacts on IAP methylation
Five isogenic mice with varying coat colour as a result of maternal supplementation with methyl donors such as betaine, choline, and folic acid. This dietary intervention affects the methylation status of the Intracisternal A Particle transposable element, which regulates the *agouti* gene. All animals have identical *agouti* genotypes, but markedly different phenotypes. Image taken from Waterland and Jirtle, 2003[109].

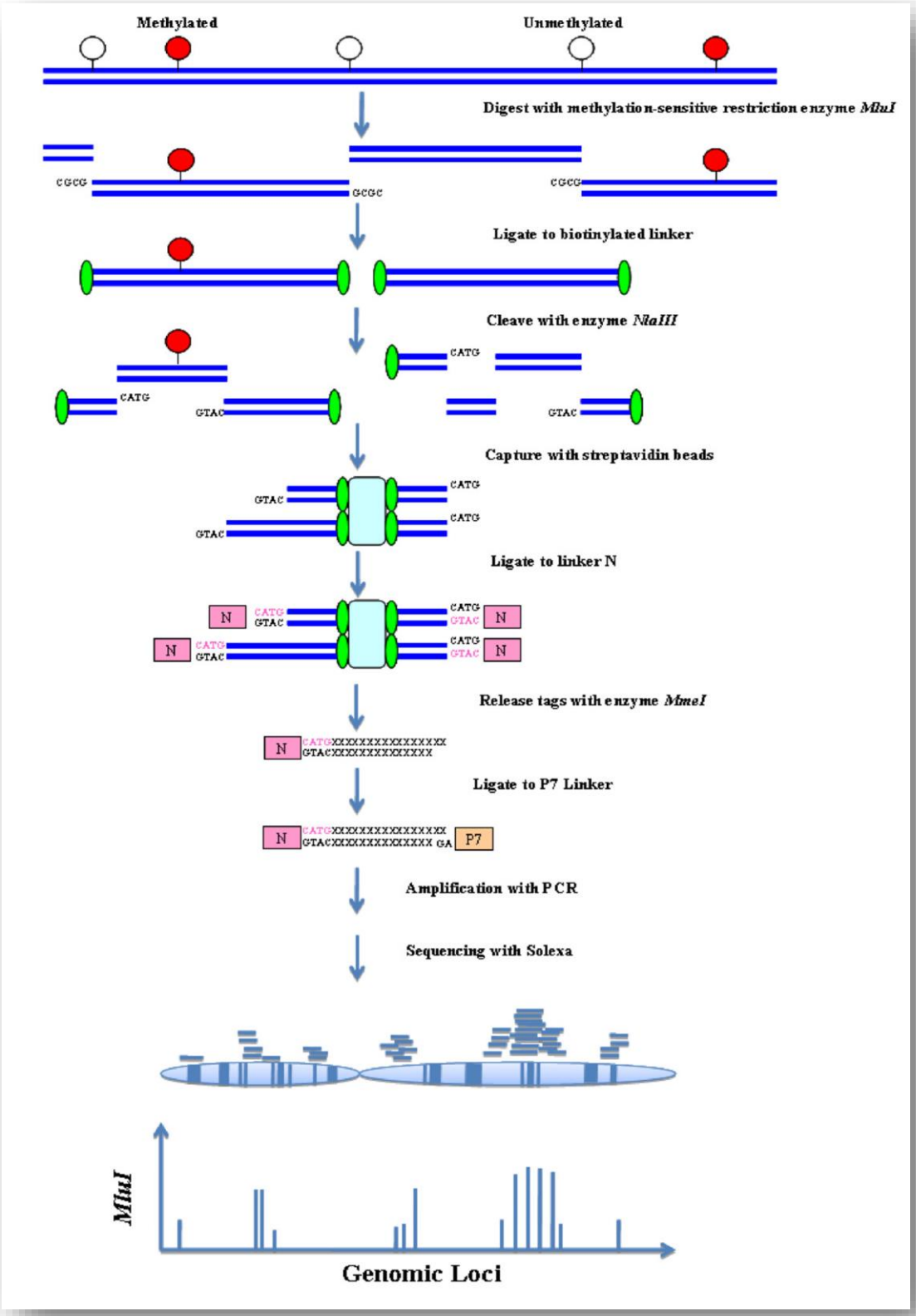

Methylated
Unmethylated
Digest with methylation-sensitive restriction enzyme MluI
CGCG
GCGC
Ligate to biotinylated linker
Cleave with enzyme NlaIII
CATG
GTAC
Capture with streptavidin beads
Ligate to linker N
N
Release tags with enzyme MmeI
Ligate to P7 Linker
P7
Amplification with PCR
Sequencing with Solexa
MluI
Genomic Loci

Figure 1.6: Modified Methylation-Specific Digital Karyotyping (MMSDK)

The MMSDK process uses three tandem enzyme digestion and linker ligation steps to generate a methylation-specific genomic library for next generation sequencing. The ligation of biotinylated oligonucleotide linkers to overhangs generated by AscI digestion allows for the separation of fragments flanked by two NlaIII sites from those with one AscI site and one NlaIII site. The latter are retained via streptavidin-bound magnetic beads, and ligated to a second oligonucleotide which contains a restriction site for the enzyme MmeI at its 5' end. The enzyme MmeI recognises this sequence, and cuts 16-18bp into the downstream unknown genomic DNA. A final linker is ligated to the overhang generated by MmeI, resulting in the flanking of an unknown 16-18bp genomic sequence by two Illumina NGS adapters. Since AscI does not cut methylated fragments, the library obtained from these reactions represents areas adjacent to unmethylated AscI sites throughout the genome. Figure taken from Li *et al*., 2009[133].

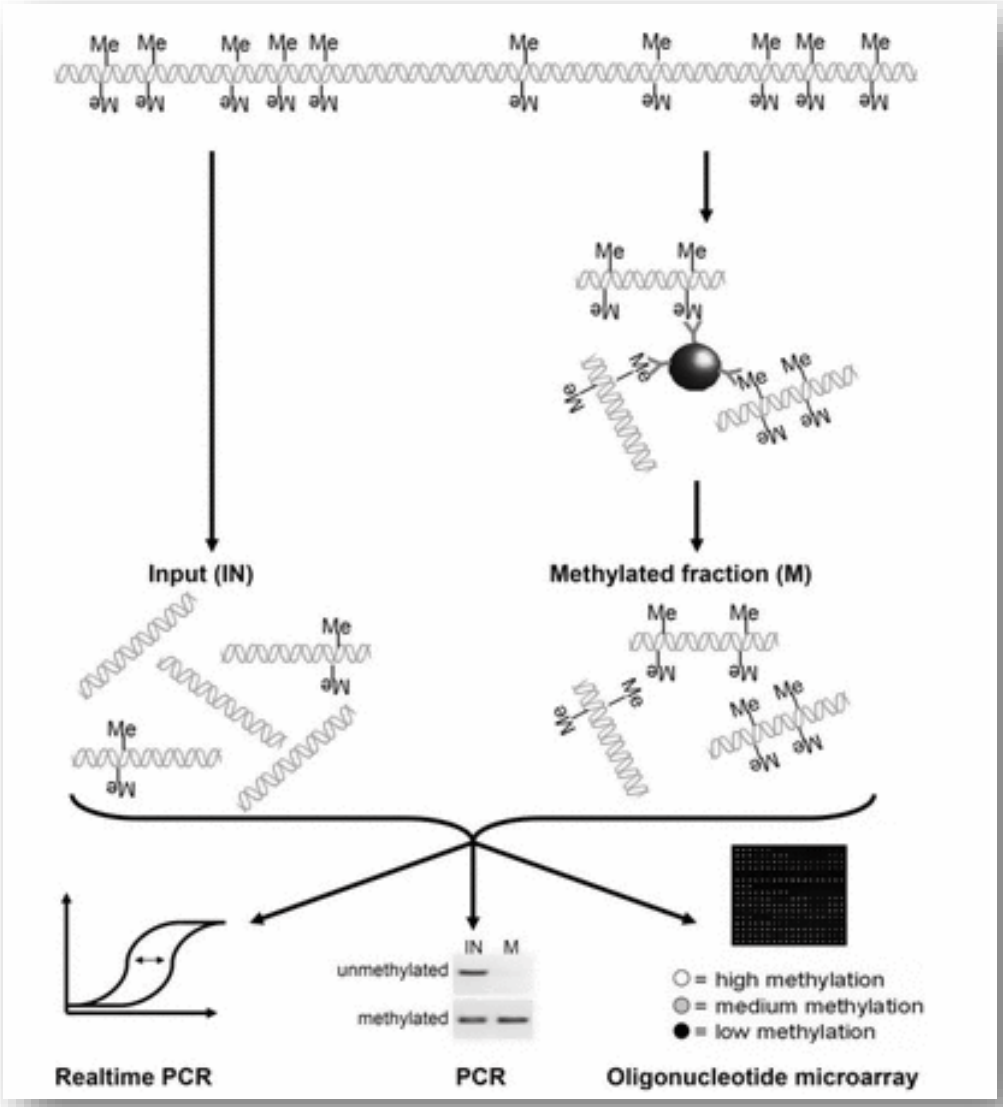


Figure 1.7: Methylated DNA Immunoprecipitation

Taken from Mohn *et al.*, (2009)[138]. Genomic DNA is sheared prior to immunoprecipitation with anti-methylcytosine bound to magnetic beads. An Input fraction is also prepared in the same manner, with immunoprecipitation omitted. Both the Methylated and Input fractions can be compared using qPCR or traditional PCR and agarose electrophoresis. Differentially labelling both fractions with Cy3 and Cy5 and hybridising to an oligonucleotide microarray can create a complete picture of DNA methylation on a genome-wide or promoter-specific scale, depending on the platform used.

Table 1.1: Clinical and Biochemical data from FASSTT cohort, taken from McNulty *et al.*, 2013[4]

	Placebo group (n = 6	Treatment group (n = !	P
General characteristics			
Age (y)	28.0 ± 4.3	29.2 ± 4.0	0.12
Ethnicity (white) (%)	100	100	—
BMI (kg/m)	24.3 ± 3.7	25.0 ± 5.0	0.39
Smokers (%)	18.3	15.3	0.93
Parity (n)	0.92 ± 1.06	0.98 ± 0.92	0.72
Gestation at baseline (wk)	13.8 ± 2.2	14.0 ± 2.2	0.59
Duration of folic acid supplement usage (wk)	13.6 ± 9.6	11.9 ± 6.8	0.26
MTHFR 677C>T genotype (%)			
CC	36.7	45.8	—
CT	45	44.1	—
TT	18.3	10.2	0.37
Dietary intakes			
Energy (kcal/d)	1927 ± 475	1861 ± 361	0.46
Total dietary folate (μg/d)	280 ± 110	298 ± 106	0.43
Folic acid added to food (μg/d)	95 ± 78	112 ± 96	0.37
Natural food folate (μg/d)	185 ± 57	186 ± 46	0.89
Vitamin B-12 (μg/d)	4.1 ± 1.9	3.9 ± 1.8	0.69
B vitamin status			
Plasma homocysteine (μmol/L)	6.6 ± 2.4	6.3 ± 1.5	0.29
Serum folate (nmol/L)	45.7 ± 21.3	47.0 ± 21.0	0.74
Red blood cell folate (nmol/L)	1106 ± 746	1203 ± 639	0.45
Serum vitamin B-12 (pmol/L)	211 ± 82	235 ± 94	0.13

Table 1.2: Overview of whole-genome methods of DNA methylation analysis

Method	Platform	Genomic Coverage	Notes
Bisulfite Sequencing[123]	NGS	Whole Genome	Single-base pair resolution, full coverage. Considered to be the 'gold standard' technique.
MeDIP-Seq[124]	NGS	Enriched Methylated DNA	Genome-wide analysis lacking single-base-pair resolution
Anti-5mC MBDiGS[125]	NGS	Enriched Methylated DNA	Genome-wide analysis lacking single-base-pair resolution
MRE-Seq[125]	NGS	Size selected fraction	Low coverage, but capable of finding novel DMRs
MMSDK[133]	NGS	Representative genome tags	Low coverage, but capable of finding novel DMRs
MRE[133]	CGI microarray	CG islands	High coverage, but limited to array
MeDIP[133]	SMRT array	Whole Genome (low resolution)	High coverage, but limited to array
MeDIP[133]	Promoter Array	All biologically relevant promoters	High coverage, but limited to array
Sodium Bisulfite Treatment[133]	Illumina beadchip	Whole genome (high resolution)	Similar to Bisulfite Sequencing. High resolution coverage across genome.

NGS: Next Generation Sequencing, MeDIP: Methylated DNA Immunoprecipitation, MBDiGS: Methyl-binding domain isolated genome sequencing, MRE: Methylation-sensitive Restriction Endonucleases, CGI: CpG Island, SMRT, Submegabase resolution tiling array

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Biological Materials

Abcam: Rabbit polyclonal antibody to dihydrofolate reductase, (Cat. no. ab85056); Rabbit polyclonal antibody to GST, (Cat. no. ab19256); Mouse monoclonal antibody [33D3] to 5-Methylcytosine ChIP Grade, (Cat. no. ab10805).

ATCC: HEK293 cell line, (Cat. no. CRL-1573).

Bioline: Alpha Gold Standard cells [*Genotype: F- deoR endA1 recA1 relA1 gyrA96 hsdR17(rk -, mk +) supE44 thi-1 phoA Δ(lacZYA argF)U169 Φ80lacZΔM15λ*], (Cat. no. Bio-85027); BioScript, (Cat. no. Bio-27036); Human Genomic DNA, (Cat. no. Bio-35025); Isolate II RNA Extraction kit, (Cat. no. Bio-52072); ISOLATE Plasmid Mini Kit, (Cat. no. Bio-52026); ISOLATE RNA Mini Kit, (Cat. no. Bio-52043); MyTaq DNA Polymerase 2500 Units, (Cat. no. Bio-21106); Oligo dT, (Cat. no. Bio-38029); Random Hexamers, (Cat. no. Bio-38028); SensiFast ROX Step Kit, (Cat. no. Bio-76001); Velocity DNA polymerase, (Cat. no. Bio-21099).

Biosera: Foetal Bovine Serum (FBS), (Cat. no. S1900).

Gibco: Dulbecco's Modified Eagle Medium (DMEM) 1X, (Cat. No. 41965).

Greiner Bio-One: 25cm^2 flasks, (Cat. no. 690175CI); 75cm^3 flasks, (Cat. no. 658175CI); Cell Scrapers, (Cat. no. 541070G); 24-well plate, (Cat. no. 662160).

Invitrogen: Dynabeads M-280 Sheep Anti-Mouse IgG, (Cat. no. 112-01D); Dynabeads M-280 Streptavidin, (Cat. no. 112-05D); T4 DNA Ligase HC, (Cat. no. 15224-041); TA Cloning Kit, (Cat. no. k202020); UltraPure Phenol:Chloroform:Isoamyl Alcohol (PC8) (25:24:1, v/v), (Cat. no. 15593-03).

New England Biolabs: 100bp Ladder, (Cat. no. N3231S); AscI with 10x NEBuffer 4, (Cat. no. R0558); BL21(DE3) Competent Cells [*Genotype: fhuA2 [lon] ompT gal (λ DE3) [dcm] Δhsds λ DE3= λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*], (Cat. no. C2527H); MluI with 10x NEBuffer 3, (Cat. no. R0198); MmeI with 10x NEBuffer 4 & 32mM SAM, (Cat. no. R0637); MseI with 10x NEBuffer 4 & 100x BSA, (Cat. no. R0525S); NlaIII with 10x NEBuffer 4 & 100x BSA, (Cat. no. R0125); SOC Outgrowth Medium, (Cat. no. B9020S).

Promega: CellTiter 96 AQueous One Solution Cell Proliferation Assay (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium; MTS), (Cat. no. G3581).

Qiagen: EpiTect Fast Sodium Bisulfite Kit, (Cat. no. 59826); EpiTect PCR Control DNA Set, (Cat. no. 59695). MaXtract High Density Gel Column, (Cat. no. 129046); Qiaquick Gel Extraction Kit, (Cat. no. 28704); Qiagen Flexigene DNA kit, (Cat. no. 51204).

Roche: Glycogen, (Cat. no. 901-393); Lightcycler 480 High Resolution Melting Master, (Cat. no. 04909631001); Proteinase K recombinant PCR grade solution, (Cat. no. 03115887001).

Sigma: Acrylamide/Bis-acrylamide 30% Solution, 29:1, (Cat. no. A3574); Agarose, (Cat. no. A9539); Ammonium Acetate (7.5M), (Cat. no. A-2706); Carbenicillin disodium salt, (Cat. no. C1389); Dihydrofolate Reductase Assay Kit (Cat. no. CS0340); DL-Dithiothreitol, (Cat. no. D9779); DNase kit, (Cat. no. AMPD1); Ethanol, (Cat. no. E7023); Ethidium Bromide (Cat. no. E1385); Ethylenediaminetetraacetic acid, EDS; Formaldehyde (37%), (Cat. no. F-1268); Genomeplex Whole Genome Amplification kit, (Cat. no. WGA2); Glycerol, (Cat. no. G5516); Guanadine Hydrochloride, (Cat. no. G4505); Imprint DNA Modification Kit, (Cat. no. MOD50); L-Glutamine Solution, (Cat. no. 59202C); Magnesium Chloride, (Cat. no. M8787); N-lauroyl sarcosine, (Cat. no. L5125); Orange G, (Cat. no. O3756); PCR Buffer without Magnesium Chloride (10X), (Cat. no. P2317); Potassium Chloride, (Cat. no. P9541); Potassium Phosphate Dibasic, (Cat. no. P3786); Potassium Phosphate Monobasic, (Cat. no. P9791); Ribonuclease, (Cat. no. R4642); RNA Loading Buffer, (Cat. no. R1386); Sodium Pyruvate, (Cat. no. S8636); Taq Polymerase 5U/µl, (Cat. no. D4545); TEMED (Cat. no. T7024); Triethanolamine, (Cat. no. 90279); Tris, (Cat. no. T1503); Triton X, (Cat. no. X100); Trypan Blue, (Cat. no. T8154); Trypsin EDTA, (Cat. no. T4049).

Thermo Scientific: 4-20% Precise Tris-Glycine gel, (Cat. no. 0025269); Acetic Acid Glacial, (Cat. no. A0360PB17); Filter paper; (Cat. no. 84783); Glutathione Agarose, (Cat. no. 16100); Glutathione Reduced, (Cat. no. BP25215); O'Range Ruler 10bp DNA ladder, (Cat. no. SM1313); Page Ruler Plus Pre-stained Protein Ladder, (Cat. no. 26619); Pierce 4-20% Tris-Glycine Gel, (Cat. no. 0025269); Pierce Page Ruler Plus Pre-Stained Protein Ladder, (Cat. no. 26619); Pierce Sugar-Signal West Femto Maximum Sensor, (Cat. no. 34096); PVC Film, (Cat. no. 88518).

2.1.2: Stock Solutions and Reagents

10x Digestion Buffer: 50mM Tris (pH 8.0), 10mM EDTA, 0.5% SDS

10x Immunoprecipitation Buffer: 100mM sodium phosphate (pH 7.0), 1.4M sodium chloride, 0.5% Triton X100

Bacteria Lysis Buffer: 50mM Potassium Phosphate, 400mM NaCl, 100mM KCl, 10% glycerol, 0.5% Triton X-100, 10mM Imidazole

Coomassie Blue De-stain: 100ml Acetic Acid, 300ml Methanol, 600ml H_20

Coomassie Blue Stain: 2.5g Coomaisse, 100ml Acetic Acid, 300ml Methanol, 600ml H_20

Crush & Soak Elution Buffer: 0.5M ammonium acetate, 10mM magnesium acetate, 1mM EDTA (pH8.0), 0.1% SDS

Lamelli Buffer (10X): 15.14g Tris-HCl, 5g 1% SDS, 71.3g Glycine, in 500ml dH_2O

LB Agar: 3.75g agar added to 250ml LB broth

LB Broth: 250ml dH_2O, 2.5g Butotrpytone, 2.5g NaCl, 1.25g yeast extract

Low Tris Buffer (LoTE): 3mM tris-HCl pH 7.5, 0.2 mM EDTA, pH 7.5

Methylated DNA Immunoprecipitation (MeDIP) Ethanol Precipitation Solution: 2 volumes of ice-cold ethanol (100%), 0.3M sodium acetate, and 2µl glycogen

Modified Methylation Specific Digital Karyotyping (MMSDK) Ethanol Precipitation Solution: 2 volumes of ice-cold Ethanol (100%), 2.5M ammonium acetate, 2µl glycogen

MMSDK Isopropanol Precipitation Solution: 0.75 volumes of isopropanol and 6.33M sodium perchlorate

Orange G (10x): 20g Sucrose, 100mg Orange G in 50ml dH_2O

PBS-BSA (0.1%): 0.01g of BSA in 10ml of PBS

Protein Purification GST Equilibration / Wash Buffer: 50mM Tris, 150mM NaCl, (pH 8.0)

Protein Purification GST Elution Buffer: 50mM Tris, 150mM NaCl, 10mM Reduced Glutathione, (pH 8.0)

Resolubilisation Buffer: 1mM EDTA (pH 8.0), 1.5% N-lauryl sarcosine, 25mM Triethanolamine.

Protein Unfolding Buffer: 20mM potassium phosphate buffer (pH 7.0), 0.1mM EDTA, 10mM DTT, 0.2M KCl, 6M guanidine HCl

Protein Refolding Buffer: 20mM potassium phosphate buffer (pH 7.0), 0.1mM EDTA, 10mM DTT, 0.2M KCl, 20% Glycerol

TBE (10x): 48.44g Tris-HCl, 12.37g, Boric Acid, 1.5g EDTA, in 500ml dH_2O (pH 8.2)

TBS (10x): 43.83g NaCl, 6.06g Tris, in 500ml dH_2O (pH 8)

TBST: 100ml TBS (10x), 900ml dH_2O, 0.5ml Tween

Wash Buffer D: 5 mM Tris, 0.5 mM EDTA, 1 M NaCl, 200 μg/ml BSA (pH 7.5)

2.1.3 Oligonucleotide and primer sequence list

Modified Methylation-Specific Digital Karyotyping*:

Biotinylated Linker, Sense:	5′Biotin-TTTGCAGAGGTTCGTAATCGAGTTGGGTGG 3′
Biotinylated Linker, Antisense:	5′Phos-CGCGCCACCCAACTCGATTACGAACCTCTGC 3′
N-linker Sense:	5′ ACAGGTTCAGAGTTCTACAGTCCGACCATG 3′
N-linker Antisense:	5′ Phos-GTCGGACTGTAGAACTCTGAAC 3′
P7 Linker Sense:	5′ TCGTATGCCGTCTTCTGCTTG 3′
P7 Linker Anti-Sense:	5′ CAAGCAGAAGACGGCATACGANN 3′
P5 Long Primer:	5′AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA′3
P7 Primer:	5′CAAGCAGAAGACGGCATACGA 3′

All sequences taken from Li et al.*, 2009* [133]

Methylated DNA Immunoprecipitation:

H19 Forward:	5′ GAGCCGCACCAGATCTTCAG 3′
H19 Reverse:	5′ TTGGTGGAACACACTGTGATCA 3′
H3b Forward:	5′ CCCACACTTCTTATGCGACA 3′
H3b Reverse:	5′ CTGTGCCTGGTTGCAGATTA 3′

All sequences taken from Buckley et al., *2009*[148]

Sensitive Melting After Real Time Methylation-Specific PCR*:

COL2A1 Forward	5′ GTAATGTTAGGAGTATTTTGTGGGTA 3′
COL2A1 Reverse	5′ CTACCCCAAAAAAACCCAATCCTA 3′
Chr9ORF44 Forward	5′ GGGATGTTTTTGAGAATGGCGG 3′
Chr9ORF44 Reverse	5′ ACCTCAAAACCCATCGTACTCT 3′
GPS2 Forward	5′ GTTAGGAGGTTTTTTAAGGGTAGC 3′
GPS2 Reverse	5′ TTTTTCTTATCTAACAAATTCCGAA 3′
IP6K1 Forward	5′ TGTTAGTAGGAAGTATTTTTTTTGGCG 3′
IP6K1 Reverse	5′ AACTACAAAAAACCCGCCCCCG 3′
RASA4 Forward	5′ TGTTTTTGTGGGTGGATTTCGG 3′
RASA4 Reverse	5′ GAAAATAAAAACCCCAACTTTCCG 3

**All sequences designed using Methprime and Netprimer software, except for COL2A1 Forward and Reverse*.

Reverse Transcriptase-Quantitative PCR

IP6K1 Forward	5′ CAACAGTGGCTTGAGTTCTGA 3′
IPK1 Reverse	5′ TTCTCAAGCAGGAGGAACTTG 3′
RASA4 Forward	5′ CGTGGAGGGGAAGAACCT 3′
RASA4 Reverse	5′ TCCACCTTCACGATGCAGTA 3′
SURF1 Forward	5′ AACTCAGAGTGGGGCCTATG 3′
SURF 1 Reverse	5′ CCTGGGAACGAACCCTCTAT 3′
GPS2 Forward	5′ AGTCGGGCTTTGCAGCTA 3′
GPS2 Reverse	5′ GATGGTCACTTGTGGTAGAATCG 3′

All sequences designed using UPL design centre

cDNA Genomic Contamination Assay

MTHFDRQ Forward:	5′ CACTCCAGTGTTTGTCCATG 3′
MTHFDRQ Reverse	5′ GCATCTTGAGAGCCCTGA 3′

Primer sequences taken from Brody et al., *2002*[149].

2.2: Cell Culture Methods

2.2.1: HEK293 Cell Culture

Human Embryonic Kidney Cells 293 (HEK293) were cultured in GIBCO's Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% (v/v) foetal bovine serum, 1% (v/v) L-glutamine (Sigma-Aldrich) and 1% (v/v) sodium pyruvate (Sigma-Aldrich). Cells were grown in a 5% CO_2 incubator at 37°C in 10ml of media for 75cm^2 flasks, 6ml of media in T25cm^2 flasks, 500µl in 24-well plates, and 100µl in 96-well plates. Prior to each cell passage, the monolayer was rinsed with 2ml of Trypsin-EDTA (Sigma-Aldrich) before incubating with 3ml Trypsin-EDTA in 37°C for 10 minutes. Trypsin was inactivated with one volume of DMEM, and cells were centrifuged at 500 g for 5 minutes.

2.2.2: HEK293 Cell Counts

Cells were collected by centrifugation as described above (Section 2.2.1) prior to resuspension of the pellet in 6ml DMEM. A 1:6 dilution of cell-suspension was made in trypan blue (Sigma-Aldrich) and left at room temperature for 5 minutes before loading into a haemocytometer. At 100x magnification, viable cells were counted in four corners of the square grid. Average cell counts were calculated by multiplying the average across the four squares with the dilution factor and 10^4.

2.2.3: HEK293 Cell Growth Curve

Cells were counted as described above (Section 2.2.2) seeded at 1 x10^3, 2 x10^3, 4 x10^3, 8 x10^3, 1.6 x10^4, 3.2 x10^4, 6.4 x10^4, 1 x10^5 cells per ml in a 96-well plate for 7 days (37°C, 0.5% CO_2). Media was removed and 20µl MTS from Promega's CellTiter 96 AQueous One Solution Cell Proliferation Assay kit was added to each well and incubated at 37°C for 3 hours. Absorbance readings were taken at A_{273nm} on a Nanoquant Infinite M200 Tecan plate reader.

2.2.4: Cytotoxicity assay of 5'Azacytidine for HEK293 Cells

Using an optimal seeding density determined from a growth curve (Section 2.2.3), cells were grown in a 24-well plate with multiple concentrations of 5'azacytidine (5aC). Control wells were included on the plate containing no cells, cells below the optimal seeding density, and cells above the optimal seeding density. 5aC was prepared by making a 4mM solution in 1ml 50% acetic acid:water. Concentrations of 5aC including 0.5µM, 3µM, 6µM, 12µM, and 25µM were made from this stock, diluting the 4mM 5aC in DMEM. The highest concentration of acetic acid – for 25µM 5aC – was made by diluting 62.5µl of the 4mM stock into 10ml

DMEM. Wells with cells growing in 62.5µl 50% acetic acid:water diluted into 10ml DMEM were also included to control for the effect of the maximum concentration of acetic acid on cell proliferation. Prior to crystal violet staining, cells were rinsed with 1x PBS and incubated with 500µl of 10% formalin for 10 minutes at room temperature. Formalin was removed and wells were dried before adding 500µl 0.25% crystal violet and incubating for 10 minutes at room temperature. The stain was eluted in 300µl 33% glacial acetic acid (Thermo Scientific) for 30 minutes at room temperature with gentle rocking. Eluted crystal violet stain was removed and added to a 96-well plate in 100µl triplicates and analysed on a Tecan plate reader at A_{620nm}.

2.3: Molecular Biological Laboratory Techniques

2.3.1: Nucleic Acid Extraction from HEK293 Cells

For RNA extraction, cells were centrifuged at 500 g for 5 minutes, then washed with 1x PBS. Cells were lysed using the Lysis Buffer RLY supplied by the Isolate II RNA Mini Kit (Bioline). The rest of the protocol adhered to the manufacturer's instructions, with RNA eluted in 60µl RNase-free water. Although an on-column DNase I digestion step is included in the Isolate II RNA Mini Kit procedure, a second digestion was carried out using the DNase I Kit from Sigma-Aldrich. Briefly, 2µg of RNA was digested in 11µl reaction at room temperature for 15 minutes with 2 units of enzyme. The DNase I reaction was stopped using a Stop Solution supplied with the kit, incubated at 70^{o}C for 10 minutes.

DNA extraction was carried out using Qiagen's Flexigene DNA kit. Cells were counted and 2 x10^{6} were used in the protocol outlined in the manufacturer's instructions. Following this, samples were treated with Ribonuclease (Sigma-Aldrich) to remove any contaminating RNA. In a 200µl reaction, DNA was incubated for 1 hour at 37^{o}C with 1 unit of enzyme. Residual enzyme was removed via isopropanol precipitation with 3M sodium acetate as described in Section 2.3.6.

2.3.2 Non-denaturing Polyacrylamide Gel Electrophoresis

The 12% PAGE gels for DNA analysis were made using 30% bis-acrylamide (Sigma-Aldrich). With a total volume of 51.4ml, each gel contained 21.3ml bis-acrylamide, 6ml 1xTBE, 25.7ml dH_2O, 378µl 10% APS and 35µl TEMED (Sigma-Aldrich), added in that order. The gel was cast using glass plates and air-dried for over an hour as the gel polymerised. A GibcoBRL vertical electrophoresis rig (Model V15-17) with 1x TBE buffer was used to run

the gel for 90 minutes at 250 volts, with power supplied by an EC600-90 apparatus from Artisan Scientific. The gel was stained in an ethidium bromide bath of 1µg/ml for 30 minutes and visualised using a DNA Mini-Bis Pro Bio-imaging system.

2.3.3: Agarose Gel Electrophoresis

For a 1% agarose gels, 0.5g of agarose was dissolved in 50ml 1xTBE and heated in a microwave for one minute until the liquid began to boil. From here, the gel was cooled under a cold tap until it reached approximately 55^{o}C and 1µl ethidium bromide (10mg/ml) was added. For each run, 5µl of DNA (typically ranging from 100ng/µl to 500ng/µl) was mixed with 2µl with Orange G (Section 2.1.2). DNA fragments were separated at 90 volts for 30 minutes on a Watman Horizon S8 electrophoresis rig, with power supplied from an EC600-90 apparatus from Artisan Scientific.

2.3.4: SDS PAGE and Western Blot

Prior to electrophoresis, a solution containing 18µl of protein extract and 6µl 4X NuPage LDS was denatured for 10 minutes at 99^{o}C. On a 4-20% Precise Tris-Glycine gel (Thermo Scientific), 5µl Page Ruler Plus Pre-stained Protein Ladder (Thermo Scientific) and 24µl of the protein samples were added and run for 1 hour at 130 volts. Gels were stained overnight in coomassie blue, and de-stained with de-staining buffer at room temperature until the background colour was removed.

Protein transfer was carried out using the Pierce G2 Fast Blotter following the manufacturer's instructions. Briefly, gels were sandwiched with a sheet of PVC pre-deionised membrane (achieved by incubation for 30 seconds in methanol) between the cathode and anode, with four sheets of pre-wet filter paper separating gel and film from the electrodes. Transfer was carried out at 25V for 7 minutes. Once this was complete, the membrane was blocked with 5% fat-free milk in TBST for 2 hours. The primary antibody was added to the membrane in 5% fat-free milk (ratio variable depending on antibody used) and left to incubate overnight at 4^{o}C with gentle shaking. The film was washed three times with TBST for 10 minutes before adding the secondary anti-rabbit HRP-conjugated antibody at a ratio of 1:50,000 in TBST for 1 hour at room temperature. Membranes were washed three times again with TBST before imaging with SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific) on a Gene Gnome SynGene Bio Imaging System.

2.3.5: Estimation of nucleic acid concentration

Concentrations of both DNA and RNA were determined using a nanodrop ND-1000 spectrophotometer from Mason Technology measuring at A_{260nm}. For each sample, 1µl was used. Quality of nucleic acid was determined by measurement of the A260:280 ratios, with pure DNA falling between 1.8 and 2.

2.3.6: Phenol Extraction and Precipitation of DNA

This protocol used to extract and precipitate DNA was derived from Sambrook's molecular cloning[150]. DNA was mixed with one volume of UltraPure Phenol:Chloroform:Isoamyl Alcohol and spun down at 19,000 g in a Qiagen Maxtract gel extraction-tube, separating the aqueous nucleic acid phase from the organic phase. The former, upper phase was removed mixed with one of the three precipitation mixes listed in Section 2.1.2, made up fresh each time. Incubated on ice for 30 minutes and spun at 19,000 g at 4ºC for 30 minutes, the pellet produced was ready for resuspending in the appropriate solution after two washing steps with 70% ethanol. The centrifuge used was a Hettich Mikro 200R Zentrifugen machine.

2.3.7: TA Cloning and Transformation of Competent Cells

MMSDK PCR products were ligated into PCR 2.1 plasmid in a 10µl ligation reaction with 10x ligation buffer, 4.0 Weiss units of T4 ligase, and 50ng plasmid DNA. The concentration of PCR product used — in each case, 1.15ng — was determined by the calculation given in Invitrogen's TA Cloning manual, taking into account the length of the PCR product (90bp) and the size of the empty vector (3,900bp). This ligation reaction took place overnight at 14ºC. Competent alpha-gold standard *E. coli* cells from Bioline were transformed with the pCR 2.1 plasmid by heat shocking the cells as follows: For each 50µl vial of competent cells, 2µl of ligation was added and incubated on ice for 30 minutes. Immediately following this, the cells were incubated at 42ºC for 30 seconds then briefly placed back on ice. These cells were then added to 250µl LB media with no antibiotics and incubated at 37ºC for 1 hour while shaking. Each transformation was then spread onto LB plates containing ampicillin (100µg/ml in agar), X-gal (40µl of 40mg/ml) and IPTG (40µl of 100mM). After an overnight incubation, white colonies indicated positive recombinant clones and were analysed further and confirmed.

Transformation of DHFR recombinant clones followed a similar transformation procedure as above, but OneShot BL21A1 cells were transformed with 100ng of the GST-tagged plasmid previously prepared by Dr. Mari Ozaki. Agar plates and overnight cultures contained both contained 50µg/ml of carbenicillin.

2.3.8: Recombinant Protein Production

Once transformation was confirmed on carbenicillin plates, a single colony was removed and incubated in 120ml of carbenicillin broth (50μg/ml) overnight. Afterwards, a 1/20 dilution of the bacteria was made in LB broth with carbenicillin (50μg/ml) up to 2L. This was split into two 1L flasks (400ml in each) and six 500ml flasks (200ml each) and incubated at 37ºC until the OD of each flask reached 0.4 at A_{600nm}. At this point, the temperature was brought down to 30ºC for 1 hour. All incubations involving broth were set to shake at 220 RPM. Cells were centrifuged at 4,000 g at 4ºC for 5 minutes and the supernatant was removed. Pellets were weighed and stored at -20ºC.

2.3.9: Recombinant Protein Fractionation

Pellets were thawed and lysed in bacterial lysis buffer at 8ml per 0.4g cells. A homogenous mixture was obtained by pipetting cells up and down and passing the fluid through a 27-gauge needle. Cells were then put through three freeze-thaw cycles in liquid nitrogen and a 42ºC water-bath. Lysed cells were centrifuged at 4,000 g for 15 minutes at 4ºC. Both the soluble and insoluble fractions were retained and stored at -20ºC.

2.3.10: Concentration of DHFR soluble fraction

Protein lysates were concentrated using 30 kDa centrifugal filter unites (Millipore), spinning at 4,000 g for 1 hour at 4ºC. These were then diluted 1 in 5 in bacterial lysis buffer and syringed-filtered through 0.45μM filters.

2.3.11: Resolubilisation of DHFR insoluble fraction via Folding and Refolding

To remove DHFR from the insoluble fraction, it was necessary to denature the enzyme with an Unfolding Buffer (20mM potassium phosphate buffer (pH 7.0), 0.1mM EDTA, 10mM DTT, 0.2M KCl, 6M guanidine HCl), and return it to its active state with a Refolding Buffer (20mM potassium phosphate buffer (pH 7.0), 0.1mM EDTA, 10mM DTT, 0.2M KCl, 20% glycerol). Pellets were resuspended in 10ml of the former, and stirred slowly at 4ºC for 1 hour. The protein was then added drop-wise to a 20 fold dilution of the Refolding Buffer. Aggregates began forming after storing samples overnight at 4ºC. This method was carried out according to Sirawaraporn e*t al.*, (1993)[151].

2.3.12: Resolubilisation of DHFR insoluble fraction via Resuspension with N-lauroyl-sarcosine and triethanolamine

Another method for resuspending the insoluble fraction of DHFR was taken from Molecular Probe's product insert for their product, Glutathione Agarose Linked through Sulfur (G2879). Recommended by the manufactures of the GST system if the protein of interest was found in the insoluble fraction, this method involved a single resuspension step in a buffer containing 1.5% N-lauroyl-sarcosine, 25mM triethanolamine, and 1mM EDTA (pH 8.0) and mixing for ten minutes before centrifugation for 10 minutes at 4^{o}C.

2.3.13: Gravity-Flow Column Purification of GST-tagged Protein

Protein purification was carried out using Pierce Glutathione Agarose (Thermo Scientific). Prior to this, the column was washed with 1x PBS and packed with an appropriate volume glutathione beads. The storage buffer was drained slowly from the beads, reducing the volume inside the column by half. The column was then equilibrated with 10 volumes Equilibration/Wash Buffer (50mM Tris, 150mM NaCl, pH 8.0). The protein sample was put through the column twice, ensuring the resin remained wet and undisturbed. The beads were then washed with 10 volumes of Equilibration/Wash Buffer, until the absorbance the wash fraction reached baseline at A_{280nm}. The targeted protein – tagged with GST – was then eluted with 2 volumes of Elution Buffer (50mM Tris, 150mM NaCl, and 10mM Reduced Glutathione, pH 8.0). Multiple fractions were taken at this point, and the absorbance at A_{280nm} was tested. Protein purification manifested as a spike in absorbance in a particular fraction (usually the fourth of fifth). Purified fractions were pooled and concentrated as aforementioned, and analysed by SDS-PAGE and Western Blot analysis.

2.3.14: Batch Purification of GST-tagged Protein

For smaller batches of lysate, binding to the glutathione column was carried out in a 15ml tube. Tubes were centrifuged with 2ml of glutathione-beads inside to reduce the resin volume to 1ml. Washing and preparation of the column from this step forward was similar as above, with the beads washed with 10ml of Equilibration/Wash Buffer. Binding took place overnight shaking at 4^{o}C. Afterwards, the resin was washed twice with the same volume of Equilibration/Wash Buffer. Elution was carried out multiple times with 1ml Elution Buffer. Absorbance was observed at A_{280nm} with each elution, and analysed by SDS-PAGE and Western Blotting.

2.3.15: Bradford Protein Assay

Bovine Serum Albumin standards from 0mg/ml to 1.25mg/ml were prepared in distilled water. On a 96-well plate, 5µl of each standard was added to 250µl Bradford reagent in triplicate. Samples were prepared in the same way, with the relevant background medium used as a blank. The plate was shaken for 30 seconds and incubated for 10 minutes at room temperature. Absorbance was read at A_{595nm} on a Tecan plate-reader.

2.3.16: DHFR Kit assay for enzyme activity

Enzyme activity of DHFR was determined using Sigma-Aldrich's Dihydrofolate Reductase Assay kit with minor changes to the manufacturer's protocol. Each 500µl reaction was set up in a cuvette with the 1x Assay Buffer, DHFR enzyme (variable), 60µM NADPH, and 50µM DHF, added in that order. At a concentration of 0.044µg/µl, 500ng of DHFR was used in these reactions. Absorbance was measured on a bench-top Biochrom Libra S12 spectrophotometer at A_{340nm} under a kinetic programme reading every 15 seconds for 2.5 minutes. Inhibition assays were carried out with 0.2µM methotrexate under the same conditions to ensure NADPH depletion was cause by DHFR activity alone.

2.3.17: *In vitro* SUMOylation assay

SUMOylation assays were carried out using the SUMO-1 link from Active Motif according to the manufacturer's instructions. A total of seven reactions were carried out per experiment, with tubes blocked with 50ng/ml BSA (Sigma-Aldrich) for 1 hour at 37ºC before starting. Each 20µl reaction contained 5x SUMOylation Buffer, 20x Protein Buffer, and 1µl each of E1 and E2 conjugating enzymes. To assess the SUMOylating capacity of the kit, supplied p53 protein was used as a control, with the protein known to contain a SUMOylation site. As a negative control, a reaction containing no protein of interest was set up to control for self-SUMOylation of the conjugating and activating enzyme complex. For each protein assessed, a second reaction containing a mutant SUMO1 protein was included — this is a protein incapable of being conjugated to the peptide of interest.

Reactions were incubated at 30ºC for 3 hours and stopped with 2X SDS-PAGE Loading Buffer (130mM Tris, pH 6.8%, 4% SDS, 0.02% Bromophenol blue, 20% glycerol, 100mM DTT). Samples were analysed via PAGE and Western blotting with supplied anti-SUMO and anti-p53 antibodies at ratios of 1:4000 and 1:5000 respectively. Probing with anti-DHFR, was carried out a ratio of 1:10,000. Anti-rabbit HRP conjugated secondary antibody was used at 1:25,000.

2.3.18: Plasmid DNA Preparation from Bacterial Cells

Using the ISOLATE plasmid mini-kit from Bioline, plasmids from transformed cells were extracted and analysed. Cells in the 15ml tubes were pelleted by spinning at 19,000 g for 1 minute and as much supernatant as possible was removed. The pellet was resuspended in 250µl resuspension buffer and mixed thoroughly. To this, 250µl lysis buffer was added and mixed by inversion. Then, 350µl neutralisation buffer was added to the sample and mixed carefully again by inversion. Afterwards, the samples were centrifuged for 10 minutes at full speed. This was all transferred to spin column P with a collection tube and centrifuged at full speed for 1 minute. The filtrate was discarded while the spin column was reused and washed with 500µl wash buffer AP with full speed spinning for 1 minute. Removing the filtrate, the same spin column was washed again with 700µl wash buffer BP and centrifuged at full speed for 1 minute. With the filtrate removed, the column was centrifuged at full speed for 2 minutes to remove any residual liquid. Finally, the column was placed into a fresh collection tube and the DNA was eluted with Elution Buffer by centrifugation for 1 minute. Isolated DNA was quantified via the ND-1000 nanodrop spectrophotometry from Mason Technology.

2.3.19: Sodium Bisulfite Treatment of DNA

For each Qiagen FAST DNA Bisulfite reaction, 1µg of DNA was used—whether obtained from the FASSTT (Folic Acid Supplementation during the Second and Third Trimester) cohort or from HEK293 cells. Following the manufacturer's instructions, samples were eluted in 100µl H_2O. Due to the nature of sodium bisulfite treated DNA, converted samples could not be quantified or qualified through agarose gel electrophoresis or through nanodropping. The COL2A1 SMART-MSP assay (Section 2.3.20) was used to confirm that converted DNA was obtained.

2.3.20: Sensitive Melting After Real Time Methylation Specific PCR

For each SMART-MSP assay, primers were designed to preferentially amplify methylated DNA that has been treated with sodium bisulfite. The COL2A1 assay, however, acted as a measure of the total amplifiable DNA in each sample. With primers containing no CpG sites in their sequence or in their product, all converted DNA — methylated or not — would be targeted. Primers were designed using MethPrime and NetPrimer, following the criteria listed by[146]. Secondary structure formation was kept to a minimum where possible. A melting analysis was carried out after each qPCR run, which was used to determine if the targeted product did indeed amplify. Each 10µl reaction contained 25ng DNA, 1mM MgCl, and 5x HRM Master Mix from Roche. Primer concentrations and Tm values varied from assay to

assay (described in more detail in Chapter 5). All runs were carried out on 96-well plates designed to be used by the Roche Lightcycler 480™ Instrument.

2.3.21: cDNA Synthesis of HEK293 extracted RNA

For each reverse transcription reaction, 5.5µl of DNase-treated RNA was used—half of the 11µl of the DNase I reaction outlined in Section 2.3.1. Following the manufacturer's instructions for the Bioline SuperScript Kit, 5.5µl DNase-treated RNA (1µg) was transferred to a fresh 0.2µl tube containing a mix of 2µl random hexamers (50ng/µl) and 4µl oligo-DT (270ng/µl). After incubation for 70ºC for 5 minutes, samples were cooled on ice and added to a reaction containing 4µl Reaction Buffer (5x), 1µl RiboSafe RNase inhibitor (40u/µl) 1µl Bioline Bioscript reverse transcriptase enzyme (200u/µl), 1µl dNTPs (10mM), and 1.5µl nuclease-free water. Samples were incubated under the following programme: 10 minutes, 25ºC; 60 minutes, 42 ºC; and 15 minutes, 70 ºC.

2.3.22: Genomic DNA Contamination Test

Although two DNase-treatment steps are included in the cDNA synthesis process, it was imperative to ensure that no genomic DNA was contaminating the RT qPCR reactions. A simple PCR assay was designed to flank an intron, generating a larger fragment for genomic DNA than for cDNA. The genomic region chosen, within the coding region of the *MTHFD1* gene, yielded a 232bp band for cDNA, and 330bp for genomic DNA. This difference can be easily visualised on an agarose gel. Only samples containing no contaminating DNA were brought forward to RT qPCR analysis. The 50µl PCR reaction containing 1x PCR buffer, 0.3µM forward primer, 0.4µM reverse primer (sequences in Section 2.1.3), 0.2mM dNTPs, and 1.5mM MgCl was carried out using the following programme: 95ºC for 3 minutes; 35 cycles of 94ºC for 30 seconds, 58ºC for 1 minute, 72ºC for 1 minute; and 72ºC for 1 minute.

2.3.23: Reverse Transcriptase-Quantitative PCR

All RT qPCR reactions were carried out on the Roche 480 Lightcycler machine with assays designed by the Universal ProbeLibrary Design Centre. Each reaction contained a 1/10 dilution of the supplied and recommended probe, 2x Probe Master, 2.5mM $MgCl_2$, and 1µl cDNA produced from Section 2.3.21.

2.3.24 Data Analysis and Statistics

All data analysis was carried out using Microsoft Excel 2013. Macro and sub macro codes for DNA microarray analysis were written in Visual Basic 8 (more detail in Section 3.3).

Significance values in Chapter 4 were obtained using single-factor ANOVA analysis from Excel's Analysis Toolpack add-in (Section 4.2.8).

Chapter 3:

DNA Methylation Method Development and Identification of Folate Sensitive Differential Methylation Regions (FS-DMRs)

3.1 Introduction

3.1.1: The Human DNA Methylome

Cytosine methylation is known to occur primarily at CpG sites in mammals: a dinucleotide typically underrepresented in their genomes. In general, about 60-80% of these CpGs are methylated[152]. A smaller fraction are located in dense regions called CpG islands (CGIs), and are classically associated with transcription start sites of housekeeping, developmental, or regulatory genes, and thus are mainly devoid of DNA methylation[153].

However, as technology improved and the field of epigenetics progressed, many cases have been found that violate this traditional model. For example, half of all CGIs are known to occur outside annotated promoter regions[154]. These 'orphan' CGIs show similar epigenetic features to promoter CGIs, and are both intergenic and intragenic in mammalian genomes[154,155]. At the same time, some tissue-specific and cancer-related differentially methylated regions have been found to occur more frequently in CGI 'shores' – regions flanking CGIs with relatively low CpG levels[156,157]. Recently, gene-body methylation has been associated with increased levels of expression in many tissue and cell-types; something that was once thought to be restricted to specific cases such as the active X-linked human chromosome and the genome of *Aradopsis thaliana*[158–160].

The role of DNA methylation at a tissue-specific and disease-specific context still remains elusive. With variations in DNA methylation appearing between different cells, tissues, and individuals, mapping all of these changes is proving to be a significant challenge. Still, building on a collective understanding of DNA methylation distribution, projects like "DiseaseMeth" and "mPod" are striving towards the identification of how aberrant methylation changes can have such a negative impact on a healthy individual, all while shedding light on the nature of DNA methylation itself[161,162].

There is still much to be learned, even outside the context of DNA methylation distribution in disease. In a review published by this research group, nutrient sensitive differentially methylated regions were identified from the literature, several of which were observed in response to folic acid supplementation in humans[3].

Using a genome-wide method of DNA methylation analysis, the work presented here will expand on this list by examining regions that alter their methylation status in response to folic acid supplementation in pregnant women.

3.1.2: Selection of a genome-wide method of DNA methylation analysis

The first step of the project was to select a method for genome-wide methylation analysis of the FASSTT samples (Folic Acid Supplementation during the Second and Third Trimester), described in Section 1.2.2. It was imperative that the method selected would be of adequate efficiency and sensitivity with respect to the aims of the project. It was also essential that the method undergo vigorous testing in the laboratory before processing the FASSTT samples, due to the precious nature of the DNA. In 2011, we published a review in *Frontiers in Genetics* outlining the major differences between various strategies of DNA methylation analysis in a historical context[5]. This was used as a framework for selecting the appropriate protocol

Various methods of genome-wide DNA methylation analysis have been discussed in Section 1.7. Two that were considered for our initial analysis were modified methylation-specific digital karyotyping (MMSDK) and methylated DNA immunoprecipitation (MeDIP).

MMSDK uses methylation-specific restriction enzymes to differentiate between unmethylated and methylated DNA, with the regions themselves being identified using Next Generation Sequencing. On the other hand, MeDIP uses antibodies raised against 5'methylcystosine to precipitate methylated DNA, which is then hybridised to a promoter microarray to measure enrichment levels across the genome. MeDIP will not yield single-base-pair resolution data, as the hybridisation process will only be used to identify CGIs that have been enriched during precipitation. Although it could be argued that this shouldn't be an issue – with CpG islands spanning hundreds of base-pairs co-methylated in human cells anyway[163] – our analysis of the FASSTT cohort aims to find novel FS-DMRs; some of which may lie outside traditional CpG islands. As a result, MMSDK was given precedence over MeDIP as a method of genome-wide methylation analysis.

Two pilot studies were carried out on MMSDK. First, it was found that this method was capable of generating a genomic library for lambda DNA. However, when this was scaled up for human genomic DNA, no products could be obtained. After extensive optimisation, MMSDK could not be applied to the FASSTT samples, and the second method, MeDIP, was considered instead.

The MeDIP protocol requires genomic DNA to be sheared and denatured prior to enrichment with anti-methylcytosine[138]. Although many protocols use sonication for the former, we used the restriction enzyme MseI to generate fragments of DNA between 200bp to 1000bp in size. This approach was favoured over sonication as the restriction digestion approach does not require excessive optimisation and testing to achieve the desired outcome. Also, with MseI

digestion being specific to a particular sequence motif, the sizes and contexts of fragments generated could be determined ahead of time.

Following immunoprecipitation, quality control analysis was carried out after enrichment to demonstrate that MeDIP successfully separated methylated DNA from unmethylated DNA. This analysis demonstrated that MeDIP had successfully enriched for methylated DNA from the FASSTT cohort – something that could not be achieved from MMSDK. As a result, MeDIP was chosen for the genome-wide methylation analysis.

MeDIP DNA was hybridised on a Nimblegen Delux 2.1M Promoter Array and compared against "Input DNA" – DNA prepared in the same manner as the MeDIP samples, but without immunoprecipitation.

Based on the selection criteria outlined in Section 3.2.10, DNA from 6 participants who exhibited a typical response to folic acid supplementation based on circulating folate levels were precipitated through MeDIP and hybridised to a DNA microarray covering every biologically relevant promoter region of the human genome known at the time of the experiment. The vast amount of data received from this analysis provided a significant challenge in the data analysis outlined in Section 3.3. Once potential FS-DMRs are identified, they'll be analysed further in a gene-specific context, as described in Figure 3.1.

3.1.3: Aims and Objectives

Aim:

To carry out an assessment of the feasibility of applying the genome-wide DNA methylation analysis methods MMSDK and MeDIP to the FASSTT cohort (n=119). Following this, the most suitable method will be used to find potential FS-DMRs from a subset of the FASSTT cohort (n = 6).

Objectives:

- To carry out a pilot study on lambda DNA using MMSDK to see if it is capable of generating a genomic library at a small scale.
- To carry out a second pilot study on human genomic DNA using MMSDK to see if it is capable of generating a genomic library at a larger scale.
- To assess the precipitation of methylated DNA via MeDIP using qPCR.

- To carry out a genome-wide method of DNA methylation analysis (MMSDK or MeDIP) on suitable subset of the FASSTT cohort (n=6), selected to represent a typical response to folic acid supplementation based on serum and red-cell folate levels.

3.2: Methods

3.2.1: The FASSTT Study

In an intervention study carried out by the University of Ulster, 119 pregnant women were recruited and given either a folic acid supplement or a placebo during their second and third trimesters. The study, called Folic Acid Supplementation during the Second and Third Trimester (FASSTT), aimed to analyse the effect of folic acid supplementation on circulating folate and homocysteine levels in mother and child without interfering with the beneficial effects of supplementation in the first trimester[4].

Blood samples were taken from women before intervention, after 36 weeks gestation, and from the umbilical cord of the child upon delivery (Chapter 1, Figure 1.1). From these samples, plasma homocysteine, serum folate, and red-cell folate levels were determined using immunological and microbiological assays, carried out by Dr. Breige McNulty in the University of Ulster. DNA was extracted from the buffy coats of the blood in our laboratory by Dr. Mari Ozaki using the Flexigene DNA extraction kit (Qiagen).

3.2.2: Annealing of MMSDK Ligation linkers

The MMSDK protocol requires double-stranded ligation linkers. Individual strands were synthesised by Integrated DNA Technologies [http://eu.idtdna.com/site] and annealed via incubation at 95ºC for 5 minutes on an Applied Biosystems GeneAmp PCR 9700 thermal cycler. The samples were cooled slowly to room temperature with the following incubations: 65ºC for 20 minutes, 37ºC for 20 minutes, and 25ºC for 20 minutes. Linkers were stored at -20ºC once annealed.

3.2.3: Modified methylation specific digital karyotyping

The MMSDK process involves three restriction enzyme digestions and three linker ligation steps[133,134]. In the first reaction, 5µg of DNA (lambda or human) was digested with 120 units of AscI and 1x NEB buffer in a 200µl reaction containing LoTE for 90 minutes at 37ºC. Afterwards, the reaction was placed in a Maxtract gel tube (Qiagen) with 200µl UltraPure Phenol:Chloroform:Isoamyl Alcohol (25:24:1) and centrifuged for 5 minutes. The 200µl

nucleic acid phase (top) was extracted and ethanol precipitated as described in Section 2.3.6, and the pellet was resuspended in 5µl LoTE. Overnight, 2.5µl of 1µM annealed biotinylated linker was ligated to the AscI fragments with 5 units of T4 DNA ligase and the supplied ligase buffer in a 12µl reaction. Before adding the enzyme, the reaction was heated to 50ºC for 2 minutes and cooled to room temperature for 10 minutes. Ligation took place overnight at 16ºC. Afterwards, 188µl LoTE was added to the ligation reaction, followed by phenol extraction and isopropanol precipitation (Section 2.3.6). The pellet was then resuspended in 172µl LoTE. A 200µl NlaIII digestion was carried out for 1 hour at 37ºC containing 1x NEB buffer 4, 1x BSA and 60 units of NlaIII. Before binding the NlaIII fragments to streptavidin Dynabeads, the beads were isolated from suspension using a magnetic stand. The isolated beads were washed with 400µl Wash Buffer D. The 200µl NlaIII digest was added directly to the beads and made up to a further 600µl with more Wash Buffer D. The beads and DNA fragments were incubated at room temperature for 20 minutes with gentle mixing to ensure that the mixture remained homogenous. After this binding step, the beads were washed with 600µl wash buffer D three times, and once with 300µl 1x ligation buffer. Next, 2.5µl of 2µM Annealed N-linkers were ligated directly onto the DNA-bound beads at 16ºC overnight with gentle mixing. This reaction was 40µl in volume made up with LoTE and 1x ligation buffer. As before, the mix was incubated at 50ºC for 2 minutes and cooled to room temperature for 10 minutes before adding the enzyme, in this case 12.5 units. Afterwards, the beads were washed three times with 600µl wash buffer D and transferred into another tube. Here, they were washed once more with 600µl wash buffer D and 200µl 1x NEBuffer 4. Tags were then released from the streptavidin Dynabeads through MmeI digestion. This was a 150µl reaction with 30 units of enzyme, 15µl of 500µM SAM (S-adenosylmethionine) and 10x NEBuffer 4 in LoTE. The tube was incubated for 1 hour and 10 minutes with gentle mixing at 37ºC. Samples were centrifuged for 2 minutes at full speed and made up to 300µl with LoTE. This was added to a pre-spun Maxtract tube as before with 300µl Phenol:Chloroform:Isoamyl and then centrifuged for 5 minutes (full speed) at room temperature. A final ethanol precipitation step was carried out, similarly to before, but at 2ml total volume. This contained 300µl of the tags, 4µl glycogen, 200µl 7.5M ammonium acetate and 1.5ml 100% pre-chilled 100% ethanol. The pellet was resuspended in 12µl LoTE. The final ligation step had the same conditions as the previous reactions, with 0.13µM linker P7. A reaction without the ligase enzyme was included to control for downstream amplification of templates without P7 linkers.

3.2.4: PCR amplification of MMSDK products

Once obtained, the MMSDK final product from lambda DNA was amplified using the MyTaq polymerase from Bioline and 1µM each of P5 and P7 primers (Section 2.1.3)[133]. The reaction

was as follows: 98ºC for 30 seconds; 25 cycles of 98ºC for 10 seconds, 60ºC for 30 seconds, 75ºC for 15 seconds; 75ºC for 10 minutes and cooling at 4ºC. Products were run on a 12% PAGE gel, with an expected product size of 88bp. For human genomic DNA, Velocity polymerase was used in a 50µl reaction containing 1mM dNTPs, 1µM of each P5 and P7 primers, 3% DMSO, and 2 units of enzyme.

3.2.5: DNA extraction form PAGE gel

DNA from bands on PAGE gels were extracted using an adapted version of the Crush and Soak method described in Molecular Cloning by Sambrook *et al.*, (1989)[150]. The bands were cut out using a sterile scalpel, transferred to micro-centrifuge tubes, and crushed using pipette tips. From here, two volumes of Crush and Soak elution buffer were added to the gel slices and incubated at 37ºC while rotating for 3 hours. At 4ºC, the gel extracts were centrifuged for 1 minute, with the supernatant removed afterwards. From the Qiaquick Gel Extraction Kit from Qiagen, 3 volumes of Buffer QG were added to this supernatant. This was applied to QIAquick Spin columns in provided 2ml collection tubes and centrifuged for 60 seconds at room temperature at full speed. Discarding the flow-through, the columns were washed with 750µl Buffer PE and centrifuged again for 60 seconds. After discarding this flow-through, the empty column was spun again at full speed to remove any residual ethanol from the tube (contained in buffer PE). DNA was eluted using 50µl of water and quantified using the nanodrop spectrophotometer (Section 2.3.5).

3.2.6: Methylated DNA Immunoprecipitation

For each DNA sample analysed, 2µg was sheared using the enzyme MseI in a 100µl reaction containing BSA and 10x NEBuffer 4 at 37ºC for 1 hour. After this, the reaction volume was brought to 225µl using 1x TE Buffer. The samples were denatured at 95ºC for 10 minutes in order to increase the affinity of the anti-5'methylcytosine antibody (Abcam) to the fragmented DNA. Before immunoprecipitation, 25µl of 10x IP Buffer was added to the samples. Overnight, 4µg of anti-5'methylcytosine was incubated at 4ºC with the fragmented DNA. After immuno-capturing, the samples were enriched using a secondary antibody pre-bound to magnetic sheep anti-mouse Dynabeads (4×10^8 beads/ml, Invitrogen). The anti-mouse sheep monoclonal antibodies have an affinity to the mouse anti-5'methylcytosine, and the beads can be purified quickly and efficiently using a magnetic stand. Dynabeads were pre-blocked before this process with 800µl PBS-BSA (0.1%) by rotation for 5 minutes at room temperature and eluted in 20µl 1X IP buffer. This elution was added directly to the immuno-captured DNA, and incubated for 2 hours at room temperature while shaking. While methylated DNA was bound to the beads via an antibody-antibody complex, the beads were washed three times by

gentle shaking for 5 minutes at room temperature with 700µl 1x IP buffer to remove the unmethylated fragments. The beads were then suspended in 250µl digestion buffer and incubated at 95ºC for 5 minutes. Following this, 315µg Proteinase K (Roche) was added and digestion occurred at 55ºC overnight while rotating. The DNA was precipitated from here to remove residual proteins, including the proteinase K enzyme itself. In accordance to the manufacturer's instructions, 50µl Phenol:Chloroform:Isoamyl Alcohol and 250µl of the sample and mixed in a Maxtract tube (Qiagen) and the organic phase was extracted. The 250µl of DNA removed from this was brought to 500µl with 0.2M NaCl. Finally, 2 volumes of ice-cold ethanol and 3µl glycogen were added to the sample and precipitated at -20ºC for 1-2 hours. The DNA pellet was formed by spinning at 19,000 g for 30 minutes at 4ºC, and this was washed twice with 70% ethanol with 15 minute spinning in between. The final DNA pellet was resuspended in 30µl of molecular grade H_2O, and the final concentration was determined using the ND-100 nanodrop spectrophotometer (Mason).

3.2.7: Whole genome-wide amplification of MeDIP DNA

The precipitated DNA was then whole-genome-amplified using Sigma-Aldrich's Genomeplex Whole Genome Amplification kit. As the DNA here has already been fragmented by the MseI enzyme, the shearing step in the manufacturer's instructions was left out. The genomic library was prepared from a 1ng/µl solution of DNA by adding 2µl of Library Preparation Buffer and 1µl of Stabilisation Solution. The samples were vortexed and centrifuged before incubating at 95ºC for 2 minutes. After cooling on ice, 1µl of Library Preparation Enzyme was added and the following reaction took place in the Applied Biosystems GeneAmp PCR 9700 thermal cycler: 16ºC for 20 minutes, 24ºC for 20 minutes, 37ºC for 20 minutes, 75ºC for 20 minutes, and 4ºC hold. Whole genome-wide amplification was carried out by adding the following solution to the 15µl library: 7.5µl of 10x Amplification Master Mix, 47.5µl dH2O, and 5µl WGA DNA polymerase. The PCR programme was as follows: 95ºC for 3 minutes; 94ºC for 15 seconds and 65ºC for 5 minutes over 14 cycles. Amplified DNA was analysed on a 1% agarose gel, quantified on a nanodrop spectrophotometer and qualified using the qPCR assays described below.

3.2.8: Quality Control Analysis of Immunoprecipitated DNA

After the methylated fragments were precipitated, the quality of the assay was assessed. Two qPCR assays were carried out on the MeDIP DNA using primers directed at regions that are known to be methylated and unmethylated. The long non-coding RNA gene *H19* was selected to represent the former, while the histone gene *H3B* was chosen to select for the latter[148]. Both assays were run on the following programme: 95ºC 10 minutes; 95ºC for 10 seconds, 56ºC for

15 seconds, 72ºC for 15 seconds for 45 cycles; and 40ºC for 10 seconds. Primer concentrations were at 0.2µM per reaction

For qPCR, DNA amplification is quantified as C_P values: the number of cycles taken for fluorescence to reach a defined threshold. When comparing the C_P levels of MeDIP against Input DNA, given that both samples will have the same concentration, positive enrichment of methylated DNA would manifest as a lower C_P value in H19, and diminishment of the unmethylated portion manifest as an increase in C_P values for H3B.

Over the twelve FASSTT samples in each QC assay, a mean value was taken and compared to the C_P of each Input. With the following formula, the mean C_P value and the mean Input C_P value were used to separately calculate the fold change for the assays:

$$Fold\ Change = 2\ \hat{}\ (Input - FASSTT)$$

DNA quality was assessed on a 1% agarose gel (Section 2.3.3). A thick smear from 200 – 1000bp in length is representative of successfully precipitated DNA.

3.2.9: Nimblegen Microarray Analysis

The following was carried out by Nimblegen: MeDIP samples were labelled using Cy5 random nanomers, while Input DNA was labelled with Cy3 nanomers. Input and MeDIP fractions were mixed and hybridised to the array using a Precision Mixer Alignment Tool (PMAT). Each PMAT was washed and incubated at 42ºC in a Roche Hybridisation system for 16 hours. All 12 samples were loaded into a NimbleGen MS 200 Microarray Scanner in one Slide Magazine. Image files were extracted and processed using DEVA software.

Data was returned in the form of pair files, processed peak files, Signal Map GFF files, and promoter reports. The data analysis discussed here is based on the promoter reports, which contain annotated information for each peak, along with the $\log_2$ ratios for the corresponding probes. Raw data analysis on the peak-probe relationship is based on the Signal Map GFF files.

3.2.10: Selection of samples for genome-wide analysis

From the entire FASSTT cohort, 12 samples were selected from 6 women to reflect the most dramatic responses to intervention. As the variation of DNA methylation between individuals is unknown (either caused by genetic diversity or environmental differences)[164], the pre-intervention samples of each of the 6 participants was paired with their respective post-

intervention samples. MeDIP analysis was used to investigate the effect of intervention on DNA methylation for the same individuals first, reducing the amount of DNA methylation 'noise' that is likely to exist between individuals.

One limit of the FASSTT study itself is the fact that intervention occurred during the second trimester, after the known beneficial effects of folic acid supplementation have taken place. As it would have been highly unethical to recruit participants at the periconceptional stage of pregnancy, women from both groups who displayed low baseline levels of red cell and serum folate before intervention were considered for MeDIP analysis. Selection was also based on response to folic acid supplementation: those from the supplementation group who had high levels of red cell and serum folate after intervention were chosen, along with those from the placebo group who exhibited the typical decline in red cell and serum folate during pregnancy (Table 3.1). All women were taking folic acid before recruitment to this trial. The aim of this selection process was to find that would represent the entire cohort in terms of response to intervention. Although bias may be introduced as a result of this selection – ruling out women who may not have a dramatic response to folic acid – the purpose of our genome-wide analysis is not to find definitive FS-DMRs, but to find potential FS-DMRs for further examination in Chapter 4. This selection process was carried out in consultation with Dr. Kristina Pentieva in the University of Ulster, Coleraine.

3.3: DNA Microarray Data Analysis

3.3.1: Introduction

DNA Microarray data was received in several raw and processed file formats, including Excel files containing annotated information for each "peak" — groups of probes across a genomic region reaching a significant level of enrichment. These were used to obtain our initial FS-DMR lists, while the raw data was used to narrow down the candidates further.

Using Visual Basic 6.0, three looping macros and several sub macros were written in-lab to sort through the Excel data and compare the peak scores of each processed locus over all the samples. The aim of this analysis was to generate a list of genes that significantly changed their DNA methylation signature in response to intervention

First, fold change values were compared between each of the 12 samples (6 participants, pre- and post-intervention) and a "peak change" score in response to intervention was obtained (Section 3.3.2). After this, the recurring genes between the three folic acid group participants were compared with those from the placebo group (Section 3.3.3). Those that also appeared in both groups were removed, leaving only FS-DMRs. Figure 3.2 illustrates this process.

In order to examine the most dramatic differences between both groups—i.e. regions that either completely lost or gained methylation in response to intervention—a similar approach was taken. Genomic regions that were found to be enriched pre-intervention while not being present in the post-intervention dataset for a particular individual were placed in an "ON/OFF" group. Those that did not appear in the pre-intervention dataset and found to be enriched post-intervention were placed in an "OFF/ON" group (Section 3.3.4). Further analysis was carried out as described above. Figure 3.3 illustrates this process

Methylation changes common in both the folic acid and the placebo groups were then implemented in a separate list of pregnancy related DMRs.

Focusing on the FS-DMRs obtained from these steps, two further stages of analysis were carried out, eliminating hits less significant than others. First, the levels of enrichment for every region were examined for each participant, and any outliers were removed – those that did not exhibit changes represented by the mean value, discussed in further detail in Section 3.3.6. Then, a thorough evaluation of the probe distribution within each peak was carried out, correlating each peak value used in the above analysis to CpG methylation at a much higher resolution (Section 3.3.7).

3.3.2: Comparison of Pre- versus Post-intervention for each Individual

From Nimblegen, a gene list with peak values representing levels of methylation was supplied. There was a cut-off of 2 for peak scores, so any gene that had less than a four-fold change in methylation against the input fraction has been omitted. It was necessary to write macros in Visual Basic 6 in order to sort through the data, separating genes with peak scores in both groups from those that only reached the cut-off in one. The aim of this step was to compare the peak scores of every genomic region pre- and post-intervention for each of the 6 FASSTT participants.

Excel data was arranged in a manner similar to Figure 3.4. Here, the rows were sorted based on their accession numbers and chromosomal regions so the loci line up as close as possible. In the first few rows, the data from "pre" (red) and "post" (blue) matched up evenly. However, a frame shift was introduced at B000007 in the "pre" set. To cut and paste an entry like this away from the dataset to be analysed later, two sub macros were written (full scripts and descriptions are included in Appendix A.1 and A.2). NuRedCut was written to remove a row of the "before" set, with NuBlueCut removing data from "after." The three lines of instructions on each dictate where Excel should copy, paste, and delete the necessary data.

As a result, the macro matches together all entries on the array that reach the threshold for enrichment before and after intervention. Those that only have an entry pre-intervention are put aside, as are ones that only have an entry post-intervention. These are brought back into the analysis in Section 3.3.4 for the ON/OFF lists. Using the fold enrichment observed in the pre-intervention group as a normaliser, fold changes in response to intervention were calculated for each locus.

3.3.3: FS-DMR Comparison between folic acid intervention and placebo groups

After the data was sorted by the looping macro described in Section 3.3.2, all of the genes that remained were arranged by their accession number for each sample (Figure 3.5), and the looping macro "RemoveRow" was written to remove every entry but those common across all three (Appendix A.3, Appendix A.4).

This approach was used to generate two separate lists of genes that exhibited an increase or a decrease in methylation in both the placebo and intervention groups. The submacros and looping macro "PlaceboCompare" (Appendices A.5 and A.6) were used to compare the two lists, removing entries that were common in both.

3.3.4: Obtaining lists for complete DNA methylation responses to folic acid supplementation

The final analysis for finding genes that completely gained or lost methylation in response to intervention was similar to that in Section 3.3.2. The entries initially removed by NuBlueCut and NuRedCut were selected because they did not appear in the opposing group. Instead of focusing on the matching rows that were left behind –as Section 3.3.2 did – the new lists generated by the cutting process were the focus for this part of the analysis. For the ON/OFF groups, genes present pre-intervention and not post-intervention in the folic acid supplementation group were compared to those of the placebo group. The selection criteria here was focused on genes that went from ON to OFF in the folic acid group, but remained ON pre- and post-intervention in placebo group. For the OFF/ON list, genes that went from unmethylated to methylated in the folic acid group and those that also stayed OFF both sides of the intervention in the placebo group were retained.

3.3.5: Pregnancy related Differential Methylated Regions

Finally, by applying these macros to the dataset without discriminating between the intervention and placebo groups, a list of genes that changed their methylation status in response to pregnancy alone was obtained.

3.3.6: In-depth analysis of FS-DMRs: Individual Enrichment Levels across the Subset

When comparing the enrichment scores for the initial FS-DMR list, an average value for the fold change across the three individual folic acid supplementation groups was used to find the most dramatic changes. However, with such a wide range of fold changes for each loci across the six participants, it was necessary to examine each individual fold change to determine if the mean change was being influenced by a single sample.

For loci found with positive enrichment, all three participants from the folic acid supplementation group needed to have a fold change greater than or equal to 1.20 to be considered for the final FS-DMR list, while all three in the placebo group needed to display either a decline or no change in methylation. Likewise, those in the negative enrichment group needed to be diminished by a factor of 0.8 or more, while having no change or an increase in all three placebo samples. Values between 0.8 and 1.2 enrichment were deemed to have no change. For example, for the DMR Solute carrier family 38, member 3, the mean enrichment across the three folic acid samples was 0.51. However, one sample was found to have an enrichment of 0.88, or “no change” under these criteria. As a result, this DMR would be

removed from the dataset. The limits of 0.8 and 1.2 are quite stringent, and led to the removal of many samples from the dataset, but these numbers were chosen to ensure that only the highest scoring FS-DMRs – those with the most dramatic change in response to intervention – would be brought forward for the gene-specific screening across the rest of the cohort (n=119).

For the ON/OFF samples, no genes were removed from this shorter list, as the selection criteria outlined in Section 3.3.4 had already taken individual scores into account.

3.3.7: In depth analysis of FS-DMRs: Probe distribution across remaining peaks

In calculating the initial peak values used in this analysis, groups of four or more probes reaching a significant threshold of enrichment on the microarray were combined and mapped back to a reference genome (Figure 3.6). One limit of this approach is that in many cases, the group of probes mapped back to a particular gene in one sample may not necessarily overlap with the group of probes mapped to the same gene in another sample i.e., the original software (DEVA) can call different sets of probes by the same gene name. By going back to the raw data, this next step aimed to eliminate any false positives found due to this limitation.

3.4: Results

3.4.1: MMSDK Pilot Study 1, Lambda Phage DNA

Two pilot studies were carried out to assess the feasibility of applying the MMSDK method to the precious FASSTT samples. First, an MMSDK library was generated from lambda phage DNA in order to test the efficiency of the process. Lambda genomic DNA contains 2 AscI sites, generating three fragments of 31853, 13127, and 3522 base pairs in length when cut with enzyme. The simplicity of the library that MMSDK generated from lambda DNA meant that direct Sanger sequencing could be used to determine the methylation status of the AscI sites. A PAGE gel (Figure 3.7) revealed that the PCR stage of the pilot study was successful, as indicated by a band at 88bp, representative of the desired MMSDK PCR product.

This band were cut from the gel (Section 3.2.5), cloned using the TA cloning system (Section 2.3.7), and sequenced using Sanger sequencing. Three MMSDK products were sequenced from this band. Each was found to have identical upstream and downstream P5 and P7 adapter sequences with different 16bp genomic sequences in between. Two of the three 16bp sequences were of reverse complement to one another, and both derived from the same AscI site in the genome. The third was matched back to the other AscI site, indicating that both

were unmethylated upon initial digestion (Figure 3.8). In Table 3.2, the MMSDK product starts at the highlighted portion, with blue representing the sequence of the P5 primer, green from the P7 primer, and yellow as the unique MMSDK read. Comparing this to the lambda genome, it was found that the first two MMSDK products were derived from the NlaIII site at approximately position 16400 of the genome, with the third from the NlaIII site at position 3500.

3.4.2: MMSDK Pilot Study 2, Human Genomic DNA

With proof that MMSDK can be used to obtain a full genomic library form a simple genome, our next goal was to demonstrate that this could be scaled up for human DNA. The human genome contains 4675 AscI sites, and will generate fragments of approximately 600,000bp after digestion[133]. The second pilot study aimed to generate MMSDK tags of human genomic DNA using the same process as the first pilot study.

Not only did amplification of human genomic DNA introduce a significant challenge based on its increased template size, but also for the processes downstream from amplification. The P5 and P7 sequences integrated into the MMSDK product are based on the adapters used for Illumina NGS. While using a conventional *Taq* polymerase to introduce A-overhangs to these products for TA-cloning was a valid option for Sanger Sequencing, the increased complexity of an MMSDK library derived from human genomic DNA required a high-fidelity polymerase for the PCR stage.

Velocity DNA Polymerase (Bioline) is capable of generating PCR products with a low number of cycles. Too many PCR cycles can cause long concatemers of DNA to form as multiple products are integrated into one another. This is phenomenon is evident on the PAGE gel in Figure 3.9. When the recommended concentration of 0.13µM P7 linker was used in the final MMSDK ligation reaction, no 88bp product was formed from the PCR step (lane 2), possibly from the linker forming self-dimers during ligation and preventing PCR amplification from taking place. When the ligase enzyme was omitted from this reaction (lane 4), a smear of non-specific amplification occurred, with separate DNA fragments containing sequences complementary to both primers now present in the reaction. When the concentration of P7 linker was reduced to 0.06 µM (lane 3, with ligase; lane 5, without ligase), concatemers over 100bp formed, along with more non-specific products.

All reactions on the PAGE gel from Figure 3.9 were carried out with 25 PCR cycles. Due to concatemer formation, the number of reactions carried out was decreased to 12, 15, 18 and 20

cycles under the same conditions (with 0.06µM P7 linker). This, however, resulted in no amplification (Figure 3.10).

After rigorous optimisation of the previous MMSDK steps and the final PCR steps, no amplification other than non-specific smears or high molecular weight concatemers could be obtained. A high-quality product is required for Illumina Sequencing, but this could not be obtained from MMSDK using human genomic DNA under the conditions described above. It was decided not to pursue this method on our FASSTT samples, given that the protocol could not be successfully executed on control human DNA.

3.4.3: Methylated DNA Immunoprecipitation, another method for genome-wide methylation Analysis

Returning to the literature, other methods for genome-wide methylation analysis were examined. Amongst these were promoter DNA microarrays by Roche Nimblegen, designed to analyse DNA fragments obtained from methylated DNA immunoprecipitation (MeDIP). Initially, there was concern that using a DNA microarray would limit the study and introduce bias. However, combined with the MeDIP technique, the Nimblegen Delux 2.1M promoter array can assess the DNA methylation status of up to 28,266 CpG islands in close proximity to gene and miRNA promoters[165]. In comparison, the enzyme AscI, when used in MMSDK, will only asses the methylation status of 4,675 sites; 3,028 of which lie in CpG islands[134]. Although MMSDK did have the advantage of finding completely new FS-DMRs outside of already annotated CGIs, MeDIP has a higher chance of linking changes in DNA methylation patterns to gene expression.

3.4.4: Quality Control of MeDIP DNA

To test the quality of the DNA obtained from the MeDIP process (Section 3.2.6), each of the twelve samples were run on a 1% agarose gel via electrophoresis. Each lane contained a smear of DNA running from 200bp to 1000bp in length, as expected from the initial MseI digestion. The difference in smearing patterns across each individual indicates that there is a variance in DNA quality across each sample. Although these also passed an independent QA step carried out by Nimblegen (based on similar criteria described here), special attention will be paid the sample “FASSTT 1 Methylated” and the data that arises from it (Figure 3.11).

Using qPCR and the assays described in Section 3.2.8, the extent of which the MeDIP process enriched for methylated DNA was determined to be 3.41 fold compared to the Input fraction.

For unmethylated DNA, the histone H3B assay exhibited a 25 fold diminishment for MeDIP versus Input (Figure 3.12). Raw data is included in Appendix B.

Once confirmed that enriched DNA was of adequate quality, the 12 MeDIP samples along with their corresponding untreated Input fractions were sent to Roche to be hybridised on a Nimblegen 2.1M Deluxe Promoter Array. Prior to labelling and hybridisation, Nimblegen carried out their own QC analysis on the samples and confirmed that all 12 DNA samples had concentrations between 250ng/µl and 1000ng/µl, with A260/A280 and A260/230 ratios greater than 1.7 and 1.6, respectively.

3.4.5: Novel FS-DMRs

Following the data analysis process described in Section 3.3, a total of 19 preliminary FS-DMRs have been identified, as depicted in Tables 3.3-3.5. None of these genomic regions have been implicated to change their methylation status in response to environmental factors before. Before finalising this list, however, it was necessary to narrow down the candidates down further, given the wide variety of variations introduced with these FASSTT samples (genetic, environmental, etc.). With such a large number of genes probed by the microarray to begin with, and DNA methylation itself capable of being affected by a wide range of factors — many of which the field hasn't fully elucidated — only the candidate regions that remained after the two-pronged approach described in Section 3.3.6 and Section 3.3.7 can be considered as novel folate sensitive-DMRs. The list of 19 FS-DMRS were reduced to 5 after taking into account their individual fold changes (Tables 3.5 – 3.7). In this analysis, individuals from the folic acid supplementation group (represented by the numbers (1085, 1156, and 1099) were compared to those of the placebo group (represented by the numbers 1061, 1010, and 1074).

However, on analysing the raw data for each of these 5 regions (Figures 3.13 – 3.17), none were found to significantly and consistently change their methylation status in response to folic acid supplementation when examined at this resolution. Each is discussed individually below. The main issue arose from peaks being compared between samples that did not overlap. All peak positions are listed in Table 3.8.

IP6K1, Figure 3.13: There is no single area covered by all twelve peaks in this locus. Thus, the changes in methylation observed here cannot be pinned down to a single probe. In samples 1085A and 1085B (four-digit numbers refer to the FASSTT participant, letters A and B refer to pre-and post-intervention samples, respectively) for example, the probes grouped together as a peak all lie in the same region, so their scores may be compared accurately. On the other hand, the regions compared between samples 1099A and 1099B do not overlap, though their

peaks were scored as 4.62 and 25.50, respectively. These figures are inflated due to the region for 1099B being further down the chromosome. At this point, the correlating region for 1099A actually has a similar level of enrichment, higher than what was called as a peak in the initial analysis. If this particular region was examined in both samples instead, IP6K1 may not have been found to be significant at all. The region 3:49824013-49824101 was chosen for further analysis in Chapter4.

Chr9ORF44, Figure 3.14: Here, there are only two samples, 1061A and 1061B, that lie outside the region Chr9:94904005-94904144 covered by every other peak. These two probes—marked as a full peak for sample 1099B—fail to hold the same trends as the other scored peaks for each sample. The goal of this phase of the project was to take the peak scores for each gene and narrow their corresponding methylation changes to a smaller group of CpGs. Given how the data has been presented, this approach would render Chr9ORF44 as a void entry. The region 9:94905691-94905940 was chosen for further analysis in Chapter 4.

RASA4, Figure 3.15: As a gene found in the ON/OFF list, RASA4 was not subjected to the same setbacks as the other potential DMRs. It is clear that the region Chr7:102156300-102156800 in samples 1099A and 1099B does not necessarily exhibit a complete loss of methylation, as suggested by no peak being scored for 1099B. The region 7:10215800-102158200 was chosen for further analysis in Chapter 4.

SGPL1, Figure 3.16: Following a similar trend to the others, there is no single probe that has a common trend across the dataset. Although 1085B was called for having no peaks, the enrichment levels of the probes at the region Chr10:72575627-72576074 are too similar to draw any further conclusions from.

Chr19ORF75, Figure 3.17: With only 3 out of the 12 samples tested showing a peak, finding a single region to represent the enrichment scores given to Chr19ORF75 proved more difficult than anticipated. No probe in the region Chr19:51767100-51769200 was found to reflect this relationship.

3.5: Discussion

The method initially selected for genome-wide methylation analysis was Modified Methylation Specific Digital Karyotyping (MMSDK) which has been successfully applied in a number of studies. Steenbergen *et al.,* (2013) used MSDK (the parent method of MMSDK) to identify 34 genes with increased methylation in response to human papillomavirus-induced

transformation[166]. Likewise, Li *et al.*,, (2013) found significant genome-wide changes using MMSDK on breast cancer cell line models resistant to the therapeutic agent tamoxifen[167].

However, after a series of pilot studies, MMSDK was not found to be reliable and consistent enough to use on the FASSTT samples. While MMSDK was successfully applied to a simple genome i.e., lambda, applying the method to human DNA did not yield the desired PCR product that would have been subject to NGS. MMSDK was considered for the initial genome-wide analysis for our samples due to it being an economical, reliable, and non-biased approach. Moreover, a related method had been implemented in the lab previously and thus, previous experience was considered an additional advantage for choosing this method. From the lambda DNA pilot study, it was found that MMSDK succeeded in generating sequencing tags for both of the possible sequencing sites in the genome, adjacent to the two AscI sequences in its genome. However, scaling this up to the size of the human genome proved problematic.

The technical challenges that MMSDK imposed meant that other methods needed to be considered. A perfect, gold standard method of DNA methylation analysis has yet to appear in the literature. The best method would be whole genome bisulfite sequencing (although with reduced complexity, reassembling the genome is proving challenging[168]), but the cost implications even at the beginning of this project prevented this as a viable option. As a result, many different approaches to genome-wide and gene-specific methylation analysis were examined, each with their own pros and cons.

It was at this point that MeDIP and DNA microarray analysis was found to be an alternative method to MMSDK given its technical challenges. Although the microarray analysis approach was considered and rejected during the early stages of the project due to the large amount of bias introduced by the array along with its inability to find completely novel sites or those outside CGIs, it was re-evaluated and found that it can potentially cover far more CGIs than MMSDK, although limited to promoter regions. This method has been successfully applied to several studies investigating DNA methylation changes arising from tumourigenisis[169–171], Alzheimer's disease[172], and across different species of primates[173,174]. Methylated DNA Immunoprecipitation (MeDIP) coupled with microarray analysis was subsequently applied to a subset of the FASSTT samples to generate a list of potential FS-DMRs.

In a literature review published by this research group[3], a list of genomic regions known to alter their methylation status in response to nutritional factors such as methyl-donor availability was put together to address our initial hypothesis: does folic acid supplementation

have an impact on DNA methylation in pregnant women. Because the supplementation strategy employed in the FASSTT study closely resembles the national medicinal guidelines[4], these samples provide a unique insight to what is happening on a molecular level to the DNA of women taking folic acid supplements. The data obtained here brings the field one step closer to linking folate availability to health and disease through epigenetic mechanisms.

This approach is not without its limits, however. Throughout the literature, methods to detect DNA methylation have been employed to learn more about tumourigeneis, disease states, and differences between related species, but rarely applied to healthy individuals[170,172,173]. MeDIP results found under these conditions tend to be far more dramatic, with larger gene lists obtained exhibiting more prominent changes in methylation levels[169,171,175].

When the initial list of 19 FS-DMRs was narrowed down based on the individual peak scores for each participant in the study, only 5 loci remained. On examining these regions further, comparing the enrichment values for the individual probes within each peak, it was clear that the analysis falls short in finding any novel FS-DMRs at this point in the study. Had the DNA methylation changes observed here been as significant as what has been reported in the literature, the seemingly contradictory nature of the peak-probe relationship would not have been so detrimental.

Most studies looking at DNA methylation using MeDIP take a qualitative approach; genomic regions are reported to be either hypomethylated or hypermethylated based on the microarray data[169,174]. In other words, all analysis in the literature either has taken the same approach as our ON/OFF and OFF/ON lists. From our own qualitative analysis, the number of FS-DMRs we identified is far lower than what has been found in other studies. Of the papers cited here, none have examined the peak-probe relationship at the same level of detail we have.

Throughout the literature, thousands of DMRs have been found by examining MeDIP data from experiments comparing disease states or different species. In most cases, a separate gene-specific method of analysis has been used to confirm these findings[169,171,173,174]. However, MeDIP results in some cases have been published without confirmation with a different method of analysis at all[170,172,175]. Nonetheless, further confirmation of our candidate FS-DMRs will be carried out in the next chapter, aiming to expand on our dataset and confirm that the methylation changes observed here are consistent with folic acid supplementation, as opposed to confirming what has been already found across the same samples already tested.

For example, Morris *et al.,* (2011)[171] combined a MeDIP and a whole-genome methylation array with high-density expression array analysis to find genes that were methylated and

transcriptionally silenced in renal cell carcinoma. For confirming methylation, both sodium bisulfite sequencing and combined bisulfite restriction analysis were carried out on the regions of interest, but the sample set was not expanded from n = 11. In this case, validation was carried out to confirm the findings of the MeDIP analysis, rather than investigating if these changes persist in other samples under the same conditions. The other approach was taken by Zhang *et al.*, (2013)[169], who discovered 2654 DMRs over 351 genes in 9 patients with tongue squamous cell carcinoma. Validation was carried out on an expanded sample set of 20 patients on three candidate genes found from the MeDIP analysis, with results consistent with the microarray study.

However, dramatic changes in DNA methylation status are known to occur during tumourigenesis. In contrast to the above studies, a similar magnitude of change was not observed for the FASSTT study which carried out over a relatively short period of time with just a nutritional intervention. Thus, confirming methylation changes at a more subtle level using this approach has not been reported in the literature.

Validation of these MeDIP results will follow the same course as Zhang *et al.*, (2013)[169]. The number of FASSTT samples examined for methylation changes will be expanded from 6 to 119 for gene-specific methylation analysis. Three of the final 5 FS-DMRs will be examined in this manner: *IP6K1*, Chr9ORF44 and *RASA4*.

Inositol hexakisphosphate kinase I (*IP6K1*) is a gene encoding for one of three kinase enzymes involved in synthesising inositol pyrophosphates, signalling molecules involved in multiple cellular processes such as chromatin remodelling and cell aging[176,177]. *IP6K1* itself has been implicated in a variety of roles associated with insulin sensitivity, homologous DNA recombination repair, and spermatogenesis[178–180]. More recently, IP6K1 has been found to mediate the assembly and disassembly of the CRL4-signalosome: a complex of ubiquitin ligases involved in cell cycle regulation and DNA repair[181]. To our knowledge, aberrant DNA methylation in the regulatory region of IP6K1 has not yet been reported in the literature.

The gene encoding Ras p21 protein activator 4 (*RASA4*) has been found to be deleted in primary effusion lymphoma cell lines, along with two well-established tumour suppressor genes[182]. The gene's product is involved in regulating Ras, a small GTP-binding signalling molecule involved in gene expression and cell cycle control[183]. Hypermethylation of RASA4 has been recently found to be a characteristic juvenile myelomonocytic leukemia, correlated with poor prognosis[184].

The approach taken by this chapter has been limited by the nature of the FASSTT study; healthy women divided into two groups were either given folic acid or a placebo over the course of several weeks. In comparison to reported DMRs arising from cancer or other disease states, it is no surprise that the changes in methylation observed so far have not been as significant. Unlike the other studies cited here, however, these experiments did not aim to find a list of FS-DMRs from MeDIP and microarray analysis alone. By concentrating on a subset of women from the FASSTT cohort who had the most pronounced response to folic acid supplementation, the work presented here has highlighted regions of the genome that may alter their methylation status in response to folic acid supplementation. Despite the negative finding of this microarray analysis, the highest scoring FS-DMRs will still be carried on to the gene-specific analysis over the entire FASSTT cohort in Chapter 4. With a more specific method of analysis, analysing DNA methylation patterns in an entirely different manner than MeDIP, these potential sites may still be confirmed as FS-DMRs.

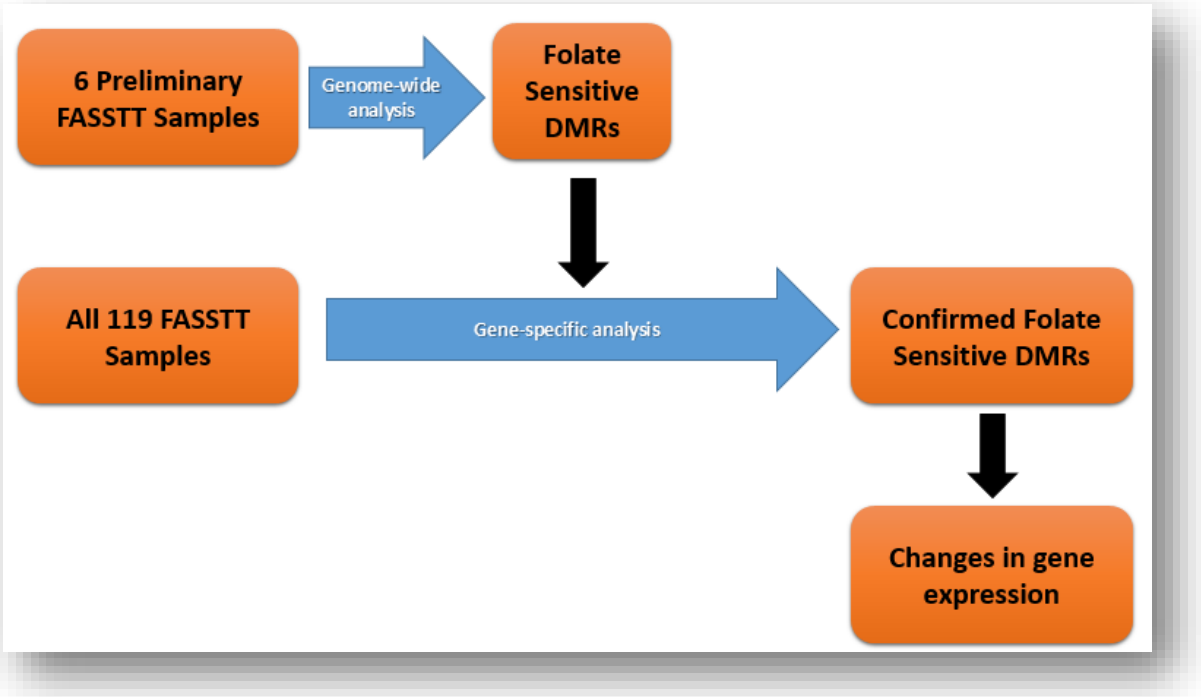


Figure 3.1: Project Overview

A genome-wide analysis will be carried out on a subset (n=6) of the FASSTT cohort to find potential folate sensitive differential methylation regions (FS-DMRs). These samples were selected from the full FASSTT cohort (n=114) based on the criteria listed in Table 3.1. Confirmation of these FS-DMRs across the rest of the FASSTT cohort (n=114) will be carried out using a gene-specific method of analysis (Chapter 4). Finally, changes in DNA expression patterns in response to these FS-DMRs will be analysed in a cell culture model (Chapter 4). In this Chapter, the method for genome-wide analysis will be selected based on evidence from a literature review and experimental data.

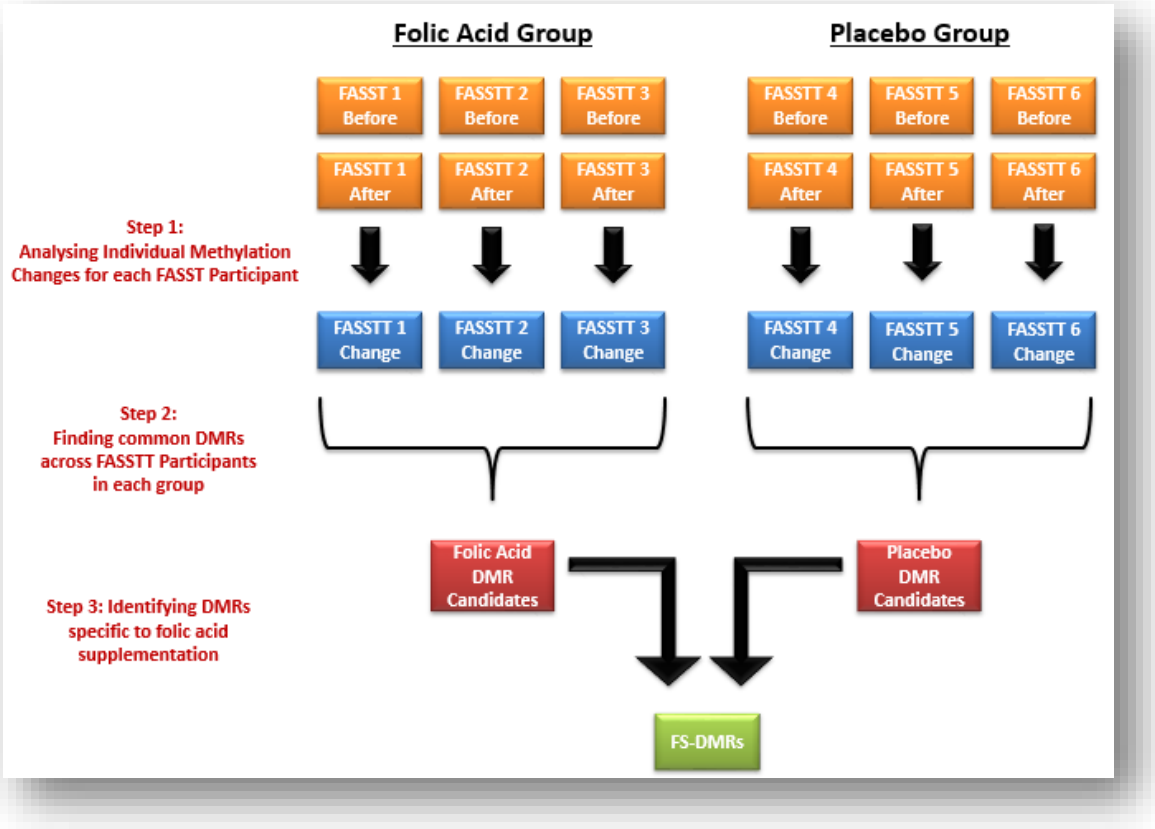


Figure 3.2: Analysis overview of Microarray Data: Quantitative Methylation Changes

The data analysis process was divided into three steps. Changes in enrichment levels of genes within the same individual (FASSTT 1, FASSTT 2, etc.) were first assessed (Step 1), generating a list of loci with changes in methylation calculated as a fold change of post-intervention against pre-intervention (FASSTT 1 Change, FASSTT 2 Change, etc.). Following this, the three participants from each group were compared to one another, and loci that were common amongst all three were examined further (Step 2). Finally, FS-DMRs were obtained by comparing those common in the folic acid group against those in the placebo group i.e., a region was identified as a FS-DMR if it only showed a change in the intervention group and not the placebo.

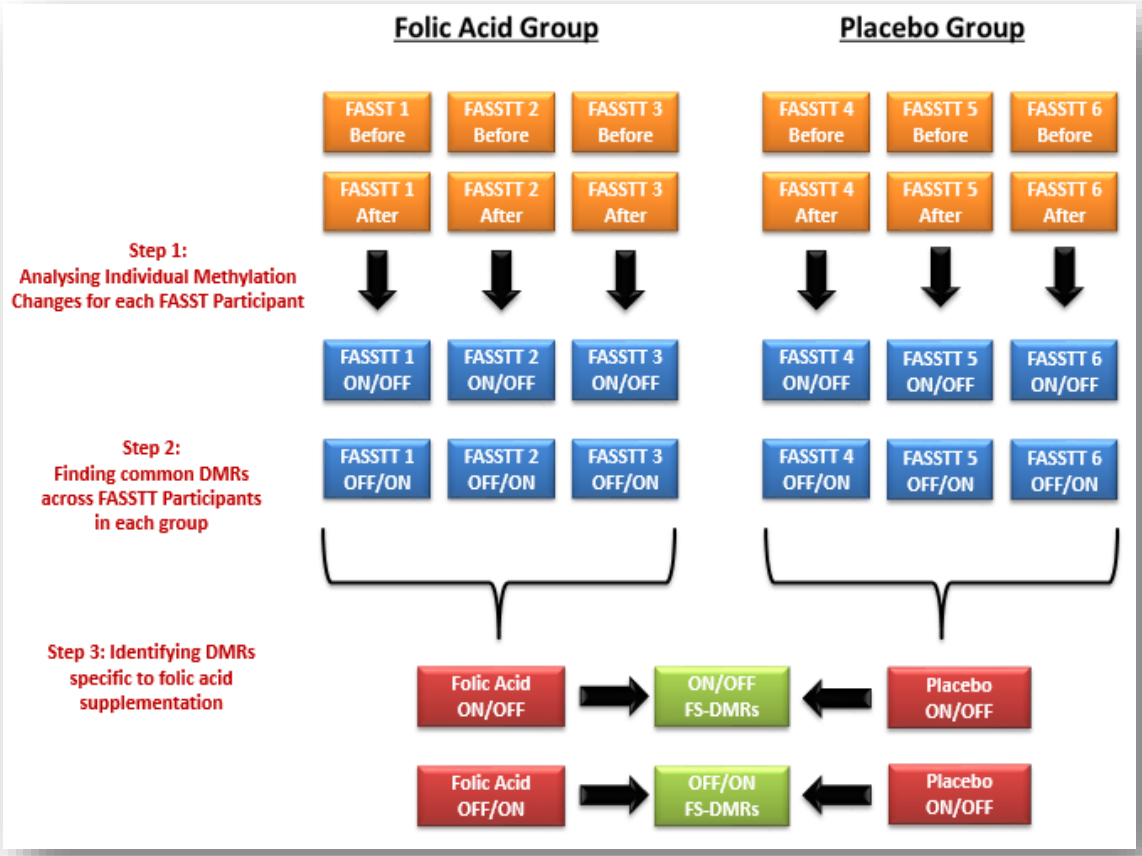


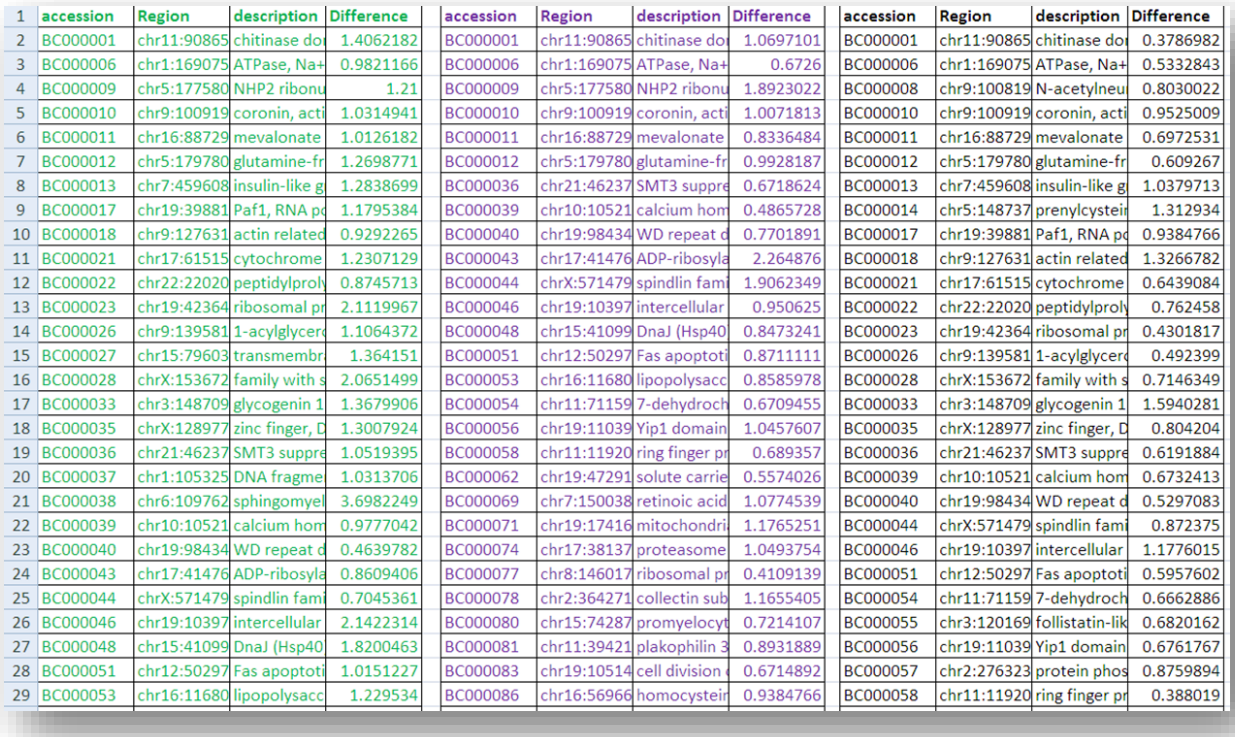
Figure 3.3: Analysis overview of Microarray Data: Complete Methylation Changes

The data analysis process for complete DNA Methylation changes was divided into three steps. Complete changes (from ON to OFF and from OFF to ON) in enrichment levels of genes within the same individual (FASSTT 1, FASSTT 2, etc.) were first assessed (Step 1), generating a list of loci that either completely lose (ON/OFF) or gain (OFF/ON) methylation in response to intervention. Following this, the three participants from each group were compared to one another, and loci that were common amongst all three were examined further (Step 2). Finally, FS-DMRs were obtained by comparing DMRs common in the folic acid group against those in the placebo group i.e., a region was identified as a FS-DMR if it only showed a change in the intervention group and not the placebo.

	Fold Change	Region	description	accession	accession	description	Region	Fold Change
2	14.2129	chr11:908659-908659	chitinase domain containing 1	BC000001	BC000001	chitinase domain containing 1	chr11:908659-908659	5.3824
3	9.2416	chr1:169075981-169075981	ATPase, Na+/K+ transporting, beta 1 polypeptide	BC000006	BC000006	ATPase, Na+/K+ transporting, beta 1 polypeptide	chr1:169075981-169075981	4.9284
4	9.61	chr5:176730813-176730813	PRELI domain containing 1	BC000007	BC000008	N-acetylneuraminic acid synthase	chr9:100819012-100819012	4.2849
5	5.3361	chr9:100819012-100819012	N-acetylneuraminic acid synthase	BC000008	BC000010	coronin, actin binding protein, 2A	chr9:100919944-100919944	4.1209
6	4.3264	chr9:100919944-100919944	coronin, actin binding protein, 2A	BC000010	BC000011	mevalonate (diphospho) decarboxylase	chr16:88729518-88729518	6.1504
7	8.8209	chr16:88729518-88729518	mevalonate (diphospho) decarboxylase	BC000011	BC000012	glutamine-fructose-6-phosphate transaminase 2	chr5:179780325-179780325	7.8961
8	12.96	chr5:179780325-179780325	glutamine-fructose-6-phosphate transaminase 2	BC000012	BC000013	insulin-like growth factor binding protein 3	chr7:45960837-45960837	10.5625
9	10.1761	chr7:45960837-45960837	insulin-like growth factor binding protein 3	BC000013	BC000014	prenylcysteine oxidase 1 like	chr5:148737633-148737633	10.89
10	5.7121	chr9:35079659-35079659	Fanconi anemia, complementation group G	BC000032	BC000036	SMT3 suppressor of mif two 3 homolog 3	chr21:46237969-46237969	5.76
11	6.7081	chr3:148709338-148709338	glycogenin 1	BC000033	BC000039	calcium homeostasis modulator 2	chr10:105212075-105212075	12.3904
12	9.61	chrX:128977435-128977435	zinc finger, DHHC-type containing 9	BC000035	BC000040	WD repeat domain 18	chr19:984342-984342	6.0516
13	9.3025	chr21:46237969-46237969	SMT3 suppressor of mif two 3 homolog 3	BC000036	BC000044	spindlin family, member 2B	chrX:57147924-57147924	13.5424
14	6.25	chr1:10532533-10532533	DNA fragmentation factor, alpha polypeptide	BC000037	BC000046	intercellular adhesion molecule 4	chr19:10397649-10397649	11.8336
15	17.9776	chr6:109762058-109762058	sphingomyelin phosphodiesterase 2,	BC000038	BC000051	Fas apoptotic inhibitory molecule 2	chr12:50297658-50297658	13.1044
16	18.4041	chr10:105212075-105212075	calcium homeostasis modulator 2	BC000039	BC000054	7-dehydrocholesterol reductase	chr11:71159402-71159402	7.3441
17	11.4244	chr19:984342-984342	WD repeat domain 18	BC000040	BC000055	follistatin-like 1	chr3:120169825-120169825	8.6436
18	9.7344	chr17:41476372-41476372	ADP-ribosylation factor-like 4D	BC000043	BC000056	Yip1 domain family, member 2	chr19:11039262-11039262	5.5696
19	15.5236	chrX:57147924-57147924	spindlin family, member 2B	BC000044	BC000057	protein phosphatase 1G magnesium-dependent	chr2:27632399-27632399	6.9169
20	10.0489	chr19:10397649-10397649	intercellular adhesion molecule 4	BC000046	BC000058	ring finger protein 26	chr11:119205440-119205440	6.8121
21	9.9856	chr15:41099662-41099662	DnaJ (Hsp40) homolog, subfamily C, member 17	BC000048	BC000062	solute carrier family 1, member 5	chr19:47291362-47291362	9.3025
22	5.9049	chr9:108456849-108456849	transmembrane protein 38B	BC000049	BC000069	retinoic acid receptor responder 2	chr7:150038694-150038694	8.1796
23	21.9961	chr12:50297658-50297658	Fas apoptotic inhibitory molecule 2	BC000051	BC000071	mitochondrial ribosomal protein L34	chr19:17416646-17416646	6.5025
24	8.4681	chr16:11680788-11680788	lipopolysaccharide-induced TNF factor	BC000053	BC000075	high-mobility group nucleosome binding domain 1	chr21:40720967-40720967	7.1824
25	11.0224	chr11:71159402-71159402	7-dehydrocholesterol reductase	BC000054	BC000076	cyclin D1	chr11:69455938-69455938	7.9524
26	12.6736	chr3:120169825-120169825	follistatin-like 1	BC000055	BC000077	ribosomal protein L8	chr8:146017766-146017766	11.8336
27	8.2369	chr19:11039262-11039262	Yip1 domain family, member 2	BC000056	BC000080	promyelocytic leukemia	chr15:74287092-74287092	5.4756
28	7.8961	chr2:27632399-27632399	protein phosphatase 1G, magnesium-dependent	BC000057	BC000081	plakophilin 3	chr11:394216-394216	8.7616
29	17.5561	chr11:119205440-119205440	ring finger protein 26	BC000058	BC000083	cell division cycle 37 homolog	chr19:10514195-10514195	5.3824
30	11.4244	chr19:47291362-47291362	solute carrier family 1, member 5	BC000062	BC000088	glutathione S-transferase mu 3	chr1:110282924-110282924	6.5536
31	16.81	chr7:150038694-150038694	retinoic acid receptor responder 2	BC000069	BC000091	ATG7 autophagy related 7 homolog	chr3:11314094-11314094	6.4009
32	7.9524	chr19:17416646-17416646	mitochondrial ribosomal protein L34	BC000071	BC000094	ribosomal protein S6 kinase, 70kDa, polypeptide 2	chr11:67195961-67195961	6.9696
33	10.1124	chr17:38137120-38137120	proteasome 26S subunit, non-ATPase, 3	BC000074	BC000095	N-acetylgalactosaminidase, alpha-	chr22:42466382-42466382	4.7961
34	7.0225	chr21:40720967-40720967	high-mobility group nucleosome binding domain 1	BC000075	BC000096	HtrA serine peptidase 2	chr2:74757105-74757105	9.1809
35	6.25	chr11:69455938-69455938	cyclin D1	BC000076	BC000097	transforming growth factor, beta-induced, 68kDa	chr5:135364690-135364690	7.3441
36	14.44	chr8:146017766-146017766	ribosomal protein L8	BC000077	BC000101	NIMA-related kinase 6	chr9:127019885-127019885	7.8961
37	11.4244	chr2:3642711-3642711	collectin sub-family member 11	BC000078	BC000103	NCK adaptor protein 2	chr2:106468319-106468319	8.6436
38	11.3589	chr15:74287092-74287092	promyelocytic leukemia	BC000080	BC000107	guanine nucleotide binding protein-like 2	chr1:38061522-38061522	6.4009
39	14.0625	chr11:394216-394216	plakophilin 3	BC000081	BC000114	megakaryocyte-associated tyrosine kinase	chr19:3786414-3786414	10.24
40	9.9856	chr19:10514195-10514195	cell division cycle 37 homolog	BC000083	BC000115	guanine nucleotide binding protein beta polypeptide 3	chr12:6950104-6950104	16.1604

Figure 3.4: Sample dataset for pre- and post-intervention comparison

Genes were sorted by their accession number and chromosomal location. Red represents the loci pre-intervention, and blue, post-intervention. Where the rows matched (rows 2 and 3), the "fold change" score from the post-intervention column was divided by that of the pre-intervention column to obtain a "fold difference" score. Whenever a frame-shift was introduced (like at BC000007, in row 4) the appropriate macro was called to remove and re-align the dataset. In this example, the looping macro would call a sub-macro to remove entry for BC000007 in red, aligning both red and blue entries for BC000008.



	accession	Region	description	Difference	accession	Region	description	Difference	accession	Region	description	Difference
2	BC000001	chr11:90865	chitinase do	1.4062182	BC000001	chr11:90865	chitinase do	1.0697101	BC000001	chr11:90865	chitinase do	0.3786982
3	BC000006	chr1:169075	ATPase, Na+	0.9821166	BC000006	chr1:169075	ATPase, Na+	0.6726	BC000006	chr1:169075	ATPase, Na+	0.5332843
4	BC000009	chr5:177580	NHP2 ribonu	1.21	BC000009	chr5:177580	NHP2 ribonu	1.8923022	BC000008	chr9:100819	N-acetylneu	0.8030022
5	BC000010	chr9:100919	coronin, acti	1.0314941	BC000010	chr9:100919	coronin, acti	1.0071813	BC000010	chr9:100919	coronin, acti	0.9525009
6	BC000011	chr16:88729	mevalonate	1.0126182	BC000011	chr16:88729	mevalonate	0.8336484	BC000011	chr16:88729	mevalonate	0.6972531
7	BC000012	chr5:179780	glutamine-fr	1.2698771	BC000012	chr5:179780	glutamine-fr	0.9928187	BC000012	chr5:179780	glutamine-fr	0.609267
8	BC000013	chr7:459608	insulin-like g	1.2838699	BC000036	chr21:46237	SMT3 suppre	0.6718624	BC000013	chr7:459608	insulin-like g	1.0379713
9	BC000017	chr19:39881	Paf1, RNA pc	1.1795384	BC000039	chr10:10521	calcium hom	0.4865728	BC000014	chr5:148737	prenylcystein	1.312934
10	BC000018	chr9:127631	actin related	0.9292265	BC000040	chr19:98434	WD repeat d	0.7701891	BC000017	chr19:39881	Paf1, RNA pc	0.9384766
11	BC000021	chr17:61515	cytochrome	1.2307129	BC000043	chr17:41476	ADP-ribosyla	2.264876	BC000018	chr9:127631	actin related	1.3266782
12	BC000022	chr22:22020	peptidylproly	0.8745713	BC000044	chrX:571479	spindlin fami	1.9062349	BC000021	chr17:61515	cytochrome	0.6439084
13	BC000023	chr19:42364	ribosomal pr	2.1119967	BC000046	chr19:10397	intercellular	0.950625	BC000022	chr22:22020	peptidylproly	0.762458
14	BC000026	chr9:139581	1-acylglycero	1.1064372	BC000048	chr15:41099	DnaJ (Hsp40	0.8473241	BC000023	chr19:42364	ribosomal pr	0.4301817
15	BC000027	chr15:79603	transmembr	1.364151	BC000051	chr12:50297	Fas apoptoti	0.8711111	BC000026	chr9:139581	1-acylglycero	0.492399
16	BC000028	chrX:153672	family with s	2.0651499	BC000053	chr16:11680	lipopolysacc	0.8585978	BC000028	chrX:153672	family with s	0.7146349
17	BC000033	chr3:148709	glycogenin 1	1.3679906	BC000054	chr11:71159	7-dehydroch	0.6709455	BC000033	chr3:148709	glycogenin 1	1.5940281
18	BC000035	chrX:128977	zinc finger, D	1.3007924	BC000056	chr19:11039	Yip1 domain	1.0457607	BC000035	chrX:128977	zinc finger, D	0.804204
19	BC000036	chr21:46237	SMT3 suppre	1.0519395	BC000058	chr11:11920	ring finger pr	0.689357	BC000036	chr21:46237	SMT3 suppre	0.6191884
20	BC000037	chr1:105325	DNA fragme	1.0313706	BC000062	chr19:47291	solute carrie	0.5574026	BC000039	chr10:10521	calcium hom	0.6732413
21	BC000038	chr6:109762	sphingomyel	3.6982249	BC000069	chr7:150038	retinoic acid	1.0774539	BC000040	chr19:98434	WD repeat d	0.5297083
22	BC000039	chr10:10521	calcium hom	0.9777042	BC000071	chr19:17416	mitochondri	1.1765251	BC000044	chrX:571479	spindlin fami	0.872375
23	BC000040	chr19:98434	WD repeat d	0.4639782	BC000074	chr17:38137	proteasome	1.0493754	BC000046	chr19:10397	intercellular	1.1776015
24	BC000043	chr17:41476	ADP-ribosyla	0.8609406	BC000077	chr8:146017	ribosomal pr	0.4109139	BC000051	chr12:50297	Fas apoptoti	0.5957602
25	BC000044	chrX:571479	spindlin fami	0.7045361	BC000078	chr2:364271	collectin sub	1.1655405	BC000054	chr11:71159	7-dehydroch	0.6662886
26	BC000046	chr19:10397	intercellular	2.1422314	BC000080	chr15:74287	promyelocyt	0.7214107	BC000055	chr3:120169	follistatin-lik	0.6820162
27	BC000048	chr15:41099	DnaJ (Hsp40	1.8200463	BC000081	chr11:39421	plakophilin 3	0.8931889	BC000056	chr19:11039	Yip1 domain	0.6761767
28	BC000051	chr12:50297	Fas apoptoti	1.0151227	BC000083	chr19:10514	cell division	0.6714892	BC000057	chr2:276323	protein phos	0.8759894
29	BC000053	chr16:11680	lipopolysacc	1.229534	BC000086	chr16:56966	homocystein	0.9384766	BC000058	chr11:11920	ring finger pr	0.388019

Figure 3.5: Sample dataset for comparison between three participants

Along the top row, the macro described in Section 3.3.2 was designed to delete any four-column rows in the green, purple, or black sections that did not have a match (each colour representing a different FASSTT participant). Those that did were removed and pasted elsewhere for further analysis. In cases were the three samples did not match up (like in row 8), the appropriate submacro was called to adjust the dataset accordingly, creating more three-way matches for further analysis.

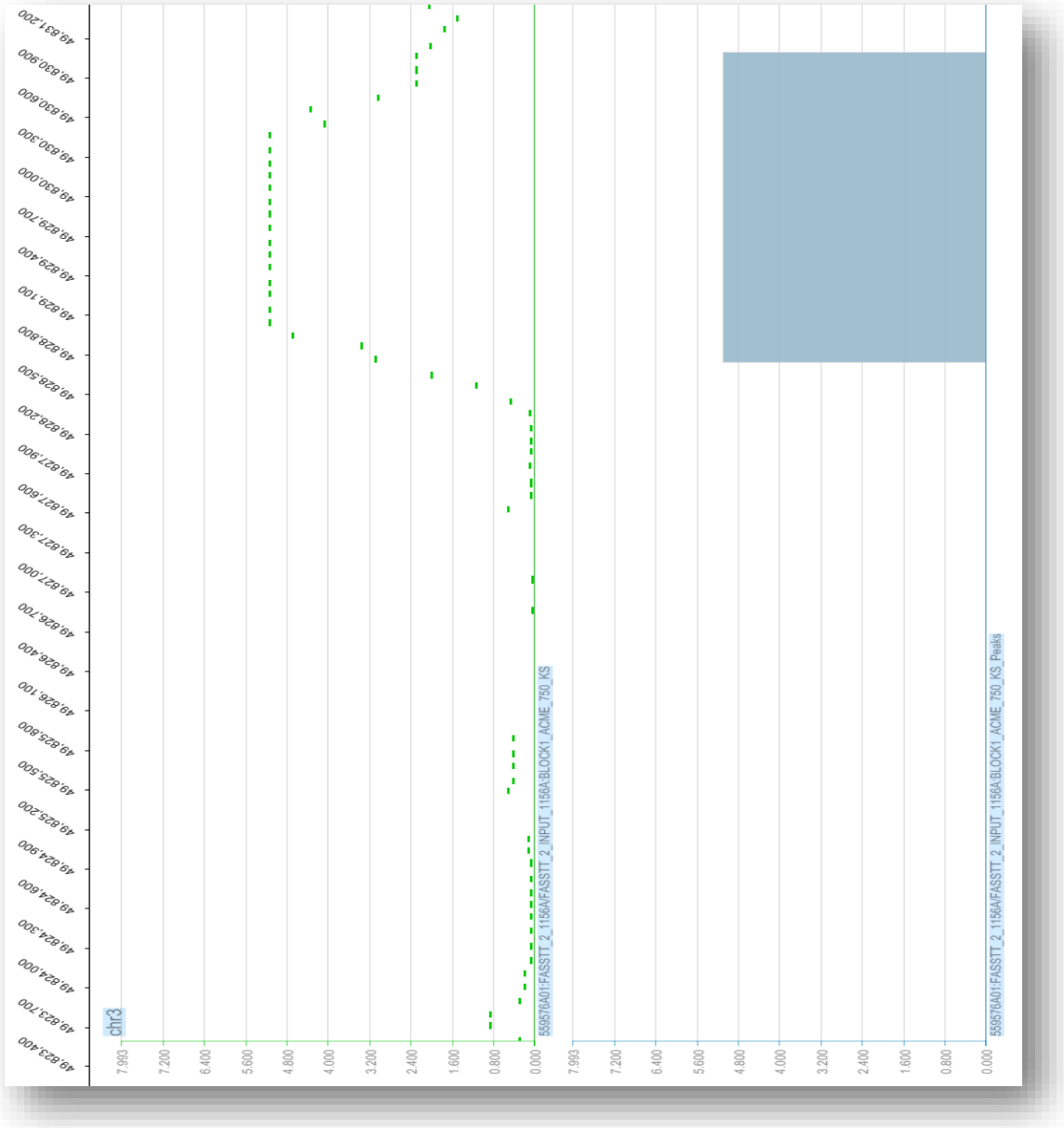


Figure 3.6: Sample illustration between probe and peak enrichment scores

Each 100bp probe is represented by a green line, with enrichment scores on the Y axis, and chromosomal position on the X axis. Adjacent probes that reach a significant threshold are grouped together and identified as a "peak" (blue rectangle), with the "peak score" given as the average enrichment value for the probes. Peak scores were converted to fold change differences for the initial analysis on the dataset to generate an initial list of FS-DMRs. In the process of narrowing these down to the final list, raw probe enrichment values making up each peak were re-examined to ensure that the relationship between a participants peak score pre- and post-intervention reflected that of their probe scores.

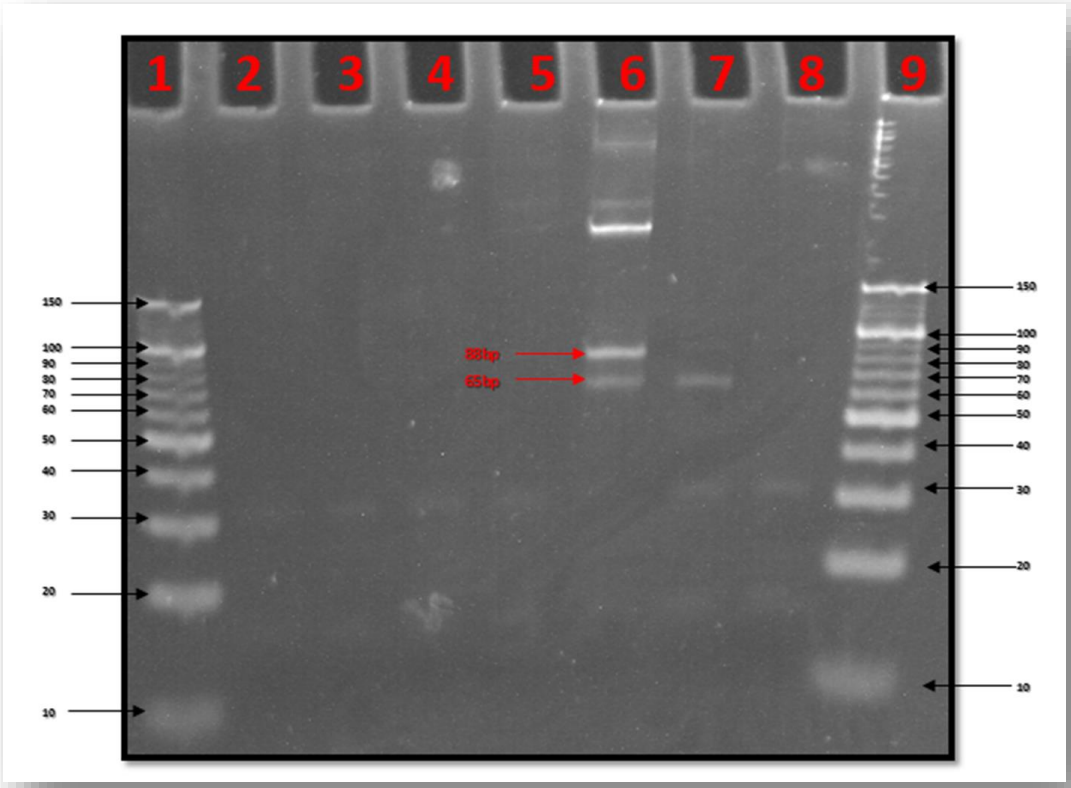


Figure 3.7: MMSDK Pilot Test on Lambda DNA: PCR amplification step

After the PCR step of MMSDK, the final product was visualised on a 12% polyacrylamide gel. Lane 1 contains a 10bp O'Range Ruler (Thermo Scientific). Lanes 2-6 contain PCR products after 8, 12, 16, 20 and 25 cycles, respectively. Lane 7 contains an MMSDK PCR product after 25 cycles, but with the ligase enzyme from the final ligation step omitted to control for amplification of templates without P7 linkers. Lane 8 contains a no template negative control. Lanes 9 contains another 10bp O'Range Ruler. In lane 6, two bands at ~88bp and ~65bp appear together. The former is equal to the expected size of the MMSDK product, and was cut out from the gel. The latter is an amplified product of DNA unligated to the P7 linker, equal to the only band to appear in lane 7, i.e., the non-desirable PCR product. The largest band here, at over 100bp in length, is possible the result of concatemer formation, resulting from extensive amplification of a short product.

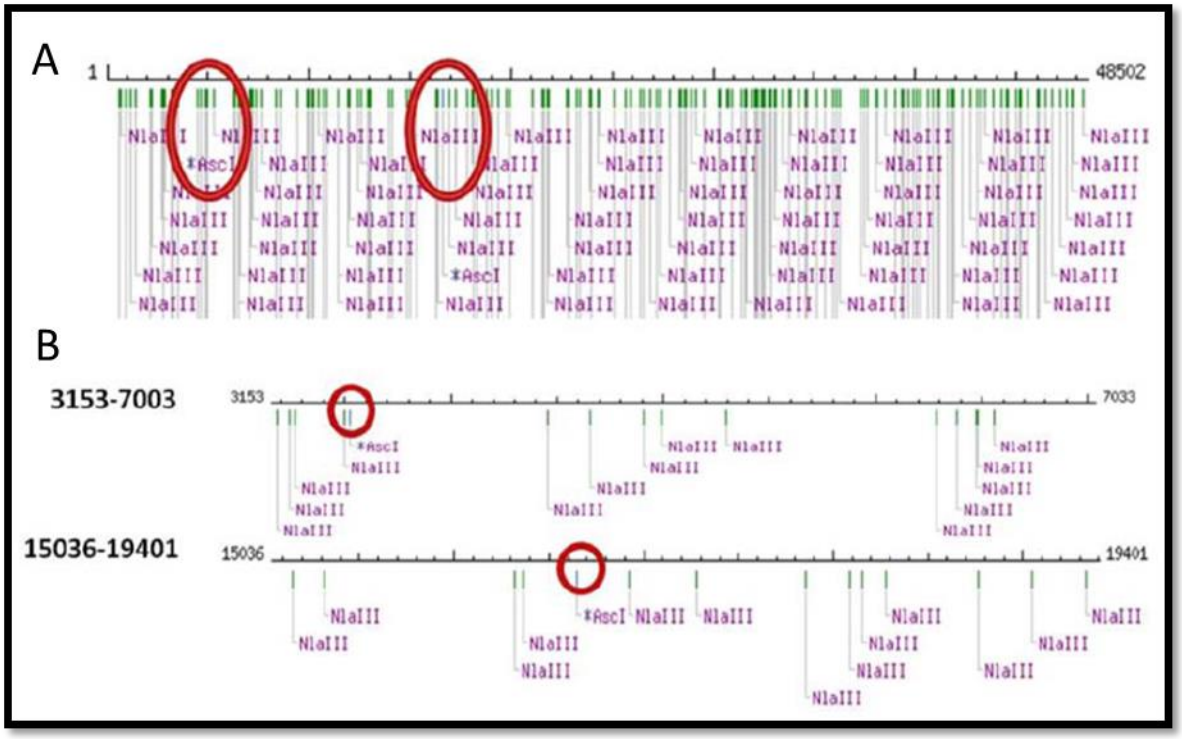


Figure 3.8: Restriction sequences of the Lambda Genome.

MMSDK is designed to target the NlaIII sites that flank the AscI sites of the input DNA. Image A depicts a map of the entire lambda genome, with two regions circled in red. Image B shows these two regions in more detail. The four-base-pair recognition sequence of NlaIII is far more numerous than the eight-base-pair recognition sequence of AscI. With only two AscI sites, marked here, MMSDK can only generate four possible tags from lambda genomic DNA, in comparison to approximately 9,000 of that in human DNA. It was the simplicity of lambda DNA resulted in the success of the first pilot study.

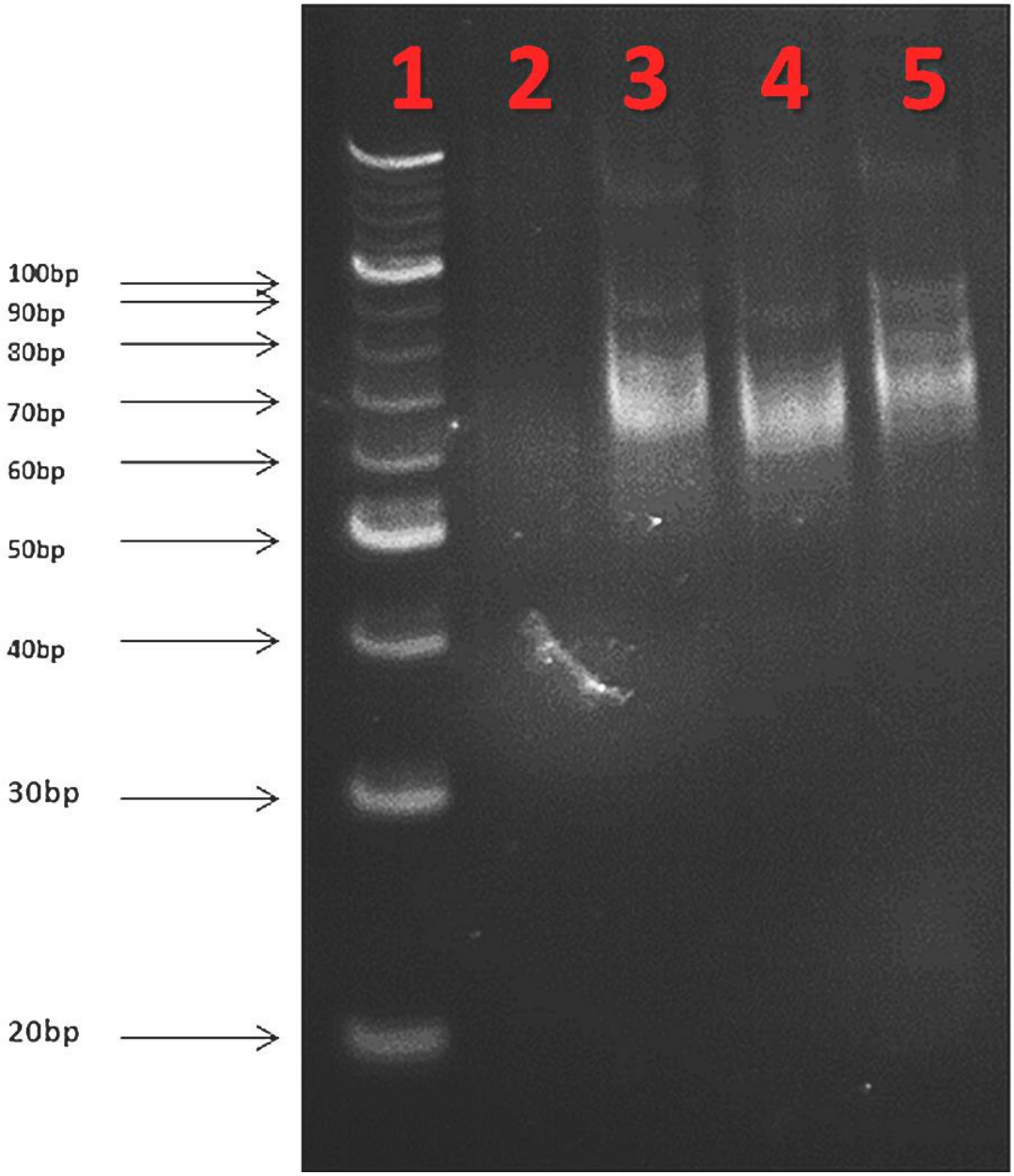


Figure 3.9: MMSDK PCR Products on Human Genomic DNA

Products of MMSDK PCR carried out on human genomic DNA. Two concentrations of P7 linker were used in the final ligation step, (0.13µM, lanes 2 and 3; 0.06µM, lanes 4 and 5), with no ligase negative controls for both (lane 3 and lane 5). When ligase was included in the reactions, this PCR only produced products for 0.06µM P7, with a very weak band visible at ~88bp. Omitting the ligase enzyme generated more non-specific products, including high molecular-weight concatemers greater than 100bp in length.

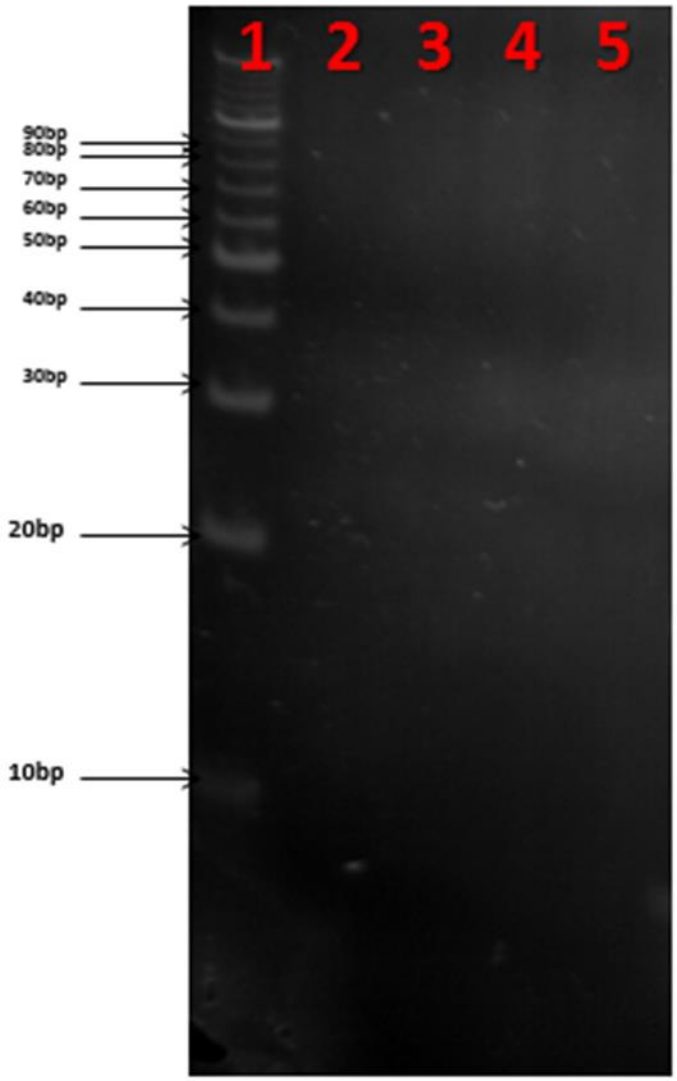


Figure 3.10: MMSDK PCR Products on Human Genomic DNA

With a P7 linker concentration of 0.06µM, the cycle number of the final MMSDK PCR was reduced to 12 (lane 2), 15 (lane 3), 18 (lane 4) and 20 (lane 5). No amplification occurred in either reaction.

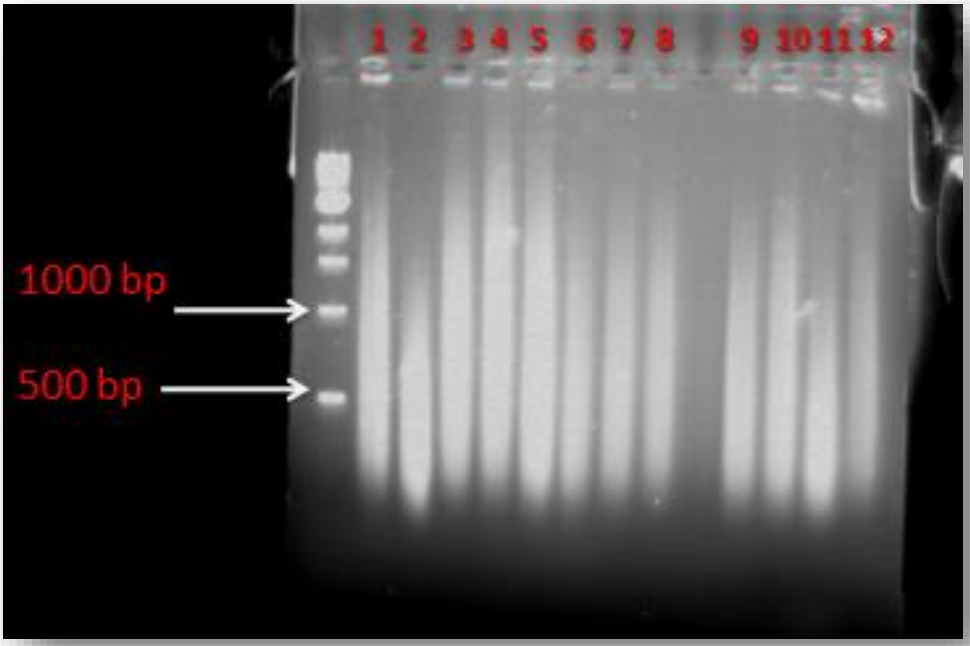


Figure 3.11: Quality Control analysis of MeDIP DNA, gel electrophoresis

All 12 MeDIP samples were run on a 1% agarose gel for 45 minutes alongside a 1kb ladder. A smear for each sample between 200bp and 2000bp in length represents DNA digested with the MseI enzyme. The MeDIP process yielded a Methylated and Input Fraction of the 6 FASSTT participants, each of which were loaded in numerical order on the gel, from left to right (Lane 1, FASSTT 1 Input; Lane 2, FASSTT 2 Methylated; Lane 3, FASSTT 2 Input, etc.)

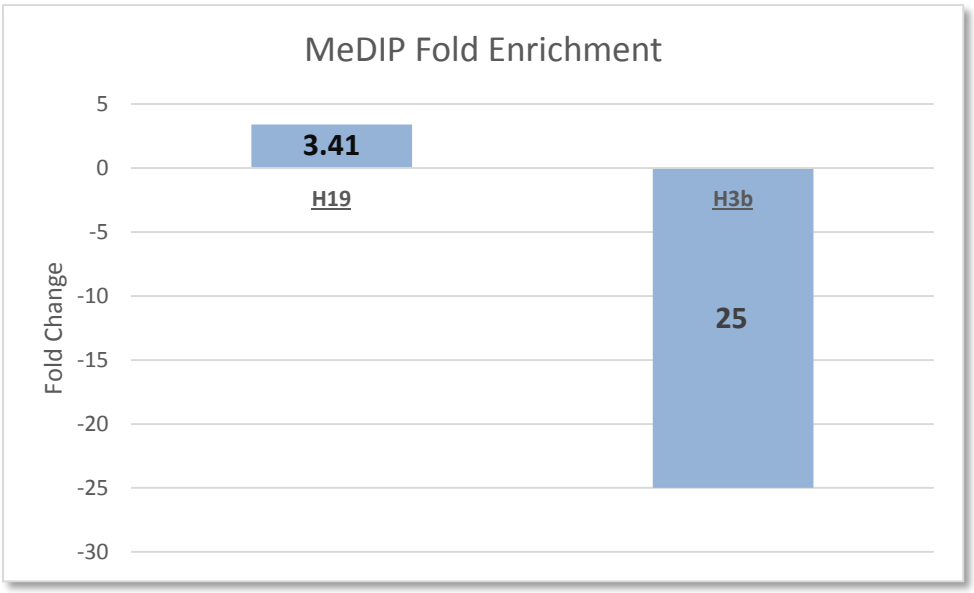


Figure 3.12: Quality Control analysis of Immunoprecipitated DNA by qPCR

Using qPCR, the methylated lncRNA H19 gene was found to be enriched 3.41 fold, while the unmethylated histone gene H3b decreased by 25 fold during the enrichment process. The mean C_P values over all FASSTT samples for each assay were compared with that of the Input samples using the formula: Fold Change = 2 ^ (Input – FASSTT).

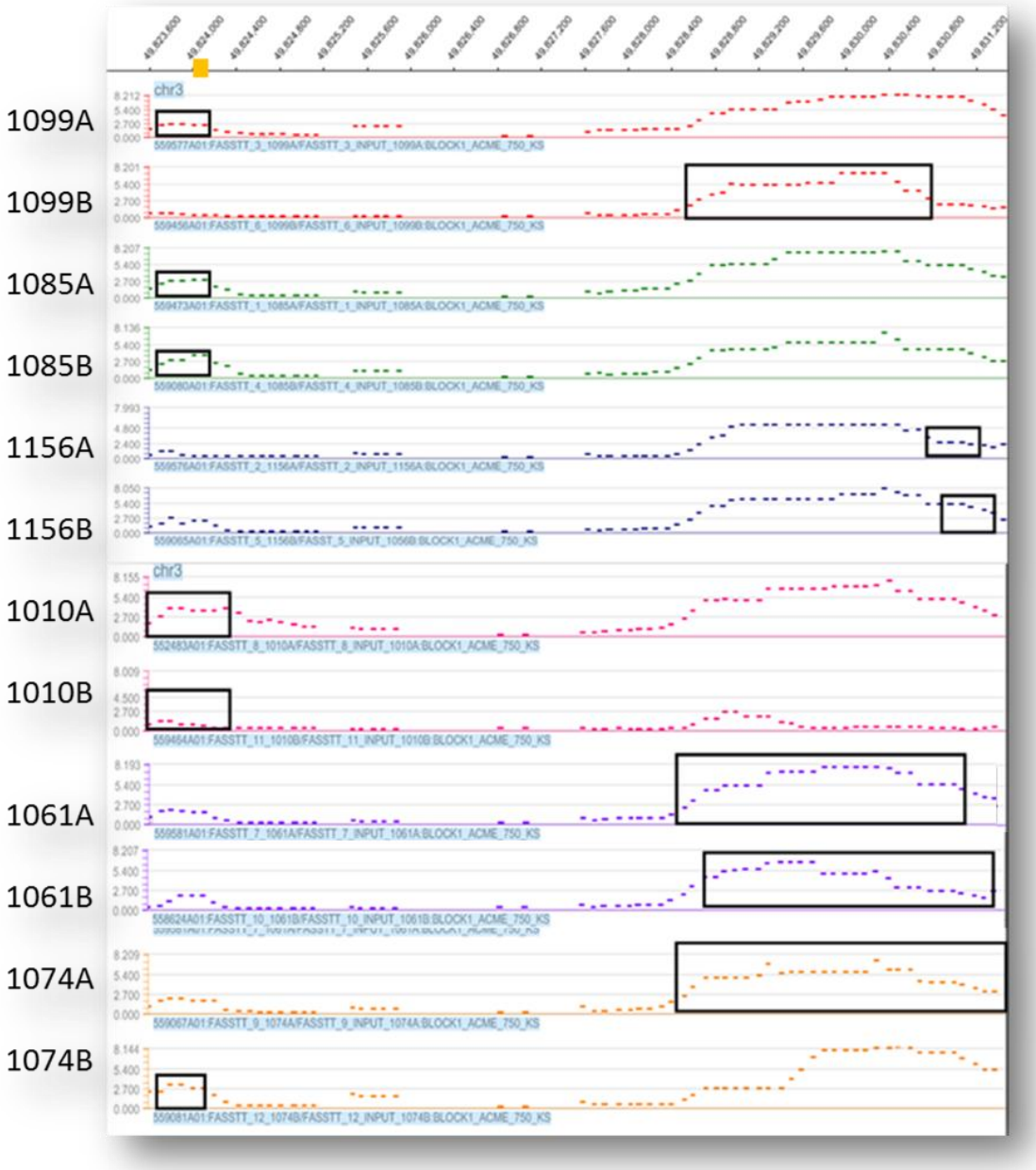


Figure 3.13: IP6K1 MeDIP probe positions

Complete peak overlap is evident in sample 1085A and 1085B (green). Although the peak scores for 1099A and 1099B (red) exhibited a significant increase after intervention, there is little difference between the actual probe enrichment values. In fact, a decrease in enrichment is evident at the region marked by the peak in 1099A. The region 3:49824013-49824101 (orange square) was chosen for further analysis in Chapter 4, as this site has the most overlap in probes that represent the trend originally depicted by the peaks.

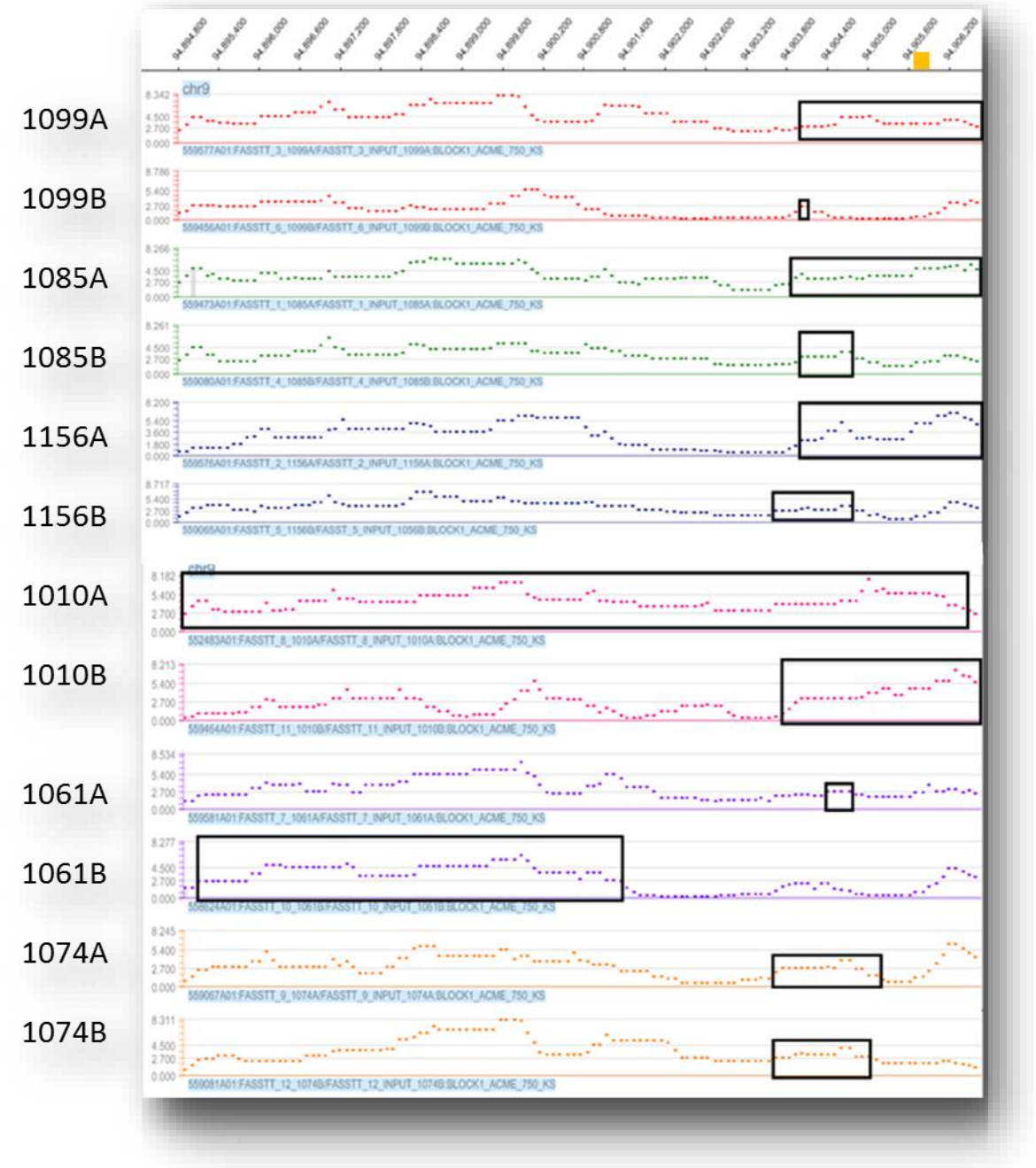


Figure 3.14: Chr9ORF44 MeDIP probe positions

Only the probes for the peak of sample 1099B (red) are present in 10 out of the 12 samples here. At this region, however, the trends of enrichment do not follow that of the full peak scores allocated to the samples in the first stage of analysis. The region 9:94905691-94905940 (orange square) was chosen for further analysis in Chapter 4, as the single probes there exhibit the decline in enrichment observed in the initial analysis.

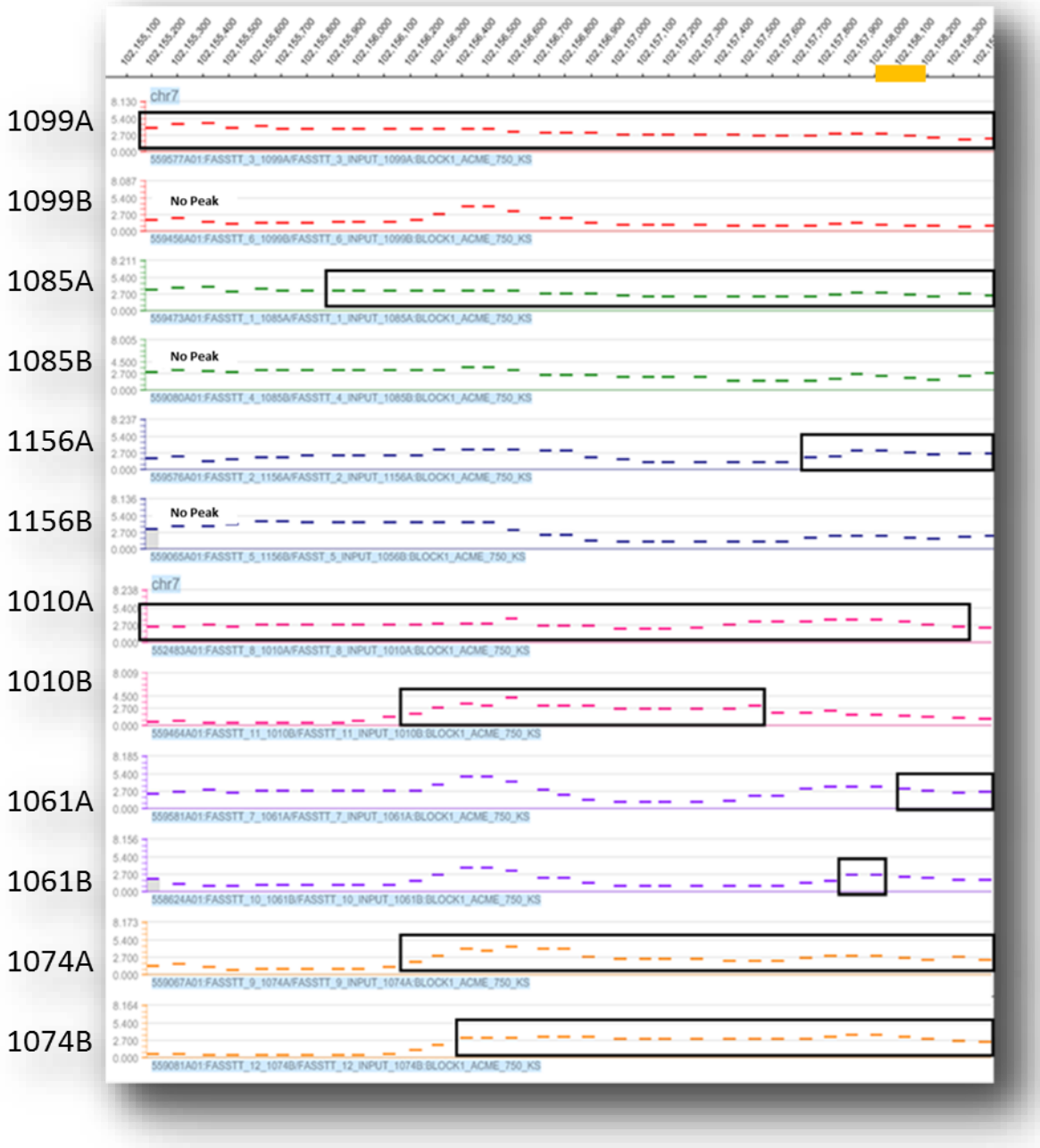


Figure 3.15: RASA4 MeDIP Probe positions

As a gene from the ON/OFF list, *RASA4* does not suffer from the same setbacks as IP6K1 and Chr9ORF44 did. The region 7:10215800-102158200 (orange rectangle) was chosen for further analysis in Chapter 4, as the peaks at this region reflect the ON/OFF trend observed in the initial peak values.

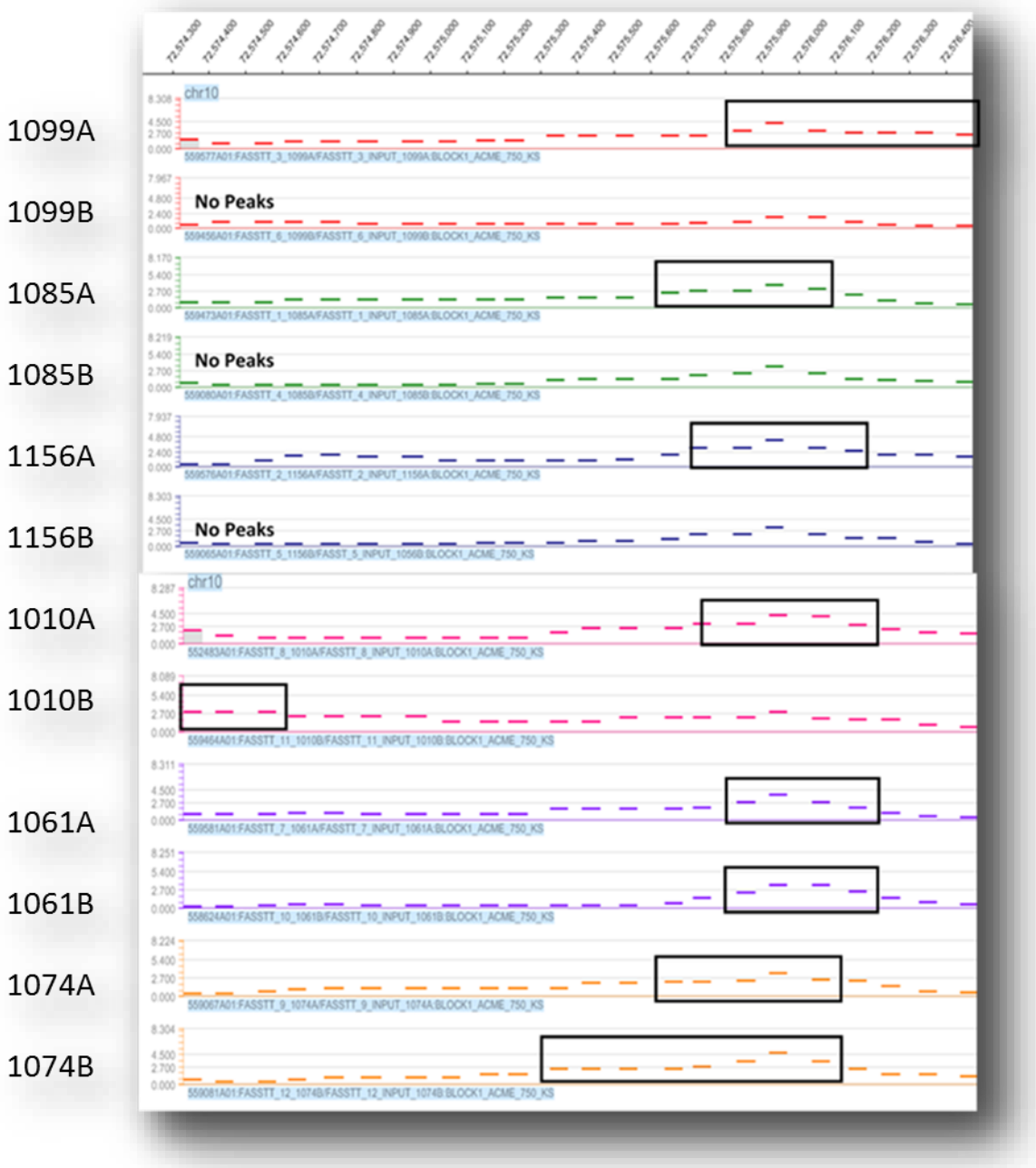


Figure 3.16: SGPL1 MeDIP probe positions

Similarly to RASA4, SGPL1's peak-probe distributions mirrored what was depicted by the peak score values used in the initial analysis. However, region at 1085B (green) expected to have no peak had enrichment at the region scored as a peak for 1085A. As a result, no conclusions from SGPL1 can be drawn, and no further analysis was not carried out on this locus.

Figure 3.17: Chr19ORF75 MeDIP probe positions

It appears that there was not a significant OFF/ON relationship between the probe values here, despite what the peak data had shown in the initial analysis. Although many of the samples here were scored with no peaks, there is a significant level of enrichment towards the right-hand side of the region. Further analysis was not carried out on Chr19ORF75.

Table 3.1: Criteria for selecting FASSTT samples for MeDIP analysis.

Folic Acid Intervention group	**Placebo Group**
Participants taking folic acid supplementation for less than 6 weeks before recruitment	Participants taking folic acid supplementation for less than 6 weeks before recruitment
Participants with lowest red cell folate at baseline	Participants with lowest red cell folate at baseline
Participants with highest red cell folate and serum folate response	Participants with typical decline of red cell folate and serum folate response
Participants with matching cord blood red cell folate and serum folate levels	Participants with matching cord blood red cell folate and serum folate levels
Participants age-matched to Placebo group	Participants age-matched to Folic Acid Supplemented group

Table 3.2: Sequences of Sanger reads obtained for MMSDK.

MMSDK Lambda Product 1	NNNNNNNNAGGGCGANTGGGCCCTCTAGATGCATGCTCGAGCGGCCGCCA GTGTGATGGATATCTGCAGAATTCGGCTT**AATGATACGGCGACCACCGAC AGGTTCAGAGTTCTACAGTCCGACCATGGGGAAACGTCTTGGTGCTCG TATGCCGTCTTCTGCTTG**AAGCCGAATTCCAGCACACTGGCGGCCGTTAC TAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTGT TTCCTGTGTGAAATT…
MMSDK Lambda Product 2	NNNNNNNNNNNNGGCGATTGGGCCCTCTAGATGCATGCTCGAGCGGCCGC CAGTGTGATGGATATCTGCAGAATTCGGCTT**CAAGCAGAAGACGGCATAC GAGCACCAAGACGTTTCCCCATGGTCGGACTGTAGAACTCTGAACCTG TCGGTGGTCGCCGTATCATT**AAGCCGAATTCCAGCACACTGGCGGCCGTT ACTAGTGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCT GTTTCCTGTGTGAAATT…
MMSDK Product 3	NNNNNNNNNGGCGATTGGGCCNCTAGATGCATGCTCGAGCGGCCGCCAGT GTGATGGATATCTGCAGAATTCGGCTT**AATGATACGGCGACCACCGACA GGTTCAGAGTTCTACAGTCCGACCATGCTGGCAATATGCGGGATCGTA TGCCGTCTTCTGCTTG**AAGCCGAATTCCAGACACTGGCGGCCGTTACTAG TGGATCCGAGCTCGGTACCAAGCTTGGCGTAATCATGGTCATAGCTGTTTC CTGTGTGAAATT…

Blue, P5 primer sequence; Yellow, Unique MMSDK read from genomic DNA; Green, P7 primer sequence

Table 3.3: Genes that exhibited changes in methylation at their promoter regions in all individuals tested

	Folic Acid Mean[1]	Folic Acid Standard[2]	Placebo Mean[1]	Placebo Standard[2]
Inositol hexakisphosphate kinase 1 [*NM_001006115*]	2.79	2.3637	0.48	0.1347
Macrophage stimulating 1 [*BC048330*]	3.04	1.2897	0.98	0.6907
Lipocalin 15 [*NM_203347*]	2.02	0.9968	1.01	0.3202
Chromosome 22 open reading frame 34 [*NR_026997*]	2.29	1.2985	1.06	0.2880
Lymphocyte-specific protein 1 pseudogene [*NR_027238*]	2.15	1.3780	1.12	0.2588
Ubiquitin-conjugating enzyme E2M pseudogene 1 [*NR_002837*]	2.02	0.6113	1.00	0.2219
Solute carrier family 38, member 3 [*NM_006841*]	0.51	0.3293	0.78	0.0945
Chromosome 9 open reading frame 44 [*NR_027341*]	0.49	0.0579	1.50	0.7752
Amiloride binding protein 1 [*BC014093*]	0.47	0.0531	1.33	0.6466
Ciliary rootlet coiled-coil, rootletin-like 1 [*NR_026752*]	0.44	0.1381	1.17	0.8183
Death-domain associated protein [*BC109073*]	0.60	0.2900	1.76	1.1806
G protein pathway suppressor 2 [*BC103901*]	0.45	0.1488	0.98	0.3883
Melanoma antigen family A, 4 *[NM_001011549*]	0.59	0.0915	1.05	0.2911
Surfeit 1 [*BC028314*]	0.46	0.1159	0.91	0.2550
Transmembrane protein 145 [*NM_173633*]	0.49	0.1127	0.90	0.3264
Hypothetical LOC149134 [*NR_015422*]	0.50	0.2944	0.87	0.2287

1: average fold change across 3 samples in either group; 2: standard deviation of fold changes across 3 samples in either group.

Green: Increased enrichment. Red: decreased enrichment.

Table 3.4: Genes that changed their methylation status from ON to OFF across all three individuals

	Folic Acid Mean[1]	Folic Acid Standard[2]	Placebo Mean[1]	Placebo Standard[2]
RAS p21 protein activator 4 [*NM_001079877*]	7.27*	1.486	0.96	0.2648
sphingosine-1-phosphate lyase 1 [*BC052991*]	7.17*	0.517	1.10	0.1601

1: average fold change across 3 samples in either group; 2: standard deviation of fold changes across 3 samples in either group.

***** *mean value taken from "before" samples only, as no peak values was scored for the "after" group*

Table 3.5: Genes that changed their methylation status from OFF to ON across all three individuals

	Folic Acid Mean[1]	Folic Acid Standard[2]	Placebo Mean[1]	Placebo Standard[2]
Chromosome 19 open reading frame 75 [*NM_001079877*]	9.97*	1.2433	0.00	0.00

1: average fold change across 3 samples in either group; 2: standard deviation of fold changes across 3 samples in either group.

***mean value taken from "after" samples only, as no peak values was scored for the "before" group. Also, no peaks were scored in any placebo samples.*

Table 3.6: Narrowing down of FS-DMRs based on individual peak scores

	1085 (FA[1])	1156 (FA)	1099 (FA)	Mean (FA)	Standard (FA)		1061 (P[2])	1010 (P)	1074 (P)	Mean (P)	Standard (P)		Verdict
Inositol hexakisphosphate kinase 1	1.36	1.49	5.52	**2.79**	2.3637		0.60	0.50	0.33	**0.48**	0.1347		**Pass**
Macrophage Stimulating 1	2.07	4.50	2.54	**3.04**	1.2897		~~1.77~~	0.51	0.66	**0.98**	0.6907		**Fail**
Lipocalin 15	1.19	3.12	1.75	**2.02**	0.9968		0.68	1.02	~~1.32~~	**1.01**	0.3202		**Fail**
Chromosome 22 open reading frame 34	1.32	3.77	1.79	**2.29**	1.2985		1.14	~~1.29~~	0.73	**1.06**	0.2880		**Fail**
Lymphocyte-specific protein 1 pseudogene	1.45	3.74	1.27	**2.15**	1.3780		~~1.26~~	0.82	~~1.27~~	**1.12**	0.2588		**Fail**
Ubiquitin-conjugating enzyme E2M pseudogene 1	2.71	1.56	1.79	**2.02**	0.6113		0.79	~~1.23~~	0.98	**1.00**	0.2219		**Fail**
Solute carrier family 38, member 3	~~0.88~~	0.42	0.24	**0.51**	0.3293		0.79	~~0.69~~	0.88	**0.78**	0.0945		**Fail**
Chromosome 9 open reading frame 44	0.56	0.47	0.45	**0.49**	0.0579		2.38	0.95	1.15	**1.50**	0.7752		**Pass**
Amiloride binding protein	0.53	0.45	0.42	**0.47**	0.0531		1.81	~~0.59~~	1.59	**1.33**	0.6466		**Fail**
Ciliary rootlet coiled-coil, rootletin-like 1	0.60	0.35	0.37	**0.44**	0.1381		0.88	~~0.54~~	2.10	**1.17**	0.8183		**Fail**
Death-domain associated protein	~~0.86~~	0.66	0.29	**0.60**	0.2900		3.12	1.12	1.03	**1.76**	1.1806		**Fail**
G protein pathway suppressor 2	0.44	0.60	0.30	**0.45**	0.1488		~~0.62~~	1.39	0.92	**0.98**	0.3883		**Fail**
Melanoma antigen family A,	0.63	0.49	0.66	**0.59**	0.0915		1.21	1.24	~~0.72~~	**1.05**	0.2911		**Fail**
Surfeit 1 Fold Change	0.50	0.32	0.54	**0.46**	0.1159		0.84	~~0.69~~	1.19	**0.91**	0.2550		**Fail**
Transmembrane protein 145	0.57	0.53	0.36	**0.49**	0.1127		~~0.77~~	~~0.67~~	1.28	**0.90**	0.3264		**Fail**
Hypothetical LOC149134	0.33	~~0.84~~	0.33	**0.50**	0.2944		~~0.62~~	0.93	1.07	**0.87**	0.2287		**Fail**

1: folic acid intervention group; 2, placebo group. Figures with strikethroughs have been removed from the dataset for not adhering to the criteria listed in Section 3.3.6

Table 3.7: Final List of FS-DMRs

	1085 (FA[1])	1156 (FA)	1099 (FA)	Mean (FA)	Standard (FA)		1061 (P[2])	1010 (P)	1074 (P)	Mean (P)	Standard (P)
Inositol hexakisphosphate kinase 1	1.36	1.49	5.52	**2.79**	2.3637		0.60	0.50	0.33	**0.48**	0.1347
Chromosome 9 open reading frame 44	0.56	0.47	0.45	**0.49**	0.0579		2.38	0.95	1.15	**1.50**	0.7752
RAS p21 protein activator 4	6.76	6.10	8.94	7.27	1.4860		0.65	1.08	1.14	0.96	0.2673
sphingosine-1-phosphate lyase 1	7.08	7.73	6.71	7.17	0.5168		1.09	0.96	1.27	1.11	0.1557
chromosome 19 open reading frame 75	8.82	11.29	9.80	9.97	1.2433		N/A	N/A	N/A	N/A	N/A

1: folic acid intervention group; 2, placebo group

Table 3.8: Probe positions of FS-DMR Peaks

RASA4	Peak Start	Peak End
1085A	Chr7:102155177	Chr7:102158474
1085B	-	-
1156A	Chr7:102157905	Chr7:102158474
1156B	-	-
1099A	Chr7:102155177	Chr7:102158160
1099B	-	-
1061A	Chr7:102157726	Chr7:102158474
1061B	Chr7:102157905	Chr7:102158160
1010A	Chr7:102155177	Chr7:102158371
1010B	Chr7:102156289	Chr7:102157575
1074A	Chr7:102156289	Chr7:102158474
1074B	Chr7:102156401	Chr7:102158474

SGPL1	Peak Start	Peak End
1085A	Chr10:72575627	Chr10:72576074
1085B	-	-
1156A	Chr10:72575703	Chr10:72576174
1156B	-	-
1099A	Chr10:72575823	Chr10:72576476
1099B	-	-
1061A	Chr10:72575823	Chr10:72576074
1061B	Chr10:72575909	Chr10:72576174
1010A	Chr10:72575403	Chr10:72576174
1010B	Chr10:72574322	Chr10:72574573
1074A	Chr10:72575823	Chr10:72576174
1074B	Chr10:72575315	Chr10:72576074

IP6K1	Peak Start	Peak End
1085A	Chr3:49823693	Chr3:49824138
1085B	Chr3:49823779	Chr3:49824236
1156A	Chr3:49828748	Chr3:49831097
1156B	Chr3:49828548	Chr3:49831381
1099A	Chr3:49823779	Chr3:49824138
1099B	Chr3:49828626	Chr3:49830779
1061A	Chr3:49828626	Chr3:49831097
1061B	Chr3:49828626	Chr3:49831097
1010A	Chr3:49823693	Chr3:49824464
1010B	Chr3:49828924	Chr3:49829073
1074A	Chr3:49828548	Chr3:49831473
1074B	Chr3:49823779	Chr3:49824138

Chr9ORF44	Peak Start	Peak End
1085A	Chr9:94903907	Chr9:94906773
1085B	Chr9:94904005	Chr9:94904952
1156A	Chr9:94904005	Chr9:94906773
1156B	Chr9:94903615	Chr9:94904952
1099A	Chr9:94903907	Chr9:94906773
1099B	Chr9:94904005	Chr9:94904144
1061A	Chr9:94904393	Chr9:94904764
1061B	Chr9:94895004	Chr9:94901361
1010A	Chr9:94894792	Chr9:94906644
1010B	Chr9:94903907	Chr9:94906773
1074A	Chr9:94903715	Chr9:94904952
1074B	Chr9:94903615	Chr9:94905046

Table 3.8: Probe positions of FS-DMR Peaks (continued)

Chr19ORF75	Peak Start	Peak End
1085A	-	-
1085B	Chr19:51767838	Chr19:51769287
1156A	-	-
1156B	Chr19:51767838	Chr19:51769287
1099A	-	-
1099B	Chr19:51755588	Chr19:51756034
1061A	-	-
1061B	-	-
1010A	-	-
1010B	-	-
1074A	-	-
1074B	-	-

Chapter 4:

Confirmation of candidate FS-DMR sites and assessment of their impact on gene expression

4.1: Introduction

4.1.1: Overview

The Methylated DNA Immunoprecipitation (MeDIP) and promoter microarray analysis described in Chapter 3 has yielded three potential folate sensitive differentially methylated regions (FS-DMRs): *IP6K1*, Chr9ORF44, and *RASA4*. Although five preliminary FS-DMRs have been found, none have passed the data qualification steps outlined in Section 3.3.7. However, these three exhibited the most dramatic change in response to folic acid supplementation.

The genome-wide analysis has not been carried out on a sample-set large enough to make any valid conclusions, nor do they contain any quantitative data regarding the extent of DNA methylation at these regions. For these reasons, it was necessary to design a set of experiments to confirm these findings using a gene-specific DNA methylation analysis on an expanded dataset.

In this chapter, three assays for gene-specific analysis of *IP6K1*, Chr9ORF44, and *RASA4* were designed, optimised, and validated (Section 4.3.1), then used to screen the entire FASSTT cohort for changes in DNA methylation (Section 4.3.3). Following this, a cell culture model treated with a demethylating agent was prepared (Section 4.3.5) to assess the effect of DNA methylation at these loci on gene expression (Section 4.3.7).

Confirmation of these FS-DMRs and aligning the dataset with gene expression data concludes the investigation of the impact of folic acid supplementation on DNA methylation levels during pregnancy.

4.1.2: The impact of DNA methylation changes on gene expression

Promoter DNA methylation is classically associated with silencing of associated genes[185]. Although recent advancements in the field regarding gene-body and non-CpG complicate this simplified model[159,186], its implications are still relevant to this study. All of the regions to be analysed in this chapter lie within or near the promoter region of their respective genes. In order to make the statement that these methylation changes are of biological relevance, aberrant methylation in these regions will need to be associated with gene expression.

For this analysis, a cell culture model based on HEK293 cells treated with 5-azacytidine (5aC) was designed. As a chemical analogue of cytosine, 5aC severely inhibits the action of the DNA methyltransferase (DNMT) enzymes, leading to global passive demethylation[187]. Previous studies have examined the effect 5aC (trade name Vidaza) exposure to mammalian

cells, measuring gene expression and DNA methylation up to several days of treatment[188–191], with concentrations varying depending on the cell lines used. Falck *et al.*, (2012)[190] observed a decrease of global DNA methylation in HEK293 cells treated with 1µM 5aC for 72 hours, from 4.61% before treatment to 1.38% after treatment, measured with micellular electrokinetic chromatography combined with laser-induced fluorescence detection. On the other hand, Komashko *et al.*, (2010)[191] treated HEK293 cells with 5µM 5aC for 8 days, confirming demethylation at selected loci with MeDIP PCR. Interestingly, Komashko and colleagues found that most genes that altered their gene expression in response to 5aC were not regulated by promoters displaying DNA methylation prior to treatment. This implies that 5aC treatment of HEK293 cells can have an effect on gene expression independent of its demethylating capabilities.

Another demethylating agent and analogue of cytidine is 5-aza-2'-deoxycytidine (trade name Decitibine)[192]. Today, both drugs are used for the treatment of myelodysplastic syndrome (MDS)[193]. Although both appear to be very similar in functionality, 5aC has a stronger cytotoxic affect[194]. Moreover, while 5-aza-2'deoxycytidine integrates 100% into DNA, 5aC is also capable of integrating into RNA, possibly affecting gene expression without impacting on DNA methylation[195].

In order to examine the effect of DNA methylation on gene expression, the growth of HEK293 cells exposed to 5aC was optimised to retain them in the log stage for 7 days, which is especially important due to its strong cytotoxic affect. This process is outlined in Section 4.2.4. By ensuring cells are exposed to the highest concentration of 5aC for a prolonged amount of time, significant loss of DNA methylation is expected at their promoter regions. It is imperative that any change in gene expression is mirrored by a loss of methylation, due to the potential for 5aC to interfere with mRNA. A loss in methylation should cause a symmetrical increase in gene expression, but as these studies indicate, the impact of DNA methylation on gene expression may not be as straightforward as previously understood.

4.1.3 Aims and Objectives

Aim

To validate the FS-DMRs found in Chapter 4 with a gene-specific method of DNA analysis and examine the effect of DNA demethylation on gene expression *in vitro* in these regions.

Objectives

- To design and optimise SMART-MSP assays for the candidate FS-DMRs identified from the MeDIP analysis.
- To screen the FASSTT dataset across said SMART-MSP assays and confirm/refute the findings of Chapter 3.
- To design a HEK293 cell-culture model to assess the effect of demethylating agent 5'azacytidine on gene expression at these loci.
- To confirm DNA demethylation of HEK293 DNA has occurred as a result of exposure to 5'azacytidine.
- To carry out a gene expression analysis on RNA extracted from cells treated with 5'azacytidine using RT-qPCR.

4.2: Methods

4.2.1: SMART MSP assay design

Primers for SMART MSP analysis were designed based on the criteria outlined in Kristensen *et al.* (2008)[146]. Briefly, two CpGs were included in the primer sequences wherever possible, with one placed close to the 3` end of the oligonucleotide. In order to control for incomplete bisulfite conversion, the amplification region of each assay was designed to contain non-CpG Cs. With unmethylated Cs deaminating to Ts after sodium bisulfite treatment, a lower melting temperature would be expected in unmethylated DNA compared to that of the fully methylated standard. This is due to the fact that the base pair G-C is held together by three hydrogen bonds while A-T contains two; the former requires more energy to break than the latter.

Primers were designed to keep both annealing temperatures as similar to one another as possible. Sometimes not all of these guidelines could be adhered by, due to the decrease in sequence complexity as a result of sodium bisulfite treatment and the limited regions being analysed from Chapter 3.

MethPrimer (http://www.urogene.org/methprimer/) was used to sodium bisulfite treat the sequence of interest *in silico*. Primers with matching annealing temperatures were found using NetPrimer (http://www.premierbiosoft.com/netprimer/), with as little primer-dimer and self-dimer forming capabilities as possible.

4.2.2: Rationale for FS-DMR selection for Confirmation in the FASSTT cohort and gene expression analysis

The following FS-DMRs were selected for confirmation in the FASSTT cohort as they each represented a different FS-DMR from the three groups presented in Chapter 3: Increase, Decrease, and On/Off: *IP6K1*, Chr9ORF44, and *RASA4*. *IP6K1* and *RASA4* were chosen for the gene expression experiments, as the FS-DMR occurs within or near the regulatory region of these genes and therefore will inform on whether methylation changes in the promoter regions of these genes equals changes in gene expression. Since the Chr9ORF44 FS-DMR is not known to occur within a regulatory region of a given gene, another gene from the preliminary Decrease in methylation group was chosen for the gene expression analysis: GPS2.

All assays were carried out using the Roche Lightcycler™ 480 system in 10µl reactions containing 2X Lightcycler 480 High Resolution Melting Master, 2.5mM magnesium chloride, and variable concentrations of forward and reverse primer (Table 4.1). For each reaction, the following programme was used: Pre-incubation, 95ºC for 10 seconds; Amplification (45 cycles), 95ºC for 10 seconds, variable annealing temperature (Table 4.1) for 15 seconds, and 72ºC for 15 seconds; Melting, 95ºC for 1 minute, 40ºC for 1 minute, 70ºC for 5 seconds, and 95ºC continuous with a ramp-rate of 0.02ºC/s and 25 acquisitions every 1ºC; Cooling, 40ºC for 10 seconds.

4.2.3: SMART MSP Optimisation

The optimisation of the SMART-MSP assays followed a three-pronged approach. First, using 100% and 0% DNA methylation standards (Qiagen), an appropriate annealing temperature that caused the reaction to preferentially amplify the methylated template was determined. Increasing the temperature by 1ºC increments was found to slowly increase C_P values (i.e. inhibiting amplification) while eliminating primer-dimer formation. When a suitable annealing temperature was found, a 9-reaction primer curve containing every combination of forward and reverse primer concentrations of 0.2µM. 0.3µM, and 0.4µM was used. Primer concentrations that produced a single product favouring methylated DNA and producing as little non-specific amplification were chosen.

To confirm that the assays were working as expected, two standard curves were run on each: one with a decreasing quantity of DNA to determine the PCR efficiency of the reactions, and another serial dilution of the 100% standard against the 0% standard. After each assay was optimised, it was apparent that each was behaving differently and would require a unique approach in the subsequent analysis, discussed further in Section 4.3.1.

4.2.4: SMART MSP FASSTT Screening

Prior to screening, each FASSTT sample was sodium bisulfite treated using the Qiagen EpiTect kit according to the manufacturer's instructions with one alteration: 1µg DNA was used in each reaction and eluted in 100µl dH_2O. Levels of amplifiable DNA were determined using the COL2A1 assay described by Kristenisn *et al.,* (2008)[146]. This is an assay with no CpGs in its primer or product region, and therefore does not differentiate between methylated and unmethylated DNA. For the ΔΔCT calculations described later, COL2A1 was treated like an endogenous control. This assay was also optimised and tested for performance based on the criteria outlined in Section 4.2.3. Each assay was performed in duplicate. Standards of 100% and 0% methylation were included on each plate, along with a negative no-template control. Melting peak data was used to confirm that the correct product had amplified for each sample, while methylation percentages were calculated from the C_P data using the following formula:

*100 * 2^ - (Sample [Cp Target – Cp COL2A1] – 100% [Cp Target – Cp COL2A1])*

Here, 'Target' refers to the gene of interest, and 'Calibrator' refers to the COL2A1 assay. This formula is accurate where PCR efficiencies for both the gene of interest and the COLA1 assay are approximately identical[146]. Where PCR efficiencies were not similar, some calculations yielded a value greater than 100%. To address this, the C_P and Tm values of each of these samples were examined to determine if the methylation percentage should be adjusted to 100%, or discarded from the dataset entirely.

4.2.5: HEK293 5'Azacytidine cytotoxicity assay

The toxicity assay for HEK293 cells treated with 5aC was carried out on a 24-well plate (6 $x10^3$ cells per well). A 40µM stock of 5'azacytidine was made in 50% acetic acid, which was then diluted down to various concentrations in DMEM (0.5µM, 3µM, 6µM, 12µM and 25µM). Media containing acetic acid was added after the cells attached to the flask. Every day, each well containing 5aC had its media changed. To control for the effect of acetic acid, wells containing acetic acid alone – proportional to the amount used in the 25µM 5aC sample – were included on the plate. Control-wells with cells seeded below and above $6x10^3$ cells per well were also included to ensure cells grew correctly over the 7 day period. Cell density was measured using the crystal violet assay described in Section 2.2.4.

For the 5aC experiments themselves, cells were grown in T75 flasks for 7 days with the media being changed every 24 hours due to the short half-life of the drug[191]. Control flasks containing acetic acid and no drug were maintained in parallel to the 5aC flasks. Once the time-course

was complete, DNA and RNA were extracted from cells treated with both 5aC and acetic acid in biological duplicates.

4.2.6: Reverse-Transcription Quantitative-PCR

RNA was extracted from HEK293 cells using Bioline's Isolate II RNA mini kit as described in Chapter 2, Section 2.3.1 and reverse-transcribed as described in Section 2.3.21. cDNA was analysed using a genomic contamination assay already established in the lab (Section 2.3.22). Once a band at 232bp was observed for amplified cDNA without genomic DNA contamination, gene-expression levels were analysed using the UPL RT-qPCR assays described in Section 2.3.23. All samples were run in duplicate, with negative-RT and no template negative controls on each plate.

4.2.7: DNA Extraction from HEK293 cells

DNA was extracted from the HEK293 cells using Qiagen's Flexigene kit (Section 2.3.1), and quantified on the ND-1000 spectrophotometer from Mason Technology measuring at A_{260nm} (Section 2.3.5). Quality was assessed by agarose gel electrophoresis, with a thick high molecular weight band representing genomic DNA (Section 2.3.3).

Prior to sodium bisulfite treatment, DNA was RNase-treated with 25 µg Ribonuclease (Sigma-Aldrich) for 30 minutes at 37°C. DNA was then ethanol precipitated with 0.3M sodium acetate and two volumes of isopropanol. DNA was quantified and checked for quality again after RNase treatment. Since RNase is an incredibly formidable enzyme, it was necessary to carry this out in another lab, followed by an ethanol precipitation step to ensure the DNA was clean of any of the contaminating enzyme. Sodium bisulfite treatment of DNA was carried out as described in Section 2.3.19.

4.2.8: Data analysis and statistics

P-values from FASSTT cohort methylation screening were obtained using single factor ANOVA. Values for methylation changes were placed in three bins: Up (>10%), Down (<-10%), and No Change (-10% $\leq x \leq$10%). The spread of data across these three bins was tested for significance. Net changes across the cohort were analysed by combining values from the Up and Down bins together, with single-factor ANOVA testing for significance. ANOVA was selected to test for significance due to there being three independent groups in this analysis.

All HEK293 experiments were carried out with technical and biological duplicates. Mean and standard deviation values for both RT-qPCR and DNA methylation analysis were obtained from four datapoints for each experiment. Single-factor ANOVA was used to test for significance. RT-qPCR analysis was carried out on the Roche Lightcycler™.

4.3: Results

4.3.1: Design and Optimisation of SMART-MSP assays

SMART MSP assay sequences are described in Section 2.1.3, with their relative genomic position detailed in Appendix D, Figure D.1-D.4. Each of the four SMART MSP assays—IP6K1, Chr9ORF44, RASA4, and GPS2—were optimised against methylated and unmethylated standards, with optimum conditions listed in Table 4.1. In each case, a standard curve of methylated DNA diluted into unmethylated DNA was obtained to evaluate how the assays worked on a heterogeneous mix of DNA; typical of the samples in the FASSTT study. After this, a second standard curve was found for methylated DNA diluted into water, in order to obtain a PCR efficiency for each assay (Appendix D, Figure D.5-D10). As described below, each assay had its own unique qualities which needed to be taken into account during the screening process. Some assays amplified methylated DNA only, while others amplified both methylated and unmethylated DNA. A third type of peak is also sometimes observed: a primer-dimer peak which is identifiable from the negative control. A careful examination of the melting curves (which shows the different types of PCR products that are being amplified) was required to assess which was the case for each assay before the C_P data could be interpreted accurately.

IP6K1: The melting curve data for the IP6K1 assay indicates that there are just two types of peaks amplified for this assay (Figure 4.1). A methylated peak (Tm approx. 76.6^{o}C), and a primer-dimer peak (Tm approx. 74.7 ^{o}C). C_P values arising only from the methylated peak were used in subsequent calculations.

Chr9ORF44: The melting curve data for the Chr9ORF44 assay indicates that there are just two types of peaks amplified for this assay (Figure 4.2). A methylated peak (Tm approx. 78.4^{o}C) and an umethylated peak (Tm approx. 77.4^{o}C). C_P values arising only from the methylated peak were used in subsequent calculations. No amplification was observed for no-template negative controls. Samples containing methylated DNA as low as 10% manifest as a methylated peak for this assay, with those at 1% generating a peak consistent with no methylation. As a result, this assay cannot differentiate between 1% and 0% DNA methylation, but can be used to determine methylation percentages above 10%.

RASA4: The melting curve data for the RASA4 assay indicates that there are just two types of peaks amplified for this assay (Figure 4.3). A methylated peak (Tm approx. 77.8^{o}C) and a primer-dimer peak (Tm approx. 71.5^{o}C). The SMART MSP calculation will be used only in products that exhibit methylated peaks.

GPS2: The melting curve data for the GPS2 assay indicates that there are just two types of peaks amplified for this assay (Figure 4.4). A methylated peak (Tm approx. 81^{o}C) an unmethylated peak (Tm approx. 78.5^{o}C). At 10%, methylation, a double-peak with both products forms. No amplification was evident in the no-template negative control. The SMART MSP calculation will be used only in products that exhibit methylated peaks.

4.3.2: Assessment of the control Col2A1 assay in the FASSTT cohort

The level of amplifiable DNA in each of the sodium bisulfite-treated FASSTT samples was determined using the COL2A1 assay described by Kristentisen *et al*[146]. With no CpGs in either primer sequence, this assay does not discriminate between methylated and unmethylated DNA. Also, with no CpGs present in the PCR product itself, the Tm for both methylated and unmethylated products are the same. As a result, no CpG methylation can occur between the COL2A1 primers. The only variable in the COL2A1 assay is the amount of amplifiable, bisulfite-treated DNA present in the sample.

Of all the samples treated, 10 out of 238 did not amplify with this assay indicating that they did not contain enough amplifiable DNA following bisulfite conversion. Of these, some had low concentrations of DNA before treatment, but even those with high concentrations initially failed to amplify after treatment. Repeated conversions of these samples failed to produce DNA capable of being amplified by the COL2A1 assay.

As each methylation assay would be normalised to the COL2A1 data, it was vital that the consistency of this data was verified. To this end, the COL2A1 assay was run in duplicate for each sample from the FASSTT study and run on two separate occasions with different batches of reagents. Relative ratios of the result of the first run was compared to the second run, and those outside the range 0.8 – 1.20 were considered inconsistent and discarded. A second screening of COL2A1 across the cohort was carried out, and values were compared to the first. Of the 230 samples remaining, 28 produced inconsistent results and were also excluded from the analysis (Table 4.2. Raw data supplied in Appendix E.1).

4.3.3: Methylation analysis of the FASSTT cohort

Methylation percentage values were calculated for each sample based on the calculation described in Section 4.2.4. Similarly to how we identified FS-DMRs in Chapter 3, differences in methylation were assessed for each sample individually first, comparing methylation levels either side of intervention for each participant. The methylation change determined by subtracting the percentage methylation in the pre-intervention sample from the post-intervention sample. This percentage difference value was then placed into one of three groups: Increased methylation, decreased methylation, and no change. These groups were

further subdivided as to whether they were from the folic acid intervention group or placebo. Having assessed the sensitivity of each assay using the standard curves, it was deemed that samples exhibiting a difference of less than 10% had no change in methylation.

Inositol hexakisphosphate kinase 1: *MeDIP Group: Increased methylation*. The majority of samples analysed with the IP6K1 assay showed no change in DNA methylation before and after intervention (Figures 4.5-4.6). Differences that were observed failed to reach statistical significance ($p = 0.95$). Samples from the Up and Down bins were divided into their respective groups, folic acid (n=6), and placebo (n=4), and net methylation change was measured for each. Differences observed here failed to reach statistical significance, ($p=0.19$). Raw data supplied in Appendix E.2.

Chromosome 9 Open Reading Frame 44: *MeDIP Group: Decreased methylation*. Although more variance in the data was observed overall for Chr9ORFF44 than IP6K1 (Figure 4.7), most samples exhibited no change in methylation in response to intervention (Figure 4.8). Those in the placebo group did respond more dramatically than the intervention group for both an increase and decrease in methylation, but this difference did not reach statistical significance ($p = 0.74$). Net changes were measured from samples showing increased or decreased methylation, but when divided into their respective groups, folic acid (n=20) and placebo (n=28), the differences observed failed to reach statistical significance ($p=0.33$). Raw data supplied in Appendix E.3.

RAS p21 protein activator 4: *MeDIP Group: ON/OFF methylation*. Across the majority of the cohort, no DNA methylation was observed for RASA4 (Figure 4.9). Similarly to the two examples above, the differences that were observed did not reach statistical significance (Figure 4.10 A, $p = 0.9$). For the folic acid (n=6) and placebo (n=7) groups, net methylation changes were calculated, but failed to reach statistical significance ($p=0.67$). Raw data supplied in Appendix E.4.

Along with examining differences in methylation between the folic acid and placebo group, the clinical data of samples that exhibited the most dramatic changes were compared to one another. For IP6K1, the single sample with the greatest decline in methylation was found to have low B_{12} levels when compared to the rest of the cohort. Likewise, the sample showing the second-largest decline in methylation for Chr9ORF44 was also found to have low levels of $B_{12.}$ However, the sample with the largest decline in methylation for this region did not match the emerging trend. For RASA4, two samples showed a considerable increase in DNA methylation. One of these two individuals was homozygous for the rare TT variant of the MTHFR polymorphism, while the other had the wild type CC variant.

4.3.4: Assessment of HEK293 Growth Curve in the presence of varying concentrations of 5'Azacytidine

On a 24-well plate, a crystal violet assay was carried out to -find the optimum concentration of 5aC that maintained cells in the log phase of growth. At a concentration of 5µM, the drug was found to have the most significant effect on growing cells while keeping them in the exponential phase by Day 7. Higher concentrations lead to significant cell death, while lower concentrations arrested growth before the 7 day time-course was complete (Figures 4.11-4.12).

4.3.5: Nucleic Acid Extraction and Integrity

After extractions were carried out based on the protocols outlined in Section 2.3.1, the concentrations of nucleic acid was assessed using the nanodrop ND-1000 spectrophotometer from Mason Technology measuring at A_{260nm}, and integrity was assessed via gel electrophoresis. Figure 4.13 shows intact, undegraded RNA on a 1% gel, while a demonstration that the resulting cDNA was free from contaminating genomic DNA was shown using an intron flanking PCR based assay, (Figure 4.14).

4.3.6: Gene Expression Analysis of HEK293 Cells treated with 5'Azacytidine by RT-qPCR

RT-qPCR assays for *IP6K1*, *RASA4*, and *GPS2* were designed using the UPL assay design centre website from Roche (http://lifescience.roche.com/shop/products/universal-probelibrary-system-assay-design). An assay designed for *GUS* was used as an endogenous control, with a relative ratio between both experimental conditions established to be 0.92 (Appendix G.1). Standard curves and PCR efficiencies were obtained for each assay (Appendix G.2-G.4), which were incorporated into the Lightcycler 480's E Calculations.

Data is displayed from Figures 4.15 to 4.17. Experiments were carried out in biological and technical duplicates, with mean and standard deviations derived from four data-points. All assays showed the same results, with 5'Azacytidine increasing expression levels of each gene. Experiments for *IP6K1* exhibited an increase in expression exceeding 10-fold ($p=0.007$). The increases in gene expression of *RASA4* were not as dramatic as that for *IP6K1*, but was significantly increased ($p=0.015$). For *GPS2*, an increase in gene expression was also observed ($p=0.003$). The results indicate that 5aC treatment of HEK293 cells has a significant and consistent impact on gene expression at these particular loci.

4.3.7: Gene specific DNA methylation analysis of 5'Azacytidine treated HEK293 cells

The gene expression results from Section 4.3.6 cannot be concluded to be caused by changes in methylation alone, however. As described by Komashko *et al.*, (2010)[191], treatment of

HEK293 cells with 5aC results in a genome-wide impact on expression, altering the transcription of genes regardless of their methylation status. Indeed, as demonstrated by the cytotoxicity assay in Section 4.3.4, exposure to the drug alone results in cell death. Given the plasticity of the transcriptome and its capacity to adapt to changes in the environment, more evidence that changes in gene expression observed through RT-qPCR were caused by demethylation of their promoter regions is required.

SMART-MSP assays *IP6K1*, *RASA4*, and *GPS2* were carried out on each DNA sample extracted from HEK293 cells treated with 5aC. Results were obtained using the standard SMART MSP calculations outlined in Section 4.2.4. In each case, 5aC treatment of HEK293 cells successfully resulted in a decrease in DNA methylation (Figures 4.18-4.20).

In cells treated with 5μM 5aC for 7 days, methylation at the *IP6K1* locus decreased significantly ($p = 2.23E-05$). For *RASA4*, the decrease in methylation was less pronounced, but still statistically significant ($p = 0.009$). For *GPS2*, the loss of methylation was also statistically significant ($p=0.019$). In each of these cases, a decrease in DNA methylation at a promoter region is mirrored by an increase in gene expression.

It is clear that 5'azacytidine treatment did not completely eliminate DNA methylation across the genome, as predicted by Falck *et al.*, (2012)[190]. Only the region analysed by the IP6K1 assay reached levels lower than 10% when exposed to the drug. Nonetheless, high levels of demethylation were found in each region, each consistent with an increase in gene expression for their associated genes.

4.4: Discussion

The primary aim of this project was to assess the effect of folic acid supplementation on DNA methylation during pregnancy. As a methyl donor for DNA methyltransferase reactions, folic acid and its folate derivatives have been hypothesised and demonstrated to influence DNA methylation on both a genome-wide and gene-specific level[3,111,196,197]. With the beneficial effects of folic acid during pregnancy being well established but poorly understood, DNA methylation has been considered to be a candidate for linking the cause and effect of periconceptional supplementation.

The FASSTT study was designed to study the effects of folic acid supplementation after week 12 of pregnancy[4]. During pregnancy, higher levels of circulating homocysteine have been associated with an increased risk of conditions including NTDs, preeclampsia, and recurrent early pregnancy loss[198–200] relative to unaffected pregnancies[198–200]. Folate status is a major

determinant of homocysteine levels, as it is required for the methylation of homocysteine to methionine[201]. Accelerated folate catabolism has been reported to occur during a typical pregnancy[202], which may contribute to the decrease in circulating folate levels found in other studies[26]. Recommended folic acid supplementation procedures during the first trimester aim to prevent this decline, but little has been known about its past week 12 of pregnancy. The FASSTT study has found that folic acid supplementation can prevent the decline in serum and red cell folate levels and the increase in homocysteine levels that occurs otherwise during pregnancy[4].

In order to examine the methylation levels of DNA extracted from the blood samples used in this study, a literature survey was first carried out to examine the different methods of genome-wide and gene-specific DNA methylation analysis used in the field[5]. Although successful in finding methods suitable for the scope and scale of this project, it was clear from the review that the vast majority of methods established for DNA methylation analysis were conceived for the purpose of finding aberrant methylation patterns in cancer[134,138,146]. For each of the three methods discussed in this thesis – MMSDK, MeDIP, and SMART MSP – the "proof-of-concept" experiments carried out in all of their respective publications were on DNA derived from cancerous tissue where dramatic DNA methylation differences are known to occur.

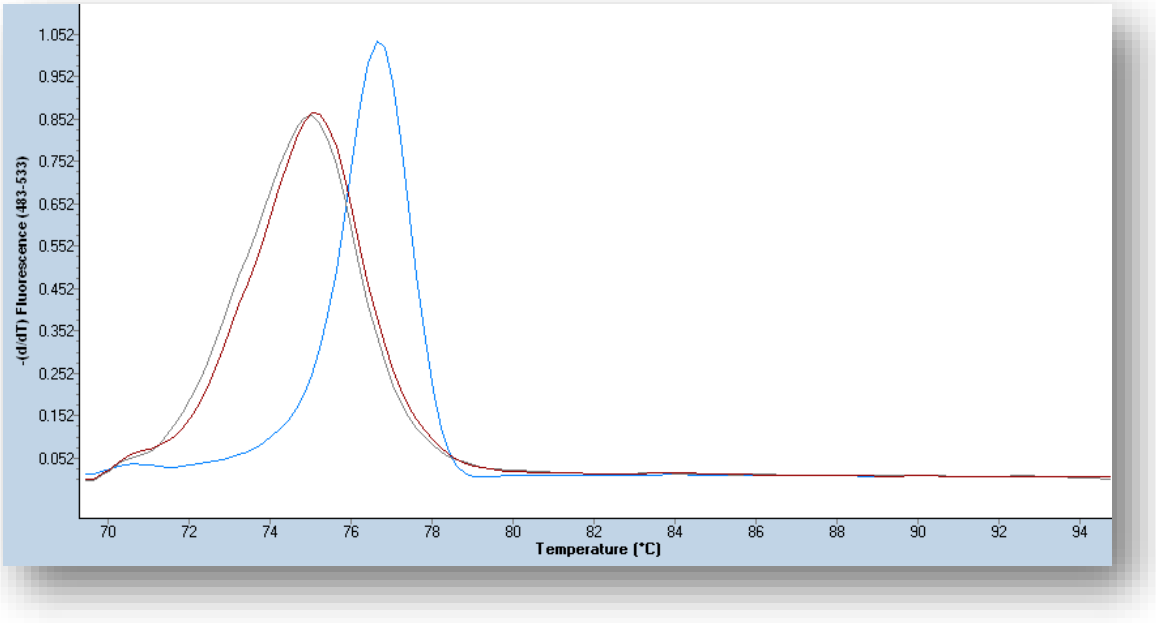
From the subset of the FASSTT cohort subjected to MeDIP and promoter DNA microarray hybridisation, a preliminary list of FS-DMRs was obtained. Using highly astringent selection criteria, this list was narrowed down to five; far fewer than what is typically reported in studies using MeDIP to find DMRs between cancerous and healthy cell-lines[170,203,204]. In the subsequent analysis, when three of these FS-DMRs were examined over an extended cohort using a different, gene-specific method, no significant difference was observed. Some individual samples exhibited major changes in DNA methylation levels for these sites in comparison to the rest of the cohort, and clinical data for each was examined. Of the five individuals examined, 2 were found to have lower B_{12} levels when compared to the rest of the cohort, while 1 was a TT homozygote for the MTHFR 677 C>T polymorphism. Without a consistent trend amongst these samples, however, no conclusion can be made from these findings.

Although these methods were not specific for examining tumourigeneis, the changes in methylation observed in the FASSTT study were not as significant as those presented in the cancer case-studies used to demonstrate their efficiency and specificity. The difficulty lies in the nature of the FASSTT samples themselves: the differences in DNA methylation levels

affected by supplementation in a healthy individual over a short period of time would never be as dramatic as those reported from tumourgenesis[32,169,171,175]. This, along with the differences in DNA methylation patterns that exist between individuals prior to intervention, has made analysing the effect of folic acid supplementation on DNA methylation particularly challenging.

The cell culture analysis did yield conclusive results, however. When the promoter regions of IP6K1, RASA4, and GPS2 are subjected to demethylation *in vitro*, a consistent increase in gene expression is observed demonstrating that DNA methylation changes have a direct impact on the gene expression of these genes.

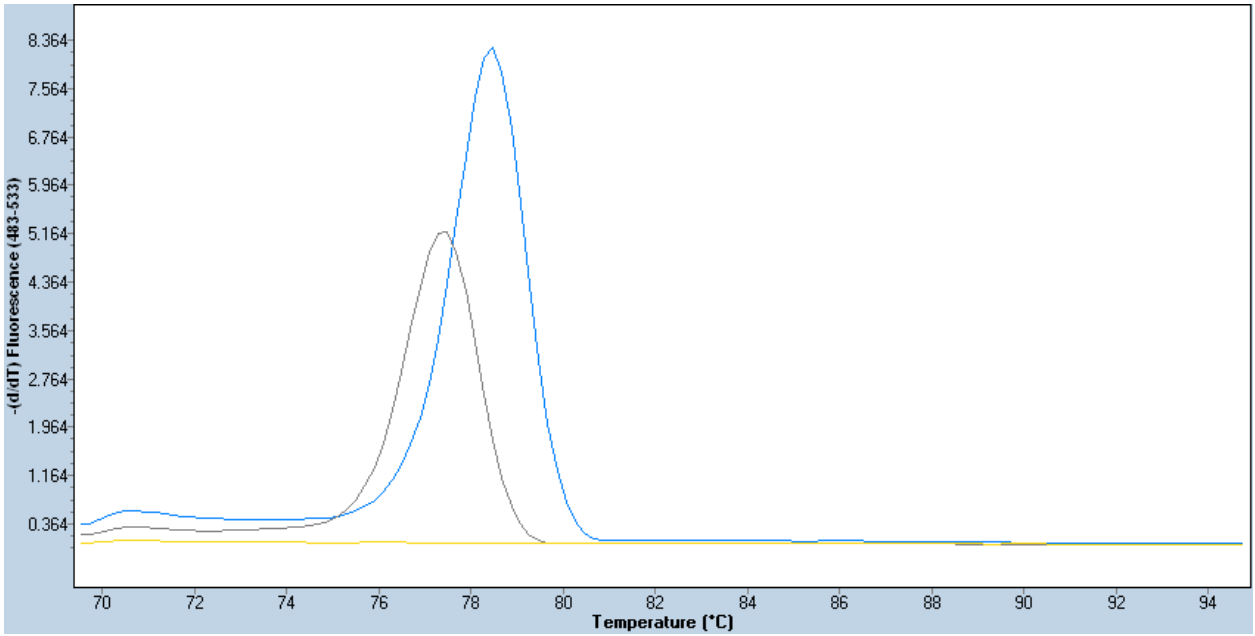
The DNA methylation analysis on the FASSTT cohort did not find measurable changes in DNA methylation in response to folic acid supplementation, but if those differences did exist on a smaller scale, other methods of genome-wide methylation analysis using next generation sequencing and sodium bisulfite treatment to generate data at a single-base-pair resolution could be used to elucidate them. With evidence here that DNA methylation does in fact affect the expression of these genes, there is potential for a follow-up study to re-examine these proposed FS-DMRs using more sophisticated technology, such as sodium bisulfite pyrosequencing.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
100%	26.10	26.19	26.15	0.0636	76.63	76.64	76.64	0.0071
50%	27.30	27.09	27.20	0.1485	76.62	76.49	76.56	0.0919
10%	29.73	29.61	29.67	0.0849	76.49	76.42	76.46	0.0495
1%	32.09	33.52	32.81	1.0112	76.65	76.56	76.61	0.0636
0%	35.62	36.09	35.86	0.3323	74.79	74.50	74.65	0.2051
neg	35.50	35.85	35.68	0.2475	74.99	74.96	74.98	0.0212

Figure 4.1: Melting profile of IP6K1 SMART MSP Assay

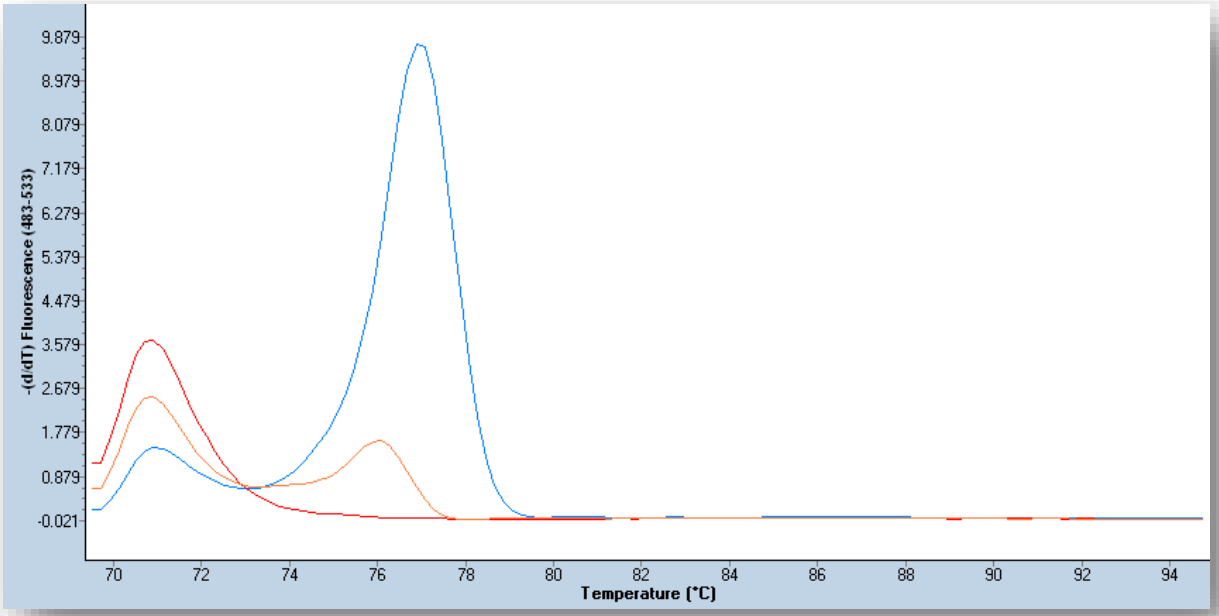
Melting peaks of products from IP6K1 assay. Blue Peak = methylated DNA (~76.6°C); Red peak = unmethylated DNA (74.7°C); Grey peak = no-template negative control (~74.7°C). Methylated and unmethylated DNA PCR products generate distinct melting peaks. The product of the no-template negative control has a peak similar to that of the unmethylated standard, both arising from primer-dimer formation. Only Cp values from the methylated peak (~76.6°C) were considered for the SMART-MSP methylation percentage calculation.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
100%	27.43	27.90	27.67	0.3323	78.41	78.35	78.38	0.0424
50%	28.64	29.19	28.92	0.3889	78.32	78.30	78.31	0.0141
10%	30.22	30.74	30.48	0.3677	78.17	78.22	78.20	0.0354
1%	31.08	31.12	31.10	0.0283	77.51	77.31	77.41	0.1414
0%	30.28	30.33	30.31	0.0354	77.37	77.32	77.35	0.0354
neg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure 4.2: Melting profile of Chr9ORF44 SMART MSP Assay

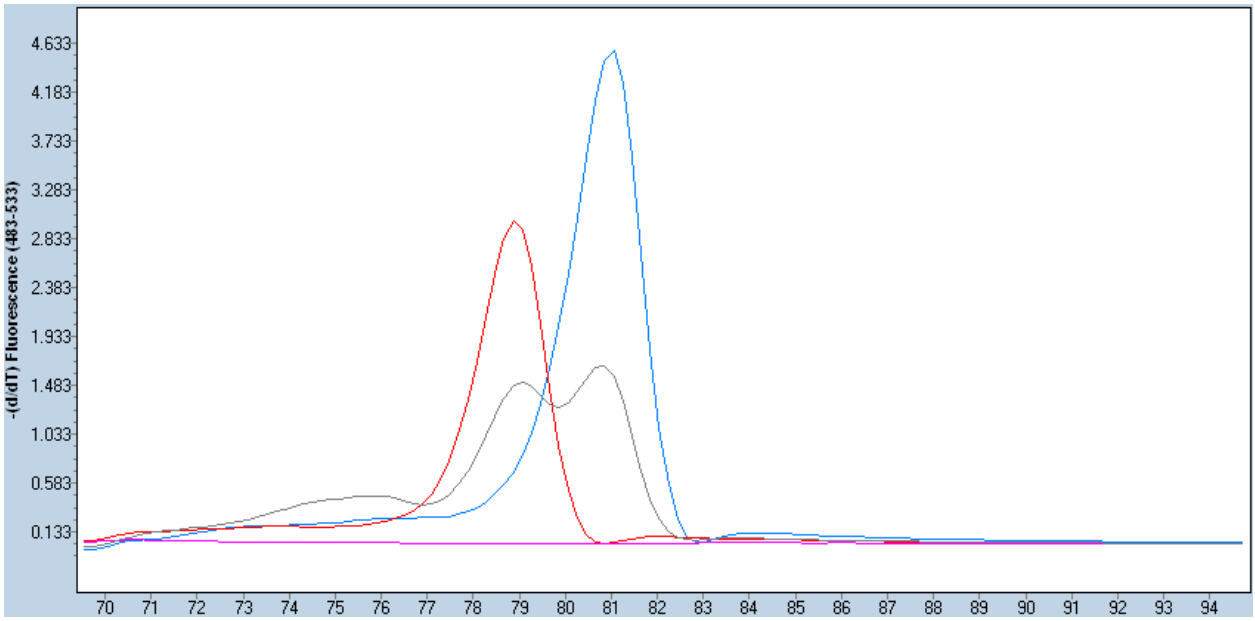
Melting peaks of products from Chr9ORF44 assay. Blue Peak = methylated DNA (~78.3ºC); Grey peak = unmethylated DNA (77.5ºC); Yellow peak = no-template negative control (no amplification). Methylated and unmethylated DNA PCR products generate distinct melting peaks. T_m scores indicate that samples with 100% to 10% methylated DNA will generate a product at ~78.3ºC, while those from 1% to 0% produced a peak at 77.5ºC. Only Cp values from the methylated peak (~78.3ºC) were considered for the SMART-MSP methylation percentage calculation.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
100%	24.54	24.8	24.67	0.1838	76.96	76.8	76.88	0.1131
50%	26.47	26.21	26.34	0.1838	76.77	76.69	76.73	0.0566
10%	27.48	27.36	27.42	0.0849	76.02	76.08	76.05	0.0425
1%	28.86	28.22	28.54	0.4526	Early Peak	Early Peak	Early Peak	Early Peak
0%	35.65	35.91	35.78	0.1838	Early Peak	Early Peak	Early Peak	Early Peak
neg	34.83	33.50	34.165	0.9405	Early Peak	Early Peak	Early Peak	Early Peak

Figure 4.3: Melting profile of RASA4 SMART MSP Assay

Melting peaks of products from RASA4 assay. Blue Peak = methylated DNA (~77.8ºC, with a shoulder of ~71.5ºC); Red peak = unmethylated DNA and no template negative controls (71.5ºC); Orange peak = 10% methylated DNA, (peaks at 76.8ºC and 71.5ºC). Only Cp values from the methylated peak (~76.8ºC) were considered for the SMART-MSP methylation percentage calculation.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
100%	31.67	31.79	31.73	0.0849	80.86	80.83	80.85	0.0212
50%	32.56	32.52	32.54	0.0283	80.71	80.66	80.69	0.0354
10%	35.72	35.53	35.63	0.1344	79.97	79.98	79.98	0.0071
1%	35.32	37.14	36.23	1.2869	78.98	78.92	78.95	0.0424
0%	35.02	35.71	35.37	0.4879	78.83	78.89	78.86	0.0424
neg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure 4.4: Melting profile of GPS2 SMART MSP Assay

Melting peaks of products from GPS2 assay. Blue Peak = methylated DNA (~80.8ºC); Red peak = unmethylated DNA (78.5ºC); Grey peak = 10% methylated DNA, (peaks at 81ºC and 78.5ºC). Only Cp values from the methylated peak (~81ºC) were considered for the SMART-MSP methylation percentage calculation. No template negative controls exhibit zero amplification.

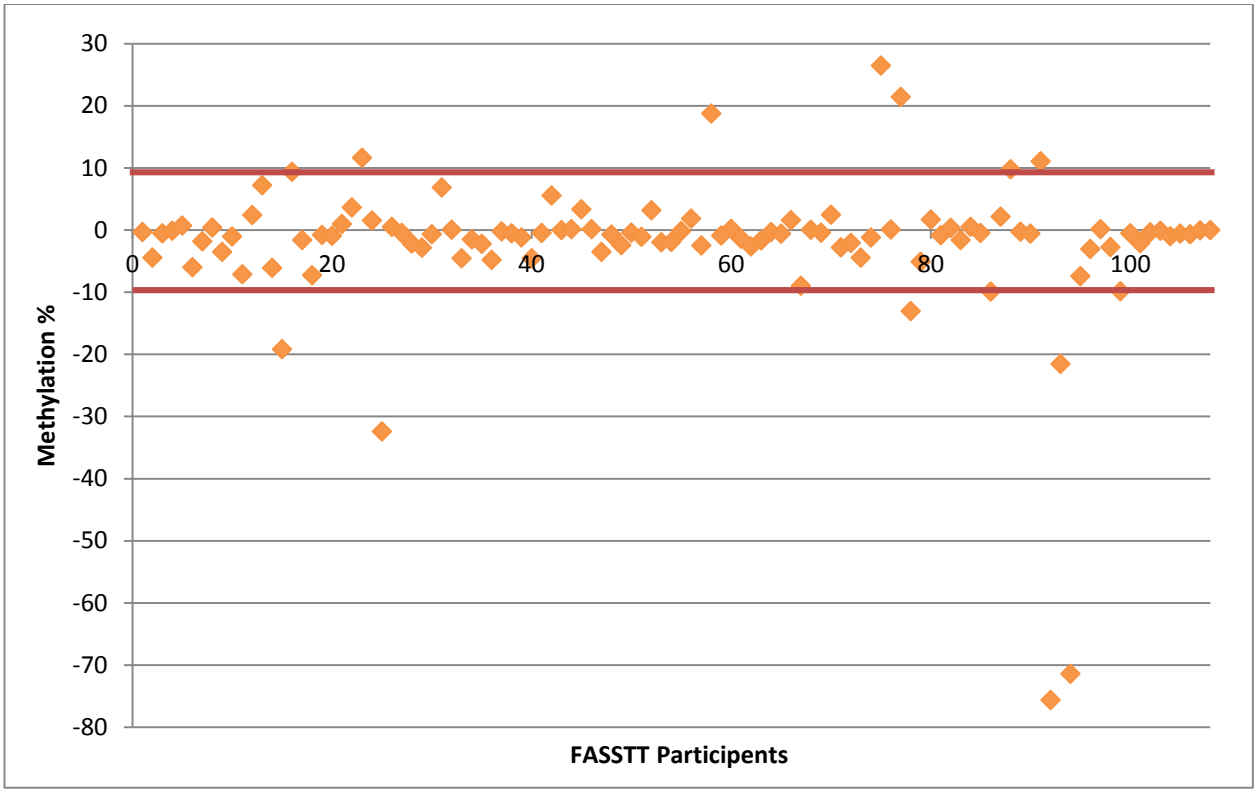


Figure 4.5: Scatterplot of IP6K1 methylation changes for all FASSTT samples

Regardless of their grouping, the change in methylation at the *IP6K1* locus before and after intervention (Y-axis) was plotted for each participant of the FASSTT cohort. Changes of 10% increase or decrease (red lines) were deemed no change due to the insensitivity of the assay to detect consistent changes in this range. The single patient that showed the greatest decline in methylation was found to also have low B_{12} levels when compared to the rest of the cohort.

A

Distribution of methylation changes for IP6K1 across FASSTT cohort

Intervention	Up	Down	No Change	Total
FA	2	4	42	48
Placebo	3	1	48	52
Total	5	6	96	107

B

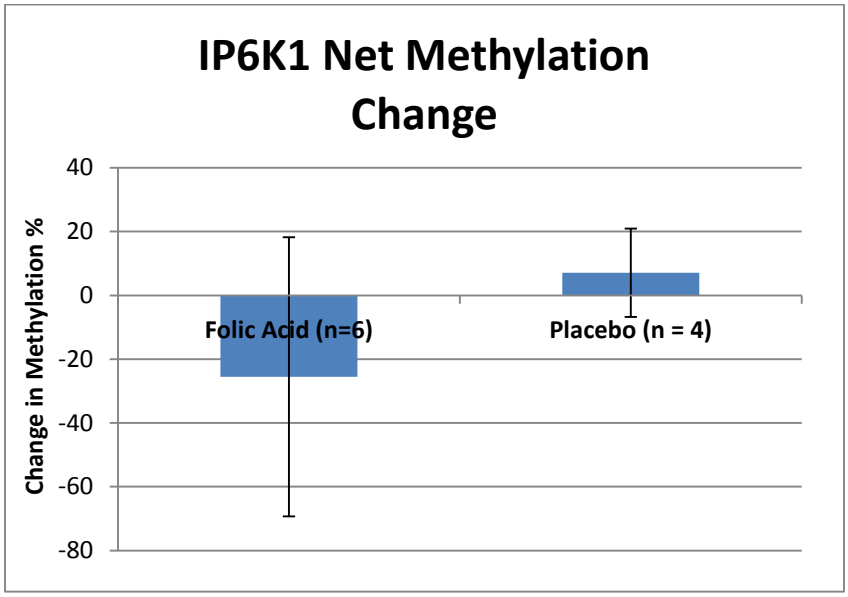


Figure 4.6: FASSTT cohort methylation changes for IP6K1

A: All methylation % changes for IP6K1 across the FASSTT cohort were divided into three groups. Up (>10%), Down (<-10%) and No Change (-10 ≤x ≤10). The vast majorty of these samples fall into the No Change group. Differences observed between the FA and Placebo group failed to reach statistical significance (p=0.95) according to single-factor ANOVA analysis. **B:** All samples excluding those exhibiting no change were divided into either their respective groups, folic acid (n=6), and placebo (n=4), and mean net methylation levels were measured, with the range represented by error bars. Differences observed here failed to reach statistical significance, (p=0.19) according to single-factor ANOVA analysis.

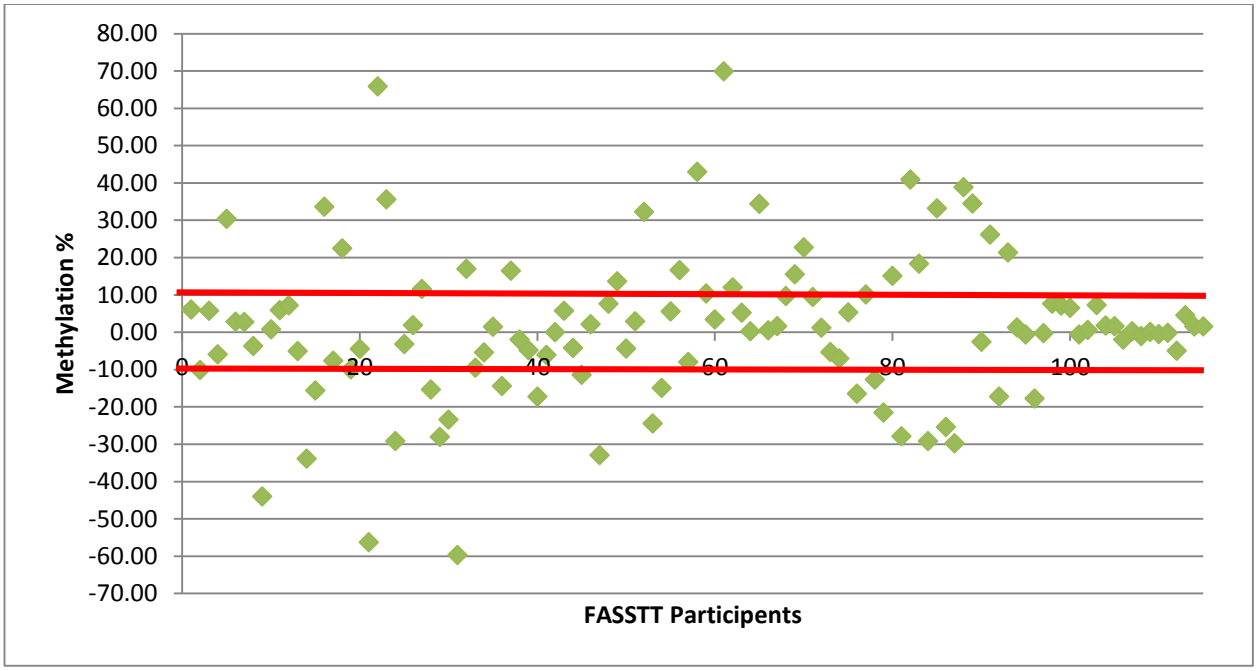


Figure 4.7: Scatterplot of Chr9ORF44 methylation changes for all FASSTT samples
Regardless of their grouping, the change in methylation at the Chr9ORF44 locus before and after intervention (Y-axis) was plotted for each participant of the FASSTT cohort. Changes below 10% (red lines) were deemed to exhibit no change. The single sample with the second-largest decline in methylation was found to have low levels of B_{12}.

A

Distribution of methylation changes for Chr9ORF44 across FASSTT cohort

Intervention	Up	Down	No Change	Total
FA	11	9	30	**50**
Placebo	13	15	31	**59**
Total	**27**	**25**	**63**	**115**

B

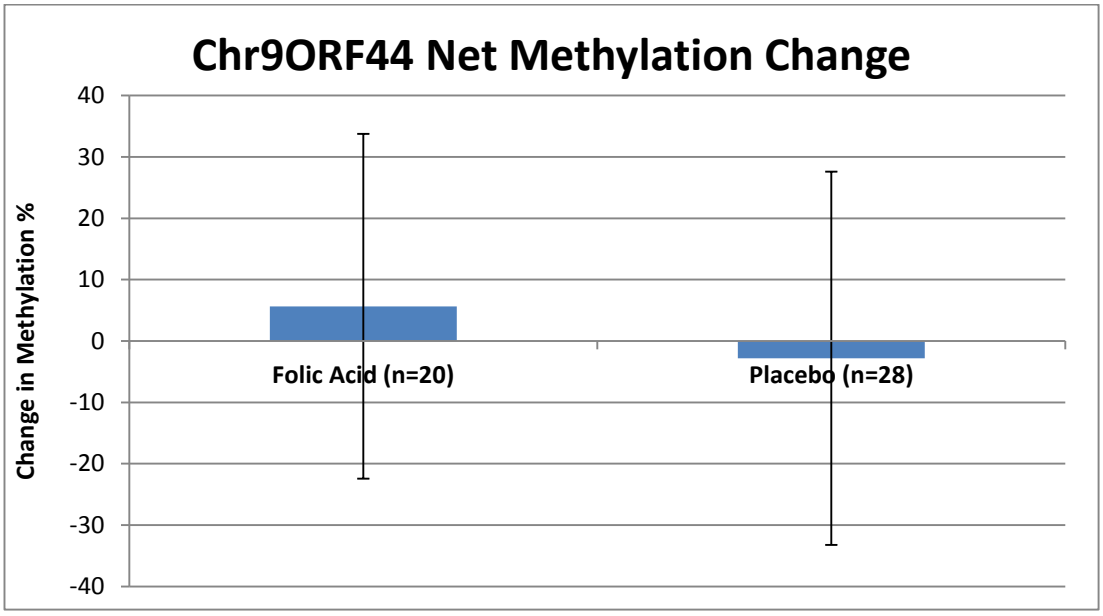


Figure 4.8: FASSTT cohort methylation changes for Chr9ORF44

A: All methylation percentage changes for Chr9ORF44 across the FASSTT cohort were divided into three groups. Up (>10%), Down (<-10%) and No Change (-10 ≤x ≤10). Again, the majorty of these samples fall into the No Change group. For the placebo group, more individuals exhibited measurable changes in methylation percentages than those taking folic acid, though these numnbers failed to reach statistical significance (p=0.75) according to single-factor ANOVA analysis. **B:** All samples excluding those exhibiting no change were divided into either their respective groups, folic acid (n=20), and placebo (n=28), and mean net methylation levels were measured, with the range represented by error bars.. Differences observed here failed to reach statistical significance, (p=0.33) according to single-factor ANOVA analysis.

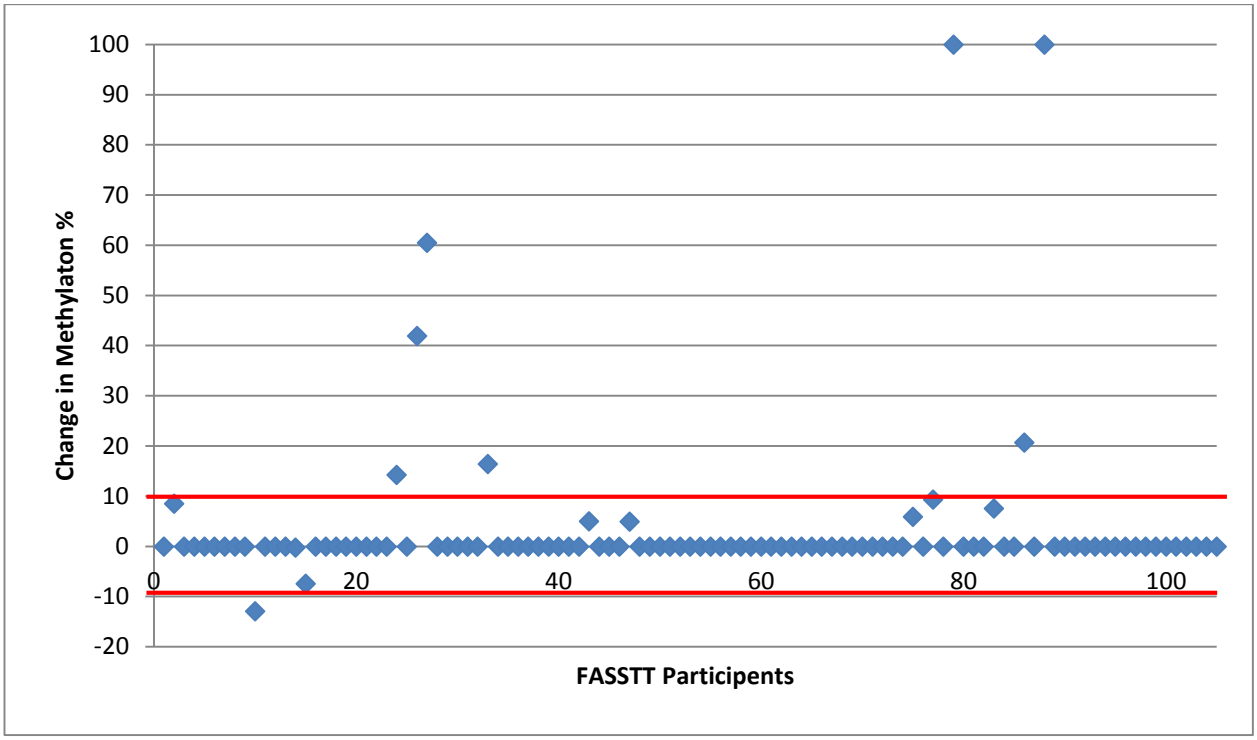


Figure 4.9: Scatterplot of RASA4 methylation changes for all FASSTT samples

Regardless of their grouping, the change in methylation at the RASA4 locus before and after intervention (Y-axis) was plotted for each participant of the FASSTT cohort. Changes below 10% (red lines) were deemed to exhibit no change. Two samples exhibited a greater increase in methylation in comparison to the rest of the cohort. One of these samples was TT homozygous for the MTHFR C>T polymorphism, while the other was found to have low levels of serum folate when compared to the rest of the cohort.

A

Distribution of methylation changes for RASA4 across FASSTT cohort

Intervention	Up	Down	No Change	Total
FA	4	2	43	**49**
Placebo	6	1	50	**57**
Total	**10**	**4**	**98**	**112**

B

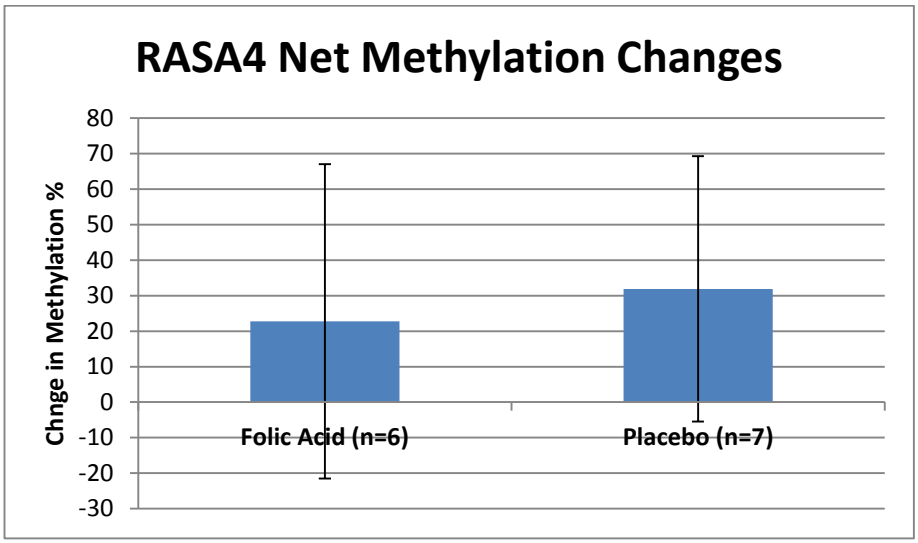


Figure 4.10: FASSTT cohort methylation changes for RASA4

A: All methylation % changes for RASA4 across the FASSTT cohort were divided into three groups. Up (>10%), Down (<-10%) and No Change (-10 ≤x ≤10). Again, the majorty of these samples fall into the No Change group. No changes here reached statistical significance (p=0.9) according to single-factor ANOVA analysis. **B:** All samples excluding those exhibiting no change were divided into either their respective groups, folic acid (n=6), and placebo (n=7) and mean net methylation levels were measured, with the range represented by error bars. Differences observed here failed to reach statistical significance, (p=0.67) according to single-factor ANOVA analysis.

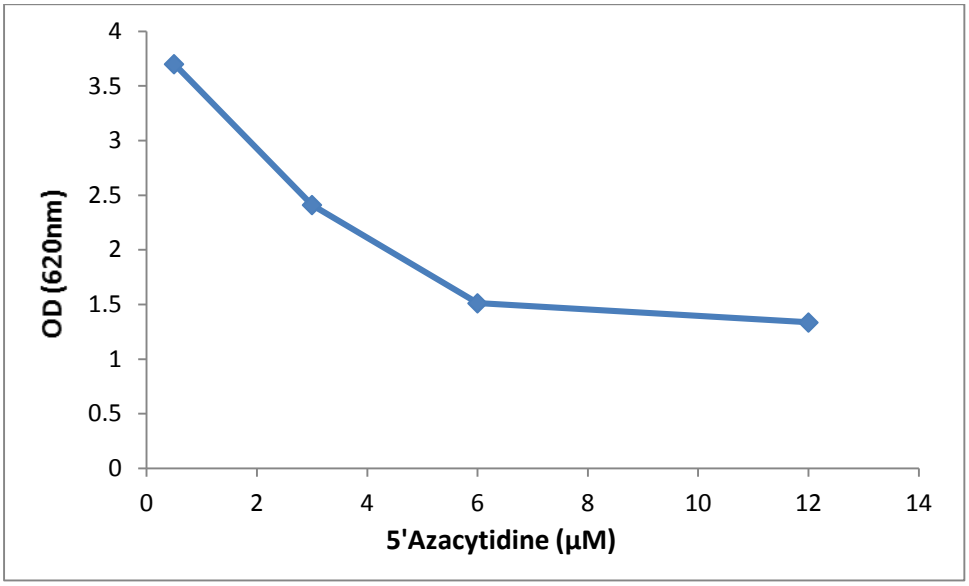


Figure 4.11: HEK293 Cytotoxcity Assay with 5'Azacytidine Treatment

Cells were seeded in a 24-well plate at a seeding denstiy of 3 x10^5 with varying concentreations of 5aC for 7 days. Control wells containing cells above and below seeding density were included to ensure uniform cell proliferation throughout the timecourse. Baseline readings were taken from cells grown in acetic acid, the vehical for 5aC. This data suggests a uniform death at concentrations above 6µM, and little effect below 3µM. As a result, DNA and RNA will be extracted from cells grown in 5µM 5aC over 7 days. Raw data included in Appendix F.1.

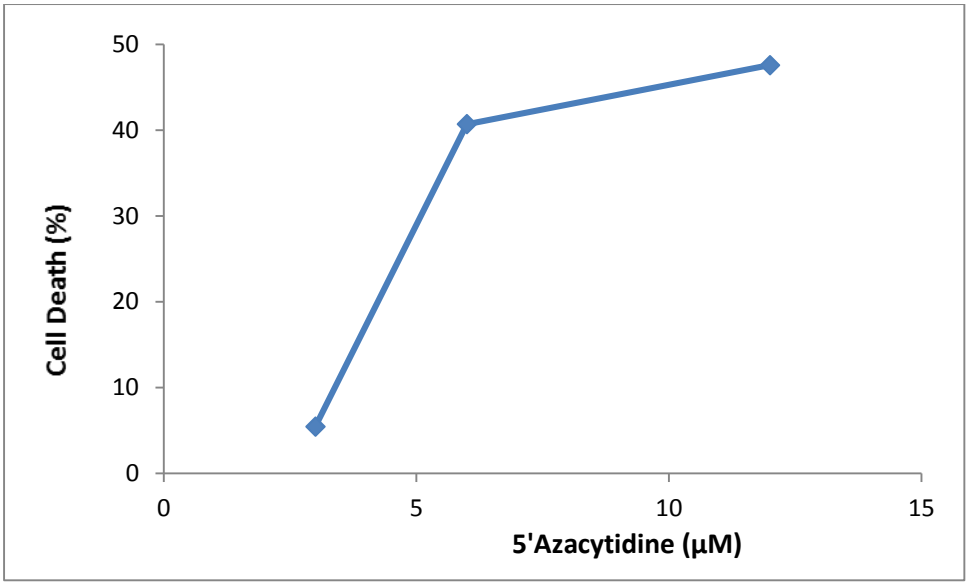


Figure 4.12: HEK293 Percentage Cell Death with 5'Azacytidine

Data from cytotoxicity assay represented as a percentage of overall cell death. Although absolute cell death was not observed in the higher concentrations of 5aC, it was necessary to maintain cells in the log phase of growth over the 7-day timecourse to ensure passive demethylation occurred with each successive cell cycle. For this reason, 5µM 5aC was chosen for the DNA methylation and gene expression experiments.

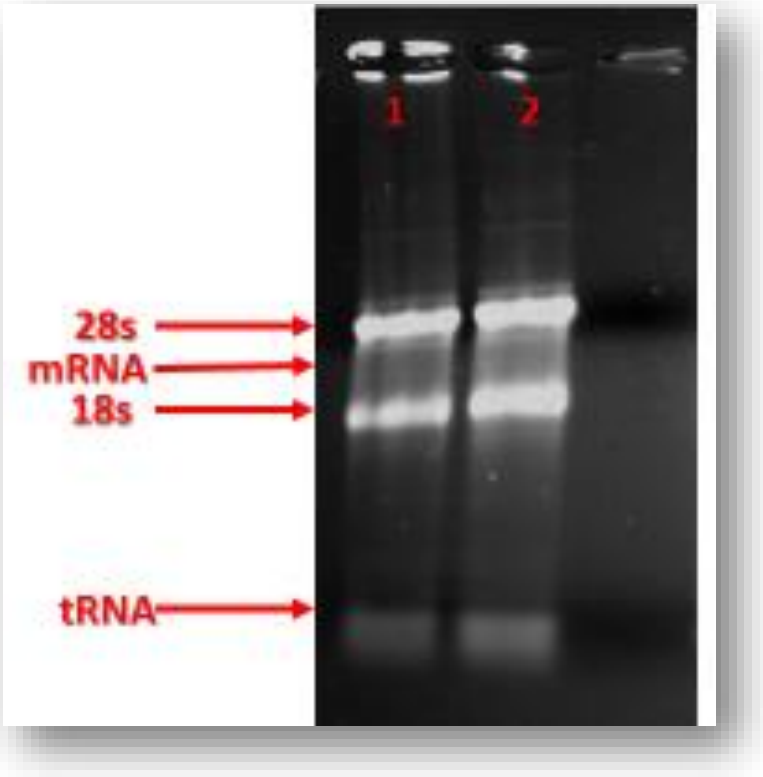


Figure 4.13: RNA extracted from 5aC-treated HEK293 Cells

RNA extracted from HEK293 cells on a 1% agarose gel. Two subunites of ribosomal RNA (28s and 18s) are visible along with mRNA (smear) and tRNA. Samples in lane 1 and 2 are denatured RNA extracted from HEK293 cells treated with 5'azcytidine.

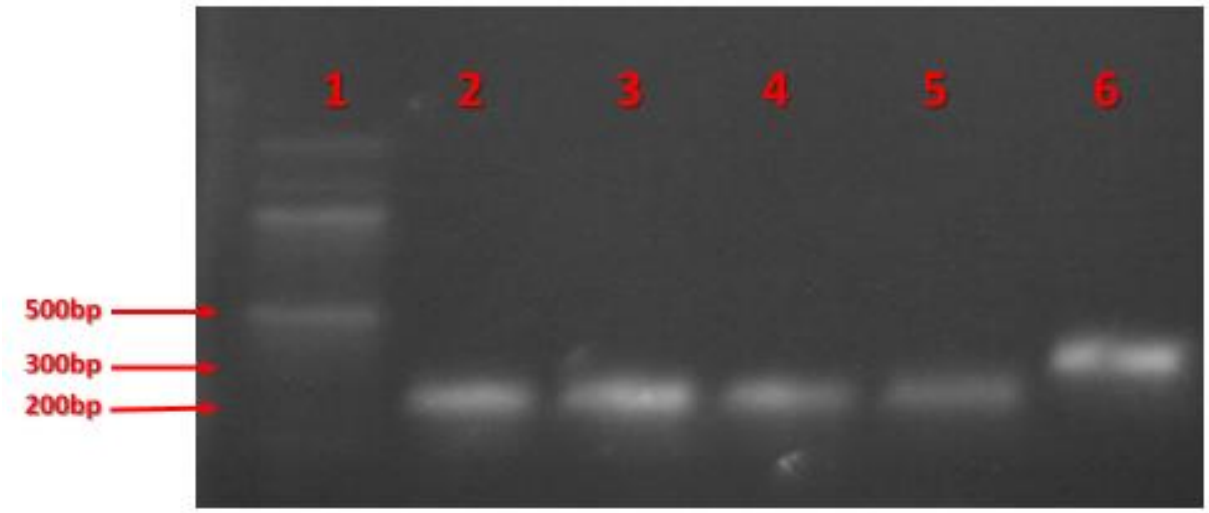


Figure 4.14: cDNA Genomic Contamination test

Products of a PCR that flanks an intron in the *MTHFD1* gene separated on a 1% agarose gel. With primers spanning an intron, cDNA fragments amplified at 232bp, while contaminating genomic DNA manifests as a band at 330bp. Lanes 1 to 4 were loaded with cDNA samples from HEK293 cells treated with 5'azacytidine (samples A0, A5, B0, and B5 respectively). A positive control (lane 6) was included with HEK293 genomic DNA amplified with the same primers. The former products appear on the gel at 232bp, with the latter at 330bp. There is no genomic contamination in these cDNA samples. The 100bp ladder from New England Biolabs was loaded in lane 1.

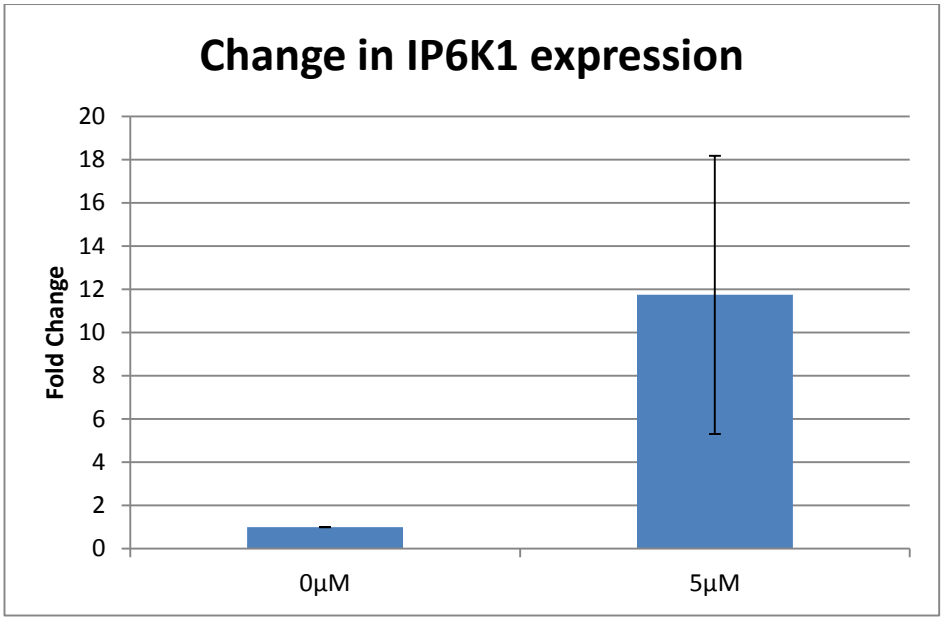


Figure 4.15: IP6K1 Gene Expression in response to 5aC Treatment

IP6K1 expression was found to increase when cells were treated with 5'azacytidine. Cells treated with acetic acid and no 5aC, as described in Section 4.3.4, were used as a baseline. Data taken from two biological replicates. Results reached statistical significance according to single-factor ANOVA analysis ($p = 0.007$). Raw data presented in Appendix G.5

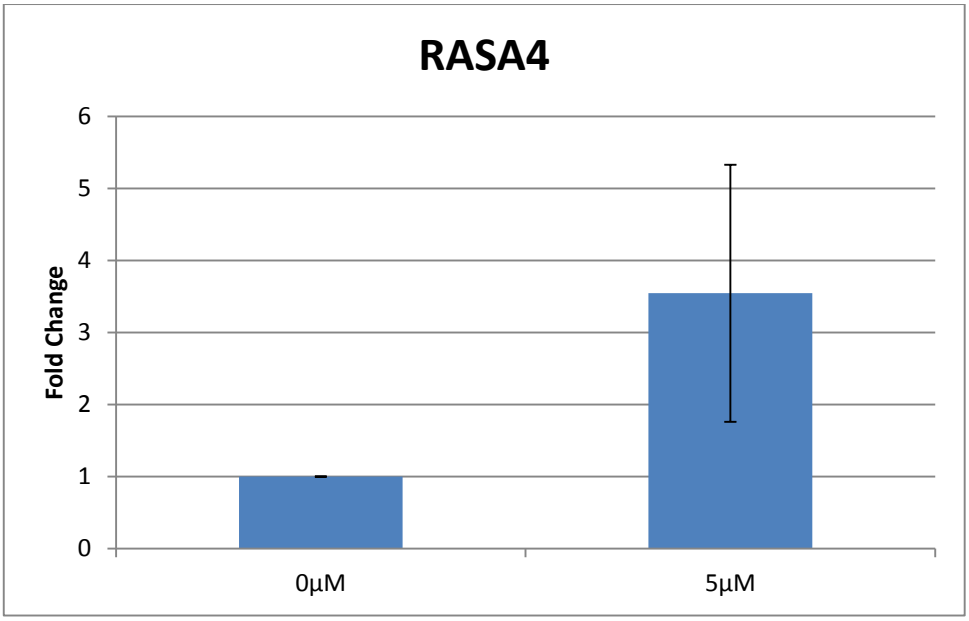


Figure 4.16: RASA4 Gene Expression in response to 5aC Treatment

RASA4 expression was found to increase when cells were treated with 5'azacytidine. Control cells (0µM) treated with acetic acid and no 5aC, as described in Section 4.3.4, were used as a baseline. Data taken from two biological replicates. Results reached statistical significance according to single-factor ANOVA analysis ($p = 0.015$). Raw data presented in Appendix G.6.

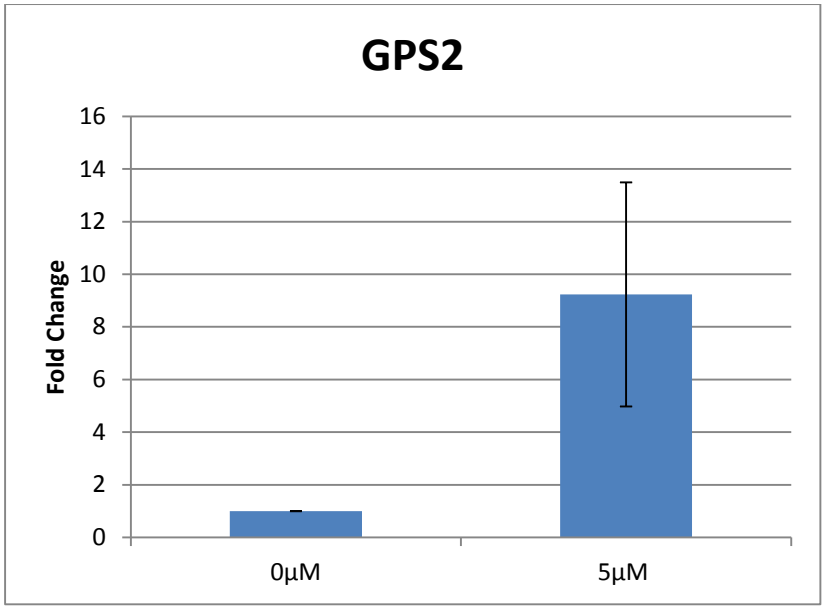


Figure 4.17: GPS2 Gene Expression in response to 5aC Treatment

GPS2 expression was found to increase when cells were treated with 5'azacytidine (5µM). Control cells (0µM) treated with acetic acid and no 5aC, as described in Section 4.3.4, were used as a baseline. Data taken from two biological replicates. Results reached statistical significance according to single-factor ANOVA analysis ($p = 0.003$). Raw data presented in Appendix C, Table C.6.2.

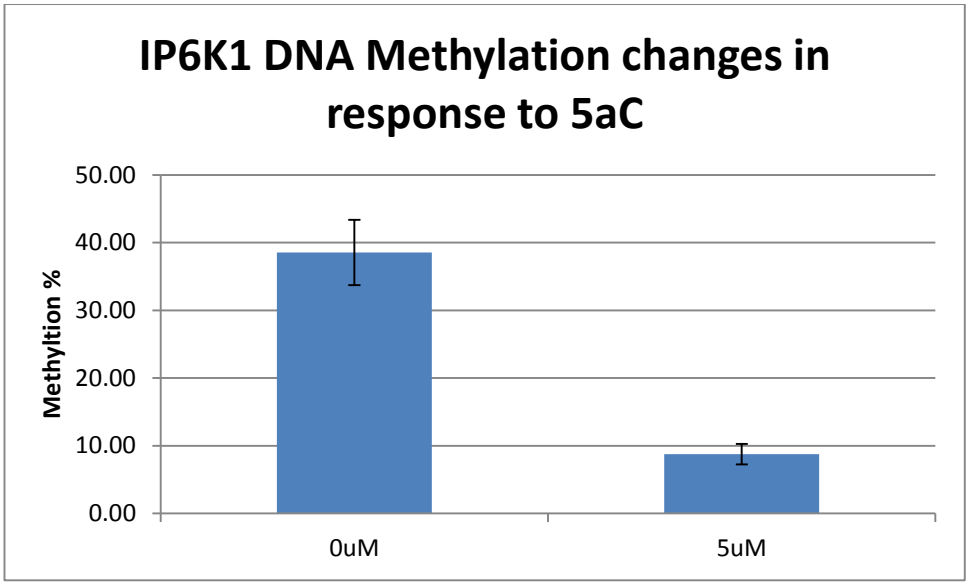


Figure 4.18: IP6K1 DNA Methylation in response to 5aC Treatment

IP6K1 SMART MSP assay confirms that HEK293 treatment with 5'azacytidine resulted in passive demethylation in biological replicates A and B. DNA was extracted from cells treated with 5µM 5'azacytidine (A5 and B5) and acetic acid (A0 and B0) for 7 days and sodium bisulfite treated using Qiagen's Epitect Fast kit. Results reached statistical significance according to single-factor ANOVA analysis (p= 2.23E-05).

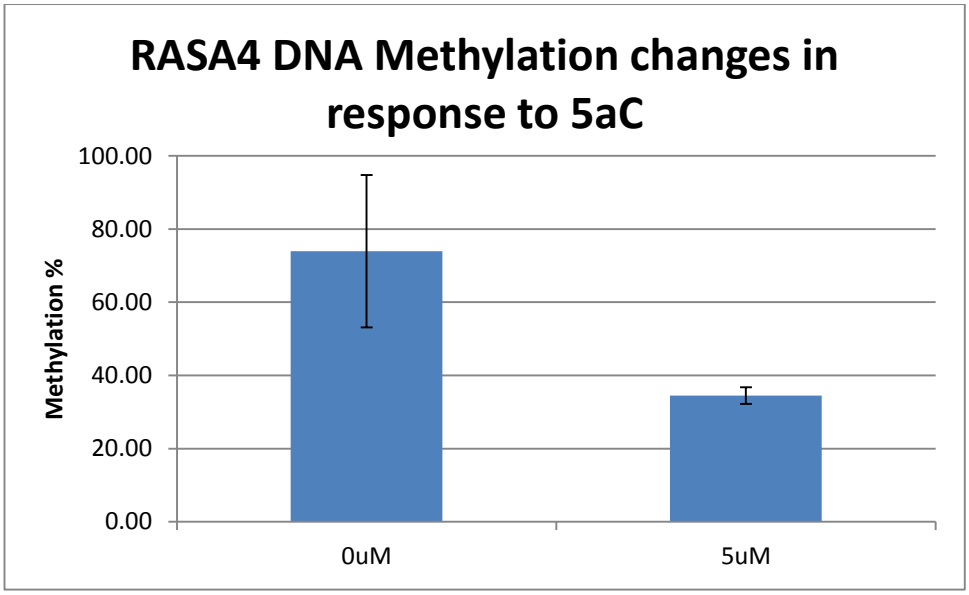


Figure 4.19: RASA4 DNA Methylation in response to 5aC Treatment

RASA4 SMART MSP assay confirms that HEK293 treatment with 5'azacytidine resulted in passive demethylation. DNA was extracted from cells treated with 5µM 5'azacytidine (5µM) and acetic acid (0µM) for 7 days and sodium bisulfite treated using Qiagen's Epitect Fast kit. Results reached statistical significance according to single-factor ANOVA analysis ($p = 0.009$).

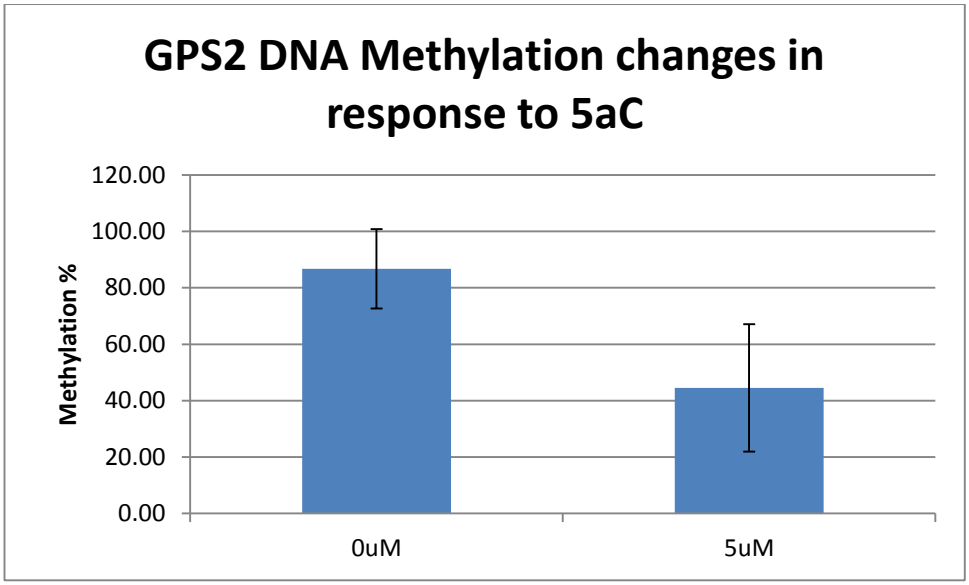


Figure 4.20: GPS2 DNA Methylation in response to 5aC Treatment

GPS2 SMART MSP assay confirms that HEK293 treatment with 5'azacytidine resulted in passive demethylation in biological replicates A and B. DNA was extracted from cells treated with 5µM 5'azacytidine (A5 and B5) and acetic acid (A0 and B0) for 7 days and sodium bisulfite treated using Qiagen's Epitect Fast kit. Results reached statistical according to single-factor ANOVA analysis significance (p=0.019).

Table 4.1: SMART MSP Assay Conditions

	Forward Primer	Reverse Primer	Annealing Temp
IP6K1	0.4µM	0.3 µM	63^{o}C
Chr9ORF44	0.4 µM	0.4 µM	59^{o}C
RASA4	0.4 µM	0.4 µM	56^{o}C
GPS2	0.4 µM	0.4 µM	56^{o}C

Table 4.2: COL2A1 Consistency check across the FASSTT cohort

Verdict		Number of Samples	Percentage
Pass		200	84.03
Fail		28	11.76
N/A		10	4.20

Chapter 5:

Investigation of Dihydrofolate Reductase Post-Translational Modifications

5.1: Introduction

5.1.1: The Epigenome beyond DNA methylation

DNA methylation is not the only epigenetic modification capable of altering gene expression without affecting the sequence itself. Although DNA methylation is the most stable and heritable epigenetic modification, the word 'epigenetics' is now typically used in a broader sense than Waddington's original definition of what he called the Epigenetic Landscape in 1942: "the interaction between genes and their products to bring the phenotype into being[205]." Today, epigenetics has become a loose term referring to any modification made to DNA or chromatin structure that impacts gene expression[206]. As a result, many cellular processes such as histones modification (including acetylation, methylation, ubiquitination and SUMOylation)[103,207], nucleosome positioning[208,209] and non-coding RNA localisation[210,211] have all been defined as 'epigenetic modifications[119],' despite their lack of heritability in both a trans-generational and mitotic context.

There is some controversy in the field regarding histone modifications, as a clear molecular mechanism to illustrate the self-propagation of these marks has yet to be elucidated[212]. To further complicate the issue, some cases have been reported where histone methylation has been found to be unstable from generation to generation, while the transcriptional effects established by it remain heritable[213]. A bimodal 'buffer' model has been proposed based on the heterologous nature of histone methylation to explain this. The model proposes that transcriptional inactivation is the result of high levels of methylation at a particular lysine residue – with lysine being capable of being mono-, di-, and tri- methylated. This creates a spectrum of histone methylation across a genetic domain, resulting in transcriptional repression unless a critical 'low' is reached. As a result, the transcriptional effects associated with histone methylation can be inherited intact mitotically, even if methylation levels are not completely stable from generation to generation[214]. This model indicates that post-translational modifications of histones may constitute epigenetic programming to the same extent as DNA methylation. Interestingly, the demethylating agent 5'aza-2'cytidine described in Chapter 4 can impact on both DNA methylation and histone methylation[195].

Another histone modification is acetylation. Histone acetylation – classically associated with gene repression – is directly opposed by enzymatic deacetylation[58]. The dynamic relationship between histone acetylation and deacetylation, coupled with histone methylation, creates an image of what has been called the 'histone code,' – the sum of modifications made to histones that affect expression of their associated genetic material[55].

Larger polypeptides such as ubiquitin and SUMO (small ubiquitin-like modifier) have also been implemented in gene activation and repression, respectively, as a result of histone modification[215]. The complementary relationship between ubiquitin and SUMOylation seems to mirror that of acetylation and deacetylation, which together build a stronger argument for a coherent system of genetic control via histone modification. Even if they cannot be identified as being 'epigenetic' in the strictest sense of the term, these unique and dynamic molecular marks have attracted considerable attention in the field, bringing us closer to decoding the histone code.

It has recently been found that LSD1 – a histone demethylase – is a folate binding protein[216]. With DHFR, an enzyme involved in one-carbon metabolism, localising from the cytoplasm to the nucleus in response to SUMOylation[43], it is possible that a strong link exists between folate, its metabolic enzymes, and gene control. It is already well established that acetylation of non-histone proteins can regulate transcription[217], alluding to the possibility that post-translational modifications of DHFR may have an impact on gene expression.

5.1.2: Post translational modifications of Dihydrofolate Reductase

Non-histone lysine acetylation is one of the most common covalent modification in eukaryotic cells, possibly analogous to phosphorylation: the regulatory "master-switch" of protein biology[218,219].

Broadly speaking, acetylation of a lysine residue neutralises its positive charge, and has been associated with increasing enzyme stability[218]. Some studies have demonstrated nuclear localisation of particular proteins after lysine acetylation[218,220,221], along with SUMOylation[222]. In 2007, Anderson *et al.*, showed that dihydrofolate reductase is capable of being SUMOylated *in vitro*[43].

As described in Chapter 1, DHFR (dihydrofolate reductase) plays a crucial role in folate metabolism, generating the active form of folate utilised from its dietary and synthetic analogues in the presence of NADPH[223]. Due to its significance in cell proliferation, DHFR is the target of various anti-folate chemotherapeutics, such as methotrexate and pemetrexed[28,224].

No evidence has been found for acetylation of DHFR *in vitro*, with Anderson *et al.*, (2007)[43] being the only publication to show *in vitro* SUMOylation of the same enzyme. This, together with recent data suggesting an interaction between folate and LSD1, opens up the possibility that acetylation and SUMOylation compete for the same lysine residue, potentially playing a unique role in the compartmentalisation of enzymes involved in one-carbon metabolism.

5.1.3: Aims and Objectives

Aim:

This chapter aims to investigate the SUMOylation and acetylation capabilities of DHFR to begin to unravel the varied roles it's likely to play during nuclear localisation.

Objectives:

•To identify potential SUMOylation and acetylation sites for DHFR *in silico.*

• To produce a purified GST-tagged DHFR protein.

• To confirm that the purified fusion protein exhibits reductase activity.

• To compare SUMOylation and acetylation of DHFR *in vitro.*

5.2: Methods

5.2.1: *In Silico* Post-translational modification analysis

Two programs were used to carry out our own *in silico* analyses: SeeSUMO (bioinfo.ggc.org/seesumo/) and SUMOplot (http://www.abgent.com/tools/). SeeSUMO uses random forest algorithms to compare up to 40 amino acid residues adjacent to the target lysine against other motifs found to be SUMOylated in the literature – approximately 35% of known SUMOylation sites do not contain the ΨKXE motif[225]. SUMOplot, on the other hand, aligns the input amino acid sequence against a SUMO consensus sequence known to bind to SUMO conjugating enzyme UBC9, incorporating possible amino acid substitutions with those exhibiting similar hydrophobicity in the case of mis-matching[226]. For acetylation sites, the software Scan-X was used, which takes a similar approach as SeeSUMO, but with a position weight matrix used instead of a random forest algorithm[227].

5.2.2: Recombinant protein production

OneShot BL21-DE3 cells (New England Biolabs) were transformed as described in Section 2.3.7 using a GST-tagged DHFR cloned in the pDEST15™ Gateway Cloning vector. Cells were grown overnight and prepared as described in Section 2.3.8 before lysis and fractionation (Section 2.3.9). From this, soluble and insoluble fractions of protein were collected. The

soluble fraction was concentrated and diluted in bacterial lysis buffer as described in Section 2.3.10.

5.2.3: Recombinant protein analysis

Soluble fractions were analysed using SDS PAGE and Western blotting (Section 2.3.4). Blots were probed with rabbit anti-DHFR (Abcam) at 1:10,000 in 5% fat free milk, followed by a HRP-conjugated anti-rabbit secondary antibody (Abcam) at a ratio of 1:50,000 in TBST (Section 2.1.2). DHFR fused to GST has an expected molecular weight of 48 kDa, while the purified recombinant DHFR protein from Sigma-Aldrich (used here as a positive control) has a molecular weight of 25 kDa. Unless otherwise stated, all blots were exposed for 15 – 30 seconds with SuperSignal West Femto Chemiluminescent Substrate (Thermo Scientific) for imaging.

5.2.4: Solubilisation of DHFR-GST insoluble fractions

Following fractionation and Western blotting, it could be seen that most of the DHFR fusion protein was remaining in the insoluble fraction so two strategies to resolubilise the protein were undertaken.

The first has been previously described by Sirawaraporn *et al*., (1993)[151]. Briefly, the protein is denatured using 6M guanidine HCl, and then refolded with 20% glycerol (Section 2.3.11), solubilising the protein in the process. The second method, taken from the Molecular Probes manual for Glutathione agarose (G2879)[228], requires the protein's insoluble fraction to be resuspended in a buffer containing 1.5% N-lauroyl-sarcosine, 25mM triethanolamine, and 1mM EDTA (pH 8.0) (Section 2.3.12). As *in vitro* SUMOylation analysis was to be carried out on resolubilised protein, it was imperative that it retained its enzyme activity, as this an indication that it has refolded into its functional conformation.

5.2.5: Purification of GST-bound DHFR

Purification of recombinant, GST-bound protein was carried out using affinity chromatography with glutathione agarose beads. This process exploits the strong binding affinity between GST and glutathione. Two methods were used to purify GST-bound protein: gravity-flow purification (for larger volumes), and batch purification (for small volumes) outlined in Section 2.3.13 and 2.3.14, respectively. Both processes produced three protein fractions: an unbound fraction, with protein that passed through the column without binding; a wash fraction, with loosely bound protein removed by the wash buffer; and an elution fraction, which should contain the target protein bound to GST. In most cases, the elution

fractions were concentrated using 35 kDa centrifugal filter-units (Millipore) before SDS PAGE analysis (Section 2.3.4). In these instances, the eluted protein has been divided further into three separate fractions; elution-neat, representative of the elution fraction prior to concentration; elution-concentrated, the concentrated form of the former; and elution-runoff, the by-product of the concentration process, which contains all protein molecules < 35 kDa, which excludes DHFR-GST at 48 kDa.

All purified protein samples were stored at -20ºC, and quantified using the Bradford assay described in Section 2.3.15.

5.2.6: In vitro SUMOylation

All *in vitro* SUMOylation reactions were carried out using the Active Motif SUMO-link kit, following the manufacturer's instructions with any variations outlined in Section 2.3.17. Briefly, three reactions were set up for each protein of interest (POI): POI + SUMO, to test if the protein of interest is capable of being modified; POI + *mut*SUMO, which contains the protein of interest in the presence of a mutant SUMO molecule with a lower capacity to bind to its target; and a no-protein negative control, with no protein of interest present in the reaction. The negative control contains all the required components of the reaction, including the active SUMO protein.

Recombinant p53 was supplied in the kit as an internal positive control, along with anti-p53 and anti-SUMO antibodies. Antibody concentrations for these were prepared as per the kit protocol (1:5000 and 1:4000, respectively), with Western blots otherwise carried out as described in Section 2.3.4.

5.2.7: Dihydrofolate Reductase Activity Assay

Activity of purified and partially purified DHFR was analysed using the Dihydrofolate Reductase kit from Sigma-Aldrich (CS0340). The kit uses the decrease in absorbance at A_{340nm} that occurs as NADPH is utilised by DHFR to measure the rate of the reaction. Taking into account the change in OD over time with the concentration of enzyme used and the extinction coefficient of the reaction, the specific activity of DHFR can be calculated. This calculation is summarised in the following formula:

$$\text{Units/mg P} = \frac{(\Delta\text{OD/min.sample}) - (\Delta\text{OD/min.blank}) \text{ X d}}{12.3 \text{ x V x mg P/ml}}$$

Where ΔOD/min represents the change in optical density over time; sample and blank refer to the ΔOD of the protein of interest and the same reaction without DHF, respectively; d refers to the enzyme's dilution factor; V and mg P/ml refer to the volume and concentration of protein used, respectively. Finally, 12.3 is the extinction co-efficient of this reaction.

Details of the protocol itself are described in Section 2.3.16. As mentioned previously, the blank used in the formula contains all the components of the reaction except for DHF in order to control for the background reduction of NADPH independent of DHFR. Inhibition assays were carried out on samples containing partially purified protein to control for impurities that may also reduce NADPH to NAD^+. These assays involved the addition of methotrexate into the test reaction to see if the decrease in absorbance was completely arrested, thus confirming all changes previously observed were caused by DHFR activity alone.

5.3: Results

5.3.1 *In silico* analysis of DHFR post-translational modifications

In silico analysis was carried out on the DHFR protein sequence to determine if it contained potential SUMOylation sites. Anderson *et al.,* (2007) concluded that DHFR contains the known SUMOylation motif ΨKXE conserved across numerous species[43], where Ψ is any hydrophobic amino acid, K is the lysine to be SUMOylated, X is any other amino acid, and E is glutamic acid. The lysine they found to be SUMOylated was K179.

Both programmes (SeeSUMO and SUMOplot, Section 5.2.1) found lysine K81 to have the highest probability of being SUMOylated, with a confidence level of 43.68 from SeeSUMO and a prediction score of 0.8 from SUMOplot (Figure 5.1). Scan-X found K179 to have the highest match for acetylation (Scan-X score of 6.156).

A Clustal alignment of DHFR found the K81 SUMOylation motif to be conserved across mammalian species, along with K179 SUMOylation motif described by Anderson *et al.,* (2007)[43]. The K179 acetylation motif – being longer than the SUMOylation motif – does not share the same level of sequence homology across multiple species (Figure 5.2).

5.3.2: DHFR is primarily found in the insoluble fraction when the human protein is expressed by BL21-DE3 cells

After lysis and fractionation of BL21-DE3 cells transformed with the pDEST15™ Gateway Cloning vector (Section 5.2.2), a significant portion of DHFR-GST was found to be in the

insoluble fraction when analysed via Western blot. This may be the result of the protein being misfolded and stored as inclusion bodies when produced too quickly for the cells to process properly. Attempts to optimise the induction of DHFR-GST by varying temperature, time, and concentration of IPTG did not improve solubility. As the expression vector pDEST15 with the *DHFR* gene contains a "leaky" promoter – producing a relatively small amount of the protein when uninduced – removing the induction step entirely resulted in the production of a small amount of soluble DHFR-GST (Figure 5.3). Large cultures of bacteria (2L) were used to compensate for the low levels of soluble protein produced when the culture is not induced.

5.3.3: Glutathione agarose purification of soluble DHFR and Western blot analysis

Eluted fractions from purified DHFR-GST were concentrated and 150ng of protein was run out on an SDS-PAGE gel with a positive control: the DHFR supplied by Sigma-Aldrich's Dihydrofolate Reductase kit. Following Western blotting as outlined in Section 2.3.4 and under optimised exposure for the DHFR antibody (15-30 seconds) no DHFR-GST was visible on the blot, though a clear band at 25 kDa representing the positive control was visible. By increasing the exposure time to 5 minutes, 48 kDa bands indicative of DHFR-GST were visualised for the concentrated protein fraction (Figure 5.4). The concentration of DHFR-GST was found to be 0.0031µg/µl when quantified by Bradford assay (Figure 5.5). *In vitro* SUMOylation analysis requires 0.5µg in a 500µl reaction.

5.3.4: Resolubilisation of DHFR insoluble fraction with Guanidine-HCl and Glycerol

When 2L of bacterial culture did not produce enough soluble DHFR-GST to carry out the SUMOylation experiments, two methods for recovering DHFR-GST from the insoluble fraction were identified from the literature.

The first[151], described in Section 5.2.4, involves denaturing and refolding the protein in two separate steps. In both the soluble and insoluble fractions generated during the refolding process, the majority of DHFR-GST remained insoluble while a small portion was successfully solubilised (Figure 5.6).

Purification of this soluble fraction, however, resulted in the same issue as before. Each fraction obtained from the purification process was analysed on a Western blot probed with DHFR antibody. Significant portions of DHFR-GST were present in the unbound and wash fractions, with very little in the eluted fraction, even when concentrated (Figure 5.7). Bradford

analysis on this product found the protein concentration to be too low for SUMOylation analysis (0.00182µg/µl).

5.3.5: Resolubilisation of DHFR with N-lauroyl-sarcosine and triethanolamine

Since the first resolubilisation method didn't produce enough protein for the SUMOylation experiment, a different method was undertaken[228]. The insoluble fraction containing DHFR was resuspended in a 'resolubilisation buffer' containing N-lauroyl-sarcosine and triethanolamine (Section 5.2.4) and subjected to glutathione batch-purification. Fractions obtained from the purification process were analysed on a Western blot probed with anti-DHFR antibody. Strong bands in the unbound and wash fractions indicate that very little DHFR-GST has bound to the column, further demonstrated by the weak band in the eluted-concentrated fraction (Figure 5.8).

As the protein visible in Figure 5.8 was from 14% of the total 2L culture, one final attempt to purify soluble DHFR was carried out by combining all of the remaining pellets from this culture. When purified the resolubilised protein failed to bind to the glutathione column. However, a significant portion of DHFR-GST was found in the wash fraction, in comparison to those eluted and concentrated (Figure 5.9). As 18µl of 25ml was loaded into lane 3 (in comparison to 18µl of 300µl in lane 5), this wash fraction was taken and concentrated in a 35 kDa centrifugal filter-unit to be analysed on a separate gel.

Figure 5.10 shows this concentrated wash-fraction with a significant portion of DHFR when compared to a positive control. ImageJ – a semi-quantitative image processing programme – was used to estimate the concentration of protein in this band. The process involved comparing the intensity of the DHFR-GST band to that of the positive control. With the concentration of the recombinant DHFR from Sigma-Aldrich known to be 0.044µg/µl, the concentration of DHFR-GST was estimated at 0.032µg/µl. In a total volume of 1ml, a total yield of 32µg partially purified DHFR has been successfully produced from 2L of bacterial culture.

When subjected to the DHFR reductase assay, the purified enzyme was found to be inactive. In a reaction containing DHF and NADPH, no depletion of NADPH was observed over the course of 2 minutes with DHFR-GST (Figure 5.11). In comparison to the positive control supplied by the kit with a specific activity of 2.822µmol/min/mg, recombinant DHFR-GST was found to have activity of 0.203µmol/min/mg. When examined in the presence of methotrexate – an anti-folate that inhibits DHFR – the activity of the reaction was found to be 0.121µmol/min/mg, representative of background NADPH depletion. Subtracting these two values gives the negligible value of 0.083µmol/min/mg for recombinant DHFR-GST activity.

5.3.6: *In vitro* SUMOylation of DHFR

As reductase activity could not be retained with the solubilisation of DHFR-GST, we moved forward with the *in vitro* SUMOylation experiments on commercially available DHFR. The recombinant DHFR used in these experiments was supplied by Sigma-Aldrich – the same DHFR enzyme used as a positive control on the Western blots and enzyme kinetics assays throughout this chapter.

The SUMOlink kit supplied by Active Motif uses a set of SUMOylation and conjugating enzymes to add a SUMO moiety to the protein of interest, which can be visualised on a Western blot as a 15 kDa band shift. A mutant SUMO protein (*mut*SUMO) incapable of forming a thioester with UBC9 – which required for UBC9-catalysed SUMOylation[43] – acts as a negative control. A second negative control recommended by the kit includes all of the required components for SUMOylation with the protein of interest omitted. This is to control for self-SUMOylation of the kit's own components.

On a Western blot probed with DHFR antibody, a strong shift of 15 kDa was present in a DHFR-SUMO reaction when compared to a reaction containing DHFR and *mut*SUMO. In a blot probed with a SUMO antibody, the same 40 kDa band appeared in the DHFR-SUMO reaction, with 15 kDa bands representing free SUMO. On the no-protein negative control of this blot, however, the same 40 kDa band was visible, along with many other bands consistent with self-SUMOylation of the kit's components. The free-SUMO band at 15 kDa appears fainter here, further confirmation that SUMOylation has taken place (Figure 5.12).

The band at ~40 kDa in the no-protein negative control complicates these results. The nature of the band at this size in the DHFR-SUMO reaction of the anti-SUMO gel is now unclear, as this could be the result of the same self-SUMOylation event that produced the band in the no protein negative control lane. The role of the anti-SUMO Western was to confirm that the shift observed in the anti-DHFR gel was indeed caused by SUMOylation. As a result of this inconsistency, it cannot be concluded that the shift of the DHFR band was caused by SUMOylation alone.

A gel from another SUMOylation experiment was stained with coomassie blue. The image illustrates the range and quantities of other proteins present in the reaction. Bands for DHFR (25 kDa) and p53 (53 kDa) are visible, along with many other bands representing the conjugating and SUMOylating enzymes of the reaction, including one at ~18 kDa in each lane, possibly for UBC9 (Figure 5.13).

The second SUMOylation experiment aimed to examine the effects of this post-translational modification on a protein known to be capable of being SUMOylated *in vitro*. Active Motif provides aliquots of recombinant p53 protein with the SUMOylation kits for this very purpose. When probed with anti-p53, a shift consistent with what was depicted in the SUMO-link manual was observed – from 53 kDa to ~70 kDa. Likewise, the position of this SUMOylated protein is confirmed on the anti-SUMO blot. Note that there is a band present in this location on both the DHFR-SUMO1 and no-protein negative lanes, but the p53-SUMO lane is the only one to contain a doublet, suggesting one band is SUMOylated p53, and the other a consistent, non-specific, self-SUMOylation event (Figure 5.14)

All of the DHFR experiments were carried out on Sigma-Aldrich's commercially available recombinant DHFR. The enzyme was supplied at a concentration of ~0.044µg/µl. In order to carry out a 20µl reaction with 0.5µg total DHFR, the protein was vacuum-concentrated and resuspended in H_2O. It was hypothesised that this drying and resuspending process may damage the enzyme thereby eliminating its SUMOylation capacity. However, an enzyme activity assay (Figure 5.15) shows that even when processed in this manner, DHFR still retains its dihydrofolate reductase activity (4.3µmol/min/mg before drying down, 2.13µmol/min/mg afterwards).

All of the above suggests that the components of the kit are being SUMOylated, with one in particular exhibiting a shift in molecular weight similar to DHFR. Two of the components of the kit, activating enzyme UBA2 and conjugating enzyme UBC9, have molecular weights of 72 kDa and 18 kDa, respectively (Uniprot). An 18 kDa band for what could be UBC9 is visible in Figure 5.13, and close enough in mass to DHFR that the SUMOylation of the former could be mistaken for that of the SUMOylation of the latter. An *in silico* analysis of UBC9 showed that not only is this conjugating enzyme in possession of SUMOylation motifs, but scores higher than DHFR as a candidate for SUMOylation according to both algorithms (Figure 5.16).

5.3.7: *In vitro* Acetylation of DHFR

Given the difficulties in purifying active DHFR-GST, an *in vitro* acetylation assay could not be optimised. Using the Sigma-Aldrich recombinant DHFR, a "one-shot" experiment was carried out following the protocol described by Levy *et al.,* (2004)[229], using commercial p300 acetlytransferase and [^{14}C]acetyl-CoA. Proteins were resolved on an SDS-PAGE gel, but no bands were observed with auto-radiographic imaging (data not shown).

5.4: Discussion

Post-translational modifications such as SUMOylation and acetylation have been implicated in a variety of biological processes such as localisation and degredation[218,219,222]. In 2007, it was discovered that DHFR – an enzyme with a pivotal role in one-carbon metabolism – is capable of being SUMOylated *in vitro*[43]. Our own *in silico* analysis on DHFR found that the same lysine residue found to be SUMOylated by Anderson *et al.,* (2007) can also be acetylated, opening up the possibility that both these post-translational modifications are in direct competition with one another for DHFR.

The *in silico* analysis completed here found K81 to have the highest potential to be SUMOylated, with K179 (cited by Anderson *et al.*, (2007)[43]) not appearing in either analysis. This may be due to the fact that leucine – the amino acid at the Ψ position of the K81 site – is far more hydrophobic at pH 7 than tyrosine, the Ψ amino acid at the K179 site. Anderson *et al.*, (2011), however, has put more weight on K179, as the ΨKXE motif is conserved across multiple species[43], this was also found to be the case for K81.

The aim of this work was to first produce recombinant DHFR, then carry out *in vitro* SUMOylation and acetylation analysis. Achieving this first objective, however, proved to be more difficult than anticipated.

The major issues in producing recombinant GST-tagged DHFR were caused by the inability of the protein to fold correctly when expressed in BL21-DE3 cells. Initially, the vast majority of the protein was produced in the form of insoluble inclusion bodies, a typical response for bacteria being forced to express non-bacterial proteins at a high level[230]. By increasing the size of the bacterial cultures, enough DHFR was produced in the soluble factions for use in the intended downstream analysis.

However, efficient purification of DHFR from the soluble fraction could not be obtained, due a lack of binding between GST and the glutathione column. After many attempts to re-solubilise the aforementioned insoluble fractions, the same problem persisted: no binding between GST and glutathione was taking place implying the tag was misfolded and thereby unavailable to the resin. The SUMOylation analysis required 500ng protein in a 20µl. The most protein successfully purified in was ~3ng/µl in a volume of ~100µl.

It was initially speculated that this was also due to misfolding of DHFR while it was processed as an inclusion body. Analysis of partially-purified DHFR that failed to bind to the glutathione

column found the enzyme to have no reductase activity, further proof that GST-DHFR fusion proteins have been misfolded.

Although McEntee *et al.*, (2011)[40] generated enough GST-DHFR for enzyme kinetic analysis, the concentrations required for the SUMOylation experiments could not be obtained. Other publications have had more success with generating recombinant DHFR through His-nickel binding[43,231,232].

The failure of recombinant proteins to fold correctly is a major obstacle in the field, even though the exact mechanism for inclusion body formation is poorly understood[233]. As a result, methods for correctly refolding proteins are typically tailored to the target protein itself. Still, most strategies in the field follow the general protocol of using the first resolubilisation process described here[233].

There are plenty of other methods available, but not necessarily relevant to DHFR. For example, proteins that contain cysteine residues may form disulfide bonds as part of their three-dimensional structure. In the cytoplasm of *E.coli*, however, disulfide bonds could form erroneously. This can be corrected by adding a reducing agent like DTT to the lysis buffer, and preventing the formation of disulfide bonds until later on in the purification process[234]. DHFR, however, does not form disulfide bonds, so this is a non-issue for this protein[235].

Another way to remove protein from bacterial inclusion bodies is to expose insoluble fractions to high hydrostatic pressure[236]. Commercially available pressure-cells can be used to exert a large amount of pressure on the aggregates to free protein from the inclusion bodies without damaging the enzyme itself. Methods based on this technology, however, require specialised equipment.

As no recombinant protein was obtained, the effect of *in vitro* SUMOylation on commercially purchased DHFR was examined. The results of this experiment were inconclusive. While the appropriate band shift was observed (Figure 5.14), the negative control indicated that the putative SUMOylated DHFR band also contained a mixture of self-SUMOylated proteins.

The manufacturers of the kit recommended including a no-protein negative control in all experiments, due to the reaction's susceptibility to produce false positives. Indeed, SUMOylation of UBC9 – a conjugating protein involved in the reaction – manifests as a shift identical to what would be expected with DHFR. By taking into account evidence from anti-SUMO and anti-p53 Western blots, it is clear that shifts observed from the former cannot be interpreted to be caused by SUMOylation of DHFR if a similar shift occurs in the no-protein

lane. Furthermore, shifts in anti-DHFR Western blots cannot be conclusively interpreted to represent SUMOylation events without a complementary shift in anti-SUMO blots. Although a shift from ~25kD to ~40kD was observed when probed with anti-DHFR, this shift being caused by some other artefact – like a self-dimer of DHFR – cannot be ruled out. Furthermore, self-SUMOylation of K14 of UBC9 as predicted by the SeeSUMO software may be the source of the band in the no protein negative control. UBC9 has also been established to be capable of SUMOylation *in vitro*[237]. To fully prove that the DHFR ~40kD band shift observed is due to SUMOylation, mass spectrometry could be carried out on the purified equivalent sized band from the anti-SUMO blot.

The *in silico* acetylation analysis of DHFR found that K179 – the same lysine shown to be SUMOylated by Anderson *et al*., (2007)[43]– is also a potential acetylation site. There are no commercially available kits for *in vitro* acetylation analysis, with methods described in the literature using recombinant p300 acetyltransferase produced in-lab with acetyl coenzyme A bound to ^{14}C, with detection carried out using auto-radiographic imaging or scintillation counting used for detection[238]. Although this method was attempted, given the limited amount of available protein and time restriction, the technical details of the method could not be optimised to address whether DHFR is in fact acetylated *in vitro*.

Although we couldn't prove that DHFR is subject to both SUMOylation and acetylation *in vitro,* this area is still warrants further investigation. Our *in silico* analyses shows that the K179 residue of DHFR is susceptible to acetylation: the same lysine stated to be SUMOylated by Anderson *et al.,* (2007)[43]. Similarly, both acetylation and ubiquitination have been shown to compete for the same lysine of p53, with the former found to increase its DNA binding affinity[239]. Ubiquitination of p53, however, marks the protein for degradation[240], which can be prevented by acetylation, hinting towards a complex system of competing post-translational modifications. If applied to DHFR, this suggests that SUMOylation of DHFR would localise it to the nucleus, with acetylation then triggering a yet unknown role in gene control.

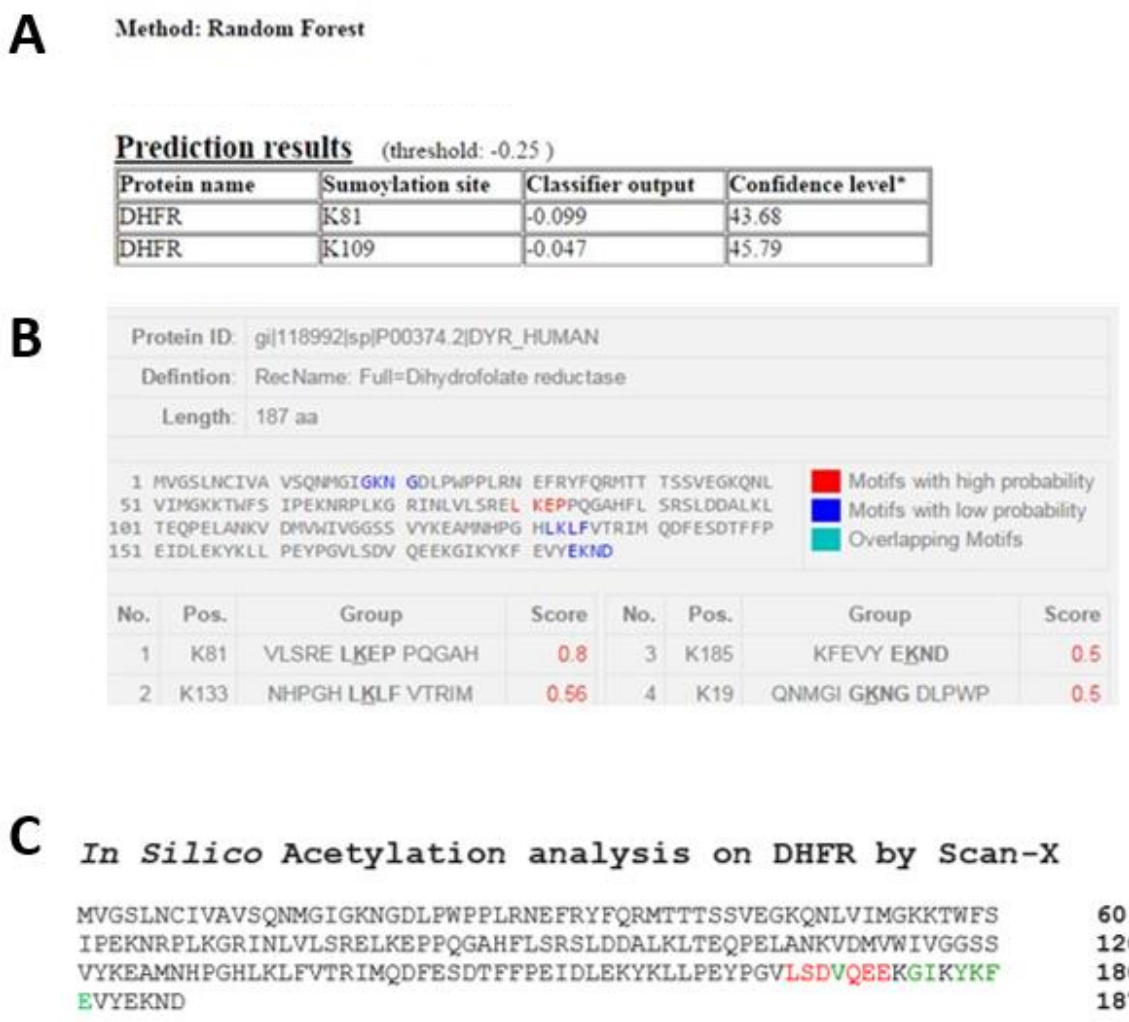


Figure 5.1: *In Silico* Post Translation Modification Analysis of DHFR

The peptide sequence of DHFR was examined for potential SUMOylation sites using SeeSUMO (A) and SUMOplot (B). Both the K81 and K109 lysine residues were predicted to be SUMOylated according to SeeSUMO, with K81 scoring highest for SUMOplot. *In silico* acetylation analysis was carried by Scan-X (C), which found K179 to be a potential acetylation site. K179 was also demonstrated to be capable of SUMOyaltion according to Anderson *et al.*, (2007)[43]. Green amino acids are represented in the motif, with those in red under-represented..

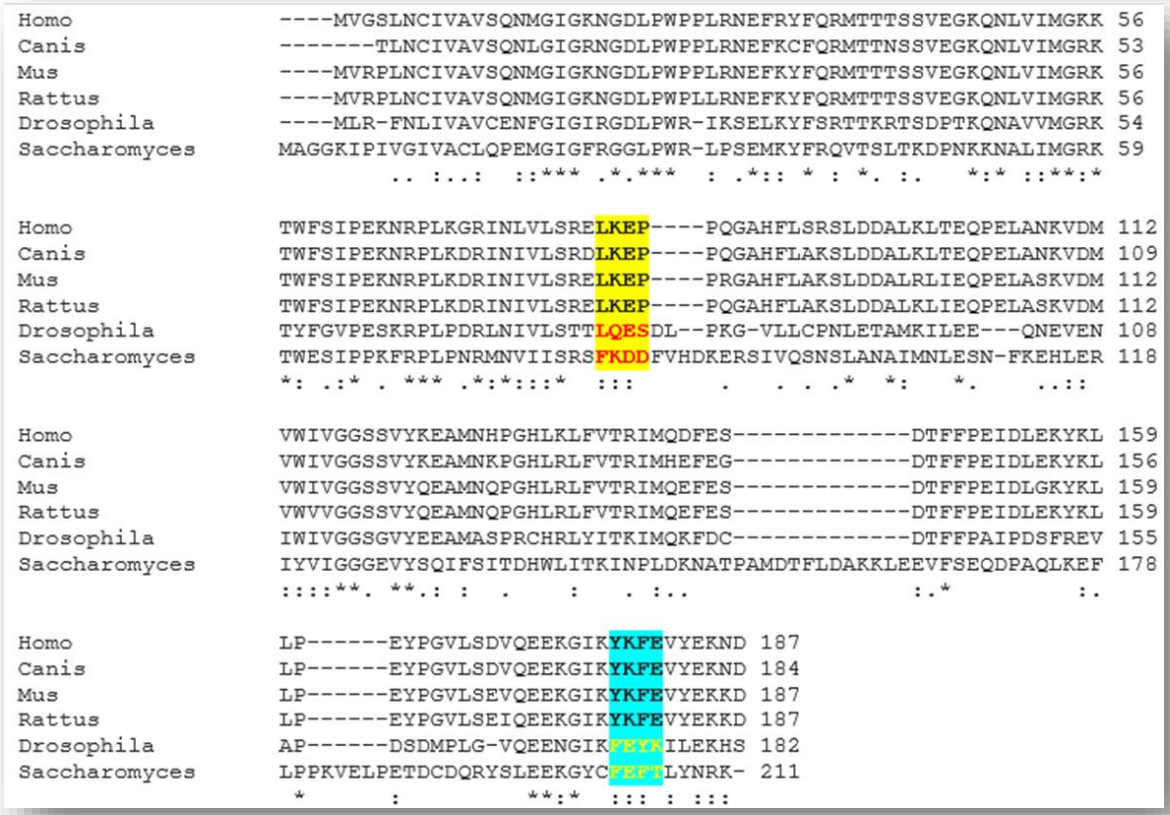


Figure 5.2: Peptide sequence alignment and conserved SUMOylation sites in mammals

The SUMOylation motif for K179 (blue) as described by Anderson *et al*[43]., is conserved across *Homo sapiens, Canis familiaris, Mus Musculus* and *Rattus norvegicus* DHFR peptide sequences. The motif for K81 (yellow) as identified by SeeSUMO and SUMOplot is also conserved across the same species. The acetylation site identified by Scan-X does not contain the same level of homology across the mammalian species, though the region that differs was not found to be representative of an acetylation motif from Scan-X analysis.

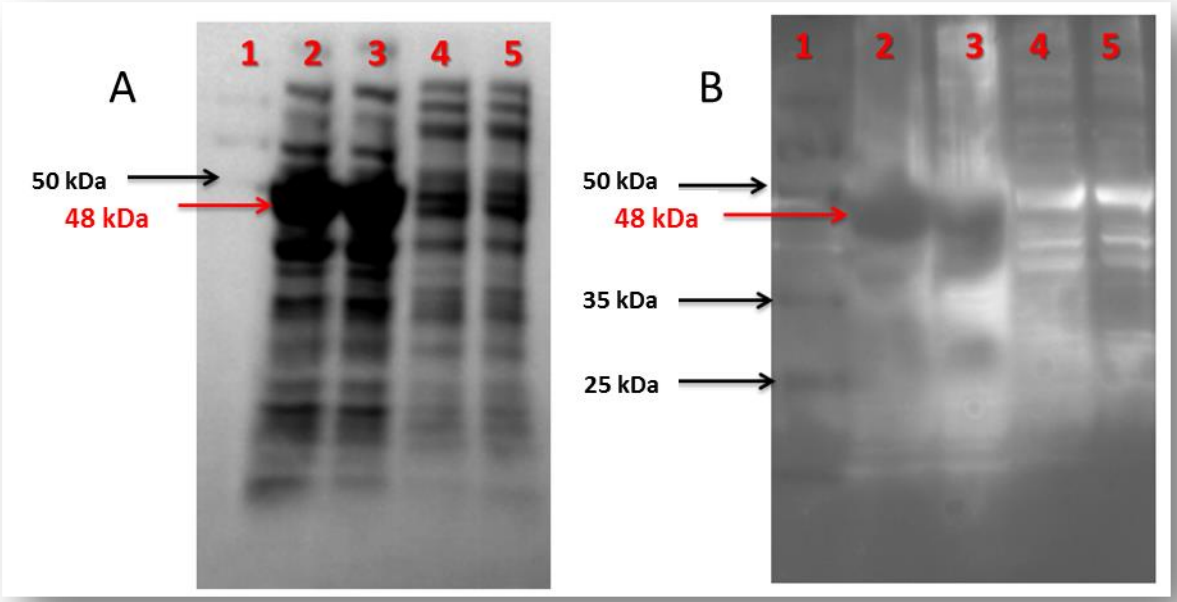


Figure 5.3: Soluble and insoluble fractions of BL21-DE3 cell lysates

Coomassie stained PAGE SDS gel (A) and Western blot probed with DHFR antibody (B) with insoluble fractions (lanes 2 and 3) and soluble fractions (lanes 4 and 5) of BL21-DE3 cell lysates. The expected band-size for DHFR-GST is 48 kDa, which is present in both images. A higher concentration of our protein is evident in the insoluble fraction. When exposed for 15 – 30 seconds, very high levels of protein in the insoluble fraction caused scorching of the membrane, which appears as a black band in lanes 2 and 3.

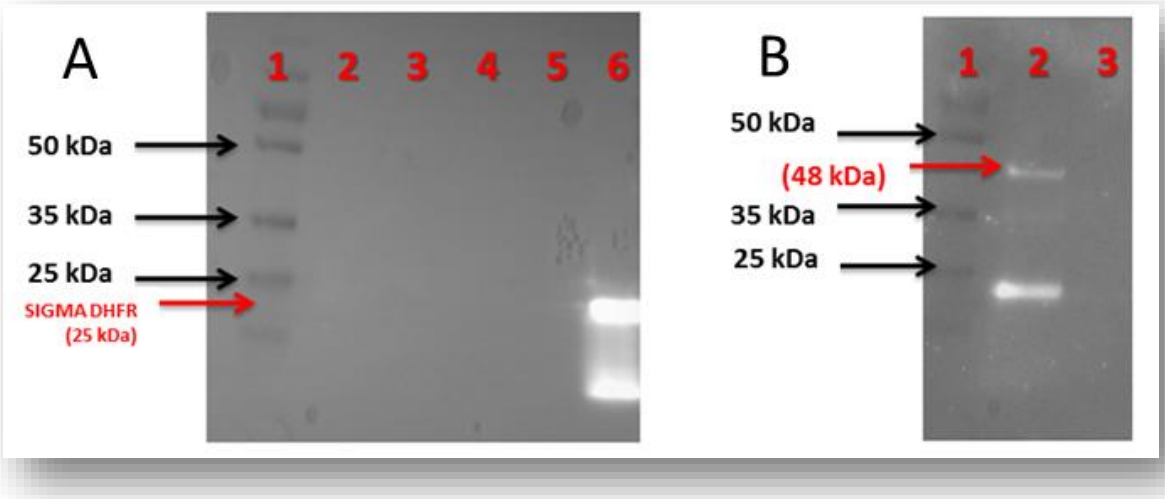


Figure 5.4: Eluted soluble DHFR

Western blot probed with DHFR antibody. (A)At a normal exposure (15 – 30 seconds), no bands for DHFR-GST are visible for the eluted-concentrated protein (lane 2) and eluted-unconcentrated protein (lane 3). A strong band for recombinant DHFR (positive control Sigma-Aldrich) is visible in lane 6. (B) At prolonged exposure (5 minutes), a weak band for DHFR-GST is visible for the eluted-concentrated protein (lane 2) but not for the eluted-unconcentrated protein (lane 3).

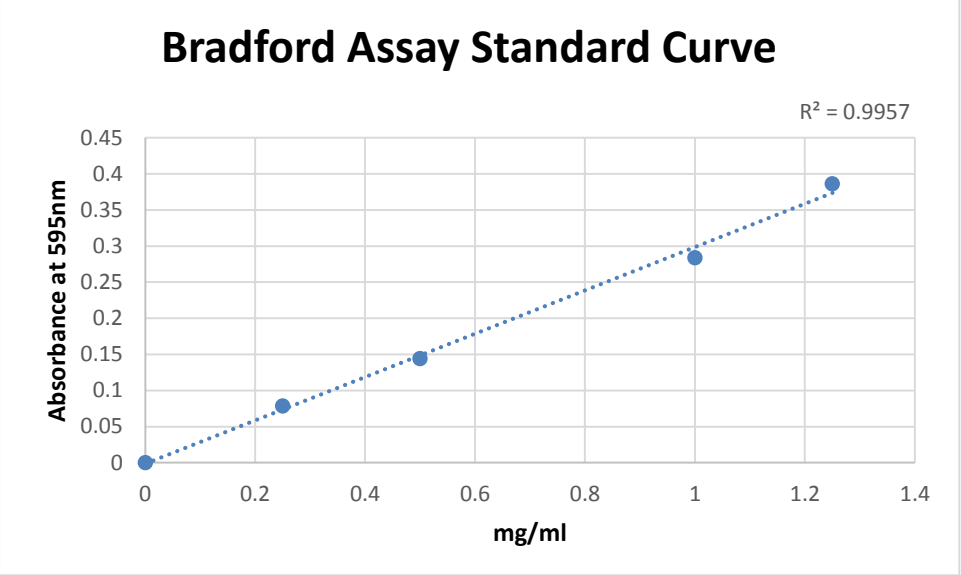


Figure 5.5: Standard Curve for Bradford Assay

A sample standard curve on a set of BSA standards (0.25mg/ml – 1.25mg/ml) measured on a Tecan Plate Reader at A_{595nm}. Only standard curves with R^2 values > 0.99 were used to determine protein concentration.

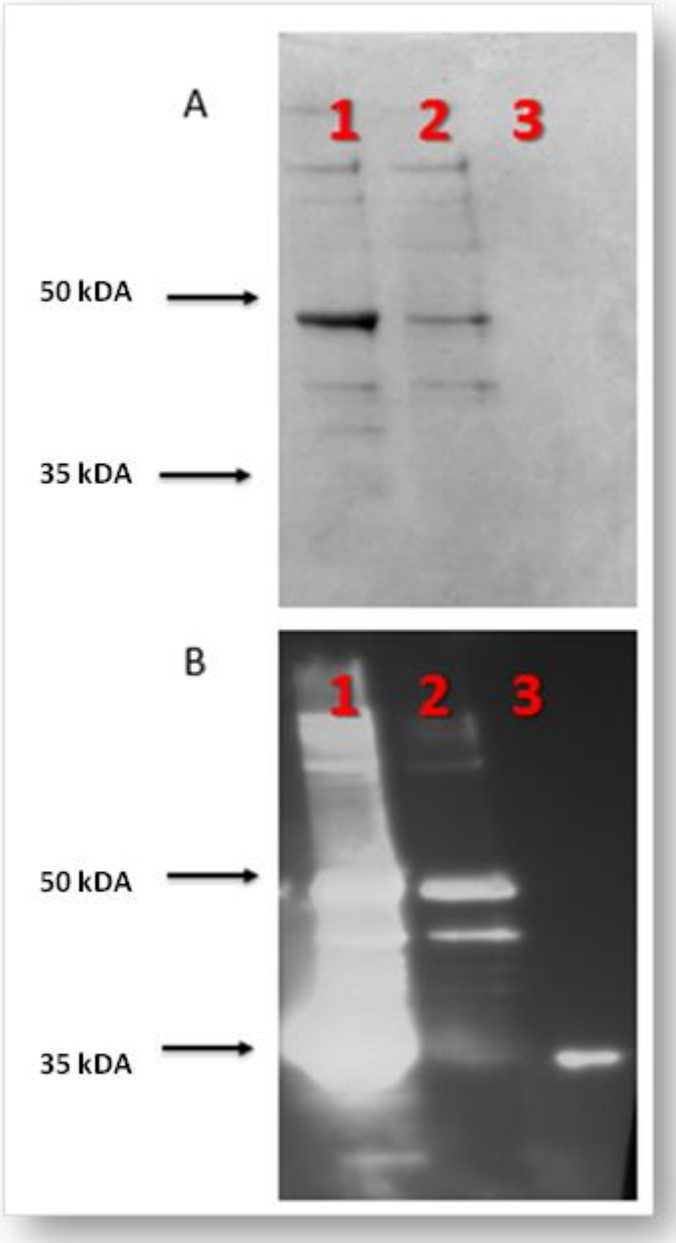


Figure 5.6: Resolubilisation of DHFR using Guanidine-HCl and Glycerol

Coomassie stained SDS PAGE gel (A) and Western blot probed with DHFR antibody (B). Resolubilisation of DHFR-GST generated a new insoluble fraction (lane 1) and soluble fraction (lane 2). Majority of recombinant protein lies in the insoluble fraction, with a strong band at ~48 kDa. Sharp bands at the same molecular weight in lane 2 indicate DHFR has been successfully resolubilised. Recombinant DHFR from Sigma-Aldrich (25 kDa) was used as a positive control (lane 3).

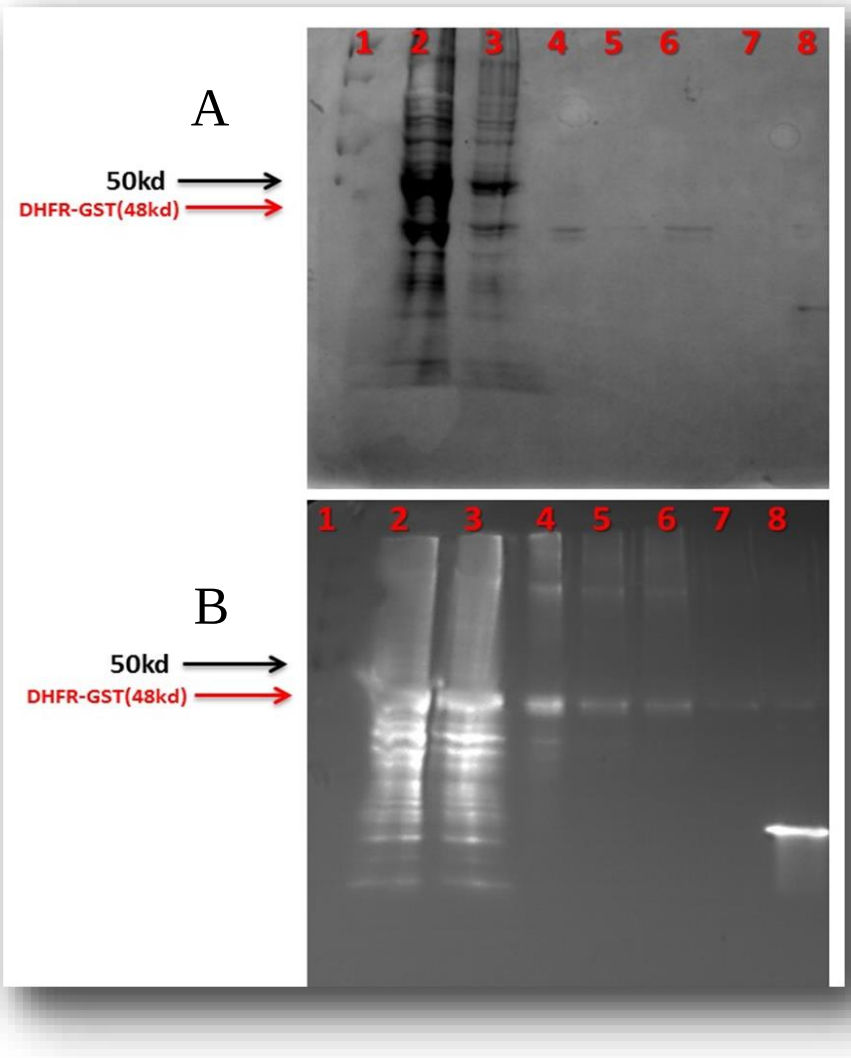


Figure 5.7: Purification of resolubilised DHFR-GST using Glutathione Agarose Beads

Coomassie stained SDS PAGE gel (A) and Western blot probed with DHFR antibody (B). Images depict products from gravity-flow affinity chromatography of soluble DHFR-GST with glutathione agarose beads. A strong band at ~48 kDa in the unpurified fraction (lane 2) indicates very high levels of protein. The unbound fraction (lane 3) is almost identical to the unpurified fraction. The wash fraction (lane 4) contains a weaker band at 48 kDa, as is the case in the eluted-unconcentrated (lane 5) the eluted-concentrated (lane 6) and the concentration-runoff fractions (lane 7). Recombinant DHFR from Sigma-Aldrich was used as a positive control (lane 8).

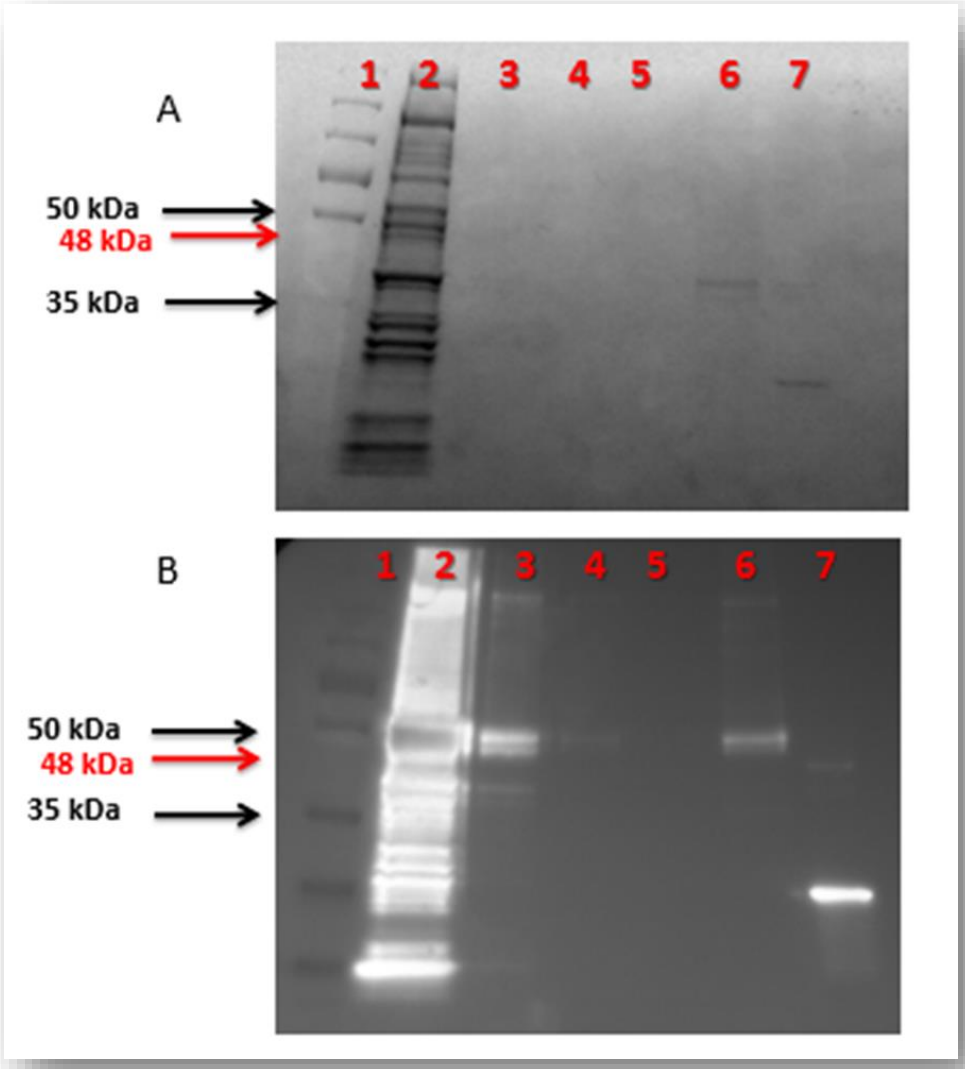


Figure 5.8: Glutathione purification of DHFR-GST resolubilised with N-lauroyl-sarcosine and triethanolamine

Coomassie stained SDS PAGE gel (A) and Western blot probed with DHFR antibody (B). Images depict products from batch affinity chromatography of soluble DHFR-GST with glutathione agarose beads. A scorched band in unbound fraction (lane 2) indicates very high levels of GST-DHFR and other protein from non-specific binding of the secondary antibody. The wash fraction (lane 3) contains a weak 48 kDa band. No protein is visible in the eluted-unconcentrated (lane 4) the concentration-runoff (lane 5) and the eluted-concentrated fractions (lane 6). Recombinant DHFR from Sigma-Aldrich was used as a positive control (lane 7).

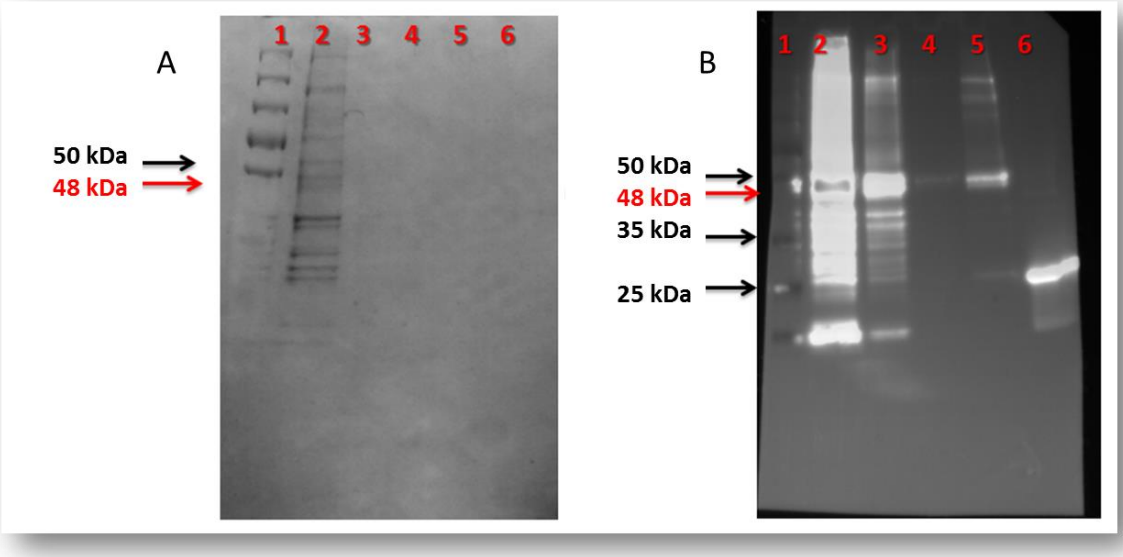


Figure 5.9: Glutathione column-purification of combined DHFR-GST pellets resolubilised with N-lauroyl-sarcosine and triethanolamine

Coomassie stained SDS PAGE gel (A) and Western blot probed with DHFR antibody (B). Images depict products from gravity-flow affinity chromatography of soluble DHFR-GST with glutathione agarose beads. A scorched band in unbound fraction (lane 2) indicates very high levels of GST-DHFR and other protein from non-specific binding of the secondary antibody. The wash fraction (lane 3) contains a strong 48 kDa band with some low molecular weight non-specifics. Eluted-unconcentrated (lane 4) and eluted-concentrated (lane 5) contain weak 48 kDa bands in comparison to the recombinant Sigma-Aldrich DHFR used as a positive control (lane 3).

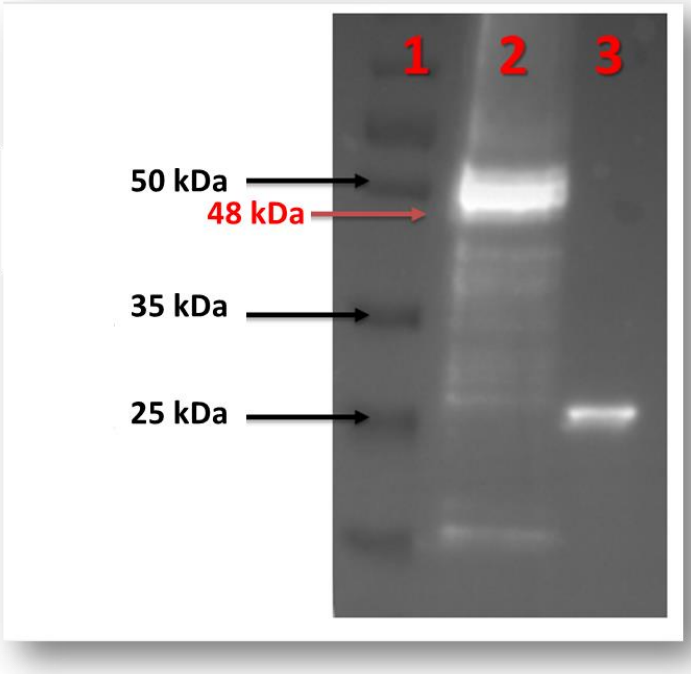
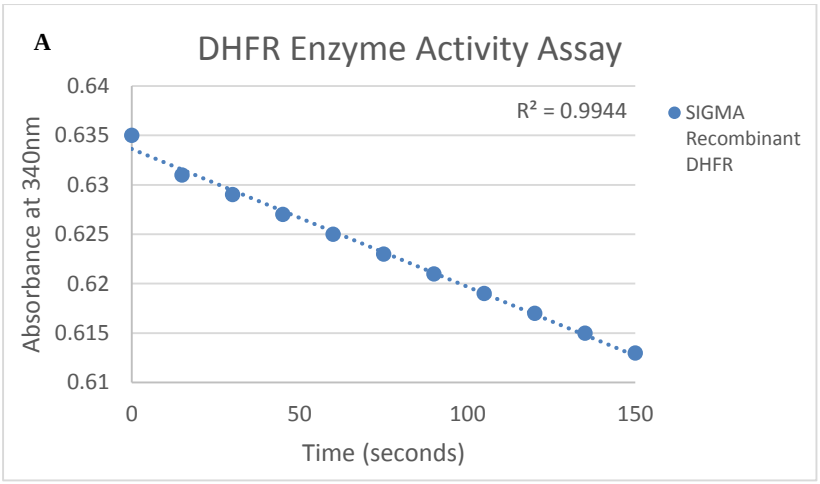


Figure 5.10: Concentration DHFR-GST wash fraction from protein resolubilised with N-lauroyl-sarcosine and triethanolamine

Western blot probed with DHFR antibody. This image depicts the concentrated DHFR-GST wash fraction obtained from gravity-column purification of protein resolubilised with N-lauroyl-sarcosine and triethanolamine (lane 2). A strong band at 48 kDa indicates DHFR-GST has been partially purified. Recombinant Sigma-Aldrich DHFR (25 kDa) used as a positive control (lane 6).



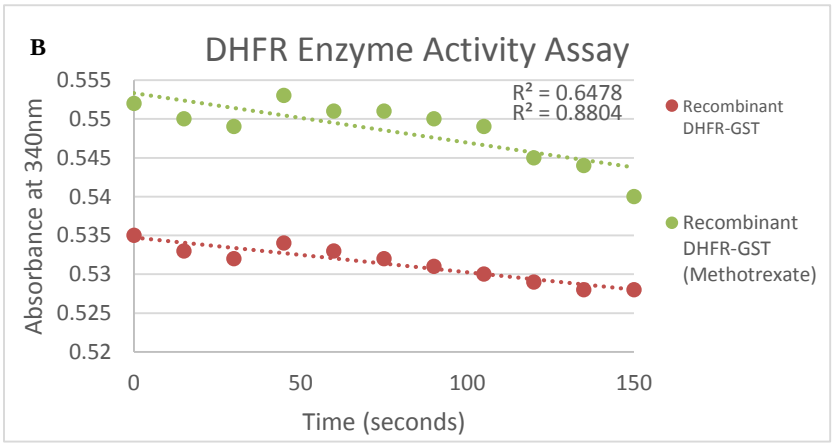


Figure 5.11: DHFR Enzyme Activity Assay on DHFR-GST Partially Purified from Resolubilised Protein Wash Fraction

Enzyme Activity of DHFR-GST partially purified from resolubilised protein wash fraction was analysed using Sigma-Aldrich's Dihydrofolate Reductase Assay Kit. Activity was measured for recombinant DHFR supplied with the kit (A) and recombinant DHFR-GST (B). An inhibition assay was also carried out for recombinant DHFR-GST (B) using methotrexate. A linear depletion NADPH over time was observed for recombinant DHFR (R^2 = 0.9944), but not for DHFR-GST (R^2 = 0.6478). The inhibition assay for DHFR-GST exhibited an absence of NADPH depletion similar to DHFR-GST alone.

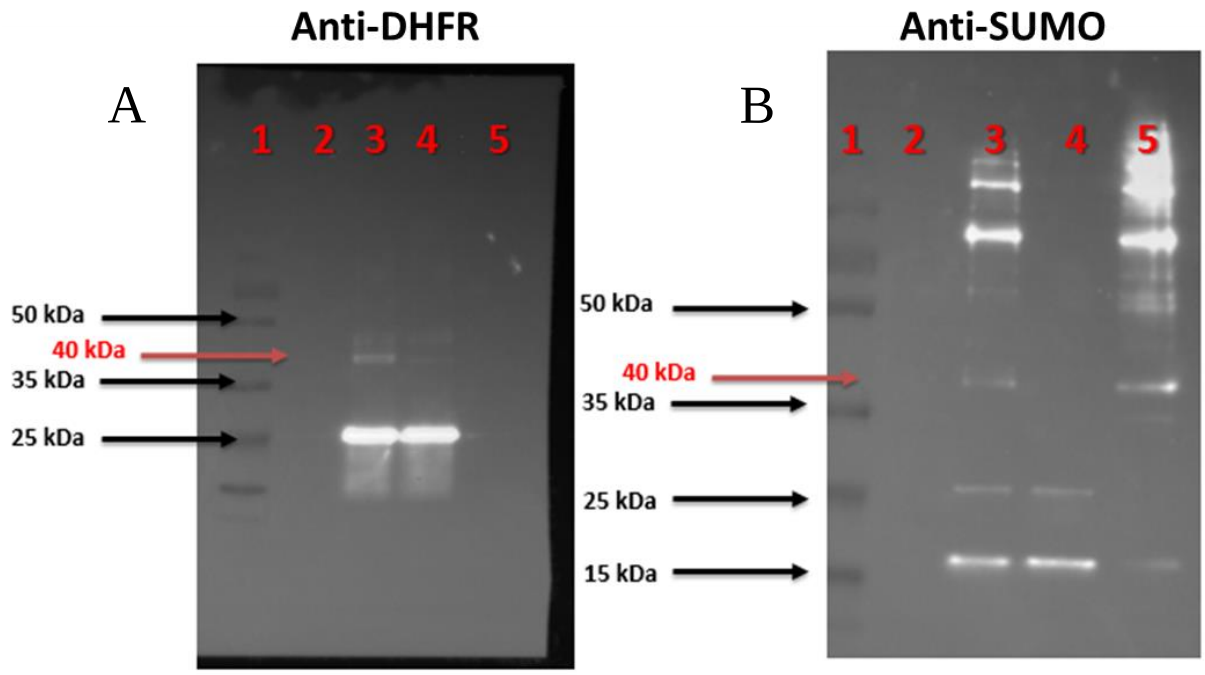


Figure 5.12: Products of *In vitro* SUMOylation assay of DHFR probed with anti-DHFR and anti-SUMO

The SUMOylation of Sigma-Aldrich's DHFR enzyme (used previously as positive control) was examined using an *in vitro* assay. Three reactions were included in each experiment. One containing DHFR with SUMO (lane 3), one with a mutant SUMO protein used instead of wild-type SUMO (lane 4), and one with the protein of interest omitted to control for self-SUMOylation of the conjugating enzymes (lane 5). Two Westerns were carried out, probed with DHFR antibody (A) and with SUMO1 antibody (B). In the anti-DHFR blot, a shift of ~15 kDa is present in the test sample (lane 3), while being less prominent in the mutant SUMO control (lane 4). Multiple Bands in no-protein negative control suggest self-SUMOylation of the reaction's components. Bands for free SUMO are also present on this blot at 15 kDa.

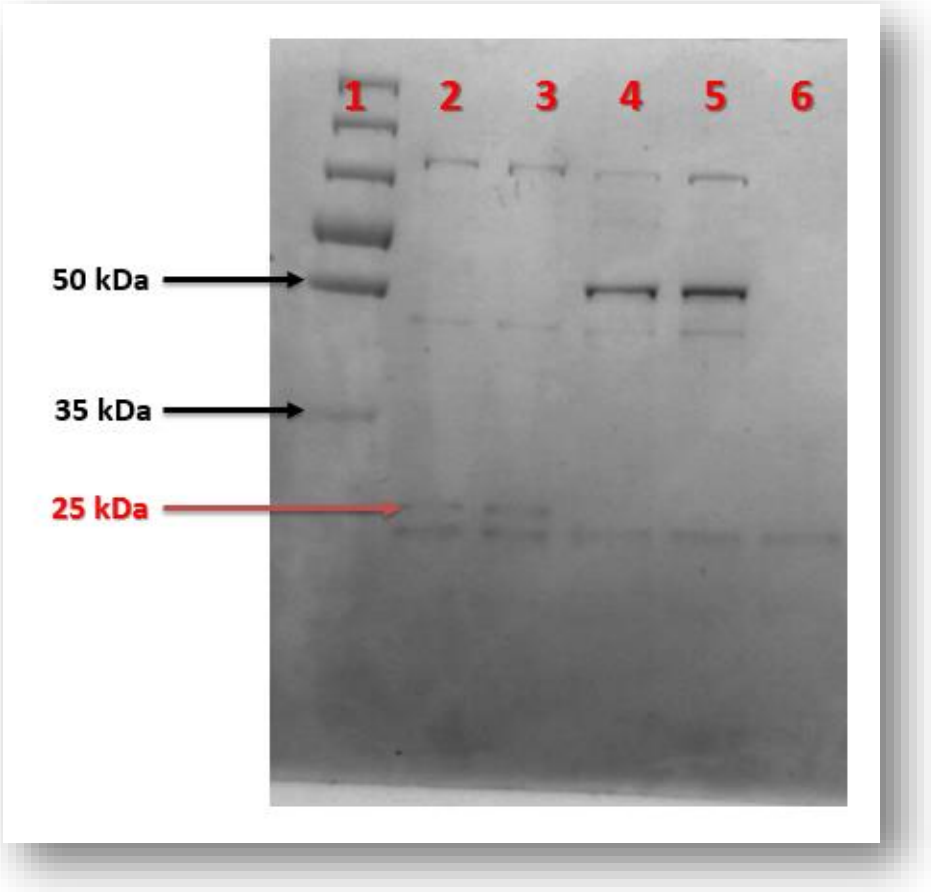


Figure 5.13: Coomassie stain of gel used in SUMOylation experiments highlights the range of proteins present in the reaction

Coomassie-stained SDS PAGE gel of *in vitro* SUMOylation assay carried out on Sigma-Aldrich's recombinant DHFR and recombinant p53 supplied with the Active Motif kit. The products of five reactions were run on this gel: DHFR +SUMO (lane 2), DHFR + *mut*SUMO (lane 3), p53 + SUMO1 (lane 4), p53 + *mut*SUMO (lane 5), and a no-protein negative control (lane 6). Bands for recombinant DHFR and p53 are visible at 25 kDa and 53 kDa in their respective reactions, while other bands represent the other components of the kit.

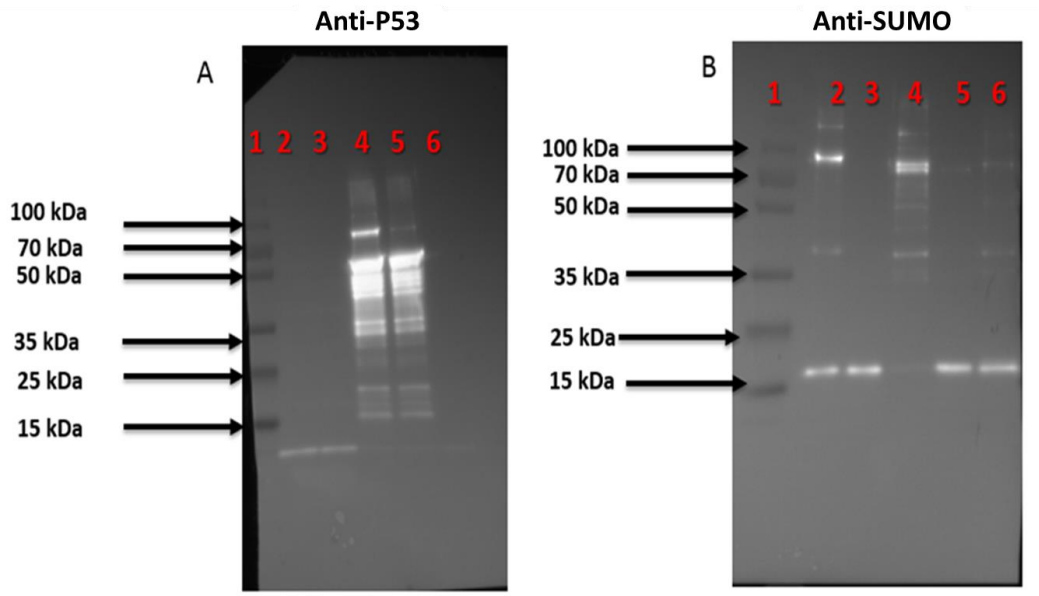


Figure 5.14: *In vitro* SUMOylation reactions on recombinant DHFR and P53 probed with anti-p53 and anti-SUMO

Two separate Western blots were carried out on products of the same set of reactions: DHFR + SUMO (lane 2), DHFR + *mut*SUMO (lane 3), P53 + SUMO (lane 4), P53 + mutSUMO (lane 5), and no-protein negative-control (lane 6). Westerns were probed with anti-p53 antibody (A) and anti-DHFR antibody (B). On the anti-P53 Western, a clear shift from 53 kDa to ~70 kDa is evident in lane 4 while being less pronounced in lane 5. In lane 4, a strong band at ~70 kDa indicates P53 has been SUMOylated. Free SUMO appears at the bottom of the gel (~15 kDa).

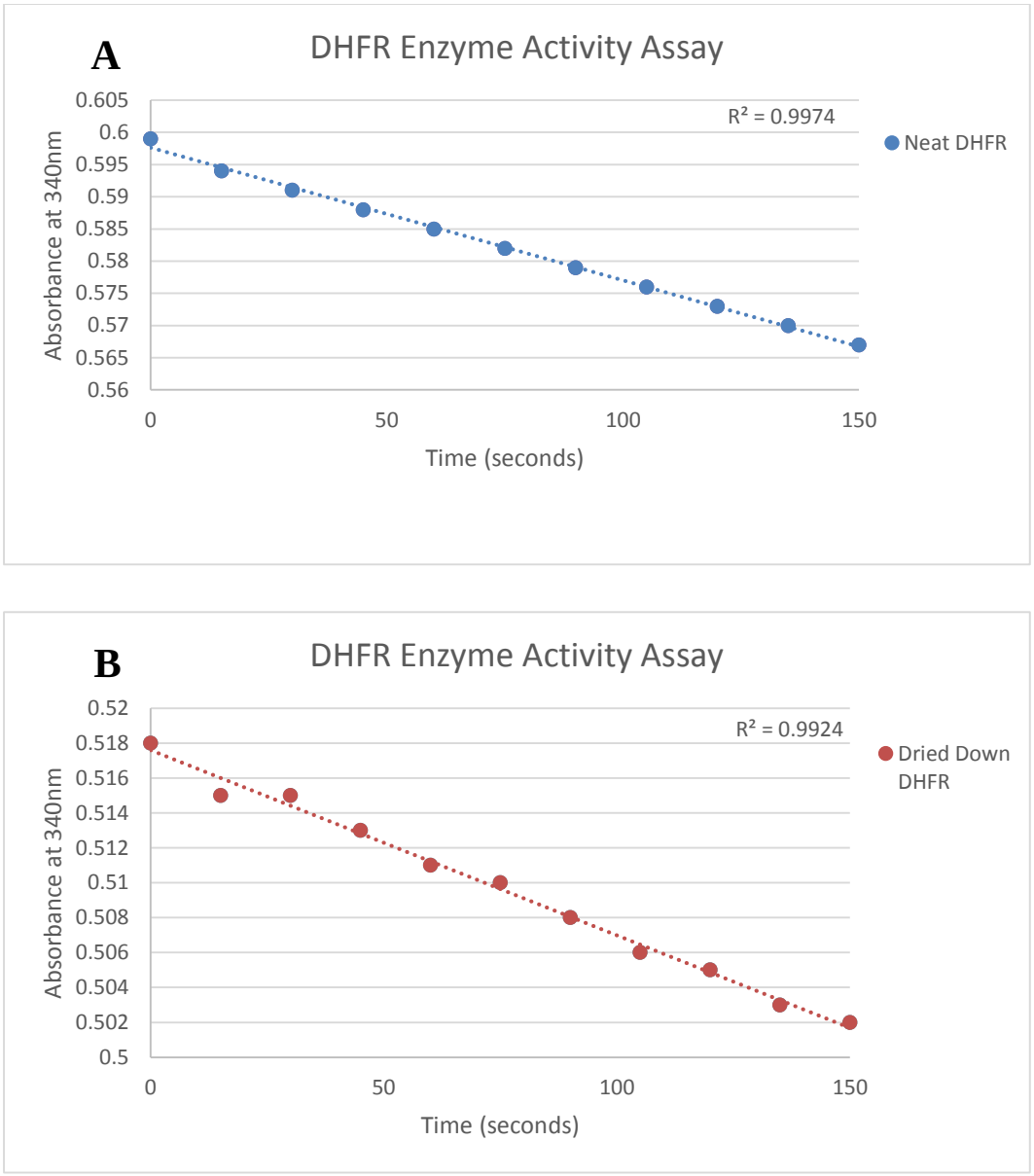


Figure 5.15: DHFR Enzyme Activity Assay on dried down recombinant DHFR from Sigma-Aldrich

Enzyme Activity of recombinant DHFR from Sigma-Aldrich dried down in the same manner as used in the *in vitro* SUMOylation reactions. Activity was measured for recombinant DHFR supplied with the kit (A) along with activity of the same enzyme dried down and resuspended in H_2O (B). A linear depletion of NADPH over time was observed in both reactions (R^2 = 0.9974, R^2 = 0.9924).

A

Method: Random Forest

There are 14 lysine residues in 1 protein.

Prediction results (threshold: -0.25)

Protein name	Sumoylation site	Classifier output	Confidence level*
UBC9	K14	0.532	91.05
UBC9	K49	0.814	99.74
UBC9	K65	0.445	84.47
UBC9	K76	0.510	90

4 lysine residues are predicted to be sumoylation sites.

* The confidence level is calculated as (1 - sensitivity) for positive predictions, and (1 - specificity) for negative predictions.

B

Protein ID:	gi|54039791|sp|P63279.1|UBC9_HUMAN
Defintion:	RecName: Full=SUMO-conjugating enzyme UBC9; AltName: Full=SUMO-protein ligase; AltName: Full=Ubiquitin carrier protein 9; AltName: Full=Ubiquitin carrier protein I; AltName: Full=Ubiquitin-conjugating enzyme E2 I; AltName: Full=Ubiquitin-protein ligase I; AltName: Full=p18
Length:	158 aa

```
  1 MSGIALSRLA QERKAWRKDH PFGFVAVPTK NPDGTMNLMN WECAIPGKKG
 51 TPWEGGLFKL RMLFKDDYPS SPPKCKFEPP LFHPNVYPSG TVCLSILEED
101 KDWRPAITIK QILLGIQELL NEPNIQDPAQ AEAYTIYCQN RVEYEKRVRA
151 QAKKFAPS
```

Motifs with high probability
Motifs with low probability
Overlapping Motifs

No.	Pos.	Group	Score	No.	Pos.	Group	Score
1	K65	KLRML FKDD YPSSP	0.85	3	K153	RVRAQ AKKF APS	0.44
2	K48	ECAIP GKKG TPWEG	0.5				

Figure 5.16: In Silico SUMOylation analysis on UBC9

Peptide sequences of UBC9 were examined for potential SUMOylation sites using SeeSUMO (A) and SUMOplot (B). According to the former, K14, K49, K65 and K76 have a high capacity to be SUMOylated, with confidence levels higher than those found in DHFR. For the same peptide sequence, SUMOplot found K65 to have the highest change of being SUMOylated, again, with a score higher than those found for the folate enzyme.

Chapter 6:

Discussion and Future Work

6.1: The Impact of Folic Acid Supplementation on DNA methylation

Folic acid supplementation has been demonstrated to prevent birth defects in numerous trials and studies over the past two decades[13,15,241,242]. Still, the biological basis through which folic acid exerts its beneficial effects is still relatively unknown, and concern is growing over studies that have linked elevated folate levels with tumourigenesis[21,24]. Although a recent meta-analysis has found no significant link between folic acid supplementation and cancer[25], it is still of paramount importance that these biological mechanisms are elucidated.

S-adenosylmethionine is a methyl donor for many methylation reactions throughout the cell, including DNA methylation, produced from folate-mediated one carbon metabolism[6]. A number of studies have shown that changes in one-carbon metabolism, such as increased levels of folate, can have a direct effect on the methylation of different genes throughout the genome[109,111]. DNA methylation is an epigenetic modification that is intricately associated with gene expression. As a result, it is plausible that increased folate levels may cause changes in DNA methylation and gene expression, thus establishing a biological mechanism by which folic acid supplementation can impact on the health status of an individual. This was the basis of this study.

Recently, an intervention study was carried out by the University of Ulster to investigate the effect of folic acid supplementation during the second and third trimester of pregnancy on folate status and homocysteine levels[4]. Blood samples were taken from 119 individuals before and after intervention, where one group received a daily supplement of folic acid, while participants from the other group received a placebo. A post-intervention sample was taken after 36 weeks of pregnancy, along with another blood sample from the umbilical cord upon delivery. Biochemical data was recorded from these blood samples, and DNA was extracted. These DNA samples made up the foundation of this study.

DNA methylation alone is a difficult modification to measure, so various strategies have been employed to measure and quantify it on both a genome-wide and gene-specific basis[5]. An in-depth review was carried out on these methods, and one that suited the aims, objectives, and resources of this study was selected: Modified Methylation Specific Digital Karyotyping (MMSDK)[133].

In a series of pilot studies using DNA from the bacteriophage lambda, MMSDK was found to work effectively on a small genome. However, scaling this up to human genomic DNA proved to be technically challenging, so an alternative genome-wide method was considered.

Methylated DNA immunoprecipitation (MeDIP) was found to efficiently enrich for methylated DNA though QC tests using qPCR, and the DNA microarray approach was selected as a route of analysis instead of MMSDK.

The DNA microarray analysis produced twelve large datasets, each of which was analysed in excel using macros scripted in Visual Basic 6. After this analysis, a preliminary list of 19 novel FS-DMRs was produced, along with 11 regions observed to change their methylation status in response to pregnancy, independent of folate status. Through examination of the microarray's raw data – the specific probe enrichment score that make up each analysed region – the preliminary list of FS-DMRs was narrowed down to five, three of which were selected for further analysis with a gene-specific method of analysis – Sensitive Melting after Real Time Methylation-Specific PCR (SMART-MSP)[146] – over the rest of the FASSTT cohort.

One of the regions analysed was adjacent to the coding region of *IP6K1*, a gene encoding a kinase responsible for synthesis of inositol pyrophosphates involved in multiple cellular processes such as chromatin remodelling and cell aging[176,177]. The second was adjacent to the coding region of *RASA4*, a small GTP binding signalling molecule involved in gene expression and cell cycle control[183], hypomethylation of which has been correlated with poor prognosis in juvenile myelomonocytic leukemia[184]. The third region, annotated as Chromosome 9, Open-Reading Frame 44, does not lie adjacent to a known protein coding region.

When analysed across the rest of the FASSTT cohort using SMART-MSP, changes in DNA methylation did not reach statistical significance. In a cell culture model treated with demethylating agent 5'azacytidine, an increase in gene expression for *IP6K1* and *RASA4* were observed in response to demethylation of their associated promoter region. *GPS2*, a gene with an FS-DMR increasing in methylation from the preliminary list of 19 regions found from the MeDIP analysis, also exhibited increased expression in response to demethylation at the region examined by the microarray.

In this investigation into the effects of folic acid supplementation on DNA methylation, no significant impact was found. Most studies from the literature using MeDIP and promoter microarray hybridisation have investigated the effect of conditions more dramatic than folic acid supplementation – healthy cells versus tumour cells, differences across species, etc.[169–173,175]. Regarding the FASSTT study, all women recruited were healthy, with folic acid supplementation being a relatively minor experimental condition imposed upon the participants. If we had the resources available to use a more sensitive method initially, i.e., next generation bisulfite sequencing producing single base-pair resolution data, robust FS-

DMR may have been found. However, using the MeDIP microarray approach did not yield consistent FS-DMRs in the samples studied.

Various studies have examined the effect of dramatic folate deficiency on mice. When examining changes in global DNA methylation – total DNA methylation across the genome without regional context – Caudill *et al.,* (2001) found no significant changes in liver hypomethylation in response to folate deficiency[243]. Using a more sensitive technique, Mejos *et al.,* (2013) found a significant correlation between hepatic folate and global DNA methylation levels in rats[244]. Of course, these methods of global analysis tell us very little about DNA methylation on a practical scale. Using gene-specific bisulfite sequencing, Geo *et al.,* (2012) found that folate deficiency in mice caused a decrease in DNA methylation of *Esr1,* the gene encoding estrogen receptor 1[245].

In human studies, examining folate deficiency is more difficult due to ethical concerns. In some instances, researchers have examined the effect of the polymorphism MTHFR 677 C>T on DNA methylation; a variation known to inhibit folate metabolism[246]. Across a numerous candidate genes, including *ESR1*, Hanks *et al.,* (2013) found no methylation changes attributed to the MTHFR 677 C>T genotype[204], mirroring results previously found by the same research group[247,248]. A recent study by Bollati *et al*., (2014) examined the impact of nutrient intake on DNA methylation in obese subjects, finding a negative correlation between TNFα methylation and folic acid intake determined from a standardised food frequency questionnaire. DNA methylation levels were determined using Pyrosequencing technology[249].

With respect to folic acid supplementation during pregnancy, Fryer *et al.,* (2009) found no correlation between folic acid intake and methylation of *LINE-1* repetitive elements (representative of global methylation), but saw an inverse correlation between plasma homocysteine levels and *LINE-1* methylation[250]. The same group later examined the relationship between homocysteine levels and genome-wide methylation using the Illumina 27k methylation bead array[111], and found 5 CpG sites directly correlated with homocysteine levels, and 12 CpG with inverse correlation. No confirmation analysis was carried out on these regions obtained from a sample size of n=12.

More recently, Amarasekera *et al*., (2014) found 7 FS-DMRs in neonatal cord blood correlated with foetal folate status. Of the regions found, the promoter upstream of the gene *ZFP57,* its product a regulator of DNA methylation during development[251]. In both the high folate and low folate groups, the study found a significant correlation between maternal serum folate levels measured at the third trimester of pregnancy, and those of the umbilical cord measured

upon delivery. This study suggests that folate levels during pregnancy – even as late as the third trimester – can have a significant impact on DNA methylation. These results mirror those of previous studies, suggesting that periconceptional folic acid supplementation – as reported by the mother after birth – is correlated with *IGF2* methylation in children[108], and inversely correlated with *HI9* methylation under similar conditions[252].

Although our MeDIP and SMART-MSP analysis found no correlation between folic acid supplementation and DNA methylation, it is likely that the changes in methylation levels were too subtle to analyse using the methods described here. Sensitive methods examining DNA methylation at a single-base pair resolution – like next generation sequencing and Pyrosequencing following sodium bisulfite treatment – may overcome this challenge.

6.2: Investigation of post-translational modifications affecting dihydrofolate reductase

Complementing the aforementioned DNA methylation analysis, an investigation was carried out on the potential post-translational modifications affecting dihydrofolate reductase (DHFR), the enzyme responsible for breaking folic acid down to its functional form.

In silico analysis of DHFR's primary structure found potential SUMOylation and acetylation sites. During the process of producing recombinant GST-bound DHFR for *in vitro* analysis, however, a number of significant challenges were encountered. On examining the soluble and insoluble fractions of BL21-DE3 cells, the majority of DHFR was found to exist in the latter. The culture was scaled up to make up for the small portion of DHFR in the soluble fraction, but this could not be purified via affinity chromatography with glutathione agarose beads.

Attempts were made to resolubilise the insoluble fraction using two different methods from the literature. Although one did succeed in partially purifying a large volume of DHFR-GST, when tested for reductase activity, the recombinant protein was found to have no enzyme activity.

SUMOylation experiments were carried out *in vitro* on commercially available DHFR from Sigma-Aldrich. This was found to be inconclusive due to self-SUMOylation of the kit's components, with a band-shift for DHFR possibly mistaken for that of UBC9: a SUMO conjugating enzyme.

These *in vitro* SUMOylation experiments were carried out as described by Anderson *et al.*, (2007)[43], but did not confirm or refute these published findings due to results obtained by our a no-protein negative controls for the *in* vitro SUMOylation experiments, and an *in silico* analysis of the UBC9 primary sequence. Further steps may be taken to confirm the identity of these bands, discussed in more detail below.

6.3: Future Work

The DNA methylation methods described here did not find a significant association between DNA methylation changes and folic acid supplementation. If such an association does exist, however, and our approach is limited by the power of these methods of analysis alone, then perhaps a more specific or sensitive platform may be used to further investigate the impact of folic acid supplementation on DNA. Apart from the limitations of the MeDIP microarray approach, a more preferable gene-specific DNA methylation analysis method would be Pyrosequencing. There is a potential for other DMRs described here to be analysed in this manner, such as the pregnancy-specific DMRs listed in Chapter 4, and the candidate genes listed in the literature review by Parle-McDermott and Ozaki[3]. With such significant strides being made technology over just the past few years, carrying out a genome-wide, single base-pair resolution DNA methylation analysis across the entire FASSTT cohort may be a viable option in the future.

Two aspects of the FASSTT cohort have not been explored in this project. Firstly, data for the *MTHFR* 677 C>T polymorphism was collected for each participant. Due to MTHFR's role in supplying the one-carbon units for the methylation cycle, a decrease in enzyme activity caused by the TT variant may amplify the effects of declining folate levels on DNA methylation. However, statistical power is likely to be an issue for such an analysis. Secondly, DNA extracted from the umbilical cord of children born to the FASSTT participants is also available for methylation analysis. Had we found a FS-DMR with a consistent change across the FASSTT cohort, examining its capacity to perpetuate transgenerationally would be a fascinating route to take. Of course, this approach would not be without its challenges, with paternal genotypic influences possibly affecting offspring DNA methylation patterns.

In the investigation into DHFR post-translational modifications, steps can be taken to overcome the most significant challenge in producing recombinant DHFR: its inability to fold correctly when bound to GST. Other studies have had more success in producing HIS-tagged DHFR for affinity purification via nickel-chromatography[43,231]. If high concentrations of

recombinant DHFR can be produced, the *in vitro* SUMOylation experiment could be repeated under the very same conditions described by Anderson *et al.*, (2007)[43], confirming or refuting the results found using commercially available DHFR. Likewise, protein recombinant technology could be used to produce DHFRL1 – the protein coded by the formally annotated pseudogene, as described by McEntee *et al.*, (2011)[40], addressing the as-yet unknown factors influencing mitochondrial localisation of DHFRL1.

With more time and resources, the *in vitro* acetylation experiments described in Section 6.3.7 could also be optimised and carried out on recombinant DHFR, allowing us to explore the potential for SUMOylation and acetylation competition for the same lysine residue in more detail. These results would aid in creating a full picture of the effects of post-translational modifications to DHFR on sub-cellular localisation.

In this investigation into the effects of folic acid supplementation on DNA methylation, a total of 3 genomic sites were screened for potential FS-DMRs and found to be inconsistent across the FASSTT cohort. However, it cannot be concluded that no FS-DMRs exist under the supplementation conditions examined. An in-depth, single base-pair resolution method of DNA methylation screening, such as bisulfite NGS, may find FS-DMRs still elusive to us. The approaches described here were made complicated by a number of factors, such as genetic variation amongst participants influencing DNA methylation patterns even before intervention. The nutritional intervention carried out as part of the FASSTT study was relatively minor compared to what these methods of DNA methylation analysis have been used to study in the literature, and it may be that MeDIP and SMART-MSP are not sensitive enough to measure potential DNA methylation changes associated with it.

Nonetheless, the work presented here constitutes a significant contribution to the field. As the aforementioned studies illustrate, examining the effects of folate levels on DNA methylation in humans is particularly challenging. This thesis has examined the limits of conventional DNA methylation techniques, but major strides in technology over the past few years indicate that we are only beginning to scale this Everest of the epigenetic landscape.

References

1. Barker, D. *Mothers, Babies and Health in Later Life*. 217 (Churchill Livingstone, Edinburgh., 1984).

2. Barker, D. J. P. The developmental origins of chronic adult disease. *Acta Paediatr. Suppl.* **93,** 26–33 (2004).

3. Parle-McDermott, A. & Ozaki, M. The impact of nutrition on differential methylated regions of the genome. *Adv. Nutr.* **2,** 463–71 (2011).

4. Mcnulty, B., McNulty, H., Marshall, B., *et al.* Impact of continuing folic acid after the first trimester of pregnancy: findings of a randomized trial of Folic Acid Supplementation in the Second and Third Trimesters. *Am. J. Clin. Nutr.* **98,** 92–8 (2013).

5. Harrison, A. & Parle-McDermott, A. DNA methylation: a timeline of methods and applications. *Front. Genet.* **2,** 74 (2011).

6. Fox, J. T. & Stover, P. J. Folate-mediated one-carbon metabolism. *Vitam. Horm.* **79,** 1–44 (2008).

7. Tibbetts, A. S. & Appling, D. R. Compartmentalization of Mammalian folate-mediated one-carbon metabolism. *Annu. Rev. Nutr.* **30,** 57–81 (2010).

8. Bailey, L. B. & Gregory, J. F. Folate metabolism and requirements. *J. Nutr.* **129,** 779–82 (1999).

9. Pitkin, R. M. Folate and neural tube defects. *Am. J. Clin. Nutr.* **85,** 285S–288S (2007).

10. Botto, L. D., Moore, C. A., Khoury, M. J. & Erickson, J. D. Neural-Tube Defects. *N. Engl. J. Med.* **341,** 1509–1519 (1999).

11. Wallingford, J. B., Niswander, L. A., Shaw, G. M. & Finnell, R. H. The continuing challenge of understanding, preventing, and treating neural tube defects. *Science* **339,** 1222002 (2013).

12. Eastern Regional Health Authority; North Eastern Health Board; South Eastern Health Board. Congenital abnormalities in the East of Ireland 1997-2001. Report of the birth defect registries of the Eastern Regional Health Authority, North Eastern Health Board, South Eastern Health Board. (2004).

13. Smithells, R. W., Sheppard, S., Schorah, C. J., *et al.* Possible prevention of neural-tube defects by periconceptional vitamin supplementation. *Lancet* **1,** 339–40 (1980).

14. Smithells, R. W., Nevin, N. C., Seller, M. J., *et al.* Further experience of vitamin supplementation for prevention of neural tube defect recurrences. *Lancet* **1,** 1027–31 (1983).

15. Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. MRC Vitamin Study Research Group. *Lancet* **338,** 131–7 (1991).

16. Mulinare, J., Cordero, J. F., Erickson, J. D. & Berry, R. J. Periconceptional use of multivitamins and the occurrence of neural tube defects. *JAMA* **260,** 3141–5 (1988).

17. Bower, C. & Stanley, F. J. Dietary folate as a risk factor for neural-tube defects: evidence from a case-control study in Western Australia. *Med. J. Aust.* **150,** 613–9 (1989).

18. Shaw, G. M., Schaffer, D., Velie, E. M., Morland, K. & Harris, J. A. Periconceptional vitamin use, dietary folate, and the occurrence of neural tube defects. *Epidemiology* **6,** 219–26 (1995).

19. Werler, M. M., Shapiro, S. & Mitchell, A. A. Periconceptional folic acid exposure and risk of occurrent neural tube defects. *JAMA* **269,** 1257–61 (1993).

20. Mason, J. B., Dickstein, A., Jacques, P. F., *et al.* A temporal association between folic acid fortification and an increase in colorectal cancer rates may be illuminating important biological principles: a hypothesis. *Cancer Epidemiol. Biomarkers Prev.* **16,** 1325–9 (2007).

21. Lucock, M. & Yates, Z. Folic acid fortification: a double-edged sword. *Curr. Opin. Clin. Nutr. Metab. Care* **12,** 555–64 (2009).

22. Hirsch, S., Sanchez, H., Albala, C., *et al.* Colon cancer in Chile before and after the start of the flour fortification program with folic acid. *Eur. J. Gastroenterol. Hepatol.* **21,** 436–9 (2009).

23. Heseker, H. B., Mason, J. B., Selhub, J., Rosenberg, I. H. & Jacques, P. F. Not all cases of neural-tube defect can be prevented by increasing the intake of folic acid. *Br. J. Nutr.* **102,** 173–80 (2009).

24. Cole, B. F., Baron, J. A., Sandler, R. S., *et al.* Folic acid for the prevention of colorectal adenomas: a randomized clinical trial. *JAMA* **297,** 2351–9 (2007).

25. Vollset, S. E., Clarke, R., Lewington, S., *et al.* Effects of folic acid supplementation on overall and site-specific cancer incidence during the randomised trials: meta-analyses of data on 50 000 individuals. *Lancet* **6736,** 1–8 (2013).

26. Hall, M. H., Pirani, B. B. K. & Campbell, D. THE CAUSE OF THE FALL IN SERUM FOLATE IN NORMAL PREGNANCY. *BJOG An Int. J. Obstet. Gynaecol.* **83,** 132–136 (1976).

27. Chanarin I, Rothman D, Ward A, P. J. Folate status and requirement in pregnancy. *Br. Med. Journal.* **2,** 390–394 (1968).

28. Bertino, J. R. Cancer research: from folate antagonism to molecular targets. *Best Pract. Res. Clin. Haematol.* **22,** 577–82 (2009).

29. Fuso, A. The “golden age” of DNA methylation in neurodegenerative diseases. *Clin. Chem. Lab. Med.* **51,** 523–34 (2013).

30. Duthie, S. J., Narayanan, S., Blum, S., Pirie, L. & Brand, G. M. Folate deficiency in vitro induces uracil misincorporation and DNA hypomethylation and inhibits DNA excision repair in immortalized normal human colon epithelial cells. *Nutr. Cancer* **37,** 245–51 (2000).

31. Finkelstein, J. D. Methionine metabolism in mammals. *J. Nutr. Biochem.* **1,** 228–37 (1990).

32. Crider, K. S., Yang, T. P., Berry, R. J. & Bailey, L. B. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv. Nutr.* **3,** 21–38 (2012).

33. Badiga, S., Johanning, G. L., Macaluso, M., *et al.* A lower degree of PBMC L1 methylation in women with lower folate status may explain the MTHFR C677T polymorphism associated higher risk of CIN in the US post folic acid fortification era. *PLoS One* **9,** e110093 (2014).

34. Stern, L. L., Mason, J. B., Selhub, J. & Choi, S. W. Genomic DNA hypomethylation, a characteristic of most cancers, is present in peripheral leukocytes of individuals who are homozygous for the C677T polymorphism in the methylenetetrahydrofolate reductase gene. *Cancer Epidemiol. Biomarkers Prev.* **9,** 849–53 (2000).

35. Blom, H. J., Shaw, G. M., den Heijer, M. & Finnell, R. H. Neural tube defects and folate: case far from closed. *Nat. Rev. Neurosci.* **7,** 724–31 (2006).

36. Askari, B. S. & Krajinovic, M. Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Curr. Genomics* **11,** 578–83 (2010).

37. Hillcoat B. L., Swett V., B. J. R. Increase of dihydrofolate reductase activity in cultured mammalian cells after exposure to methotrexate. *Proc. Natl. Acad. Sci. U. S. A.* **58,** 1632–1637 (1967).

38. Cronstein, B. N. Molecular therapeutics. Methotrexate and its mechanism of action. *Arthritis Rheum.* **39,** 1951–60 (1996).

39. Chen, M. J., Shimada, T., Moulton, A. D., *et al.* The functional human dihydrofolate reductase gene. *J. Biol. Chem.* **259,** 3933–43 (1984).

40. McEntee, G., Minguzzi, S., O'Brien, K., *et al.* The former annotated human pseudogene dihydrofolate reductase-like 1 (DHFRL1) is expressed and functional. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 15157–62 (2011).

41. Yuan, T.-T., Huang, Y., Zhou, C.-X., *et al.* Nuclear translocation of dihydrofolate reductase is not a pre-requisite for DNA damage induced apoptosis. *Apoptosis* **14,** 699–710 (2009).

42. MacFarlane, A. J., Anderson, D. D., Flodby, P., *et al.* Nuclear localization of de novo thymidylate biosynthesis pathway is required to prevent uracil accumulation in DNA. *J. Biol. Chem.* **286,** 44015–22 (2011).

43. Anderson, D. D., Woeller, C. F. & Stover, P. J. Small ubiquitin-like modifier-1 (SUMO-1) modification of thymidylate synthase and dihydrofolate reductase. *Clin. Chem. Lab. Med.* **45,** 1760–3 (2007).

44. Woeller, C. F., Anderson, D. D., Szebenyi, D. M. E. & Stover, P. J. Evidence for small ubiquitin-like modifier-dependent nuclear import of the thymidylate biosynthesis pathway. *J. Biol. Chem.* **282,** 17623–31 (2007).

45. Felsenfeld, G. A brief history of epigenetics. *Cold Spring Harb. Perspect. Biol.* **6,** (2014).

46. Waddington, C. H. The epigenotype. 1942. *Int. J. Epidemiol.* **41,** 10–3 (2012).

47. Van Speybroeck, L. From epigenesis to epigenetics: the case of C. H. Waddington. *Ann. N. Y. Acad. Sci.* **981,** 61–81 (2002).

48. Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737–8 (1953).

49. Watson, J. D. & Crick, F. H. C. The Structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* **18,** 123–131 (1953).

50. Stedman. Cell specificity of histones. *Nature* **166,** 780–1 (1950).

51. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci.* **51,** 786–794 (1964).

52. Peck, A. . *Aristotle, Generation of Animals. Translated by A.L. Peck*. 144–145 (Cambridge Harvard University Press, 1943).

53. Speybroeck, L., Waele, D. & Vijver, G. Theories in Early Embryology. *Ann. N. Y. Acad. Sci.* **981,** 7–49 (2006).

54. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128,** 669–81 (2007).

55. Nathan, D., Sterner, D. E. & Berger, S. L. Histone modifications: Now summoning sumoylation. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 13118–20 (2003).

56. Pray-Grant, M. G., Daniel, J. A., Schieltz, D., Yates, J. R. & Grant, P. A. Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. *Nature* **433,** 434–8 (2005).

57. Sims, R. J., Chen, C.-F., Santos-Rosa, H., *et al.* Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *J. Biol. Chem.* **280,** 41789–92 (2005).

58. Kuo, M. H. & Allis, C. D. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20,** 615–26 (1998).

59. Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.* **20,** 300–7 (2013).

60. Carrieri, C., Cimatti, L., Biagioli, M., *et al.* Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491,** 454–7 (2012).

61. Underwood, J. G., Uzilov, A. V, Katzman, S., *et al.* FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* **7,** 995–1001 (2010).

62. Van Dongen, J., Ehli, E. A., Slieker, R. C., *et al.* Epigenetic variation in monozygotic twins: a genome-wide analysis of DNA methylation in buccal cells. *Genes (Basel).* **5,** 347–65 (2014).

63. Daxinger, L. & Whitelaw, E. Transgenerational epigenetic inheritance: more questions than answers. *Genome Res.* **20,** 1623–8 (2010).

64. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187,** 226–32 (1975).

65. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14,** 9–25 (1975).

66. Siegfried, Z. & Simon, I. DNA methylation and gene expression. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2,** 362–71

67. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366,** 362–5 (1993).

68. Feinberg, A. P. The epigenetics of cancer etiology. *Semin. Cancer Biol.* **14,** 427–32 (2004).

69. Bickle, T. A. & Krüger, D. H. Biology of DNA restriction. *Microbiol. Rev.* **57,** 434–50 (1993).

70. Bestor, T. H. The DNA methyltransferases of mammals. *Hum. Mol. Genet.* **9,** 2395–402 (2000).

71. Bestor, T. H., Hellewell, S. B. & Ingram, V. M. Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* **4,** 1800–6 (1984).

72. Goll, M. G., Kirpekar, F., Maggert, K. A., *et al.* Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311,** 395–8 (2006).

73. Schaefer, M., Pollex, T., Hanna, K., *et al.* RNA methylation by Dnmt2 protects transfer RNAs against stress-induced cleavage. *Genes Dev.* **24,** 1590–5 (2010).

74. Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139,** 15–31 (2012).

75. Seki, Y., Hayashi, K., Itoh, K., *et al.* Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278,** 440–58 (2005).

76. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13,** 7–13 (2012).

77. Tahiliani, M., Koh, K. P., Shen, Y., *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–5 (2009).

78. Wu, H., Alessio, A. C. D., Ito, S., *et al.* Genome-wide analysis of distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* **25,** 679–684 (2011).

79. Zhu, J.-K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43,** 143–66 (2009).

80. Cortellino, S., Xu, J., Sannai, M., *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146,** 67–79 (2011).

81. Suzuki, H., Itoh, F., Toyota, M., *et al.* Quantitative DNA methylation analysis by fluorescent polymerase chain reaction single-strand conformation polymorphism using an automated DNA sequencer. *Electrophoresis* **21,** 904–8 (2000).

82. Ito, S., Shen, L., Dai, Q., *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–3 (2011).

83. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328,** 916–9 (2010).

84. Piccolo, F. M. & Fisher, A. G. Getting rid of DNA methylation. *Trends Cell Biol.* **24,** 136–43 (2014).

85. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11,** 204–20 (2010).

86. Hashimoto, H., Liu, Y., Upadhyay, A. K., *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40,** 4841–9 (2012).

87. Hahn, M. A., Szabó, P. E. & Pfeifer, G. P. 5-Hydroxymethylcytosine: A stable or transient DNA modification? *Genomics* **104,** 314–23 (2014).

88. Bachman, M., Uribe-Lewis, S., Yang, X., *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* (2014). doi:10.1038/nchem.2064

89. Roseboom, T. J., van der Meulen, J. H., Ravelli, a C., *et al.* Effects of prenatal exposure to the Dutch famine on adult disease in later life: an overview. *Twin Res.* **4,** 293–8 (2001).

90. Schulz, L. C. The Dutch Hunger Winter and the developmental origins of health and disease. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16757–8 (2010).

91. De Rooij, S. R., Wouters, H., Yonker, J. E., Painter, R. C. & Roseboom, T. J. Prenatal undernutrition and cognitive function in late adulthood. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 16881–6 (2010).

92. Stein, Z., Susser, M., Saenger, G. & Marolla, F. Nutrition and Mental Performance: Prenatal exposure to the Dutch famine of 1944-1945 seems not related to mental performance at age 19. *Science (80-.).* **178,** 708–713 (1972).

93. Ravelli, G. P., Stein, Z. A. & Susser, M. W. Obesity in young men after famine exposure in utero and early infancy. *N. Engl. J. Med.* **295,** 349–53 (1976).

94. Roseboom, T., de Rooij, S. & Painter, R. The Dutch famine and its long-term consequences for adult health. *Early Hum. Dev.* **82,** 485–91 (2006).

95. Hales, C. N. The thrifty phenotype hypothesis. *Br. Med. Bull.* **60,** 5–20 (2001).

96. Roseboom, T. J., Painter, R. C., van Abeelen, A. F. M., Veenendaal, M. V. E. & de Rooij, S. R. Hungry in the womb: what are the consequences? Lessons from the Dutch famine. *Maturitas* **70,** 141–5 (2011).

97. Heijmans, B. T., Tobi, E. W., Stein, A. D., *et al.* Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 17046–9 (2008).

98. Borgel, J., Guibert, S., Li, Y., *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* **42,** 1093–100 (2010).

99. Stöger, R. The thrifty epigenotype: an acquired and heritable predisposition for obesity and diabetes? *Bioessays* **30,** 156–66 (2008).

100. Gluckman, P. D., Hanson, M. a, Cooper, C. & Thornburg, K. L. Effect of in utero and early-life conditions on adult health and disease. *N. Engl. J. Med.* **359,** 61–73 (2008).

101. Paulsen, M. & Ferguson-Smith, a C. DNA methylation in genomic imprinting, development, and disease. *J. Pathol.* **195,** 97–110 (2001).

102. Liang, G., Salem, C. E., Yu, M. C., *et al.* DNA methylation differences associated with tumor tissues identified by genome scanning analysis. *Genomics* **53,** 260–8 (1998).

103. Urdinguio, R. G., Sanchez-Mut, J. V & Esteller, M. Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol.* **8,** 1056–72 (2009).

104. Gonzalgo, M. L., Liang, G., Spruck, C. H., *et al.* Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR. *Cancer Res.* **57,** 594–9 (1997).

105. Waterland, R. A., Dolinoy, D. C., Lin, J., *et al.* Maternal Methyl Supplements Increase Offspring DNA Methylation at Axin Fused. **406,** 401–406 (2006).

106. Marsit, C. J., McClean, M. D., Furniss, C. S. & Kelsey, K. T. Epigenetic inactivation of the SFRP genes is associated with drinking, smoking and HPV in head and neck squamous cell carcinoma. *Int. J. Cancer* **119,** 1761–6 (2006).

107. Tobi, E. W., Lumey, L. H., Talens, R. P., *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* **18,** 4046–53 (2009).

108. Steegers-Theunissen, R. P., Obermann-Borst, S. A., Kremer, D., *et al.* Periconceptional maternal folic acid use of 400 microg per day is related to increased methylation of the IGF2 gene in the very young child. *PLoS One* **4,** e7845 (2009).

109. Waterland, R. A. & Jirtle, R. L. Transposable Elements : Targets for Early Nutritional Effects on Epigenetic Gene Regulation Transposable Elements : Targets for Early Nutritional Effects on Epigenetic Gene Regulation. **23,** (2003).

110. Cropley, J. E., Suter, C. M., Beckman, K. B. & Martin, D. I. K. CpG Methylation of a Silent Controlling Element in the Murine A vy Allele Is Incomplete and Unresponsive to Methyl Donor Supplementation. **5,** 1–7 (2010).

111. Fryer, A. A., Emes, R. D., Ismail, K. M. K., *et al.* Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics* **6,** 86–94 (2011).

112. Dominguez-Salas, P., Moore, S. E., Baker, M. S., *et al.* Maternal nutrition at conception modulates DNA methylation of human metastable epialleles. *Nat. Commun.* **5,** 3746 (2014).

113. Amarasekera, M., Martino, D., Ashley, S., *et al.* Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. *FASEB J.* (2014). doi:10.1096/fj.13-249029

114. Claverie, J.-M. GENE NUMBER: What If There Are Only 30,000 Human Genes? *Science (80-.).* **291,** 1255–1257 (2001).

115. Kruglyak, L. & Nickerson, D. a. Variation is the spice of life. *Nat. Genet.* **27,** 234–6 (2001).

116. Lander, E. S., Linton, L. M., Birren, B., *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

117. D'Hont, A., Denoeud, F., Aury, J.-M., *et al.* The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature* **488,** 213–7 (2012).

118. Kuo, K. C., Mccune, R. A. & Gehrke, C. W. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* **8,** 4763–4776 (1980).

119. Sarda, S. & Hannenhalli, S. Next-generation sequencing and epigenomics research: a hammer in search of nails. *Genomics Inform.* **12,** 2–11 (2014).

120. Lister, R., O'Malley, R. C., Tonti-Filippini, J., *et al.* Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* **133,** 523–36 (2008).

121. Bibikova, M., Barnes, B., Tsan, C., *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98,** 288–95 (2011).

122. Sandoval, J., Heyn, H., Moran, S., *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6,** 692–702 (2011).

123. Cokus, S. J., Feng, S., Zhang, X., *et al.* Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452,** 215–9 (2008).

124. Maunakea, A. K., Nagarajan, R. P., Bilenky, M., *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–7 (2010).

125. Serre, D., Lee, B. H. & Ting, A. H. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* **38,** 391–9 (2010).

126. Cedar, H., Solage, A. & Hebrew, T. Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. **6,** 2125–2132 (1979).

127. Singer-sam, J., Grant, M., Lebon, J. M., *et al.* Use of a HpaII-Polymerase Chain Reaction Assay To Study DNA Methylation in the Pgk-1 CpG Island of Mouse Embryos at the Time of X-Chromosome Inactivation IG0-0. **10,** 4987–4989 (1990).

128. Hayashizaki Y, Hirotsune S, Okazaki Y, Hatada I, Shibata H, Kawai J, Hirose K, Watanabe S, Fushiki S, Wada S, et al. Restriction landmark genomic scanning method and its various applications. *Electrophoresis.* **14,** 251–8 (1993).

129. Hayatsu, H., Wataya, Y., Kai, K. & Iida, S. Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* **9,** 2858–65 (1970).

130. Kawai, J., Hirose, K., Fushiki, S., *et al.* Comparison of DNA methylation patterns among mouse cell lines by restriction landmark genomic scanning. *Mol. Cell. Biol.* **14,** 7421–7 (1994).

131. Liang, G., Gonzalgo, M. L., Salem, C. & Jones, P. A. Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* **27,** 150–5 (2002).

132. Welsh, J. & Mcclelland, M. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* **18,** 7213–8 (1990).

133. Li, J., Gao, F., Li, N., *et al.* An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines. *BMC Genomics* **10,** 223 (2009).

134. Hu, M., Yao, J. & Polyak, K. Methylation-specific digital karyotyping. *Nat. Protoc.* **1,** 1621–36 (2006).

135. Adouard, V., Dante, R., Niveleau, a, *et al.* The accessibility of 5-methylcytosine to specific antibodies in double-stranded DNA of Xanthomonas phage XP12. *Eur. J. Biochem.* **152,** 115–21 (1985).

136. Oakeley, E. J., Podestà, A. & Jost, J. P. Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 11721–5 (1997).

137. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* **241,** 172–82 (2002).

138. Mohn, F., Weber, M., Schübeler, D. & Roloff, T.-C. Methylated DNA immunoprecipitation (MeDIP). *Methods Mol. Biol.* **507,** 55–64 (2009).

139. Wang, R. Y., Gehrke, C. W. & Ehrlich, M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.* **8,** 4777–4790 (1980).

140. Frommer, M., McDonald, L. E., Millar, D. S., *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 1827–31 (1992).

141. Gonzalgo, M. L. & Jones, P. a. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* **25,** 2529–31 (1997).

142. Poduslo, S. E., Dean, M., Kolch, U. & O'Brien, S. J. Detecting high-resolution polymorphisms in human coding loci by combining PCR and single-strand conformation polymorphism (SSCP) analysis. *Am. J. Hum. Genet.* **49,** 106–11 (1991).

143. Bianco, T., Hussey, D. & Dobrovic, A. Methylation-sensitive, single-strand conformation analysis (MS-SSCA): A rapid method to screen for and analyze methylation. *Hum. Mutat.* **14,** 289–93 (1999).

144. Wilton, S. D., Lim, L., Dye, D. & Laing, N. Bandstab: a PCR-based alternative to cloning PCR products. *Biotechniques* **22,** 642–5 (1997).

145. Chabaud, R., Guenoun, M., Taddei, J. P. & Guillou, J. [Abnormal movements and shaking in geriatrics (author's transl)]. *Sem. Hop.* **55,** 465–9

146. Kristensen, L. S., Mikeska, T., Krypuy, M. & Dobrovic, A. Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res.* **36,** e42 (2008).

147. Tost, J. & Gut, I. G. DNA methylation analysis by pyrosequencing. *Nat. Protoc.* **2,** 2265–75 (2007).

148. Buckley, P. G., Das, S., Bryan, K., *et al.* Genome-wide DNA methylation analysis of neuroblastic tumors reveals clinically relevant epigenetic events and large-scale epigenomic alterations localized to telomeric regions. *Int. J. Cancer* **128,** 2296–305 (2011).

149. Brody, L. C., Conley, M., Cox, C., *et al.* A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Res. *Am. J. Hum. Genet.* **71,** 1207–15 (2002).

150. Sambrook, J., Fritsch, E. F., Maniatis, T. *Molecular Cloning: A laboratory manual, Second Edition.* (Cold Spring Harbour Laboratory Press, 1989).

151. Sirawaraporn, W., Prapunwattana, P., Sirawaraporn, R., Yuthavong, Y. & Santi, D. V. The dihydrofolate reductase domain of Plasmodium falciparum thymidylate synthase-dihydrofolate reductase. Gene synthesis, expression, and anti-folate-resistant mutants. *J. Biol. Chem.* **268,** 21637–44 (1993).

152. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14,** 204–20 (2013).

153. Costello, J. F., Frühwald, M. C., Smiraglia, D. J., *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* **24,** 132–8 (2000).

154. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25,** 1010–22 (2011).

155. Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6,** (2010).

156. Irizarry, R. A., Ladd-Acosta, C., Wen, B., *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41,** 178–86 (2009).

157. Doi, A., Park, I.-H., Wen, B., *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat. Genet.* **41,** 1350–3 (2009).

158. Jones, P. A. The DNA methylation paradox. *Trends Genet.* **15,** 34–7 (1999).

159. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315,** 1141–3 (2007).

160. Ball, M. P., Li, J. B., Gao, Y., *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27,** 361–8 (2009).

161. Lv, J., Liu, H., Su, J., *et al.* DiseaseMeth: a human disease methylation database. *Nucleic Acids Res.* **40,** D1030–5 (2012).

162. Rakyan, V. K., Down, T. A., Thorne, N. P., *et al.* An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* **18,** 1518–29 (2008).

163. Eckhardt, F., Lewin, J., Cortese, R., *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38,** 1378–85 (2006).

164. McRae, A. F., Powell, J. E., Henders, A. K., *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* **15,** R73 (2014).

165. Choi, M. R., In, Y.-H., Park, J., *et al.* Genome-scale DNA methylation pattern profiling of human bone marrow mesenchymal stem cells in long-term culture. *Exp. Mol. Med.* **44,** 503–12 (2012).

166. Steenbergen, R. D. M., Ongenaert, M., Snellenberg, S., *et al.* Methylation-specific digital karyotyping of HPV16E6E7-expressing human keratinocytes identifies novel methylation events in cervical carcinogenesis. *J. Pathol.* **231,** 53–62 (2013).

167. Lin, X., Li, J., Yin, G., *et al.* Integrative analyses of gene expression and DNA methylation profiles in breast cancer cell line models of tamoxifen-resistance indicate a potential role of cells with stem-like properties. *Breast Cancer Res.* **15,** R119 (2013).

168. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10,** 232 (2009).

169. Zhang, S., Feng, X.-L., Shi, L., *et al.* Genome-wide analysis of DNA methylation in tongue squamous cell carcinoma. *Oncol. Rep.* **29,** 1819–26 (2013).

170. Fujiwara, K., Ghosh, S., Liang, P., *et al.* Genome-wide screening of aberrant DNA methylation which associated with gene expression in mouse skin cancers. *Mol. Carcinog.* (2013). doi:10.1002/mc.22085

171. Morris, M. R., Ricketts, C. J., Gentle, D., *et al.* Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma. *Oncogene* **30,** 1390–401 (2011).

172. Cong, L., Jia, J., Qin, W., Ren, Y. & Sun, Y. Genome-wide analysis of DNA methylation in an APP/PS1 mouse model of Alzheimer's disease. *Acta Neurol. Belg.* (2013). doi:10.1007/s13760-013-0267-6

173. Wang, J., Cao, X., Zhang, Y. & Su, B. Genome-wide DNA methylation analyses in the brain reveal four differentially methylated regions between humans and non-human primates. *BMC Evol. Biol.* **12,** 144 (2012).

174. Schneider, E., El Hajj, N., Richter, S., *et al.* Widespread differences in cortex DNA methylation of the "language gene" CNTNAP2 between humans and chimpanzees. *Epigenetics* **9,** 533–45 (2014).

175. Lambertini, L., Lee, T.-L., Chan, W.-Y., *et al.* Differential methylation of imprinted genes in growth-restricted placentas. *Reprod. Sci.* **18,** 1111–7 (2011).

176. Shen, X., Xiao, H., Ranallo, R., Wu, W.-H. & Wu, C. Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science* **299,** 112–4 (2003).

177. Zhang, Z., Zhao, C., Liu, B., *et al.* Inositol pyrophosphates mediate the effects of aging on bone marrow mesenchymal stem cells by inhibiting Akt signaling. *Stem Cell Res. Ther.* **5,** 33 (2014).

178. Chakraborty, A., Koldobskiy, M. A., Bello, N. T., *et al.* Inositol pyrophosphates inhibit Akt signaling, thereby regulating insulin sensitivity and weight gain. *Cell* **143,** 897–910 (2010).

179. Jadav, R. S., Chanduri, M. V. L., Sengupta, S. & Bhandari, R. Inositol pyrophosphate synthesis by inositol hexakisphosphate kinase 1 is required for homologous recombination repair. *J. Biol. Chem.* **288,** 3312–21 (2013).

180. Bhandari, R., Juluri, K. R., Resnick, A. C. & Snyder, S. H. Gene deletion of inositol hexakisphosphate kinase 1 reveals inositol pyrophosphate regulation of insulin secretion, growth, and spermiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 2349–53 (2008).

181. Rao, F., Xu, J., Khan, A. B., *et al.* Inositol hexakisphosphate kinase-1 mediates assembly/disassembly of the CRL4-signalosome complex to regulate DNA repair and cell death. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 16005–10 (2014).

182. Roy, D., Sin, S.-H., Damania, B. & Dittmer, D. P. Tumor suppressor genes FHIT and WWOX are deleted in primary effusion lymphoma (PEL) cell lines. *Blood* **118,** e32–9 (2011).

183. Lockyer, P. J., Kupzig, S. & Cullen, P. J. CAPRI regulates Ca(2+)-dependent inactivation of the Ras-MAPK pathway. *Curr. Biol.* **11,** 981–6 (2001).

184. Poetsch, A. R., Lipka, D. B., Witte, T., *et al.* RASA4 undergoes DNA hypermethylation in resistant juvenile myelomonocytic leukemia. *Epigenetics* **9,** 1252–60 (2014).

185. Fukushige, S. & Horii, A. DNA methylation in cancer: a gene silencing mechanism and the clinical potential of its biomarkers. *Tohoku J. Exp. Med.* **229,** 173–85 (2013).

186. Pinney, S. E. Mammalian Non-CpG Methylation: Stem Cells and Beyond. *Biology (Basel).* **3,** 739–51 (2014).

187. Schneider-stock, R., Diab-assef, M., Rohrbeck, A., *et al.* 5-aza-Cytidine Is a Potent Inhibitor of DNA Methyltransferase 3a and Induces Apoptosis in HCT-116 Colon Cancer Cells via Gadd45- and p53-Dependent Mechanisms. **312,** 525–536 (2005).

188. Wong, J., Sia, Y. Y., Misso, N. L., *et al.* Effects of the demethylating agent, 5-azacytidine, on expression of the kallikrein-kinin genes in carcinoma cells of the lung and pleura. *Patholog. Res. Int.* **2011,** 167046 (2011).

189. Yuan, B.-Z., Jefferson, A. M., Popescu, N. C. & Reynolds, S. H. Aberrant gene expression in human non small cell lung carcinoma cells exposed to demethylating agent 5-aza-2'-deoxycytidine. *Neoplasia* **6,** 412–9

190. Falck, E., Groenhagen, A., Mühlisch, J., Hempel, G. & Wünsch, B. Genome-wide DNA methylation level analysis by micellar electrokinetic chromatography and laser-induced fluorescence detection after treatment of cell lines with azacytidine and antifolates. *Anal. Biochem.* **421,** 439–45 (2012).

191. Komashko, V. M. & Farnham, P. J. 5-azacytidine treatment reorganizes genomic histone modification patterns. *Epigenetics* **5,** 229–40 (2010).

192. Lee, Y.-G., Kim, I., Yoon, S.-S., *et al.* Comparative analysis between azacitidine and decitabine for the treatment of myelodysplastic syndromes. *Br. J. Haematol.* **161,** 339–47 (2013).

193. Aaron, N., Paul, H., Normand, R., *et al.* Azacitidine and decitabine have different mechanisms of action in non-small cell lung cancer cell lines. *Lung Cancer Targets Ther.* 119 (2010). doi:10.2147/LCTT.S11726

194. Hollenbach, P. W., Nguyen, A. N., Brady, H., *et al.* A Comparison of Azacitidine and Decitabine Activities in Acute Myeloid Leukemia Cell Lines. *PLoS One* **5,** e9001 (2010).

195. Pleyer, L. & Greil, R. Digging deep into "dirty" drugs - modulation of the methylation machinery. *Drug Metab. Rev.* 1–28 (2015). doi:10.3109/03602532.2014.995379

196. Jacob, R. A., Gretz, D. M., Taylor, P. C., *et al.* Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women. *J. Nutr.* **128,** 1204–12 (1998).

197. Rampersaud, G. C., Kauwell, G. P., Hutson, A. D., Cerda, J. J. & Bailey, L. B. Genomic DNA methylation decreases in response to moderate folate depletion in elderly women. *Am. J. Clin. Nutr.* **72,** 998–1003 (2000).

198. Mills, J. L., McPartlin, J. M., Kirke, P. N., *et al.* Homocysteine metabolism in pregnancies complicated by neural-tube defects. *Lancet* **345,** 149–51 (1995).

199. Cotter, A. M., Molloy, A. M., Scott, J. M. & Daly, S. F. Elevated plasma homocysteine in early pregnancy: a risk factor for the development of nonsevere preeclampsia. *Am. J. Obstet. Gynecol.* **189,** 391–4; discussion 394–6 (2003).

200. Nelen, W. L., Blom, H. J., Steegers, E. A., den Heijer, M. & Eskes, T. K. Hyperhomocysteinemia and recurrent early pregnancy loss: a meta-analysis. *Fertil. Steril.* **74,** 1196–9 (2000).

201. Homocysteine Lowering Trialists' Collaboration. Dose-dependent effects of folic acid on blood concentrations of homocysteine: a meta-analysis of the randomized trials. *Am. J. Clin. Nutr.* **82,** 806–12 (2005).

202. McPartlin, J., Halligan, A., Scott, J. M., Darling, M. & Weir, D. G. Accelerated folate breakdown in pregnancy. *Lancet* **341,** 148–9 (1993).

203. Zhang, B., Xing, X., Li, J., *et al.* Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics* **15,** 868 (2014).

204. Hanks, J., Ayed, I., Kukreja, N., *et al.* The association between MTHFR 677C>T genotype and folate status and genomic and gene-specific DNA methylation in the colon of individuals without colorectal neoplasia. *Am. J. Clin. Nutr.* **98,** 1564–74 (2013).

205. Goldberg, A. D., Allis, C. D. & Bernstein, E. Epigenetics: a landscape takes shape. *Cell* **128,** 635–8 (2007).

206. Ptashne, M. On the use of the word "epigenetic". *Curr. Biol.* **17,** R233–6 (2007).

207. Bártová, E., Krejcí, J., Harnicarová, A., Galiová, G. & Kozubek, S. Histone modifications and nuclear architecture: a review. *J. Histochem. Cytochem.* **56,** 711–21 (2008).

208. Bintu, L., Ishibashi, T., Dangkulwanich, M., *et al.* Nucleosomal elements that control the topography of the barrier to transcription. *Cell* **151,** 738–49 (2012).

209. Bell, O., Tiwari, V. K., Thomä, N. H. & Schübeler, D. Determinants and dynamics of genome accessibility. *Nat. Rev. Genet.* **12,** 554–64 (2011).

210. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322,** 750–6 (2008).

211. Koerner, M. V, Pauler, F. M., Huang, R. & Barlow, D. P. The function of non-coding RNAs in genomic imprinting. *Development* **136,** 1771–83 (2009).

212. Zhu, B. & Reinberg, D. Epigenetic inheritance: uncontested? *Cell Res.* **21,** 435–41 (2011).

213. Huang, C., Xu, M. & Zhu, B. Epigenetic inheritance mediated by histone lysine methylation: maintaining transcriptional states without the precise restoration of marks? *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368,** 20110332 (2013).

214. Xu, M., Wang, W., Chen, S. & Zhu, B. A model for mitotic inheritance of histone lysine methylation. *EMBO Rep.* **13,** 60–7 (2012).

215. Shiio, Y. & Eisenman, R. N. Histone sumoylation is associated with transcriptional repression. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 13225–30 (2003).

216. Luka, Z., Moss, F., Loukachevitch, L. V, Bornhop, D. J. & Wagner, C. Histone demethylase LSD1 is a folate-binding protein. *Biochemistry* **50,** 4750–6 (2011).

217. Das, C. & Kundu, T. K. Transcriptional regulation by the acetylation of nonhistone proteins in humans -- a new target for therapeutics. *IUBMB Life* **57,** 137–49 (2005).

218. Spange, S., Wagner, T., Heinzel, T. & Krämer, O. H. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **41,** 185–98 (2009).

219. Kouzarides, T. Acetylation: a regulatory modification to rival phosphorylation? *EMBO J.* **19,** 1176–9 (2000).

220. Pickard, A., Wong, P.-P. & McCance, D. J. Acetylation of Rb by PCAF is required for nuclear localization and keratinocyte differentiation. *J. Cell Sci.* **123,** 3718–26 (2010).

221. Lv, L., Xu, Y.-P., Zhao, D., *et al.* Mitogenic and oncogenic stimulation of K433 acetylation promotes PKM2 protein kinase activity and nuclear localization. *Mol. Cell* **52,** 340–52 (2013).

222. Hay, R. T. SUMO: a history of modification. *Mol. Cell* **18,** 1–12 (2005).

223. Ifergan, I. & Assaraf, Y. G. Molecular mechanisms of adaptation to folate deficiency. *Vitam. Horm.* **79,** 99–143 (2008).

224. Blais, N. & Hirsh, V. Chemotherapy in Metastatic NSCLC - New Regimens (Pemetrexed, Nab-Paclitaxel). *Front. Oncol.* **4,** 177 (2014).

225. Teng, S., Luo, H. & Wang, L. Predicting protein sumoylation sites from sequence features. *Amino Acids* **43,** 447–55 (2012).

226. Abgent. SUMOplot™ Analysis Program. (2014). at <http://www.abgent.com/sumoplot>

227. Chou, M. F. & Schwartz, D. Using the scan-x Web site to predict protein post-translational modifications. *Curr. Protoc. Bioinformatics* **Chapter 13,** Unit 13.16. (2011).

228. Grieco, F., Hay, J. M. & Hull, R. An improved procedure for the purification of protein fused with glutathione S-transferase. *Biotechniques* **13,** 856–8 (1992).

229. Levy, L., Wei, Y., Labalette, C., *et al.* Acetylation of -Catenin by p300 Regulates -Catenin-Tcf4 Interaction. *Mol. Cell. Biol.* **24,** 3404–3414 (2004).

230. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99,** 303–10 (2005).

231. Tai, N., Ding, Y., Schmitz, J. C. & Chu, E. Identi ® cation of critical amino acid residues on human dihydrofolate reductase protein that mediate RNA recognition. **30,** (2002).

232. Ma, L. & Kovacs, J. a. Expression and Characterization of Recombinant Human-Derived Pneumocystis carinii Dihydrofolate Reductase. *Antimicrob. Agents Chemother.* **44,** 3092–3096 (2000).

233. Burgess, R. R. Refolding solubilized inclusion body proteins. *Methods Enzymol.* **463,** 259–82 (2009).

234. Anthony, L. C., Dombkowski, A. A. & Burgess, R. R. Using Disulfide Bond Engineering To Study Conformational Changes in the ’260-309 Coiled-Coil Region of Escherichia coli RNA Polymerase during 70 Binding. *J. Bacteriol.* **184,** 2634–2641 (2002).

235. Smith, V. F. & Matthews, C. R. Testing the role of chain connectivity on the stability and structure of dihydrofolate reductase from E. coli: fragment complementation and circular permutation reveal stable, alternatively folded forms. *Protein Sci.* **10,** 116–28 (2001).

236. Qoronfleh, M. W., Hesterberg, L. K. & Seefeldt, M. B. Confronting high-throughput protein refolding using high pressure and solution screens. *Protein Expr. Purif.* **55,** 209–24 (2007).

237. Knipscheer, P., Flotho, A., Klug, H., *et al.* Ubc9 sumoylation regulates SUMO target discrimination. *Mol. Cell* **31,** 371–82 (2008).

238. Dormeyer, W., Ott, M. & Schnölzer, M. Probing lysine acetylation in proteins: strategies, limitations, and pitfalls of in vitro acetyltransferase assays. *Mol. Cell. Proteomics* **4,** 1226–39 (2005).

239. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90,** 595–606 (1997).

240. Brooks, C. L. & Gu, W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr. Opin. Cell Biol.* **15,** 164–171 (2003).

241. Berry, R. J., Li, Z., Erickson, J. D., *et al.* Prevention of neural-tube defects with folic acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. *N. Engl. J. Med.* **341,** 1485–90 (1999).

242. Food Safety Authoirty of Ireland. *National Committee on folic acid food fortification*. ISBN 1–9004465–43–9 (2006).

243. Caudill, M. A., Wang, J. C., Melnyk, S., *et al.* Intracellular S-adenosylhomocysteine concentrations predict global DNA hypomethylation in tissues of methyl-deficient cystathionine beta-synthase heterozygous mice. *J. Nutr.* **131,** 2811–8 (2001).

244. Mejos, K. K., Kim, H. W., Lim, E. M. & Chang, N. Effects of parental folate deficiency on the folate content, global DNA methylation, and expressions of FRα, IGF-2 and IGF-1R in the postnatal rat liver. *Nutr. Res. Pract.* **7,** 281–6 (2013).

245. Gao, R., Ding, Y., Liu, X., *et al.* Effect of folate deficiency on promoter methylation and gene expression of Esr1, Cdh1 and Pgr, and its influence on endometrial receptivity and embryo implantation. *Hum. Reprod.* **27,** 2756–65 (2012).

246. Harisha, P. N., Devi, B. I., Christopher, R. & Kruthika-Vinod, T. P. Impact of 5,10-methylenetetrahydrofolate reductase gene polymorphism on neural tube defects. *J. Neurosurg. Pediatr.* **6,** 364–7 (2010).

247. Pufulete, M., Al-Ghnaniem, R., Rennie, J. A., *et al.* Influence of folate status on genomic DNA methylation in colonic mucosa of subjects without colorectal adenoma or cancer. *Br. J. Cancer* **92,** 838–42 (2005).

248. Al-Ghnaniem, R., Peters, J., Foresti, R., Heaton, N. & Pufulete, M. Methylation of estrogen receptor alpha and mutL homolog 1 in normal colonic mucosa: association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia. *Am. J. Clin. Nutr.* **86,** 1064–72 (2007).

249. Bollati, V., Favero, C., Albetti, B., *et al.* Nutrients intake is associated with DNA methylation of candidate inflammatory genes in a population of obese subjects. *Nutrients* **6,** 4625–39 (2014).

250. Fryer, A. A., Nafee, T. M., Ismail, K. M. K., *et al.* LINE-1 DNA methylation is inversely correlated with cord plasma homocysteine in man: a preliminary study. *Epigenetics* **4,** 394–8 (2009).

251. Amarasekera, M., Martino, D., Ashley, S., *et al.* Genome-wide DNA methylation profiling identifies a folate-sensitive region of differential methylation upstream of ZFP57-imprinting regulator in humans. *FASEB J.* (2014). doi:10.1096/fj.13-249029

252. Hoyo, C., Murtha, A. P., Schildkraut, J. M., *et al.* Methylation variation at IGF2 differentially methylated regions and maternal folic acid use before and during pregnancy. *Epigenetics* **6,** 928–36 (2011).

Appendix A:

Visual Basic 6 Macro Scripts

A.1: Sub Macros used for removing and pasting mismatched entries

```
Sub NuBlueCut(Rng As Range)
    Rng.Resize(, 4).Copy
    Rng.Offset(, 5).Insert xlDown
    Rng.Resize(, 4).Delete xlUp
End Sub

Sub NuRedCut(Rng As Range)
    Rng.Offset(, -4).Resize(, 4).Copy
    Rng.Offset(, 10).Insert xlDown
    Rng.Offset(, -4).Resize(, 4).Delete xlUp
End Sub
```

A.2 Looping Macro for Step 1 of data analysis

```
Sub StepOne()
Dim Rng As Range
Dim i As Long
i = 2
Application.ScreenUpdating = False
While i <= 40043
Set Rng = Range("E" & i)

If Rng > Rng.Offset(, -1) Then
NuRedCut Rng

ElseIf Rng < Rng.Offset(, -1) Then
NuBlueCut Rng

ElseIf Rng = Rng.Offset(, -1) And Rng.Offset(, -3) >
Rng.Offset(, 2) Then
NuRedCut Rng

ElseIf Rng = Rng.Offset(, -1) And Rng.Offset(, -3) <
Rng.Offset(, 2) Then
NuBlueCut Rng

Else
i = i + 1


End If
Wend
End Sub
```

Both of the macros in Appendix A.1 were tied together in the *Do While…* loop named "StepOne" in Appendix A.2. The opening lines of the code define the dimensions *Rng* and *i* as a range and a long integer, respectively. The next line, "*i* = 2", defines the long integer *i* as 2 to begin. "Application.Screenupdating = False" tells excel to not update this process on-screen as it happens, making the process faster. The next line tells the code to only work while *i* is less than or equal to 40,043; the total number of rows in the dataset. Finally, the range is set to the cell at column E, row *i*, which has already been defined as 2. With this information, the looping macro begins at cell E2.

Four conditions are specified next, and a consequence is listed for each. First, if the active range – like the accession number BC000008 – is greater than that immediately to the left of it, then the sub macro NuRedCut is called. If the active range is less than the cell immediately to the left of it, then the sub macro NuBlueCut is called. The following two conditions are similar to these, but they also include instructions that the chromosomal region must also match in each case. This is to ensure multiple entries with the same accession number are not confused.

The final line, "Else", says that if the active range, at this point E2, is not greater than or less than that in the adjacent row, then the value for *i* is increased by 1 and the macro starts again in cell E3.

A.3 Sub Macros for Step 2 of data analysis

```
Sub MiddleDelete(Rng As Range)
    Rng.Resize(, 4).Delete Shift:=xlUp
End Sub

Sub LeftDelete(Rng As Range)
    Rng.Offset(, -5).Resize(, 4).Delete Shift:=xlUp
End Sub

Sub RightDelete(Rng As Range)
    Rng.Offset(, 5).Resize(, 4).Delete Shift:=xlUp
End Sub

Sub FKCutandPaste(Rng As Range)
    Rng.Resize(, 9).Copy
    Rng.Offset(, 19).Insert Shift:=xlDown
    Rng.Resize(, 9).Delete Shift:=xlUp
End Sub
```

```
Sub AKCutandPaste(Rng As Range)
    Rng.Offset(, -5).Resize(, 4).Copy
    Rng.Offset(, 45).Insert Shift:=xlDown
    Rng.Offset(, -5).Resize(, 4).Delete Shift:=xlUp
    Rng.Offset(, 5).Resize(, 4).Copy
    Rng.Offset(, 50).Insert Shift:=xlDown
    Rng.Resize(, 4).Delete Shift:=xlUp
End Sub

Sub AFCutandPaste(Rng As Range)
    Rng.Offset(, -5).Resize(, 9).Copy
    Rng.Offset(, 35).Insert Shift:=xlDown
    Rng.Offset(, -5).Resize(, 9).Delete Shift:=xlUp
End Sub

Sub AFKCutandPaste(Rng As Range)
    Rng.Offset(, -5).Resize(, 14).Copy
    Rng.Offset(, 60).Insert Shift:=xlDown
    Rng.Offset(, -5).Resize(, 14).Delete Shift:=xlUp
End Sub
```

A.4 Looping macro for Step 2 of data analysis

```
Sub RemoveRow()
Dim i As Long
Dim Rng As Range
i = 2
Application.ScreenUpdating = False

Do While i <= 4043
Set Rng = Range("F" & i)

    If Rng < Rng.Offset(, -5) And Rng < Rng.Offset(, 5)
Then
        MiddleDelete Rng

    ElseIf Rng.Offset(, 5) < Rng And Rng.Offset(, 5) <
Rng.Offset(, -5) Then
        RightDelete Rng

    ElseIf Rng.Offset(, -5) < Rng And Rng.Offset(, -5) <
Rng.Offset(, 5) Then
        LeftDelete Rng

    ElseIf Rng = Rng.Offset(, 5) And Rng <> Rng.Offset(,
-5) Then
        FKCutandPaste Rng
```

```
     ElseIf Rng.Offset(, -5) = Rng.Offset(, 5) And Rng <>
Rng.Offset(, -5) Then
        AKCutandPaste Rng

     ElseIf Rng.Offset(, -5) = Rng And Rng <> Rng.Offset(,
5) Then
        AFCutandPaste Rng

     ElseIf Rng = Rng.Offset(, -5) And Rng = Rng.Offset(,
5) Then
        AFKCutandPaste Rng

Else: Stop
End If
Loop
End Sub
```

A.5: Sub Macros for comparing Intervention DMRs against Placebo DMRs

```
Sub PlaceboCut(Rng)
    Rng.Resize(, 4).Delete Shift:=xlUp
End Sub

Sub FACut(Rng)
    Rng.Offset(, -5).Resize(, 4).Copy
    Rng.Offset(, 10).Insert Shift:=xlDown
    Rng.Offset(, -5).Resize(, 4).Delete Shift:=xlUp
End Sub

Sub BothCut(Rng)
    Rng.Offset(, -5).Resize(, 9).Copy
    Rng.Offset(, 18).Insert Shift:=xlDown
    Rng.Offset(, -5).Resize(, 9).Delete Shift:=xlUp
End Sub
```

A.6: Looping Macro to Remove DMRs Common in Placebo and Intervention Groups

```
Sub PlaceboCompare()
Dim i As Long
Dim Rng As Range

i = 2

Do While i <= 4043
    Set Rng = Range("F" & i)

    If Rng < Rng.Offset(, -5) Then
        PlaceboCut Rng
```

```
    ElseIf Rng = Rng.Offset(, -5) And Rng.Offset(, -3) <
Rng.Offset(, 2) Then
        PlaceboCut Rng

    ElseIf Rng.Offset(, -5) < Rng Then
        FACut Rng
    ElseIf Rng = Rng.Offset(, -5) And Rng.Offset(, -3) >
Rng.Offset(, 2) Then
        FACut Rng

    ElseIf Rng.Offset(, -5) = Rng And Rng.Offset(, -3) =
Rng.Offset(, 2) Then
        BothCut Rng

Else: Stop
End If
Loop
End
```

Appendix B:

Raw data from MeDIP qPCR QC analysis.

Roche

Alan qPCR HRM H3b H19 161112

Abs Quant/2nd Derivative Max for New Subset 2 (Abs Quant/2nd Derivative Max)

Results

Inc	Pos	Name	Type	CP	Concentration	Standard	Status
☑	C5	H19 Input	Unknown	26.45			
☑	C7	H19 neg	Unknown				
☑	D5	H19 Input	Unknown	26.01			
☑	D7	H19 neg	Unknown				
☑	E1	H19 FASSTT 1	Unknown	24.26			
☑	E2	H19 FASSTT 2	Unknown	24.82			
☑	E3	H19 FASSTT 3	Unknown	22.87			
☑	E4	H19 FASSTT 4	Unknown	24.27			
☑	E5	H19 FASSTT 5	Unknown	23.81			
☑	E6	H19 FASSTT 6	Unknown	23.98			
☑	E7	H19 FASSTT 7	Unknown	24.71			
☑	E8	H19 FASSTT 8	Unknown	25.37			
☑	E9	H19 FASSTT 9	Unknown	24.99			
☑	E10	H19 FASSTT 10	Unknown	24.93			
☑	E11	H19 FASSTT 11	Unknown	24.62			
☑	E12	H19 FASSTT 12	Unknown	24.23			
☑	F1	H19 FASSTT 1	Unknown	23.70			
☑	F2	H19 FASSTT 2	Unknown	24.88			
☑	F3	H19 FASSTT 3	Unknown	22.94			
☑	F4	H19 FASSTT 4	Unknown	24.35			
☑	F5	H19 FASSTT 5	Unknown	23.65			
☑	F6	H19 FASSTT 6	Unknown	25.18			
☑	F7	H19 FASSTT 7	Unknown	24.92			
☑	F8	H19 FASSTT 8	Unknown	25.28			
☑	F9	H19 FASSTT 9	Unknown	25.28			
☑	F10	H19 FASSTT 10	Unknown	25.04			
☑	F11	H19 FASSTT 11	Unknown	24.06			
☑	F12	H19 FASSTT 12	Unknown	24.83			

Statistics

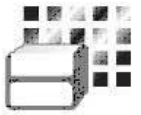
Samples	Mean Cp	Std Cp	Mean conc	Std conc
C5, D5	26.23	0.31		

Statistics

Samples	Mean Cp	Std Cp	Mean conc	Std conc
C6, D6	22.21	0.00		
C7, D7				
E1, F1	23.98	0.40		
E2, F2	24.85	0.04		
E3, F3	22.91	0.06		
E4, F4	24.31	0.06		
E5, F5	23.73	0.11		
E6, F6	24.58	0.85		
E7, F7	24.82	0.15		
E8, F8	25.33	0.06		
E9, F9	25.14	0.20		
E10, F10	24.98	0.08		
E11, F11	24.34	0.40		
E12, F12	24.53	0.42		

Figure B.1: *H19* qPCR Data

Raw data from qPCR QC analysis on MeDIP FASSTT DNA and Input DNA for the long non-coding RNA gene *H19*, known to be methylated.



Roche

Alan qPCR FASSTT H3b 271112

Abs Quant/2nd Derivative Max for New Subset 2 (Abs Quant/2nd Derivative Max)

Results

Inc	Pos	Name	Type	CP	Concentration	Standard	Status
☑	A1	FASSTT 1	Unknown	26.18			
☑	A2	FASSTT 2	Unknown	26.53			
☑	A3	FASSTT 3	Unknown	24.39			
☑	A4	FASSTT 4	Unknown	27.73			
☑	A5	FASSTT 5	Unknown	24.26			
☑	A6	FASSTT 6	Unknown	26.38			
☑	A7	FASSTT 7	Unknown	26.54			
☑	A8	FASSTT 8	Unknown	25.84			
☑	A9	FASSTT 9	Unknown	28.79			
☑	A11	FASSTT 11	Unknown	26.09			
☑	A12	FASSTT 12	Unknown	24.98			
☑	B1	FASSTT 1	Unknown	26.11			
☑	B2	FASSTT 2	Unknown	26.68			
☑	B3	FASSTT 3	Unknown	24.54			
☑	B4	FASSTT 4	Unknown	27.67			
☑	B5	FASSTT 5	Unknown	24.45			
☑	B6	FASSTT 6	Unknown	26.59			
☑	B7	FASSTT 7	Unknown	26.72			
☑	B8	FASSTT 8	Unknown	25.57			
☑	B9	FASSTT 9	Unknown	28.00			
☑	B11	FASSTT 11	Unknown	25.73			
☑	B12	FASSTT 12	Unknown	24.48			
☑	C10	FASSTT 10	Unknown	27.28			
☑	D10	FASSTT 10	Unknown	27.34			
☑	E1	Mock	Unknown	22.25			
☑	E2	Mock	Unknown	21.75			

Statistics

Samples	Mean Cp	Std Cp	Mean conc	Std conc
A1, B1	26.14	0.05		
A2, B2	26.61	0.11		
A3, B3	24.46	0.11		

Statistics

Samples	Mean Cp	Std Cp	Mean conc	Std conc
A4, B4	27.70	0.04		
A5, B5	24.36	0.13		
A6, B6	26.48	0.15		
A7, B7	26.63	0.13		
A8, B8	25.71	0.19		
A9, B9	28.39	0.56		
A11, B11	25.91	0.26		
A12, B12	24.73	0.35		
C10, D10	27.31	0.04		
E1, E2	22.00	0.35		

Figure B.2: *H3b* qPCR Data

Raw data from qPCR QC analysis on MeDIP FASSTT DNA and Input DNA for the histone *H3b* gene known to be unmethylated

Appendix C:

Pregnancy Related Differentially Methylated Regions

Appendix C.1: Genes that exhibited a decrease in methylation at their promoter regions in response to pregnancy

Description	Fold Change	STDEV
splicing factor 3b, subunit 4, 49kDa	-2.82	1.5976
death-domain associated protein	-2.32	0.9864
RhoA/RAC/CDC42 exchange factor	-2.26	1.3798
talin 2	-2.09	1.2596
melanoma antigen family D, 4	-2.01	1.2154
melanoma antigen family D, 4B	-2.01	1.2154

Appendix D:

SMART MSP Assay Data

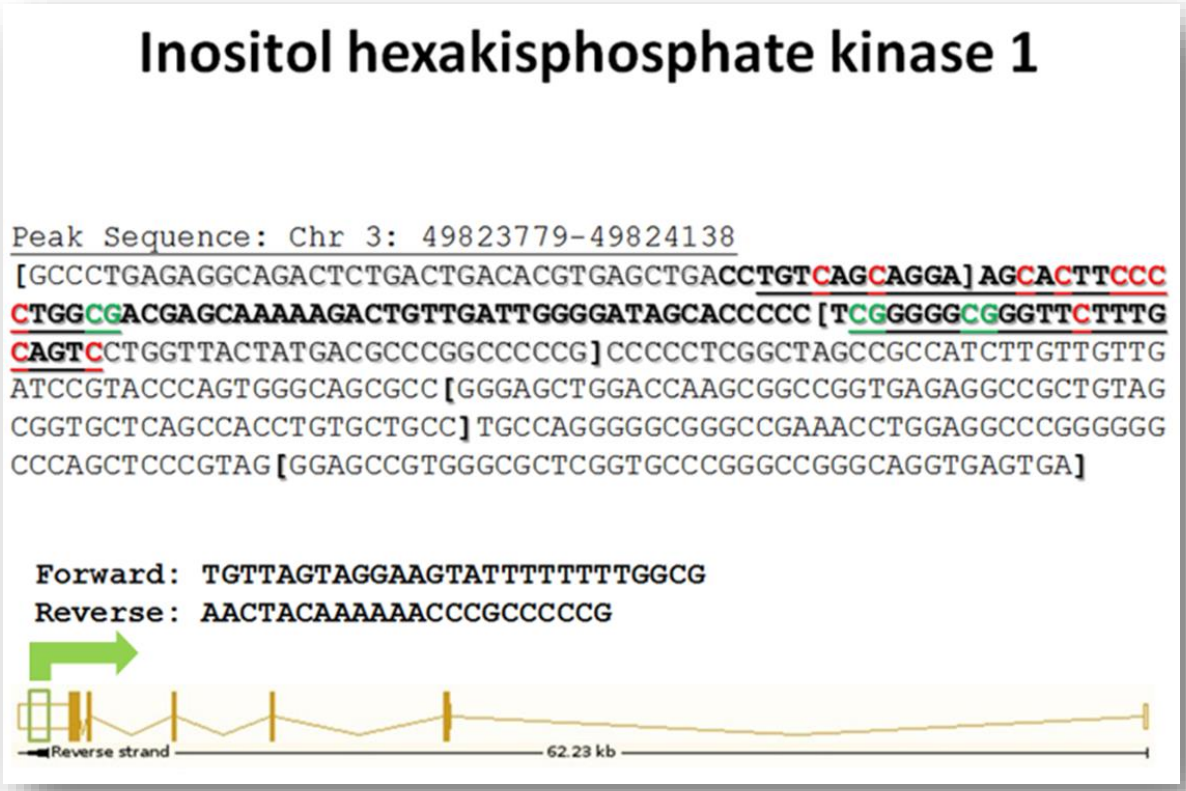


Figure D.1: IP6K1 SMART-MSP Assay

Sequence and location information on the SMART MSP assay designed for Inositol hexakisphosphate kiniase 1 (*IP6K1*). Probe data taken from the highlighted portion of Figure 3.13, with sequence data from primary assembly GRCh37.p13. Primer sequences were derived from sodium bisulfite treatment of this region, with red cytosines anticipated to deaminate to thymidines. Probe position indicated by parentheses. For fully methylated DNA, CpG dinucleotides (highlighted in green) will retain their parent sequence. In the gene map below, the green rectangle refers to where this sequence lies relative to the introns and exons of IP6K1, and direction of transcription is indicated by the green arrow. The regulatory region of IP6K1, 3:49823786-49823961 is covered by this region. The TSS for IP6K1 lies at 3:49785433.

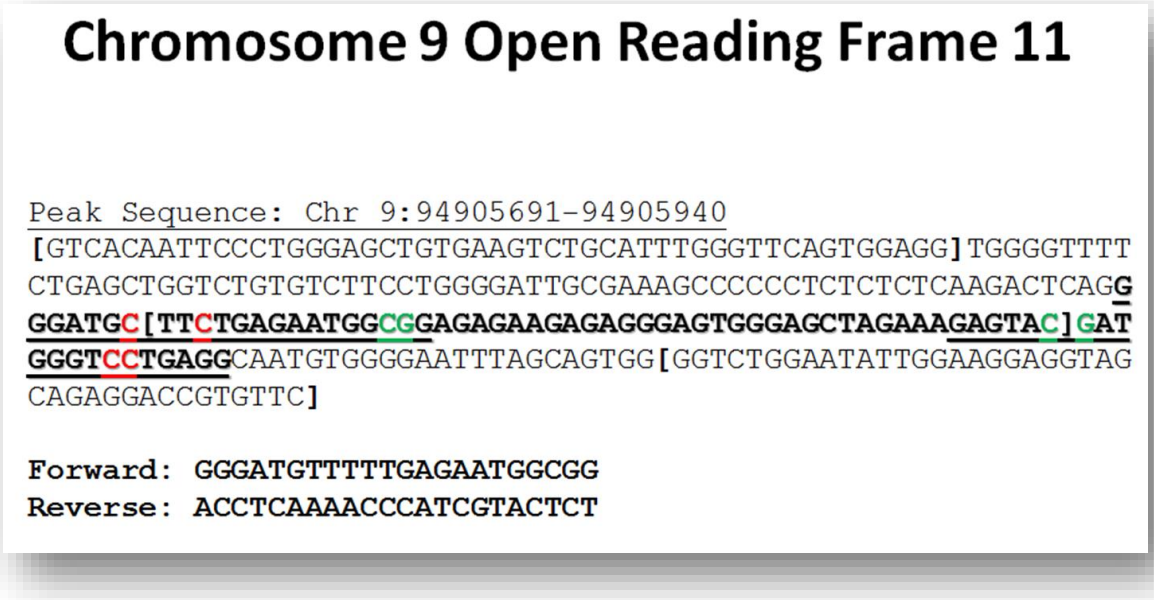


Figure D.2: Chr9ORF44 SMART MSP Assay

Sequence and location information on the SMART MSP assay designed for Chromosome 9 Open Reading Frame 44 (Chr9ORF44). Probe data taken from the highlighted portion of Figure 3.14, with sequence data from primary assembly GRCh37. Primer sequences were derived from sodium bisulfite treatment of this region, with red cytosines anticipated to deaminate to thymidines. Probe position indicated by parentheses. For fully methylated DNA, CpG dinucleotides (highlighted in green) will retain their parent sequence. No gene map is displayed, as this probe region does not lie adjacent to a known expressed gene.

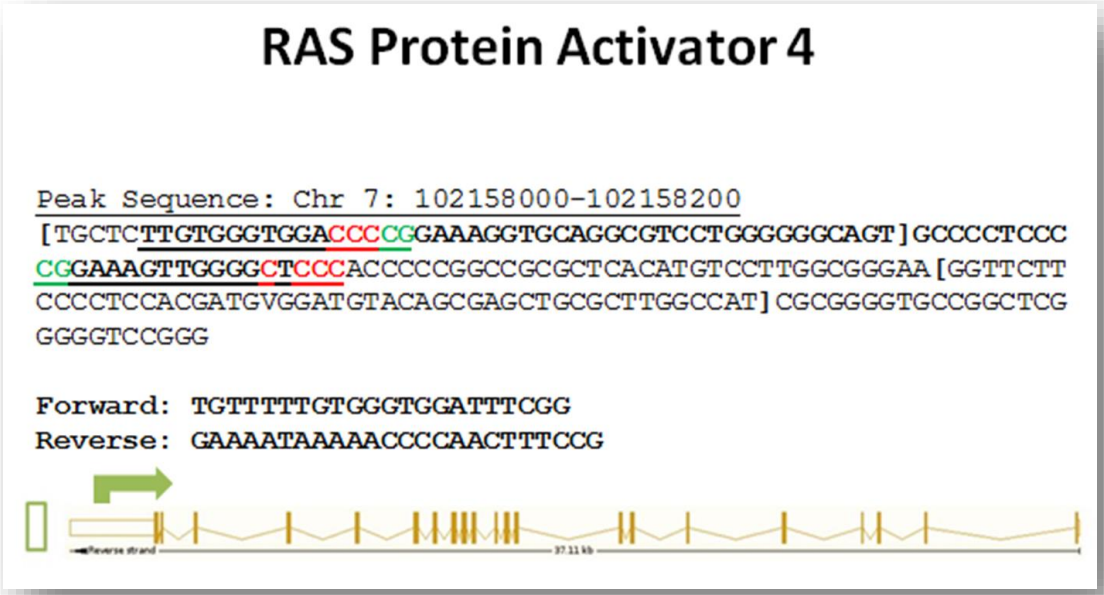


Figure D.3: RASA4 SMART MSP Assay

Sequence and location information on the SMART MSP assay designed for Ras protein activator 4 (RASA4). Probe data taken from the highlighted portion of Figure 3.15, with sequence data from primary assembly GRCh37. Primer sequences were derived from sodium bisulfite treatment of this region, with red cytosines anticipated to deaminate to thymidines. For fully methylated DNA, CpG dinucleotides (highlighted in green) will retain their parent sequence. This MeDIP probe region lies upstream of the RASA4 regulatory region, 7:102158157-102158228, with its TSS at positon 3:102158158.

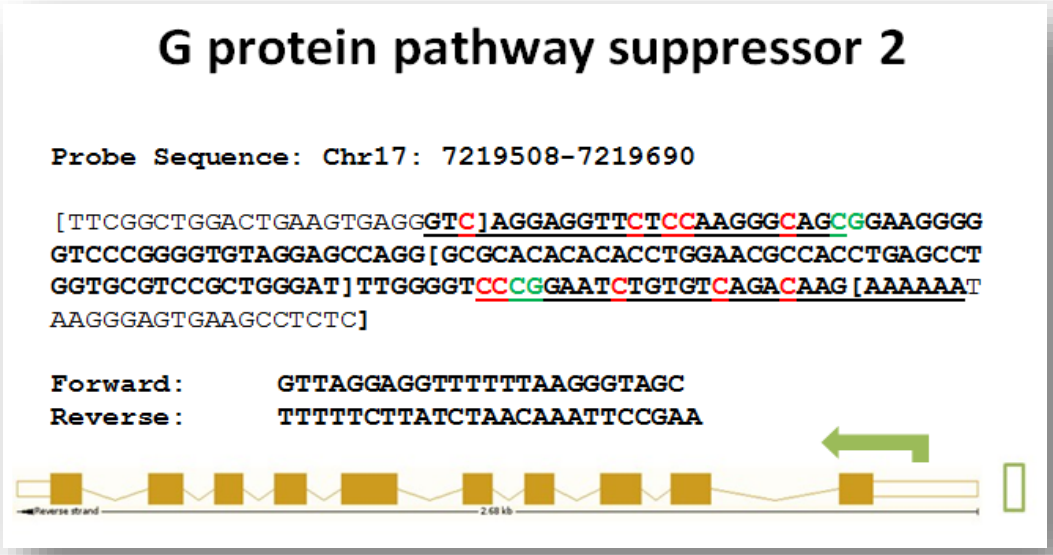
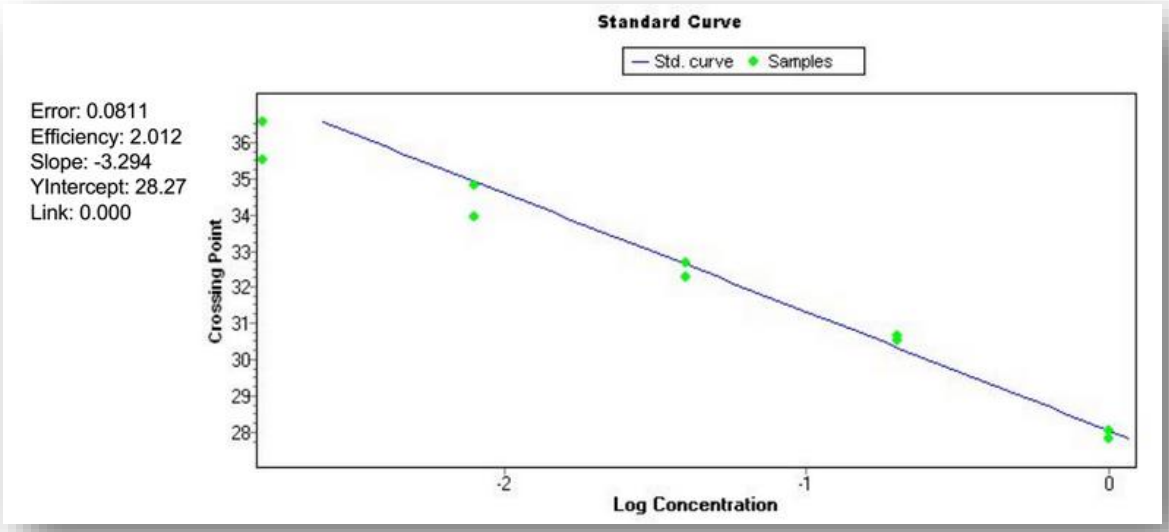


Figure D.4: GPS2 SMART MSP Assay

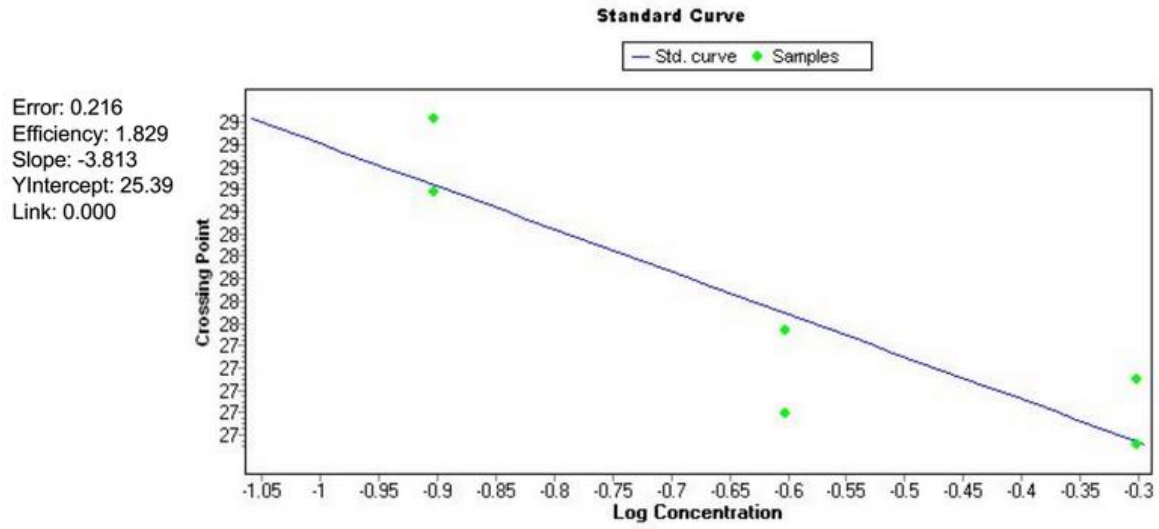
Sequence and location information on the SMART MSP assay designed for G protein pathway suppressor 2 (GPS2). Primer sequences were derived from sodium bisulfite-treatment of this region, with red cytosines anticipated to deaminate to thymidines. For fully methylated DNA, CpG dinucleotides (highlighted in green) will retain their parent sequence. This MeDIP probe region lies upstream of the GPS2 regulatory region, 17:7218371-7218663, with its TSS at positon 3:7218370.



	Cp1	Cp2	Average	STDEV
1 in 5	30.51	30.68	30.6	0.0601
1 in 25	32.66	32.3	32.48	0.1272
1 in 125	34.85	33.94	34.4	0.3217
1 in 625	35.54	36.58	36.06	0.3677

Figure D.5: COL2A1 SMART MSP Assay Standard Curve

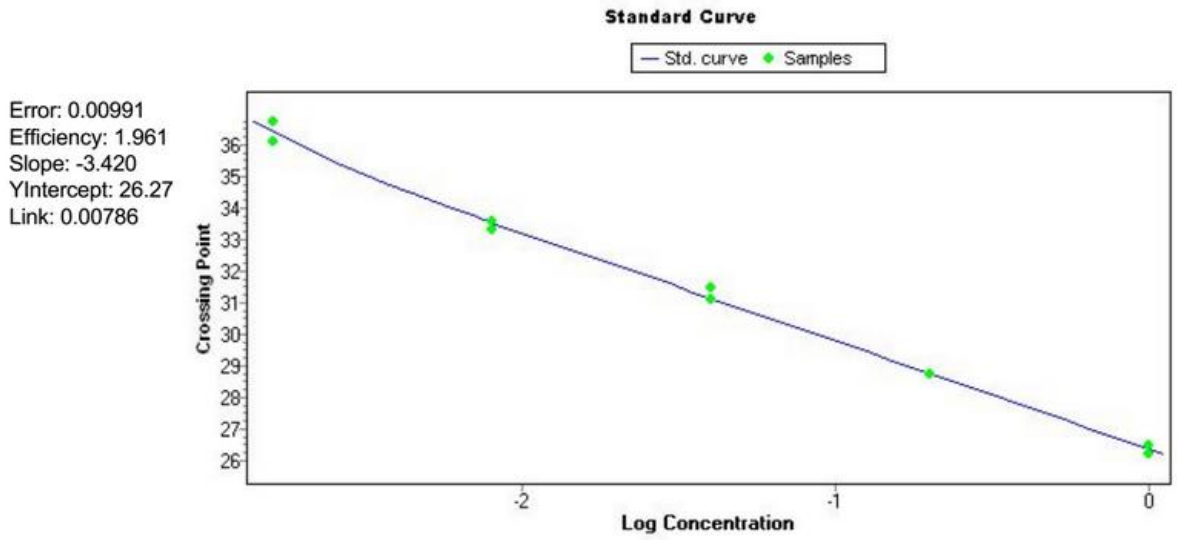
Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a standard dilution of sodium bisulfite treated DNA from 1/5 to 1/625. COL2A1 has a PCR efficiency of 2.012.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
Neat	25.66	25.90	25.78	0.1697	76.81	76.81	76.81	0.0000
1 in 2	27.55	27.70	27.63	0.1061	76.79	76.72	76.76	0.0495
1 in 4	28.17	28.03	28.10	0.0990	76.75	76.72	76.74	0.0212
1 in 8	28.97	29.04	29.01	0.0495	76.78	76.81	76.80	0.0212
1 in 16	29.70	29.93	29.82	0.1626	76.64	76.45	76.55	0.1344
neg	35.50	35.85	35.68	0.2475	74.99	74.96	74.98	0.0212

Figure D.6: IP6K1 SMART MSP Assay Standard Curve

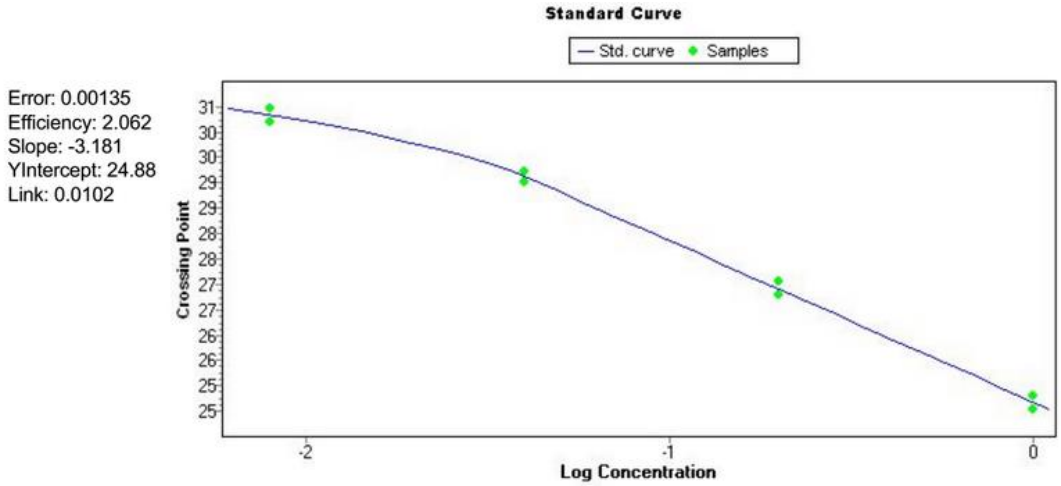
Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of sodium bisulfite treated DNA from 1/2 to 1/8. IP6K1 has a PCR efficiency of 1.829.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
Neat	26.21	26.45	26.33	0.1697	78.88	78.95	78.92	0.0495
1 in 5	28.74	28.75	28.75	0.0071	78.88	78.94	78.91	0.0424
1 in 25	31.08	31.48	31.28	0.2828	78.85	78.96	78.91	0.0778
1 in 125	33.56	33.32	33.44	0.1697	78.95	78.94	78.95	0.0071
1 in 625	36.08	36.73	36.41	0.4596	78.93	78.91	78.92	0.0141
neg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure D.7: Chr9ORF44 SMART MSP Assay Standard Curve

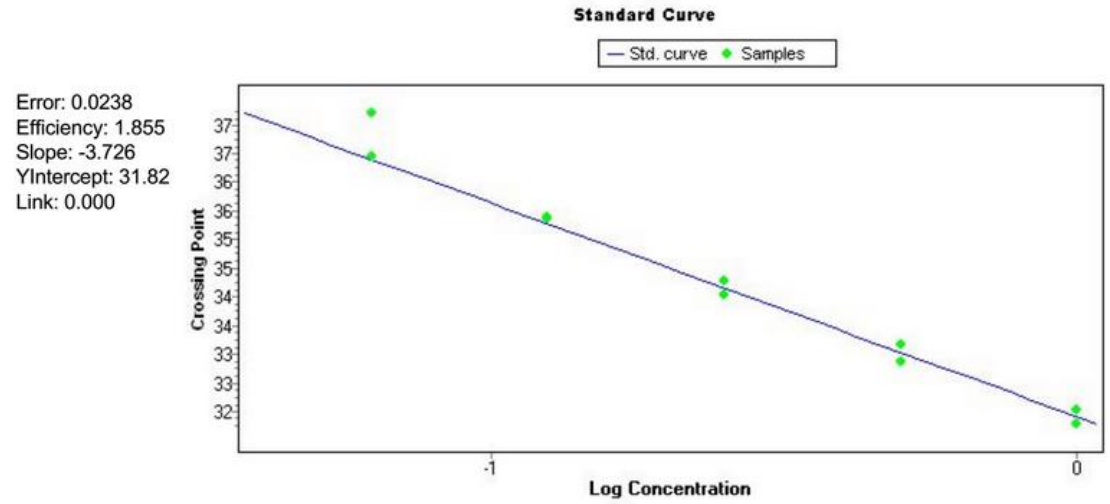
Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of sodium bisulfite treated DNA from 1/2 to 1/8. Chr9ORF44 has a PCR efficiency of 1.961.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
Neat	24.54	24.8	24.67	0.1838	76.96	76.8	76.88	0.1131
1 in 5	27.05	26.78	26.92	0.1909	76.44	76.26	76.35	0.1273
1 in 25	31.08	31.48	31.28	0.2828	76.75	76.15	76.45	0.4243
1 in 125	33.56	33.32	33.44	0.1697	Early Peak	Early Peak	Early Peak	Early Peak
1 in 625	36.08	36.73	36.41	0.4596	Early Peak	Early Peak	Early Peak	Early Peak
neg	34.83	33.50	34.165	0.9405	Early Peak	Early Peak	Early Peak	Early Peak

Figure D.8: RASA4 SMART MSP Assay Standard Curve

Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a serial dilution of sodium bisulfite treated DNA from 1/5 to 1/125. RASA4 has a PCR efficiency of 2.062.



	Cp	Cp	Average	STDEV	Tm	Tm	Average	STDEV
Neat	32.04	31.79	31.92	0.1768	80.62	80.62	80.62	0.0000
1 in 2	32.88	33.17	33.03	0.2051	80.59	80.57	80.58	0.0141
1 in 4	34.04	34.27	34.16	0.1626	80.64	80.62	80.63	0.0141
1 in 8	35.35	35.38	35.37	0.0212	80.68	80.67	80.68	0.0071
1 in 16	36.46	37.20	36.83	0.5233	80.23	80.69	80.46	0.3253
neg	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Figure D.9: GPS2 SMART MSP Assay Standard Curve

Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of sodium bisulfite treated DNA from 1/2 to 1/16. GPS2 has a PCR efficiency of 1.855.

Appendix E:

FASSTT Cohort Data

Appendix E.1: COL2A1 Screening Experiments

	Mean (A) COL2A1	STDEV (A) COL2A1	100% (A) COL2A1	Mean (B) COL2A1	STDEV (B) COL2A1	100% (B) COL2A1	Relative Ratio (A)	Relative Ratio (B)	Verdict
1074 A	29.82	0.0636	29.24	26.10	0.7920	25.11	1.02	1.34	N/A
1074 B	40.00	0.0000	29.24	26.71	0.1414	25.11			
1085 A	28.84	0.0424	29.24	26.15	0.1131	25.11	1.01	1.02	Pass
1085 B	29.54	0.1344	29.24	26.51	0.0071	25.11			
1010 A	29.24	0.0141	29.24	26.31	0.1202	25.11	1.03	1.32	N/A
1010 B	38.54	2.0718	29.24	27.06	0.0354	25.11			
1156 A	32.67	0.1909	29.24	27.13	0.0000	25.11	0.93	1.18	N/A
1156 B	38.39	2.2769	29.24	25.22	0.0354	25.11			
1099 A	29.77	0.0071	29.24	26.79	0.0849	25.11	0.94	0.90	Pass
1099 B	26.87	0.0354	29.24	25.30	0.0849	25.11			
1061 A	30.91	0.1626	29.24	27.01	0.1556	25.11	0.92	0.86	Pass
1061 B	26.72	0.0990	29.24	24.92	0.0424	25.11			
1069 A	35.48	0.0141	29.24	35.14	0.5374	29.02	0.97	0.99	Pass
1069 B	35.07	0.0566	29.24	33.93	0.0707	29.02			
1160 A	36.15	5.4447	29.24	30.85	0.0566	28.35	0.99	1.11	N/A
1160 B	40.00	0.0000	29.24	30.57	1.8243	28.35			
1092 A	32.03	0.3748	29.24	30.95	0.0849	28.35	0.93	0.95	Pass
1092 B	30.49	0.2333	29.24	28.88	0.0566	28.35			
1123 A	30.29	0.0919	29.24	28.65	0.0283	28.35	1.28	1.23	Pass
1123 B	37.31	1.6193	29.24	36.67	0.5798	28.35			

1132 A	33.67	0.1697	29.24	32.07	0.2263	28.35	0.85	0.86	Pass
1132 B	28.91	0.0707	29.24	27.26	0.1061	28.35			
1157 A	30.90	0.1344	29.24	29.35	0.1838	28.35	0.99	0.99	Pass
1157 B	30.61	0.0495	29.24	28.96	0.0636	28.35			
1082 A	33.34	0.4172	29.24	32.53	0.0636	28.35	0.92	0.94	Pass
1082 B	31.25	0.1485	29.24	29.78	0.1202	28.35			
1168 A	33.64	0.1697	29.24	32.28	0.7354	28.35	0.81	0.83	Pass
1168 B	28.03	0.0000	29.24	26.30	0.2687	28.35			
1174 A	36.00	0.2546	29.24	31.72	6.2367	28.35	0.86	0.80	Pass
1174 B	28.71	0.0071	29.24	27.15	0.0849	28.35			
1180 A	35.52	0.7637	29.24	34.89	0.5798	28.35	0.79	0.83	Pass
1180 B	29.36	0.9970	29.24	27.40	0.1768	28.35			
1222 A	35.87	5.8478	29.24	31.74	0.0636	28.35	1.26	1.09	Pass
1222 B	39.23	1.0889	29.24	40.00	0.0000	28.35			
1224 A	32.99	0.0778	29.24	30.95	0.0566	28.35	0.91	0.90	Pass
1224 B	29.82	0.0000	29.24	28.08	0.0566	28.35			
1003 A	33.50	0.1980	29.24	31.49	0.3889	28.35	0.89	0.88	Pass
1003 B	29.59	0.0071	29.24	28.00	0.0141	28.35			
1004 A	38.52	2.1001	29.24	21.01	0.1344	25.11	1.24	0.71	N/A
1004 B	27.47	0.0566	29.24	26.05	0.0071	27.98			
1005 A	27.70	0.0424	29.26	26.69	0.0919	28.35	1.00	1.04	Pass
1005 B	28.76	0.2616	29.26	26.65	0.0990	28.35			
1007 A	27.35	0.9192	29.26	25.78	0.0707	28.35	1.07	1.08	Pass
1007 B	29.63	0.4808	29.26	27.57	0.1344	28.35			
1008 A	29.53	0.7920	29.26	27.24	0.0990	28.35	1.01	0.99	Pass

1008 B	29.21	0.3606	29.26	27.58	0.0778	28.35			
1009 A	30.95	0.4243	29.26	29.07	0.2192	28.35	0.95	0.95	Pass
1009 B	29.26	0.2828	29.26	27.69	0.0354	28.35			
1012 A	30.38	0.4667	29.26	28.61	0.1202	28.35	0.94	0.94	Pass
1012 B	28.61	0.0566	29.26	26.84	0.0354	28.35			
1018 A	28.45	0.1061	29.26	26.13	0.0566	28.35	1.02	1.06	Pass
1018 B	30.06	0.0849	29.26	26.60	0.0141	28.35			
1019 A	40.00	0.0000	29.26	40.00	0.0000	28.35	0.64	0.71	Pass
1019 B	28.28	0.0849	29.26	25.77	0.1273	28.35			
1020 A	28.75	0.2192	29.26	25.68	0.1485	28.35	1.10	1.04	Pass
1020 B	29.94	0.6081	29.26	28.23	0.1273	28.35			
1022 A	29.81	0.2404	29.26	29.87	0.0354	29.02	1.03	1.08	Pass
1022 B	32.06	0.0636	29.26	30.63	0.1768	29.02			
1023 A	31.73	0.1061	29.26	28.99	0.0141	29.02	0.96	0.99	Pass
1023 B	31.40	0.1485	29.26	27.88	0.0919	29.02			
1024 A	32.84	0.2970	29.26	30.08	0.1344	29.02	0.92	0.96	Pass
1024 B	31.47	0.6223	29.26	27.62	0.0000	29.02			
1028 A	29.90	0.1980	29.26	27.93	0.0283	29.02	0.96	0.92	Pass
1028 B	27.37	0.1909	29.26	26.95	0.0495	29.02			
1030 A	25.22	0.1131	29.26	24.20	0.0071	29.02	1.22	1.21	Pass
1030 B	30.53	0.2687	29.26	29.48	0.0566	29.02			
1031 A	34.94	0.1061	29.26	31.16	0.0212	29.02	1.23	1.14	Pass
1031 B	40.00	0.0000	29.26	38.46	2.1779	29.02			
1032 A	40.00	0.0000	29.26	38.54	2.0648	29.02	1.04	1.00	Pass
1032 B	40.00	0.0000	29.26	40.00	0.0000	29.02			

1033 A	27.01	0.4738	29.26	26.25	0.0141	29.02	0.96	1.05	Pass
1033 B	28.49	0.7920	29.26	25.32	3.2032	29.02			
1035 A	31.28	0.1061	29.26	29.80	0.2333	27.98	0.91	0.95	Pass
1035 B	29.68	0.6293	29.26	27.04	0.1273	27.98			
1037 A	35.10	0.0990	29.26	21.02	0.4313	25.11	1.32	0.90	Fail
1037 B	31.44	0.0778	29.26	27.70	0.0283	27.98			
1041 A	32.72	0.2404	29.26	29.59	0.0141	27.98	0.98	1.00	Pass
1041 B	32.79	0.1838	29.26	29.11	0.5303	27.98			
1042 A	34.88	0.1273	29.26	30.86	0.1768	27.98	0.94	0.93	Pass
1042 B	32.35	0.9405	29.26	29.04	0.0990	27.98			
1043 A	30.16	0.5303	28.56	29.47	0.0071	27.98	0.92	0.95	Pass
1043 B	28.52	0.6293	28.56	27.25	0.0424	27.98			
1050 A	28.97	0.4879	28.56	29.05	0.6293	27.98	0.87	0.90	Pass
1050 B	26.10	0.7071	28.56	25.17	0.0636	27.98			
1052 A	28.28	0.6647	28.56	28.37	1.2374	27.98	0.97	1.04	Pass
1052 B	29.48	1.2940	28.56	27.65	0.0495	27.98			
1053 A	28.26	0.3465	28.56	21.52	0.1344	25.11	1.29	0.95	Fail
1053 B	26.71	0.0707	28.56	27.70	0.7142	27.98			
1062 A	31.83	0.2616	28.56	26.81	0.1414	27.98	1.03	0.90	Fail
1062 B	28.76	0.0495	28.56	27.60	0.0071	27.98			
1063 A	31.83	0.2616	28.56	31.32	0.3606	27.98	0.88	0.90	Pass
1063 B	28.78	0.2121	28.56	27.58	0.0424	27.98			
1064 A	30.61	0.0778	28.56	29.50	0.0495	27.98	0.92	0.92	Pass
1064 B	28.24	0.0424	28.56	27.27	0.0849	27.98			
1066 A	31.20	0.0990	28.56	30.25	0.0495	27.98	0.96	0.96	Pass

1066 B	29.94	0.3889	28.56	28.93	0.0141	27.98			
1067 A	29.68	0.1980	28.56	29.43	0.2051	28.50	1.03	0.97	Fail
1067 B	28.76	0.0495	28.56	30.17	0.4384	28.50			
1070 A	31.83	0.2616	28.56	24.76	0.0424	28.50	1.10	0.92	Fail
1070 B	29.33	0.7707	28.56	27.23	0.0849	28.50			
1076 A	31.06	0.3465	28.56	30.20	0.0354	28.50	0.94	0.92	Pass
1076 B	28.59	0.0778	28.56	28.51	0.0141	28.50			
1077 A	33.76	4.2497	28.56	30.18	0.0354	28.50	0.77	0.80	Pass
1077 B	27.03	0.0849	28.56	23.33	0.5728	25.11			
1078 A	28.54	0.4525	28.56	28.29	0.2475	28.50	0.96	0.98	Pass
1078 B	27.88	0.1061	28.56	27.05	0.0071	28.50			
1090 A	32.81	0.2687	28.56	29.75	0.1202	28.50	0.98	0.93	Pass
1090 B	30.40	0.0071	28.56	29.28	0.1273	28.50			
1091 A	28.84	0.0000	28.56	27.73	0.0849	28.50	0.93	0.93	Pass
1091 B	26.89	0.5162	28.56	25.83	0.0778	28.50			
1093 A	29.35	0.3394	28.56	29.57	0.0919	28.50	0.84	0.88	Pass
1093 B	25.95	0.4031	28.56	24.76	0.0636	28.50			
1095 A	31.64	0.3323	28.56	30.52	0.3394	28.50	0.93	0.99	Pass
1095 B	31.33	0.4101	28.56	28.43	0.1273	28.50			
1100 A	30.71	0.2263	28.56	28.74	0.0000	28.50	1.06	1.01	Pass
1100 B	31.11	0.2758	28.56	30.37	0.1414	28.50			
1102 A	28.47	0.2263	28.56	27.82	0.0212	28.50	1.01	1.02	Pass
1102 B	28.96	0.0212	28.56	27.98	0.1485	28.50			
1107 A	28.40	0.1909	28.56	27.98	0.0000	28.50	1.01	1.08	Pass
1107 B	30.79	0.8415	28.56	28.29	0.0566	28.50			

1108 A	35.78	0.6788	28.56	28.49	0.4384	28.50	0.98	1.00	Pass
1108 B	35.69	1.6829	28.56	27.79	0.0424	28.50			
1109 A	34.38	1.1526	28.56	29.96	0.4950	28.50	0.71	0.87	Pass
1109 B	30.08	0.6010	28.56	21.37	0.0000	25.11			
1110 A	33.11	1.9658	28.56	29.51	0.0212	28.50	0.91	0.97	Pass
1110 B	32.07	0.5869	28.56	26.95	0.4243	28.50			
1112 A	33.57	0.2899	28.56	30.27	0.0212	28.50	0.82	0.86	Pass
1112 B	28.82	0.8202	28.56	24.94	0.0354	28.50			
1113 A	31.73	0.3182	28.56	27.15	0.0283	28.50	1.01	1.07	Pass
1113 B	34.08	0.0849	28.56	27.48	0.0495	28.50			
1114 A	36.07	0.7778	28.56	29.75	0.0636	28.50	0.91	0.94	Pass
1114 B	34.05	0.0636	28.56	27.22	0.1061	28.50			
1114 B	33.41	0.2192	28.56	22.84	0.1414	25.11			
1115 A	35.49	0.9405	28.56	22.09	0.1697	25.11	1.17	0.81	Fail
1115 B	28.74	2.0718	28.56	25.89	0.0071	28.50			
1116 A	31.88	0.2121	28.56	29.95	0.1980	28.94	0.94	0.98	Pass
1116 B	31.11	0.2404	28.56	28.07	0.1061	28.94			
1118 A	34.53	0.4879	28.56	29.28	0.1131	28.94	0.77	1.06	Fail
1118 B	36.55	0.9970	28.56	22.64	0.0424	28.94			
1119 A	32.60	0.0566	28.56	27.51	0.0071	28.94	0.96	6.00	Fail
1119 B	32.10	0.3818	28.56	26.30	0.0424	28.94			
1120 A	35.51	0.6505	28.56	28.76	0.0636	28.94	0.95	1.44	Fail
1120 B	30.57	0.0566	28.56	27.41	0.0919	28.94			
1122 A	30.23	0.0707	29.20	29.60	0.0354	28.94	0.91	0.92	Pass
1122 B	27.80	0.0354	29.20	26.91	0.0212	28.94			

1126 A	30.83	0.0141	29.20	29.92	0.1626	28.94	0.94	0.92	Pass
1126 B	28.46	0.1556	29.20	28.04	0.0071	28.94			
1127 A	27.62	0.1273	29.20	26.69	0.0495	28.94	1.00	0.99	Pass
1127 B	27.41	0.0566	29.20	26.76	0.0141	28.94			
1131 A	28.38	0.1061	29.20	27.64	0.0212	28.94	1.11	1.00	Fail
1131 B	28.41	0.3818	29.20	30.67	0.8556	28.94			
1136 A	31.93	0.0636	29.20	30.37	0.1131	28.94	0.91	0.91	Pass
1136 B	29.03	0.0990	29.20	27.69	0.0354	28.94			
1138 A	33.42	3.3375	29.20	29.75	0.4031	28.94	0.85	0.79	Pass
1138 B	26.26	0.0778	29.20	25.22	0.0636	28.94			
1139 A	36.40	5.0912	29.20	30.52	0.3465	28.94	0.91	0.79	Pass
1139 B	28.80	0.6364	29.20	27.81	0.4738	28.94			
1141 A	31.01	0.3748	29.20	30.91	0.1061	28.94	0.81	0.82	Pass
1141 B	25.40	0.2192	29.20	24.99	0.0141	28.94			
1144 A	24.91	0.0424	29.20	24.10	0.0141	28.94	1.15	1.14	Pass
1144 B	28.48	0.2192	29.20	27.76	0.0071	28.94			
1145 A	30.50	0.0424	29.20	29.28	0.0636	28.94	0.92	0.91	Pass
1145 B	27.85	0.0849	29.20	26.91	0.0141	28.94			
1151 A	29.63	0.0424	29.20	28.81	0.1414	28.94	1.04	1.02	Pass
1151 B	30.21	0.0212	29.20	29.88	0.0495	28.94			
1152 A	31.57	0.0424	29.20	30.04	0.1909	28.94	0.92	0.89	Pass
1152 B	28.06	0.0495	29.20	27.66	0.0212	28.94			
1153 A	30.88	0.0424	29.20	30.40	0.1344	28.94	0.88	0.89	Pass
1153 B	27.35	0.1838	29.20	26.84	0.0566	28.94			
1154 A	31.94	0.0283	29.20	31.25	0.1768	28.94	0.81	0.81	Pass

1154 B	25.95	0.0354	29.20	25.44	0.5020	28.94			
1155 A	28.62	0.0778	29.20	28.09	0.1980	28.94	0.98	0.98	Pass
1155 B	28.04	0.2051	29.20	27.59	0.0071	28.94			
1161 A	31.07	0.0212	29.20	31.37	0.1414	28.53	0.84	0.82	Pass
1161 B	25.53	0.0354	29.20	26.44	0.0000	28.53			
1163 A	24.89	0.0495	29.20	31.08	0.3394	28.53	0.88	1.14	Fail
1163 B	28.47	0.0071	29.20	27.26	0.4525	28.53			
1166 A	30.52	0.0212	29.20	25.62	0.0778	28.53	1.08	0.91	Fail
1166 B	27.77	0.0636	29.20	27.68	0.0849	28.53			
1171 A	29.77	0.1202	29.20	26.97	0.0566	28.53	1.02	1.01	Pass
1171 B	29.99	0.1485	29.20	27.64	0.1556	28.53			
1173 A	26.72	0.0566	29.17	26.03	0.0495	28.53	1.05	1.05	Pass
1173 B	27.96	0.0849	29.17	27.25	0.1556	28.53			
1177 A	28.34	0.0283	29.17	27.75	0.1838	28.53	0.99	0.97	Pass
1177 B	27.63	0.2333	29.17	27.35	0.4313	28.53			
1183 A	30.83	0.1273	29.17	22.33	0.2758	25.11	1.29	0.96	Fail
1183 B	29.70	0.0566	29.17	28.91	0.0354	28.53			
1185 A	28.67	0.2333	29.17	27.72	0.0354	28.53	1.09	1.01	Pass
1185 B	29.06	0.2263	29.17	30.19	2.7224	28.53			
1185 B	27.35	0.0141	29.17	22.92	0.8061	25.11			
1187 A	30.02	0.2051	29.17	22.68	0.5445	25.11	1.15	0.90	Fail
1187 B	27.05	0.0141	29.17	26.17	0.0849	28.53			
1188 A	31.58	0.3323	29.17	23.83	0.2192	25.11	1.16	0.90	Fail
1188 B	28.28	0.0849	29.17	27.66	0.0212	28.53			
1189 A	33.77	0.0919	29.17	33.23	0.0636	28.53	0.88	0.89	Pass

1189 B	30.07	0.1202	29.17	29.30	0.0919	28.53			
1190 A	30.91	0.2192	29.17	30.15	0.0919	28.53	0.99	0.99	Pass
1190 B	30.54	0.0424	29.17	29.72	0.0424	28.53			
1192 A	32.92	0.0566	29.17	32.43	0.2546	28.53	0.89	0.89	Pass
1192 B	29.34	0.0071	29.17	28.91	0.0849	28.53			
1193 A	33.15	0.1909	29.17	33.00	0.2546	28.53	0.89	0.90	Pass
1193 B	29.74	0.1697	29.17	29.22	0.0707	28.53			
1194 A	31.86	0.0424	29.17	31.67	0.0283	28.53	0.96	0.96	Pass
1194 B	30.65	0.0849	29.17	30.31	0.0566	28.53			
1195 A	31.52	0.3465	29.17	31.30	0.0495	28.53	0.92	0.94	Pass
1195 B	29.48	0.0354	29.17	28.94	0.0212	28.53			
1197 A	31.44	0.0919	29.17	31.64	0.0000	28.53	0.95	0.97	Pass
1197 B	30.47	0.2828	29.17	29.98	0.0071	28.53			
1198 A	30.91	0.1202	29.17	30.58	0.1202	28.53			
1199 A	32.70	0.1061	29.17	32.91	0.1838	28.53	0.99	1.00	Pass
1199 B	32.62	0.4313	29.17	32.61	0.0071	28.53			
1200 A	35.89	0.8132	29.17	40.00	0.0000	28.53	0.75	0.92	Pass
1200 B	33.04	3.4436	29.17	30.20	0.0212	28.53			
1201 A	30.63	0.0495	29.58	32.81	0.1909	28.53	0.96	0.96	Pass
1201 B	29.35	0.0000	29.58	31.43	0.0636	28.53			
1203 A	35.60	1.2445	29.58	40.00	0.0000	28.53	0.79	0.84	Pass
1203 B	29.83	0.0354	29.58	31.77	0.0212	28.53			
1205 A	40.00	0.0000	29.58	40.00	0.0000	29.02	0.69	0.71	Pass
1205 B	28.44	0.0071	29.58	27.60	0.0071	29.02			
1206 A	27.50	0.0071	29.58	29.48	0.2970	28.53	1.01	1.01	Pass

1206 B	27.81	0.0566	29.58	29.65	0.0849	28.53			
1209 A	30.94	0.1061	29.58	32.68	0.1980	28.53	0.95	0.94	Pass
1209 B	29.07	0.0636	29.58	30.91	0.0636	28.53			
1211 A	35.63	0.8344	29.58	40.00	0.0000		0.78	0.83	Pass
1211 B	29.69	0.1414	29.58	31.12	0.0424	28.53			
1212 A	31.27	0.0283	29.58	32.78	0.0778	28.53	0.93	0.92	Pass
1212 B	28.72	0.1909	29.58	30.34	0.1838	28.53			
1214 A	29.09	0.0424	29.58	31.18	0.0636	28.53	0.96	0.96	Pass
1214 B	28.01	0.0424	29.58	29.92	0.0141	28.53			
1215 A	30.58	0.0849	29.58	32.36	0.1344	28.53	0.82	0.81	Pass
1215 B	24.83	0.0919	29.58	26.66	0.0000	28.53			
1216 A	28.55	0.1131	29.58	31.56	0.1485	28.53	0.84	0.93	Pass
1216 B	26.45	0.1909	29.58	26.38	2.6799	28.53			
1217 A	27.69	0.0354	29.58	30.06	0.1414	28.53	1.05	1.04	Pass
1217 B	28.91	0.0778	29.58	31.42	0.0849	28.53			
1218 A	30.80	0.0566	29.58	33.73	0.2404	28.53	1.19	1.19	Pass
1218 B	36.66	1.0112	29.58	40.00	0.0000				
1219 A	31.93	0.1414	29.58	33.87	0.0707	28.53	0.92	0.92	Pass
1219 B	29.37	0.1556	29.58	31.06	0.3677	28.53			
1220 A	31.33	0.2051	29.58	30.68	0.0000	28.53	1.04	0.95	Fail
1220 B	29.65	0.1273	29.58	31.80	0.2121	28.53			
1225 A	30.99	0.1909	29.58	33.73	0.2333	28.53	0.93	0.94	Pass
1225 B	29.18	0.0141	29.58	31.34	0.0212	28.53			
1226 A	28.74	0.0283	29.58	30.55	0.0566	28.53	0.99	0.98	Pass
1226 B	28.23	0.0000	29.58	30.16	0.1344	28.53			

Appendix E.2: IP6K1 SMART-MSP Raw Data

Sample Name	Methylation %	Intervention	Methylation % Change	Outcome
1003 A	1.83	2		
1003 B	1.54	2	-0.29	No Change
1005 A	11.11	2		
1005 B	6.70	2	-4.41	No Change
1007 A	15.18	1		
1007 B	14.66	1	-0.52	No Change
1008 A	12.20	1		
1008 B	12.12	1	-0.08	No Change
1009 A	14.01	2		
1009 B	14.76	2	0.75	No Change
1012 A	11.30	2		
1012 B	5.31	2	-5.99	No Change
1018 A	7.08	2		
1018 B	5.27	2	-1.81	No Change
1022 A	18.11	2		
1022 B	14.61	2	-3.50	No Change
1023 A	8.36	1		
1023 B	7.36	1	-1.01	No Change
1024 A	9.67	1		
1024 B	2.56	1	-7.11	No Change

1028 A	6.12	2		
1028 B	8.57	2	2.45	No Change
1030 A	6.06	1		
1030 B	13.26	1	7.20	No Change
1033 A	7.43	1		
1033 B	1.34	1	-6.09	No Change
1037 A	0.18	2		
1037 B	9.54	2	9.36	No Change
1041 A	12.16	2		
1041 B	10.58	2	-1.57	No Change
1042 A	25.88	1		
1042 B	18.62	1	-7.26	No Change
1043 A	1.05	1		
1043 B	0.27	1	-0.78	No Change
1050 A	1.13	2		
1050 B	0.23	2	-0.90	No Change
1052 A	2.70	1		
1052 B	3.67	1	0.96	No Change
1053 A	0.34	2		
1053 B	4.00	2	3.66	No Change
1061 A	4.42	1		
1061 B	16.04	1	11.62	Up
1062 A	2.36	1		

1062 B	3.90	1	1.54	No Change
1063 A	34.99	2		
1063 B	2.57	2	-32.42	Down
1064 A	1.68	2		
1064 B	2.23	2	0.54	No Change
1066 A	3.89	2		
1066 B	3.44	2	-0.44	No Change
1067 A	3.44	2		
1067 B	1.29	2	-2.16	No Change
1069 A	6.89	2		
1069 B	4.04	2	-2.85	No Change
1070 A	1.63	2		
1070 B	0.94	2	-0.69	No Change
1074 A	9.51	1		
1074 B	16.38	1	6.87	No Change
1076 A	3.78	2		
1076 B	3.85	2	0.07	No Change
1077 A	5.87	1		
1077 B	1.33	1	-4.54	No Change
FIP1078 A	4.74	1		
1078 B	3.25	1	-1.49	No Change
1082 A	4.80	2		
1082 B	2.58	2	-2.22	No Change

1085 A	15.02	2		
1085 B	10.26	2	-4.76	No Change
1091 A	2.07	1		
1091 B	1.55	1	-0.52	No Change
1092 A	3.73	2		
1092 B	2.56	2	-1.17	No Change
1093 A	6.23	1		
1093 B	1.72	1	-4.51	No Change
1095 A	0.63	2		
1095 B	0.16	2	-0.47	No Change
1099 A	10.29	2		
1099 B	15.88	2	5.58	No Change
1100 A	0.01	1		
1100 B	0.04	1	0.03	No Change
1102 A	0.15	2		
1102 B	0.31	2	0.17	No Change
1107 A	0.32	1		
1107 B	3.69	1	3.38	No Change
1108 A	0.84	2		
1108 B	0.99	2	0.15	No Change
1109 A	4.12	1		
1109 B	0.62	1	-3.50	No Change
1110 A	3.49	1		

1110 B	2.78	1	-0.71	No Change
1112 A	4.31	2		
1112 B	1.90	2	-2.41	No Change
1113 A	2.03	1		
1113 B	1.60	1	-0.43	No Change
1114 A	2.86	2		
1114 B	1.79	2	-1.06	No Change
1115 A	0.11	2		
1115 B	3.30	2	3.19	No Change
1116 A	4.58	1		
1116 B	2.60	1	-1.97	No Change
1118 A	2.18	2		
1118 B	0.28	2	-1.89	No Change
1119 A	1.41	1		
1119 B	1.26	1	-0.15	No Change
1120 A	0.88	1		
1120 B	2.75	1	1.87	No Change
1122 A	4.15	1		
1122 B	1.70	1	-2.45	No Change
1123 A	2.13	1		
1123 B	20.88	1	18.74	Up
1132 A	2.68	1		
1132 B	0.08	1	-2.61	No Change

1136 A	1.67	1		
1136 B	1.37	1	-0.30	No Change
1138 A	1.15	1		
1138 B	0.62	1	-0.54	No Change
1139 A	2.08	1		
1139 B	3.69	1	1.61	No Change
1141 A	10.77	1		
1141 B	1.79	1	-8.98	No Change
1144 A	1.47	2		
1144 B	1.50	2	0.03	No Change
1145 A	1.73	1		
1145 B	1.31	1	-0.42	No Change
1151 A	1.86	1		
1151 B	4.33	1	2.47	No Change
1152 A	4.17	1		
1152 B	1.40	1	-2.77	No Change
1153 A	4.04	2		
1153 B	1.97	2	-2.07	No Change
1154 A	6.14	1		
1154 B	1.73	1	-4.41	No Change
1155 A	3.47	1		
1155 B	2.30	1	-1.16	No Change
1156 A	2.55	2		

1156 B	29.02	2	26.47	Up
1157 A	2.60	1		
1157 B	2.70	1	0.10	No Change
1160 A	2.23	2		
1160 B	23.65	2	21.42	Up
1161 A	14.66	1		
1161 B	1.63	1	-13.03	Down
1163 A	6.14	2		
1163 B	1.02	2	-5.12	No Change
1166 A	0.63	2		
1166 B	2.34	2	1.70	No Change
1168 A	1.70	1		
1168 B	0.89	1	-0.81	No Change
1171 A	1.90	1		
1171 B	2.26	1	0.36	No Change
1173 A	4.58	1		
1173 B	2.97	1	-1.61	No Change
1174 A	1.03	1		
1174 B	1.53	1	0.50	No Change
1177 A	1.65	1		
1177 B	1.18	1	-0.46	No Change
1180 A	11.34	1		
1180 B	1.47	1	-9.87	No Change

1183 A	0.18	1		
1183 B	2.34	1	2.16	No Change
1185 A	2.10	2		
1185 B	11.87	2	9.77	No Change
1187 A	0.39	1		
1187 B	0.15	1	-0.24	No Change
1188 A	1.13	2		
1188 B	0.54	2	-0.58	No Change
1190 A	5.42	1		
1190 B	16.49	1	11.07	Up
1192 A	100.00	2		
1192 B	24.40	2	-75.60	Down
1194 A	100.00	2		
1194 B	78.46	2	-21.54	Down
1195 A	100.00	2		
1195 B	28.62	2	-71.38	Down
1197 A	75.26	2		
1197 B	67.83	2	-7.43	No Change
1201 A	5.11	2		
1201 B	2.08	2	-3.03	No Change
1206 A	0.54	1		
1206 B	0.70	1	0.17	No Change
1209 A	4.02	1		

1209 B	1.28	1	-2.74	No Change
1212 A	13.12	1		
1212 B	3.29	1	-9.83	No Change
1214 A	1.19	2		
1214 B	0.68	2	-0.51	No Change
1215 A	2.23	1		
1215 B	0.11	1	-2.12	No Change
1216 A	0.57	2		
1216 B	0.22	2	-0.35	No Change
1217 A	0.36	2		
1217 B	0.25	2	-0.12	No Change
1219 A	1.19	1		
1219 B	0.22	1	-0.97	No Change
1220 A	0.88	2		
1220 B	0.30	2	-0.59	No Change
1225 A	1.02	2		
1225 B	0.35	2	-0.67	No Change
1226 A	0.10	2		
1226 B	0.03	2	-0.07	No Change
Intervention: Folic Acid Supplementation 1, Placebo, 2				

Appendix E.3: Chr9ORF44 SMART-MSP Raw Data

Sample Name	Methylation %	Group	Methylation % Change	Outcome

1003 A	17.80	2		
1003 B	23.90	2	6.10	No Change
1005 A	53.77	2		
1005 B	43.68	2	-10.10	Down
1007 A	44.75	1		
1007 B	50.52	1	5.77	No Change
1008 A	41.90	1		
1008 B	35.97	1	-5.93	No Change
1009 A	25.26	2		
1009 B	55.67	2	30.41	Up
1012 A	37.89	2		
1012 B	40.75	2	2.86	No Change
1018 A	33.56	2		
1018 B	36.35	2	2.79	No Change
1022 A	61.99	2		
1022 B	17.99	2	-44.00	Down
1023 A	33.22	1		
1023 B	34.03	1	0.82	No Change
1024 A	34.51	1		
1024 B	40.47	1	5.96	No Change
1028 A	45.85	2		
1028 B	53.03	2	7.18	No Change
1030 A	41.04	1		

1030 B	35.97	1	-5.06	No Change
1033 A	45.06	1		
1033 B	11.19	1	-33.87	Down
1037 A	0.07	2		
1037 B	33.68	2	33.61	Up
1041 A	30.99	2		
1041 B	23.41	2	-7.59	No Change
1042 A	19.14	1		
1042 B	41.61	1	22.47	Up
1043 A	50.70	1		
1043 B	40.75	1	-9.94	No Change
1050 A	66.90	2		
1050 B	62.42	2	-4.48	No Change
1052 A	92.34	1		
1052 B	36.10	1	-56.24	Down
1053 A	3.87	2		
1053 B	69.74	2	65.86	Up
1061 A	64.39	1		
1061 B	100.00	1	35.61	Up
1062 A	68.07	1		
1062 B	38.96	1	-29.11	Down
1063 A	52.67	2		
1063 B	49.48	2	-3.19	No Change

1064 A	33.22	2		
1064 B	35.11	2	1.89	No Change
1066 A	29.02	2		
1066 B	40.61	2	11.60	Up
1067 A	35.85	2		
1067 B	20.52	2	-15.33	Down
1069 A	47.14	2		
1069 B	19.14	2	-27.99	Down
1070 A	54.53	2		
1070 B	31.10	2	-23.43	Down
1074 A	92.66	1		
1074 B	32.99	1	-59.67	Down
1076 A	26.06	2		
1076 B	43.08	2	17.02	Up
1077 A	32.09	1		
1077 B	22.61	1	-9.48	No Change
FIP1078 A	47.14	1		
1078 B	41.75	1	-5.38	No Change
1082 A	30.15	2		
1082 B	31.64	2	1.50	No Change
1085 A	39.37	2		
1085 B	25.00	2	-14.37	Down
1091 A	30.46	1		

1091 B	28.52	1	-1.94	No Change
1092 A	42.93	2		
1092 B	38.16	2	-4.77	No Change
1093 A	51.94	1		
1093 B	34.75	1	-17.19	Down
1095 A	24.57	2		
1095 B	18.49	2	-6.08	No Change
1099 A	100.00	2		
1099 B	100.00	2	0.00	No Change
1100 A	22.30	1		
1100 B	28.03	1	5.73	No Change
1102 A	31.32	2		
1102 B	27.17	2	-4.15	No Change
1107 A	37.89	1		
1107 B	26.52	1	-11.38	Down
1108 A	14.56	2		
1108 B	16.72	2	2.16	No Change
1109 A	42.04	1		
1109 B	9.15	1	-32.89	Down
1110 A	40.61	1		
1110 B	48.30	1	7.68	No Change
1112 A	24.91	2		
1112 B	38.56	2	13.64	Up

1113 A	27.26	1		
1113 B	22.93	1	-4.34	No Change
1114 A	26.24	2		
1114 B	29.22	2	2.98	No Change
1115 A	0.85	2		
1115 B	33.10	2	32.26	Up
1116 A	44.14	1		
1116 B	19.68	1	-24.45	Down
1118 A	36.73	2		
1118 B	21.84	2	-14.89	Down
1119 A	25.53	1		
1119 B	31.10	1	5.57	No Change
1120 A	18.49	1		
1120 B	35.11	1	16.62	Up
1122 A	30.35	1		
1122 B	22.38	1	-7.98	No Change
1123 A	37.37	1		
1123 B	80.39	1	43.01	Up
1132 A	26.52	1		
1132 B	38.56	1	12.04	Up
1136 A	15.88	1		
1136 B	21.10	1	5.22	No Change
1138 A	21.39	1		

1138 B	21.69	1	0.30	No Change
1139 A	21.61	1		
1139 B	56.06	1	34.44	Up
1141 A	47.47	1		
1141 B	47.96	1	0.50	No Change
1144 A	33.56	2		
1144 B	35.23	2	1.67	No Change
1145 A	13.82	1		
1145 B	23.57	1	9.75	No Change
1151 A	25.88	1		
1151 B	41.47	1	15.58	Up
1152 A	11.50	1		
1152 B	34.27	1	22.77	Up
1153 A	31.75	2		
1153 B	41.18	2	9.43	No Change
1154 A	32.42	1		
1154 B	33.68	1	1.26	No Change
1155 A	42.04	1		
1155 B	36.73	1	-5.32	No Change
1156 A	41.32	2		
1156 B	34.27	2	-7.05	No Change
1157 A	33.22	1		
1157 B	38.56	1	5.34	No Change

1161 A	34.27	1		
1161 B	17.80	1	-16.47	Down
1163 A	15.12	2		
1163 B	25.26	2	10.14	Up
1166 A	45.22	2		
1166 B	32.53	2	-12.69	Down
1168 A	55.86	1		
1168 B	34.39	1	-21.48	Down
1171 A	50.87	1		
1171 B	65.98	1	15.10	Up
1173 A	70.47	1		
1173 B	42.63	1	-27.83	Down
1174 A	3.47	1		
1174 B	44.44	1	40.97	Up
1177 A	34.27	1		
1177 B	52.67	1	18.40	Up
1180 A	68.54	1		
1180 B	39.37	1	-29.17	Down
1183 A	1.91	1		
1183 B	35.11	1	33.20	Up
1185 A	37.63	2		
1185 B	12.29	2	-25.35	Down
1187 A	32.53	1		

1187 B	2.80	1	-29.74	Down
1188 A	2.19	2		
1188 B	41.04	2	38.85	Up
1189 A	0.00	1		
1189 B	34.51	1	34.51	Up
1190 A	36.60	1		
1190 B	34.03	1	-2.57	No Change
1192 A	28.32	2		
1192 B	54.53	2	26.20	Up
1193 A	52.49	1		
1193 B	35.23	1	-17.25	Down
1194 A	35.48	2		
1194 B	56.84	2	21.36	Up
1195 A	52.85	2		
1195 B	54.15	2	1.30	No Change
1197 A	52.30	2		
1197 B	51.76	2	-0.54	No Change
1199 A	39.78	1		
1199 B	22.07	1	-17.71	Down
1201 A	8.63	2		
1201 B	8.36	2	-0.26	No Change
1203 A	0.10	1		
1203 B	7.78	1	7.68	No Change

1204 A	0.13	2		
1204 B	7.30	2	7.18	No Change
1205 A	2.02	1		
1205 B	8.54	1	6.52	No Change
1206 A	8.16	1		
1206 B	7.56	1	-0.60	No Change
1209 A	7.56	1		
1209 B	8.19	1	0.63	No Change
1211 A	0.10	2		
1211 B	7.43	2	7.33	No Change
1212 A	5.75	1		
1212 B	7.54	1	1.79	No Change
1214 A	5.13	2		
1214 B	6.77	2	1.64	No Change
1215 A	8.02	1		
1215 B	6.08	1	-1.94	No Change
1216 A	5.83	2		
1216 B	6.14	2	0.31	No Change
1217 A	6.96	2		
1217 B	5.95	2	-1.00	No Change
1218 A	0.00	1		
1218 B	0.13	1	0.13	No Change
1219 A	7.03	1		

1219 B	6.52	1	-0.52	No Change
1220 A	5.61	2		
1220 B	5.44	2	-0.17	No Change
1222 A	40.05	1		
1222 B	35.11	1	-4.94	No Change
1224 A	34.51	2		
1224 B	39.09	2	4.59	No Change
1225 A	5.40	2		
1225 B	6.98	2	1.58	No Change
1226 A	5.69	2		
1226 B	7.25	2	1.56	No Change
Intervention: Folic Acid Supplementation 1, Placebo, 2				

Appendix E.4: RASA4 SMART-MSP Raw Data

Sample Name	Methylation %	Group	Methylation % Change	Outcome

1003 A	0.00	2		
1003 B	<5	2	5.00	No Change
1005 A	27.93	2		
1005 B	<5	2	-22.00	Down
1007 A	19.14	1		
1007 B	<5	1	-15.00	Down
1008 A	<5	1		
1008 B	<5	1	0.00	No Change
1009 A	<5	2		
1009 B	<5	2	0.00	No Change
1012 A	0.00	2		
1012 B	<5	2	5.00	No Change
1018 A	19.28	2		
1018 B	<5	2	-14.00	Down
1022 A	0.00	2		
1022 B	0.00	2	0.00	No Change
1023 A	<5	1		
1023 B	<5	1	0.00	No Change
1024 A	0.00	1		
1024 B	0.00	1	0.00	No Change
1028 A	22.45	2		
1028 B	22.22	2	-0.23	No Change
1030 A	7.46	1		

1030 B	0.00	1	-7.46	No Change
1033 A	10.22	1		
1033 B	<5	1	-5.00	No Change
1037 A	0.00	2		
1037 B	<5	2	5.00	No Change
1041 A	0.00	2		
1041 B	0.00	2	0.00	No Change
1042 A	0.00	1		
1042 B	<5	1	5.00	No Change
1043 A	0.00	1		
1043 B	<5	1	5.00	No Change
1050 A	0.00	2		
1050 B	14.26	2	14.26	Up
1052 A	<5	1		
1052 B	0.00	1	-5.00	No Change
1053 A	4.07	2		
1053 B	46.01	2	41.94	Up
1061 A	0.00	1		
1061 B	60.50	1	60.50	Up
1062 A	32.09	1		
1062 B	<5	1	27.00	Up
1064 A	0.00	2		
1064 B	<5	2	5.00	No Change

1066 A	0.00	2		
1066 B	0.00	2	0.00	No Change
1067 A	0.00	2		
1067 B	0.00	2	0.00	No Change
1069 A	0.00	2		
1069 B	0.00	2	0.00	No Change
1070 A	0.00	2		
1070 B	16.44	2	16.44	Up
1074 A	<5	1		
1074 B	<5	1	0.00	No Change
1076 A	<5	2		
1076 B	<5	2	0.00	No Change
1077 A	<5	1		
1077 B	18.17	1	13.00	Up
FIP1078 A	<5	1		
1078 B	21.99	1	17.00	Up
1082 A	0.00	2		
1082 B	0.00	2	0.00	No Change
1085 A	0.00	2		
1085 B	<5	2	5.00	No Change
1091 A	<5	1		
1091 B	9.61	1	4.00	No Change
1092 A	0.00	2		

1092 B	0.00	2	0.00	No Change
1093 A	0.00	1		
1093 B	5.01	1	5.01	No Change
1095 A	0.00	2		
1095 B	<5	2	5.00	No Change
1099 A	0.00	2		
1099 B	<5	2	5.00	No Change
1100 A	0.00	1		
1100 B	0.00	1	0.00	No Change
1102 A	35.36	2		
1102 B	40.33	2	4.98	No Change
1107 A	0.00	1		
1107 B	0.00	1	0.00	No Change
1108 A	0.00	2		
1108 B	0.00	2	0.00	No Change
1109 A	0.00	1		
1109 B	0.00	1	0.00	No Change
1110 A	0.00	1		
1110 B	0.00	1	0.00	No Change
1112 A	0.00	2		
1112 B	0.00	2	0.00	No Change
1113 A	0.00	1		
1113 B	0.00	1	0.00	No Change

1114 A	0.00	2		
1114 B	0.00	2	0.00	No Change
1115 A	0.00	2		
1115 B	0.00	2	0.00	No Change
1116 A	0.00	1		
1116 B	0.00	1	0.00	No Change
1119 A	0.00	1		
1119 B	0.00	1	0.00	No Change
1120 A	0.00	1		
1120 B	0.00	1	0.00	No Change
1122 A	0.00	1		
1122 B	0.00	1	0.00	No Change
1123 A	0.00	1		
1123 B	0.00	1	0.00	No Change
1132 A	0.00	1		
1132 B	0.00	1	0.00	No Change
1136 A	0.00	1		
1136 B	0.00	1	0.00	No Change
1138 A	0.00	1		
1138 B	0.00	1	0.00	No Change
1139 A	0.00	1		
1139 B	0.00	1	0.00	No Change
1141 A	0.00	1		

1141 B	0.00	1	0.00	No Change
1144 A	0.00	2		
1144 B	0.00	2	0.00	No Change
1145 A	0.00	1		
1145 B	0.00	1	0.00	No Change
1151 A	0.00	1		
1151 B	0.00	1	0.00	No Change
1152 A	0.00	1		
1152 B	0.00	1	0.00	No Change
1153 A	0.00	2		
1153 B	0.00	2	0.00	No Change
1154 A	0.00	1		
1154 B	5.84	1	5.84	No Change
1155 A	0.00	1		
1155 B	0.00	1	0.00	No Change
1156 A	0.00	2		
1156 B	9.34	2	9.34	No Change
1157 A	<5	1		
1157 B	<5	1	0.00	No Change
1160 A	0.00	2		
1160 B	100.00	2	100.00	Up
1161 A	0.00	1		
1161 B	0.00	1	0.00	No Change

1163 A	0.00	2		
1163 B	0.00	2	0.00	No Change
1166 A	<5	2		
1166 B	0.00	2	-5.00	No Change
1168 A	0.00	1		
1168 B	7.54	1	-4.00	No Change
1171 A	<5	1		
1171 B	<5	1	0.00	No Change
1173 A	<5	1		
1173 B	<5	1	0.00	No Change
1174 A	0.00	1		
1174 B	20.73	1	20.73	Up
1177 A	0.00	1		
1177 B	0.00	1	0.00	No Change
1180 A	0.00	1		
1180 B	100.00	1	100.00	Up
1183 A	0.00	1		
1183 B	0.00	1	0.00	No Change
1185 A	0.00	2		
1185 B	0.00	2	0.00	No Change
1187 A	<5	1		
1187 B	0.00	1	-5.00	No Change
1188 A	0.00	2		

1188 B	0.00	2	0.00	No Change
1189 A	0.00	1		
1189 B	0.00	1	0.00	No Change
1190 A	0.00	1		
1190 B	0.00	1	0.00	No Change
1192 A	0.00	2		
1192 B	0.00	2	0.00	No Change
1193 A	0.00	1		
1193 B	0.00	1	0.00	No Change
1194 A	0.00	2		
1194 B	0.00	2	0.00	No Change
1195 A	0.00	2		
1195 B	0.00	2	0.00	No Change
1197 A	0.00	2		
1197 B	0.00	2	0.00	No Change
1199 A	0.00	1		
1199 B	0.00	1	0.00	No Change
1200 A	0.00	2		
1200 B	0.00	2	0.00	No Change
1201 A	0.00	2		
1201 B	0.00	2	0.00	No Change
1203 A	0.00	1		
1203 B	0.00	1	0.00	No Change

1204 A	0.00	2		
1204 B	0.00	2	0.00	No Change
1205 A	0.00	1		
1205 B	0.00	1	0.00	No Change
1206 A	0.00	1		
1206 B	0.00	1	0.00	No Change
1209 A	0.00	1		
1209 B	0.00	1	0.00	No Change
1211 A	0.00	2		
1211 B	0.00	2	0.00	No Change
1212 A	0.00	1		
1212 B	0.00	1	0.00	No Change
1214 A	0.00	2		
1214 B	0.00	2	0.00	No Change
1215 A	0.00	1		
1215 B	<5	1	5.00	No Change
1216 A	0.00	2		
1216 B	<5	2	5.00	No Change
1217 A	0.00	2		
1217 B	0.00	2	0.00	No Change
1222 A	0.00	1		
1222 B	0.00	1	0.00	No Change
1224 A	0.00	2		

1224 B	0.00	2	0.00	No Change
1225 A	0.00	2		
1225 B	0.00	2	0.00	No Change
1226 A	0.00	2		
1226 B	0.00	2	0.00	No Change
Intervention: Folic Acid Supplementation 1, Placebo, 2. >5 scores based on melting peak data				

Appendix F:

Cell culture raw data

Appendix F.1: HEK293 5'azacytidine Cytotoxicity Assay Raw Data

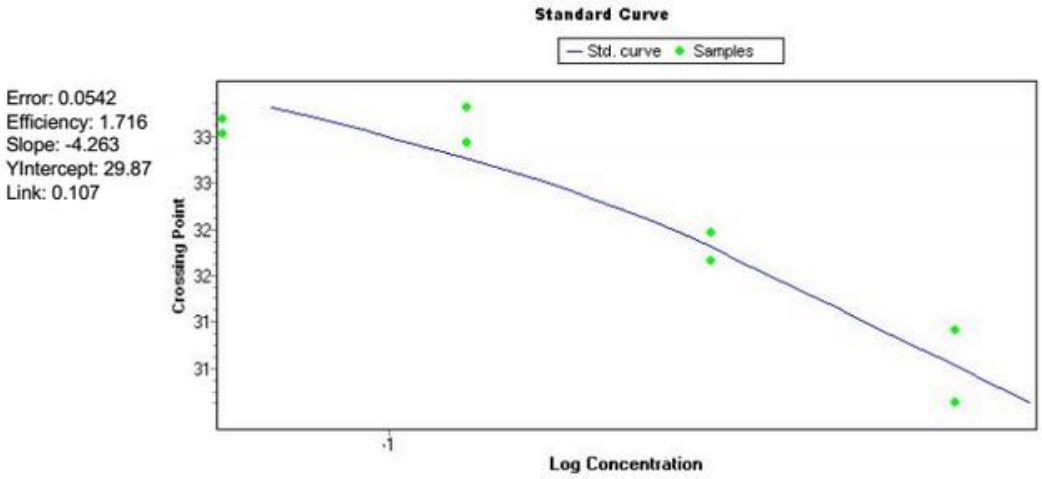
5aC Concentration	Abs1	Abs2	Mean	Standard Deviation
0.5	3.5986	3.8035	3.7010	0.1449
3	2.5505	2.2706	2.4105	0.1979
6	1.5357	1.4892	1.5124	0.0329
12	1.3412	1.3326	1.3369	0.0061

Appendix G:

RT qPCR Raw Data

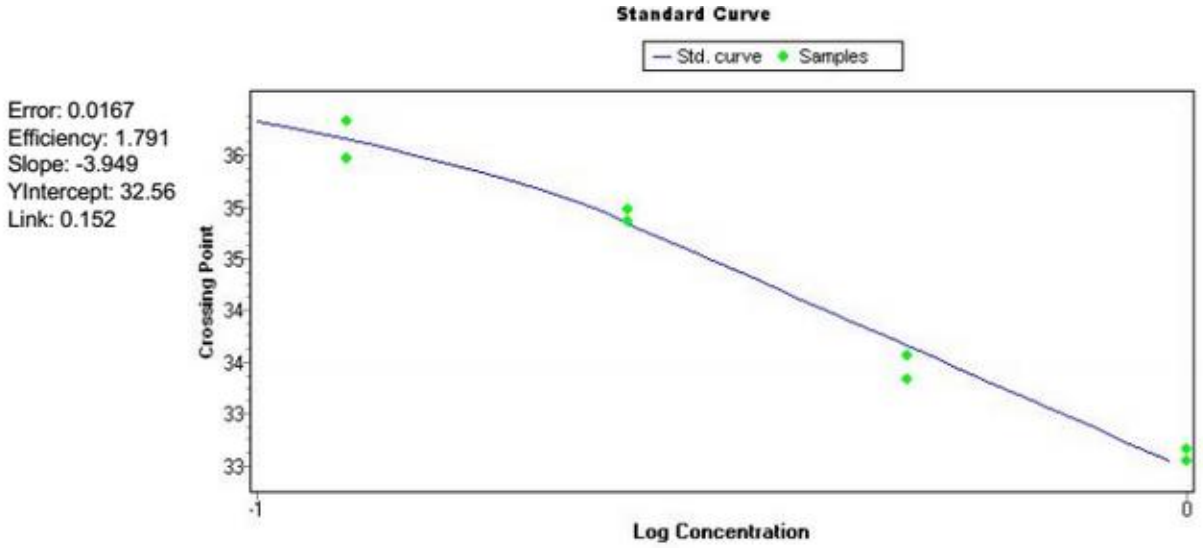
Appendix G.1 GUS RT qPCR data

Gene Name	Cp1	CP2	Average	STDEV
GUS 0uM	25.95	25.99	25.97	0.0283
GUS 5uM	27.85	28.34	28.10	0.3465



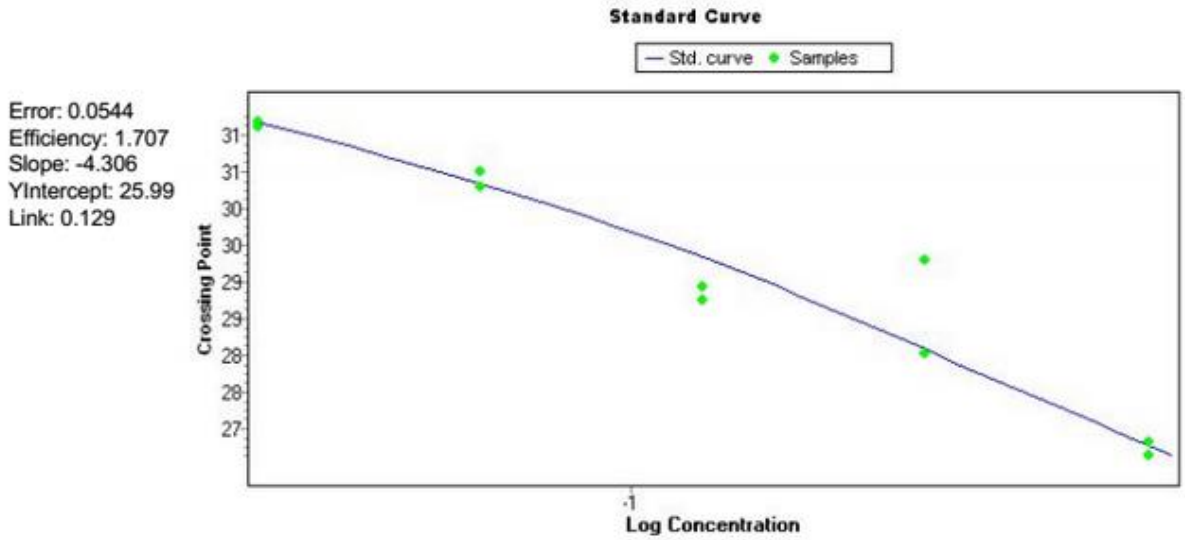
Appendix G.2 IP6K1 rt qPCR assay standard Curve

Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of cDNA from 1/2 to 1/16. IP6K1 has a PCR efficiency of 1.72.



Appendix G.3 RASA4 RTqPCR assay standard Curve

Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of cDNA from 1/2 to 1/8. RASA4 has a PCR efficiency of 1.79.



Appendix G.4 GPS2 RTqPCR assay standard Curve

Standard curve graph generated from the Roche Lightcycler 480 system. Data derived from a doubling dilution of cDNA from 1/2 to 1/8. GPS2 has a PCR efficiency of 1.71.

Appendix G.5: IP6K1 RT qPCR data

	Target Mean Cp	Target Error	Ref. Mean Cp	Ref. Error	Target/Ref	Error
A 0µM 5aC	32.92	0.16	30.8	0.21	40.02	5.583
A 5µM 5ac	32.77	0.16	33.53	0.28	287.8	50.54
B 0µM 5aC	31.65	0.01	28.95	0.05	21.94	0.6
B 5µM 5aC	31.08	0.05	32.54	0.12	357.4	26.08

Appendix G.6: RASA4 RT qPCR data

	Target Mean Cp	Target Error	Ref. Mean Cp	Ref. Error	Target/Ref	Error
A 0µM 5aC	33.14	0.21	30.8	0.21	7.833	2.532
A 5µM 5ac	34.97	0.22	33.53	0.28	17.89	4.591
B 0µM 5aC	31.76	0.12	28.95	0.05	4.851	0.264
B 5µM 5aC	33.33	0.04	32.54	0.12	23.31	2.354

Appendix G.7 GPS2 RT qPCR data

	Target Mean Cp	Target Error	Ref. Mean Cp	Ref. Error	Target/Ref	Error
A 0µM 5aC	24.99	0.04	26.1	0.13	30.17	2.345
A 5µM 5ac	24.12	0.04	28	0.15	187.8	16.37
B 0µM 5aC	25.18	0.12	27.82	0.08	88.73	6.474
B 5µM 5aC	24.91	0.01	31.2	0.1	1086	58.64

Appendix H:

“DNA Methylation: A Timeline of Methods and Applications”

frontiers in
GENETICS

REVIEW ARTICLE
published: 25 October 2011
doi: 10.3389/fgene.2011.00074

DNA methylation: a timeline of methods and applications

Alan Harrison and Anne Parle-McDermott*

Nutritional Genomics Group, School of Biotechnology, Dublin City University, Dublin, Ireland

Edited by:
Jill Ann McKay, Newcastle University, UK

Reviewed by:
Jafar Sharif, RIKEN, Japan
Ravi Goyal, Loma Linda University, USA
Lisa Jane Coneyworth, Nottingham University, UK

****Correspondence:***
Anne Parle-McDermott, Nutritional Genomics Group, School of Biotechnology, Dublin City University, Glasnevin, Dublin 9, Ireland.
e-mail: anne.parle-mcdermott@dcu.ie

DNA methylation is a biochemical process where a DNA base, usually cytosine, is enzymatically methylated at the 5-carbon position. An epigenetic modification associated with gene regulation, DNA methylation is of paramount importance to biological health and disease. Recently, the quest to unravel the Human Epigenome commenced, calling for a modernization of previous DNA methylation profiling techniques. Here, we describe the major developments in the methodologies used over the past three decades to examine the elusive epigenome (or methylome). The earliest techniques were based on the separation of methylated and unmethylated cytosines *via* chromatography. The following years would see molecular techniques being employed to indirectly examine DNA methylation levels at both a genome-wide and locus-specific context, notably immunoprecipitation *via* anti-5′methylcytosine and selective digestion with methylation-sensitive restriction endonucleases. With the advent of sodium bisulfite treatment of DNA, a deamination reaction that converts cytosine to uracil only when unmethylated, the epigenetic modification can now be identified in the same manner as a DNA base-pair change. More recently, these three techniques have been applied to more technically advanced systems such as DNA microarrays and next-generation sequencing platforms, bringing us closer to unveiling a complete human epigenetic profile.

Keywords: DNA, methylation, bisulfite, sequencing, methods

INTRODUCTION

When the Human Genome Project was completed in 2003, 50 years after the discovery of the Double Helix, it was clear that the full picture had yet to be elucidated (Claverie, 2001; Kruglyak and Nickerson, 2001; Lander et al., 2001). The sequence of bases that make up the human genome alone was not enough to account for what makes the human populace so diverse. Five years later, giant leaps in technological advances paved the way for the announcement of the 1000 Genomes Project, aiming to sequence the genomes of 1000 anonymous individuals to visualize the genomic differences that make each person unique (Kaiser, 2008; Durbin et al., 2010). However, mounting evidence from the past few decades is pointing to a new set of variables that contribute to our individuality. The Human Genome Project has already unveiled the genetic hardware needed to create a person, but the search for the biochemical software is still underway. The next major milestone in defining life is The Epigenome: the sum of heritable chemical and chromosomal modifications to genetic material that influences the development of complex organisms. We focus on DNA methylation, the incorporation of a methyl group in mostly a CpG motif, which was first found to influence gene expression in 1975 (Holliday and Pugh, 1975; Riggs, 1975).

At the time, it had been accepted that bacteria were capable of methylating both adenine and cytosine residues while higher organisms possessed mainly methylated cytosines (Wyatt, 1950; Doskocil and Sorm, 1962; Meselson et al., 1972; Smith et al., 1973). The enzyme DNA adenine methylase (Dam) in *E. coli* specifically methylates GATC sequences, and DNA cytosine methylase (Dcm) methylates the duplex sequence CCWGG (W denotes A or T; Casadesús and Low, 2006). As a defence mechanism, bacteria use a plethora of very specific DNA digesting enzymes to ward off invading phages. These enzymes cleave DNA based on a target nucleotide sequence, usually a palindrome motif of several bases, so the enzymes have no way of differentiating between viral and bacterial DNA. A restriction/modification mechanism allows bacterial cells to protect their own DNA from restriction enzymes by introducing a DNA methylation signature into newly synthesized strands (reviewed in Bickle and Krüger, 1993). It was understood that these bacteria carried out methylation in a highly specific manner, but the significance of cytosine methylation in eukaryotes was not fully realized until later. Even though there was no direct evidence of a specific methylating enzyme, Holliday and Pugh (1975) based their early DNA methylation model in eukaryotes on the mechanisms of bacterial methylating enzymes, and the fact that methyl groups are distributed about the genome in a non-random manner. Amongst their concluding remarks, they suggest "it may be significant that the doublet CpG is the most highly methylated," oblivious to how important this statement would be in context of the huge strides in the field to come in the following decades. Independently, a similar paper by Arthur Riggs presented the same hypothesis, this time focusing on the role of DNA methylation in X-inactivation and in mediating DNA binding proteins (Riggs, 1975). Both papers brought considerable attention to the phenomena of DNA methylation, whilst alluding to a new somatically heritable information system that lay within the genetic code.

DNA methylation is now considered to be an important molecular mechanism in a number of biological processes including

genomic imprinting, X-inactivation, tissue specific gene expression, and possibly trans-generational effects (Riggs, 1975; Razin and Ceder, 1991; Li et al., 1993). However, the methods to analyze genome-wide DNA methylation patterns is still evolving. We review the development of DNA methylation methodologies from the late 1970s to the present day (**Figure 1**).

EARLY NON-SPECIFIC METHODS

Early non-specific methods are summarized in **Table 1** and described in more detail below.

HPLC AND TLC METHODS

Ambitious attempts to map the Epigenome started long before the era of the Human Genome. In the fallout of the Holliday and Riggs papers (Holliday and Pugh, 1975; Riggs, 1975), methods for measuring and profiling these epigenetic variations were put forward. The earliest breaches into the epigenetic landscape were based on the separation of methylated and unmethylated deoxynucleosides. The most significant technique at the time was the separation of purines and pyrimidines by Vischer and Chargaff (1948) through paper chromatography. In the context of DNA methylation, Kuo et al. (1980) established an analytical technique to measure 5-methylcytosine (5mC) quantitatively using reversed-phase high performance liquid chromatography (RP-HPLC). This method is based on the quantitative hydrolysis of DNA using DNase I and nuclease P1, followed by treatment with alkaline phosphatase. The individual bases can then be monitored based on their UV absorbances at 254 and 280 nm. The RP-HPLC method was further improved throughout the 1980s (Gomes and Chang, 1983; Patel and Gopinathan, 1987) with incorporation of mass spectrometry with standard HPLC by Annan et al. (1989). Of course, HPLC based methods require specialized machinery, so naturally, alternative separation techniques came into use. Bestor et al. (1984) used two restriction endonucleases, *Msp1* and *Taq1* to discriminate between methylated and unmethylated-CpG residues in their restriction sites, CCGG and TCGA respectively. Digested DNA was 5′ end-labeled with a ^{32}P isotope and subsequently hydrolyzed to deoxyribonucleotide monophosphate followed by separation in two dimensions *via* thin-layer chromatography (TLC). Quantitative measurement of DNA methylation is based on the relative intensity between C and 5mC fractions after separation.

The RP-HPLC and TLC methods described above were only capable of measuring the relative ratio of methylated cytosine residues against unmethylated cytosines. Although this has been useful for many applications, such as comparing the DNA methylation amongst different animal or plant species (Wagner and Capesius, 1981; Gama-Sosa et al., 1983), fully charting the epigenome was far out of reach using these methods. More specific and informative methods are now in practice to detect 5-methylcytosine, but today, HPLC and TLC based methods are now best suited to detecting hydroxymethylcytosine, an epigenetic modification once believed to be only found in bacteriophages, but recently discovered to be abundant in humans and animals (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009).

RADIOLABELING

Instead of trying to separate and observe individually methylated bases at a high resolution, more indirect approaches have been devised. It is possible to enzymatically incorporate tritium labeled methyl groups from *S*-adenosylmethionine to unmethylated cytosines. Assays have been developed using bacterial *SssI* methyltransferase to incorporate radiolabeled *methyl* groups into CpG sites. The level of radioactivity measured is inversely proportional to the level of DNA methylation of a sample (Wu et al., 1993; Duthie et al., 2000).

ANTI-METHYLCYTOSINE

Other alternative methods include the wide range of immunological DNA methylation assays that suddenly appeared after it was

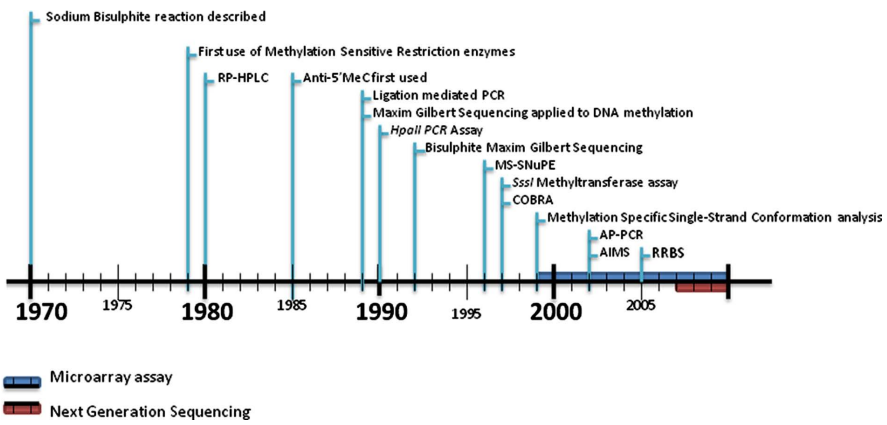


FIGURE 1 | Timeline of DNA methylation analysis. The techniques for DNA methylation analysis have developed from the ability to simply measure the amount of 5-methylcytosine within a particular genome in the early 1980s to a variety of basic comparative methods involving methylation-sensitive restriction enzymes, immunoprecipitation or bisulfite sequencing usually in combination with PCR up to the late 1990s. The introduction of microarray technology and next-generation sequencing saw the adaption of these earlier methods to these newer platforms during the 2000s. More details on microarray/beadchip technologies and next-generation sequencing are described in **Tables 1** and **2**. RP-HPLC, reversed-phase high performance liquid chromatography; 5mC, 5-methylcytosine; MS-SnuPe, methylation-sensitive single nucleotide primer extension; COBRA, combined bisulfite restriction analysis; AP-PCR, arbitrarily primed PCR; AIMS, amplification of inter-methylated sites; RRBS, reduced representation bisulfite sequencing.

Table 1 | Summary of early non-specific methods.

Method	Author
Restriction endonuclease digestion, isotope incorporation, and TLC	Bestor et al. (1984)
Polyclonal leporine antibody, radiolabeled DNA	Adouard et al. (1985)
RP-HPLC	Kuo et al. (1980)
HPLC, mass spectrometry	Annan et al. (1989)
SssI methyltransferase tritium labeling	Wu et al. (1993)
Monoclonal, isothiocyanate labeled fluorescent anti-5mC	Oakeley et al. (1997)

TLC, thin-layer chromatography; RP-HPLC, reverse phase high performance liquid chromatography; Anti-5mC, anti-5-methylcytosine.

first found that methylcytosine was accessible to specific antibodies in 1985 (Adouard et al., 1985). This vital development paved the way for the possibility to chart the DNA methylation landscape on a cell to cell basis. The Adouard paper introduces the quantification of radiolabeled DNA retained by leporine polyclonal antibodies, visualized under electron microscopy. Later, confocal fluorescence microscopy was used to detect global changes in methylation patterns. Using anti-5mC monoclonal antibodies and secondary antibodies labeled with fluorescent isothiocyanate, Oakeley et al. (1997) devised an efficient method to study global changes in DNA methylation during tobacco pollen maturation. The use of anti-5mC has been widely applied since its introduction but most notably in the investigation of DNA methylation changes during embryonic development. The mammalian genome undergoes a mass loss of DNA methylation, followed by remethylation during early embryonic development. Although this was established using methylation-sensitive restriction enzymes (Monk et al., 1987) a more precise profile was obtained with anti-5mC antibody in conjunction with confocal imaging (Santos et al., 2002). They found that the paternal genome undergoes selective demethylation immediately after sperm decondensation, and is complete after 90–120 min. The *de novo* methylases DNMT3a and DNMT3b restore DNA methylation later in development, which is maintained by DNMT1 throughout life (Bestor, 2000).

Out of the methods discussed so far, the immunological approach has seen the most significant improvements and novel applications over the past decade alone, due to advances in microarray technology. These will be discussed in more detail later.

EARLY DIFFERENTIAL GENE METHYLATION ANALYSIS

Early differential gene methylation methods are summarized in **Table 2** and described in more detail below.

METHYLATION-SENSITIVE RESTRICTION ENZYMES

Restriction enzymes cleave DNA through recognition of specific nucleotide motifs. Amongst the variety of different types of restriction enzymes that exist, only some are sensitive to DNA methylation. Given the considerable amount of redundancy amongst the many different palindrome motifs targeted by restriction enzymes, many pairs of restriction enzymes exist that both cut at the same nucleotide sequence but with differing sensitivity to DNA methylation signatures. Isoschizomer pairs like this can be used to discriminate between methylated and unmethylated regions of the genome in a laboratory setting (Bird and Southern, 1978) and initially exemplified in 1979 with *Hpa*II and *Msp*I (Cedar et al., 1979). Both recognize and cut at the same sequence, CCGG, but methylation of the second C in this motif prevents digestion by *Hpa*II. Detection of digested DNA fragments was initially by radiolabeling and two dimensional TLC. Later Southern blotting was employed (Southern, 1975) for visualization followed by the introduction of methylation-sensitive PCR-based methods in 1990 (Singer-Sam et al., 1990). However, efficiency of the restriction enzymes was a likely issue for these techniques.

Table 2 | Summary of early differential gene methylation analysis.

Method	Author
Isoschizomer digestion and isotope incorporation	Cedar et al. (1979)
*Hpa*II PCR	Singer-Sam et al. (1990)
Methylation-specific RLGS	Kawai et al. (1993), Hayashizaki et al. (1993)
AP-PCR	Liang et al. (2002)
AIMS	Frigola et al. (2002)

RLGS, restriction landmark genome scanning; AP-PCR, arbitrarily primed polymerase chain reaction; AIMS, amplification of inter-methylated sites.

DIFFERENTIAL GENOME-WIDE SCANNING

Restriction landmark genomic scanning (RLGS) is a genomic scanning method that takes advantage of the specificity of restriction endonucleases and allows a low resolution comparison of genome-wide differences between individuals (Hatada et al., 1991). Radiolabeled DNA is digested with two restriction enzymes and separated in two dimensions. This produces an autoradiograph profile of thousands of spots spread through the gel, each spot representing a restriction site. This method was adapted for DNA methylation analysis (RLGS-M) by employing methylation-sensitive restriction enzymes (Hayashizaki et al., 1993; Kawai et al., 1993) to differentiate methylation differences between individuals. Later, simpler and less expensive genome-wide screening strategies came into use. Using a single primer and two low-stringency annealing steps, Liang et al. (2002) found that methylation profiles could be obtained by digesting DNA with methylation-specific endonucleases followed by a PCR reaction with random primers. This process is known as arbitrarily primed PCR (AP-PCR), and is based on a method developed by Welsh and McClelland (1990) initially used to identify bacterial species. AP-PCR was adapted in order to scour tumor genomes for new differential methylation sites (Gonzalgo et al., 1997; Liang et al., 2002). Amplification of inter-methylated sites (AIMS) is a similar but more effective PCR-based approach. Methylation-sensitive isoschizomers are employed that cleave DNA leaving a blunt end or an overhang. These properties are exploited by the addition of linkers that only ligate to the methylated sites with subsequent PCR amplification. Fingerprints composed of multiple anonymous bands represent methylated regions of the genome are generated and can be excised out and characterized individually (Frigola et al., 2002).

It should be noted that the methods discussed so far are limited in the context of other genetic techniques at the time. This is

because *in vitro* amplification of methylated DNA strands *via* PCR causes the target strand to lose its methylation status. The methods so far have aimed to detect 5-methylcytosine as it manifests naturally. On a genome-wide or gene-specific scale, these approaches are limited. In order to advance to the stage of possibly sequencing the epigenome, a new approach was needed.

THE SODIUM BISULFITE ERA

In 1970, a chemical interaction between sodium bisulfite and pyrimidines was described that would have a colossal impact on how DNA methylation is studied (Hayatsu et al., 1970). It was found that uracil, thymidine, and deoxycytidine were subjected to sulfonation at position six of their pyrimidine rings. Ten years later, this model was extended to 5-methylcytosine although the reaction takes place at a slower rate than cytosine (Wang et al., 1980). Frommer et al. (1992) described in a classic paper that the differing reaction rates of 5mC to C could be exploited to analyze DNA methylation patterns in genomic DNA. Treating DNA with sodium bisulfite, they proposed, will deaminate cytosine residues into uracil at a much faster rate than 5mC. This phenomenon made it possible to change a chemical modification of DNA to an easily detected genetic element. At the time, Maxim and Gilbert sequencing was used to pinpoint the changes, but the methods put forward by Frommer and colleagues would be revised and refined as technological advances in the subsequent years would pave the way for large scale, next-generation sequencing.

The Frommer et al. (1992) paper marked somewhat of a revolution in the field. Now the elusive biochemical software could be converted to more tangible genetic hardware. Although it was initially described how bisulfite modification could be used to augment sequencing-based methods, the concept itself would be used to formulate entirely new methods to probe the genome for DNA methylation in the following years. These methods are based on the treatment of DNA with bisulfite such that unmethylated cytosines are converted to uracil and methylated cytosines remain as cytosines. The approaches to detect these conversions are various and are summarized in **Table 3** and described in more detail below.

Table 3 | Summary of methods using sodium bisulfite treatment.

Method	Author
Ligation-mediated PCR	Pfeifer et al. (1989)
Bisulfite sequencing	Frommer et al. (1992)
MS-PCR	Herman et al. (1996)
MS-SNuPE	Gonzalgo and Jones (1997)
MS-SSCA	Bianco et al. (1999)
MS-HRM	Wojdacz and Dobrovic (2007)
Bisulfite treatment to create new restriction sites	Sadri and Hornsby (1996)

MS-PCR, methylation-specific polymerase chain reaction; MS-SNuPE, methylation-specific single nucleotide primer extension; MS-SSCA, methylation-specific single-strand conformation analysis; MS-HRM, methylation-specific high resolution melting.

GENE-SPECIFIC APPROACHES

Methylation-specific PCR (MS-PCR) was one of the first innovative methods to incorporate bisulfite conversion outside the context of sequencing (Herman et al., 1996). Primers were designed to discriminate between methylated and unmethylated regions of DNA after bisulfite treatment, so primer sites that were originally methylated would undergo amplification only. The nature of this rapid assay eliminated the frequent false positives associated with previous PCR-based endonuclease methods; however PCR bias was an issue. Technical advances in genomics and molecular biology in more recent years have allowed MS-PCR take on a new form.

Many of the new techniques introduced during the Sodium Bisulfite Era followed a similar strategy; using well established genetic techniques to detect DNA methylation since the elusive epigenetic modification could be converted into the more tangible nucleotide variant. Methylation-sensitive single nucleotide primer extension (MS-SNuPE) is based on a conventional genotyping technique, single nucleotide primer extension (Kuppuswamy et al., 1991). MS-SNuPE (Gonzalgo and Jones, 1997) uses a PCR step after bisulfite conversion to amplify a desired fragment. Once the product is isolated, primers specific for the amplified fragments are used in another PCR stage, this time incorporating ^{32}P dNTPs which can be used to quantify the nucleotides that have been converted during bisulfite treatment, therefore quantifying the level of DNA methylation in the initial genomic DNA.

Based on another well-known PCR method for resolving single-base restriction fragment length polymorphisms (Poduslo et al., 1991), methylation-sensitive single-strand conformation analysis (MS-SSCA) is a method to screen and analyze DNA methylation in a gene-specific manner (Bianco et al., 1999). Genomic DNA is bisulfite treated and the gene of interest is amplified with PCR, and then cut with frequently cutting restriction enzymes. The digestion patterns of samples are compared to a methylation standard and variations in pattern imply changes in DNA methylation. Methylation differences are characterized using a gel stabbing technique and sequencing (Wilton et al., 1997). This method has been expanded by Suzuki et al. (2000) to include high performance capillary electrophoresis (HPCE).

High resolution melting (HRM) was originally used to genotype Single Nucleotide Polymorphisms (Wittwer et al., 2003) but was adopted to detect DNA methylation changes in bisulfite treated DNA (Wojdacz and Dobrovic, 2007). Single base differences can be detected by their distinct melting profiles utilizing specific fluorescent dyes. The difference between 5-methylcytosine and cytosine, manifests as a single base change after DNA is treated with sodium bisulfite. With careful primer design to eliminate PCR bias, it is possible to estimate the methylation levels of a test sample by comparison of its melting curve with that of a series of controls of known methylated and unmethylated percentages (Wojdacz et al., 2008).

The "bisulfite revolution" was not limited to the importation of early genetic techniques to the field of DNA methylation; methylation analysis mentioned earlier in this review would also receive a renewal. The endonuclease-based protocols used up until the mid 1990s were limited to the detection of a negative result: the absence of a band indicates a methylated site. This was first improved by

Sadri and Hornsby (1996), where DNA was first treated with bisulfite according to a revised version of the 1992 bisulfite reaction (Feil et al., 1994), then exposed to two rounds of endonuclease digestion including a newly created restriction site following bisulfite treatment. This innovation was expanded by Xiong and Liard (1997) to determine the methylation status of individual loci. Combined bisulfite restriction analysis (COBRA), is based on the creation of new methylation dependant restriction sites, or the retention of pre-existing ones, by bisulfite conversion followed by PCR. With phosphorimaging, the relative ratio of digested products can be determined. Although it is a powerful technique, COBRA is limited to the restriction sites of the enzymes used. Laird et al. (2004) devised a technique known as "hairpin-bisulfite PCR" to investigate DNA methylation symmetry at a specific locus. With bisulfite treatment, the required denaturation steps make it difficult to analyze the methylation pattern of two complementary DNA strands from one molecule. By ligating a hairpin linker to restriction-enzyme cleaved DNA, the team were able to establish a covalent bond between complementary strands of the DNA molecule, which would allow a PCR product to span the linker and cover both strands.

REGIONAL METHYLATION LEVELS

Another method already mentioned here that has received a sodium bisulfite facelift is the *SssI* methyltransferase assay. In its new incarnation, the enzymatic regional methylation assay (ERMA), genomic DNA is treated with sodium bisulfite prior to amplification of a particular region of interest with non-discriminating primers containing flanking GATC sites. These tetranucleotide sequences are required to standardize DNA quantity in this assay, as they are *dam* sites that accept methyl groups from *dam* methyltransferase. To quantify DNA methylation, *E. coli* cytosine methyltransferase *SssI* was used to specifically methylate the cytosine in all of the CpG dinucleotides that remained after sodium bisulfite treatment, using ^{3}H-labeled *S*-adenosyl-L-methionine (SAM) as a methyl donor. For the aforementioned standardization step, ^{14}C-labeled SAM was incubated along with *dam* methyltransferase so the total number of amplicons could be visualized (Galm et al., 2002).

In 2001, the working draft of the Human Genome was published in special issues of *Nature* and *Science* (Pennisi, 2001). Later that year, Human chromosome 20 was fully sequenced, the third chromosome to be completed in the Human Genome Project. This year was also an important year for DNA methylation and epigenetics too, because it was here that a new phrase entered the vocabulary of the scientific community: The Methylome (Feinberg, 2001).

The post-genome era

By the beginning of the twenty-first century, a great deal of the epigenetic landscape had been explored. While the role and mechanism of gene regulation *via* DNA methylation was well understood, the gene-specific methods described above helped bring these ideas to the context of complex diseases states, especially tumorigenesis (Jones and Laird, 1999). However, very little was known about the genome-wide distribution of 5-methylcytosine until robust array precipitation methods were devised.

COMPARATIVE METHYLATION PROFILING USING MICROARRAY TECHNOLOGY

Throughout the 1990s, the development of DNA microarray technology was responsible for a revolution in functional genomics, paving the way for high-throughput analysis of single nucleotide polymorphisms and other genomic variants (Southern et al., 1999). With the help of these novel tools, the three traditional lines of attack on the DNA methylation landscape; immunoprecipitation, endonuclease digestion, and sodium bisulfite treatment, would each receive a post-genome era transformation (**Figure 2**). These three DNA methylation differentiation and isolation methods have been the principal approaches used to compare the DNA methylation patterns between samples over the last decade. In the microarray assays discussed here, the underlying principle is the same in each: methylated and unmethylated fragments of the genome are separated and analyzed. Hybridization to a microarray of known probes allows for quantification and identification of areas of the genome that are methylated or unmethylated. All of the microarray-based techniques discussed here are listed in **Table 4** and are based on one of the three approaches described below and in **Figure 2**.

ENDONUCLEASE DIGESTION

Differential methylation hybridization (DMH) was the first array-based method for genome-wide screening of hypermethylated-CpG islands in tumor cells (Huang et al., 1999). This early array was only able to asses about 300 CpG islands at a time, and suffered from major sequence bias. Genomic DNA was first sheared with a methylation insensitive restriction enzyme, *Mse*I.

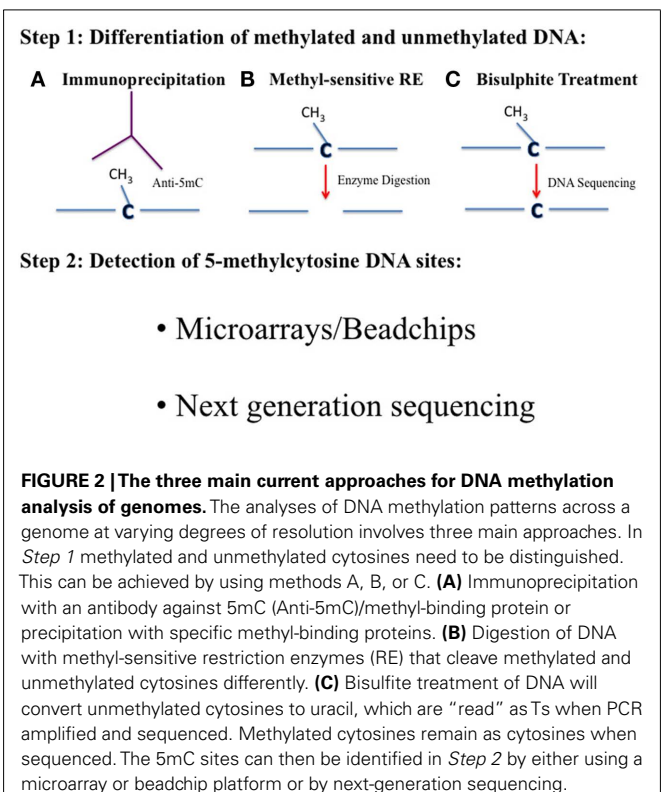


FIGURE 2 | The three main current approaches for DNA methylation analysis of genomes. The analyses of DNA methylation patterns across a genome at varying degrees of resolution involves three main approaches. In *Step 1* methylated and unmethylated cytosines need to be distinguished. This can be achieved by using methods A, B, or C. **(A)** Immunoprecipitation with an antibody against 5mC (Anti-5mC)/methyl-binding protein or precipitation with specific methyl-binding proteins. **(B)** Digestion of DNA with methyl-sensitive restriction enzymes (RE) that cleave methylated and unmethylated cytosines differently. **(C)** Bisulfite treatment of DNA will convert unmethylated cytosines to uracil, which are "read" as Ts when PCR amplified and sequenced. Methylated cytosines remain as cytosines when sequenced. The 5mC sites can then be identified in *Step 2* by either using a microarray or beadchip platform or by next-generation sequencing.

Table 4 | Applications of microarray technology to differential DNA methylation analyses.

Method	Array type	Citation
Methyl-sensitive restriction enzymes	CGI[1] library	Huang et al. (1999)
Anti-methylcytosine immunoprecipitation[3]	SMRT[2]	Weber et al. (2005)
Anti-methylcytosine immunoprecipitation	Promoter array	Keshet et al. (2006)
Methyl-binding protein precipitation[4]	CGI library	Gebhard et al. (2006)
Methyl-binding protein precipitation	CGI library	Rauch et al. (2006)
Sodium bisulfite treatment	Oligonucleotide	Gitan et al. (2002)
Sodium bisulfite treatment	Oligonucleotide	Adorjan et al. (2002)
Sodium bisulfite treatment	Oligonucleotide[5]	Reinders et al. (2008)
Sodium bisulfite treatment	Illumina beadchip	Sandoval et al. (2011)

[1] *CGI, CpG island microarray.*
[2] *SMRT, submegabase resolution tiling array.*
[3] *MeDIP, methylated DNA immunoprecipitation.*
[4] *MeCIP, methyl-CpG immunoprecipitation.*
[5] *Whole genome oligonucleotide array.*

Having a restriction recognition site at TTAA, *Mse*I was unlikely to interfere with any CpG islands. After the ligation of linkers to the end of each DNA fragment, half of the pool was treated with methylation-sensitive *Bst*UI. As a result, the methylated fragments, and those only treated with *Mse*I remained intact, and only these could be amplified *via* PCR, with primers specific to the linkers. These amplicons were differentially labeled and co-hybridized to a CpG island array. As for any array-based method, the analysis is limited to the number of genomic elements represented on the array. The array, used to determine the methylation status of CpG islands in breast cancer cells, was constructed from a physical library of CpG islands generated from a novel column separation strategy (Cross et al., 1994). Shortly after it was introduced, DMH was used to detect specific methylation profiles in breast and ovarian cancer cells (Yan et al., 2000; Ahluwalia et al., 2001). Following on from this, the same group improved on this method further (Yan et al., 2001; Chen et al., 2003).

In 2005, the Promoter-associated methylated DNA amplification DNA chip was introduced. Contrary to the use of a second wave of restriction enzymes, the restriction endonuclease *McrBC* has been used to fractionate methylated regions of DNA. In a protocol pioneered by Nouzova et al. (2004), DNA is treated with *Mse*I, and the fragments are ligated to linkers, in concurrence to the previous methods. However, the fragments are then divided into two pools: one to be treated with *McrBC*, while the other is not. Unlike the other restriction enzymes discussed so far, *McrBC only* cuts at methylated sequences. After PCR, both pools are differentially labeled with Cy3 and Cy5 fluorescent dyes and co-hybridized to a CpG island array. In the Methylscope platform (Ordway et al., 2006), DNA fragments are prepared in a similar way, but the DNA is randomly sheared in the first step.

Comprehensive high-throughput arrays for relative methylation (CHARM) is another platform for array-based methylation analysis (Irizarry et al., 2008). The workflow is based on some of the methods already mentioned in this section, and works to eliminate the disadvantages of each. CHARM was conceived while Irizarry and colleagues were comparing three already established methods for analyzing DNA methylation: methylated DNA immunoprecipitation (MeDIP), HELP, and fractionation by McrBC. The HELP assay (*Hpa*II tiny fragment enrichment by ligation-mediated PCR) is based again on the use of two sets of restriction enzymes, but the fragments are amplified *via* ligation-mediated PCR and hybridized to a custom microarray along with separate fluorochromes (Khulan et al., 2006). In the 2008 paper, Irizarry points out significant flaws with each of the array-based methods in use. MeDIP (also discussed below), was shown to have a significant bias to CpG islands, HELP had incomplete genomic coverage and McrBC fractionation displayed location imprecision. After the rigorous comparison of these methods, the second half of the paper discusses how a new platform of original array design strategies and statistical procedures involving genome-weighted averages from larger genomic areas was capable of countering these limitations (Irizarry et al., 2008).

IMMUNOPRECIPITATION

Differentiation between methylated DNA and non-methylated DNA using anti-methyl antibodies has been discussed already. However, with the requirement to enrich methylated DNA prior to microarray hybridization, immunological separation techniques became relevant again.

In 2005, MeDIP was used to immunocapture methylated cytosines with an antibody specific for methylated cytosines for array hybridization (Weber et al., 2005). Prior to immunoprecipitation, genomic DNA was randomly fragmented *via* sonication or enzyme restriction. Immunocaptured DNA and control genomic DNA were both labeled with Cy5 and Cy3 fluorescent dyes, producing a ratio of green fluorescence to red fluorescence which would be indicative of the relative levels of hypermethylation or hypomethylation. In the 2005 paper, Weber and colleagues used a submegabase resolution tiling (SMRT) array consisting of 32,433 overlapping BAC clones spanning the entire genome (Ishkanian et al., 2004; Weber et al., 2005). Independently, Keshet et al. (2006) devised a similar array-based approach: methyl-DNA immunoprecipitation (MDIP). They found that tumor specific methylated genes are found in clusters on chromosomes, and shared many structural and functional features. This reinforced the hypothesis that tumorigenesis arises as a result of *de novo* mechanisms. One of the major drawbacks of MDIP and MeDIP is their inability to pinpoint DNA methylation changes at a single base-pair resolution (Beck and Rakyan, 2008). However, some argue that since neighboring CpG islands spanning up to 1000 bp are co-methylated in healthy cells, there is no need for methylation analyses with single base-pair resolutions (Eckhardt et al., 2006). In 2006 MDIP was responsible for producing the first complete Methylome: *Arabidopsis thaliana* (Zhang et al., 2006). Although much smaller than the mammalian genome, the map of the plant's methylome represents an important milestone in epigenetics, while the data produced was of interest in itself. It was found that one third

of expressed genes contained DNA methylation in their transcribed regions, and these regions were still highly expressed and constitutionally active.

Methyl-CpG immunoprecipitation (MeCIP) is another immunoprecipitation assay similar to MeDIP in terms of the techniques used and its applications, but a recombinant protein complex with similar properties to an anti-methylcytosine antibody is used (Gebhard et al., 2006). Epigenetic gene silencing *via* DNA methylation is caused by steric hindrance, resulting from methylated DNA recruiting methyl-binding domain proteins (MBDs; Thu et al., 2010). In the 2006 paper, Gebhard and colleagues introduce a recombinant protein made up of MBD2 combined with the Fc tail of a human Ig1 with very high affinity to single-stranded methylated DNA, stronger than that of the MeDIP and MDIP methods. Also, it is possible to separate DNA fragments into fractions of increasing methylation density by eluting with a salt gradient (Schilling and Rehli, 2007). This approach allows for the quantification of tissue specific methylation differences for a wide range of DNA methylation densities.

Another protein complex, MBD3LI bound to MBD2, has also been shown to have a high affinity to methylated DNA (Rauch and Pfeifer, 2005). Methylated-CpG island recovery assay (MIRA) separates fragmented DNA by incubation with a matrix containing glutathione-*S*-transferase-MBD2b in the presence of methyl-CpG-binding domain protein 3-like-1, which increases the affinity of MBD2b when paired. CpG island methylation can be detected using PCR or array-based methods (Rauch et al., 2006, 2009).

BISULFITE TREATMENT

In 2002, the principles of DHM were expanded to the use of novel methylation-specific oligonucleotide arrays (Adorjan et al., 2002; Gitan et al., 2002). This time, DNA was prepared for hybridization by bisulfite modification and PCR amplification to convert unmethylated cytosines to thymidine, allowing the epigenetic modification to be detected *via* traditional hybridization methods. This approach has the potential to detect methylated-CpG islands at a single-base resolution, but the global conversion of cytosines to thymidines results in a reduction in sequence complexity, making it difficult to design enough unique probes to scale up to a genome-wide level (Beck and Rakyan, 2008). Although it is possible to design probes taken from amplified bisulfite treated DNA, novel approaches have been devised. For small, methylation-rich genomes, a method called bisulfite methylation profiling (BiMP) can be employed (Reinders et al., 2008). The entire genome of *A. thaliana* was amplified using a technique utilizing random tetranucleotides primers reducing the amplification bias usually associated with bisulfite treated DNA. The BiMP data from the paper was compared to the MDIP results cited earlier (Zhang et al., 2006). Data from both studies were in concordance, although the former exhibited profiles of considerably higher resolution than the latter (Reinders et al., 2008). As a result, BiMP is more likely to pick up specific, localized changes in DNA methylation patterns that could prove elusive to detection *via* MDIP.

The Illumina Beadchip technology, while technically different from the types of arrays discussed above, do fall under the microarray category. Illumina Infinium has been applied to the DNA methylation analysis. Fryer et al. (2011) examined DNA methylation patterns at 27,578 CpG sites using the Infinium HumanMethylation27K in cord blood samples and correlated to homocysteine levels and birth weight. More recently, Illumina have launched the Infinium HumanMethylation450Karray which allows the analysis of >450,000 DNA methylation sites (Sandoval et al., 2011) with up to 12 samples at a time. This is by far the most high-throughput comprehensive method available for whole genome DNA methylation analysis outside of the next-generation sequencing methods described below.

The use of microarrays for DNA methylation analysis proves to be a versatile strategy for probing the methylome. Before hybridization, methylated DNA can be purified by a number of different strategies, each with their own unique merits. The strengths and weaknesses of most of these methods have been systematically evaluated in Laird (2010). DNA methylation microarrays provide cheap and accessible genome-wide insights to the DNA methylation status of a sample, or even a large number of samples. However, as with the Human Genome Project, there needs to be a trend toward a gold standard: a perfect assay. Although none exists at time of writing, only sequencing-based assays have the potential to provide such a detailed look at the enigmatic methylome, i.e., at single-base resolution.

SEQUENCING-BASED APPROACHES

SANGER SEQUENCING

Although the bisulfite reaction itself has been adapted and applied to the conventional genetic techniques aforementioned, the technology has taken a novel route while genome sequencing platforms have improved as the Human Genome Project progressed over the subsequent years. At the moment, it is possible to directly sequence the human genome with sophisticated technology; technology that is starting to be applied to the field of DNA methylation. First, it is worth revisiting how sequencing-based DNA methylation analyses have evolved over the past two decades.

The original bisulfite sequencing protocol from Frommer et al. (1992) suffers from several difficulties. For example, relatively large quantities of genomic DNA are needed for a full profile, limiting its proficiencies in a genome-wide perspective. In 1994, the same lab (Clark et al., 1994), integrated a PCR amplification step to increase the assay's sensitivity by 10^4 fold. The old protocol also required DNA to be denatured in order to expose the individual bases to bisulfite treatment. Some workarounds have been devised to counter this, but according to a review by Oakeley (1999), the best approach at the time was to denature DNA in solution with NaOH, then mix with molten agarose. Cooling the agarose locks the DNA in the denatured conformation, allowing subsequent reactions to be performed on the agarose block (Olek et al., 1996).

In the pilot study of the Human Epigenome Project, Rakyan et al. (2004) aimed to profile the DNA methylation patterns of the human major histocompatibility complex (MHC) located on chromosome six. This region of the genome was selected as it is associated with more diseases than any other region of the genome, and also it's the most polymorphic area of the genome, so complete sequencing and annotation from the Human Genome was readily available for the study. This sequencing method was innovative as it did not require a sub-cloning step, but utilized a novel high-throughput method of direct sequencing of PCR products.

An algorithm described by Lewin et al. (2004) allows for quantitative analysis of DNA methylation from four-dye electropherogram data obtained from direct sequencing. Although such data was previously used in the human genome project and for analyzing single-base SNPs (Qiu et al., 2003), earlier applications of bisulfite treated PCR were impeded by unique technical difficulties. The new software, called epigenetic sequencing methylation analysis software (ESME), corrects for incomplete bisulfite conversion, performs quality control tests on data, and maps methylation positions to the reference sequence. This approach was also used in a related paper reporting the DNA methylation profiles of human chromosomes 6, 20, and 22 (Eckhardt et al., 2006).

Reduced representation bisulfite sequencing (RRBS) is a random sequencing-based method for analyzing and comparing DNA methylation patterns on a genome scale (Meissner et al., 2005). Size selected *Bgl*II fragments of a whole genome were fixed with ligation linkers and denatured. Bisulfite treatment of these fragments yielded single-stranded DNA, as complementarity between both strands was lost when methylated cytosines were converted to uracil residues. Converted fragments were amplified *via* PCR and cloned into plasmid vectors for sequencing. Comparison of the bisulfite treated DNA sequence to a reference sequence allows the operator to pinpoint which cytosines have been methylated, as those are the only ones to remain cytosines at this stage. Thymidines that align to cytosines during this comparison stage represent cytosines that were once unmethylated. RRBS has the obvious advantage over PCR-based bisulfite sequencing methods in generating a reproducible library of a small, defined area of a genome. This makes RRBS suitable for comparative methylation studies across different tissue or cell types.

NEXT-GENERATION SEQUENCING

Up until 2005, most whole genome DNA sequencing strategies were based on the cloning of fragments into bacterial vectors, followed by amplification and Sanger sequencing *via* chain terminating fluorescent signaling, visualized with capillary electrophoresis (Prober et al., 1987). In recent years, however, a new parallel sequencing method was developed that did not require a sub-cloning step. It involves a previously established method for genotyping known as pyrosequencing, which incidentally has also been applied to the analysis of gene-specific/local DNA methylation patterns (Tost and Gut, 2007). The emulsion based PCR method, described by Margulies et al. (2005), utilizes a pyrosequencing protocol optimized for picoliter-scale volumes in the high density picoliter "reactors" formed by the emulsion droplets. In 2007, this massively parallel sequencing system (commercialized as Roche 454 FLX) was employed for bisulfite sequencing (Taylor et al., 2007). The pilot study showed robustness and superiority of this approach by analyzing methylation in 25 gene-related CpG rich regions from over 40 primary cell lines. During the process, specific four-nucleotide tags were added to the 5′end of each primer, so each amplicon could be individually indexed, pooled, and manipulated (Taylor et al., 2007).

Further advances in next-generation sequencing including the Illumina/Solexa Genome Analyzer and the Applied Biosystems SOLiD™ System (reviewed in Mardis, 2008) has meant that going forward, most genome-wide DNA methylation protocols will feature some form of next-generation sequencing. The current gold standard is to carry out whole genome bisulfite sequencing of target samples where a reference genome is available. However, the costs for such an approach are still not trivial and adaptations of methods to produce a representation of genome-wide DNA methylation have been developed. **Table 5** describes some of the current options for DNA methylation analysis in combination with next-generation sequencing and are described below. It is worth noting that similarly to microarray analysis, for next-generation sequencing, the three principle approaches still employ sodium bisulfite treatment, immunoprecipitation, and the utilization of methyl-sensitive restriction enzymes (**Figure 2**).

Table 5 | Approaches for DNA methylation analyses by next-generation sequencing.

Method	Genome coverage	Citation
Bisulfite sequencing	Whole genome	Cokus et al. (2008)
MeDIP-seq[1] anti-5mC	Enriched Methylated DNA	Maunakea et al. (2010)
MBDiGS[2]	Enriched Methylated DNA	Serre et al. (2009)
MRE-seq[3]	Size selected fraction	Maunakea et al. (2010)
MMSDK[4]	Representative genome tags	Li et al. (2009)

[1] *Sequencing of immunoprecipitated anti-5mC DNA.*
[2] *Methyl-binding protein precipitated sequencing.*
[3] *Methyl-sensitive restriction enzyme sequencing.*
[4] *Modified methylation-specific digital karyotyping.*

WHOLE GENOME SEQUENCING-BISULFITE TREATMENT

An entire DNA methylome can be assessed at a single nucleotide resolution with sodium bisulfite treatment followed by whole genome sequencing. This approach has been taken to generate a DNA methylation map of *A. thaliana* (Cokus et al., 2008). Unlike previous genome-wide approaches, this allowed for the sensitive measurement of cytosine-methylation across the genome with sequence specific contexts. When compared to array-based methods, the authors reported the discovery of new methylation sites in previously inaccessible areas of the genome. A whole genome approach was also recently applied to mammalian cells. The first human DNA Methylome in embryonic and fetal cells at single-base resolution was recently published (Lister et al., 2009), identifying a significant proportion of non-CG methylation. Additional single-base resolution human methylomes continue to be published (Maunakea et al., 2010) highlighting the importance of intragenic DNA methylation in the regulation of gene expression. Thus, the elusive Human DNA Methylome is more complex than previously thought.

The whole genome approach is the most desirable with unlimited resources, but realistically for a lot of laboratories this is not an approach that can be taken for the analysis of numerous samples. A more cost effective approach is to reduce the complexity of the genome in order to reduce the amount of sequencing required per sample. The methods described below are some examples of how this can be done.

METHYLATED DNA IMMUNOPRECIPITATION SEQUENCING

The methylated DNA immunoprecipitation sequencing (MeDIP-seq) approach incorporates the anti-methylcytosine antibody described earlier. Briefly, methylated DNA is immunoprecipitated using the antibody against 5-methylcytosine and sequenced (Maunakea et al., 2010; **Table 5**). The portion of DNA that is immunoprecipitated represents the methylated portion of DNA and is identified by comparison to the reference genome.

METHYL-BINDING DOMAIN ISOLATED GENOME SEQUENCING

Methyl-binding domain isolated genome sequencing (MBDiGs) uses recombinant MBD and MBD2 proteins to enrich methyl-rich DNA fragments from a pool of sonicated genomic DNA (Serre et al., 2009). According to the review by Hirst and Mara (2010), MBDiGS is preferable over MeDIP-seq because a gradient in salt concentrations can be used to elute DNA fragments at different rates depending on their methylation status.

METHYL-SENSITIVE RESTRICTION ENZYME SEQUENCING

Methyl-sensitive restriction enzyme sequencing (MRE-seq), as its name suggests, involves methylation-sensitive restriction enzymes (Maunakea et al., 2010). Genomic DNA samples are digested with the restriction enzymes and the subsequent DNA fragments are size selected and sequenced. Differential DNA methylation may be identified by comparison of the fragments sequenced between samples and site specific information is identified by comparison to a reference genome. This method analyses a different portion of the genome compared to MeDIP-seq and therefore, they can be viewed as complimentary approaches.

MODIFIED METHYLATION-SPECIFIC DIGITAL KARYOTYPING

Modified methylation-specific digital karyotyping (MMSDK) or MSDK-seq (Li et al., 2009) is similar to MRE-seq in that a methyl-sensitive restriction enzyme is employed but includes additional steps that reduce the amount of sequencing required. Rather than sequencing sections of the genome, specific regions of the genome can be identified from their short sequence tags. This significantly reduces the amount of sequencing and in turn reduces the costs of this approach.

COMPARISON OF CONTEMPORARY DNA METHYLATION METHODS

In this review, we aim to detail the development and evolution of these analytical methods over time with respect to advancements made in genetics, nucleotide biochemistry, and DNA sequencing technology. As a result, many of the methods discussed are obsolete today and have been replaced by more recent technologies. However, some of the techniques described in the latter part of this review have subtle strengths and weaknesses, and careful judgment should be employed in adopting any of these methods in a research laboratory. There are many recent reviews that compare most recent methods of DNA methylation analysis as mentioned below.

Laird (2010) list the features and source of bias for various sequencing and microarray-based techniques including CpG ambiguity, fragment size bias, cross-hybridization bias. All of the methods that involve sodium bisulfite treatment, they argue, are subject to incomplete bisulfite conversion bias. Thu et al. (2010), also compare the strengths and weaknesses of each method, with a special focus on techniques based on immunoprecipitation. In more detail, a paper by Harris et al. (2010) quantitatively compares the sequencing-based methods MethylC-seq, RRBS, MeDIP-seq and MBD-seq (2010). Due to the nature of their processes, the two bisulfite-based methods yield data with single base-pair resolution, augmented with the capacity to quantify methylation levels. At a reduced coverage, the enrichment methods both have a lower cost-per-CpG in a genome-wide context, but not allowing precise quantification of methylation levels on a genome-wide scale. It appears that none of the currently available methods are without their flaws but bisulfite treated whole genome sequencing offers complete genome coverage at single-base resolution and is currently the method of choice for genome-wide DNA methylation analysis where costs are not prohibitive.

FUTURE OF METHYLOME ANALYSIS

The Next-generation sequencing approaches for DNA methylation analysis will dominate for the moment. However, the methods discussed here cannot detect non-cytosine related methylation reactions, i.e., *N*6-methyladenine nor 5-hydroxymethylcytosine and therefore, more sophisticated methods are required than currently on offer. Newer sequencing technologies such as single-molecule real-time (SMRT) sequencing (Flusberg et al., 2010) can directly detect all known DNA methylation reactions without the need for bisulfite treatment and is likely to take over from next-generation sequencing in the very near future.

The analysis of 5-hydroxymethylcytosine (5hmC) *via* HPLC based methods has been discussed briefly. Recently, two novel approaches have been described to discern the genomic distribution of 5hmC (Pastor et al., 2011). The first, called GLIB involves the glucosylation, periodate oxidation, and biotinylation of 5mhC. Biotin molecules can be added to newly formed aldehyde groups on pretreated 5hmC. A glucose moiety is added to 5hmC via a glucosyltransferase enzyme, which has its vicinal hydroxyl groups converted to aldehydes *via* treatment with sodium periodate. We have previously discussed at length how sodium bisulfite treatment of 5mC does not result in a conversion in a similar time-frame to unmethylated cytosine. However, treatment of 5hmC yields another molecule: 5-methylenesulfonate. For the second method discussed in the article, Pastor et al. (2011) have succeeded in selectively isolating biotinylated 5hmC and sodium bisulfite converted 5hmC using streptavidin and anti-5-methylenesulfonate, respectively. In another recent publication, Kinney et al. (2011) exploited an isoschizomer pair of restriction enzymes, *Msp*I and *Hpa*II, that can differentiate between 5hmC and its glucosylated form. Coupled with qPCR, the team found that ES and brain cell genomic DNA contains a considerable amount of 5hmC, and identified novel loci containing 5hmC in both mouse ES and human brain DNA (Kinney et al., 2011).

CONCLUSION

The major developments in the methodologies for profiling and fingerprinting the human methylome have followed a clear progression toward innovative sequencing techniques at a single base-pair resolution. As this technology improves, the cost of genome-wide sequencing will decrease, resulting in a new wave

of DNA methylation data as more labs become fully immersed in the field. The bioinformatic tools will continue to improve in order to accurately analyze the vast datasets that will no doubt be generated in the coming years. The precedent for this has already been set through the Human Genome Project. Earlier this year, the International Human Epigenome Consortium was launched, aiming to map 1000 Epigenome by 2020 (IHEC, 2010). This is by no means an easy task, but if we see as many technical advances in the field in the next 10 years as we have in the previous decade, it is a challenge we are more than capable of facing.

REFERENCES

Adorjan, P., Distler, J., Lipscher, E., Model, F., Muller, J., Pelet, C., Braun, A., Florl, A. R., Gutig, D., Grabs, G., Howe, A., Kursar, M., Lesche, R., Leu, E., Lewin, A., Maier, S., Muller, V., Otto, T., Scholz, C., Schulz, W. A., Seifert, H.-H., Schwope, I., Ziebarth, H., Berlin, K., Piepenbrock, C., and Olek, A. (2002). Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* 30, e21.

Adouard, V., Dante, R., Niveleau, A., Delain, E., Revet, B., and Ehrlich, M. (1985). The accessibility of 5-methylcytosine to specific antibodies in double-stranded DNA of *Xanthomonas* phage XP12. *Eur. J. Biochem.* 152, 115–121.

Ahluwalia, A., Yan, P., Bigsby, R., Hurteau, J., Jung, S., Huang, T., and Nephew, K. (2001). DNA methylation and ovarian cancer I. Analysis of CpG island hypermethylation in human ovarian cancer using differential methylation hybridization. *Gynecol. Oncol.* 82, 261–268.

Annan, R., Kresbach, G., Giese, R., and Vouros, P. (1989). Trace detection of modified DNA bases via moving-belt liquid chromatography-mass spectrometry using electrophoric derivatization and negative chemical ionization. *J. Chromatogr.* 465, 285–296.

Beck, S., and Rakyan, V. (2008). The methylome: approaches for global DNA methylation profiling. *Trends Genet.* 24, 231–237.

Bestor, T. (2000). The DNA methyltransferases of mammals. *Hum. Mol. Genet.* 9, 2395–2402.

Bestor, T., Hellewell, S., and Ingram, V. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell. Biol.* 4, 1800–1806.

Bianco, T., Hussey, D., and Dobrovic, A. (1999). Methylation-sensitive, single-strand conformation analysis (MS-SSCA): a rapid method to screen for and analyze methylation. *Hum. Mutat.* 14, 289–293.

Bickle, T., and Krüger, D. (1993). The biology of DNA restriction. *Microbiol. Rev.* 57, 434–450.

Bird, A., and Southern, E. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation patterns in ribosomal DNA from *Xenopus laevis*. *J. Mol. Biol.* 118, 27–47.

Casadesús, J., and Low, D. (2006). Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* 70, 830–856.

Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* 6, 2125–2132.

Chen, C., Chen, H., Hsiau, T., Hsiau, A., Shi, H., Brock, G., Wei, S., Caldwell, C., Yan, P., and Huang, T. (2003). Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes. *Am. J. Pathol.* 163, 37–45.

Clark, S., Harrison, J., Paul, C., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990–2997.

Claverie, J. (2001). What if there are only 30,000 human genes? *Science* 291, 1255–1257.

Cokus, S., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., Pradhan, S., Nelson, S. F., Pellegrini, M., and Jacobsen, S. E. (2008). Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219.

Cross, S., Charlton, J., Nan, X., and Bird, A. (1994). Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* 6, 236–244.

Doskocil, J., and Sorm, F. (1962). Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim. Biophys. Acta* 55, 953–959.

Durbin, R. M., Abecasis, G. R., Altshuler, D. L., Auton, A., Brooks, L. D., Gibbs, R. A., Hurles, M. E., McVean, G. A., and 1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073.

Duthie, S., Narayanan, S., Blum, S., Pirie, L., and Brand, G. (2000). Folate deficiency in vitro induces uracil misincorporation and DNA hypomethylation and inhibits DNA excision repair in immortalized normal human colon epithelial cells. *Nutr. Cancer* 37, 245–251.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., and Beck, S. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38, 1378–1385.

Feil, R., Charlton, J., Bird, A., Walter, J., and Reik, W. (1994). Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.* 22, 695–696.

Feinberg, A. (2001). Methylation meets genomics. *Nat. Genet.* 27, 9–10.

Flusberg, B., Webster, D., Lee, J., Travers, K., Olivares, E., Clark, T., Korlach, J., and Turner, S. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.

Frigola, J., Ribas, M., Risques, R., and Peinado, M. (2002). Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Res.* 30, e28.

Frommer, M., McDonald, L., Millar, D. S., Collis, C., Watt, F., Grigg, G., Molloy, P., and Paul, C. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strand. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1827–1831.

Fryer, A., Emes, R., Ismail, K., Haworth, K., Mein, C., Carroll, W., and Farrell, W. (2011). Quantitative, high-resolution epigenetic profiling of CpG loci identifies associations with cord blood plasma homocysteine and birth weight in humans. *Epigenetics* 6, 86–94.

Galm, O., Rountree, M., Bachman, K., Jair, K., Baylin, S., and Herman, J. (2002). Enzymatic regional methylation assay: a novel method to quantify regional CpG methylation density. *Genome Res.* 12, 153–157.

Gama-Sosa, M., Midgett, R., Slagel, V., Githens, S., Kuo, K., Gehrke, C., and Ehrlich, M. (1983). Tissue-specific differences in DNA methylation in various mammals. *Biochim. Biophys. Acta* 740, 212–219.

Gebhard, C., Schwarzfischer, L., Pham, T., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. (2006). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* 66, 6118–6128.

Gitan, R., Shi, H., Chen, C., Yan, P., and Huang, T.-M. (2002). Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* 12, 158–164.

Gomes, J., and Chang, C. (1983). Reverse-phase high-performance liquid chromatography of chemically modified DNA. *Anal. Biochem.* 129, 387–391.

Gonzalgo, M., and Jones, P. (1997). Rapid quantitation of methylationdifferences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 25, 2529–2531.

Gonzalgo, M., Liang, G., Spruck, C. III, Zingg, J., Rideout, W. III, and Jones, P. (1997). Identification and characterization of differentially methylated regions of genomic DNA by methylation-sensitive arbitrarily primed PCR1. *Cancer Res.* 57, 594–599.

Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y., Oishen, A., Ballinger, T., Zhou, X., Forsberg, K. J., Gu, J., Echipare, L., O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C. B., Bernstein, B. E., Hawkins, R. D., Ren, B., Chung, W.-Y., Gu, H., Bock, C., Gnirke, A., Zhang, M. Q., Haussler, D., Ecker, J. R., Li, W., Farnham, P. J., Waterland, R. A., Meissner, A., Marra, M. A., Hirst, M., Milosavljevic, A., and Costello, J. F. (2010). Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat. Biotechnol.* 28, 1097–1105.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T. (1991). A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9523–9527.

Hayashizaki, Y., Hirosune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., Wada, S., Sugimoto, T., Kobayakawa, K., Kawara, T., Katsuki, M., Shibuya, T., and Mukai, T. (1993). Restriction landmark genomic scanning method and its various applications. *Electrophoresis* 14, 251–258.

Hayatsu, H., Wataya, Y., Kai, K., and Iida, S. (1970). Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* 9, 2858–2865.

Herman, J., Graff, J., Myöhänen, S., Nelkin, B., and Baylin, S. (1996). Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. U.S.A.* 93, 9821–9826.

Hirst, M., and Mara, M. (2010). Next generation sequencing based approaches to epigenomics. *Brief. Funct. Genomics* 9, 455–465.

Holliday, R., and Pugh, J. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.

Huang, T., Perry, M., and Laux, D. (1999). Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* 8, 459–470.

Internation Human Epigenome Consortium (IHEC). (2010). Available at: http:// www.ihec-epigenomes.org/ [accessed May 16, 2011].

Irizarry, R. A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S. A., Jeddeloh, J. A., Wen, B., and Feinberg, A. P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* 18, 780–790.

Ishkanian, A., Malloff, C. A., Watson, S. K., DeLeeuw, R. J., Chi, B., Coe, B. P., Snijders, A., Albertson, D. G., Pinkel, D., Marra, M. A., Ling, V., MacAulay, C., and Lam, W. L. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.* 36, 299–303.

Jones, P. A., and Laird, P. W. (1999). Cancer epigenetics comes of age. *Nat. Genet.* 21, 163–167.

Kaiser, J. (2008). DNA sequencing – a plan to capture human diversity in 1000 genomes. *Science* 319, 395.

Kawai, J., Hirotsune, S., Hirose, K., Fushiki, S., Watanabe, S., and Hayashizaki, Y. (1993). Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method. *Nucleic Acids Res.* 21, 5604–5608.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H., and Simon, I. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* 38, 149–153.

Khulan, B., Thompson, R., Ye, K., Fazzari, M., Suzuki, M., Stasiek, E., Figueroa, M., Glass, J., Chen, Q., Montagna, C., Hatchwell, E., Selzer, R. R., Richmond, T. A., Green, R. D., Maniatis Melnick, A., and Greally, J. M. (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* 16, 1046–1055.

Kinney, S., Chin, H., Vaisvilla, R., Bitinaite, J., Zheng, Y., Estève, P., Feng, S., Stroud, H., Jacobsen, S., and Pradhan, S. (2011). Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J. Biol. Chem.* 286, 24685–24683.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.

Kruglyak, L., and Nickerson, D. (2001). Variation is the spice of life. *Nat. Genet.* 27, 234–236.

Kuo, K., McCune, R., Gehrke, C., Midgett, R., and Ehrlich, M. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* 8, 4763–4776.

Kuppuswamy, M., Hoffmann, J., Kasper, C., Spitzer, S., Groce, S., and Bajaj, S. P. (1991). Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes. *Proc. Natl. Acad. Sci. U.S.A.* 88, 1143–1147.

Laird, C., Pleasant, N., Clark, A., Sneeden, J., Hassan, K., Manley, N., Vary, J., Morgan, T., Hansen, R., and Stöger, R. (2004). Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. *Proc. Natl. Acad. Sci. U.S.A.* 101, 204–209.

Laird, P. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* 11, 191–203.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y. J., and International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Lewin, J., Schmitt, A. O., Adorjan, P., Hildmann, T., and Piepenbrock, C. (2004). Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates. *Bioinformatics* 20, 3005–3012.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365.

Li, J., Gao, F., Li, N., Li, S., Yin, G., Tian, G., Jia, S., Wang, K., Zhang, X., Yang, H., Nielson, A., and Bolund, L. (2009). An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines. *BMC Genomics* 10, 223. doi:10.1186/1471-2164-10-223

Liang, G., Gonzalgo, M., Salem, C., and Jones, P. (2002). Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* 27, 150–155.

Lister, R., Pelizzola, M., Dowen, R., Hawkins, R., Hon, G., Tonti-Filippini, J., Nery, J., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A., Thomson, J., Ren, B., and Ecker, J. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322.

Mardis, E. (2008). Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9, 387–402.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben,

L. A., Berka, J., Braverman, M. S., Chen, Y.-J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L. I., Jarvie, T. P., Jirage, K. B., Kim, J.-B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Richard, F., Begley, R. F., and Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380.

Maunakea, A., Nagarajan, R., Bilenky, M., Ballinger, T., D'Souza, C., Fouse, S., Johnson, B., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V., Rowitch, D., Xing, X., Fiore, C., Schillebeechx, M., Jones, S., Haussler, D., Marra, M., Hirst, M., Wang, T., and Costello, J. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253–257.

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* 33, 5868–5877.

Meselson, M., Yuan, R., and Heywood, J. (1972). Restriction and Modification of DNA. *Annu. Rev. Biochem.* 41, 447–466.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99, 371–382.

Nouzova, M., Holtan, N., Oshiro, M., Isett, R. B., Munoz-Rodriguez, J. F., List, A. F., Narro, M. L., Miller, S. J., Merchant, N. C., and Futscher, B. W. (2004). Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. *J. Pharmacol. Exp. Ther.* 311, 968–981.

Oakeley, E. (1999). DNA methylation analysis: a review of current methodologies. *Pharmacol. Ther.* 84, 389–400.

Oakeley, E., Podestà, A., and Jost, J. (1997). Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 11721–11725.

Olek, A., Oswald, J., and Walter, J. (1996). A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* 24, 5064–5066.

Ordway, J., Bedell, J., Citek, R., Nunberg, A., Garrido, A., Kendall, R., Stevens, J., Cao, D., Doerge, R., Korshunova, Y., Holemon, H., McPherson, J., Lakey, N., Leon, J., Martienssen, R., and Jeddeloh, J. (2006). Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* 27, 2409–2423.

Pastor, W., Pape, U., Huang, Y., Henderson, H., Lister, R., Ko, M., McLoughlin, E., Brudno, Y., Mahapatra, S., Kapranov, P., Tahiliani, M., Daley, G., Liu, X., Ecker, J., Milos, P., Agarwal, S, and Rao, A. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397.

Patel, C., and Gopinathan, K. (1987). Determination of trace amounts of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* 164, 164–169.

Pennisi, E. (2001). The human genome. *Science* 291, 1177–1180.

Pfeifer, G., Steigerwald, S., Mueller, P., Wold, B., and Riggs, A. (1989). Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* 246, 810–881.

Poduslo, S., Dean, M., Kolch, U., and O'Brien, S. (1991). Detecting high-resolution polymorphisms in human coding loci by combining PCR and single-strand conformation polymorphism (SSCP) analysis. *Am. J. Hum. Genet.* 49, 106–111.

Prober, J., Trainor, G., Dam, R., Hobbs, R., Robertson, C., Zagursky, R., Cocuzza, J. A., and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* 238, 336–341.

Qiu, P., Soder, G., Sanfiorenzo, V., Wang, L., Greene, J., Fritz, M., and Cai, X. (2003). Quantification of single nucleotide polymorphisms by automated DNA sequencing. *Biochem. Biophys. Res. Commun.* 309, 331–338.

Rakyan, V., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., and Beck, S. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2, e405. doi:10.1371/journal.pbio.0020405

Rauch, T., Hongwei, L., Xiwei, W., and Pfeifer, G. (2006). MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res.* 66, 7939–7947.

Rauch, T., and Pfeifer, G. (2005). Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab. Invest.* 85, 1172–1180.

Rauch, T., Wu, X., Zhong, X., Riggs, A., and Pfeifer, G. (2009). A human B cell methylome at 100-base pair resolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 671–678.

Razin, A., and Ceder, H. (1991). DNA methylation and gene expression. *Microbiol. Mol. Biol. Rev.* 55, 451–458.

Reinders, J., Vivier, D., Theiler, G., Chollet, D., Descombes, P., and Paszkowski, J. (2008). Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res.* 18, 469–476.

Riggs, A. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 14, 9–25.

Sadri, R., and Hornsby, P. (1996). Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. *Nucleic Acids Res.* 24, 5058–5059.

Sandoval, J., Heyn, H. A., Moran, S., Serra-Musach, J., Pujana, M. A., Bibikova, M., and Esteller, M. (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6, 692–702.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* 241, 172–182.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. *Genomics* 90, 314–323.

Serre, D., Lee, B., and Ting, A. (2009). MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* 38, 391–399.

Singer-Sam, J., Grant, M., LeBon, J., Okuyama, K., Chapman, V., Monk, M., and Riggs, A. (1990). Use of HpaII-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol. Cell Biol.* 10, 4987–4498.

Smith, J., Arber, W., and Kühnlein, U. (1973). Host specificity of DNA produced by *Escherichia coli*: XIV. The role of nucleotide methylation in in vivo B-specific modification. *J. Mol. Biol.* 63, 1–8.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–517.

Southern, E., Mir, K., and Schchepinov, M. (1999). Molecular interactions on microarrays. *Nat. Genet.* 21, 5–9.

Suzuki, H., Itoh, F., Toyota, M., Kikuchi, T., Kakiuchi, H., Hinoda, Y., and Imai, K. (2000). Quantitative DNA methylation analysis by fluorescent polymerase chain reaction single strand conformation polymorphism using an automated DNA sequencer. *Electrophoresis* 21, 904–908.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M, Liu, D. R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.

Taylor, K., Kramer, R., Davis, J., Guo, J., Duff, D., Xu, D., Caldwell, C., and Shi, H. (2007). Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res.* 67, 8511–8518.

Thu, K., Pikor, L., Kennet, J., Alvarez, C., and Lam, W. (2010). Methylation analysis by DNA immunoprecipitation. *J. Cell. Physiol.* 222, 522–531.

Tost, J., and Gut, I. G. (2007). DNA methylation analysis by pyrosequencing. *Nat. Protoc.* 2, 2265–2275.

Vischer, E., and Chargaff, E. (1948). The composition of the pentose nucleic acids of yeast and pancreas. *J. Biol. Chem.* 176, 715–734.

Wagner, I., and Capesius, I. (1981). Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. *Biochim. Biophys. Acta* 654, 52–56.

Wang, Y., Gehrke, W., and Ehrlich, M. (1980). Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.* 8, 4777–4790.

Weber, M., Davies, J., Wittig, D., Oakeley, E., Haase, M., Lam, W., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862.

Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18, 7213–7218

Wilton, S., Lim, L., Dye, D., and Laing, N. (1997). Bandstab: a PCR based alternative to cloning PCR products. *Biotechniques* 22, 642–645.

Wittwer, C., Reed, G., Gundry, C., Vandersteen, J., and Pryor, R. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clin. Chem.* 49, 853–860.

Wojdacz, T. K., and Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Res.* 35, e41.

Wojdacz, T. K., Dobrovic, A., and Hansen, L. L. (2008). Methylation-sensitive high-resolution melting. *Nat. Protoc.* 3, 1903–1908.

Wu, J., Issa, J., Herman, J., Bassett, D., Nelkin, B., and Baylin, S. (1993). Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc. Natl. Acad. Sci. U.S.A.* 90, 8891–8895.

Wyatt, G. (1950). The purine and pyrimidine composition of deoxypentose nucleic acids. *Biochem. J.* 48, 584–590.

Xiong, Z., and Liard, P. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acid Res.* 25, 2532–2534.

Yan, P., Chen, C., and Shi, H. (2001). Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.* 61, 8375–8380.

Yan, P., Perry, M., Laux, D., Asare, A., Caldwell, C., and Huang, T. H.-M. (2000). CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. *Clin. Cancer Res.* 6, 1432–1438.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S., Chen, H., Henderson, I., Shinn, P., Pellegrini, M., Jacobsen, S., and Ecker, J. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126, 1189–1201.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 07 July 2011; accepted: 04 October 2011; published online: 25 October 2011.

Citation: Harrison A and Parle-McDermott A (2011) DNA methylation: a timeline of methods and applications. Front. Gene. ***2****:74. doi: 10.3389/fgene.2011.00074*

This article was submitted to Frontiers in Epigenomics, a specialty of Frontiers in Genetics.