



UNIVERSITY OF  
**LEICESTER**

**Investigating the role of DNA methylation in abdominal  
aortic aneurysms**

**Thesis submitted for the degree of Doctor of Philosophy at  
the University of Leicester**

**By Bradley J Toghill**

**Department of Cardiovascular Sciences**

**University of Leicester**

**2018**

# Investigating the role of DNA methylation in abdominal aortic aneurysms

**Bradley Toghil**

## **Abstract**

Abdominal aortic aneurysm (AAA) is a complex disease characterised by the irreversible dilation of the abdominal aorta. Little is known about its epigenetic basis, and DNA methylation is the most extensively studied epigenetic mechanism.

Global methylation was assessed in peripheral blood mononuclear cell (PBMC) DNA from 185 people using enzyme-linked immunosorbent assays (ELISAs), identifying global hypermethylation in those with a large AAA, and a linear association with increasing AAA diameter.

The regulatory regions of genes proximal to AAA genomic risk loci were then bisulphite sequenced using next-generation sequencing (NGS) in PBMC DNA from 96 people and vascular smooth muscle cell (VSMC) DNA from 44 people. In PBMCs, hypermethylation in individuals with AAA was seen in *LDLR*, *SORT1* and *IL6R*. In VSMCs, the same region in *IL6R* was hypermethylated and differential methylation was also observed in *ERG*, *SERPINB9*, and *SMYD2*.

ELISAs were conducted on plasma from the same PBMC samples to corroborate the methylation patterns seen in *LDLR*, *IL6R* and *SORT1*, where there was a reduction in circulating IL6R. Gene expression analysis was also performed on mRNA from the VSMCs. *SERPINB9* was downregulated in AAA but independently of DNA methylation, and a relationship between *SMYD2* promoter hypo-methylation and decreased *SMYD2* gene expression was shown. Downregulation of *SMYD2* in AAA was further corroborated in 6 whole aortic tissue samples using immunohistochemistry.

This PhD has illustrated a significant original contribution to knowledge at each stage. Global and gene specific DNA methylation changes are associated with AAA and could be involved in disease pathobiology. This is the first work to assess DNA methylation in AAA using NGS, and the first to assess methylation in VSMCs. In particular, methylation status of the *SMYD2* promoter could be a causal factor of decreased *SMYD2* expression, which has previously been implicated in adverse cardiovascular physiology and increased inflammation.

## **Acknowledgements**

Primarily I would like to express gratitude to my supervisor Professor Matt Bown.

Without his continuous guidance, time and expertise I would not be where I am now, and for which I am sincerely thankful. I would also like to thank Mr Athanasios Saratzis and Dr Eamonn Mallon for acting as secondary supervisors and providing support whenever I needed it.

There are many other people who helped me throughout this process at specific stages; particularly Ana Verissimo for helping me settle in when I first started, Nicolas Sylvius and Peter Freeman for assisting me with next-generation sequencing assays, Matt Blades for bioinformatics support, Jonathan Barber for teaching me cell culture, the UKAGS team for helping with sample recruitment, and the patients who donated their blood and tissue which I used for this research.

Finally I would like to acknowledge and thank the Department of Cardiovascular Sciences, University of Leicester, for awarding me this studentship opportunity.

## Contents

Investigating the role of DNA methylation in abdominal aortic aneurysms.....	i
List of Figures.....	ix
List of Tables.....	xv
Abbreviations.....	xviii
Publications resulting from PhD studies.....	xxi
Conference presentations during PhD studies.....	xxii
1. Chapter 1 - Introduction.....	1
1.1. Introduction to aortic aneurysmal disease .....	2
1.2. Abdominal aortic aneurysm (AAA) .....	2
1.2.1. AAA pathogenesis .....	3
1.2.2. Risk factors for AAA.....	4
1.3. AAA and atherosclerosis .....	5
1.3.1. Clinical risk factors and epidemiology of AAA and atherosclerosis .....	6
1.3.2. The Protective Role of Diabetes Mellitus in AAA .....	7
1.3. Thoracic aortic aneurysm (TAA).....	10
1.4. Genetic basis of aortic aneurysmal disease .....	12
1.4.1. Genetics of AAA .....	14
1.4.2 Genetics of TAA.....	19
1.5. Epigenetics and DNA methylation .....	25
1.6. Methylation quantitative trait loci (MeQTL).....	27
1.7. DNA methylation in cardiovascular disease.....	30
1.8. DNA methylation in AAA.....	31
1.8.1. DNA methylation and matrix degradation .....	32

1.8.2. DNA methylation and inflammation .....	33
1.8.3. DNA methylation and smoking .....	34
1.8.4. DNA methylation and ageing .....	35
1.8.5. DNA methylation and homocysteine.....	36
1.9. Epigenetic treatments.....	38
1.11. Conclusion .....	40
1.12. Aims.....	41
2. Chapter 2 - Methodology.....	44
2.1. Population.....	45
2.2. Vascular smooth muscle cell culture .....	46
2.3. DNA extraction.....	47
2.4. Global methylation analysis in peripheral blood DNA .....	48
2.5. Homocysteine analysis in blood plasma.....	50
2.6. Bisulphite conversion of DNA for sequencing.....	51
2.8. Bisulphite specific PCR for sequencing .....	54
2.9. Agarose gel electrophoresis .....	55
2.10. Cleaning of PCR amplicons and pooling for sequencing library preparation .....	56
2.11. Ligation of adaptors and barcoding for multiplex sequencing .....	57
2.11.1. Adaptor ligation .....	57
2.11.2. Cleanup of adaptor ligations .....	59
2.11.3. PCR enrichment of adaptor ligated DNA (barcoding) .....	60
2.12. Agilent Bioanalyzer analysis for sequencing .....	68
2.13. Next generation sequencing (NGS) – Illumina Mi-seq .....	71
2.14. Bioinformatics analysis of sequencing data .....	77
2.14.1. Statistical analysis of sequencing data.....	84
2.15. Summary of NGS workflow .....	85

2.16. Blood plasma protein analysis (ELISAs) for functional corroboration of PBMC differential methylation .....	87
2.17. RNA extraction for gene expression analysis in VSMCs.....	88
2.18. Genomic DNA digestion to remove contaminating DNA.....	89
2.19. CDNA synthesis for qPCR .....	89
2.20. TaqMan gene expression analysis (qPCR) for functional corroboration of VSMC differential methylation .....	90
2.21. Immunohistochemistry to assess protein expression in aortic tissues .....	93
2.22. Summary of statistical analysis .....	94
3. Chapter 3 – Results section 1.....	95
3.1. Introduction.....	96
3.1.1. Aims.....	97
3.1.2. Objectives .....	97
3.1.3. Hypotheses.....	98
3.2. Population.....	98
3.3. Global DNA methylation analysis.....	100
3.4. Circulating blood plasma homocysteine analysis.....	106
3.5. Conclusion .....	108
4. Chapter 4 – Results section 2.....	110
4.1. Introduction.....	111
4.1.1. Aims.....	113
4.1.2. Objectives .....	113
4.1.3. Hypotheses.....	114
4.2. Population.....	114
4.3. Targeted bisulphite sequencing of PBMC DNA from AAA and controls .....	115
4.3.1. Genes for PBMC sequencing.....	116
4.4. DNA methylation status of candidate gene regions in PBMCs.....	120

4.4.1 <i>IL6R</i> gene.....	123
4.4.2 <i>SORT1</i> gene .....	126
4.4.3 <i>LDLR</i> gene .....	129
4.4.4 <i>CDKN2B</i> gene .....	133
4.4.5 <i>SPAG17</i> gene .....	135
4.4.6 <i>GDF7</i> gene.....	137
4.4.7 <i>MMP9</i> gene.....	139
4.4.8 <i>ERG</i> gene .....	141
4.4.9 <i>LRP1</i> gene.....	143
4.5. Conclusion .....	145
5. Chapter 5 – Results section 3.....	148
5.1. Introduction.....	149
5.1.1. Aims.....	150
5.1.2. Objectives .....	150
5.1.3. Hypotheses.....	150
5.2. Population .....	151
5.3. Genes for VSMC sequencing assay.....	152
5.4. DNA methylation status of candidate gene regions in VSMCs .....	156
5.4.1. <i>ERG</i> gene .....	159
5.4.2. <i>IL6R</i> gene.....	164
5.4.3. <i>SERPINB9</i> gene .....	167
5.4.4. <i>SMYD2</i> gene .....	170
5.4.5. <i>DAB2IP</i> gene .....	173
5.4.6. <i>LDLR</i> gene .....	177
5.4.7. <i>MMP9</i> gene.....	179
5.4.8. <i>LRP1</i> gene.....	181
5.4.9. <i>SORT1</i> gene .....	183

5.5. Comparing DNA methylation profiles in PBMC and VSMC DNA .....	185
5.5.1. <i>ERG</i> comparison.....	186
5.5.2. <i>IL6R</i> comparison.....	188
5.5.3. <i>LDLR</i> comparison.....	190
5.5.4. <i>LRP1</i> comparison.....	192
5.5.5. <i>SORT1</i> comparison .....	194
5.6. Conclusion .....	196
6. Chapter 6 – Results section 4.....	199
6.1. Introduction.....	200
6.1.1 Aims for Chapter 6 – Results section 4 .....	201
6.1.2. Objectives for Chapter 6 – Results section 4.....	201
6.1.3. Hypotheses for Chapter 6 – Results section 4 .....	201
6.2. Plasma protein analysis to corroborate PBMC methylation.....	202
6.2.1. <i>IL6R</i> blood plasma analysis.....	204
6.2.2. <i>LDLR</i> blood plasma analysis.....	206
6.2.3. <i>SORT1</i> blood plasma analysis.....	208
6.2.4. LDL blood plasma analysis .....	210
6.3. Gene expression analysis to corroborate VSMC methylation.....	213
6.3.1. Gene expression and DNA methylation analysis .....	216
6.4. Immunohistochemistry to assess <i>SMYD2</i> in aortic tissue.....	221
6.5. Conclusion .....	223
7. Chapter 7 - Discussion.....	226
7.1. Summary of the potential role of DNA methylation in AAA .....	227
7.2. Summary of findings .....	228
7.3. Discussion of value and implications of findings.....	232
7.3.1. Global DNA methylation and homocysteine.....	232
7.3.2. PBMC DNA bisulphite sequencing.....	233

7.3.3. VSMC DNA bisulphite sequencing .....	235
7.4. Limitations and future work .....	239
Appendix I – Reference sequence for PBMC NGS bioinformatics .....	244
Appendix II – Reference sequence for VSMC NGS bioinformatics.....	249
References.....	257

## List of Figures

Figure 1 – Illustration of the aorta, sites of aortic aneurysmal formation and cross section of the aorta .....	10
Figure 2 – Genes associated with AAA in large genome wide studies .....	17
Figure 3 – Karyotype schema of abdominal aortic aneurysm (AAA) and thoracic aortic aneurysm (TAA) risk loci .....	24
Figure 4 – Overview of epigenetic mechanisms .....	26
Figure 5 - Visual representation of methylation quantitative loci (meQTL) .....	29
Figure 6 – Overview of the role of DNA methylation in AAA .....	37
Figure 7 – Summary flow chart of work conducted in this PhD .....	43
Figure 8 - The effects of sodium bisulphite on DNA .....	52
Figure 9 - Visual representation of adaptor ligation during sequencing library preparation .....	58
Figure 10 - Suggested set up of dual index strategy to ‘barcode’ DNA samples .....	64
Figure 11 - Plate layout for barcoding of PBMC samples for sequencing .....	65
Figure 12 - Plate layout for barcoding of VSMC samples for sequencing .....	66
Figure 13 – Data output from 2100 Agilent Bioanalyzer .....	69
Figure 14 – Principle of ddNTP base terminators in sequencing by synthesis reactions	71
Figure 15 - Summary of how Illumina NGS works .....	73
Figure 16 – Photograph of Illumina mi-seq .....	76
Figure 17 - Bioinformatics command list .....	80
Figure 18 - Summary of bioinformatics workflow and data analysis .....	81

Figure 19 - Screenshot from IGV software used to visually map high quality sequencing reads to the reference sequence for the LDLR gene promoter .....	82
Figure 20 - A higher resolution screen shot displaying individual bases from mapped sequencing reads to the reference sequence in the LDLR gene promoter.....	83
Figure 21 – Summary flow-chart of all the methodologies adopted to complete the next-generation sequencing experiments .....	86
Figure 22 – Global DNA methylation analysis .....	100
Figure 23 - The linear relationship between AAA size and global DNA methylation ...	102
Figure 24 - Smoking status and global DNA methylation .....	104
Figure 25 – Global DNA methylation in ex-smokers vs current smokers .....	105
Figure 26 - Homocysteine analysis in AAA and controls .....	106
Figure 27 – Global DNA methylation and homocysteine in AAA and controls .....	107
Figure 28 – Sequencing coverage of candidate genes in PBMC DNA .....	121
Figure 29 - Schematic representation of the sequenced region of the IL6R gene in PBMC and VSMC DNA .....	123
Figure 30 - DNA methylation status of the IL6R gene promoter in PBMC DNA .....	124
Figure 31 - Schematic representation of the sequenced region of the SORT1 gene in PBMC and VSMC DNA .....	126
Figure 32 - DNA methylation status of the SORT1 gene promoter in PBMC DNA .....	127
Figure 33 - Schematic representation of the sequenced region of the LDLR gene in PBMC DNA .....	129
Figure 34 - DNA methylation in the LDLR gene promoter in PBMC DNA .....	130

Figure 35 - Schematic representation of the sequenced region of the CDKN2B gene in PBMC DNA .....	133
Figure 36 - DNA methylation status of the CDKN2B gene promoter in PBMC DNA ..	134
Figure 37 - Schematic representation of the sequenced region of the SPAG17 gene in PBMC DNA .....	135
Figure 38 - DNA methylation status of the SPAG17 gene promoter in PBMC DNA ...	136
Figure 39 - Schematic representation of the sequenced region of the CDKN2B gene in PBMC DNA .....	137
Figure 40 - DNA methylation status of the GDF7 gene promoter in PBMC DNA .....	138
Figure 41 - Schematic representation of the sequenced region of the MMP9 gene in PBMC and VSMC DNA .....	139
Figure 42 - DNA methylation status of the MMP9 gene promoter in PBMC DNA .....	140
Figure 43 - Schematic representation of the sequenced region of the ERG gene in PBMC DNA .....	141
Figure 44 - DNA methylation status of the ERG gene promoter in PBMC DNA .....	142
Figure 45 - Schematic representation of the sequenced region of the LRP1 gene in PBMC DNA .....	143
Figure 46 - DNA methylation status of the LRP1 gene promoter in PBMC DNA .....	144
Figure 47 - Manhattan plot from GWAS meta-analysis .....	153
Figure 48 - Sequencing coverage of candidate genes in VSMC DNA .....	157
Figure 49 - Schematic representation of the sequenced region of the ERG gene in VSMC DNA .....	159

Figure 50 - DNA methylation status of one of two ERG gene promoters in VSMC DNA .....	160
Figure 51 - DNA methylation status of the second of two ERG gene promoters in VSMC DNA .....	162
Figure 52 - DNA methylation status of the IL6R gene promoter in VSMC DNA .....	165
Figure 53 - Schematic representation of the sequenced region of the SERPINB9 gene in VSMC DNA .....	167
Figure 54 - DNA methylation status of the SERPINB9 gene in VSMC DNA .....	168
Figure 55 - Schematic representation of the sequenced region of the SMYD2 gene in VSMC DNA .....	170
Figure 56 - DNA methylation status of the SMYD2 gene in VSMC DNA .....	171
Figure 57 - Schematic representation of the sequenced region of the DAB2IP gene in VSMC DNA .....	173
Figure 58 - DNA methylation status in one of two DAB2IP gene promoters in VSMC DNA .....	175
Figure 59 - DNA methylation status of the second DAB2IP gene promoter in VSMC DNA .....	176
Figure 60 - Schematic representation of the sequenced region of the LDLR gene in VSMC DNA .....	177
Figure 61 - DNA methylation status of the LDLR gene promoter in VSMC DNA .....	178
Figure 62 - DNA methylation status of the MMP9 gene promoter in VSMC DNA .....	180
Figure 63 - Schematic representation of the sequenced region of the LRP1 gene in VSMC DNA .....	181

Figure 64 - DNA methylation status of the LRP1 gene promoter in VSMC DNA .....	182
Figure 65 - DNA methylation status of the SORT1 gene promoter in VSMC DNA .....	184
Figure 66 – Comparison of PBMC and VSMC DNA methylation status in ERG .....	186
Figure 67 – Comparison of PBMC and VSMC DNA methylation status in IL6R .....	188
Figure 68 – Comparison of PBMC and VSMC DNA methylation status in LDLR .....	190
Figure 69 – Comparison of PBMC and VSMC DNA methylation status in LRP1 .....	192
Figure 70 – Comparison of PBMC and VSMC DNA methylation status in SORT1 .....	194
Figure 71 – IL6R blood plasma concentrations of 20 AAA and 20 controls .....	204
Figure 72 - Linear relationship between IL6R blood plasma and IL6R gene promoter methylation .....	205
Figure 73 – LDLR blood plasma concentrations of 20 AAA and 20 controls .....	206
Figure 74 - Linear relationship between LDLR blood plasma and LDLR gene promoter differential methylation .....	207
Figure 75 – SORT1 blood plasma concentrations of 20 AAA and 20 controls .....	208
Figure 76 - Linear relationship between SORT1 blood plasma and SORT1 gene promoter differential methylation .....	209
Figure 77 – LDL blood plasma concentrations of 20 AAA and 20 controls .....	210
Figure 78 - Linear relationship between LDL blood plasma and SORT1 gene promoter differential methylation .....	211
Figure 79 - Linear relationship between LDL blood plasma and LDLR gene promoter differential methylation .....	212
Figure 80 - Gene expression values of differentially methylated genes in VSMCs .....	214

Figure 81 - Linear regression between gene expression and mean DNA hypermethylation in ERG .....	216
Figure 82 - Linear regression between gene expression and mean DNA hypermethylation in IL6R .....	217
Figure 83 - Linear regression between gene expression and mean DNA hypermethylation in SERPINB9 .....	218
Figure 84 - Linear regression between gene expression and mean DNA hypo-methylation in SERPINB9 .....	219
Figure 85 - Linear relationship between gene expression and mean DNA methylation in SMYD2 .....	220
Figure 86 - Immunohistochemical staining for smooth muscle actin (SMA) and Smyd2 .....	222
Figure 87 – Summary flow chart of the results from this PhD project .....	230

## List of Tables

Table 1 – Epidemiology and risk factors associated with AAA and atherosclerotic plaques .....	9
Table 2 - Summary of differences between common epidemiological and risk factors associated with abdominal and thoracic aortic aneurysms .....	13
Table 3 – Genetics of AAA summary .....	18
Table 4 – Genetics of TAA summary .....	23
Table 5 - Bisulphite specific PCR cycling conditions .....	55
Table 6 - i5 primer index sequences for multiplex sequencing .....	61
Table 7 – i7 primer index sequences for multiplex sequencing .....	62
Table 8 – Cycling conditions for PCR enrichment of adaptor ligated DNA .....	67
Table 9 – NormFinder results for potential housekeeper genes .....	91
Table 10 - TaqMan gene expression qPCR cycling conditions .....	92
Table 11 - Demographic summary of samples used for peripheral blood global DNA methylation analysis .....	98
Table 12 - Demographic summary of samples used for homocysteine analysis in blood plasma.....	99
Table 13 - Demographic summary of samples where peripheral blood cell DNA was bisulphite treated and used for next generation sequencing .....	114
Table 14 - Summary of provisional GWAS meta-analysis results .....	116
Table 15 - Candidate genes for peripheral blood DNA bisulphite sequencing.....	118
Table 16 - Primers designed at targeted gene loci that were not successful .....	119
Table 17 – Descriptive statistics of the differentially methylated CpGs in IL6R after PBMC sequencing .....	125

Table 18 – Descriptive statistics of the differentially methylated CpGs in LDLR after PBMC sequencing .....	128
Table 19 – PBMC sequencing descriptive statistics of the differentially methylated CpGs in LDLR .....	132
Table 20 - Demographic summary of samples where VSMC DNA was used for bisulphite NGS .....	151
Table 21 - Candidate genes for vascular cell DNA bisulphite sequencing .....	154
Table 22 - Unsuccessful primers designed for bisulphite PCR at targeted gene loci ...	155
Table 23 – Descriptive statistics of the differentially methylated CpGs in ERG after VSMC sequencing .....	163
Table 24 – Descriptive statistics of the differentially methylated CpGs in IL6R after VSMC sequencing .....	166
Table 25 – Descriptive statistics of the differentially methylated CpGs in SERPINB9 after VSMC sequencing .....	169
Table 26 – Descriptive statistics of the differentially methylated CpGs in SMYD2 after VSMC sequencing .....	172
Table 27 – Descriptive statistics of the differentially methylated CpGs in ERG after comparison of PBMC and VSMC sequencing .....	187
Table 28 – Descriptive statistics of the differentially methylated CpGs in IL6R after comparison of PBMC and VSMC sequencing .....	189
Table 29 – Descriptive statistics of the differentially methylated CpGs in LDLR after comparison of PBMC and VSMC sequencing .....	191

Table 30 – Descriptive statistics of the differentially methylated CpGs in LRP1 after comparison of PBMC and VSMC sequencing .....	193
Table 31 – Descriptive statistics of the differentially methylated CpGs in SORT1 after comparison of PBMC and VSMC sequencing .....	195
Table 32 - Demographic summary of samples which were used to assess blood plasma protein concentrations of SORT1, LDLR, LDL and IL6R .....	203
Table 33 - Descriptive statistics of the blood plasma assays to corroborate methylation in PBMCs .....	203
Table 34 - Descriptive statistics of the gene expression assays to corroborate methylation in VSMCs .....	215
Table 35 – Timeline of PhD project .....	231

## Abbreviations

AAA	Abdominal aortic aneurysm
CAD	Coronary artery disease
CDKN2B	Cyclin dependent kinase inhibitor 2B
CDNA	Complementary DNA
CKD	Chronic kidney disease
CRP	C reactive protein
CpG	Cytosine-phosphate-Guanine
Ct	Cycle threshold
CVD	Cardiovascular disease
DAB2IP	DAB2 interacting protein
DNTP	Deoxynucleotide
DDNTP	Di-deoxynucleotide
DM	Diabetes Mellitus
DMR	Differentially methylated region
DNMT	DNA methyltransferase
ECM	Extra-cellular matrix
ERG	ETS-related gene
EQTL	Expression quantitative trait loci
EWAS	Epi-genome wide association study
FBN1	Fibrilin 1
FDR	False discovery rate

GDF7	Growth Differentiation Factor 7
GWAS	Genome wide association study
HAT	Histone acetyltransferase
HRP	Horseradish peroxidase
IHD	Ischemic heart disease
IL6R	Interleukin 6 receptor
IL-6	Interleukin 6
IGV	Interactive genome browser
IQR	Interquartile range
LDL	Low density lipoprotein
LDLR	Low-density lipoprotein receptor
LINE-1	Long interspersed nucleotide element-1
LRP1	Low density lipoprotein receptor-related protein 1
LS	Lost significance
MeQTL	Methylation quantitative trait loci
MFS	Marfan syndrome
MMP	Matrix metalloproteinase
NGS	Next-generation sequencing
NIHR	National institute for health research
OD	Optical density
OR	Odds ratio
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate buffer saline

PCR	Polymerase chain reaction
QPCR	quantitative PCR
RT-qPCR	Reverse transcription PCR
SAM	S-adenosylmethionine
SD	Standard deviation
SE	Standard error
SERPINB9	Serpin family B member 9
SMYD2	SET and MYND domain containing 2
SPAG17	Sperm Associated Antigen 17
SNP	Single nucleotide polymorphism
SORT1	Sortilin 1
TAA	Thoracic aortic aneurysm
TAD	Thoracic aortic dissection
TAAD	Thoracic aortic aneurysm and dissection
TAE	Tris acetate EDTA
TIMP	Tissue inhibitor of matrix metalloproteinase
TGF	Transforming growth factor
TNF	Tumour necrosis factor
UKAGS	UK aneurysm growth study
VSMC	Vascular smooth muscle cell
WGBS	Whole genome bisulphite sequencing

## **Publications resulting from PhD studies**

**Toghill BJ**, Saratzis A, Sylvius N, Freeman PJ, UKAGS collaborators, Bown MJ (2018). *SMYD2* promoter hypo-methylation is associated with abdominal aortic aneurysm and *SMYD2* expression in vascular smooth muscle cells. *Clinical epigenetics*. 10: 29. doi: 10.1186/s13148-018-0460-9.

**Toghill BJ**, Saratzis A, Bown MJ (2016). Abdominal Aortic Aneurysm – an independent disease to atherosclerosis? *Cardiovascular pathology*. 27: 71-75.

**Toghill BJ**, Saratzis A, Liyanage LS, Sidloff D, Bown MJ (2016). Genetics of aortic aneurysmal disease. *Encyclopedia of life sciences (ELS)*. DOI: 10.1002/9780470015902.a0026851.

**Toghill BJ**, Saratzis A, Harrison SC, Verissimo AR, Mallon EB, Bown MJ (2015). The potential role of DNA methylation in the pathogenesis of abdominal aortic aneurysm. *Atherosclerosis*. 241: 121–129.

## Conference presentations during PhD studies

Midlands Academy of Medical Sciences Research Festival – Loughborough University –  
23/03/2018

Abstract – oral presentation: **Toghill BJ**, Saratzis A, Sylvius N, Freeman PJ, UKAGS collaborators, Bown MJ (2018). *SMYD2* promoter hypo-methylation is associated with abdominal aortic aneurysm and *SMYD2* expression in vascular smooth muscle cells. *Clinical epigenetics*. 10: 29. doi: 10.1186/s13148-018-0460-9.

5<sup>th</sup> International Meeting on Aortic Diseases 2016 – Crowne Plaza Liege, Belgium –  
15/09/2016

Abstract - oral presentation: **Toghill BJ**, Saratzis A, Sayers R, UKAGS collaborators, Bown MJ. Global and gene specific DNA hypermethylation is associated with Abdominal Aortic Aneurysms. *Aorta*. <http://aorta.scienceinternational.org/abstracts/2016-imad-abstracts>.

The Society of Academic & Research Surgery – Royal College of Surgeons London –  
06/01/2016

Abstract - oral presentation: **Toghill BJ**, Saratzis A, Harrison SC, Samani NJ, Thompson JR, Sayers R, Sweeting MJ, Bown MJ (2016). The role of DNA methylation in the pathology of abdominal aortic aneurysms. *SARS 2016 Meeting Abstracts, British Journal of surgery*. 103 (S3): 1-56.

Vascular society ASM 2015 – Bournemouth International Centre – 11/11/2015

Abstract - oral presentation: **Toghill BJ**, Saratzis A, Harrison S, Sayers R, Samani N, Bown MJ (2016). Genome-wide DNA methylation is associated with abdominal aortic aneurysm. *SARS 2016 Meeting Abstracts, British Journal of surgery*. 103 (S5): 5-30.

**1. Chapter 1 - Introduction**

## 1.1. Introduction to aortic aneurysmal disease

Aortic aneurysmal disease is characterised by chronic mural degradation and expansion of the aorta, leading to eventual rupture and death if left untreated. Aneurysms are typically defined by an increase in arterial diameter of 1.5 times that of the normal artery, and increased aneurysm size is strongly associated with increased risk of rupture<sup>1-5</sup>.

Aortic aneurysmal disease is a life threatening condition for which there is currently no effective non-surgical treatment. Rupture is characterised by massive internal aortic hemorrhage and is an event that is frequently fatal. Patients are often asymptomatic until rupture occurs<sup>6</sup>. Aneurysms present in the abdominal and thoracic aorta, referred to as abdominal and thoracic aortic aneurysms (AAA and TAA respectively)<sup>1,2</sup>.

## 1.2. Abdominal aortic aneurysm (AAA)

This thesis will concentrate specifically on AAAs, which are responsible for over 3,000 deaths per annum in England and Wales, and approximately 2-4% of deaths in all Caucasian males over the age of 65 years<sup>7</sup>. One strategy to prevent high mortality is community screening to detect AAAs in asymptomatic patients before they cause death by rupture<sup>5</sup>. The majority of AAAs pose no immediate risk; however, these may grow over a number of years to a size where they do.

Individuals with AAA are typically asymptomatic before rupture<sup>8</sup>, and due to the long period between AAA development and rupture<sup>9</sup>, screening using ultrasound can detect AAA early and rupture can be prevented by surgical repair. A person with an AAA can remain under surveillance until it reaches a diameter of >5.4 cm, when surgical repair is often effective at preventing AAA-related mortality<sup>10,11</sup>, and screening can reduce AAA related mortality by 50%<sup>12</sup>.

Previous studies have demonstrated that annual rupture rates are as high as 11% and 26% for infrarenal AAAs with diameters of 5.5cm and 6.5cm respectively, and approximately 80% of all patients whose AAAs do rupture, die from the event<sup>13, 14</sup>. A therapeutic medical treatment target is therefore desired to remove the reliance on surgical treatment. However as of yet, no successful pharmaco-therapeutic strategies have been identified. The further identification of adverse genetic or epigenetic modifications directly linked to patients with AAA could offer a more comprehensive understanding of AAA pathobiology, and an alternative research avenue in the search for a potential future treatment strategy<sup>15, 16</sup>.

### 1.2.1. AAA pathogenesis

Chronic inflammation, extracellular matrix (ECM) degradation, thrombosis and haemodynamic forces are all important pathological hallmarks of AAA formation<sup>17-23</sup>. Inflammation is the main driver of AAA development, contributing towards vascular ECM remodelling and vascular smooth muscle cell (VSMC) death. The structural ECM is a complex network of protein interactions with an essential need for constant regulation. Degradation of the ECM is directly dependent on the counter balance between matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs (TIMPS)<sup>24</sup>. MMPs are central in the regulation of the inflammatory response through pro-inflammatory cytokine, chemokine and membrane receptor modification<sup>25</sup>.

A chronic activation, expression, or secretion of MMPs at localised inflammatory sites may have long term detrimental effects in the aorta. Excessive remodelling might explain the severe medial thinning, but not the inflammation seen in biopsies of the walls of advanced AAA<sup>26</sup>. Several inflammatory cells have been demonstrated in AAA biopsies, including; T and B lymphocytes, macrophages, neutrophils and mast cells<sup>27</sup>. Specifically,

monocytes/macrophages are involved in chronic vascular remodelling and subsequent AAA expansion<sup>19, 28</sup>.

MMP-9 is a major secretion product of macrophages<sup>29</sup>. Monocyte activity is modulated by pro-inflammatory cytokines, growth factors and chemokines. These mediators allow vessel wall endothelium interaction through the up-regulation of adhesion molecules<sup>30</sup>. The increased production and activation of MMPs is a hallmark feature of AAA. MMPs previously implicated include: MMP-1, -2, -3, -8, -9, -10, -12 and -13<sup>18</sup>. Of particular importance are elastases: MMP-2, -9 and -12. MMP-2 and -9 have catalytic ability to cleave various substrates, including elastin and collagens. MMP-12 however, has a higher specificity for elastin<sup>31, 32</sup>. Aneurysmal formation is the end product of a complex pathological process characterised by aortic wall connective tissue destruction. A range of evidence suggests that chronic destruction/remodeling of the ECM, tunica media elastin structures and reduction in VSMCs all have roles in AAA pathogenesis<sup>18, 31-33</sup>. The presence of activated MMP-2 and -9 contributes to degradation of the vascular ECM and the development of AAAs<sup>20</sup>. MMP-12 knockout in mice demonstrate that lack of MMP-12 expression results in the attenuation of aneurysm growth by a possible decrease in macrophage recruitment<sup>17</sup>.

### **1.2.2. Risk factors for AAA**

AAA is a highly complex and multifactorial disease. Smoking, family history, male-sex, age, and Caucasian ethnicity are the main risk-factors associated with AAA presence<sup>34-37</sup>.

The strongest independent risk factor for developing AAA is smoking, which has been identified in several epidemiological analyses of AAA and is a stronger risk factor for AAA than for atherosclerosis<sup>38-40</sup>. In a large study including small (3-5.4cm) and large AAAs (>5.4cm), 91.8% of 1090 AAA patients with aortic diameters of 4cm to 5.5cm

were smokers or had a history of smoking<sup>14</sup>. In another study, of 15,475 patients with small AAAs, current smoking was associated with an increased rate of expansion of 0.35 mm/year<sup>39</sup>.

First degree relatives of patients with AAA have an increased risk of developing the disease, and previous estimates have assessed overall heritability of at least 70%<sup>41,42</sup>. The most recent study was conducted on 414 twins with AAA, of which 69.8% were men and 30.2% women. The concordance rate in monozygotic twins was 30% compared with 12% in dizygotic twins. In the heritability analysis, 77% of the total variance was explained by additive genetic components and 23% was explained by non-shared environmental factors<sup>42</sup>. The genetic basis of AAA will be discussed extensively later in this Chapter.

Other important risk factors for AAA include age, ethnicity and male sex. Males are more likely to develop AAA, and the general male: female ratio is around 6:1. In an analysis of risk factors in a cohort of more than 3 million individuals the overall prevalence of AAA was estimated at 2.8% for men aged 65 to 79 years of age<sup>43</sup>. The same study also identified that Blacks, Hispanics, and Asians had lower risk of AAA than Whites and Native Americans. Other well-known risk factors were reaffirmed, including male sex, family history, and cardiovascular disease<sup>43</sup>.

### **1.3. AAA and atherosclerosis**

Atherosclerosis is an important risk factor for AAA, which commonly co-exists in the aneurysmal wall. The two diseases share a common aetiological basis, with the strongest factors being genetic background, increasing age and smoking<sup>38,44-46</sup>. Previous studies have also indicated the importance of increased haemodynamic force and hypercholesterolemia in AAA risk<sup>26,38,47</sup>.

Collectively, all these factors promote destruction/weakening of the aortic wall, predisposing to the effects of blood pressure and mechanical wall stress on the aorta, resulting in its gradual irreversible dilation. The traditional notion that AAA is just a result/manifestation of atherosclerosis has never been fully confirmed, and it is now more commonly suggested that factors other than atherosclerosis can independently and synergistically contribute towards AAA formation. This concept is debateable, but since the topic was last reviewed in 2010<sup>26</sup>, a variety of research has been published supporting the idea.

### 1.3.1. Clinical risk factors and epidemiology of AAA and atherosclerosis

There are a range of studies based around the similarities and differences in epidemiology and clinical risk factors associated with AAA and atherosclerosis. AAA prevalence was significantly higher in patients with CAD versus those without, suggesting that CAD represents a significant independent risk factor of AAA development<sup>48</sup>. Two common risk factors that are observed in AAA and atherosclerosis are cigarette smoking and male gender, however these are much more dominant for AAA than they are for atherosclerosis, and have been identified as risks for AAA independently of atherosclerosis<sup>26, 49</sup>.

In addition, hypercholesterolaemia is weakly associated with AAA but strongly linked to atherosclerosis<sup>47</sup>. Palazzuoli *et al.*, (2008)<sup>44</sup> conducted a study on 98 AAA patients vs 82 people who were viewed as being at high cardiovascular risk (2 or more risk factors) and observed that AAA patients displayed different risk factors, and also less risk factors than those in the cardiovascular risk group with respect to atherosclerosis. Ito *et al.*, (2008)<sup>50</sup> studied the epidemiological differences in atherosclerotic profiles between 343 patients with thoracic and abdominal aortic aneurysms (132 TAA and 211 AAA). The incidence

of coronary artery disease (CAD) in people with AAA was 53%, highlighting that atherosclerosis is not likely a causal event in AAA<sup>50</sup>.

The Tromsø study in 2010 investigated atherosclerosis and abdominal aortic diameter, and whether atherosclerosis was a risk marker for AAA in 6,446 people (3,164 men and 3,282 women) from a general population. There was absence of a consistent dose-response relationship between atherosclerosis and AAA diameter, and it was concluded that the findings from this study indicate that atherosclerosis is not causal in AAA, but more likely develops in parallel with or secondary to aneurismal dilatation<sup>51</sup>, which appears consistent with many studies conducted in relation to the clinical risk factors and epidemiology of the two diseases.

### **1.3.2. The Protective Role of Diabetes Mellitus in AAA**

For several years epidemiological data has suggested that patients with diabetes mellitus (DM) have a lower incidence of AAA<sup>52-54</sup>. Traditionally those with DM have at least a tenfold risk for developing cardiovascular disease (CVD) in their lifetime<sup>55</sup>. The real insight here is the contrasting role of DM in AAA to its causal role in atherosclerotic disease, providing a convincing argument against the view that AAA is just a manifestation of atherosclerosis.

A systematic review revealed that the prevalence of DM in patients with AAA was 6-14%, whereas in control patients without AAA, prevalence ranged from 17-36%<sup>56</sup>. The full biological aspect of this relationship is not yet known. Relevant research has illustrated the involvement of DM mediated changes in extracellular matrix biology<sup>28</sup>, where the aortic media may be protected from degeneration in a hyperglycaemic environment<sup>54</sup>. The main hallmark of diabetes is hyperglycaemia, which is associated

with increased risk of atherosclerosis<sup>57</sup>, but has interestingly been found to limit experimental aortic aneurysm progression<sup>58</sup>.

Miyama *et al.*, (2010)<sup>58</sup> induced AAA in hyperglycaemic mice to investigate blood glucose and AAA progression. 14 days after AAA induction, hyperglycaemic mice displayed a reduced level of AAA enlargement compared to euglycaemic mice, and the lowering of serum glucose levels with insulin treatment diminished this protective effect<sup>58</sup>. This study was the first to identify candidate mechanisms of hyperglycaemic AAA suppression *in vivo* characterised by the attenuation of physiological features of AAA including; reduced AAA diameter, mural neovascularization, macrophage infiltration and medial elastolysis<sup>58</sup>.

A full understanding of the processes involved in diabetic AAA suppression could potentially be utilised as a therapeutic strategy i.e. as a preventative measure of AAA progression and therefore reduce AAA associated mortality. This research is particularly helpful in our understanding of the pathogenic mechanisms of AAA and atherosclerosis and is possibly some of the strongest evidence that the two entities are to some extent independent.

Table 1 illustrates a comparison of epidemiological and clinical risk factors associated with populations of people with AAA and with established atherosclerotic plaques.

<b>Epidemiology and risk</b>	<b>AAA</b>	<b>Atherosclerosis</b>
Smoking	88-92% <sup>14, 50</sup>	40% <sup>59</sup>
Gender	~6:1 male: female ratio <sup>60</sup>	66% male <sup>59</sup>
Ethnicity	White (90%) <sup>43</sup>	White (58%) <sup>59</sup>
Hypercholesterolemia	28% <sup>61</sup>	46% <sup>59</sup>
Hypertension	81% <sup>50</sup>	82% <sup>62</sup>
Diabetes mellitus	11% <sup>43</sup>	31% <sup>63</sup>
Co-morbidity	Coronary artery disease in people with AAA was 27-53% <sup>43, 50</sup>	
Genetic basis	Shared and independent small effect genomic risk loci (susceptibility)	
Epigenetic basis	Global hypermethylation common, gene specific loci different	

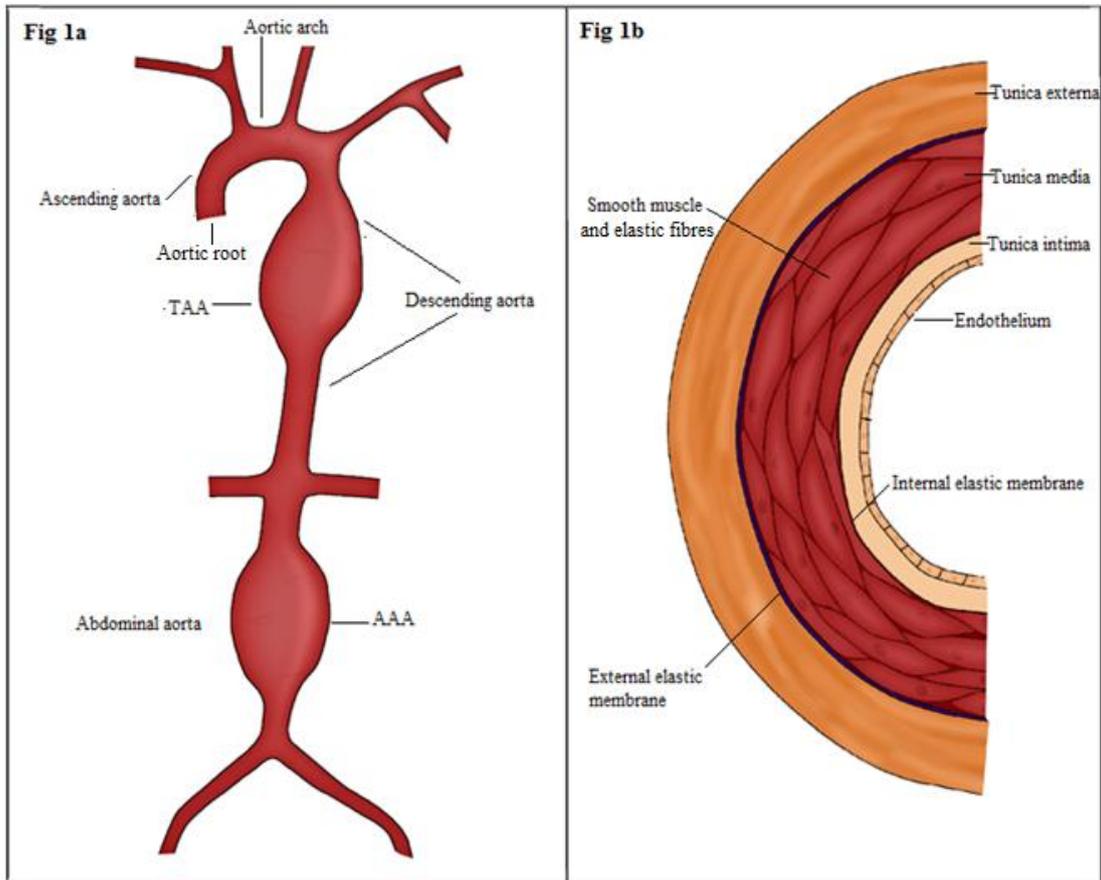
**Table 1 – Epidemiology and risk factors associated with AAA and atherosclerotic plaques:** Smoking, being white and being male are all more pronounced risk factors (% of study population) in AAA than atherosclerosis. However, hypercholesterolemia and diabetes mellitus appear higher risk for developing atherosclerosis.

### 1.3. Thoracic aortic aneurysm (TAA)

TAA is the enlargement of the thoracic aorta affecting the root, ascending aorta, or less frequently the arch and descending aorta. TAA is commonly accompanied by thoracic aortic dissection (TAD), which is often preceded by TAA. In TAD, blood is diverted from its usual location within the lumen of the aorta into a false lumen within the media through a tear in the intima<sup>35, 64</sup>. TAD is a dangerous complication of thoracic aortic disease, and is a medical emergency associated with high morbidity and mortality<sup>65</sup>. TAD is often preceded by TAA, even when minimal dilation is evident and aneurysm diameter is less than 5.5cm<sup>66</sup>. Therefore, risk of death is exacerbated in those with TAA when accompanied with TAD.

In addition to this fundamental difference between AAA and TAA, the epidemiology, risk factors, and genetic basis of disease are also different<sup>65</sup> due to changes in composition of the aortic wall from thoracic to abdominal regions. The thoracic aorta has a larger concentration of elastin, enabling it to withstand the increased pulse pressure exerted against it, and allows conformational dilation during systole and contraction during diastole. The thoracic regions of the aorta are derived from the neural crest during development, whereas the abdominal aorta has a mesodermal origin. Additionally, the physiological composition of collagen to elastin ratio decreases from the ascending aorta through the descending aorta, whilst the media becomes thinner<sup>67</sup>. These factors play a role in the presentation of the two distinct pathologies, which is reflected in the genetic basis of the two diseases.

For a diagrammatical illustration of AAA and TAA sites of development and a cross section of the structure of the aorta, see Figure 1.



**Figure 1 – Illustration of the aorta, sites of aortic aneurysmal formation and cross section of the aorta**  
**- Figure 1a:** Labelled representation of abdominal aortic aneurysm (AAA) and thoracic aortic aneurysm (TAA). TAA can develop in the aortic root, ascending aorta, or less frequently the arch and descending aorta. **Figure 1b:** Labelled cross section of the structure of the aorta. The epidemiology, risk factors, and genetic basis of disease are different between AAA and TAA due to change in composition of the aortic wall from thoracic to abdominal regions, with the thoracic having a larger concentration of elastin, enabling it to withstand the increased pulse pressure exerted against it, and allowing conformational dilation during systole and contraction during diastole.

#### 1.4. Genetic basis of aortic aneurysmal disease

Classical studies of inheritance before the genomics revolution were based on approaches of finding a single disease-causing gene. Since then it has become increasingly obvious that only a small minority of diseases follow a Mendelian mode of monogenic inheritance, and that several genes can interact and predispose to specific disease phenotypes. Hence, there are monogenic diseases (caused by single gene mutations of very large effect) and polygenic diseases (underpinned by many small effect loci).

Genome wide association studies (GWAS), whole genome sequencing, family linkage studies and candidate gene studies adopting a case-control cohort design are the most appropriate methodologies to study associations between gene variation and disease.

Family based linkage approaches (where entire multigenerational pedigrees are assessed in regards to genetic variants) are powerful, but are unrealistic with regards to aneurysmal disease given that it develops in older age, and sample numbers are generally low<sup>68</sup>.

For a summary of the key differences between risk factors and epidemiology commonly associated with AAA and TAA, see table 2.

<b>Abdominal Aortic Aneurysm</b>	<b>Thoracic Aortic Aneurysm</b>
<b>Prevalence/Incidence</b>	
~5% of population <sup>69</sup>	16.3/100,000 of population <sup>70</sup>
<b>Male/Female Ratio</b>	
~ 6:1 <sup>60</sup>	~ 2:1 <sup>70</sup>
<b>Risk of Rupture</b>	
6.5 cm = 26% <sup>14</sup>	6cm = 3.6% <sup>71</sup>
<b>Screening in the UK</b>	
Yes (since 2013)	No
<b>Family history</b>	
Less common - 15% of 1st degree relatives <sup>72</sup>	More common - 20% of all TAAs are familial <sup>73</sup>
<b>Genetic Basis</b>	
Several small effect loci (susceptibility)	Several large effect loci (often causal)
<b>Coronary Artery Disease in those with AAA/TAA</b>	
53% <sup>50</sup>	23% <sup>50</sup>
<b>Smoking</b>	
88-92% <sup>14, 50</sup>	76% <sup>50</sup>
<b>Hypertension</b>	
81% <sup>50</sup>	91% <sup>50</sup>

**Table 2 - Summary of differences between common epidemiological and risk factors associated with abdominal and thoracic aortic aneurysms.** AAA is more common in the general population, and develops more in men than women. Genetic causal factors are more prominent in TAA aetiology, whereas AAA is more multifactorial and significantly influenced by non-genetic risk factors, such as smoking.

### 1.4.1. Genetics of AAA

Much of the literature relating to the genetic basis of AAA focuses on candidate gene studies such as those encoding MMPs, lipid metabolism, interleukins and other key mediators of the inflammatory response due to the underlying pathophysiology of AAA<sup>74-76</sup>. However, many of these studies are not sufficiently powered or have failed to yield significant loci of interest with high odds ratios (OR) after meta-analyses or follow up<sup>68</sup>. For example, a variant in the *IL10* promoter was identified as associated with AAA with an OR of 1.8 in 100 AAA cases vs 100 controls, but when followed up in a larger, independent study (389 AAA cases, 404 controls) and adjusted for demographic covariates, no association was found<sup>68</sup>. This appears to be a recurring theme with smaller scale candidate gene studies and clearly emphasises the advantage of higher powered methodological techniques for studying disease association, such as GWASs.

By adopting a high-throughput genome wide approach, collaborative groups have conducted a range of studies and identified true effect variation at a total of 10 genomic loci for AAA. Gretarsdottir *et al.*, (2010)<sup>77</sup> identified that the A allele of rs7025486 on 9q33 was associated with AAA, which is located within *DAB2IP* and encodes an inhibitor of cell growth and survival. This analysis was conducted on 1,292 AAA and 30,503 controls, which was further validated in 3,267 AAA and 7,451 controls.

A discovery GWAS conducted by Bown *et al.*, (2011)<sup>78</sup> on 1866 patients with AAA and 5435 controls identified an association in intron 1 of the low-density lipoprotein receptor-related protein 1 (*LRP1*) gene (rs1466535), which was validated in a combined follow up of 6228 AAA and 49182 controls. LRP1 is a receptor involved in LDL cholesterol metabolism. CC homozygotes displayed an increased *LRP1* expression level compared to TT homozygotes.

Bradley *et al.*, (2013)<sup>45</sup> observed that variation at rs6511720 in the low-density lipoprotein receptor (*LDLR*) gene, which has previously been implicated in coronary artery disease (CAD)<sup>46</sup>, was also associated with AAA. The study was conducted on 1830 AAA cases with infrarenal aorta diameters of >3cm or ruptured AAA and 5435 controls, and rs6511720 was significantly associated overall in 3 of 5 individual replication studies.

Jones *et al.*, (2013)<sup>79</sup> identified a genetic variant (rs599839) involved with AAA risk. The rs599839 G allele near the Sortilin 1 (*SORT1*) gene, reached genome-wide significance in 11 combined independent cohorts (meta-analysis with 7048 AAA cases and 75 976 controls). This has been previously associated with abnormal *SORT1* protein expression, LDL blood cholesterol profile and CAD risk, but this association with AAA was independent of other known cardiovascular risk factors, and therefore similarly to *LRP1*, appears to be an independent risk for developing AAA.

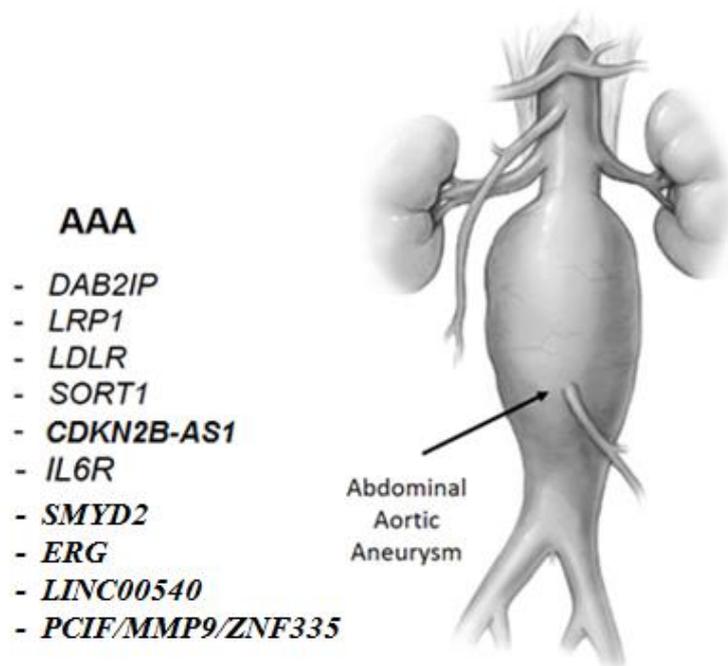
Harrison *et al.*, (2013)<sup>80</sup> demonstrated that signalling via the IL-6R pathway is likely involved in AAA, after a meta-analysis of 869 cases and 851 controls. Another meta-analysis of 4524 AAA and 15,710 controls demonstrated that *IL6R* rs7529229 was associated with lower risk of AAA.

Genetic variants at chromosomal region 9p21.3<sup>81</sup> are located away from known genes, and *CDKN2B-AS1* is proposed as the best candidate (also known as *ANRIL*), which is a non-coding RNA that regulates the expression of *CDKN2B* and *CDKN2A* epigenetically. Leeper *et al.*, (2013)<sup>82</sup> reported that *CDKN2B* knockout mice developed larger AAA as a result of increased VSMC apoptosis.

More recently the most extensive genomic study of the genetic basis of AAA was conducted in the form of a GWAS meta-analysis<sup>83</sup>. This study identified 4 new disease

specific loci associated with AAA, in addition to corroborating the previously known loci. These newly identified regions were 21q22.2 (*ERG*), 1q32.3 (*SMYD2*), 13q12.11 (*LINC00540*) and 20q13.12 (near *PCIF1/MMP9/ZNF335*).

Genetic predisposition is one of the largest risk factors for AAA, and it is clear here how the genes involved all relate to the underlying pathophysiology of the abdominal aortic structure. Small scale studies with low power, or those that have failed to be replicated, potentially indicating false positive results, have not been reported here. A summary of the genetic basis of AAA can be observed in Figure 2 and Table 3.



**Figure 2 – Genes associated with AAA in large genome wide studies:** Summary of genomic risk loci associated with abdominal aortic aneurysm (AAA): DAB2 interacting protein (*DAB2IP*), low-density lipoprotein receptor-related protein 1 (*LRP1*), low-density lipoprotein receptor; (*LDLR*), sortilin 1 (*SORT1*), interleukin-6 receptor (*IL6R*), SET and MYND domain containing 2 (*SMYD2*), cyclin-dependent kinase inhibitor 2B antisense RNA 1 (*CDKN2B-AS1*), ETS-related gene (*ERG*), long intergenic non-protein coding RNA 540 (*LINC00540*) and Matrix metalloproteinase 9 (*MMP9*) / Zinc Finger Protein 335 (*znf335*) / PDX1 C-Terminal Inhibiting Factor 1 (*PCIF1*). Figure adapted from Saratzis and Bown (2014)<sup>75</sup>.

Gene symbol	Gene location	NCBI gene ID	Gene function
<i>DAB2IP</i>	9q33.1	153090	DAB2IP is a Ras GTPase-activating protein (GAP) that acts as a tumour suppressor.
<i>LDLR</i>	19p13.2	3949	Receptor involved in endocytosis of LDL, which is normally bound at the cell membrane and taken into the cell lysosomes where the protein is degraded.
<i>LRP1</i>	12q13.3	4035	Another LDLR receptor involved in several cellular processes e.g. intracellular signalling, lipid homeostasis, and clearance of apoptotic cells.
<i>SORT1</i>	1p13.3	6272	Cell surface receptor involved in trafficking of different proteins to either the cell surface, or subcellular compartments (key role in lipid metabolism).
<i>CDKN2B-AS1</i>	9p21.3	100048912	Functional RNA molecule that interacts with PRC1 and PRC2, leading to epigenetic regulation of other genes in this cluster (CDKN2A/B).
<i>IL6R</i>	1q21	3570	Encodes a subunit of the interleukin 6 receptor complex, which is a potent cytokine that regulates cell growth differentiation and plays a role in the immune response.
<i>ERG</i>	21q22.2	2078	Member of transcription factor family expressed in the nucleus. ERG is a key regulator of several biological processes. ERG is required for vascular cell remodelling
<i>LINC00540</i>	13q12.11	100506622	Non-coding RNA gene (not translated into a protein) with no currently known function.
<i>PCIF1</i>	20q13.12	63935	May play a role in transcription elongation or in coupling transcription to pre-mRNA processing through its association with the phosphorylated C-terminal domain.
<i>MMP9</i>	20q13.12	4318	Breakdown of fibrous tissue during extracellular matrix homeostasis. MMP9 degrades type IV and V collagens - abnormal MMP9 activity is implicated in disease.
<i>ZNF335</i>	20q13.12	63925	Enhances transcriptional activation by ligand-bound nuclear hormone receptors and may function by altering local chromatin structure.
<i>SMYD2</i>	1q32.3	56950	Responsible for catalysing lysine methylation and is essentially involved in gene expression and chromatin organization pathways.

**Table 3 – Genetics of AAA summary:** Genes where mutations have been identified and are associated with the disease are listed with corresponding chromosomal locations, NCBI gene ID and gene function. Gene specific information was acquired from the NCBI database: <http://www.ncbi.nlm.nih.gov/gene>.

### 1.4.2 Genetics of TAA

TAA can also be multifactorial but is mainly a disease characterised by interaction between genetics and haemodynamics and does not follow an aetiology of genetic predisposition triggered by a range of alternative risk factors, such as those observed in AAA. As a result, there are no high-yield risk factors that can be used for screening the general population for TAA, and many people will continue to remain asymptomatic until rupture or dissection causes death. Numerous studies have been conducted in the pursuit of establishing the genetic basis of TAA and have resulted in the identification of a range of genomic loci with higher individual effect sizes than those observed in AAA. For non-syndromic and syndromic forms of TAA, a genetic basis has now been identified<sup>84</sup>. Genetic screening for certain types of TAAs with a fully established genetic aetiology can result in better identification, intervention and clinical outcome of patients<sup>85</sup>.

#### 1.4.2.1. Syndromic TAA

Syndromic TAA is associated with several inherited connective tissue disorders as mentioned previously<sup>84</sup>. Marfan syndrome (MFS) was associated with TAA first and is historically known to be a result of mutations in the Fibrilin 1 (*FBNI*) gene, which encodes an essential protein for the structure and subsequent functionality of the ECM. Almost all the TAAs that result directly from MFS occur in the aortic root. Around 600 mutations were observed in *FBNI* in patients with MFS<sup>86</sup>, suggesting that the multi-systemic features of the condition are a direct result of this. More recently it has become apparent that, not only do *FBNI* mutations directly result in dysregulation of the ECM structure, but ECM homeostasis also plays a significant role in MFS derived TAA, through transforming growth factor-beta (TGF- $\beta$ ) signalling<sup>87</sup>. TGF- $\beta$  molecules are cytokines that regulate embryonic development and adult tissue homeostasis through

complex signalling cascades<sup>84</sup>. *FBNI* mutations can disrupt normal TGF- $\beta$  signalling by increasing binding of ligand/receptor complexes, inducing TGF $\beta$  signalling events and thus promoting aneurysm development<sup>88</sup>.

Other gene mutations involved in transcriptional growth factor signalling pathways have also been observed in a more recently defined condition called Loeys-Dietz syndrome (LDS). Specifically, LDS is caused by mutation in TGF- $\beta$  receptors 1 and 2 (*TGFBR1* and *TGFBR2*)<sup>89</sup>. The main shared symptom of MFS and LDS is TAA, and those with LDS have other marfanoid symptoms, but LDS typically presents a more diverse phenotypic array e.g. vascular disease in LDS is not confined to the aorta. Dissections tend to occur in smaller aneurysms and at a younger age than MFS, and phenotypes arising from mutations in the *TGFBR2* gene display greater adversity than *TGFBR1*<sup>90</sup>. A range of research has been conducted in subsequent years, highlighting the role of gene variations that are a part of an inherited syndrome, and display a range of phenotypic effects, but where TAA is one of the primary adverse consequences. The main genes identified were: *SMAD2*, *3*, *4*, *SKI*, *TGFB2*, *TGFB3*, *COL3A1*, *FLNA* and *EFEMP2*<sup>84, 91-93</sup>.

There is strong evidence from studies investigating several inherited syndromes that suggest a causal, Mendelian role involving a range of genetic variants in the aetiology of TAA. The dysfunction of these genes disrupts the TGF- $\beta$  pathway, in addition to the disruption of the ECM and smooth muscle contractile apparatus, which results in structural damage to the thoracic aorta, causing aneurysm/dissection.

### 1.4.2.2. Non-syndromic TAA

Other studies have investigated the non-syndromic basis of TAA, which can be arbitrarily split in two further groups; familial (more than one family member affected) and sporadic (one family member affected). In the absence of a syndromic cause, it has been estimated that 20% of TAA cases are familial<sup>73</sup>, meaning sporadic cases make up the majority of all cases. Keramati *et al.*, (2010)<sup>94</sup> conducted a GWAS on just 15 family members using an autosomal dominant model of inheritance with incomplete penetrance. Three first degree relatives had TAA, and none of the 15 had syndromic features. The study mapped TAA to a single genomic locus on chromosome 15q21 with a peak LOD score of 3.6 at *FBNI*, suggesting that a familial, non-syndromic version of TAA had been observed. Following this, in a multistage GWAS, LeMaire *et al.*, (2011)<sup>95</sup> compared 765 people with sporadic TAA and 874 controls, again identifying variation at the 15q21 region which was associated with the disease. These SNPs all fell in to linkage disequilibrium within a large region within *FBNI*, also responsible for causing syndromic TAA in MFS, therefore suggesting a common pathogenesis between sporadic, familial and syndromic TAA. It is possible that any effective treatments designed to target syndromic TAA could also apply more widely to other forms of the disease in the future as a result of these findings.

Several other genes have been implicated in non-syndromic familial TAA. The *ACTA2* gene produces an  $\alpha$ -actin isoform protein involved in VSMC contraction which subsequently regulates aortic blood pressure and flow. Mis-sense mutations in this gene are responsible for around 14% of all cases of familial ascending TAAs. Three TAA families of European descent had mutations altering Arg258 to either a histidine or cysteine, and tissues from affected people displayed aortic medial degeneration, VSMC hyperplasia and disarray and stenotic arteries as a result of VSMC proliferation<sup>96,97</sup>. A

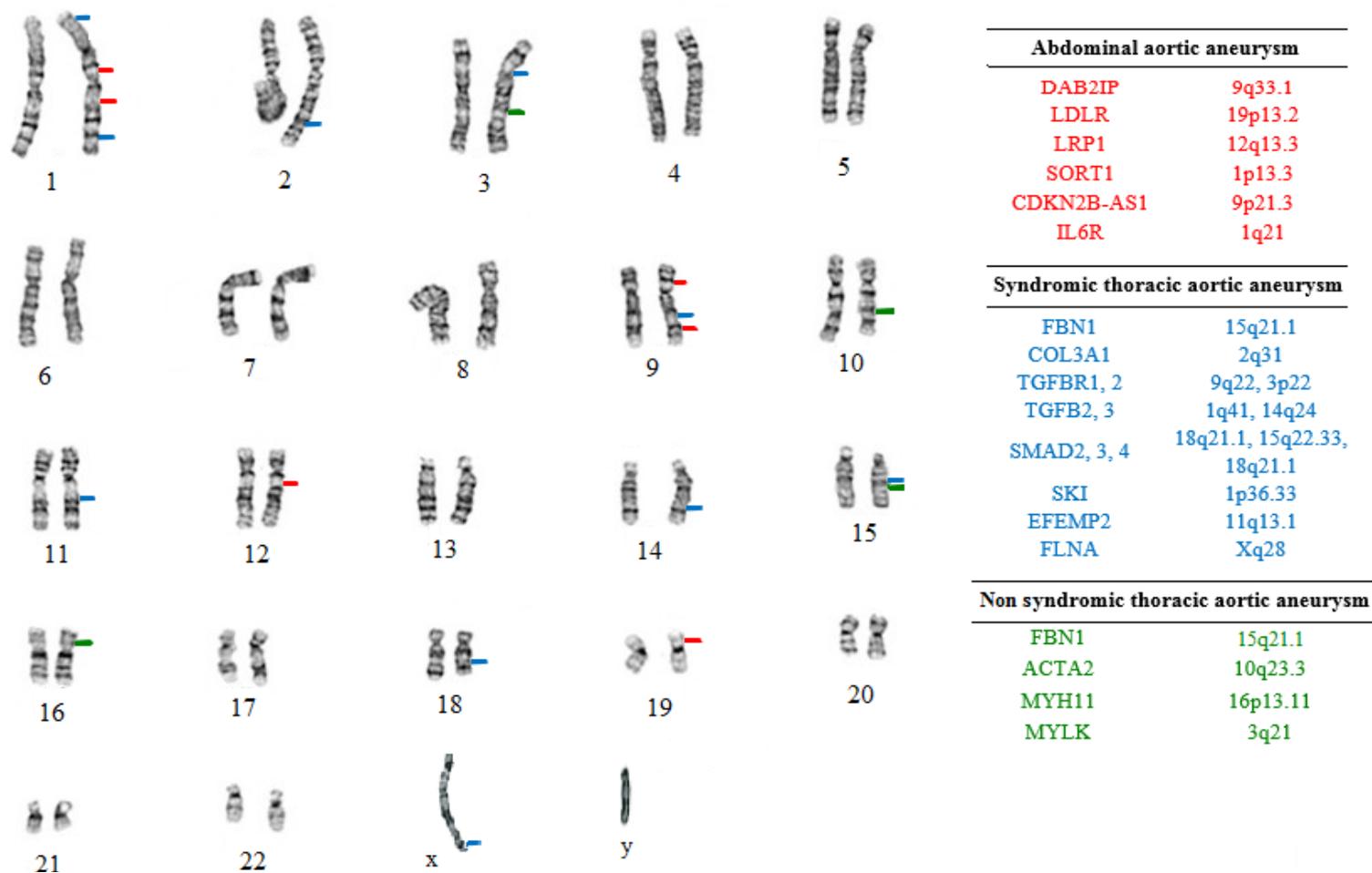
further study also showed *ACTA2* mutation plays a causal role in coronary artery disease, stroke, and Moyamoya disease<sup>96,97</sup>. *MYH11* and *MYLK* are further genes identified in non-syndromic TAA, and both are also involved in pathways relating to VSMC contraction. *MYH11* encodes a myosin heavy chain and *MYLK* gene encodes myosin light chain kinase. Disease causing mutations in *MYLK* do not occur as commonly as in *MYH11*, and *MYH11* is also a factor in ductus arteriosus<sup>85,98</sup>. The novel *MYH11* and *ACTA2* mutations observed revealed a role for enhanced TGF $\beta$  pathway signalling, as previously noted in other TAA associated mutations. However different molecular defects in TAA may account for a different pathogenic mechanism of enhanced TGF $\beta$  signalling<sup>99</sup>. Collectively, these genes indicate the importance of VSMC contraction in maintaining the structural integrity of the ascending aorta.

The continued study of TAA has given insight to its strong genetic, and often Mendelian basis. Syndromic TAA, non-syndromic familial TAA and non-syndromic sporadic TAA all display this pattern to an extent. The main difference between familial and sporadic TAA is the origin of mutation (having family history or not), rather than the actual site of mutation causing the disease, as causal pathways, such as *FBNI* gene variation and TGF- $\beta$  signalling can be common in the pathology of each type of TAA. Other genes can also be involved, and have been described here, which effect VSMC contractile and proliferation events. Other risk factors are generally not involved in the disease but haemodynamic force is an obvious reason for dilation of a weakened aorta, resulting from adverse gene mutations.

For a summary of the genetic basis of TAA, see Table 4, and for an illustrative karyotype-like schema of the AAA and TAA genomic risk loci described see Figure 3.

Disease	Gene symbol	Gene location	NCBI gene ID	Gene function
Syndromic TAA	<i>FBN1</i>	15q21.1	2200	Extracellular matrix glycoprotein is a structural component of calcium-binding micro fibrils. Provides force-bearing structural support in connective tissue.
	<i>COL3A1</i>	2q31	1281	Type III collagen, found in extensible connective tissues such as skin, lung, uterus, intestine and the vascular system, associated with type I collagen.
	<i>TGFBRI, 2</i>	9q22, 3p22	7046, 7048	Transmembrane proteins form a heteromeric complex with TGF-beta receptors, transducing the TGF-beta signal from the cell surface to the cytoplasm.
	<i>TGFB2, 3</i>	1q41, 14q24	7042, 7043	Family of cytokines, which are multifunctional peptides that regulate proliferation, differentiation, adhesion, migration and more, through TGFBRI/2 and SMAD proteins.
	<i>SMAD2, 3, 4</i>	18q21.1, 15q22.33,	4087, 4088,	Family of signal transducer and transcriptional modulator proteins that mediate multiple signalling pathways, including TGF-beta.
		18q21.1	4089	
	<i>SKI</i>	1p36.33	6497	Nuclear proto-oncogene - protein repressor of TGF-beta signalling, and may play a role in neural tube development and muscle differentiation.
	<i>EFEMP2</i>	11q13.1	30008	EGF containing fibulin like extracellular matrix protein 2 essential for elastic fibre formation and connective tissue development.
<i>FLNA</i>	Xq28	2316	Actin-binding protein that crosslinks actin filaments and links them to glycoproteins involved in remodelling the cytoskeleton to effect changes in cell shape and migration.	
Non-syndromic TAA	<i>FBN1</i>	See above		
	<i>ACTA2</i>	10q23.3	59	Alpha actin protein that that plays a role in cell motility, structure and integrity, constituting a major role in the contractile apparatus of smooth muscle cells.
	<i>MYH11</i>	16p13.11	4629	Smooth muscle heavy chain myosin that functions as a major contractile protein, converting chemical energy into mechanical energy through the hydrolysis of ATP.
	<i>MYLK</i>	3q21	4638	Encodes myosin light chain kinase which is a calcium/calmodulin dependent enzyme - facilitates myosin interaction with actin filaments to produce contractile activity.

**Table 4 – Genetics of TAA summary:** For syndromic thoracic aortic aneurysms (TAA) and non-syndromic TAA, genes where mutations have been identified and are involved in the diseases are listed with corresponding chromosomal locations, NCBI gene ID's, normal gene functions and study references. Gene specific information was acquired from the NCBI database: <http://www.ncbi.nlm.nih.gov/gene>.



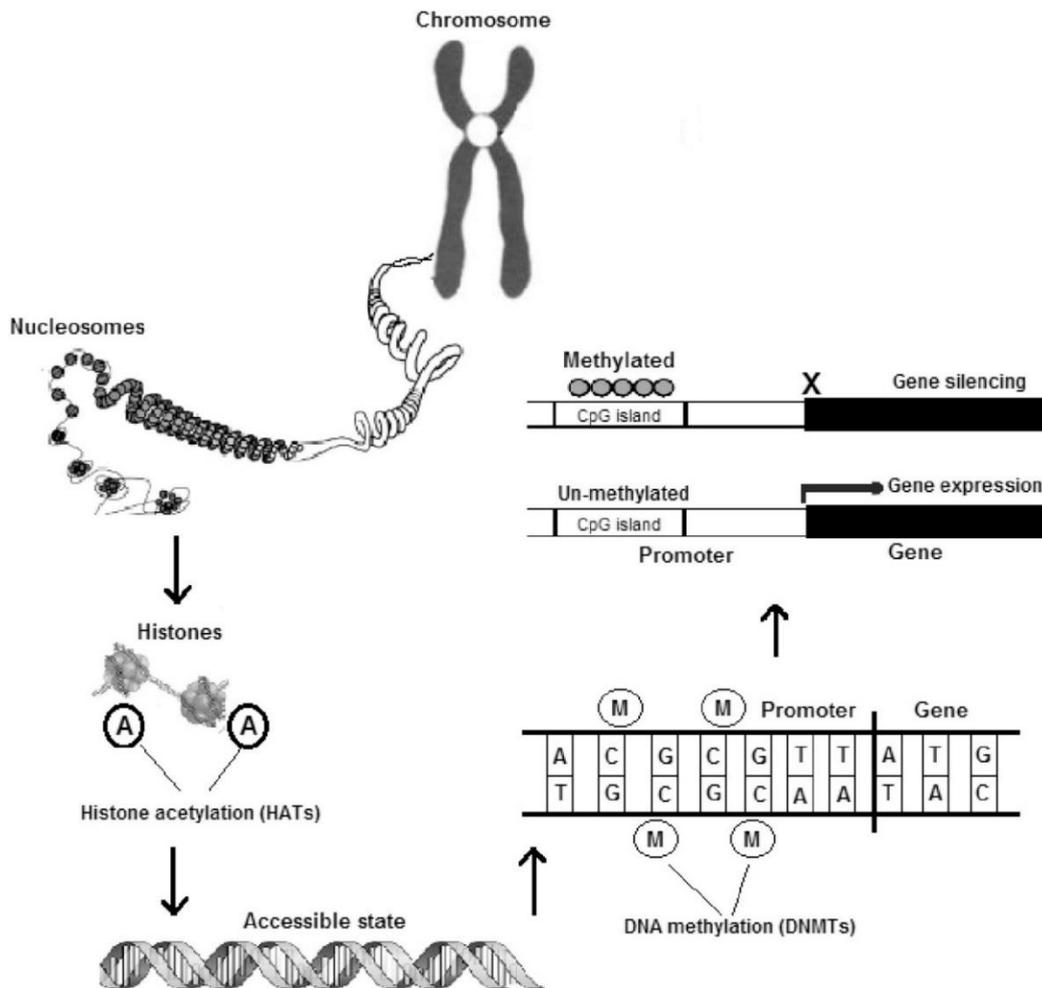
**Figure 3 – Karyotype schema of abdominal aortic aneurysm (AAA) and thoracic aortic aneurysm (TAA) risk loci:** All genes that are implicated in AAA or TAA are listed in a colour coded karyotype-like schema. AAA risk loci are illustrated in red, syndromic TAA in blue and non-syndromic TAA in green.

## 1.5. Epigenetics and DNA methylation

Epigenetics refers to the study of modifications to the genome that are not exclusively a result of change in the primary DNA sequence, and include DNA methylation, histone modifications and non-coding RNAs<sup>100</sup>. These mechanisms are essential for cell, tissue and organ development and directly interact with DNA sequence/structure to manipulate gene expression at the transcriptional and post-transcriptional level. However, adverse epigenetic effects have been implicated in complex disease, including cancers and cardiovascular disease<sup>101-103</sup>. There are a range of technical tools and methodologies available to study epigenetic processes and the epigenetic basis of disease. These methodologies are extensive in their application, and it is particularly important to adopt the best experimental design possible<sup>104, 105</sup>.

DNA methylation is the most extensively studied epigenetic modification to DNA and involves the addition of a methyl group to a cytosine base 5' to a guanine (CpG dinucleotide). Approximately 1% of the genome consists of CpGs, and 70% of CpGs are thought to be methylated in the mammalian genome<sup>106, 107</sup>. DNA methylation patterns induce long term gene expression pathways in selectively targeted genes during development, and methylation patterns are usually maintained throughout each subsequent cell division<sup>108</sup>. However, methylation is dynamic and changes in methylation can be induced through environmental shifts, such as a change in nutritional status or environmental exposures (e.g. smoking)<sup>109</sup>. CpG methylation status is particularly vulnerable to environmental change during development<sup>110</sup>, or as people age<sup>111</sup>, and DNA methylation can be inhibited, induced or reversed<sup>101, 103, 112, 113</sup>. Changes in gene promoter methylation result in direct changes to gene expression, and in particular

hypermethylation can reduce gene expression (Figure 4). However increased levels of methylation have also been associated with increased gene expression <sup>114</sup>.



**Figure 4 – Overview of epigenetic mechanisms:** Chromatin is the structure in which DNA is bound to histone proteins for the packaging of the genome into nucleosomes. Histone acetylation enzymes (HATs) are responsible for the opening up of chromatin into euchromatin, a more transcriptionally accessible state. Histone modifications enable the winding and unwinding of chromatin, which is linked closely to the regulation of gene expression<sup>115</sup>. CpG methylation is characterised by the addition of a methyl group (CH<sub>3</sub>) to the 5<sup>th</sup> carbon of a cytosine base that is 5' to a guanine. DNA methyltransferase (DNMT) enzymes are responsible for this. CpG islands (dense regions of CpG dinucleotides within gene promoter regions) can have abnormally high (hyper) or low (hypo) levels of DNA methylation in many disease processes<sup>116</sup>. DNA methylation of gene promoter regions can alter gene transcription due to interference of the transcriptional binding complex essential for the recruitment of RNA polymerases<sup>108</sup>.

DNA methyltransferase (DNMT) enzymes induce and maintain DNA methylation. There are two types: DNMT1 (maintenance enzymes) and DNMT3 (*de novo* enzymes).

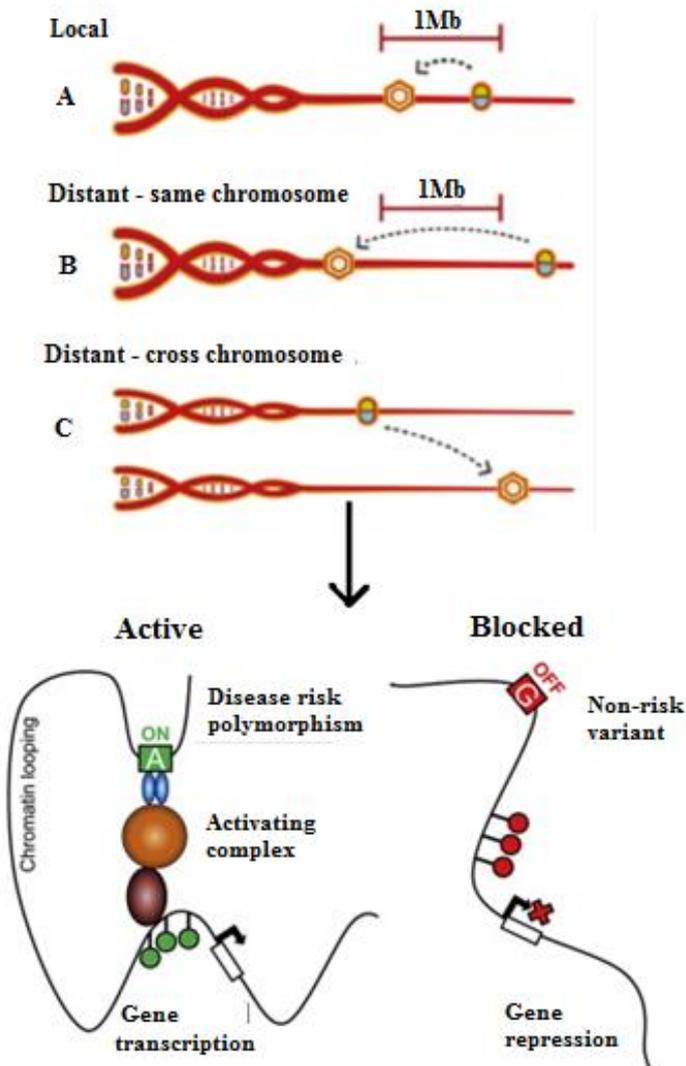
Research has demonstrated that both types work together to establish and sustain long term epigenetic patterns<sup>117</sup>. DNMT3 enzymes are categorised into two groups: DNMT3A and DNMT3B. DNMT3A is thought to have a preference for un-methylated DNA, but DNMT3B is equally efficient at methylating un-methylated and hemi-methylated CpG sites<sup>117</sup>. DNMT1 recognises hemi-methylated sites opposite newly synthesised DNA strands during mitosis, meaning newly synthesised strands are methylated, maintaining methylation symmetry<sup>118</sup>. Methylation patterns are therefore established during gametogenesis and embryogenesis with DNMT3A and DNMT3B, and subsequently maintained with DNMT1 and partly with DNMT3B<sup>117, 118</sup>. Non CpG related DNA methylation or hydroxymethylation may also be important in some processes<sup>119</sup>, but is not the focus of this thesis.

### **1.6. Methylation quantitative trait loci (MeQTL)**

Recent studies have revealed that individual genotypes at specific loci can result in different patterns of DNA methylation, known as methylation quantitative trait loci (meQTL), which are indicative of a direct genetic influence on the level of DNA methylation at specific CpG sites<sup>120</sup>. This is essentially a mechanism that bridges the gap between genetics and epigenetics and can influence phenotypic output. It has been specifically observed that single nucleotide polymorphisms (SNPs) located at CpG sites modulate genome-epigenome interactions<sup>121</sup>. There is some knowledge of meQTLs/meSNPs, but this is still very scarce and there is currently no large comprehensive database highlighting phenotype specific associations between genotypes and methylation. There have been studies that have successfully identified such

phenotypic specific interactions<sup>121-124</sup> and it is likely that there are still many existing meQTLs that are unreported.

Genetic variants can alter transcription factor levels and methylation of their binding sites<sup>125</sup>. There is a breadth of recent high profile evidence relating to disease specific and non-disease specific genetic variants and their regulatory effects on DNA methylation patterns, which subsequently regulate gene functionality<sup>121-123, 125-128</sup>. The direct mechanisms of the polymorphic risk loci associated with AAA are not fully known, but many are intronic<sup>83</sup> and it is therefore likely that their effects are regulatory in nature. It is feasible that meQTL exist in AAA at these loci, and considering many meQTL act in *cis*, their likely effects on methylation may be in the genes in proximity to/surrounding AAA risk loci. A schematic illustration of meQTL activity can be seen in Figure 5.



**Figure 5 - Visual representation of methylation quantitative loci (meQTL):** Polymorphic loci and methylated CpG sites can interact locally (a) and distantly (b and c). These interactions can result in the regulation of the functionality of genes and therefore contribute towards phenotypic diversity. Specific meQTL may be exclusive to certain disease phenotypes. Figure adapted from Heyn *et al.*,<sup>129</sup> and McClay *et al.*,<sup>130</sup>.

## 1.7. DNA methylation in cardiovascular disease

DNA methylation can influence disease onset and progression as a result of changes induced by environmental factors or genetic mutation, which have been studied extensively in many diseases<sup>109, 120-123, 125, 129-137</sup>. Genetic variation and environmental factors both have an influence on arterial phenotype, and cardiovascular diseases are truly multifactorial and complex<sup>138</sup>. Identifying the molecular epigenetic mechanisms and other non-genetic influences on disease are important because as of yet, they are relatively unresolved<sup>102</sup>, whereas the genetic basis of disease is better established. High throughput epigenome-wide association studies (EWAS) have allowed for informative detection of differences in the methylomes of different cells, tissues and diseased states. A case control approach has been widely adopted for such analyses, similarly to GWAS and candidate gene studies in genomics. Specifically, DNA methylation has been implicated in CVDs such as atherosclerosis and hypertension.

Zaina *et al.*, (2014)<sup>139</sup> created a human DNA methylation map of atherosclerosis using Whole-Genome Bisulfite Sequencing (WGBS) and DNA methylation micro-array analysis. WGBS identified a significant hyper methylation at numerous genomic loci of the atherosclerotic part of the aorta compared to healthy matched controls. With the use of Illumina's 450K micro-array, specific loci of differential DNA methylation were defined. 1895 candidate CpG sites were differentially methylated between the groups, and functional expression data (qPCR) corroborated the findings from 7 genes containing differentially methylated regions (DMRs) from the same samples<sup>102</sup>.

Hypertension is another disorder characterised by a multifactorial etiology that affects a large proportion of the western population<sup>140</sup>. Genome wide methylation analysis has suggested a key role for methylation in the pathogenesis of the disease. Wang *et al.*,

(2013)<sup>140</sup> were the first to conduct such an analysis. Using an age matched case-control design, a micro-array chip was conducted to determine differentially methylated regions (DMRs) between cases and controls. Select CpG sites were subsequently chosen in the *SULF1* gene for validation. It was observed that the methylation levels of 2 CpG sites in the promoter region of the *SULF1* gene were higher in hypertensives than in controls<sup>140</sup>. The differences in blood leukocyte DNA methylation sites between the groups suggests that changes in DNA methylation may play an important role in the pathogenesis of hypertension. In addition, it has also been shown that the observed changes in methylation have a direct influence on *SULF1* gene expression<sup>141</sup>.

### 1.8. DNA methylation in AAA

As illustrated previously, DNA methylation plays an important role in a range of CVDs<sup>102, 140, 142</sup>, but a very limited investigation into the epigenetic basis of AAA currently exists.

The only published AAA methylation study to date is that of Ryer *et al.*, (2015)<sup>143</sup> who conducted an Illumina 450k bead chip methylation micro-array on peripheral blood mononuclear cell (PBMC) DNA. They assayed 20 AAA patients (11 smokers and 9 non-smokers) and 21 control samples (10 smokers and 11 non-smokers). The sample groups contained smoking and non-smoking sub-groups since cigarette smoking is an important mediator of DNA methylation in tobacco-related cardiovascular disease.

They identified differentially methylated regions in CpG islands within four genes: *ADCY10P1*, *CNN2*, *KLHL35*, and *SERPINB9*. mRNA gene expression analysis was conducted for the differentially methylated genes identified, but only *SERPINB9* and *CNN2* exhibited transcriptional changes. Immuno-histological analysis then provided further evidence that the expression changes were also present in aortic tissues at the

aneurysmal site. Overall, the study findings were novel and illustrate a potential role for DNA methylation in the pathobiology of AAA.

Other than this study, much of the evidence for the potential role of DNA methylation in AAA has to be inferred from studies of atherosclerosis and hypertension, which share risk factors with AAA, and often co-exist with AAA<sup>102, 140, 144</sup>. Evidence can also be inferred from the molecular pathogenic mechanisms of AAA, in addition to major environmental risk factors<sup>145</sup>. The key processes known to be involved in the development of AAA are the immune response, ECM degradation and vascular remodeling<sup>146</sup>.

### 1.8.1. DNA methylation and matrix degradation

Changes to the methylation status of any gene can result in abnormal expression patterns and can subsequently contribute towards disease. Changes to DNA methylation in genes controlling the regulatory cycles of ECM proteolysis may influence phenotypic output of the structure and function of the ECM. Up-regulated production of MMPs are linked to tissue damage in degenerative inflammatory disorders other than AAA, including cancer metastasis, chronic kidney disease (CKD) and rheumatoid arthritis<sup>147-149</sup>. Fibronectin induced promoter de-methylation in the *MMP-2* gene increases expression in non-invasive breast cancer cell lines<sup>150</sup>. Additionally, Yuan *et al.*, (2014) observed that DNA de-methylation at the *MMP-9* promoter region enhances its expression in ectopic endometrial stromal cells<sup>151</sup>. MMP-2 and MMP-9 are two of the main enzymes involved in aortic remodelling and degradation. The feedback loop between the balance of MMPs and TIMPs has been investigated in the epigenomes of cancer cell lines. Proteolytic degradation of the ECM is directly dependent on the counter balance between MMPs and TIMPs<sup>147</sup>. Zhang *et al.*, (2014) identified that MMP-1, -2 and -7 were all up-regulated in cancer cell lines, whilst TIMP-3 and -4 were down-regulated. Targeted gene methylation

analysis (methylation specific PCR) revealed a correlation between aberrant promoter methylation status of the genes and their subsequent expression levels<sup>147</sup>. Transcription, expression and the relationship with promoter methylation status in normal and diseased states is actively investigated in many chronic inflammatory disorders<sup>102, 147, 152</sup>, with the exception of AAA.

### 1.8.2. DNA methylation and inflammation

Changes in epigenetic status of genes regulating inflammation and associated cells/proteins, may have significant functional effects in the vascular phenotype. It is known that altered DNA methylation profiles are linked to the risk and severity of several chronic inflammatory conditions<sup>153</sup>. For instance, a genome wide methylation study demonstrated that global DNA hypermethylation is associated with inflammation and increased mortality in CKD<sup>148</sup>.

Global DNA hypermethylation is commonly a hallmark of chronic inflammation<sup>102, 147-149, 154, 155</sup>, however it is unclear whether DNA methylation changes are a causal factor of inflammatory disease or vice versa. Just as monocytes are involved in AAA, they are also associated with inflammatory phenotypes of asthma.

Gunawardhana *et al.*, (2014) hypothesised that DNA methylation caused by environmental exposures may contribute to the heterogeneous inflammatory response in asthma<sup>156</sup>. To investigate this, monocytes were purified from blood taken from adults with eosinophilic, paucigranulocytic and neutrophilic asthma phenotypes. DMR's were identified in each phenotype compared to healthy controls using a methylation micro-array. Epigenomes of asthma phenotypes were significantly hypermethylated at DMR's in comparison to controls and 9 genes were common to all three phenotypes, some of which were implicated in an ECM-receptor proteolysis network<sup>156</sup>. This research

illustrates how potential environmental exposures can give rise to long term epigenetic modifications that take effect by inducing the inflammatory response. It is plausible that the induction of similar changes to methylation status within the genome may contribute to the development and progression of AAA.

### 1.8.3. DNA methylation and smoking

A range of risk factors discussed in this thesis so far are known to increase the risk of AAA development. The strongest of which are family history and smoking<sup>157</sup>. An important factor about the exposure to prolonged tobacco smoke is the effect it can have on the epigenomes of cells, especially blood leukocytes, which are heavily involved in AAA.

DNA methylation status is vulnerable to environmental shift, and studies have recently mapped where and how a change to CpG status occurs, and whether this is reversible after smoking cessation<sup>158</sup>. EWAS have demonstrated a mechanistic link between DNA methylation, current smoking, prenatal cigarette-smoke exposure and the development of adult chronic diseases<sup>159</sup>. Using a Human Methylation 450K array, Tsaprouni *et al.*, (2014) illustrated that cigarette smoking results in the hypo-methylation of PBMC DNA, as all but one of 30 probes (*CG23480021*) showed that smokers have lower methylation levels than non-smokers. In addition, smoking cessation only resulted in partial restoration of DNA methylation status, but DNA methylation was never completely reversed to non-smoking levels<sup>109</sup>.

Smoking plays a role in pathogenesis and increases the rate of expansion and risk of rupture of established AAAs<sup>40</sup>. The relationship between smoking, pathology and AAA growth is being investigated with the use of rodent models. Bergoeing *et al.*, (2007)<sup>36</sup> were the first to implicate smoking as a synergistic factor of aneurysm growth in an

animal model. There was no increase in MMP-9 or-12 expression when mice were exposed to tobacco smoke after a minor aortic elastase injury, which increased tissue inflammation, elastin degradation and subsequent aneurysm size by 134.5% (+/- 7.9%)<sup>36</sup>.

### 1.8.4. DNA methylation and ageing

It is possible that DNA methylation is another key contributor to age induced AAA.

Horvath, (2013) proposed a biological ageing clock, where DNA methylation age measures the cumulative effect of an epigenetic maintenance system<sup>160</sup>. This was developed as a multi-tissue predictor of age allowing the estimation of DNA methylation age of most tissues and cell types. The predictor was developed using 8,000 samples from 82 Illumina DNA methylation array datasets, encompassing 51 healthy tissues and cell types. As a result of this analysis, it was concluded that DNA methylation age has the following properties: it is close to zero for embryonic and induced pluripotent stem cells, it correlates with cell passage number, and it gives rise to a highly heritable measure of age acceleration.

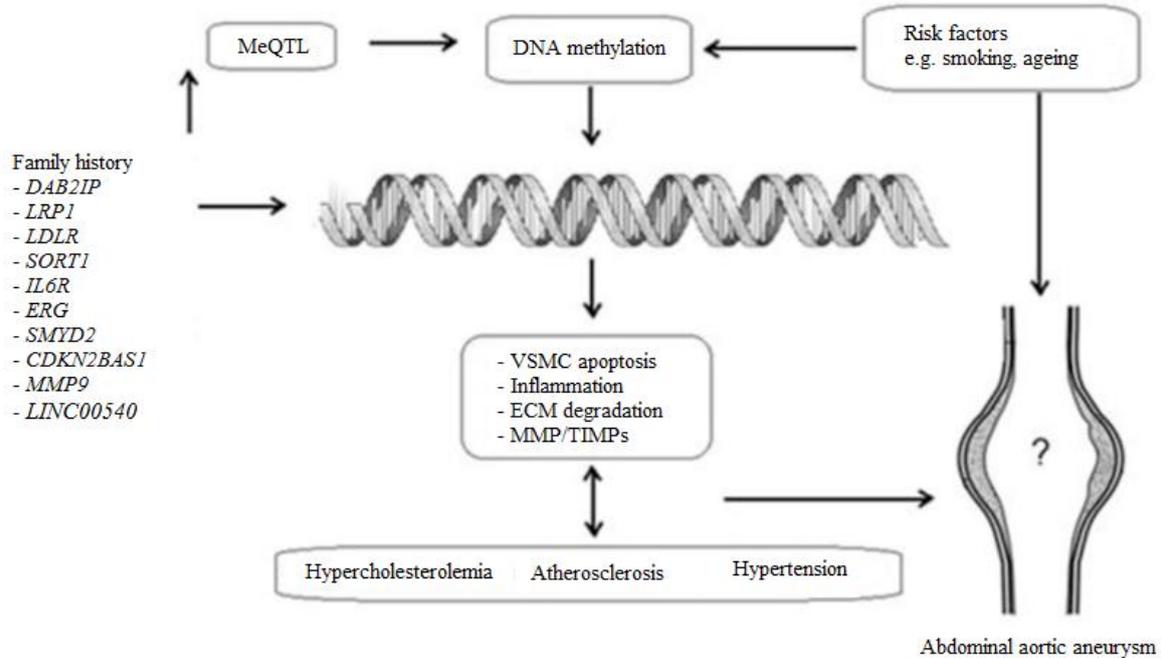
It is known that ageing has a strong effect on DNA methylation levels at thousands of CpG sites throughout the genome<sup>160, 161</sup>. The fact that DNA methylation patterns are only significantly altered during development, as a result of environmental shift, or stochastically as people age, suggest that the ageing process results in methylation differences that influence expression of genes involved in diseased phenotypes, such as AAA<sup>110, 111, 161</sup>. Age dependent methylation and subsequent changes in gene expression levels can be a large mediating factor in disease development. Generally, this is characterised by genome-wide hypo-methylation and promoter-specific hypermethylation<sup>162</sup>. There is a range of research relating to ageing and DNA methylation in CVD and associated risk factors, in addition to how the two interlink<sup>160, 161, 163</sup>.

### 1.8.5. DNA methylation and homocysteine

Krishna *et al.*,<sup>164</sup> reviewed the role of homocysteine mediated DNA methylation and associated epigenetic changes in AAA, which appears to be important in disease.

Homocysteine is a non-protein forming  $\alpha$ -amino acid that is synthesised from methionine and is strongly linked to the regulation of DNA methylation (homocysteine is produced as a by-product of the methyltransferase reaction)<sup>165</sup>. Aberrations to normal levels of homocysteine have been associated with adverse changes in DNA methylation, which can subsequently contribute towards disease<sup>165</sup>. Hyperhomocysteinemia is characterised by unusually high levels of blood circulating homocysteine that is associated with CVD, and is considered a significant, independent biomarker to determine cardiovascular risk, especially in atherosclerosis<sup>155, 166</sup>.

Genetic polymorphisms that alter enzymes involved in homocysteine metabolism, such as *MTHFR*, and vitamin deficiency can result in hyperhomocysteinemia<sup>167</sup>, and *MTHFR* rs1801133 was found to be significantly associated with AAA in a candidate gene association study (P=0.088) in 400 AAA and 400 controls<sup>168</sup>. However this loci was replicated with an established genome-wide significance. Some studies have demonstrated a significant association between homocysteine and AAA<sup>169</sup>, similarly to that observed in atherosclerosis, this is far from consistent across different studies<sup>170</sup>, and has been described as an enigma<sup>171</sup>. Lindqvist *et al.*,<sup>172</sup> concluded that homocysteine does not appear to be a useful biomarker in AAA, as it does for atherosclerosis. Additionally, no studies assessing homocysteine in AAA have also investigated the relationship between global methylation and homocysteine, which could be potentially insightful. Figure 6 displays a summary of the potential role of DNA methylation in AAA.



**Figure 6 – Overview of the role of DNA methylation in AAA:** DNA methylation may be the central mediating factor in AAA pathology. Environmental influences such as smoking and ageing predominantly take effect through the epigenome, but can also independently influence AAA. Family history directly influences disease predisposition, and several genes have been identified through genome wide analysis. DNA sequence variation can also associate with the epigenome through interactions known as methylation quantitative trait loci (meQTL). Changes in DNA methylation have been shown to affect expression of various genes, which have been implicated in vascular smooth muscle cell (VSMC) apoptosis, inflammation and extra-cellular matrix (ECM) degradation by an imbalance in matrix metalloproteinases (MMPs) and tissue inhibitors of MMPs (TIMPs). These processes are all pathogenic hallmarks of abdominal aortic aneurysms (AAA), and are also associated with hypercholesterolemia, atherosclerosis and hypertension, which are major risk factors of, and often co-exist with AAA.

## 1.9. Epigenetic treatments

As discussed, there is no pharmaco-therapeutic treatment for AAA, and disruption to the balance of epigenetic networks has the potential to cause major pathologies. Advances in new diagnostic tools might reveal other diseases that are not as well-mapped, that are also caused by epigenetic alterations, such as AAA. These factors have led to the development of new epigenetic therapies, since DNA methylation can be inhibited, induced or reversed<sup>101, 103, 112, 113, 173</sup>, meaning that pathogenic regions that are hyper or hypo-methylated can be selectively targeted for treatment. A clear example of this is the targeted induction of methylation at CpG sites in gene promoters of cancer cells as a targeted treatment by repressing over-expressed genes<sup>100, 103</sup>.

In addition to this, diet is emerging as a potential epigenetic regulator of CVD. The role that diet plays in the development of the molecular mechanisms underlying CVD is not fully known, but DNA methylation may account for the effect of dietary factors. Certain nutrients are needed for the production of S-adenosylmethionine (SAM), which is a methyl donor and can induce DNA methylation. These methyl nutrients include vitamins (folate, riboflavin, vitamin B12, vitamin B6, choline) and amino acids (methionine, cysteine, serine, glycine), and an imbalance in their metabolism can alter DNA methylation. It is proposed that a calorie restricted diet and adopted epigenetic diet (correct balance of methyl nutrients) may contribute to increased human longevity and prevention of chronic disease<sup>110, 174, 175</sup>.

Epigenetic therapies have been found to be effective in the treatment of cancer<sup>176</sup>.

Pathogenic changes to DNA methylation status are capable of silencing tumour suppressor genes, promoting cell and tumour growth and increasing the rate of metastasis.

It was observed that a methyl supplementation donor (SAM) inhibits cancer associated

skeletal metastasis by suppressing gene transcription of its selected target by inducing DNA hypermethylation (cancer associated skeletal metastasis is a result of DNA hypomethylation). The treatment on invasive prostate cancer cells resulted in dose dependent inhibition of cell proliferation, invasion and cell migration<sup>176</sup>. This research highlights a possible mechanism where pharmacological agents inducing DNA methylation at hypomethylated sites may prevent cancer metastasis and pathogenic methylation changes in other diseases.

Epigenetic treatments have also been extensively examined in CVDs. Inflammation is a chronic hallmark of atherogenesis in mice, which in turn is symbolised by genome wide and gene specific DNA hypermethylation<sup>102, 113</sup>. Cao *et al.*, (2014) investigated whether 5-aza-2'-deoxycytidine would mitigate atherosclerosis in knockout mice deficient in the *LDLR* gene, which is also associated with AAA<sup>75</sup>. 5-aza-2'-deoxycytidine is a demethylating compound and useful in the application of increasing the expression of down-regulated genes.

Atherosclerosis development was decreased in the Cao *et al.*, (2014) study as a result of the treatment without change to body weight, plasma lipid profile, macrophage cholesterol levels and plaque lipid content of the mice. Instead, the change was due to decreased macrophage inflammation. Macrophages treated with 5-aza-2'-deoxycytidine were found to have decreased levels of expression in inflammatory genes (TNF- $\alpha$ , IL-6, IL-1 $\beta$ ). By de-methylating and increasing the expression of loci promoting anti-inflammation (LXR $\alpha$  and PPAR $\gamma$ 1), the genes involved in inflammation were down-regulated. If such therapeutic solutions become common in reversing pathogenic, aberrant DNA methylation patterns, common methylation patterns in AAA could potentially be targeted in a similar way.

## 1.11. Conclusion

AAA is characterised by the chronic degradation and gradual, irreversible dilation of the abdominal aorta. Smoking, genetics, male sex, increased age and atherosclerosis are major factors associated with developing AAA. Rupture contributes to 2-4% of deaths in all Caucasians over 65, and there is no current pharmaco-therapeutic treatment strategy.

Methylation is an epigenetic modification to DNA, where a methyl group is added to a cytosine base 5' to a guanine (CpG dinucleotide). Methylation patterns are long term, inherited signatures that can induce changes in gene transcription, and can be affected by both genetic and environmental factors. Methylation changes are involved in hypertension and atherosclerosis, both of which are risk factors of, and often coexist with AAA. ECM degradation and inflammation, both important pathological hallmarks of AAA, are also promoted by changes in CpG methylation in other diseases. Additionally, the adverse effects of smoking and ageing take place largely through epigenetic manipulation of the genome.

Many factors associated with AAA are associated with DNA methylation, yet there is very limited investigation addressing the role of DNA methylation in AAA. Such an investigation to identify a link between aberrant DNA methylation changes and AAA could offer a more comprehensive understanding of AAA pathobiology, in addition to an alternative research avenue in the search for a potential treatment strategy. Epigenetic therapies are already being investigated to target pathogenic CpG methylation changes in other diseases such as atherosclerosis and hypertension, and it is feasible that these therapies may also be applicable to AAA in the future.

## 1.12. Aims

As discussed throughout chapter 1, it is not fully known whether abnormal epigenetic mechanisms play a role in the development and/or growth of AAA. There is a clear rationale and justification for further study, which this thesis will address with a series of aims.

### **Aims for Chapter 3 – Results section 1: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

1. To establish whether changes in global DNA methylation are evident in DNA taken from the peripheral blood of patients with AAA compared to healthy controls.
2. To establish whether changes in circulating homocysteine levels are evident in blood plasma taken from AAA patients compared to healthy controls.
3. To establish whether there is a direct relationship between global DNA methylation levels and circulating homocysteine.

### **Aims for Chapter 4 – Results section 2: Peripheral blood mononuclear cell DNA bisulphite sequencing for CpG methylation analysis**

1. To identify whether changes in peripheral blood CpG methylation exist in the regulatory regions of genes proximal to AAA genomic risk loci in those with AAA compared to healthy controls using targeted bisulphite next generation sequencing.

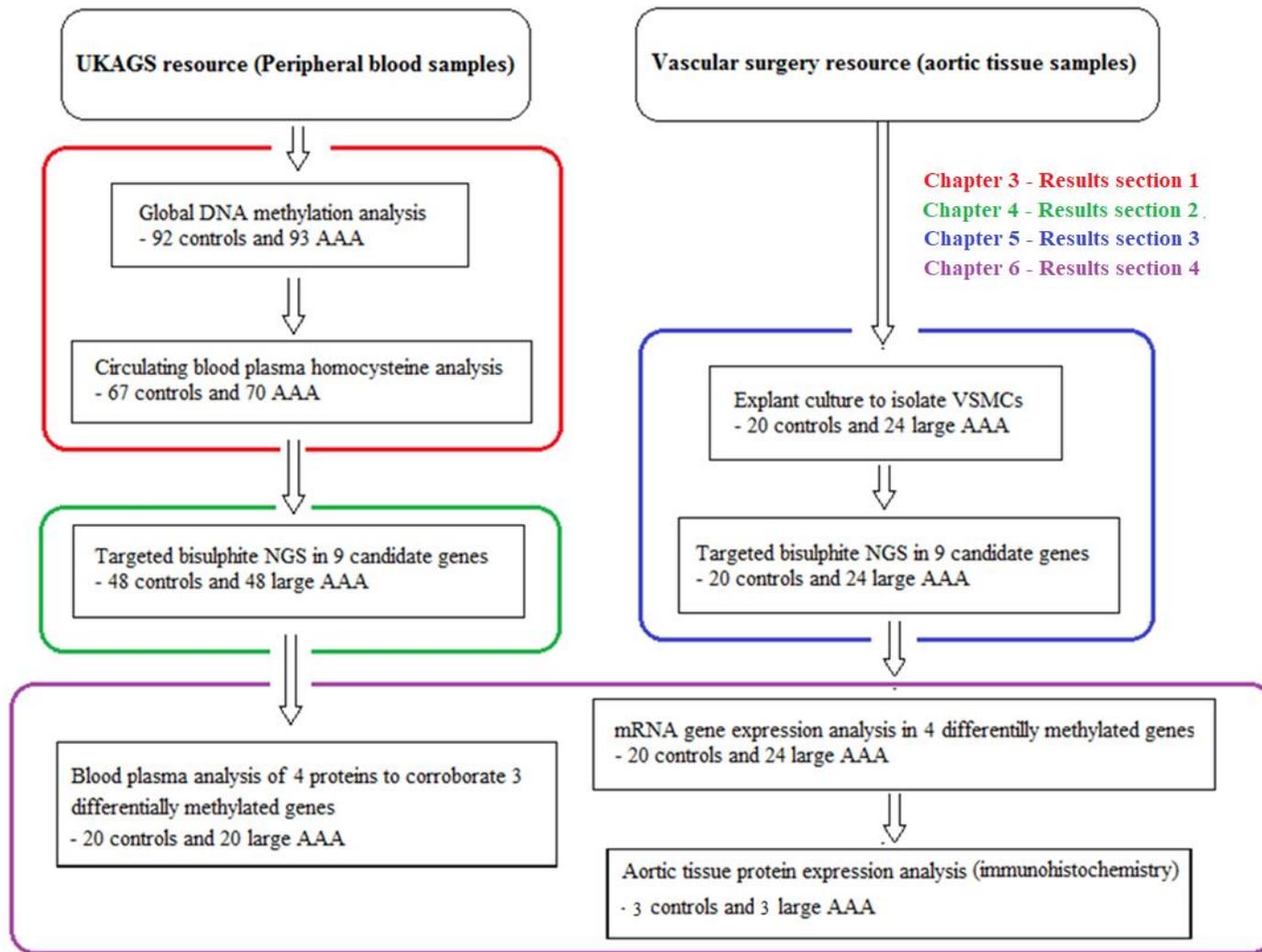
### **Aims for Chapter 5 – Results section 3: Vascular smooth muscle cell DNA bisulphite sequencing for CpG methylation analysis**

1. To identify whether changes in VSMC DNA methylation exist in the regulatory regions of genes (promoters and transcriptional start sites) proximal to AAA genomic risk loci in those with AAA compared to healthy controls using bisulphite next generation sequencing.
2. To compare the methylation profiles of commonly sequenced CpG sites in PBMCs and VSMCs.

### **Aims for Chapter 6 – Results section 4: Functional corroboration of observed methylation differences in AAA**

1. To determine whether changes to CpG methylation in peripheral blood DNA are directly associated with blood plasma concentration of the respective gene product.
2. To determine whether changes to CpG methylation in vascular cell DNA are directly associated with mRNA expression of the same gene.
3. To establish whether protein expression in whole aortic tissue is evident in genes where methylation and gene expression differences are observed.

Figure 7 displays a summary work flow of the project. The illustration highlights the overall work process from both sample resources used (peripheral blood and aortic tissues), and additionally highlights in which Chapter the work will be described. More detailed aims, objectives and hypotheses are stated in the introduction to each results chapter.



**Figure 7 – Summary flow chart of work conducted in this PhD:** DNA was isolated from peripheral blood and aortic tissues of those with AAA and controls. The chart illustrates the work process for each sample resource and highlights in which Chapter the results will be described.

**2. Chapter 2 - Methodology**

### 2.1. Population

Patients and controls in this research were recruited from two sources: The UK aneurysm growth study (UKAGS) (<http://www2.le.ac.uk/projects/ukags>) and the local AAA research programme based at the NIHR Leicester Biomedical Research Centre. The UKAGS collects peripheral blood samples from men with AAA (aortic diameter 3cm or greater) and healthy controls (all screened for AAA) recruited from the English NHS AAA Screening Programme. Furthermore, AAA biopsies (aortic wall) are taken from patients undergoing open AAA repair in the regional vascular unit, where there is also an established aortic tissue collection from cadaveric organ donors. Ethical approval by an NHS Research Ethics Committee was obtained for both studies.

A total of 185 blood samples from the UKAGS were selected for global methylation analysis (Chapter 3). Of these 185 samples, 96 (48 large AAA and 48 controls) were also selected for PBMC DNA bisulphite NGS (Chapter 4). In addition, 24 aortic biopsies were collected from people who were undergoing open AAA repair, and 20 aortic tissue samples were collected from cadaveric organ donors with the intention of isolating VSMCs via explant culture for bisulphite NGS (Chapter 5). There were no signs of AAA or atherosclerosis in the abdominal aortas obtained from organ donors, which were therefore used as controls. Detailed descriptions of sample group demographics for each experiment are provided in the appropriate results chapters.

**2.2. Vascular smooth muscle cell culture**

Aortic tissues were collected by vascular surgeons within the vascular surgery group at the University of Leicester. For each aortic tissue sample collected from those undergoing AAA surgery (n=24), and from cadavers (n=20), explant culture was performed to isolate and grow VSMCs *in vitro*. This was particularly important for this study (Chapter 5), given that independent cell lines have different epigenetic profiles, and the aim of this experiment was to assess the methylation profiles of VSMCs within the aorta without other cell types potentially confounding the results. VSMCs are a good surrogate for such analysis considering aneurysmal formation is characterised by inflammation and VSMC apoptosis in the tunica media of the aortic wall. As a result, any changes observed in methylation in the DNA from VSMCs would provide potential insight to the pathobiology of AAA.

Thin sections of the tunica media were carved from the whole aortic tissues no longer than 48 hours after surgery (AAA samples), or death (cadaveric samples), and placed in T80 cell culture flasks containing 10ml smooth muscle cell Medium 231 (ThermoFisher scientific - M2315005) with added Smooth Muscle Growth Supplement (ThermoFisher scientific - S00725). A 25ml bottle of growth serum was added to each 500ml 231 media prior to use. The flasks containing the media, growth supplement and aortic tissues were left to incubate at 37°C until visible primary cell growth was observed (~2 weeks). After primary growth of VSMCs, confluent cells were detached from the flask with the addition of 2x trypsin-EDTA solution and incubated at 37°C for 3 minutes. 5ml sterile phosphate buffered saline (PBS) and fetal bovine serum solution (20:1 ratio) was added to neutralise the trypsin. Solutions containing detached cells were aspirated from the flasks into sterile universal tubes and centrifuged at 400xg for 6 minutes at 20°C. Supernatants were

discarded and cell pellets were re-suspended in 3ml media. 1ml of the 3ml suspensions were added to 3 new T80 cell culture flasks, each containing 10ml of media. This process was repeated until sufficient amounts of isolated VSMCs (two clearly visible cell pellets) were available for DNA and RNA extraction, however never beyond passage 3 to attenuate any potential changes in cellular gene expression or DNA methylation patterns because of the artificial media/culture environment. Isolated cultured cells stored by cryogenic freezing at  $-80^{\circ}\text{C}$  until needed for experimental analysis.

### 2.3. DNA extraction

DNA was extracted from each PBMC (n=185) and VSMC (n=44) sample used in this research with the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturers standard protocol. Isolation of DNA from cells is essential for all genetic based research, and the most recent and optimised technique was adopted in this project. The technique uses the premise of cell lysis to 'free' endogenous cellular DNA in a solution, which is then transferred to a spin column containing a silica gel membrane. If the pH and salt concentration of the binding solution are optimal (provided by the manufacturer), the DNA will bind to the silica gel membrane as the solution passes through during centrifugation, resulting in high quality and purified yields of DNA.

The protocol for DNA extraction from PBMCs and VSMCs is as follows:

Frozen buffy coats (PBMC samples) and cultured cell pellets (VSMC samples) were removed from the freezer and thawed at room temperature. 20 $\mu\text{l}$  proteinase K and 200 $\mu\text{l}$  of the buffy coats were added to 1.5mL Eppendorf tubes, whereas 200 $\mu\text{l}$  PBS and 20 $\mu\text{l}$  proteinase K were added to the cell pellets in 1.5mL Eppendorf tubes. 200 $\mu\text{l}$  buffer AL was subsequently added to all tubes, which were vortexed and incubated for 10 minutes at  $56^{\circ}\text{C}$ . The tubes were centrifuged briefly before adding 200 $\mu\text{l}$  100% ethanol, vortexed,

and again centrifuged briefly. Entire solutions were transferred via pipet to labelled DNeasy mini spin columns in 2mL collection tubes and centrifuged at 6000 x g for 1 minute. Flow through was discarded, 500µl buffer AW1 was added and the tubes were centrifuged for 1 minute at 6000x g. This step was repeated but buffer AW2 was added in place of AW1 and centrifuged for 3 minutes at 6000 x g. Flow through was discarded and the columns were centrifuged for a further 1 minute to ensure the spin columns were dry. Spin columns were placed in new collection tubes and 200µl buffer AE was added to elute the DNA. These were left to stand at 5 minutes at room temperature and finally centrifuged at 8000 x g for 1 minute. The eluted solutions contained the purified DNA. DNA concentrations were measured using a NanoDrop Spectrophotometer and some of each sample was diluted to 25ng/µl using sterilised distilled water in a 100µl final volume as a working stock. All samples were stored at -20°C until further use.

### **2.4. Global methylation analysis in peripheral blood DNA**

Whole genome global DNA methylation levels were assessed in DNA derived from 185 PBMC samples from the UKAGS resource using colorimetric enzyme linked immunosorbent-assays (ELISAs) following the manufacturers' protocol (Epigentek - Methyl flash DNA quantification (colorimetric)). A maximum of 185 (93 AAA and 92 controls) samples were chosen for this assay based on how many 96-well plates were to be used in total (n=5). This number was based on assaying a sufficient sample size to yield accurate, reliable results, whilst also considering the limited funds available for the research. Hence, 5 separate assay plates were conducted, comfortably allowing for positive and negative control reactions, repeats of unsuccessful sample reactions and additional repeats to assess for potential batch effects.

The protocol adopted was as follows:

100ng (4µl working stock) of each DNA sample was bound to a well by adding it to 80µl binding solution and incubating at 37°C for 90 minutes in a polypropylene 96 well microplate. Solutions were removed and the wells were washed 3 times with 150µl wash buffer. Each well was then treated with 50µl (1:1000) anti-5 methyl cytosine primary antibody for 60 minutes at room temperature, binding all methylated genomic DNA of each sample to the foot of the well. The solutions were removed from the wells and the same washing process was performed as previously described. 50µl (1:2000) of a secondary antibody attached to an reporter enzyme was added and left to incubate for 30 minutes at room temperature for primary anti-body detection. The solutions were removed and the wash step was conducted 4 times before 100µl substrate solution was added for secondary antibody detection. The detection solution was left to incubate for 3 minutes in the dark before 100µl stop solution was added to halt the enzymatic reactions.

All reactions were performed in duplicate, and for each new 96 well plate (total = 5), new standardization controls were conducted. Mixed plate reactions were also performed to assess for batch effects. Absorbance (optical density (OD)) was measured with a microplate colorimetric spectrophotometer (Bio-Tek ELX808IU Ultra Microplate Reader), and ODs were converted to genome methylation percentage with the use of positive (5ng – 50% methylated DNA) and negative control (20ng– 50% un-methylated DNA) reactions, which were also conducted in duplicate. This standardisation method was used to acquire differences in relative levels of DNA methylation.

To convert ODs to global methylation percentage the following equation was used:

$$5\text{-mC \%} = \frac{(\text{Mean sample OD} - \text{Negative control OD}) \div 100\text{ng: input sample DNA}}{(\text{Positive control OD} - \text{Negative control OD}) \times 2^* \div 5\text{ng: input control DNA}} \times 100$$

Statistical analysis included column statistics in Graph Pad Prism 7 to conduct a one-way ANOVA and Tukey's B post-Hoc group comparison. Linear regression was then performed on all AAA cases to assess the association between aneurysm diameter and global methylation. The regression P value was further adjusted for differences in age between small AAA and large AAA groups.

### **2.5. Homocysteine analysis in blood plasma**

A total of 137 blood plasma samples from the global methylation assay were available (67 controls and 70 AAA) for circulating homocysteine analysis, which was performed with colorimetric ELISAs following the manufacturers' protocol (Cell Biolabs Inc – Homocysteine ELISA kit – STA-670). The total number of samples used for this experiment was based on available resources. Of the 185 DNA samples used in the global methylation assay, blood plasma was available for only 137.

The protocol for circulating blood plasma homocysteine uses the same concept as the global methylation assay (antibody binding specificity to quantify the presence of a substance), and was as follows:

100µl homocysteine conjugate was added to each well of a 96 well plate and left to incubate for 2 hours at 37°C. The solutions were removed and the plate was washed 3 times with 200µl PBS prior to adding 200µl assay diluent which was left at room temperature for 1 hour. The diluent was removed and 50µl plasma sample/control standard was added to the wells. The plates were incubated at room temperature for 10 minutes, and 50µl (1:500) homocysteine primary antibody was added to the wells and incubated for 1 hour at room temperature. The plate was washed 4 times with 200µl wash buffer and 100µl (1:1000) secondary antibody was added. After a 1 hour incubation at

room temperature the plate was washed in the same manner as the time before and 100µl substrate solution was added and incubated for 3 minutes at room temperature before 100µl stop solution was finally added.

All reactions were conducted in duplicate with new controls and mixed plate repeats for each new ELISA plate. ODs were measured with a micro-plate colorimetric spectrophotometer at 450nm and homocysteine levels were determined using the standard curve method.

Graphpad Prism 7 statistical analysis included unpaired t-test to assess differences between group means and linear regression to determine relationships between global methylation, homocysteine and aneurysm diameter.

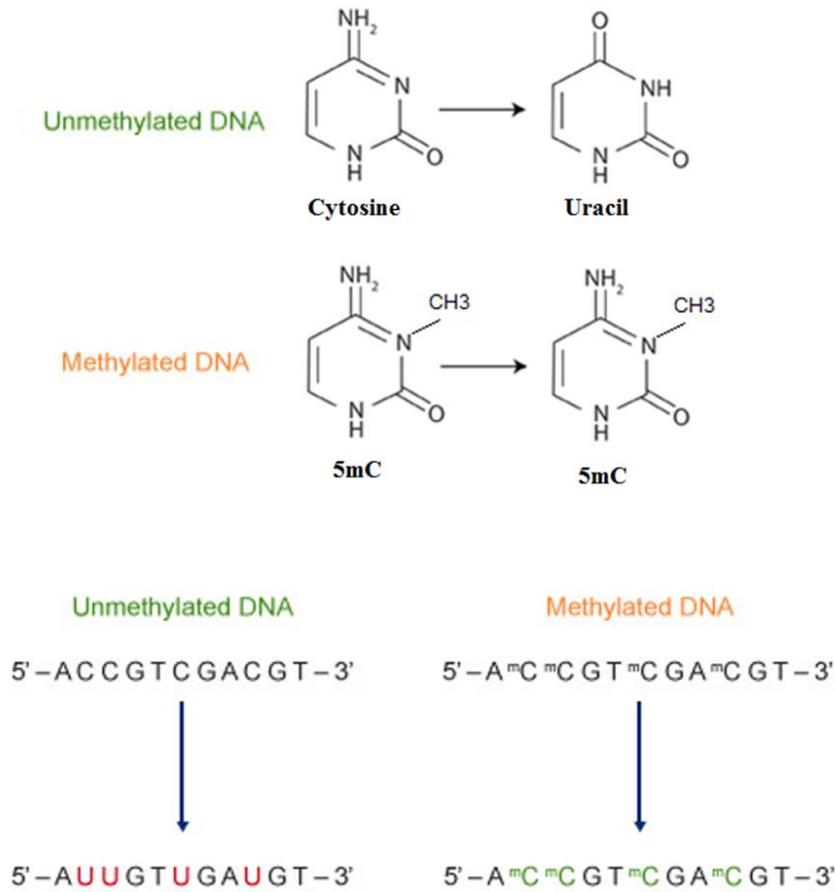
### **2.6. Bisulphite conversion of DNA for sequencing**

Each of the DNA samples required for PBMC (n=96) and VSMC (n=44) NGS sequencing during this PhD (Chapter 4 and 5 respectively) were bisulphite-converted using the MOD50 kit (Sigma–Aldrich) according to the manufacturer's standard protocol.

Bisulphite treatment of DNA deaminates un-methylated cytosines to uracil, but not methylated cytosines, allowing the distinction between methylated and un-methylated DNA after sequencing (Figure 8)<sup>177</sup>. This essentially means that un-methylated cytosines become uracil after bisulphite exposure, and methylated cytosines remain unchanged.

Considering uracil is unstable and is not a suitable base for DNA replication, uracil subsequently becomes thymine after PCR. Therefore, after bisulphite treatment, PCR, and sequencing, all known cytosine bases become thymine unless they are methylated (Figure 8). This methodology is regarded as the gold standard technique of DNA methylation

analysis, providing single CpG base resolution and higher accuracy in comparison to other methylation assays such as the Illumina 450k methylation micro-array<sup>178</sup>.



**Figure 8 - The effects of sodium bisulphite on DNA:** Bisulphite treatment of DNA deaminates un-methylated cytosines to uracil, whereas methylated cytosines stay unchanged. It is possible to distinguish between methylated and un-methylated CpG sites after sequencing, as all un-methylated CpGs become uracil and subsequently thymine.

The bisulphite assay protocol is as follows:

20 $\mu$ l (25ng/ $\mu$ l) of working stock DNA from each purified sample (500ng conversions) was added to 110 $\mu$ l DNA bisulphite modification solution and incubated at 99°C for 6 minutes, immediately followed by incubation at 65°C for 90 minutes. 300 $\mu$ l capture solution was added to spin columns placed in collection tubes and left to stand for 1 minute. The modified DNA solutions from the incubation step were then added to the capture solutions in the spin columns. These were centrifuged at 12,000 x g for 20 seconds and the flow through was discarded. 200 $\mu$ l cleaning solution was added to each spin column and centrifuged at 12,000 x g for 20 seconds. Flow through was discarded and 50 $\mu$ l of balance/ethanol wash solution was added before incubation for 8 minutes at room temperature. The tubes were centrifuged for 20 seconds at 12,000 x g and flow through was discarded. 200 $\mu$ l 90% ethanol was added to the spin columns which were centrifuged for 40 seconds at 12,000 x g, and the spin columns were placed in new 1.5mL collection tubes. Finally, 40 $\mu$ l elution solution was added to the spin columns, left to stand for 1 minute and centrifuged for 20 seconds at 12,000 x g to elute the bisulphite converted DNA at 10-15ng/ $\mu$ l. The converted DNA samples were used immediately or stored at -20°C for up to 1 month until further use.

## 2.8. Bisulphite specific PCR for sequencing

PCR is a common technique in molecular biology designed to amplify and isolate specific loci of interest within the genome. Thermo-cycling a genomic DNA template with polymerase enzymes, specific oligonucleotide primers and dNTPs in an optimised buffer allows amplification and enrichment of a desired product. Over many cycles, exponential amplification of a DNA sequence into the millions is observed<sup>179</sup>.

The technique was adopted in this research to target regulatory regions within candidate genes with bisulphite specific PCR primers after bisulphite treatment of DNA, as previously described. For each bisulphite treated DNA sample in Chapter 4 (48 AAA and 48 controls) and Chapter 5 (24 AAA and 20 controls), each targeted locus was isolated.

The optimised PCR protocol is as follows:

Each 20µl bisulphite specific PCR reaction consisted of 8µl sterilised distilled water, 10µl 2x jumpstart red-taq polymerase ready mix (SIGMA), 1µl mixed forward and reverse primers (10uM - SIGMA) and 1µl bisulphite converted DNA (10-15ng/ul). Each bisulphite specific primer pair was optimised based on annealing temperature. Specific targeted gene loci, primers and annealing temperatures are shown in the results chapters specific to each separate assay (Chapter 4 and 5). See Table 5 for thermo-cycling parameters.

Temperature	Time	Function	Cycles
94°C	2 minutes	Denature	1
94°C	15 seconds	Denature	40
52°C - 58°C	35 seconds	Anneal	40
72°C	35 seconds	Extension	40
72°C	5 minutes	Final extension	1

**Table 5 - Bisulphite specific PCR cycling conditions.**

## 2.9. Agarose gel electrophoresis

Agarose gel electrophoresis is primarily used to differentiate between varied sizes of amplified DNA fragment and to validate PCR products. Dissolved agarose moulded into a gel provides a structural foundation on which DNA can be electrophoresed, utilising the uniformed negative charge of the phosphate back bone of all DNA molecules<sup>180</sup>. When ran in a salt buffer, DNA loaded into an agarose gel migrate toward the positive charge. DNA migrates through the porous agarose gel based on size, allowing sufficient fractionation at specific resolutions. Smaller DNA fragments migrate further, and a higher percentage of agarose increases fractionation resolution<sup>180</sup>.

In this project, during the optimisation of primers, amplicons were checked by fractionation using agarose gel electrophoresis to validate PCR products and the protocol is as follows:

For each gel, 2g agarose powder (Melford, UK) was dissolved in a solution of 100ml 1X TAE (tris, acetate, EDTA) by microwaving for 100 seconds at full power (2% gel). After all agarose had dissolved, the solution was left to cool for 5 minutes before 10µl SYBR safe gel dye (Thermo-Fisher, UK) was added, mixed, and the whole solution was poured

into a gel cast. Gels were left to solidify and loaded with PCR amplicons. There was no need to add loading dye as the 2x jumpstart red-taq polymerase ready mix already contained the necessary dye. Gels were run in 900ml 1X TAE buffer at 120 volts for 50 minutes, using a Biometra powerpack (P25 standard). The gels were placed on a GeneFlash (Syngene, UK) UV-transilluminator and the resulting banding patterns of loaded samples were visualised.

### **2.10. Cleaning of PCR amplicons and pooling for sequencing library preparation**

All bisulphite specific PCR reactions were cleaned with ExoSAP-IT PCR Clean-up (Affymetrix). This was important because when using the PCR amplicons for further downstream application, which is in this case was adaptor ligation for Illumina sequencing library preparation, contaminating primers and dNTPs from the PCR reactions can adversely affect the process.

The clean-up protocol is as follows:

On ice, 5µl of the PCR reactions were mixed with 2µl ExoSAP-IT solution and incubated at 37°C for 15 minutes, then at 80°C for 15 minutes. These solutions were taken forward for pooling and the remaining uncleaned PCR reactions were stored at -20°C as back up samples. Individual PCR reactions from were pooled together, resulting in 96 individual samples that contained each genomic region of interest for the sequencing assay prior to library preparation.

## 2.11. Ligation of adaptors and barcoding for multiplex sequencing

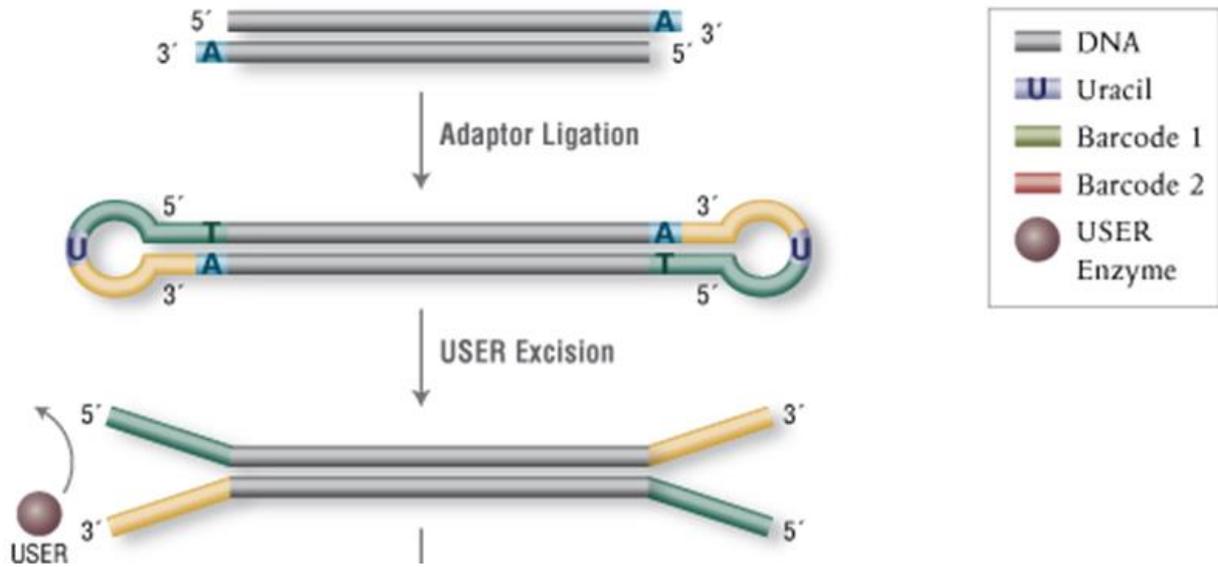
The NEBNext Ultra™ DNA Library Prep Kit for Illumina (E7370) and NEBNext Multiplex Oligos for Illumina dual index kit (E7600) were used for sequencing library preparation. The concept of the approach adopted in this project for both PBMC and VSMC sequencing assays was to conduct multiplex sequencing of all individual samples at each targeted genomic loci in one single sequencing reaction. This means that for each individual sample containing the pooled PCR amplicons described previously, ligation of adaptors and subsequent PCR enrichment with sequence specific primers was necessary to multiplex the samples. This is known as barcoding, and essentially allows the distinction of individual samples by the Illumina platform after they have been pooled in to one sequencing sample.

### 2.11.1. Adaptor ligation

The protocol for adaptor ligation of DNA samples was as follows:

Each of the pooled samples were concentrated to 1ug in a 55.5µl solution with sterilised distilled water and mixed in a 200µl PCR tube with 3µl end prep enzyme mix and 6.5µl 10x end repair reaction buffer (total volume 65ul). The samples were incubated at 20°C for 30 minutes followed by 65°C for 30 minutes and finally cooled to 4°C. 15µl blunt/TA ligase master mix, 2.5µl NEBNext adaptor for Illumina and 1µl ligation enhancer were subsequently added to the reactions and incubated for 15 minutes at 20°C. 3µl of USER enzyme was finally added to the reactions prior to clean up of adaptor ligated DNA without size selection (total volume 83.5ul).

Figure 9 shows a visual representation of adaptor ligation and why it is necessary.



**Figure 9 - Visual representation of adaptor ligation during sequencing library preparation:** Bisulphite specific PCR amplicons in each sample are exposed to conditions in which adaptor sequences are ligated to the 5' and 3' ends. The unstable base 'uracil' is cleaved from the adaptor loop by the 'USER' enzyme provided by the manufacturer. This results in both strands of the double stranded DNA amplicons having individual adaptors in preparation for the PCR enrichment (barcoding) phase of library preparation. Figure adapted from the NEB sequencing library preparation protocol.

### 2.11.2. Cleanup of adaptor ligations

The adaptor ligated reactions were then cleaned without using fragment size selection. Hence, fragmentation of DNA, which is commonly adopted in sequencing assays<sup>124, 181</sup>, was not necessary. It was my decision not to fragment DNA based on size selection because my amplicon sequences were not too large for the capacity of the sequencer. Avoiding DNA fragmentation additionally prevented the reduction in DNA yield and quality.

Amplicon cleaning without size selection was performed using the following protocol:

Clean up of adaptor ligated DNA was performed using Agencourt AMPure XP - PCR Purification (A63880). 86.5µl AMPure beads were suspended in the reactions (1:1 volume ratio), mixed at least 10 times and incubated for 5 minutes at room temperature. The tubes were briefly centrifuged and placed on a magnetic stand to separate the beads from the supernatant. The supernatant was removed and discarded, 180µl of 80% ethanol was added to the tubes containing the magnetic beads and left to stand for 30 seconds before the supernatant was removed and discarded. This step was repeated once more before the beads were left to air dry for 5 minutes in the tubes on the magnetic stand with the lids open. The tubes were removed from the magnetic stand and 17µl of 0.1x Tris-EDTA solution was added, mixed and left to incubate for 2 minutes at room temperature to elute the DNA. The tubes were placed back on the magnetic stand for 5 minutes and 15µl of the clear solutions were transferred to new PCR tubes for amplification.

**2.11.3. PCR enrichment of adaptor ligated DNA (barcoding)**

At this stage, all individual samples containing the isolated genomic loci of interest now had adaptor sequences attached to 5' and 3' ends of the double stranded bisulphite PCR amplicons. These adaptors are specific to the primers used in the next stage of the barcoding procedure, which adopts a dual index primer strategy.

Dual indexing is enabled by adding a unique index to both ends of a sample to be sequenced. Up to 96 different samples can be uniquely indexed by combining each of the 12 i7 primers with each of the 8 i5 primers.

Table 6 and 7 display the sequence specific primers used for barcoding of the individual samples.

NEB #	PRODUCT	INDEX PRIMER SEQUENCE	EXPECTED INDEX PRIMER SEQUENCE READ
#E7603A: 0.060 ml	NEBNext i501 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACTATAGCCTTACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	TATAGCCT
#E7604A: 0.060 ml	NEBNext i502 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACATAGAGGCACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	ATAGAGGC
#E7605A: 0.060 ml	NEBNext i503 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACCCTATCCTACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	CCTATCCT
#E7606A: 0.060 ml	NEBNext i504 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACGGCTCTGAACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	GGCTCTGA
#E7607A: 0.060 ml	NEBNext i505 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACAGGCGAAGACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	AGGCGAAG
#E7608A: 0.060 ml	NEBNext i506 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACTAATCTTAACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	TAATCTTA
#E7609A: 0.060 ml	NEBNext i507 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACCAGGACGTACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	CAGGACGT
#E7610A: 0.060 ml	NEBNext i508 Primer	5'-AATGATACGGCGACCACCGAGATC-TACACGTA CTGACACACTCTTTCCCTA-CACGACGCTCTCCGATC*T-3'	GTACTGAC

**Table 6 - i5 primer index sequences for multiplex sequencing:** Each individual DNA sample for sequencing is ligated with adaptor sequences to the 5' and 3' ends. After this, index primers are attached to the adaptor sequences which are unique and essentially allow identification of individual samples after sequencing. Table adapted from the NEB sequencing library preparation protocol.

NEB #	PRODUCT	INDEX PRIMER SEQUENCE	EXPECTED INDEX PRIMER SEQUENCE READ
#E7611A: 0.040 ml	NEBNext i701 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TCGAGTAATGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	ATTACTCG
#E7612A: 0.040 ml	NEBNext i702 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TTCTCCGGAGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	TCCGAGAA
#E7613A: 0.040 ml	NEBNext i703 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TAATGAGCGGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	CGCTCATT
#E7614A: 0.040 ml	NEBNext i704 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TGGAATCTCGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	GAGATTCC
#E7615A: 0.040 ml	NEBNext i705 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TTTCTGAATGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	ATTCAGAA
#E7616A: 0.040 ml	NEBNext i706 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TACGAATTCGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	GAATTCGT
#E7617A: 0.040 ml	NEBNext i707 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TAGCTTCAGGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	CTGAAGCT
#E7618A: 0.040 ml	NEBNext i708 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TGCGCATTAGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	TAATGCGC
#E7619A: 0.040 ml	NEBNext i709 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TCATAGCCGGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	CGGCTATG
#E7620A: 0.040 ml	NEBNext i710 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TTTCGGGAGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	TCCGCGAA
#E7621A: 0.040 ml	NEBNext i711 Primer	5'-CAAGCAGAAGACGGCATAACGAGAT-GCGCGAGAGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	TCTCGCGC
#E7622A: 0.040 ml	NEBNext i712 Primer	5'-CAAGCAGAAGACGGCATAACGAGA-TCTATCGCTGTGACTGGAGTTCAGAC-GTGTGCTCTCCGATC*T-3'	AGCGATAG

**Table 7 – i7 primer index sequences for multiplex sequencing:** Each individual DNA sample for sequencing is ligated with adaptor sequences to the 5' and 3' ends. After this, index primers are attached to the adaptor sequences which are unique and essentially allow identification of individual samples after sequencing. Table adapted from the NEB sequencing library preparation protocol.

A specific laboratory set up of these reactions is suggested by the manufacturer New England Bio labs, and is as follows:

Ensure that a valid combination of i7 and i5 primers is used.

Arrange (orange) i7 primers in increasing order horizontally, so that i701 is in column 1, i702 is in column 2, i703 is in column 3, etc.

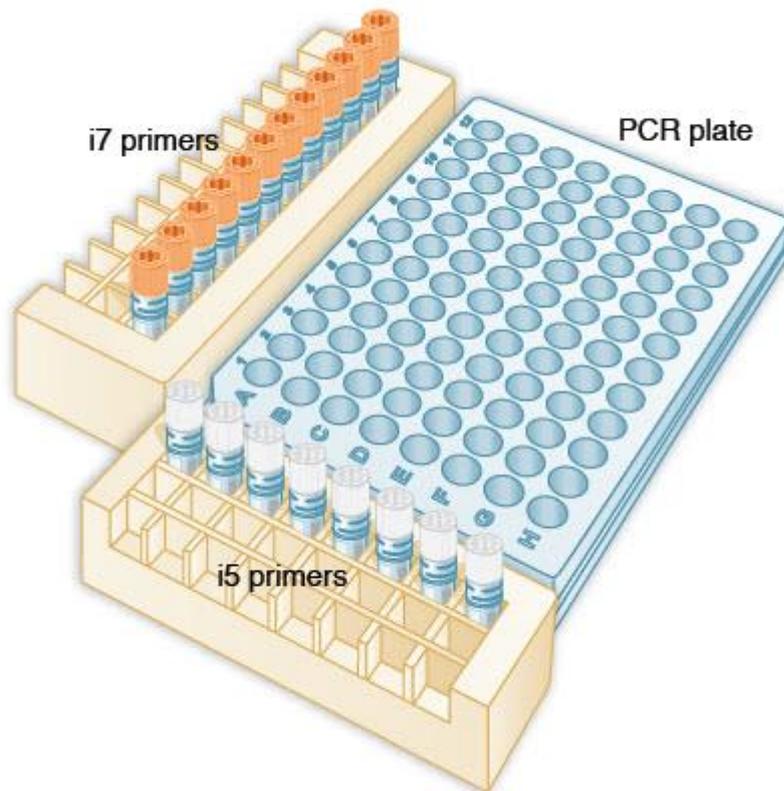
Arrange the (white) i5 primers in increasing order vertically, so that i501 is in row A, i502 is in row B, i503 is in row C, etc.

Record their positions on the PCR setup template.

Using a multichannel pipette, add desired volume of (white) i5 primers to every column of the PCR plate. It is critical to change tips between columns to avoid cross contamination.

Using a multichannel pipette, add desired volume of (orange) i7 primers to every row of the PCR plate. It is critical to change tips between rows to avoid cross-contamination.

Figure 10 visually displays the recommended set up of the PCR enrichment reactions.



**Figure 10 - Suggested set up of dual index strategy to ‘barcode’ DNA samples:** During PCR enrichment of adaptor ligated DNA fragments in the sequencing library, only a unique combination of i7 and i5 index primers should be used for each DNA sample for identification after sequencing. Figure adapted from the NEB sequencing library preparation protocol.

Following these guidelines, Figure 11 and 12 show how the samples were set up when conducted for this project on PBMC and VSMC DNA for bisulphite sequencing assays (Chapter 4 and 5 respectively).

Primer		i701	i702	i703	i704	i705	i706	i707	i708	i709	i710	i711	i712
	position	1	2	3	4	5	6	7	8	9	10	11	12
i508	A	<b>8c</b>	<b>16C</b>	<b>24C</b>	<b>32c</b>	<b>41c</b>	<b>49c</b>	<b>56</b>	<b>69</b>	<b>78</b>	<b>86</b>	<b>94</b>	<b>102</b>
		i508/i701	i508/i702	i508/i703	i508/i704	i508/i705	i508/i706	i508/i707	i508/i708	i508/i709	i508/i710	i508/i711	i508/i712
i507	B	<b>7c</b>	<b>15C</b>	<b>23C</b>	<b>31c</b>	<b>40c</b>	<b>48c</b>	<b>55</b>	<b>68</b>	<b>77</b>	<b>85</b>	<b>93</b>	<b>101</b>
		i507/i701	i507/i702	i507/i703	i507/i704	i507/i705	i507/i706	i507/i707	i507/i708	i507/i709	i507/i710	i507/i711	i507/i712
i506	C	<b>6c</b>	<b>14C</b>	<b>22C</b>	<b>30c</b>	<b>39c</b>	<b>47c</b>	<b>54</b>	<b>67</b>	<b>76</b>	<b>84</b>	<b>92</b>	<b>100</b>
		i506/i701	i506/i702	i506/i703	i506/i704	i506/i705	i506/i706	i506/i707	i506/i708	i506/i709	i506/i710	i506/i711	i506/i712
i505	D	<b>5c</b>	<b>13C</b>	<b>21C</b>	<b>29c</b>	<b>38c</b>	<b>46c</b>	<b>53</b>	<b>65</b>	<b>75</b>	<b>83</b>	<b>91</b>	<b>99</b>
		i505/i701	i505/i702	i505/i703	i505/i704	i505/i705	i505/i706	i505/i707	i505/i708	i505/i709	i505/i710	i505/i711	i505/i712
i504	E	<b>4c</b>	<b>12C</b>	<b>20C</b>	<b>28c</b>	<b>37c</b>	<b>45c</b>	<b>52</b>	<b>64</b>	<b>74</b>	<b>82</b>	<b>90</b>	<b>98</b>
		i504/i701	i504/i702	i504/i703	i504/i704	i504/i705	i504/i706	i504/i707	i504/i708	i504/i709	i504/i710	i504/i711	i504/i712
i503	F	<b>3c</b>	<b>11C</b>	<b>19C</b>	<b>27c</b>	<b>36</b>	<b>44c</b>	<b>51</b>	<b>61</b>	<b>73</b>	<b>81</b>	<b>89</b>	<b>97</b>
		i503/i701	i503/i702	i503/i703	i503/i704	i503/i705	i503/i706	i503/i707	i503/i708	i503/i709	i503/i710	i503/i711	i503/i712
i502	G	<b>2c</b>	<b>10C</b>	<b>18C</b>	<b>26c</b>	<b>35c</b>	<b>43c</b>	<b>50</b>	<b>59</b>	<b>71</b>	<b>80</b>	<b>88</b>	<b>96</b>
		i502/i701	i502/i702	i502/i703	i502/i704	i502/i705	i502/i706	i502/i707	i502/i708	i502/i709	i502/i710	i502/i711	i502/i712
i501	H	<b>1c</b>	<b>9C</b>	<b>17C</b>	<b>25c</b>	<b>34c</b>	<b>42c</b>	<b>49</b>	<b>58</b>	<b>70</b>	<b>79</b>	<b>87</b>	<b>95</b>
		i501/i701	i501/i702	i501/i703	i501/i704	i501/i705	i501/i706	i501/i707	i501/i708	i501/i709	i501/i710	i501/i711	i501/i712

**Figure 11 - Plate layout for barcoding of PBMC samples for sequencing:** The first sequencing assay in this PhD was conducted on 96 peripheral blood DNA samples (48 AAA and 48 controls). The bar coding/adaptor ligation for sample identification purposes can be observed.

Primer		i701		i702		i703		i704		i705		i706
	position	1		2		3		4		5		6
i508	A	<b>8c</b>		<b>16C</b>		<b>4</b>		<b>12</b>		<b>20</b>		
		i508/i701		i508/i702		i508/i703		i508/i704		i508/i705		
i507	B	<b>7c</b>		<b>15C</b>		<b>3</b>		<b>11</b>		<b>19</b>		
		i507/i701		i507/i702		i507/i703		i507/i704		i507/i705		
i506	C	<b>6c</b>		<b>14C</b>		<b>2</b>		<b>10</b>		<b>18</b>		
		i506/i701		i506/i702		i506/i703		i506/i704		i506/i705		
i505	D	<b>5c</b>		<b>13C</b>		<b>1</b>		<b>9</b>		<b>17</b>		
		i505/i701		i505/i702		i505/i703		i505/i704		i505/i705		
i504	E	<b>4c</b>		<b>12C</b>		<b>20C</b>		<b>8</b>		<b>16</b>		<b>24</b>
		i504/i701		i504/i702		i504/i703		i504/i704		i504/i705		i504/i706
i503	F	<b>3c</b>		<b>11C</b>		<b>19C</b>		<b>7</b>		<b>15</b>		<b>23</b>
		i503/i701		i503/i702		i503/i703		i503/i704		i503/i705		i503/i706
i502	G	<b>2c</b>		<b>10C</b>		<b>18C</b>		<b>6</b>		<b>14</b>		<b>22</b>
		i502/i701		i502/i702		i502/i703		i502/i704		i502/i705		i502/i706
i501	H	<b>1c</b>		<b>9C</b>		<b>17C</b>		<b>5</b>		<b>13</b>		<b>21</b>
		i501/i701		i501/i702		i501/i703		i501/i704		i501/i705		i501/i706

**Figure 12 - Plate layout for barcoding of VSMC samples for sequencing:** The second sequencing assay in this PhD was conducted on 44 peripheral blood DNA samples (24 AAA and 20 controls). The bar coding/adaptor ligation for sample identification purposes can be observed.

The components for PCR enrichment of adaptor ligated DNA are as follows:

25µl NEBNext Q5 hot start PCR master mix, 5µl i5 Multiplex primers and 5µl i7

Multiplex primers were added to the 15µl adaptor ligated DNA solutions to make a total volume of 50ul. Combinations of i5 and i7 primers were specific to each sample (total of 96 combinations) for identification after sequencing.

Cycling conditions for PCR enrichment of adaptor ligated DNA are shown in Table 8.

<b>Temperature</b>	<b>Time</b>	<b>Function</b>	<b>Cycles</b>
98°C	30 seconds	Denature	1
98°C	10 seconds	Denature	4
65°C	75 seconds	Anneal/extension	4
65°C	5 minutes	Final extension	1
4°C	hold		

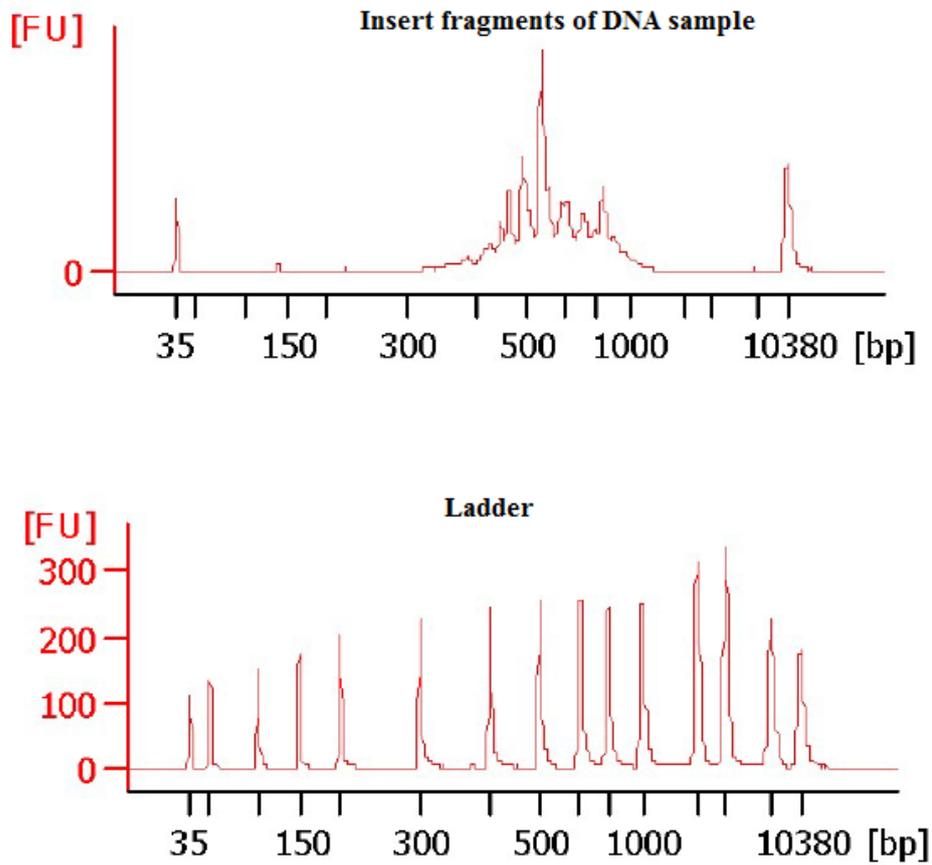
**Table 8 – Cycling conditions for PCR enrichment of adaptor ligated DNA.**

Finally, cleanup of the bar coded PCR amplicons was necessary, which was performed using the Agencourt AMPure XP - PCR Purification method as previously described using a 1:1 volume ratio of beads.

### 2.12. Agilent Bioanalyzer analysis for sequencing

Each of the adaptor ligated, bar coded DNA samples were ran on an Agilent 2100 Bioanalyzer Instrument<sup>182</sup> (expert High Sensitivity DNA chip) to assess sample DNA concentration and to check the presence and distribution of correct fragment sizes following standard protocol (Agilent High Sensitivity DNA Assay Protocol). With automated electrophoresis, the Agilent 2100 Bioanalyzer system provides sizing, quantitation, and purity assessments for DNA, RNA, and protein samples<sup>182</sup>.

The Bioanalyzer works the same as other electrophoretic equipment, such as gel tanks, by utilising the electrophoretic potential of DNA molecules, which are negatively charged. Therefore, when a sample for sequencing is loaded on to the same chip as a DNA ladder containing varied DNA fragment sizes, it is possible to gauge DNA fragment size of the inserts (Figure 13). In addition, the bioanalyzer accurately assesses concentrations of DNA fragments.



**Figure 13 – Data output from 2100 Agilent Bioanalyzer:** Figure displays an example of a single pooled DNA sample (above) that has been ran on a bioanalyzer chip and contains a variety of fragment insert sizes. The sample is compared against a DNA ladder (below) of standard DNA fragment sizes. When comparing the fluorescent signals of the samples and the ladder, it is generally easy to acquire fragment sizes and concentrations of samples.

The protocol for the Agilent Bioanalyzer is as follows:

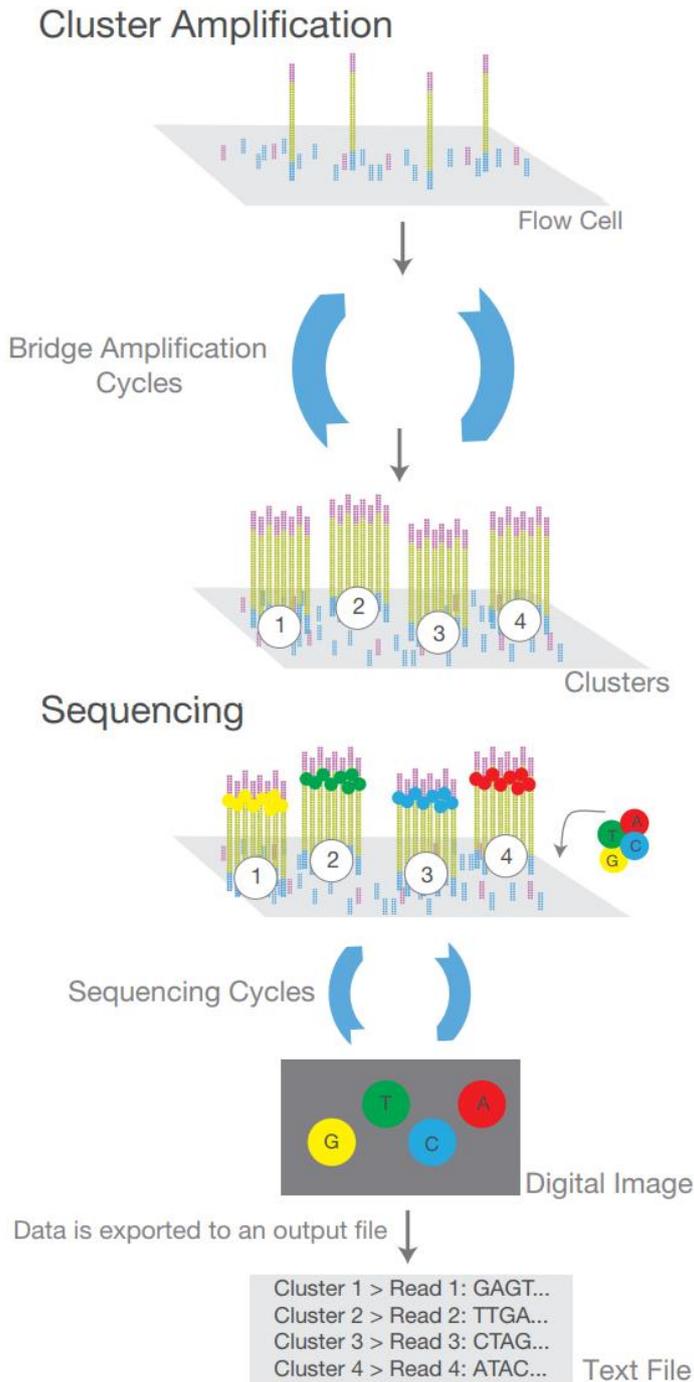
For each chip used (capacity for 11 DNA samples), 9.0 $\mu$ l of gel-dye mix was added at the bottom of the well (marked 'G') and dispersed with a syringe plunger. 9 $\mu$ l of gel-dye mix was then added to each of the other wells marked 'G'. 5  $\mu$ L of green-capped High Sensitivity DNA marker was added into the well-marked with the ladder symbol and into each of the 11 sample wells. 1  $\mu$ l of the High Sensitivity DNA ladder vial was added in the well-marked with the ladder symbol, and in each of the 11 sample wells 1  $\mu$ L of sample was added. The chips were vortexed for 60 seconds at 2400 rpm and placed in the Bioanalyzer. The select-region function was used to determine the concentration of each sample spanning the range of insert fragment sizes.

The insert fragment sizes and estimated concentrations of samples given by the Bioanalyzer were used to pool the individual samples prior to sequencing at 4nM.



Illumina NGS, which was adopted in this PhD project, operates similarly to Sanger sequencing, as it also uses the sequencing by synthesis method. However, this is conducted on a massively parallel scale, yielding substantially more throughput than Sanger sequencing. As a result of this technology, millions of sequencing reads can be obtained from a single run, capable of assessing the entire genome, or, more applicable to this project, enabling deep sequencing of many different targeted loci in a number of multiplexed samples, all in a single sequencing run.

As previously described, NGS samples for sequencing are prepared by ligating specialised adaptors to the ends of the 5' and 3' ends of the DNA fragments. The library prepared DNA is loaded on to the flow cell of the sequencer, where the reaction takes place. Each adaptor ligated DNA fragment hybridises to the surface of the flow cell, and each bound fragment is amplified (using sequencing by synthesis) in to a clonal cluster. There are millions of clonal cluster reactions on a single flow cell, which allows for such high throughput. On very high performance machines (Illumina Hi-seq platform), several flow cells can be used in each sequencing run, demonstrating just how much high throughput potential the technology has. Figure 15 provides a visual demonstration of how Illumina NGS works.



**Figure 15 - Summary of how Illumina NGS works:** The sequencing library is loaded in to the flow cell and the fragments hybridise to the surface. Each bound fragment is amplified in to a cluster. During each cycle, amplification reagents, including fluorescently labelled nucleotides, are added. Each incorporated base is read by the sequencer, which is based on the specific wavelength emitted by each different base. Sequencing runs can consist of up to ~600 cycles if using paired-end sequencing. Figure adapted from Illumina: [https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina\\_sequencing\\_introduction.pdf](https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf)

This methodology is regarded as the gold standard technique of DNA methylation analysis, providing single CpG base resolution and higher accuracy in comparison to other methylation assays such as the Illumina 450k methylation micro-array<sup>178</sup>. Using this methodology therefore has the potential to yield very high impact research, and was by far the methodology of choice in this PhD project.

However sequencing can be expensive, and compromises were made based on resources, funding, and time available to conduct and finish the work during the timeframe of the PhD. This is primarily why targeted sequencing was chosen over whole genome sequencing. As a result of the specific experimental approach required for targeted bisulphite sequencing of candidate genes, the Illumina Mi-seq was chosen as the sequencing platform. The Mi-seq is ideal for targeted sequencing, whereas more high performance machines such as the Hi-seq (most commonly used for whole genome sequencing), would be more costly and would yield more data than necessary.

A paired-end multiplexed sequencing (2x310bp) approach was adopted, meaning that the DNA fragments were sequenced on both strands. When thinking of DNA as a single linear fragment, this essentially means that each fragment was read from the beginning towards the end, and from the end towards the beginning, for 310 base pairs each way. Therefore the total capacity for the sequencing runs were a total of around 620bp fragments.

The sequencing reactions were conducted following optimised protocols for the Mi-seq platform, provided by Illumina. The specific sequencing kit used was the MiSeq Reagent Kit v3: Firstly, the pooled multiplex sequencing library (5µl of 4nM library) was denatured by mixing with 5µl of 0.2 N NaOH. This solution was vortexed and centrifuged at 240 x g for 1 minute and incubated at room temperature for 5 minutes.

Then, 990µl pre-chilled HT1 was added to the tube containing the denatured library and mixed. This subsequent denatured 20pM library was further diluted to a concentration of 15pM by mixing 450µl 10pM library and 150µl pre-chilled HT1. This solution was left on ice while a PhiX internal sequencing control was created. 10nM PhiX library (2µl) was mixed with 10mM Tros-Cl, PH 805 with 0.1% Tween 20 (3µl) and denatured by combining with 5µl 0.2 N NaOH. The 10µl solution was vortexed, centrifuging for 1 minute at 280 x g and left to incubate for 5 minutes at room temperature. The denatured PhiX library was added to 990µl pre-chilled HT1 to dilute to 20pM. 570µl of the original multiplexed denatured sample library was mixed with 30µl of the PhiX control library before loading into the reagent cartridge which had been thawing at room temperature for 90 minutes. The reagent cartridge was then inserted into the sequencer through the reagent chiller door. The on screen instructions were then followed based on a paired end 310 cycle run.

Two separate library preparations and sequencing assays were conducted in this PhD; one assay to sequence DNA from PBMCs and the other to sequence DNA from VSMCs.

Figure 16 is a photograph of the Mi-seq used for this research, which is based in the Department of Genetics, University of Leicester.



**Figure 16 – Photograph of Illumina mi-seq:** Platform used to conduct targeted bisulphite sequencing of multiplexed samples.

## 2.14. Bioinformatics analysis of sequencing data

The data output from a NGS run is very extensive, and consists of millions of individual sequencing reads for each individual sample. These reads are automatically dedicated to individual sample files by the sequencing platform (fast-q files), based on the index sequences which are attached during sequencing library preparation. The data load of these files is too large to manually assess, which is why bioinformatics software has been developed over the years since the advent of NGS. The relevant software necessary for this analysis is primarily run on the command line due to the processing power that is required. For this project I performed the bioinformatics analysis of sequencing data on the ALICE High Performance Computing Facility at the University of Leicester.

The work flow is as follows:

A provisional data quality filter was applied with a programme called Trimmomatic to remove all low quality raw sequencing reads<sup>184</sup>. This process removed contaminating Illumina adaptor sequences from raw sequencing reads, and trimmed the ends of all sequencing reads at a Phred quality score of 15 (Q15). Q scores are standard measures of quality produced by Illumina sequencers, which essentially highlights the probability of the base being called correctly. Q15 is equivalent to an accuracy of 95%. The same software was additionally used for the sliding window function, which clips and discards sequencing data in each read at Q15, 4 bases at a time.

The specific commands for the trimmomatic functions were:

Remove adaptors:

```
trimmomatic PE -phred33 -threads 2 -trimlog logfile RAW-SAMPLE-1.fastq RAW-  
SAMPLE-2.fastq left-paired.fastq left-unpaired.fastq right-paired.fastq right-  
unpaired.fastq ILLUMINACLIP:~/TruSeq3-PE-2.fa:2:40:15 MINLEN:36
```

Trim low quality:

```
trimmomatic PE -phred33 -threads 2 -trimlog logfile2 left-paired.fastq right-paired.fastq  
80-1-final.fastq left-trim-unpaired.fastq 80-2-final.fastq right-trim-unpaired.fastq  
LEADING:15 TRAILING:15 SLIDINGWINDOW:4:15 MINLEN:36
```

Reference alignment of the quality filtered sequencing reads was then conducted for each individual sample using software called BWA-meth: <https://github.com/brentp/bwa-meth>.

The reference sequences used in this project were compiled by me, and contained only the targeted gene regions of interest to this study. The specific candidate gene sequences for Chapter 4 and 5 (GRCh38) were compiled in to separate files for use as reference sequences, which can be viewed in Appendix I and II. BWA-meth is specifically designed to align bisulphite sequencing data by converting the reference sequence similarly to how bisulphite treatment of DNA would. This subsequently produces sequencing read alignments to the reference sequences that are more suitable for methylation data than a standard genome aligner, and the alignments are therefore more accurate.

The specific command used for BWA-meth reference alignment is:

```
python /alice/scratch/aortaneu/bt96/bt96/bwa-meth-0.10/bwameth.py --reference ~/REF-  
SEQ.fa SAMPLE-PAIR1.fastq SAMPLE-PAIR2.fastq --prefix OUTPUT
```

PCR duplicates for all BAM files were marked and excluded with the Picard utility

MarkDuplicates: <http://broadinstitute.github.io/picard/> and the final processed files were

sorted and indexed with Samtools<sup>185</sup> using the following commands:

```
java -jar /cm/shared/apps/picard/1.93/MarkDuplicates.jar INPUT=SAMPLE.bam
```

```
OUTPUT=SAMPLE.bam METRICS_FILE=dupl_metrics.txt
```

```
samtools sort SAMPLE.bam OUTPUT-SAMPLE
```

```
samtools index OUTPUT-SAMPLE.bam
```

Finally, for each sample, the methylation values (normalised as methylation %) of each sequenced CpG in each gene were extracted to a bedGraph file with high quality filter thresholds using PileOMeth: <https://github.com/dpryan79/PileOMeth> software.

Read alignment quality was filtered at an accuracy of Q50 (99.999%).

Base quality was filtered at an accuracy of Q20 (99%).

A minimum read depth of 5 was also applied.

Sequencing data that did not meet these criteria were excluded from the study, rendering the final methylation calls incredibly accurate.

The specific command for methylation calling was:

```
/alice/scratch/aortaneu/bt96/bt96/PileOMeth/PileOMeth extract --mergeContext --
```

```
minDepth 5 --keepSingleton --keepDiscordant -q 50 -p 20 ~/REF-SEQ.fa SAMPLE.bam
```

The idea behind bioinformatics programming is that whatever you tell the computer to do, it will do, providing you know how to ask it. Therefore it is very important that these commands are produced perfectly, considering a single character change or phrases in the wrong order will result in a failed command, and there will be no output.

A summary of the commands used for bioinformatics analysis of the NGS data acquired in this project are displayed in Figure 17, whilst a flow chart summary of bioinformatics work flow is shown in Figure 18.

```
REMOVE ADAPTORS
trimmomatic PE -phred33 -threads 2 -trimlog logfile PAIR1.fastq PAIR2.fastq left-paired.fastq left-
unpaired.fastq right-paired.fastq right-unpaired.fastq ILLUMINACLIP:~/TruSeq3-PE-2.fa:2:40:15
MINLEN:36

TRIM LOW QUALITY:
trimmomatic PE -phred33 -threads 2 -trimlog logfile2 left-paired.fastq right-paired.fastq final-1.fastq
left-trim-unpaired.fastq final-2.fastq right-trim-unpaired.fastq LEADING:15 TRAILING:15
SLIDINGWINDOW:4:15 MINLEN:36

REFERENCE ALIGNMENT:
python /alice/scratch/aortaneu/bt96/bt96/bwa-meth-0.10/bwameth.py --reference ~/ngsref2.fa final-
1.fastq final-2.fastq --prefix SAMPLE

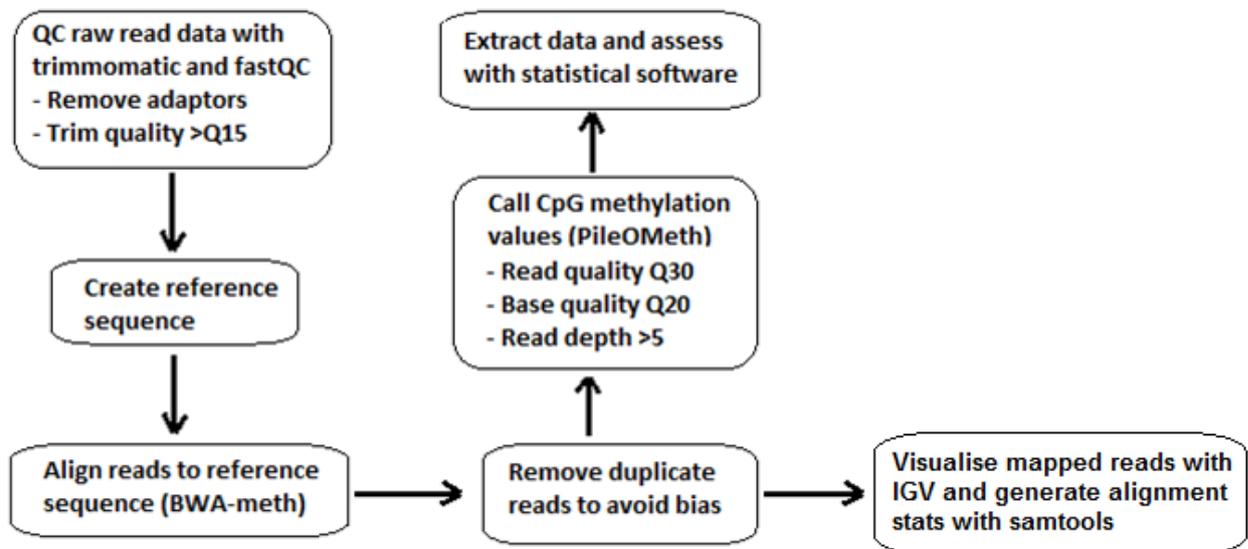
REMOVE DUPLICATES:
java -jar /cm/shared/apps/picard/1.93/MarkDuplicates.jar INPUT=SAMPLE.bam OUTPUT=sample-
dup.bam METRICS_FILE=dupl_metrics.txt

samtools sort sample-dup.bam sample-final

samtools index sample-final.bam

CALL VARIANTS
/alice/scratch/aortaneu/bt96/bt96/PileOMeth/PileOMeth extract --mergeContext --minDepth 5 --
keepSingleton --keepDiscordant -q 50 -p 20 ~/ngsref2.fa sample-final.bam
```

**Figure 17 - Bioinformatics command list:** The specific commands which were used during the bioinformatics analysis of next-generation sequencing data during this project are listed. The parts highlighted in red represent file names, which were variable for each different sample file.



**Figure 18 - Summary of bioinformatics workflow and data analysis:** Each stage of the analysis for next-generation sequencing data is presented in a flow chart with the specific filter criteria that was applied.

Bioinformatics analysis of sequencing data is commonly performed without visualisation, considering the process results in vast amounts of genetic data which has been quantified to numerical data. However, there is a commonly used software called Interactive Genomics Viewer (IGV), which helps provide better understanding of the bioinformatics process by allowing the visualisation of the sequencing alignments to the reference sequence. Figure 19 displays sequencing read alignments from this project (Chapter 4) in the *LDLR* gene across the entire sequenced promoter region. Figure 20 then displays the same region, which has been zoomed in so that the individual primary base sequence can be seen. In these examples, all sequenced cytosines have been converted to thymines, except for those which are methylated (in blue). This is how methylation percentage is determined, the proportion of methylated vs un-methylated sequencing reads at each individual CpG site.

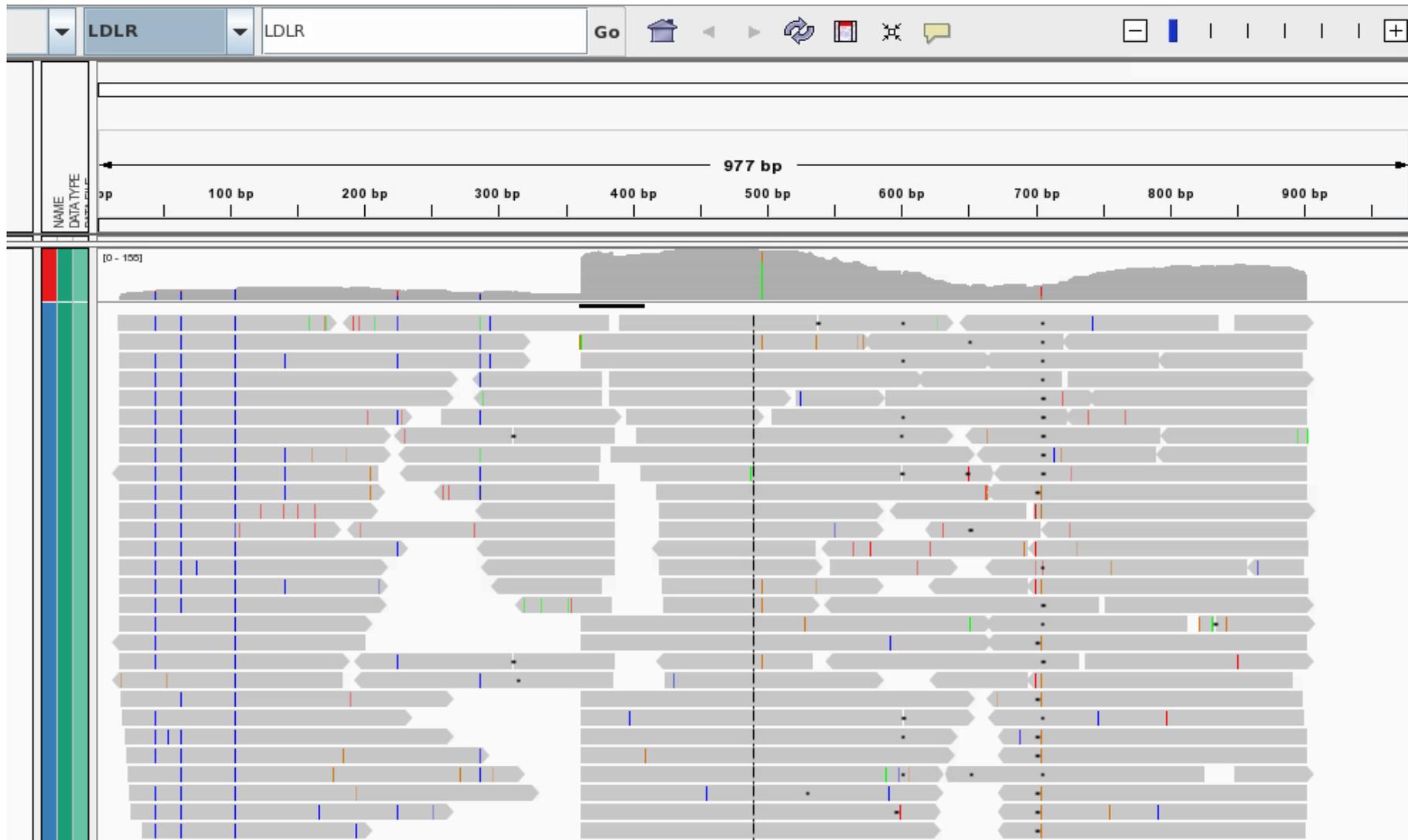


Figure 19 - Screenshot from IGV software used to visually map high quality sequencing reads to the reference sequence for the *LDLR* gene promoter.



Figure 20 - A higher resolution screen shot displaying individual bases from mapped sequencing reads to the reference sequence in the *LDLR* gene promoter.

### 2.14.1. Statistical analysis of sequencing data

The methylation data extracted using PileOMeth was taken forward for statistical analysis using IBM SPSS Statistics 24 and GraphPad prism 7.

To correct for non-normal distribution in the acquired data, each methylation value at CpG sites in each individual were ranked with inverse normal transformations (Blom normal scores in SPSS). Many complex traits studied in genetics, including in this project, have non-normal distributions, and inverse normal transformations have gained popularity in genetic analysis and are implemented as an option to correct non-normal distribution<sup>186</sup>. By transforming all data in to ranks, subsequent analysis can be conducted under the parametric assumption of even distribution of data.

Unpaired multiple t-tests were then conducted on the transformed data (ranked normal scores) at each sequenced CpG site between cases vs controls for each gene separately. For multiple comparison testing the false discovery rate approach was adopted. Discovery was determined using the two-stage linear step-up procedure of Benjamini, Krieger and Yekutieli, with  $Q = 10\%$ . CpG sites with significant P values and Q values were then manually checked to ensure only sites with a noticeably visual difference in methylation between cases and controls were included to investigate further. Logistic regression was then performed using SPSS at each significant CpG site to adjust for age differences between cases vs controls.

### **2.15. Summary of NGS workflow**

Several different methodologies have been discussed and described as they were carried out during this PhD project. A visual, simplified sequencing work flow is shown in Figure 21 for greater understanding of the bigger picture of the methodologies that have been described in depth.

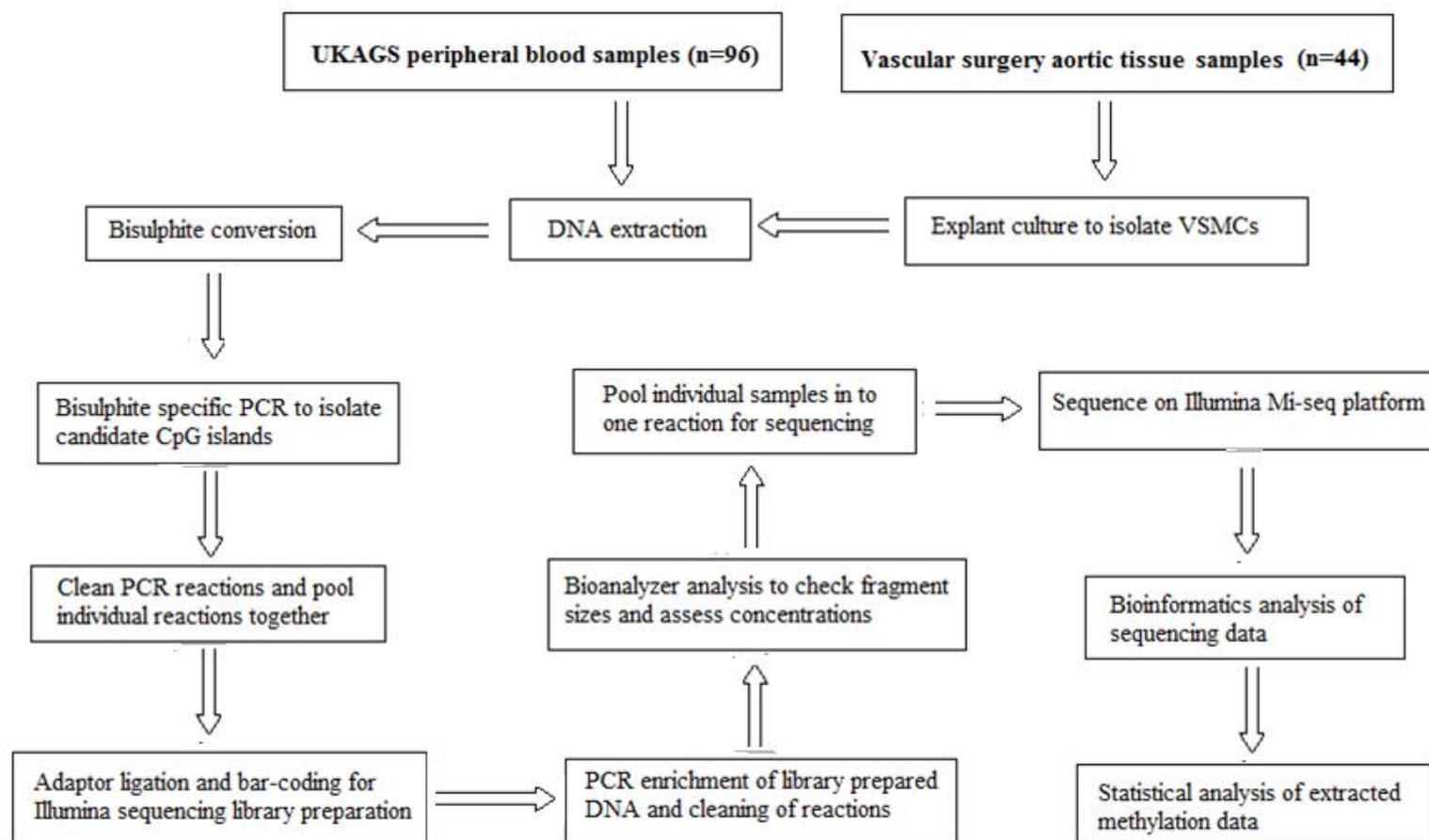


Figure 21 – Summary flow-chart of all the methodologies adopted to complete the next-generation sequencing experiments.

## **2.16. Blood plasma protein analysis (ELISAs) for functional corroboration of PBMC differential methylation**

MRNA gene expression analysis is the conventional way to study the relationship between differential methylation and gene function. However, RNA could not be isolated from the PBMC samples to measure gene expression of the differentially methylated genes from Chapter 4. As a result, functional corroboration was studied with the conduct of ELISAs to assess the blood plasma concentrations of proteins related to the genes that were differentially methylated in Chapter 4 (LDLR, SORT1, LDL and IL6R).

Each assay was purchased from the manufacturer 'Elabscience', and the manufacturer's standard protocol was adopted for each assay.

This analysis was performed on 20 AAA and 20 control blood plasma samples for which there was methylation data. For each assay, blood plasma from each sample was diluted 100x with sterilised distilled water for the effective assay detection range. 100µl diluted sample, or standard, was added to each well and incubated for 90 minutes at 37°C. The liquid was removed and without washing, 100µl (1:100) biotinylated detection (primary) antibody was added and left to incubate for 1 hour at 37°C. The liquid was removed and the plates were washed with wash buffer 3 times before 100µl (1:100) HRP conjugate secondary antibody was added. The plates were incubated for 30 minutes at 37°C prior to 5 plate washes with wash buffer, and the addition of the substrate reagent, which was left to incubate for 5 minutes at 37°C. Stop solution was finally added, and the reactions were measured with a micro-plate colorimetric spectrophotometer at 450nm. All reactions were conducted in duplicate, and circulating blood plasma concentrations of LDLR, SORT1, LDL and IL6R were determined using the standard curve method.

Graphpad Prism 7 statistical analysis included unpaired t-test to assess differences between group means and linear regression to determine relationships between CpG methylation and plasma concentrations.

### **2.17. RNA extraction for gene expression analysis in VSMCs**

RNA isolation was conducted in this study for gene expression analysis, and was performed following the manufacturers protocol (QIAGEN RNeasy Mini Kit).

Differential methylation of CpG sites in gene promoters, as previously discussed, can alter gene functionality by influencing gene expression. Therefore, gene expression analysis was necessary to corroborate the VSMC methylation differences observed from Chapter 5 (*ERG*, *IL6R*, *SMYD2* and *SERPINB9*).

The protocol was as follows:

Isolated cultured VSMCs (n=44) were removed from -80°C storage (from explant culture), immediately treated with 350µl RLT buffer and mixed vigorously before 350µl 70% ethanol was added. Thawing of cells is avoided as much as possible to prevent RNA degradation. The lysates were vortexed before being transferred to spin columns placed in collection tubes. The tubes were centrifuged for 15 seconds at 8,000 x g, the flow through was discarded, and 700µl RW1 buffer was added. The same centrifugation step was performed again, and 500µl RPE buffer was added. The tubes were centrifuged again, and the columns were placed in new collection tubes. 30µl RNase free water was added to the spin columns, left to stand for 5 minutes, and centrifuged at 8000 x g for 1 minute to elute the RNA.

### **2.18. Genomic DNA digestion to remove contaminating DNA**

The purification of total RNA does not guarantee 100% removal of genomic DNA, so it was necessary to digest any contaminating DNA in the RNA samples before synthesising cDNA, which if not removed, results in false positive and inconsistent results when conducting downstream analysis. DNase I digests double and single stranded DNA into oligo and mononucleotides.

Genomic DNA digestion in the 44 RNA samples was performed following the manufacturers' standard protocol (DNase I - RNase-Free: New England BioLabs M0303). The 30 $\mu$ l RNA samples were mixed with 1 $\mu$ l (2 units/ $\mu$ l) DNase, 5 $\mu$ l 10x DNase buffer solution and 14 $\mu$ l RNase free water (50 $\mu$ l final volume). The solutions were incubated at 37°C for 10 minutes, and 1 $\mu$ l 0.5M EDTA was added before a final incubation of 75°C for 10 minutes.

### **2.19. CDNA synthesis for qPCR**

Finally, the RNA was synthesised to cDNA (more stable for analysis than RNA). CDNA was synthesised from the DNase digested RNA using the ThermoFisher High-Capacity cDNA Reverse Transcription Kit (4368814) according to the manufacturers protocol.

Reactions consisted of 2 $\mu$ l 10x buffer, 0.8 $\mu$ l 25x dntp mix, 2 $\mu$ l 10x random primer mix, 1 $\mu$ l reverse transcriptase, 4.2  $\mu$ l RNase free water and 10 $\mu$ l DNase treated RNA (final volume 20 $\mu$ l). The cycle conditions for cDNA synthesis were: 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes and 4°C hold. Converted RNA (cDNA) was stored at -20°C until required for gene expression analysis.

## 2.20. TaqMan gene expression analysis (qPCR) for functional corroboration of VSMC differential methylation

Real time qPCR is a high throughput methodology which can assess gene expression levels in cDNA derived from purified mRNA. The technique is also commonly used for genotyping target DNA at variant loci<sup>187, 188</sup>, and involves amplification and target detection in a single step, allowing quantitative data collection throughout the reaction.

The process involves using fluorescent chemistry which correlates with the exponential product of the PCR reaction<sup>188</sup>. This is determined by the cycle threshold (point in the reaction where linear exponential amplification of the target sequence is first detected).

Resultantly, the larger the amount of starting product in the reaction, the lesser amount of cycles it takes to reach this Ct value<sup>188</sup>. By using a housekeeper gene (expressed in all cells of an organism under normal and patho-physiological conditions) as an internal reference for normalisation, the relative differences in expression of genes from diseased and healthy samples can be established.

For each gene where differential methylation was observed after VSMC bisulphite sequencing (*SMYD2*, *IL6R*, *SERPINB9* and *ERG*) pre-designed TaqMan gene expression assays were purchased from Fisher Scientific UK (which were all exon spanning):

Hs01554629\_m1 (*ERG* gene FAM-MGB dye)

Hs00220210\_m1 (*SYMD2* gene: FAM-MGB dye)

Hs01075666\_m1 (*IL6R* gene: FAM-MGB dye)

Hs00394497\_m1 (*SERPINB9* gene: FAM-MGB dye)

Hs02758991\_g1 (*GAPDH* gene: FAM-MGB dye) – normalization reference gene.

*GAPDH* was chosen as the housekeeper gene for this analysis, as a PhD student in my lab assessed which genes are most appropriate for VSMCs from a range of potential candidates. Using a software called ‘NormFinder’, which is an algorithm for identifying the optimal normalization gene among a set of candidate genes, it was identified that *GAPDH* was the most suitable out of 6.

The results from this analysis are below (Table 9), where lower stability values represent more consistent expression patterns between samples, and therefore highlights the most appropriate candidates for potential use as a housekeeper gene.

	Hu PGK1	Hu TBP	Hu GAPDH	Hu B2M	Hu 18S RN	Hu ELF1
	Ct Mean	Ct Mean				
Mean	26.09669	34.19445	23.26022	22.88179	16.50121	27.21769
Stability value	0.173689	0.209648	0.163216	0.307518	0.420254	0.194112

**Table 9 – NormFinder results for potential housekeeper genes:** After a series of candidate genes were tested to see which the most stably expressed in VSMCs were, GAPDH was chosen.

The TaqMan qPCR assays were all performed on an Applied Biosystems Step One Plus qPCR system in duplicate and the qPCR constituents were as follows in a 20µl total volume: 10µl 2x TaqMan Gene Expression Master Mix (ThermoFisher), 1µl 20x TaqMan Gene Expression primers/probes (ThermoFisher), 7µl sterilised distilled water and 2µl cDNA. The TaqMan qPCR cycling conditions are displayed in Table 10.

Temperature	Time	Cycles
50°C	2 minutes	1
95°C	20 seconds	1
95°C	1 second	40
60°C	20 seconds	40

**Table 10 - TaqMan gene expression qPCR cycling conditions.**

Mean Ct values from duplicate reactions of the target gene were subtracted from the mean Ct values of the housekeeper gene to acquire inverse delta Ct values (higher value represents higher expression). Delta Ct values of cases vs controls were compared using two tailed unpaired t-tests in GraphPad Prism 7. The relationship between gene expression and DNA methylation % at CpG sites where significant differential methylation was observed were assessed in each respective gene using linear regression analysis.

### 2.21. Immunohistochemistry to assess protein expression in aortic tissues

Immunohistochemical analysis of aortic tissue was conducted in this project as part of the functional corroboration of observed differential methylation (Chapter 6) from Chapter 4 and 5. The technique utilises antibody binding specificity to target certain proteins, enabling the identification of translated expression profiles of target genes. Specifically, staining of SMYD2 was conducted during this PhD to determine and corroborate whether differences in protein expression would exist between 3 AAA and 3 control abdominal aortic tissue samples, which were identified as differentially methylated and expressed during previous stages of the project.

Three assays were conducted for each individual aortic tissue sample, meaning a total of 18 slides were prepared for microscopy. Six slides (3 controls and 3 AAA) were prepared with a SMYD2 polyclonal rabbit antibody at 50: 1 (ThermoFisher – PA5-51339), six control slides (3 controls and 3 AAA) were prepared with no primary antibody, and six slides (3 controls and 3 AAA) were also prepared with a smooth muscle actin polyclonal rabbit antibody at 100: 1 (ThermoFisher - PA5-19465). This was so that the regions densely populated with smooth muscle cells could be identified in the aorta (tunica media), and the SMYD2 staining analysis was concentrated on where the highest density of smooth muscle fibres was seen. The primary antibodies were detected using the NOVOLINK polymer detection system (Novocastra RE7140-K) and visualisation of the slides was performed on the Hamamatsu Nanozoomer 2.0HT Slide Scanner with the use of NDP VIEW2 software (U12388-01).

The in depth protocol for immunostaining, which was carried out on cut frozen sections from aortic tissues was as follows:

Air dry the frozen tissue cuts for 30 minutes and fix in Acetone for 10 minutes. Rinse in PBS before incubating in peroxidase block for 15 minutes (block endogenous peroxidase activity). Wash slides in PBS for 2 x 5 minutes, then incubate with protein block (blocking of non-specific antibody binding) for 1hour. Wash slides in PBS for 2 x 5 minutes and incubate with optimally diluted primary antibody for one hour at room temperature. Wash the slides in PBS for 2 x 5 minutes and incubate with Post Primary Block for 30 minutes. Wash slides in PBS for 2 x 5 minutes and incubate with Post Primary NovoLink Polymer (anti-rabbit Poly-HRP-IgG) for 30 minutes. Wash slides in PBS for 2 x 5 minutes and finally develop peroxidase activity with DAB working substrate solution for 5 minutes.

### **2.22. Summary of statistical analysis**

Specific descriptions of data and statistical analysis for each experiment have been described under the relevant methodology sections. In summary, statistical analyses were conducted using GraphPad Prism 7 (GraphPad Software, Inc., CA, USA) and IBM SPSS Statistics 24 (IBM, NY, USA). Continuous parametric data are presented as mean value  $\pm$  standard deviation (SD) and non-parametric data are presented as median value and range; categorical data are presented as absolute value or percentage. The chi-square test was used to compare categorical data. T-test was used to compare continuous parametric data. Pearson's correlation coefficient was calculated to assess linear dependence between two variables. A P value of  $<0.05$  was considered statistically significant. Where applicable, a Q value has been reported ( $Q<0.1$ ), adjusted for False Discovery Rate (FDR).

**3. Chapter 3 – Results section 1**

**Peripheral blood mononuclear cell global DNA methylation  
and circulating homocysteine analysis in AAA**

## **Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

### **3.1. Introduction**

It was highlighted in Chapter 1 of this thesis (introduction) that the major pathological hallmarks of AAA are inflammation, VSMC apoptosis and ECM degradation.

Essentially, AAA is considered an inflammatory disease due to the significant impact irregular inflammatory responses play in ECM degradation, vascular remodelling, weakening of the aorta and subsequent aortic dilation as a result of haemodynamic force exerted against the aortic wall<sup>18, 24-26</sup>. It has also been observed that aberrant changes in DNA methylation are in turn associated with irregular inflammatory responses in other major inflammatory diseases<sup>153</sup>, but has never been investigated in regards to AAA. It is plausible that the induction of similar changes to methylation status within the genome may contribute to the development and progression of AAA.

Homocysteine is synthesised from methionine and is strongly linked to the regulation of DNA methylation (homocysteine is produced as a by-product of the methyltransferase reaction)<sup>165</sup>. Abnormal levels of homocysteine have been associated with adverse changes in DNA methylation, which can subsequently contribute towards disease<sup>165</sup>.

homocysteine has previously been studied in AAA and is considered a potential biomarker of cardiovascular disease<sup>169</sup>, however homocysteine has never been studied in relation to global DNA methylation in AAA, which was reviewed by Krishna *et al.*,<sup>164</sup> and is considered an important potential avenue of research.

## **Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

### **3.1.1. Aims**

The aims of the first results Chapter (Chapter 3) are as follows:

1. To establish whether changes in global DNA methylation are evident in DNA taken from the peripheral blood of patients with AAA compared to healthy controls.
2. To establish whether changes in circulating homocysteine levels are evident in blood plasma taken from AAA patients compared to healthy controls.
3. To establish whether there is a direct relationship between global DNA methylation levels and circulating homocysteine.

### **3.1.2. Objectives**

1. Conduct enzyme linked immunosorbent assays (ELISAs) on purified DNA taken from the peripheral blood of 93 AAA and 92 controls to assess global DNA methylation.
2. Perform ELISAs on blood plasma taken from 70 AAA and 67 controls to determine circulating homocysteine concentrations from the same individuals used for the global DNA methylation assay.
3. Perform linear regression to establish whether global DNA methylation is directly related to circulating homocysteine in AAA.

## Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA

### 3.1.3. Hypotheses

1. It is hypothesised that aberrant levels of global DNA methylation will be observed in those with AAA compared to controls.
2. It is hypothesised that there will be changes in the concentration of circulating homocysteine in those with AAA compared to controls.
3. It is hypothesised that there will be a linear relationship between aberrant global DNA methylation levels and aberrant circulating homocysteine concentrations.

### 3.2. Population

Whole genome DNA methylation levels were assessed in DNA derived from 185 PBMC samples from the UKAGS resource (48 large AAA, 45 small AAA and 92 controls) using a colorimetric ELISA. Risk factors that are known to be associated with AAA can be confounders of DNA methylation and include age, smoking, gender, and ethnicity. To help prevent any confounding effects of these factors, white males over the age of 65 with at least a 10 year history of smoking were selected for the study. Table 11 displays the sample group demographics.

	48 Large AAA (>5.5cm)	45 Small AAA (3.0-5.5cm)	92 Controls (<2.5cm)
Median age (IQR)	74 (69-77)	70 (68-74)	68 (65-71)
Mean aortic diameter (cm (+/- SD))	6.8 (0.96)	4.4 (0.68)	1.9 (0.1)

**Table 11 - Demographic summary of samples used for peripheral blood global DNA methylation analysis.**

### Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA

A total of 137 blood plasma samples of those used in the global methylation assays were available (67 controls and 70 AAA) for homocysteine analysis, again performed using colorimetric ELISAs. Table 12 displays group demographics of samples used for homocysteine analysis.

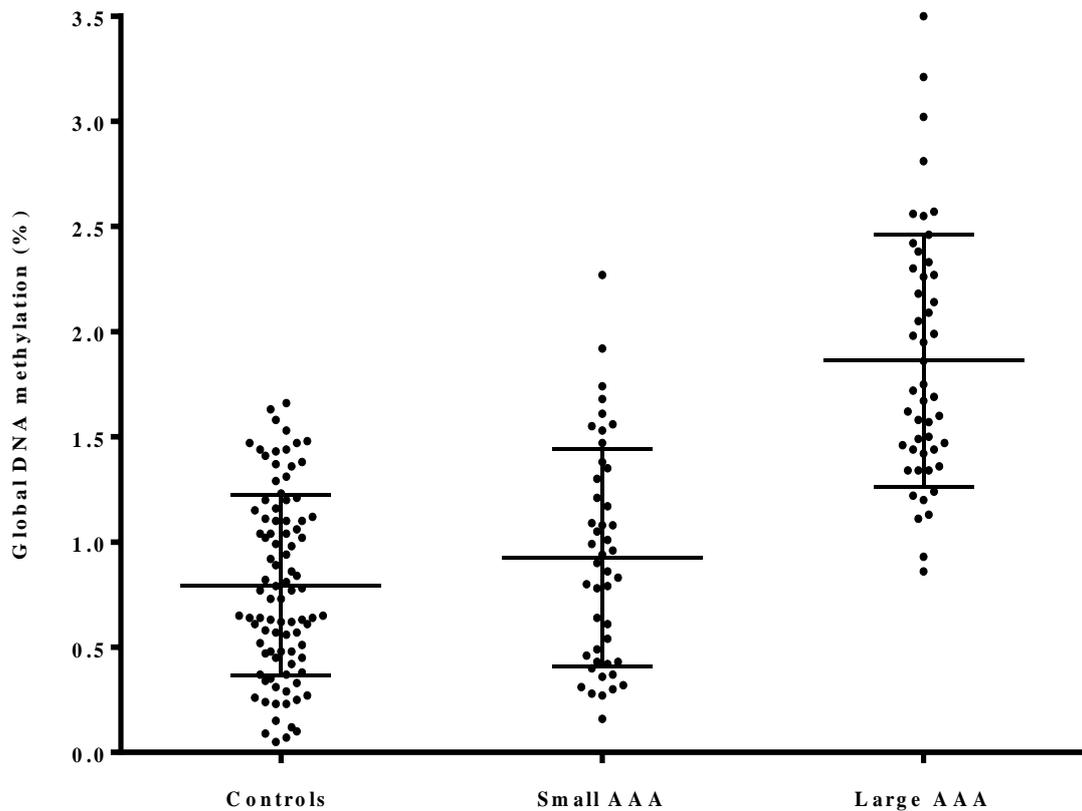
	70 AAA (30-71mm)	67 Controls (<25mm)
Median age (IQR)	73 (68-76)	68 (65-72)
Mean aortic diameter (cm (+/- SD))	5.1 (1.2)	1.9 (0.1)

**Table 12 - Demographic summary of samples used for homocysteine analysis in blood plasma.**

### **3.3. Global DNA methylation analysis**

After global methylation values were acquired for each individual DNA sample (normalised as %), the controls, small AAA and large AAA values were compared.

Figure 22 displays the results from the global methylation assay.



**Figure 22 – Global DNA methylation analysis:** Levels of peripheral blood DNA methylation in controls (1.8-2.3cm, n=92), small aneurysms (3.0-5.5cm, n=45) and large aneurysms (>5.5cm, n=48).

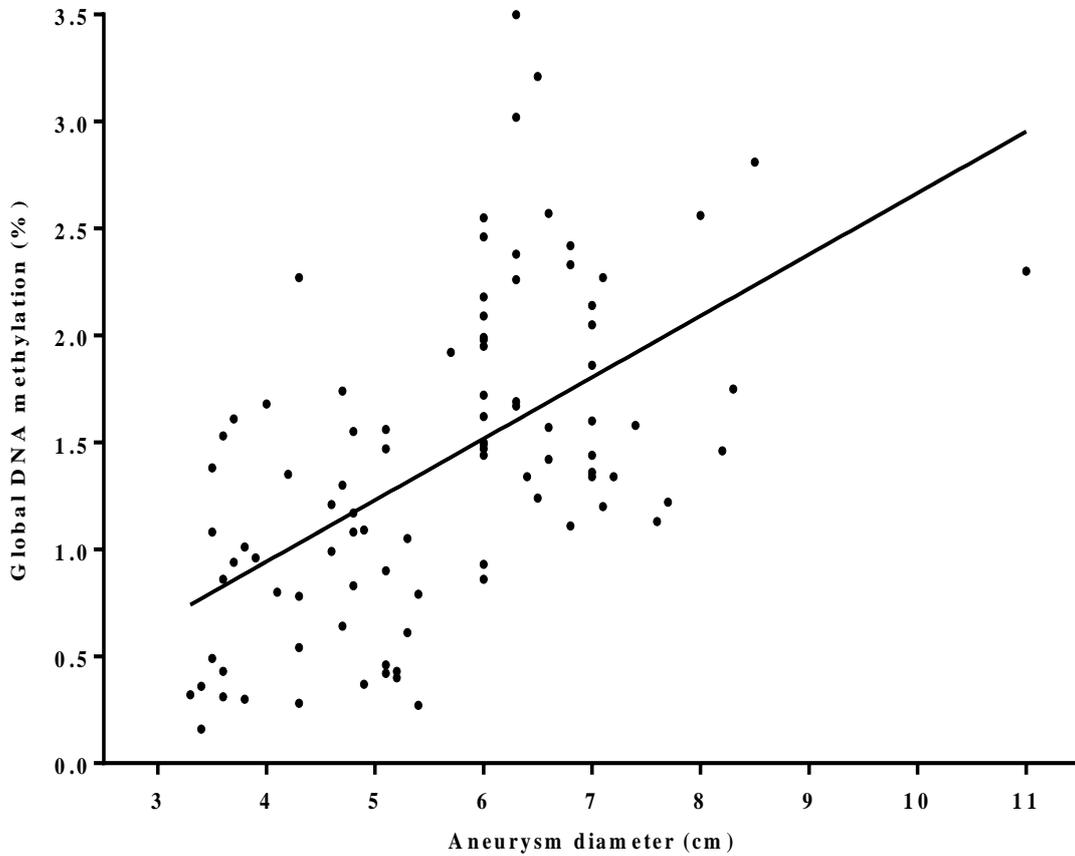
### **Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

Global genomic DNA methylation was significantly higher in men ( $P < 0.0001$ ) with large AAA ( $> 5.5$ cm,  $n = 48$ , global DNA methylation  $1.86\% (\pm 0.6\%)$ ) compared to men with small AAA ( $3.0$ - $5.5$ cm,  $n = 45$ , global DNA methylation  $0.93\% (\pm 0.52)$ ) and controls ( $< 2.5$ cm,  $n = 92$ , global DNA methylation  $0.79\% (\pm 0.43\%)$ ). There was no significant difference between small AAA and controls, however a small visual difference is noticeable.

Further to this analysis, the methylation values of each of the AAA DNA samples were compared to their respective aortic diameters to assess whether there was an association between AAA size and global DNA methylation.

Figure 23 illustrates that there was a clear, positive linear relationship between AAA size and global DNA methylation. The P value was adjusted for differences in age between groups using linear regression analysis and remained statistically significant (Pearson coefficient R square value =  $0.3175$ ,  $p < 0.0001$ ).

**Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**



**Figure 23 - The linear relationship between AAA size and global DNA methylation:** The global methylation values of 45 small AAA and 48 large AAA are plotted against their respective aortic diameters. There was a significant positive correlation between the two ( $R^2=0.3175$  and  $P<0.0001$ ) after being adjusted for age.

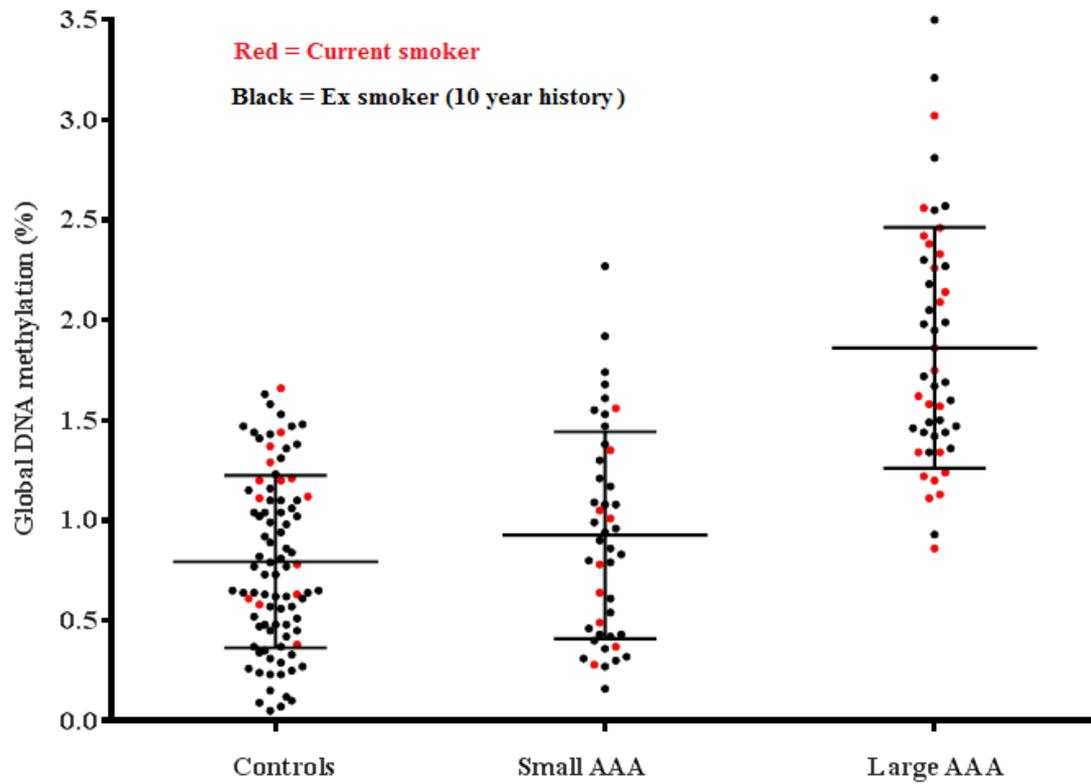
### **Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

The methylation differences observed in those with large AAA which are presented in Figure 22 and 23 are independent of age, gender, ethnicity and smoking based on the criteria of inclusion for this study (previously described).

However the smoking criteria was not based on being a smoker or ex-smoker, it was based on having at least a 10 year history of smoking, resulting in a cohort of current and ex-smokers for each group. Some studies suggest that those who quit smoking immediately start to reduce the risk of AAA growth<sup>40</sup>. However, more relevant to this work, it is conversely suggested that smoking causes irreversible epigenetic switching and long term modification of methylation patterns<sup>109</sup>.

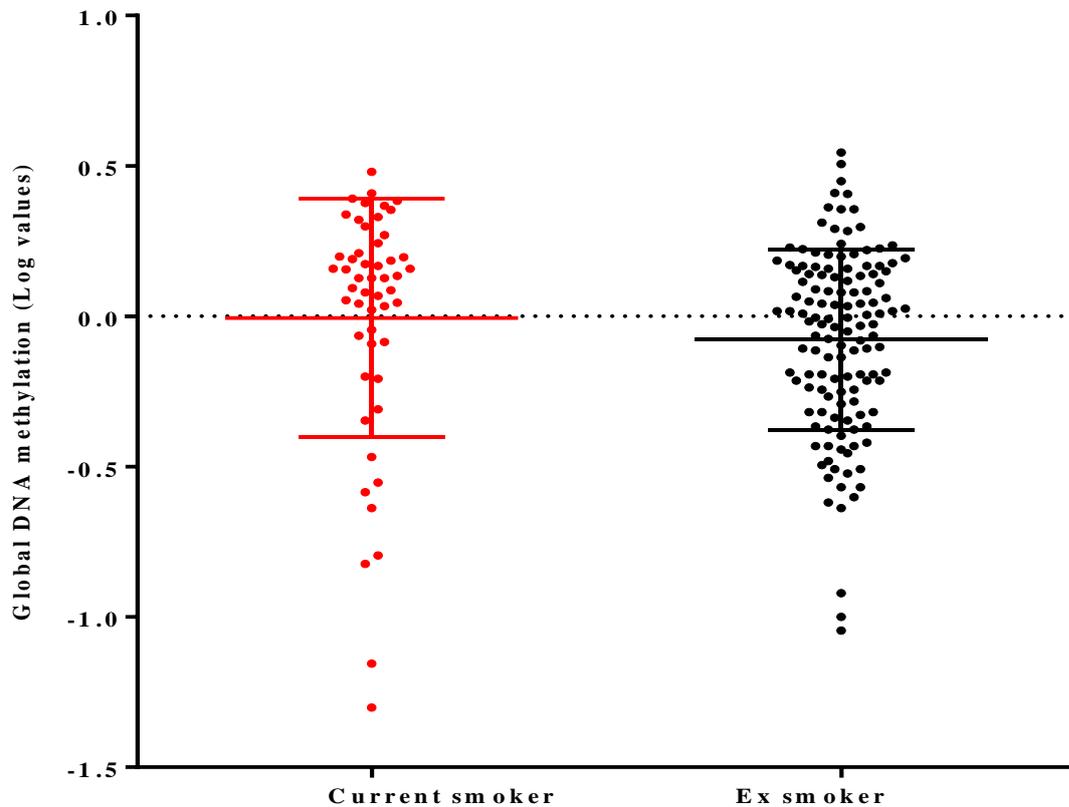
To ensure that there was no bias or confounding effect of the smoking inclusion criteria in this experiment, global DNA methylation of all samples included in this study were split in to two groups (smoker and ex-smoker) and compared (Figure 24 and 25).

**Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**



**Figure 24 - Smoking status and global DNA methylation:** There was an even distribution of current and ex-smokers within and between sample groups where global DNA methylation values were compared.

### Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA



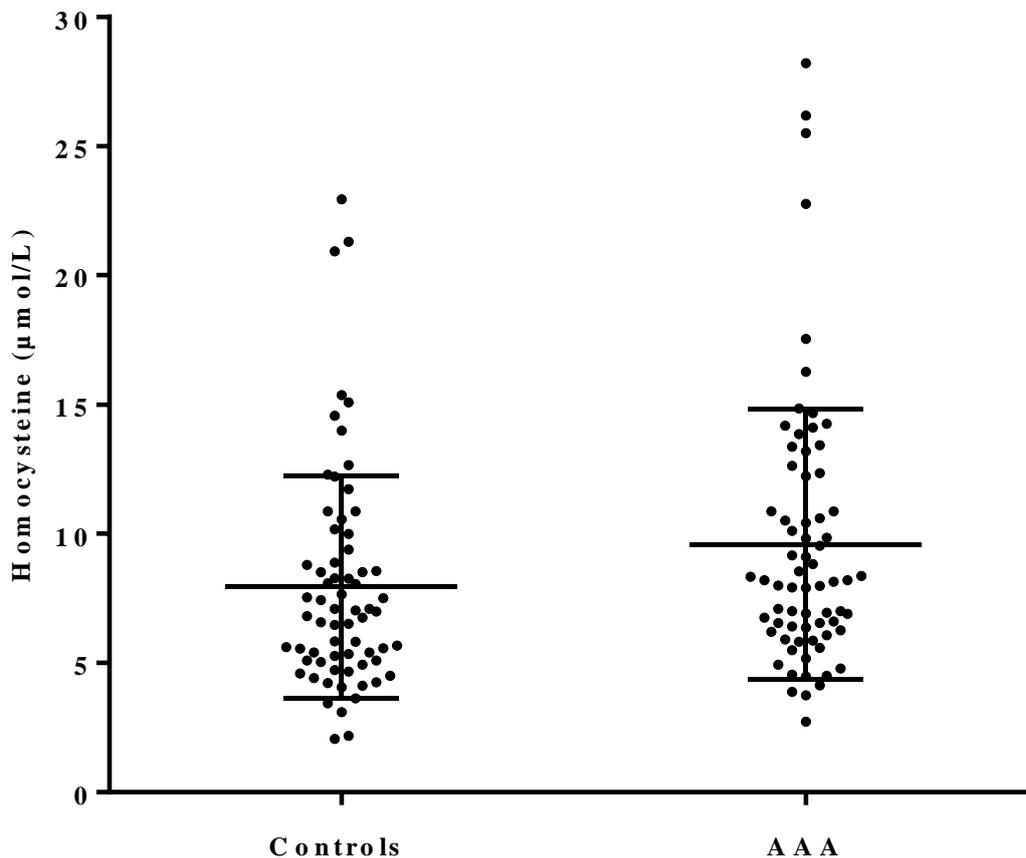
**Figure 25 – Global DNA methylation in ex-smokers vs current smokers:** There was no difference ( $P=0.24$ ) between the log values (to correct for non-normal distributions) of quantified global DNA methylation in ex-smokers and current smokers with at least a 10 year history of smoking.

There was no significant difference ( $P=0.24$ ) between the global methylation of those who still currently smoke, and those who have now ceased smoking but have at least a 10 year history, highlighting that any differences in global DNA methylation in this study are not a result of differential smoking status, which justifies the smoking inclusion criteria for this study.

### 3.4. Circulating blood plasma homocysteine analysis

Circulating homocysteine levels were then measured in blood plasma from the same samples as the global methylation assays in 67 controls and 70 individuals with AAA.

Figure 26 shows that plasma homocysteine levels were significantly higher in men with AAA compared to controls ( $9.6 \mu\text{mol/L} \pm 0.62$ ,  $n=70$ , vs  $7.94 \mu\text{mol/L} \pm 0.52$ ,  $n=67$ ,  $p=0.0433$ ).

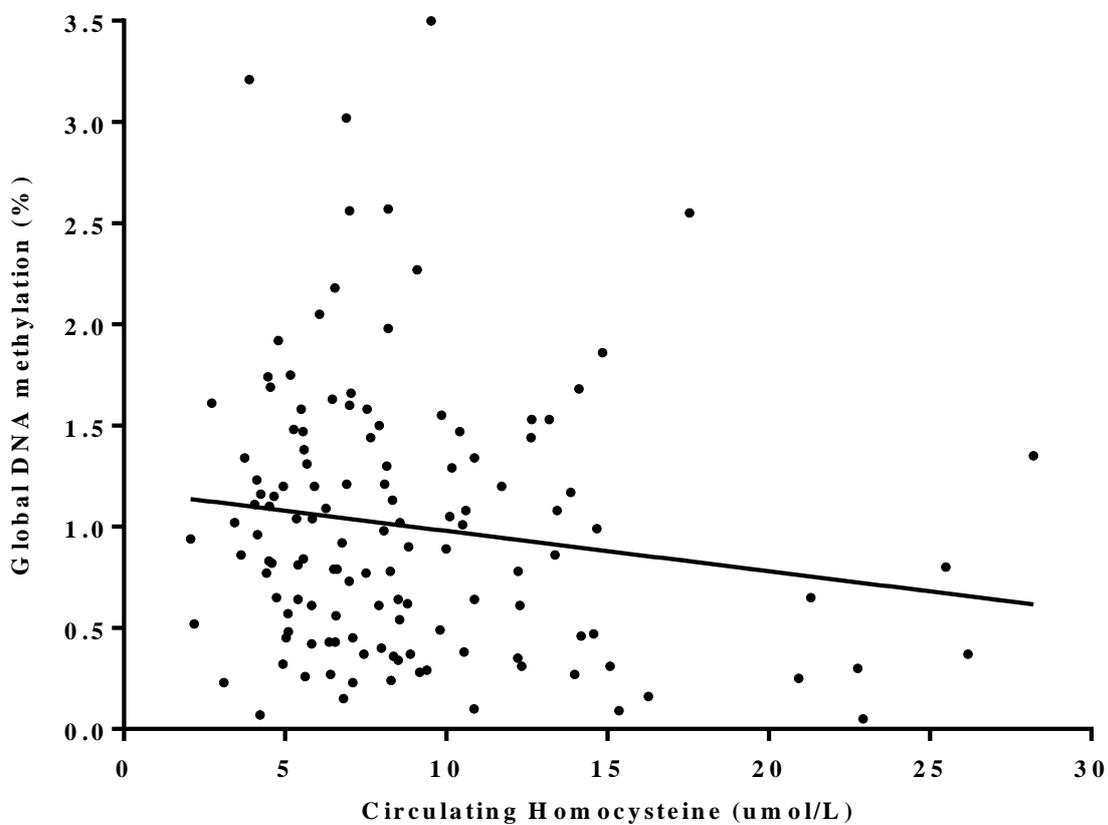


**Figure 26 - Homocysteine analysis in AAA and controls:** Circulating homocysteine levels in matched (from global methylation assay) blood plasma of controls ( $n=67$ ) and AAA ( $n=70$ ). Plasma homocysteine levels were higher in men with AAA compared to controls ( $9.6 \mu\text{mol/L} \pm 0.62$ ,  $n=70$ , vs  $7.94 \mu\text{mol/L} \pm 0.52$ ,  $n=67$ ,  $p=0.0433$ ).

### Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA

However, Figure 27 shows that there was no significant association between global DNA methylation and circulating homocysteine ( $P=0.095$ ,  $R^2=0.021$ ) when linear regression was used to assess the potential relationship.

As a result, this data indicates that it is likely the differences observed in global DNA methylation in those with AAA and increasing AAA size was independent of circulating homocysteine.



**Figure 27 – Global DNA methylation and homocysteine in AAA and controls:** The linear relationship between global methylation percentage and homocysteine levels ( $n=137$ ). There was no significant association between global methylation and homocysteine levels ( $P=0.095$ ,  $R^2=0.021$ ).

### **3.5. Conclusion**

Global DNA hypermethylation is commonly a hallmark of chronic inflammation and has been observed as a potential pathological marker of disease<sup>102, 147-149, 154, 155</sup>. Inflammation in turn is a pathological hallmark of AAA, yet up until now, no work has investigated global methylation in AAA. This chapter has illustrated for the first time that global DNA hypermethylation in PBMC DNA is associated with large AAA and increasing AAA size. These results were independent of smoking, gender, age and ethnicity, highlighting the likely effects of chronic inflammation on global methylation in AAA.

homocysteine levels have been previously associated with global methylation<sup>189</sup> and is a proposed biomarker for cardiovascular disease<sup>166</sup>. homocysteine has also been suggested to play a role in AAA<sup>164</sup>. However, the differences seen in circulating blood plasma homocysteine between patients with AAA and controls in this study were not clinically different (hyperhomocysteinemia is described above 15  $\mu\text{mol/L}$ ), and supports the idea that homocysteine is in fact not an appropriate biomarker for AAA, which is also suggested by Lindqvist *et al.*,<sup>172</sup>. There was also no relationship between homocysteine and global DNA methylation in this study, suggesting that global DNA hypermethylation is associated with AAA independently of circulating blood homocysteine, which is in contrast to the case argued by Krishna *et al.*,<sup>164</sup>.

The work presented in this chapter provides novel associations between global DNA methylation and the presence of AAA and AAA size, and suggests that similarly to other inflammatory diseases, increased levels of global DNA methylation may be a factor in the pathobiology of AAA. However one limitation of this work is the use of ELISAs to assess genome-wide DNA methylation. This technique is not as informative as other available methodologies, such as bisulphite NGS, Illumina's 450k human methylation

### **Chapter 3: Peripheral blood mononuclear cell global DNA methylation and circulating homocysteine analysis in AAA**

micro-array and Methylated DNA immuno-precipitation (MeDIP). Such alternative methodologies provide more informative, base specific DNA methylation values at any desired location in the genome. However, these more advanced techniques greatly increase experimental cost and time, whilst also decreasing sample size.

This was a foundational study to identify a potential epigenetic basis to the aetiology of AAA, and was the first work to do so, providing significant rationale for further, more in depth analysis of DNA methylation and its potential role in the development and progression of AAA. Further, more informative analysis will be presented in subsequent chapters of this PhD and will concentrate on individual CpG methylation changes and the functional effects of these changes in genes located at AAA genomic risk loci.

**4. Chapter 4 – Results section 2**

**Peripheral blood cell DNA methylation analysis in AAA**

### 4.1. Introduction

The results presented in Chapter 3 – Results section 1 illustrated for the first time that global DNA hypermethylation is associated with the presence of large AAA and increasing AAA size, a clear justification and rationale for further study of the role of DNA methylation in the aetiology of AAA.

It was considered likely at this point during the project that aberrant DNA methylation changes would exist throughout the genome in individuals with AAA, and that methylation status of those with AAA would need to be assessed at individual CpG sites using more accomplished methodological techniques than those adopted in the previous Chapter.

A genome-wide approach to assess DNA methylation (e.g. WGBS, MeDIP or 450k methylation micro-array) was considered beyond the scope of this PhD due to the amount of time, resources and funding that would be required for even a small sample size, which would therefore yield low powered results. As a result of this, a hypothesis driven candidate gene sequencing approach was adopted for this Chapter, which would allow a larger sample size whilst still adopting the best available methodology to study CpG specific DNA methylation in AAA and controls.

Variations in DNA sequence at polymorphic loci can result in variable patterns of DNA methylation, known as methylation quantitative trait loci (meQTL)<sup>120, 126 121</sup>. Disease variants can alter transcription factor levels and methylation of their binding sites<sup>125</sup>. The direct mechanisms of the variants associated with AAA are not fully known, but many are intronic and it is likely that their effects are regulatory in nature. It is therefore feasible that meQTL exist in AAA, and considering many meQTL act in *cis*<sup>126</sup>, their likely effects on methylation are in the genes surrounding the risk loci identified in previous GWASs.

#### **Chapter 4: Peripheral blood cell DNA methylation analysis in AAA**

Therefore in this study, targeted bisulphite NGS was performed in PBMC DNA from individuals with AAA and controls to investigate the methylation status of CpG islands in regulatory regions of genes located at AAA genomic risk loci. This approach provides a more refined way to assess the methylation status of genes in proximity to AAA risk loci than other methodologies. The identification of adverse epigenetic modifications directly linked to patients with AAA could offer a more comprehensive understanding of AAA pathobiology, and an alternative research avenue in the search for a future treatment strategy<sup>15, 16</sup>. Nobody had ever studied the methylation status of PBMCs using NGS in AAA prior to this work.

### 4.1.1. Aims

1. To identify whether changes in peripheral blood CpG methylation exist in the regulatory regions of genes (promoters and transcriptional start sites) proximal to AAA genomic risk loci in those with AAA compared to healthy controls.

### 4.1.2. Objectives

1. Extract DNA from the peripheral blood of 48 AAA and 48 controls and isolate the regulatory regions of 7 genes proximal to AAA genomic risk loci using bisulphite specific PCR in each of the 96 samples. Targeted genes include: *LRP1*, *GDF7*, *SPAG17*, *ERG*, *CDKN2B*, *SORT1*, *LDLR*, *IL6R* and *MMP9*.

2. Conduct library preparation for sequencing, which involves ligating unique adapter sequences to each individual DNA sample prior to multiplexing.

3. Run samples for quality check on bioanalyzer and conduct bisulphite specific next-generation sequencing on the Illumina Mi-seq platform.

4. Use a bioinformatics platform (command line) and open source software to analyse raw sequencing data. Extract methylation values (normalised as %) at each CpG in each gene for each individual sample after applying high quality data filters.

5. Statistically analyse the extracted methylation data using multiple t-test analysis to establish whether changes to CpG methylation exist between AAA and controls.

### 4.1.3. Hypotheses

1. Is hypothesised that significant differences in CpG specific DNA methylation will be observed in those with AAA compared to controls at multiple CpG sites in several targeted genes.

### 4.2. Population

Bisulphite NGS was used to assess the methylation status of the regulatory regions of genes associated with AAA in PBMC derived DNA from 96 of the samples used in the global DNA methylation assay (48 controls and 48 AAA, Table 13).

	48 Large AAA (>5.5cm)	48 Controls (<2.5cm)
Median age (IQR)	74 (69-77)	69 (67-72)
Mean aortic diameter (mm (+/- SD))	6.8 (0.96)	1.9 (0.1)

**Table 13 - Demographic summary of samples where peripheral blood cell DNA was bisulphite treated and used for next generation sequencing.**

Similarly to Chapter 3, white males over the age of 65 with at least a 10 year history of smoking comprised each sample group to avoid potential confounding variable effects of these factors on DNA methylation.

### 4.3. Targeted bisulphite sequencing of PBMC DNA from AAA and controls

A candidate gene approach was adopted in this project to address whether CpG specific methylation changes existed in those with AAA compared to healthy controls. More specifically, the gene promoters and transcriptional start sites of genes proximal to AAA genomic risk loci were targeted.

Gene promoters and transcriptional start sites of candidate genes were identified using the transcriptional regulatory element promoter database: <https://cb.utdallas.edu/CGi-bin/TRED/tred.CGi?process=searchPromForm> and the NCBI gene database: <https://www.ncbi.nlm.nih.gov/gene>. This simply involved searching for the gene names in the regulatory element sections within these databases. All identified candidate gene sequences identified within these databases were acquired from the GRCh38 genome assembly.

Promoter sequences were then copy and pasted to a word document and bisulphite specific PCR primers were designed, based on these acquired sequences, to isolate candidate regions of interest (CpG islands) using a programme called methprimer<sup>190</sup> as previously described in Chapter 2.

### 4.3.1. Genes for PBMC sequencing

At the time when this experiment was designed and conducted, the complete results from the most recent GWAS meta-analysis<sup>83</sup> was not complete. Due to this, the candidate genes chosen for this experiment were based on the provisional results of the GWAS meta-analysis. Hence, there are additional loci/genes now known to be associated with AAA, which I did not know about at the time of this experiment.

Based on the rationale presented previously, and the provisional results from the recent GWAS, the candidate genes chosen and targeted for this experiment included: *SPAG17*, *GDF7*, *LRP1*, *ERG*, *MMP9*, *CDKN2B*, *LDLR*, *IL6R* and *SORT1*.

The provisional GWAS results that the genes were based on are shown in Table 14.

Lead SNP	Region	Discovery phase P-value	Validation phase P-value
rs2765283	<i>SPAG17</i>	2.71E-09	Not reported
rs4129267	<i>IL6R</i>	1.74x10 <sup>-10</sup>	1.81x10 <sup>-4</sup>
rs602633	<i>PSRC1-CELSR2-SORT1</i>	3.12x10 <sup>-8</sup>	9.83x10 <sup>-3</sup>
rs13382862	<i>GDF7</i>	4.78E-09	0.36
rs10757274	<i>CDKN2B-AS1</i>	2.71x10 <sup>-14</sup>	1.02x10 <sup>-21</sup>
rs1385526	<i>LRP1</i>	2.10E-09	6.4x10 <sup>-7</sup>
rs9316871	<i>LINC00540</i>	1.23x10 <sup>-6</sup>	8.28x10 <sup>-5</sup>
rs6511720	<i>LDLR</i>	8.60x10 <sup>-13</sup>	6.02x10 <sup>-4</sup>
rs3827066	<i>PCIF1/MMP9/ZNF335</i>	1.88x10 <sup>-10</sup>	2.00x10 <sup>-8</sup>
rs2836411	<i>ERG</i>	2.51x10 <sup>-8</sup>	1.13x10 <sup>-2</sup>

**Table 14 - Summary of provisional GWAS meta-analysis results:** The discovery phase of these results were used as a guide to include candidate genes for Chapter 4 of this PhD project. Validation phase P values are presented retrospectively, as these were only reported after the PBMC sequencing had been conducted. Lead SNP, genes where the lead SNPs are located, and the GWAS P-values are displayed.

#### Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

As shown in Table 14, when the lead SNP is not within an individual identified gene, the two nearest genes to the lead SNP are displayed. *SORT1* is already known to be associated with AAA, and was therefore included in this study as opposed to *CELSR2*. *MMP9* was also chosen as a candidate over *ZNF335*, as *MMP9* has been extensively linked to AAA development. Finally, *LINC00540* was, at the time of this study, identified as a non-coding RNA with no known function and was not included.

The specific targeted loci for the PBMC sequencing assay are displayed in Table 15.

**Chapter 4: Peripheral blood cell DNA methylation analysis in AAA**

<b>Gene symbol</b>	<b>Chromosome and NCBI sequence ID</b>	<b>NCBI gene co-ordinates</b>	<b>Primer sequences 5'-3'</b>	<b>Annealing temp (°C)</b>	<b>CpG coverage</b>
<i>LRPI</i>	Chr: 12 NG_016444.1	4274-4640	F: GGGTATTAAGGTGGGTTTTATTTT R: TACTCTAAAATTTCAAACCTCCCTCC	56.5	25
		4680-5036	F: GTATTAGGGGAGGAGGGTTTAGTTAG R: TCCTCAATACATAAACCTAAAACCTC	54.4	24
<i>ERG</i>	Chr: 21 NG_029732.1	282787-283319	F: TTTTATTAGGTAGTGGTTAGATTTAGTTTT R: TACCCCAATTAATAAAATCCAATATA	54.4	22
		283331 -283680	F: TTGGAATTTATTAATTGGGGGTATATAT R: AACTATCTTTTACAAAATCAATCCAC	57.7	6
<i>MMP-9</i>	Chr: 20 NG_011468.1	4040 -4416	F: TATATTTGGTTTTGAATTTTGGGTT R: ACTAAAACAACCCCTATCTTCC	57	5
<i>CDKN2B</i>	Chr: 9 NG_023297.1	4309 -4749	F: TTTTGTTTTATGTGTGTTAGGTTGTTATT R: AAATATTCACCACTACCCCAACTC	58.5	8
		4771 -5267	F: AGTTGAGGGTAGTGGTGAATATTT R: TCTCCTCCTAAAAAACCTAAAACCTC	57	45
<i>LDLR</i>	Chr: 19 NG_009060.1	4376-4725	F: TTTTTAAGGGGAGAAATTAATATTTA R: CACAAAAAATAACAACAACCTTTC	54.4	5
		4776 -5355	F: GAAAGGTTGTTGTTATTTTTTTGTG R: AAACCTCCCTCTCAACCTATTCTAAC	57.7	28
<i>IL6R</i>	Chr: 1 NG_012087.1	2781 -3061	F: TTTTGTTTAGGTTGGAGTGTAGTG R: CAAAATTTAATATTATAATTCACATAAAA	51	8
		3095 -3729	F: GTTTGTTTTGGTGTAGAGATAGGTG R: ATAAAACCTCCCAAATAAACAAAACC	57.7	7
<i>SORT1</i>	Chr: 1 NG_028280.1	4363 -4804	F: AGATTATTTTTTAGGTTTGTAGGAGTTA R: AACAAAACTACTAAAATCAACCCC	54.4	24
<i>SPAG17</i>	Chr: 1 NG_053041.1	4415-4664	F: ATTGGAGTTGTTTGGGTAGATAAGTATAT R: CCAAATTAATAACTAACCCCTTCA	57.7	4
		4684-5105	F: TTTTTGAAGGGGTTAGTATTTTAATTTT R: AATTCCTCCTTCTCCTTCTTAAATAC	56.5	25
<i>GDF7</i>	Chr: 1 NC_000002.12	20650525-	F: ATTGTGTTTTTATGGTAATTTGAGT	54.4	24
		20650919	R: AAACCTAATACAAATCTCCCAAC		
		20650947 - 20651336	F: GTTGGGGAGATTTGTATTAGTTTT R: CCAAACCTAACTTCTTCTAAATAATTAC	54.4	17

**Table 15 - Candidate genes for peripheral blood DNA bisulphite sequencing:** Corresponding genomic locations, primer sequences, optimised annealing temperatures and CpG coverage are displayed for each amplicon.

Unsuccessful PCR primers which were designed to isolate candidate CpG islands of interest are shown in Table 16. For each of these primer sets, I conducted rigorous amounts of PCR optimisation but could not get the reactions to work. Overall the whole process took many weeks.

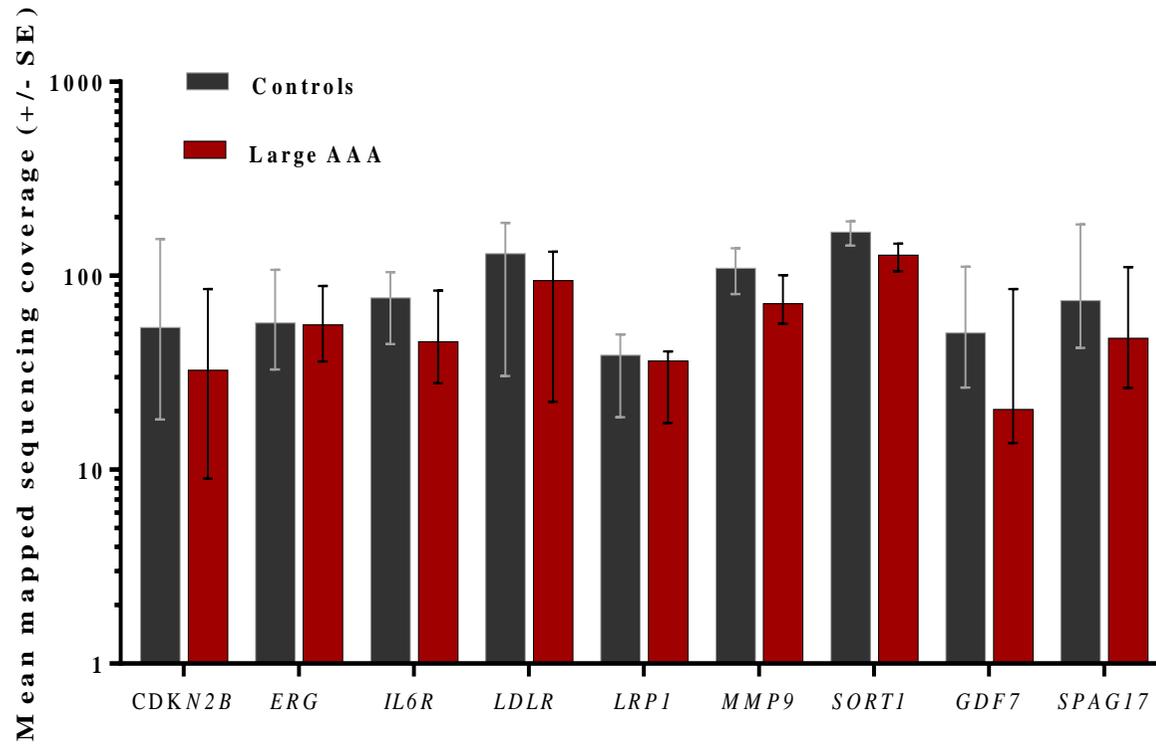
Gene Symbol	Primer sequence 5'-3'
<i>LRP1</i>	F: TTTTAAATTAGGGTTTGTTTTTTT R: AACTAAACCTCCTCCCTAATACTC
<i>LRP1</i>	F: TTTTTTAAATTAGGGTTTGTTTTTTT R: CCTAACTAAACCTCCTCCCTAATA
<i>LRP1</i>	F: TTTTTTAAATTAGGGTTTGTTTTTTT R: TACTCTAAAATTTCAAACCTCCCTCC
<i>MMP-9</i>	F: GAATTTTGGGTTTTGGTTTTAGTAAT R: TAACCCATCCTTAACCTTTTACAAC
<i>MMP-9</i>	F: TTTTGGTTTTAGTAATTAATAATTATT R: TCCTCTCCCTACTTCATCTAAAAAC
<i>MMP-9</i>	F: GATGGGGGATTTTTTTAGTTTTATT R: TACCCACCTCTACCAACTACCTATC
<i>MMP-9</i>	F: TGGGGGATTTTTTTAGTTTTATTTT R: AAACAACAACCCAACACCAA
<i>MMP-9</i>	F: TTTATAAGTTTTGTAGTTTGTAATAATTTA R: AAACACCAAAACCAAAAACCTACC
<i>MMP-9</i>	F: TAGTTTGTAATAATTTATTTTTTTTT R: AAACACCAAAACCAAAAACCTACC
<i>LDLR</i>	F: TTTTTTAAAGGGGAGAAATTAATATTTA R: CACAAAAAATAACAACAACCTTTC
<i>LDLR</i>	F: TTTTTAAGGGGAGAAATTAATATTTA R: TTATCCTCTCTACCTCAAAAAAAA
<i>IL6R</i>	F: TTTAGGTTGGAGTGTAGTGGTGTTAT R: ATCATACAAATCCCAACTTTACCAC
<i>IL6R</i>	F: TTTTGTTTAGGTTGGAGTGTAGTG R: ACCCCTATTTCTAATTTATAAAAATAAAC

Table 16 - Primers designed at targeted gene loci that were not successful.

#### 4.4. DNA methylation status of candidate gene regions in PBMCs

After bisulphite sequencing (described in Chapter 2) of the regulatory regions of the 9 candidate genes in 96 PBMC DNA samples (48 AAA and 48 controls), data analysis was performed. Bioinformatics analysis was conducted on raw sequencing data which was then extracted and further statistically analysed. Only sequencing reads with a base quality of  $>Q20$ , read quality of  $>Q50$  and read depth of  $>5$  were mapped to the reference sequence and used for data analysis, essentially meaning the data obtained from this experiment was very accurate. Statistical analysis was then based on a significance of  $P<0.05$  and false discovery rate of  $Q<10\%$  after multiple t-test comparisons.

A summary of the mapped sequencing read depth of each gene from the PBMC sequencing assay is shown in Figure 28. Sequencing coverage was consistently high in each gene and there were no significant differences in depth between AAA and controls in any gene, suggesting that there was no sequencing bias between AAA and controls. As the mean sequencing coverage was considerably high for each gene in AAA and controls, it further suggests that the number of high quality mapped reads signify a high reliability of results in addition to accuracy.



**Figure 28 – Sequencing coverage of candidate genes in PBMC DNA:** Mean coverage of high quality sequencing reads in candidate genes from PBMC DNA bisulphite sequencing of 48 AAA and 48 controls. Bars represent standard deviation.

#### Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

After the methylation status (normalised as percentage) of each CpG, in each gene, in each of the 96 individual samples was determined, cases and control values were compared, and significant differential methylation was observed in 3 genes (*LDLR*, *IL6R* and *SORT1*).

Conversely, there were no differences in methylation status of the regulatory regions sequenced in all of the other 6 genes (*SPAG17*, *GDF7*, *CDKN2B*, *ERG*, *LRP1* and *MMP9*).

The results from each gene will now be described individually.

#### 4.4.1 *IL6R* gene

A schematic representation of the *IL6R* gene region that was sequenced in this study is shown in Figure 29. This shows the amplicon sequencing coverage spanning the gene promoter shortly upstream of the exon 1 transcriptional start site.

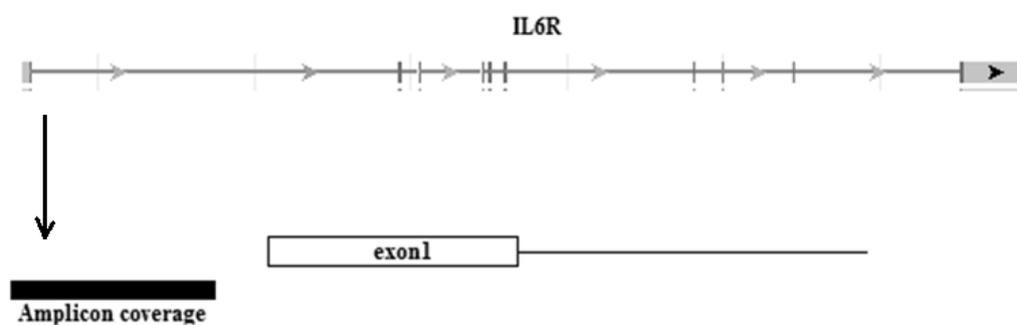
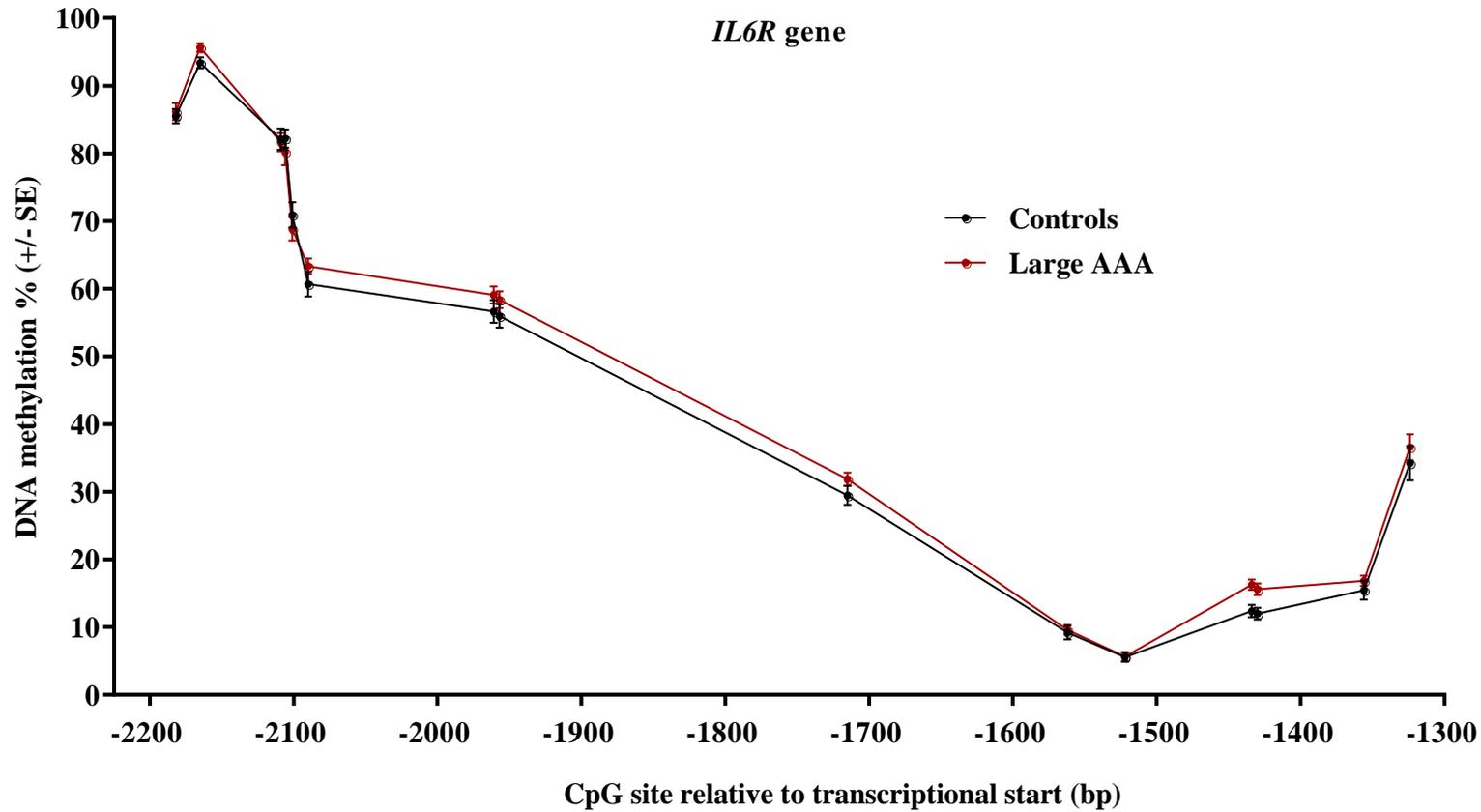


Figure 29 - Schematic representation of the sequenced region of the *IL6R* gene in PBMC and VSMC DNA.

Two amplicons were designed to isolate a total of 15 CpGs in the *IL6R* gene promoter at the genomic coordinates: NG\_012087.1: 2781-3729bp. The DNA methylation status of the entire sequenced promoter is shown in Figure 30.



**Figure 30 - DNA methylation status of the *IL6R* gene promoter in PBMC DNA:** DNA methylation status of 16 sequenced CpG sites in the *IL6R* gene promoter isolated from the peripheral blood of 48 AAA and 48 controls. 2 CpG sites were differentially methylated in AAA (-1434 (P=0.0016) and -1430 (P=0.0037)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

There was differential methylation at 2 CpGs spanning 4bp upstream of the transcriptional start site at the following loci: NG\_012087.1:g.3566\_3570|gom.

Mean methylation at these differentially methylated CpG sites was 15.93%  $\pm$  0.49 vs 12.19%  $\pm$  0.27 (AAA vs controls). The mean increase in methylation in those with AAA was 3.74%  $\pm$  0.19 (P=0.0027, Q=0.026).

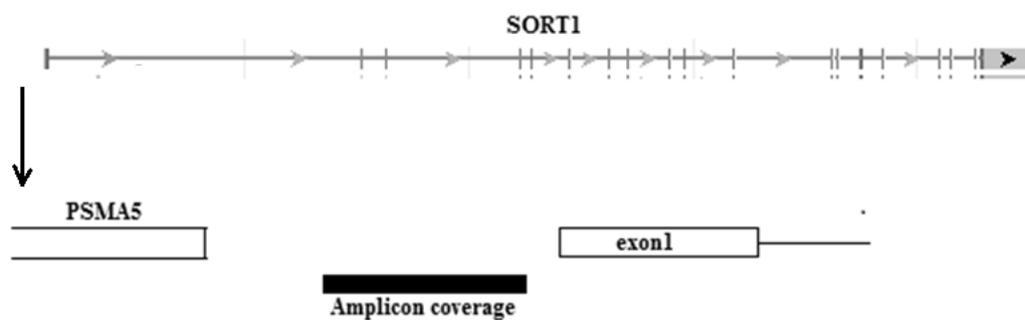
Descriptive statistics of the differential methylation observed in *IL6R* after bisulphite sequencing of PBMC DNA in 48 AAA and 48 controls are shown in Table 17.

Gene and location	Region (CpG site)	Relative to Transcription start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>IL6R</i> Chr: 1	3566	-1434	12.38	16.27	3.89	0.0016	0.0247	2.39
NG_012087.1	3570	-1430	12	15.58	3.58	0.0037	0.0276	2.3

**Table 17 – Descriptive statistics of the differentially methylated CpGs in *IL6R* after PBMC sequencing:** The sequenced gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported.

#### 4.4.2 *SORT1* gene

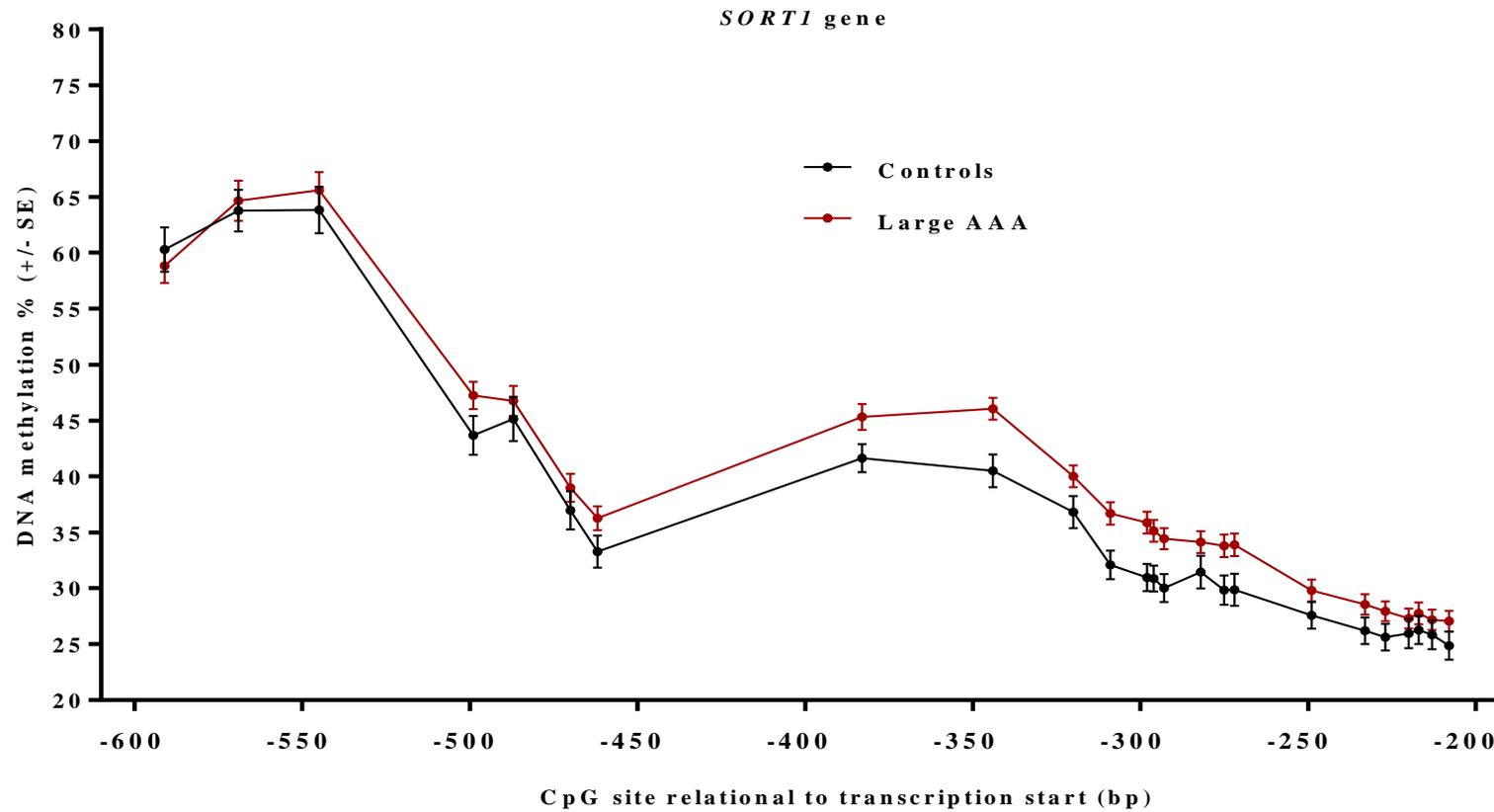
A schematic representation of the *SORT1* gene region that was isolated and sequenced in this study is shown in Figure 31. This shows the amplicon sequencing coverage spanning the gene promoter shortly upstream of the exon 1 transcriptional start site, which is located downstream from the 3' end of the *PSMA5* gene.



**Figure 31 - Schematic representation of the sequenced region of the *SORT1* gene in PBMC and VSMC DNA.**

A single amplicon was designed to isolate a total of 24 CpGs in the *SORT1* gene promoter at the genomic coordinates: NG\_028280.1: 4363-4804bp.

The DNA methylation status of the entire sequenced promoter is shown in Figure 32.



**Figure 32 - DNA methylation status of the *SORT1* gene promoter in PBMC DNA:** DNA methylation status of 24 sequenced CpG sites in the *SORT1* gene promoter isolated from the peripheral blood of 48 AAA and 48 controls. 7 CpG sites were differentially methylated in AAA (-345 (P=0.0009), -310 (P=0.0049), -299 (P=0.0023), -297 (P=0.0087), -294 (P=0.0070), -276 (P=0.0227) and -273 (P=0.0176)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

In Figure 32 there is a clear visual increase in methylation status of the *SORT1* promoter in individuals with AAA compared to controls, which is evident across a large section of the whole CpG Island.

Within the sequenced *SORT1* promoter, 7 CpG sites were significantly hypermethylated (P=0.0092, Q=0.045) in people with AAA (mean - 36.55% ± 4.32) compared to controls (mean – 32.03% ± 3.82) and the mean increase in methylation was 4.53% ± 0.56.

The differentially methylated region spans 72bp and lies in the *SORT1* gene promoter upstream of the transcriptional start site (RefSeq gene coordinate NG\_028280.1:g.4646\_4718|gom).

Descriptive statistics of the differential methylation observed in *SORT1* after bisulphite sequencing of PBMC DNA in 48 AAA and 48 controls are shown in Table 18.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>SORT1</i> Chr: 1 NG_028280.1	4646	-345	40.5	46.06	5.56	0.0009	0.0214	2.82
	4681	-310	32.08	36.69	4.61	0.0049	0.0389	2.71
	4692	-299	30.98	35.88	4.9	0.0023	0.0276	2.76
	4694	-297	30.88	35.13	4.25	0.0087	0.0417	2.44
	4697	-294	30.03	34.44	4.41	0.007	0.0417	2.64
	4715	-276	29.83	33.79	3.97	0.0227	0.0779	2.71
	4718	-273	29.88	33.88	4	0.0176	0.0702	2.14

**Table 18 – Descriptive statistics of the differentially methylated CpGs in *LDLR* after PBMC sequencing:**

The sequenced gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported.

#### 4.4.3 *LDLR* gene

A schematic representation of the section of *LDLR* gene that was isolated and sequenced in this study is shown in Figure 33. This shows the amplicon sequencing coverage spanning the gene promoter upstream of transcriptional start site in addition to the whole of exon 1.

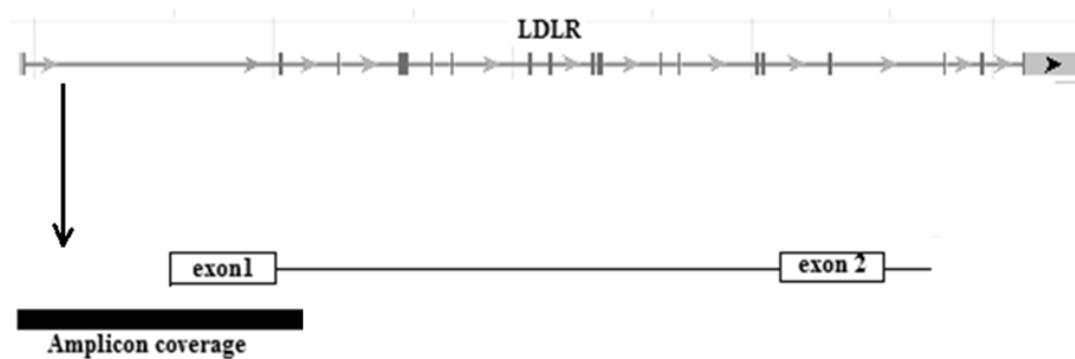
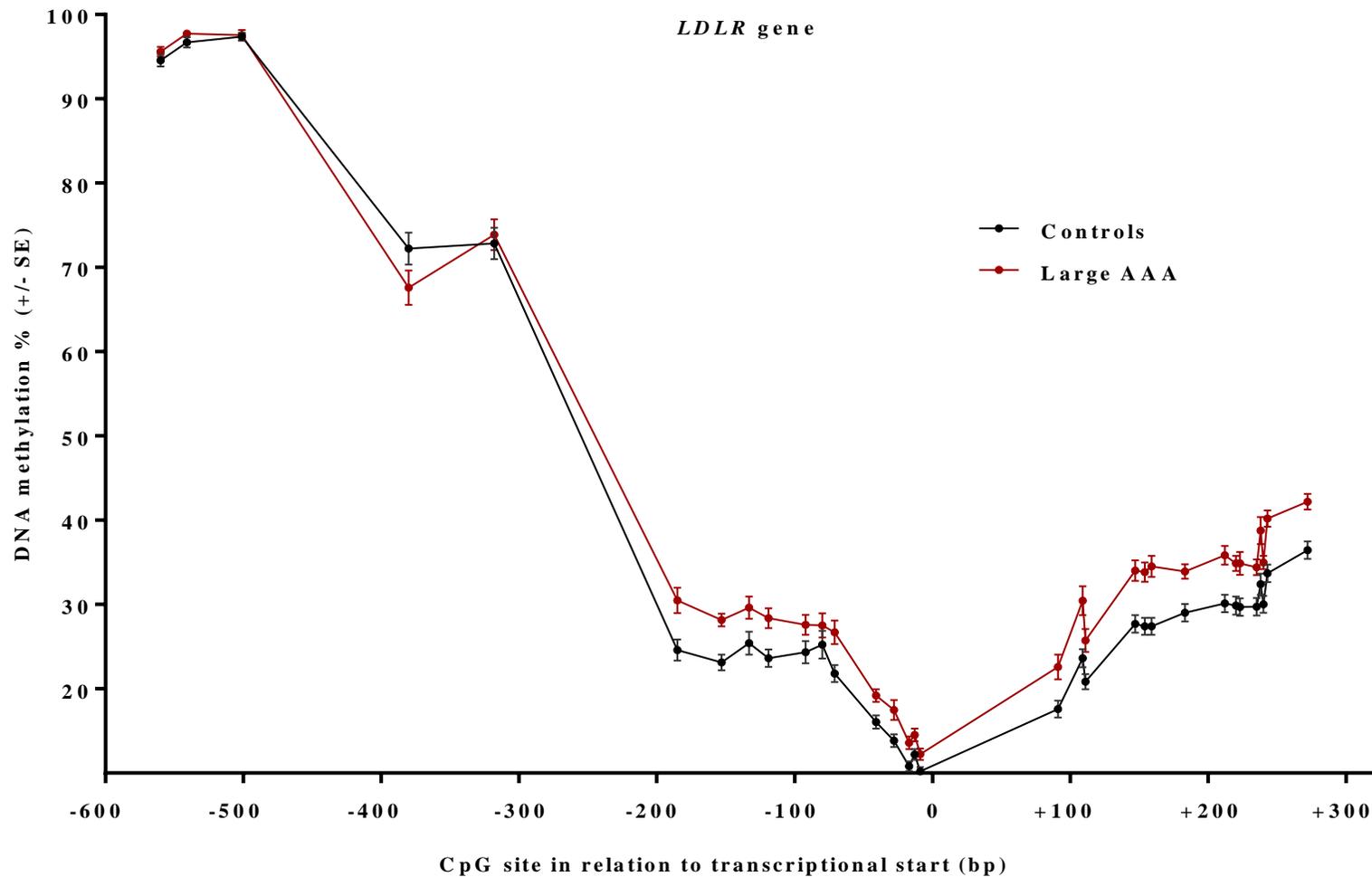


Figure 33 - Schematic representation of the sequenced region of the *LDLR* gene in PBMC DNA.

Two amplicons spanning this region for a total of 33 CpGs at NG\_009060.1: 4376-5355 were sequenced. The DNA methylation status of the whole sequenced CpG Island is shown in Figure 34.



**Figure 34 - DNA methylation in the *LDLR* gene promoter in PBMC DNA:** DNA methylation status of 33 sequenced CpG sites in the *LDLR* gene isolated from the peripheral blood of 48 AAA and 48 controls. 20 CpG sites were differentially methylated in AAA (-187 (P=0.0003), -155 (P=0.0001), -135 (P=0.0054), -121 (P=0.0008), -73 (P=0.0018), +89 (P=0.0127), +107 (P=0.0008), +109 (P=0.0054), +145 (P=0.0001), +152 (P=0.00001), +157 (P=0.00001), +181 (P=0.0003), +210 (P=0.0002), +218 (P=0.0004), +221 (P=0.0012), +233 (P=0.0007), +237 (P=0.0003), +239 (P=0.0001), +242 (P=0.00001), +271 (P=0.0001)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

Here, a large portion of the sequenced *LDLR* CpG Island is hypermethylated in those with AAA compared to controls. In total, 20 individual CpG sites were consistently and significantly hypermethylated ( $P=0.0015$ ,  $Q=0.0025$ ) in PBMC DNA from 48 people with AAAs, displaying a mean overall hypermethylation ( $32.73\% \pm 4.94$ ) compared to 48 controls (mean -  $27.21\% \pm 4.64$ ).

The mean increase in methylation was  $5.52\% \pm 0.83$ . The hypermethylated region of *LDLR* spans 457bp (RefSeq gene coordinate NG\_009060.1: g.4795\_5252| gom), located partially in the promoter region upstream of the genes transcriptional start, completely spans exon 1 and ends in the intronic region between exons 1 and 2.

Descriptive statistics of the differentially methylated CpG sites in *LDLR* are shown in Table 19.

Chapter 4: Peripheral blood cell DNA methylation analysis in AAA

Gene and location	Region (CpG site)	Relational to transcription Start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>LDLR</i> Chr: 19 NG_009060.1	4795	-187	24.58	30.48	5.9	0.000286	0.000884	2.75
	4827	-155	23.13	28.15	5.02	0.000059	0.000385	3.61
	4847	-135	25.42	29.63	4.21	0.005384	0.007842	2.15
	4861	-121	23.63	28.38	4.75	0.000774	0.001743	2.93
	4909	-73	21.81	26.71	4.9	0.00182	0.003426	2.42
	5071	+89	17.58	22.6	5.02	0.012705	0.01694	2.06
	5089	+107	23.64	30.46	6.82	0.000817	0.001743	2.39
	5091	+109	20.85	25.73	4.88	0.005392	0.007842	2.04
	5127	+145	27.69	34.02	6.33	0.000127	0.000579	2.8
	5134	+152	27.4	33.85	6.46	0.000042	0.000385	3.2
	5139	+157	27.4	34.5	7.1	0.000015	0.00024	3.52
	5163	+181	29.02	33.9	4.88	0.000254	0.000884	2.66
	5192	+210	30.13	35.83	5.71	0.000173	0.000693	2.66
	5200	+218	29.88	34.88	5	0.000435	0.00116	2.41
	5203	+221	29.71	34.88	5.17	0.001165	0.002331	2.3
	5215	+233	29.73	34.42	4.69	0.000742	0.001743	2.39
	5218	+237	32.42	38.77	6.35	0.000304	0.000884	2.81
	5220	+239	30.02	34.98	4.96	0.000077	0.00041	2.9
5223	+242	33.69	40.19	6.5	0.000002	0.000074	3.56	
5252	+271	36.46	42.19	5.73	0.00006	0.000385	2.86	

**Table 19 – PBMC sequencing descriptive statistics of the differentially methylated CpGs in *LDLR*:** The sequenced gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported.

#### 4.4.4 *CDKN2B* gene

A schematic representation of the *CDKN2B* gene promoter that was isolated and sequenced in this study is shown in Figure 35. This shows the amplicon sequencing coverage spanning the gene promoter upstream of transcriptional start site in addition to part of exon 1.

*CDKN2B* is located on the sense strand, opposite to the anti-sense non-coding RNA known as *CDKN2B-AS1*.

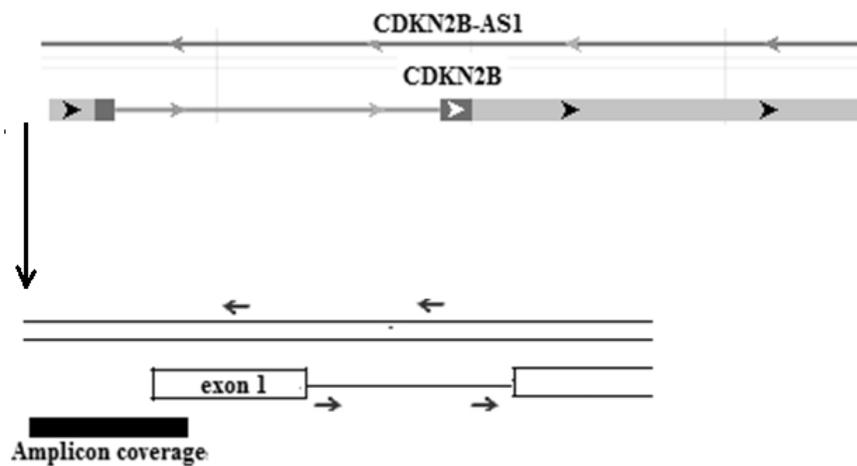
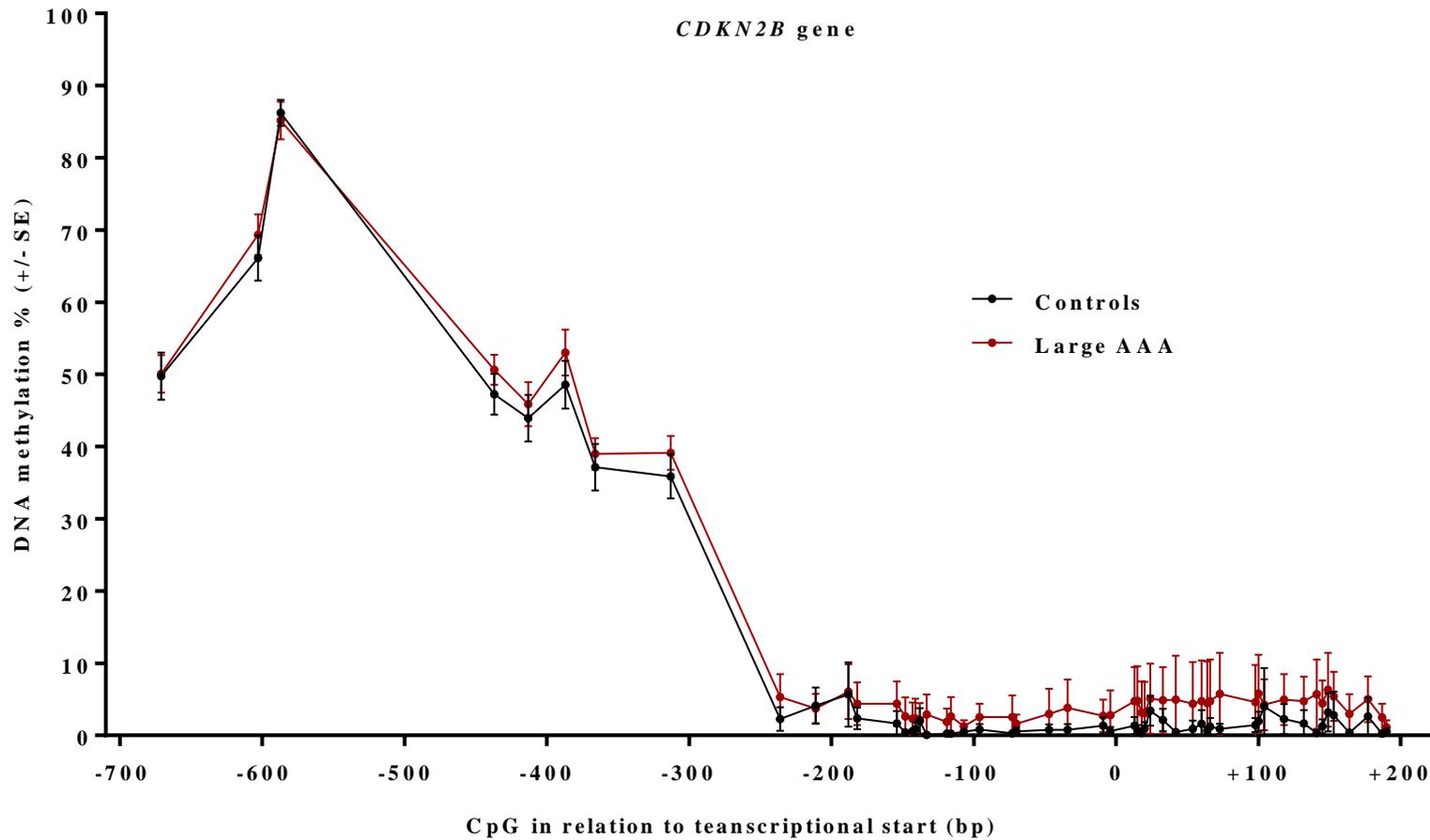


Figure 35 - Schematic representation of the sequenced region of the *CDKN2B* gene in PBMC DNA.

Two amplicons were designed to isolate and sequence the regulatory region of *CDKN2B*, which spanned a total of 53 CpG sites between the following genomic coordinates:

NG\_023297.1: 4309-5267.

The DNA methylation status of the entire sequenced promoter is shown in Figure 36. The methylation values were not significantly different in AAA and controls across the whole CpG Island.



**Figure 36 - DNA methylation status of the *CDKN2B* gene promoter in PBMC DNA:** DNA methylation status of 53 sequenced CpG sites in the *CDKN2B* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 4.4.5 *SPAG17* gene

A schematic representation of the *SPAG17* gene promoter that was isolated and sequenced in this study is shown in Figure 37. This shows the amplicon sequencing coverage spanning a CpG island in the gene promoter upstream and including the transcriptional start site.

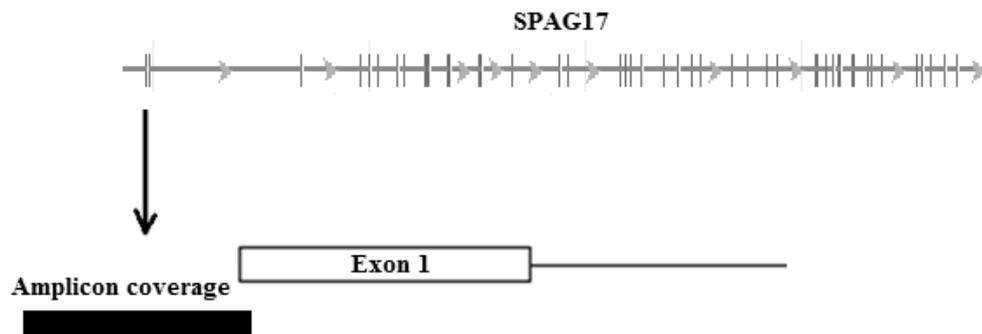
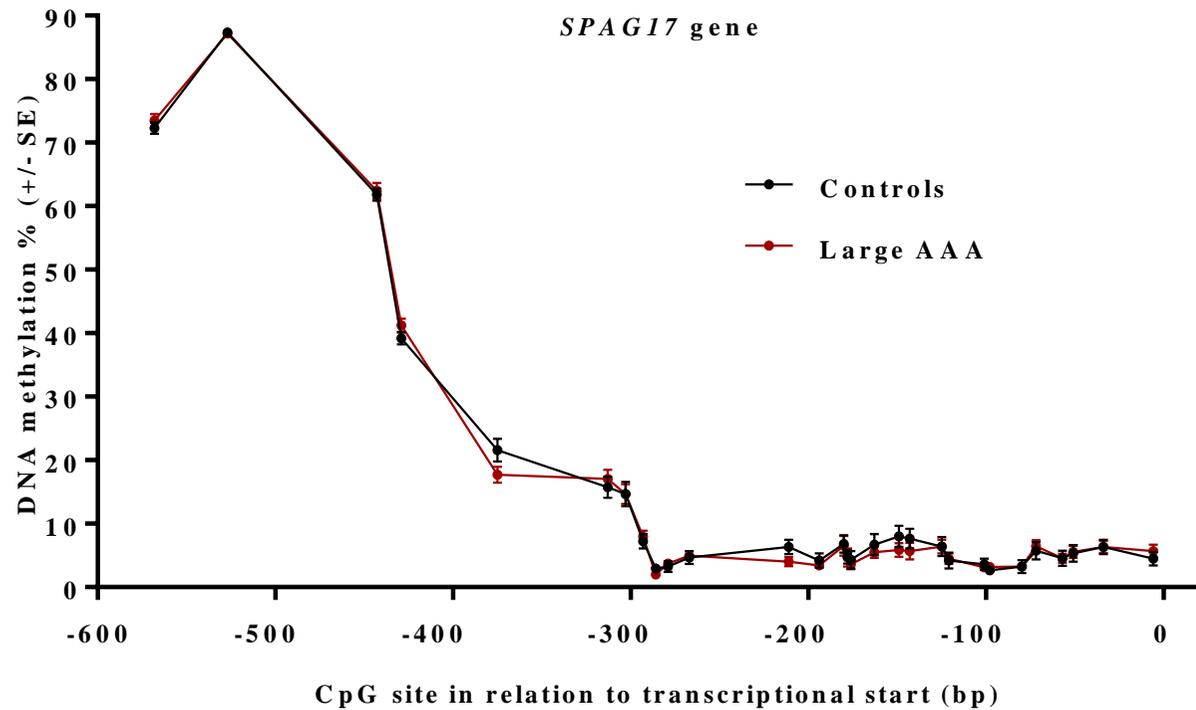


Figure 37 - Schematic representation of the sequenced region of the *SPAG17* gene in PBMC DNA.

Two amplicons were used to target and isolate the CpG Island with a total coverage of 29 CpGs spanning the region: NG\_053041.1: 4415-5105.

The DNA methylation status of the sequenced promoter is shown in Figure 38, where there were no statistically significant regions of differential methylation between those with AAA and controls.



**Figure 38 - DNA methylation status of the *SPAG17* gene promoter in PBMC DNA:** DNA methylation status of 29 sequenced CpG sites in the *SPAG17* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 4.4.6 *GDF7* gene

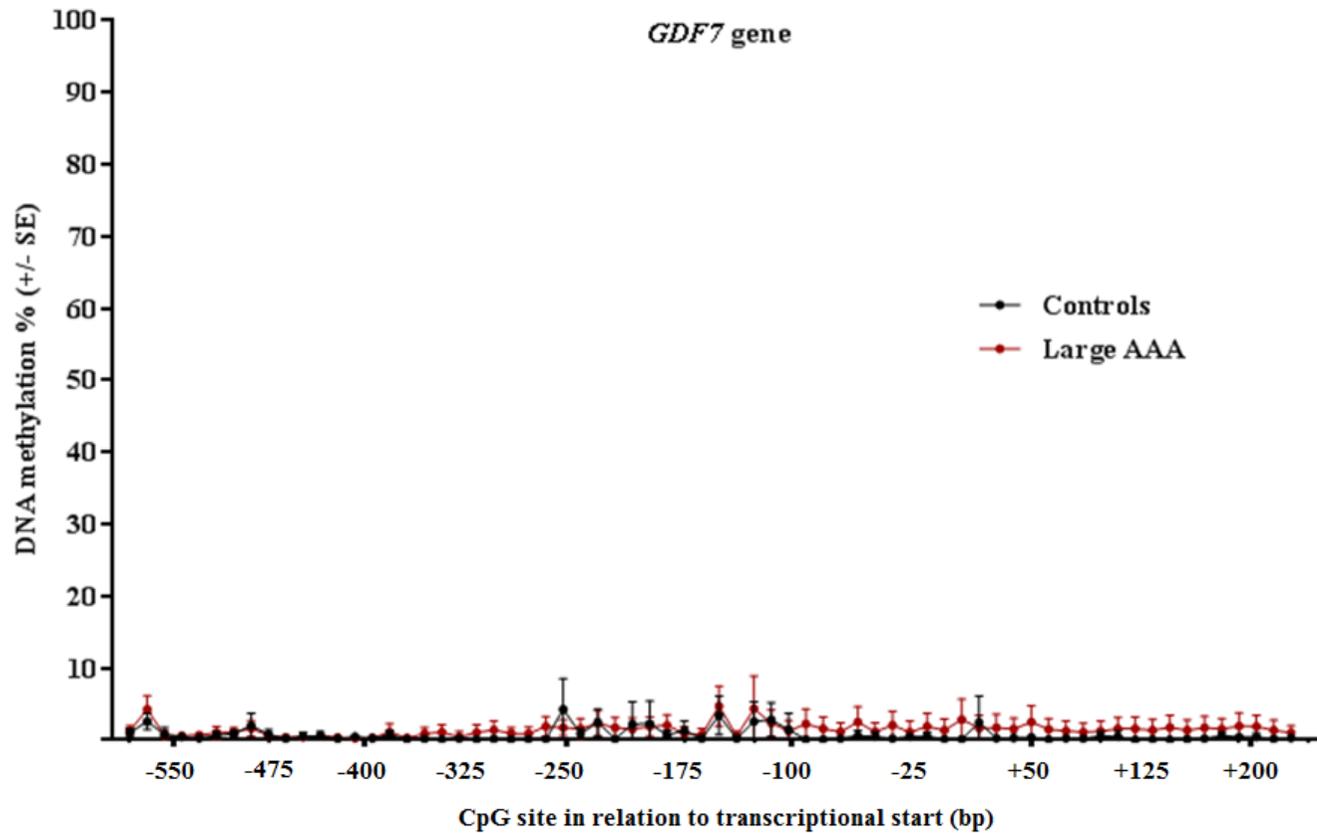
A schematic representation of the *GDF7* gene promoter that was isolated and sequenced in this study is shown in Figure 39. This shows the amplicon sequencing coverage spanning a CpG island in the gene promoter covering exon 1 and both intronic regions upstream and downstream of the exon.



Figure 39 - Schematic representation of the sequenced region of the *CDKN2B* gene in PBMC DNA.

A total number of 41 CpG sites across the whole CpG Island within the promoter and across the first exon of the *GDF7* gene was targeted with two sets of bisulphite specific primers, which covered the region: NC\_000002.12: 20650525-20651336.

The DNA methylation status of the sequenced region of the *GDF7* gene is shown in Figure 40. There were low levels of DNA methylation in the *GDF7* promoter CpG Island, and there were no statistically significant differences in methylation between AAA and controls.



**Figure 40 - DNA methylation status of the *GDF7* gene promoter in PBMC DNA:** DNA methylation status of 41 sequenced CpG sites in the *GDF7* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 4.4.7 *MMP9* gene

A schematic representation of the *MMP9* gene region that was isolated and sequenced in this study is shown in Figure 41.

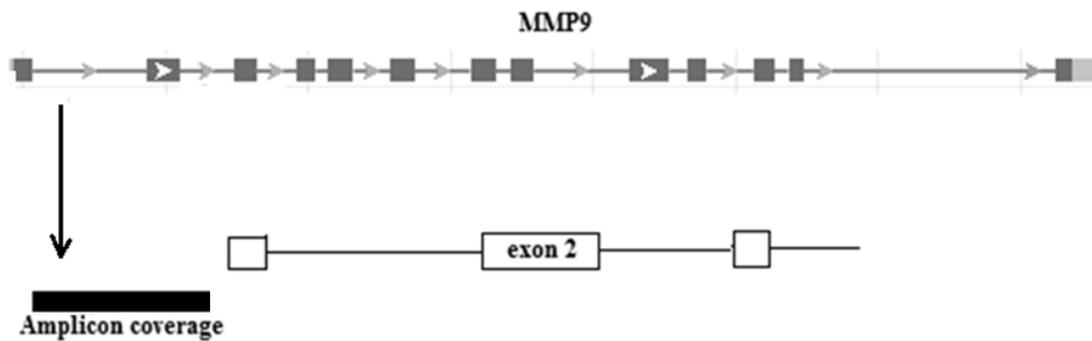
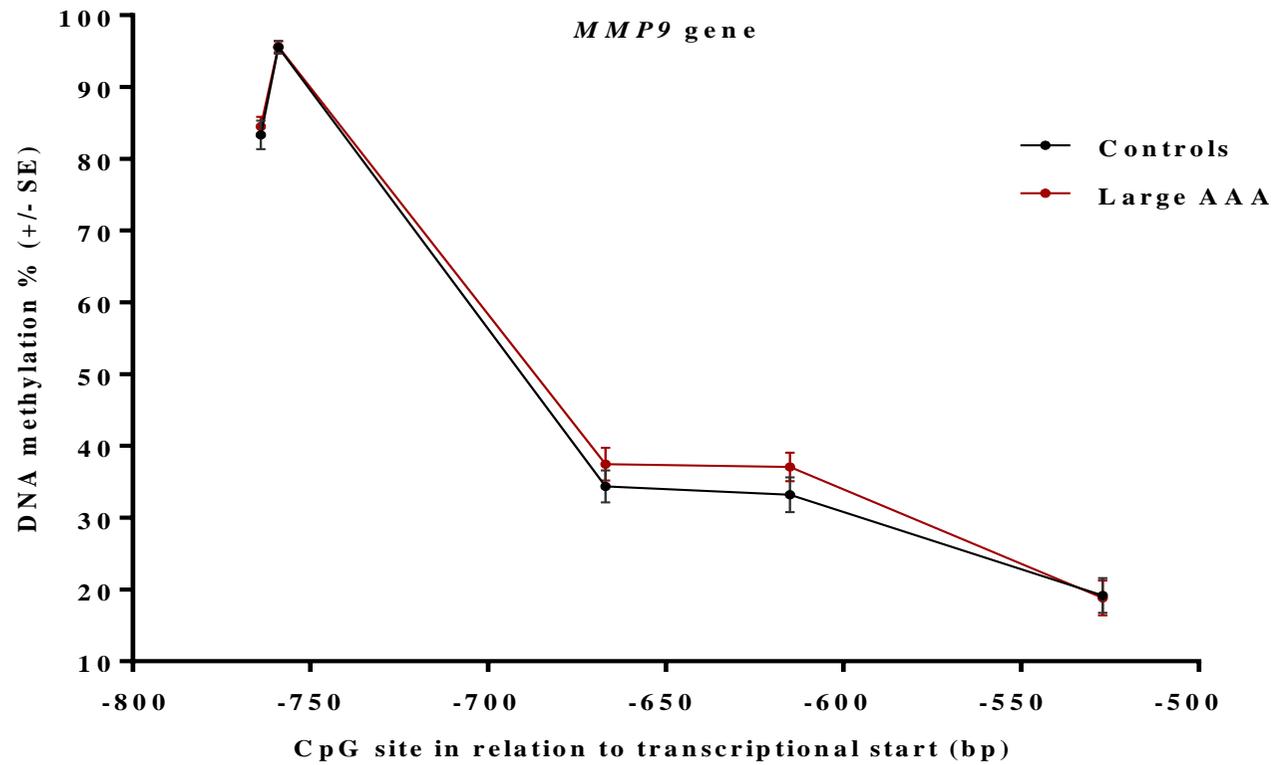


Figure 41 - Schematic representation of the sequenced region of the *MMP9* gene in PBMC and VSMC DNA.

One amplicon spanning 5 CpG sites was successfully used to sequence the regulatory region of the *MMP9* gene, which is located at NG\_011468.1: 4040-4416.

The DNA methylation status of the sequenced promoter is shown in Figure 42. The methylation values in *MMP9* were not significantly different between AAA and controls at any CpG site.



**Figure 42 - DNA methylation status of the *MMP9* gene promoter in PBMC DNA:** DNA methylation status of 5 sequenced CpG sites in the *MMP9* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 4.4.8 *ERG* gene

A schematic representation of the *ERG* gene region that was isolated and sequenced in this study is shown in Figure 43. This shows the amplicon sequencing coverage spanning the regulatory CpG Island in the intron upstream to exon 12, which also stretches in to exon 12.

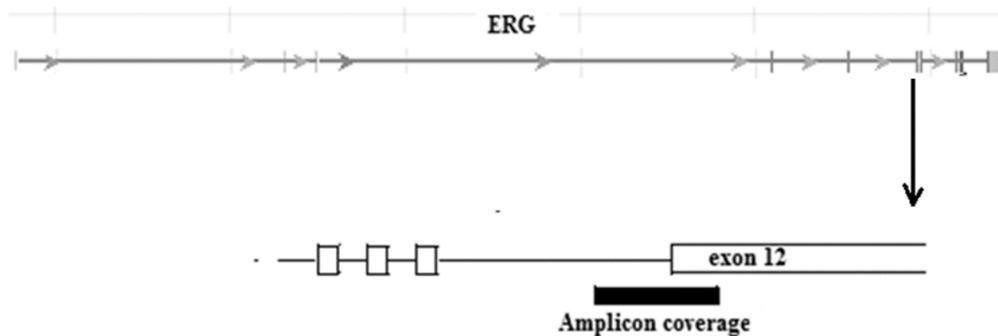
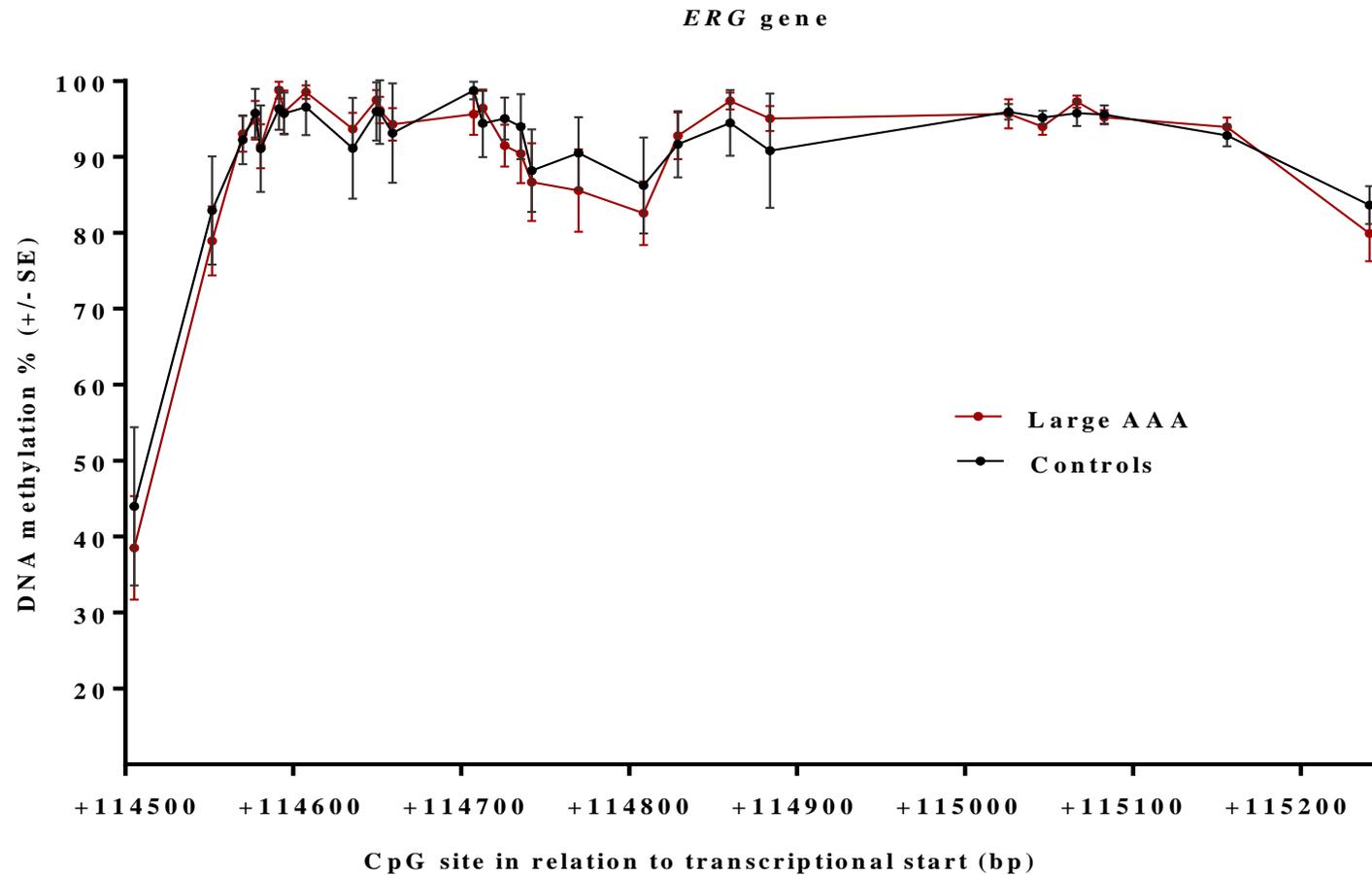


Figure 43 - Schematic representation of the sequenced region of the *ERG* gene in PBMC DNA.

Two amplicons were designed to isolate and sequence the regulatory region of *ERG*, spanning a total of 28 CpG sites between the following genomic coordinates: NG\_029732.1: 282787-283680.

The DNA methylation status of the sequenced promoter is shown in Figure 44. There were no statistically significant sites of differential methylation in this sequenced promoter, although there were clear visual fluctuations.



**Figure 44 - DNA methylation status of the *ERG* gene promoter in PBMC DNA:** DNA methylation status of 28 sequenced CpG sites in the *ERG* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 4.4.9 *LRP1* gene

A schematic representation of the *LRP1* gene region that was isolated and sequenced in this study is shown in Figure 45. This shows the amplicon sequencing coverage spanning the gene promoter shortly upstream of, and including, the exon 1 transcriptional start site.

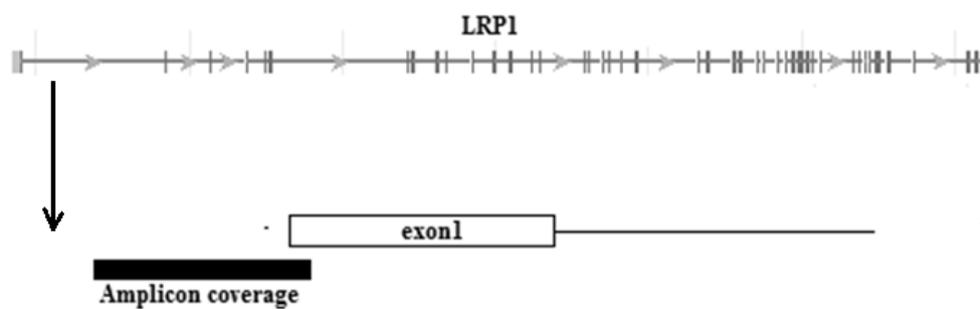
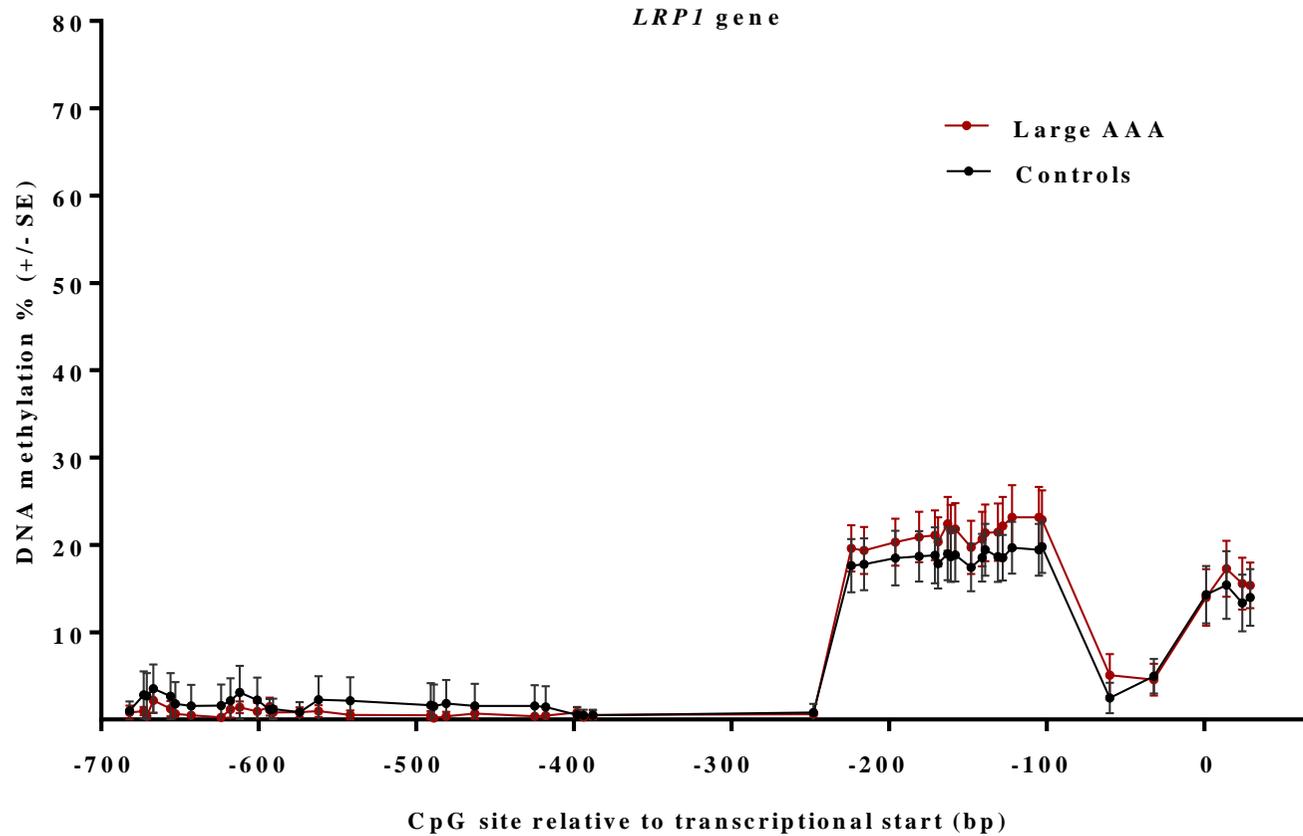


Figure 45 - Schematic representation of the sequenced region of the *LRP1* gene in PBMC DNA.

Two amplicons were designed to isolate and sequence the regulatory region of *LRP1*, spanning a total of 33 CpG sites between NG\_009060.1: 4376-5355.

The DNA methylation status of the sequenced promoter is shown in Figure 46. After statistical analysis of each methylation value of each CpG in AAA and controls, there were no significant differences in methylation at any region of the *LRP1* gene promoter.



**Figure 46 - DNA methylation status of the *LRP1* gene promoter in PBMC DNA:** DNA methylation status of 33 sequenced CpG sites in the *LRP1* gene isolated from the peripheral blood of 48 AAA and 48 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## 4.5. Conclusion

This study is novel, and is the first to assess DNA methylation in PBMC DNA from people with AAA using bisulphite NGS, which is the gold standard methodology for studies of DNA methylation<sup>191</sup>. The results presented in this chapter demonstrate further evidence that there is an epigenetic basis to AAA, and is proof of concept for larger scale analysis in the future. It was identified in this study that significant differentially methylated CpGs exist in the regulatory regions of *LDLR*, *SORT1* and *IL6R* in PBMC DNA from people with AAA compared to controls.

By analysing the epigenetic status of the peripheral blood, it allows the direct identification of potential sites that may be responsible for the regulation of biological pathways potentially involved in adverse inflammatory processes that contribute towards AAA development. PBMC DNA is derived from the whole blood, and is an important, easily obtainable source of material. Several inflammatory cells which circulate in the blood have been demonstrated in AAA biopsies, including; T and B lymphocytes, macrophages, neutrophils and mast cells<sup>27</sup>. Specifically, monocytes/macrophages are involved in chronic vascular remodelling and subsequent AAA expansion<sup>19, 28</sup>. MMP-9 is a major secretion product of macrophages and increased MMP-9 activity is responsible for the degradation of the ECM and weakening of the aortic wall<sup>29</sup>. Monocyte activity is modulated by pro-inflammatory cytokines, growth factors and chemokines.

The functional roles of the genes identified as differentially methylated in this study are well characterised. The *IL6R* protein is a common pleiotropic cytokine receptor that induces pro-inflammatory signalling in addition to cell growth and differentiation, whereas the *LDLR* and *SORT1* proteins are well known cell membrane receptors involved in the transport and

degradation of lipids during lipid metabolism<sup>45, 79, 80</sup>. Both inflammation and lipid metabolism are pathologically linked to AAA development. However, the exact mechanistic manner in which the polymorphisms located within these genes, identified in previous GWASs, predisposes to the development of AAA is still not fully understood. For instance, polymorphisms within *LDLR* and *SORT1* are intronic, and suggest regulatory role of action. Further investigation could yield valuable insight for potential meQTL association, as it is plausible that genetic variation at these AAA risk loci regulate methylation and subsequent adverse gene functionality.

There are significant limitations to the work presented in this Chapter. Due to the manner in which the UKAGS samples are stored after collection (patient blood samples are centrifuged and frozen), it was not possible to isolate specific peripheral blood cells for the sequencing of homogenous cell lines (e.g. monocytes or lymphocytes), because after freeze-thaw cycles, cells burst and cannot be isolated or cultured. In addition, the UKAGS does not measure blood cell counts, and it was not possible to retrospectively acquire cell counts, meaning there was no way to proportionately determine how many of which mononuclear cells were present in each sample, and subsequently no way to tell how much DNA from each cell type there was after each extraction. This is said to be an important factor in epigenetic studies, as heterogeneity in white blood cells has the potential to confound DNA methylation measurements<sup>192</sup>, even though peripheral blood cell DNA has been an essential source for genetic experiments for the past decades. Adalsteinsson *et al.*, (2012)<sup>192</sup> state that tissues are ensembles of cells that may each have their own epigenetic profile, and therefore inter-individual cellular heterogeneity may compromise these studies, even in peripheral blood<sup>192</sup>. There may be some likelihood that this factor could confound DNA methylation levels to some extent in this study, but not to the extent where methylation changes would not be

#### **Chapter 4: Peripheral blood cell DNA methylation analysis in AAA**

observed in the whole PBMC population, as clearly demonstrated by the work conducted in this chapter. The data presented is therefore limited, as each cell type has its own independent epigenetic profile, but does still represent novel data that requires further investigation, especially with regards to the isolation of homogenous cell lines to assess DNA methylation in AAA.

The work presented in this chapter utilised the best available methodologies to address the hypothesis driven research question of whether specific CpG methylation changes are present in genes that are associated with AAA at genomic risk loci, yielding novel data. The next chapter will present data that was acquired using the same methodological approach, but was conducted on VSMC DNA to establish the epigenetic profiles of cells derived directly from the aortic wall at AAA genomic risk loci.

**5. Chapter 5 – Results section 3**

**Vascular smooth muscle cell DNA methylation analysis in AAA**

## 5.1. Introduction

The previous chapter (Chapter 4) of this thesis was based on the CpG methylation analysis of PBMC DNA using bisulphite NGS. The results illustrated novel associations between differential CpG island methylation in the regulatory regions of genes located at AAA genomic risk loci in individuals with AAA compared to controls (*IL6R*, *SORT1* and *LDLR*). Further functional investigation is required; however it is feasible that these methylation changes contribute towards the pathobiology of AAA.

It is well known that different cell types have independent epigenetic profiles which are involved in phenotypic diversity and can contribute towards disease<sup>125, 126, 189, 193</sup>. This chapter will therefore adopt the same methodological approach and experimental analyses as Chapter 4, but will instead concentrate on the methylation profiles of VSMC DNA isolated from the aortic tissues of individuals with AAA and compared to cadaveric organ donors used as controls.

VSMCs are the main constituents of the tunica media, where cellular apoptosis occurs as a hallmark of AAA during inflammation, ECM degradation and vascular remodelling during aneurysm development and growth<sup>33</sup>. This analysis will determine if any of the same or alternative methylated regions observed in PBMC DNA also exists in VSMC DNA taken directly from the site of aneurysm. Nobody has ever studied the methylation status of VSMCs in AAA prior to this work, which essentially allows an alternative perspective to the potential pathobiology of AAA from what can be established from PBMCs.

### **5.1.1. Aims**

1. To identify whether changes in VSMC DNA CpG methylation exist in the regulatory regions of genes (promoters and transcriptional start sites) proximal to AAA genomic risk loci in those with AAA compared to healthy controls.

### **5.1.2. Objectives**

1. Extract VSMC DNA (isolated from aortic tissues) from 24 AAA and 20 controls and isolate the regulatory regions of 9 candidate genes proximal to AAA genomic risk loci using bisulphite specific PCR in each of the 44 samples. Targeted genes include: *LRP1*, *ERG*, *SORT1*, *LDLR*, *IL6R*, *MMP9*, *DAB2IP*, *SMYD2* and *SERPINB9*.

2. Conduct library preparation for sequencing, which involves ligating unique adapter sequences to each individual DNA sample prior to multiplexing.

3. Run samples for quality check on bioanalyzer and conduct bisulphite specific next-generation sequencing on the Illumina Mi-seq platform.

4. Use a bioinformatics platform (command line) and open source software to analyse raw sequencing data. Extract methylation values (normalised as %) at each CpG in each gene for each individual sample after applying high quality data filters.

5. Statistically analyse the extracted methylation data using multiple t-test analysis to establish whether changes to CpG methylation exist between AAA and controls.

### **5.1.3. Hypotheses**

1. Is hypothesised that significant differences in CpG specific DNA methylation will be observed in those with AAA compared to controls at multiple CpG sites in several targeted genes.

## 5.2. Population

In this Chapter, bisulphite NGS was performed to assess the methylation status of CpG islands in the regulatory regions of genes associated with AAA in VSMC derived DNA from 44 individuals (Table 20). VSMCs were isolated from aortic biopsies taken from individuals who had undergone open AAA surgery (n=24) and cadaveric controls (n=20) using explant culture. DNA was then extracted from isolated VSMCs and used for bisulphite methylation analysis. Due to the difficulty in obtaining sufficient numbers of aortic tissues from people, sample selection was limited and it was not possible control for gender, smoking and age as effectively as the PBMC sequencing study described in Chapter 4. Nor was it possible to obtain a higher sample number.

	24 Large AAA (>5.5cm)	20 Cadaveric controls (Exact aortic diameters unknown)
No of males	24	16
No of females	0	4
Ethnicity	Caucasian	Caucasian
Smoker	22	15
Median age (IQR)	68 (65-73)	56 (49-62)
Mean aortic diameter (mm (+/- SD))	6.5 (1.2)	<3

**Table 20 - Demographic summary of samples where VSMC DNA was used for bisulphite NGS.**

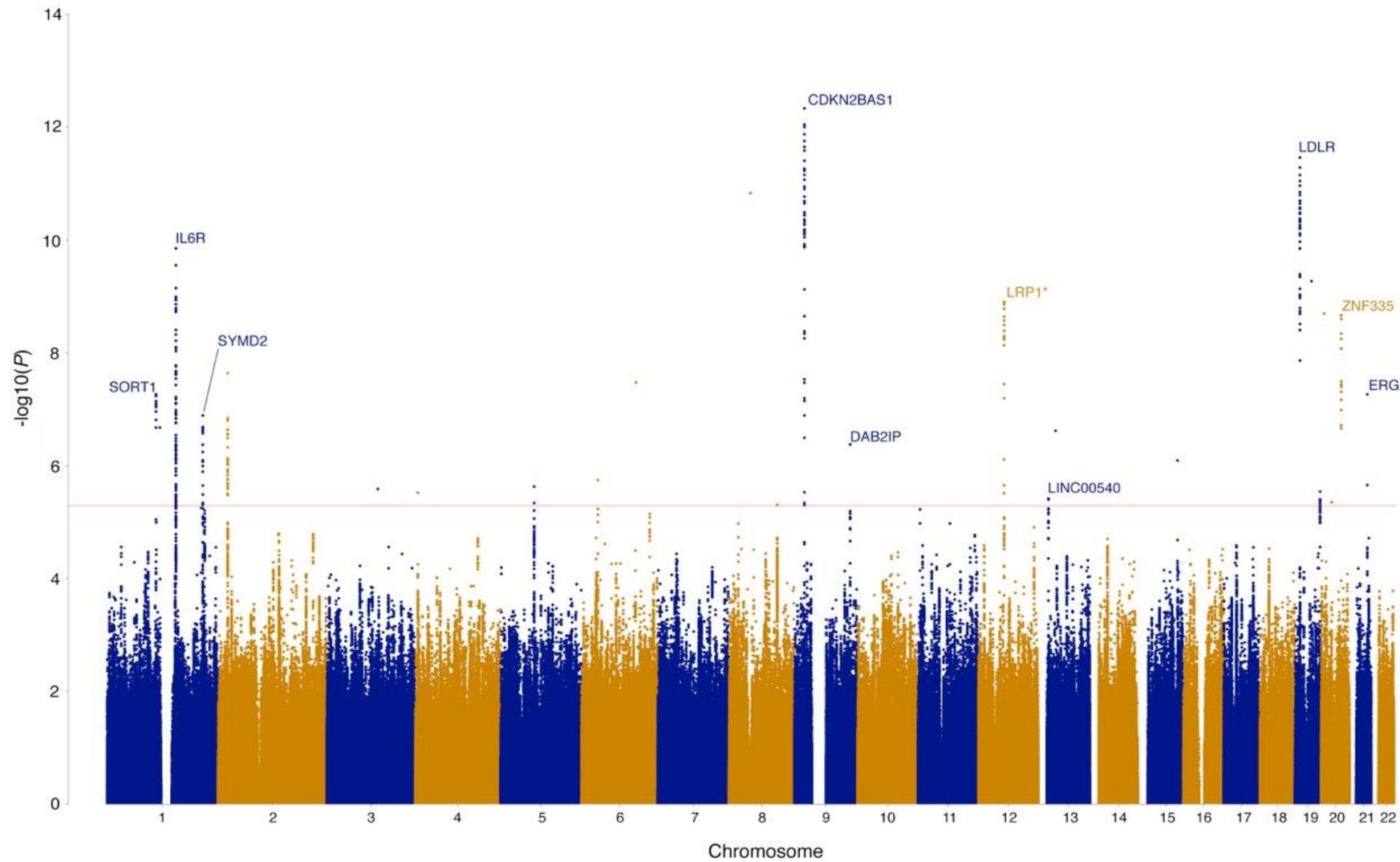
### 5.3. Genes for VSMC sequencing assay

Following from the PBMC sequencing assay in which the regulatory regions of 9 genes were bisulphite sequenced, the same analysis was performed on DNA isolated from VSMCs. However, for this assay, the complete GWAS meta-analysis results previously discussed had become available. The final GWAS results did not include *SPAG17* or *GDF7*, indicating that they were previously false positives. These genes were subsequently dropped from this Chapter. In addition, *CDKN2B* was also excluded since it was mistakenly included in the previous experiment (Chapter 4) in place of *CDKN2B-AS1*. *CDKN2B-AS1* was then also not included in the VSMC sequencing assay because a regulatory region could not be identified.

As a result, the final list of candidate genes included in Chapter 5 are as follows: *LRP1*, *ERG*, *MMP9*, *LDLR*, *IL6R*, *SORT1*, SET and MYND Domain Containing 2 (*SMYD2*), Serpin family B member 9 (*SERPINB9*) and DAB2 interacting protein (*DAB2IP*).

*DAB2IP* was not known to me at the time of the PBMC assay (Chapter 4), but was a known gene associated with AAA. *SMYD2* is a novel AAA locus identified in the GWAS meta-analysis. *SERPINB9* was not identified in any GWAS but was identified during an EWAS published by Ryer *et al.*, (2015)<sup>143</sup>. Due to this, the regulatory regions of *DAB2IP* and *SMYD2* were targeted for the VSMC sequencing assay, and so was the previously observed differentially methylated gene body CpG Island in *SERPINB9*.

In summary (Figure 47), the Manhattan plot extracted from the Jones *et al.*,<sup>83</sup> recent GWAS meta-analysis of AAA illustrates the most up to date analysis of genomic risk loci associated with AAA. In addition, the specific targeted loci for this Chapter are displayed in Table 21.



**Figure 47 - Manhattan plot from GWAS meta-analysis:** Polymorphic loci associated with AAA identified in the GWAS meta-analysis (Jones *et al.*,<sup>83</sup>) were used as a guide to select candidate genes for Chapter 5 of this PhD, except for *CDKN2B-AS1* and *LINC00540*. The locus identified as *ZNF335* was included, but as *MMP9*.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

Gene symbol	Chromosome and RefSeq ID	RefSeq coordinates	Primer sequence 5'-3'	Annealing temp (°C)	CpG coverage
<i>LRP1</i>	Chr: 12 NG_016444.1	3875-4240	F: TAGAAGGGGGTAGTGATTA AAAAGTA R: AAAACAAACCCTAATTTAAAAAAA	54	13
		4274-4640	F: GGGTATTAAGGTGGGTTTTATTT R: TACTCTAAAATTTCAAACCTCCCTCC	57	25
		4680-5036	F: GTATTAGGGAGGAGGTTTAGTTAG R: TCCTCAATACATAAACTAAAACTC	54	24
<i>ERG</i>	Chr: 21 NG_029732.1	282787-283319	F: TTTTATTAGGTAGTGGTTAGATTTAGTTTT R: TACCCCAATTAATAAATTC CAATATA	54	22
		283331-283680	F: TTGGAATTTATTAATTGGGGGTATATAT R: ACACTATCTTTTACAAAATCAATCCAC	57	6
		5162-5540	F: AAAATTTTGGAAAGGGGTTTAGTT R: ATAATATTTTCCAACCTCATTA AAAAC	56	37
<i>MMP-9</i>	Chr: 20 NG_011468.1	3980-4579	F: TTGGGTTAAGTAATTTTTTTATTT R: TAACCCATCCTTAACCTTTTACAAC	52	7
		4640-5600	F: GATGGGGGATTTTTTTAGTTTATT R: TACCATTCTAACCATCACTACTC	57	8
<i>LDLR</i>	Chr: 19 NG_009060.1	4376-4725	F: TTTTAAAGGGGAGAAATTAATATTTA R: CACAAAAAATAACAACAACCTTTC	54	5
		4776-5355	F: GAAAGGTTGTGTATTTTTTTGTG R: AAACCTCCCTCAACCTATTCTAAC	57	28
		5464-5972	F: TTTAAGTTTTATAGGGTGAGGGAT R: ACACCCAAC TCAAAATAACAATAAC	56	31
		6054-6785	F: AATTTTATGGGTGTAGTTAATAGGTTAT R: CTCAAAATCATACTAACCAACCTC	54	49
<i>IL6R</i>	Chr: 1 NG_012087.1	2781-3061	F: TTTTGTTTAGGTTGGAGTGTAGTG R: CAAAATTTAATATTATAATTCACATAAAA	52	8
		3095-3729	F: GTTGTTTGGTGTAGAGATAGGTG R: ATAAAACTCCCAAATAAAACAAAACC	58	7
<i>SORT1</i>	Chr: 1 NG_028280.1	4363-4804	F: AGATTATTTTTAGGTTGTAGGAGTTA R: AACAAAACTACTAAAATCAACCCC	55	24
<i>SMYD2</i>	Chr: 1 NC_000001.11	214280141-214280630	F: TTTTATTTGAAGTAGTGGTTTTTGTAA R: TTTTCCAAAATTA AAAATTTTAAACCT	55	13
		214280631-214281330	F: AAGGTTTAAAAATTTTTAATTTGGAAA R: AAATCCCCACCTAAAAAACTAC	58	82
		214281331-214281733	F: TTTTTTAGGTGGGGGATTT R: CCCACATTTAAAAACAAAACCTAC	55	46
<i>SERPIN9</i>	Chr: 6 NC_000006.12	2891544-2892079	F: GGTGTTGAAAATTTTTGGAGGTA R: AAAATCTACCATTCACTAACTAACAAA	57	26
		2890564-2891263	F: TTAGAGGGTGGGATTAGAGGTAGTT R: TTTCCACCTAAAAAACCAAAATTA	54	9
		2889934-2890423	F: GTATTTGGGAATTTGTTGATGTTTT R: ACAACAAAAAATCATTATAATCTATTTCT	55	11
		2889425-2889933	F: AGAATTTTATATGTAATTTATTTTGG R: TATAATCCCAACACTTTAAAAAAC	54	31
<i>DAB2IP</i>	Chr: 9 NC_000009.12	121689899-121690502	F: TTGGAGTAGTTTGTGTTGTGTT R: TAAAAAATTAATAACCCAAAACCTC	57	10
		121690573-121691062	F: TTGTTATATTTAAGTTGAGATTTTGGG R: TAACCACATAAAAAACATTCCAACA	57	6
		121767430-121767872	F: GGAGAGGAGATAGGAAAGTTTTTAG R: CTAACCTTATCCCTTACAACCACTAC	57	6
		121768273-121768692	F: AATGGAGTGGGAGTTTGTATAGTG R: TAATAACAAC TTTAACCCACCTC	57	11

**Table 21 - Candidate genes for vascular cell DNA bisulphite sequencing:** Genomic locations, primer sequences, annealing temperatures and CpG coverage are displayed for each amplicon.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

The *CNN2* gene was previously observed as differentially methylated in the Ryer *et al.*,<sup>143</sup> EWAS, and was also targeted in this project, similarly to *SERPINB9*. However bisulphite specific PCR primers could not be successfully designed due to the repetitive nature of the target sequence.

Unsuccessful primers for other targeted gene loci include regions within *LRP1*, *ERG*, *LDLR*, *IL6R*, *SERPINB9* and *SMYD2*. Similarly, to the previous Chapter, the optimisation of those amplicons included in the study (Table 21), in addition to those which I could not get to work (Table 22) after extensive testing, took many weeks.

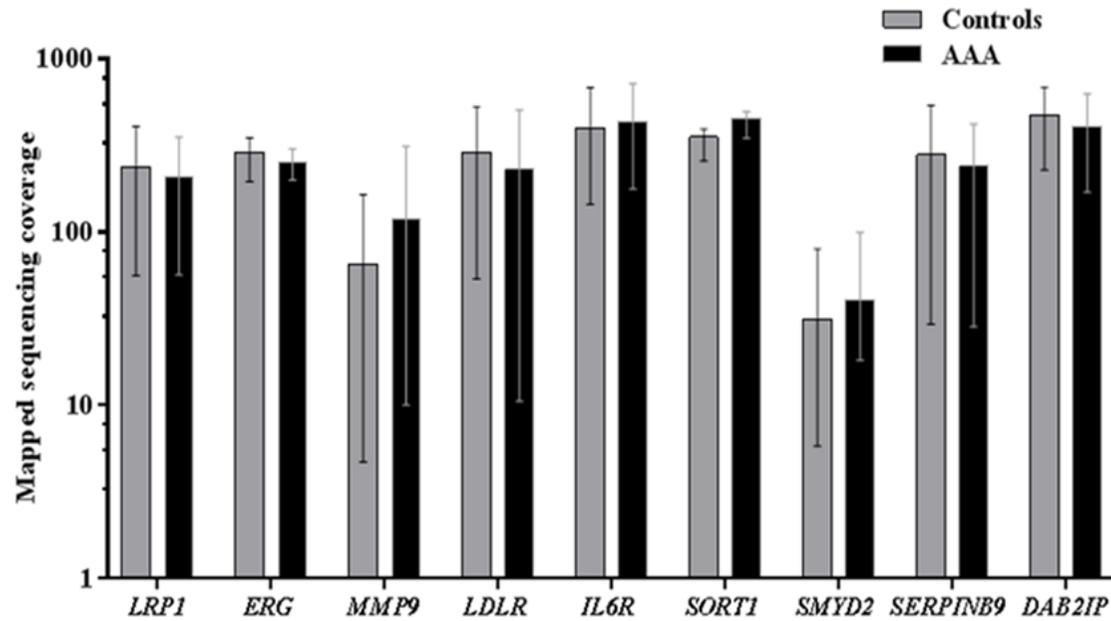
Gene Symbol	Primer sequence 5'-3'
<i>LRP1</i>	F: AGAAGGGGGTAGTGATTAAGTA R: CCTAACTAAACCTCCTCCCTAATA
<i>LRP1</i>	F: GTTTTAGGTTTATGTATTGAGGAGG R: ATAAAAAACCTCTTCTCTACCC
<i>ERG</i>	F: TTGTTTGTTGAAATTTTTGGTTGT R: AAAAATAATATTTTTCCAACCTCATTAA
<i>ERG</i>	F: ATGAGGTTGGAAAAATATTATTTTT R: AAACCCTACAAAATAAACCCCTAAA
<i>LDLR</i>	F: TTAGAATAGGTTGAGAGGGAGTTTT R: ACACCCAACCTCAAATAACAATAAC
<i>IL6R</i>	F: TTTATTTTAAGGGTAAGAGGAATAT R: AAAAACACAATCCTATACACAAACC
<i>SMYD2</i>	F: TTTTTTAGGTGGGGGATTTGT R: CCCAAACTTAAAACAAACCTAACTC
<i>SERPINB9</i>	F: GAATTTTATATGTAATTTATTTTGG R: TCTATCCCCTTATATCCCAATAC
<i>CNN2</i>	F: GGAGGTTGAGTTTAGGAGTTAAGAG R: ATCTACACCAAACAATATATCCCAC
<i>CNN2</i>	F: GTGGGATATATTGTTTGGTGTAGAT R: ACCACAAAAACATCTAAAAACACC

Table 22 - Unsuccessful primers designed for bisulphite PCR at targeted gene loci.

#### **5.4. DNA methylation status of candidate gene regions in VSMCs**

After bisulphite sequencing (described in Chapter 2) of the regulatory regions of 9 candidate genes in 44 VSMC DNA samples (24 AAA and 20 controls), data analysis was performed, which included bioinformatics and data quality filtering of sequencing reads at; base quality score of >Q20, read quality of >Q50 and read depth of >5.

Total mapped sequencing coverage for this assay was consistently high in each gene and there were no significant differences in mapped sequencing read depth between AAA and controls in any of the genes (Figure 48), suggesting there was no sequencing bias between AAA and controls, and the final data had a high level of accuracy and reliability.



**Figure 48 - Sequencing coverage of candidate genes in VSMC DNA:** Mean coverage of high quality sequencing reads in candidate genes from VSMC DNA bisulphite sequencing of 24 AAA and 20 controls. Bars represent standard deviation.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

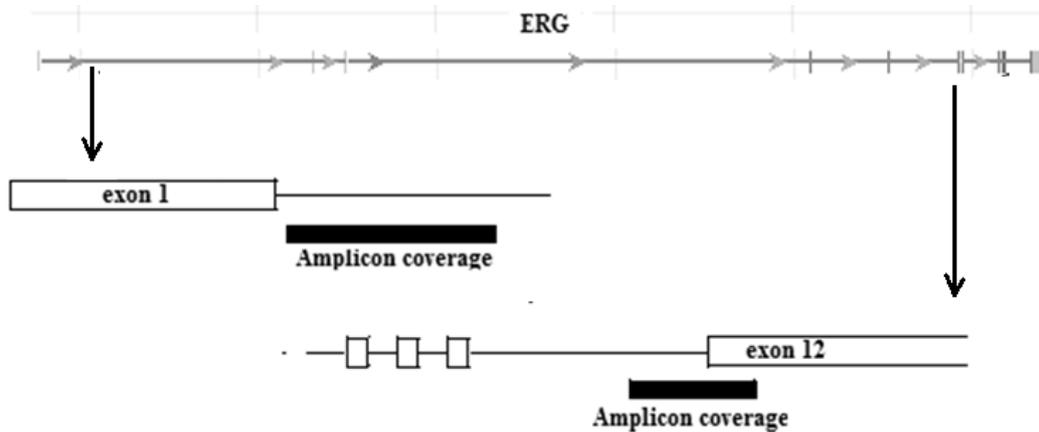
The methylation status (normalised as percentage) of each CpG, in each gene, in each of the 44 individual samples was determined and statistically assessed. Statistical significance was based at  $P < 0.05$  and a false discovery rate of  $Q < 10\%$  after multiple t-test comparisons.

Methylation profiles of AAA and controls were compared, and significant differential methylation was observed at many CpGs in 4 genes (*IL6R*, *ERG*, *SERPINB9* and *SMYD2*). No significant differences in methylation were observed in any of the other genes (*LDLR*, *LRP1*, *SORT1*, *MMP9*, *ERG* and *DAB2IP*).

The methylation sequencing results from each gene will now be described individually.

### 5.4.1. *ERG* gene

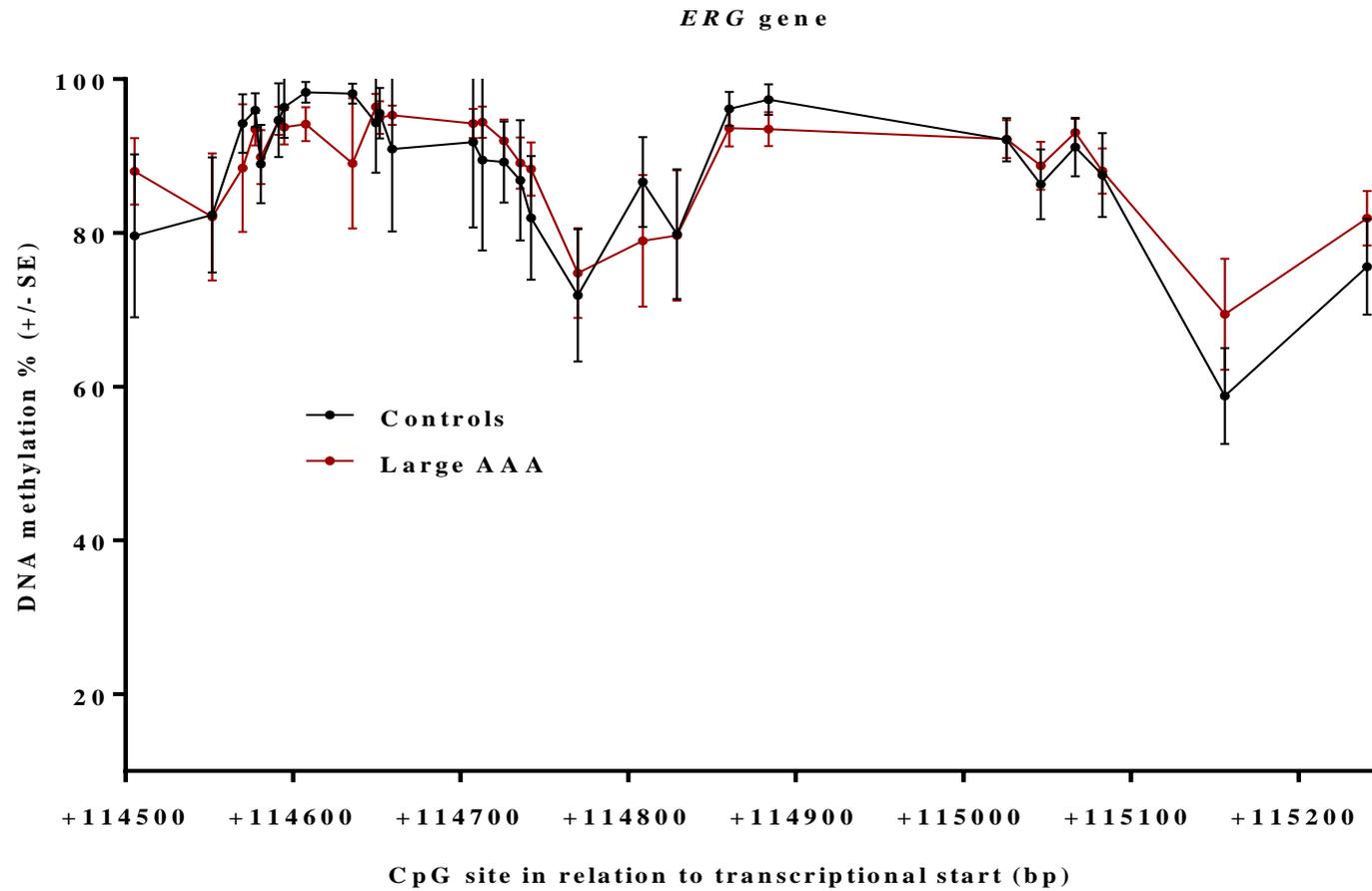
A schematic representation of the *ERG* gene loci that were isolated and sequenced in this Chapter is shown in Figure 49. Amplicon coverage was increased from the previous Chapter, which only sequenced the CpG Island near exon 12. The newly targeted region in the current experiment is located in the intron directly downstream from the first exon. This region was included in this experiment after I realised there were 2 known promoters for different transcript variants of *ERG*.



**Figure 49 - Schematic representation of the sequenced region of the *ERG* gene in VSMC DNA.**

As a result of the inclusion of the second *ERG* gene promoter, which was targeted with a single amplicon, a total of three amplicons were used to isolate and sequence both promoter regions. The promoter near exon 12, as in the previous Chapter, spanned a total of 28 CpG sites at NG\_029732.1: 282787-283680. The new addition near exon 1 covered a total of 37 CpG sites at NG\_029732.1: 5162-5540.

The DNA methylation status of the promoter located near exon 12, which was sequenced in the previous Chapter in PBMC DNA, is shown in Figure 50, where there were no statistically significant differences in methylation between AAA and controls.



**Figure 50 - DNA methylation status of one of two *ERG* gene promoters in VSMC DNA:** DNA methylation status of 28 sequenced CpG sites in *ERG* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

The DNA methylation status of the promoter located near exon 1, which was not sequenced in the previous Chapter (in PBMC DNA), but was included for this experiment, is shown in Figure 51.

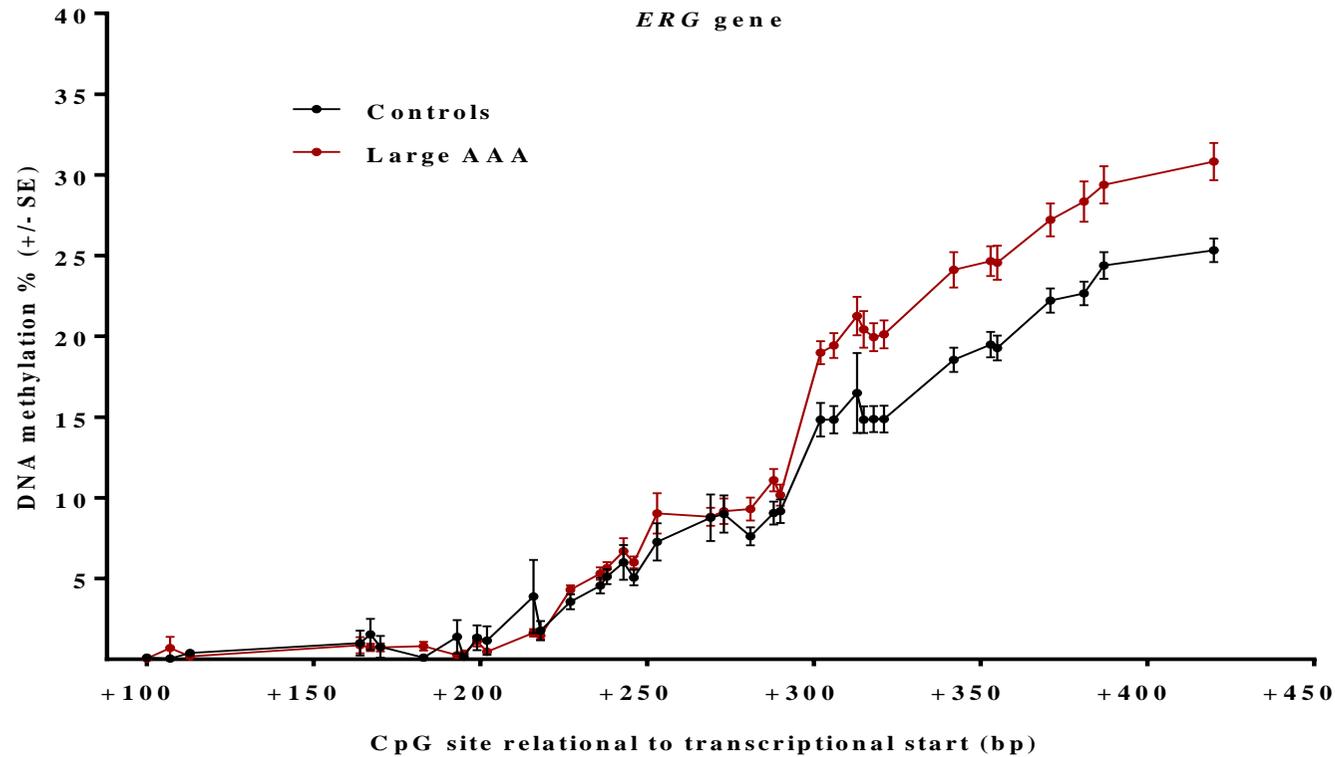
This Figure shows that a large portion of the newly targeted *ERG* promoter, which consists of 37 CpG sites at NG\_029732.1: 5162-5540, was considerably differentially methylated in those with AAA compared to controls.

Overall 13 CpGs were consecutively hypermethylated in AAAs compared to controls.

The mean methylation of these sites was  $18.67\% \pm 3.92$  in controls versus  $23.8\% \pm 4.11$  in cases with a mean increase in methylation of  $5.13\% \pm 0.43$  ( $P=0.0005$ ,  $Q=0.0014$ ).

The 13 CpGs in *ERG* span a region of 118bp (NG\_029732.1:g.5389\_5507|gom) which is located within an intronic region directly downstream from the first *ERG* exon.

Descriptive statistics of the differential methylation observed in *ERG* after bisulphite sequencing of VSMC DNA in 24 AAA and 20 controls are shown in Table 23.



**Figure 51 - DNA methylation status of the second of two *ERG* gene promoters in VSMC DNA:** DNA methylation status of 37 sequenced CpG sites in the *ERG* gene promoter sequenced in 24 AAA and 20 controls. 13 CpG sites were hypermethylated in AAA: (+302 (P=0.0019), +306 (P=0.0004), +313 (P=0.0002), +315 (P=0.0001), +318 (P=0.0001), +321 (P=0.0001), +342 (P=0.0003), +353 (P=0.0001), +355 (P=0.0002), +371 (P=0.0003), +381 (P=0.0015), +387 (P=0.0012) and +420 (P=0.0004)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

Gene and location	Region (CpG site)	Relative to start of transcription (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>ERG</i> Chr: 21 NG_029732.1	5389	+302	14.83	19	4.17	0.0019	0.0037	LS
	5393	+306	14.83	19.43	4.6	0.0004	0.001	LS
	5400	+313	16.5	21.26	4.76	0.0002	0.001	6.7
	5402	+315	14.83	20.43	5.6	0.0001	0.0009	LS
	5405	+318	14.89	19.96	5.07	0.0001	0.0008	6.18
	5408	+321	14.89	20.13	5.24	0.0001	0.0008	6.19
	5429	+342	18.56	24.13	5.58	0.0003	0.001	LS
	5440	+353	19.5	24.65	5.15	0.0001	0.0008	LS
	5442	+355	19.28	24.57	5.29	0.0002	0.0009	LS
	5458	+371	22.22	27.22	5	0.0003	0.001	LS
	5468	+381	22.67	28.35	5.68	0.0015	0.0032	LS
	5474	+387	24.39	29.39	5	0.0012	0.0027	LS
5507	+420	25.33	30.83	5.49	0.0004	0.001	LS	

**Table 23 – Descriptive statistics of the differentially methylated CpGs in *ERG* after VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported. LS = lost significance.

### 5.4.2. *IL6R* gene

The *IL6R* gene promoter, located directly upstream from the genes transcription start site, was sequenced in the previous Chapter in PBMC DNA. An illustration of the sequenced region is reproduced below in Figure 29. The exact same region was also sequenced in this Chapter in VSMC DNA from 24 AAA and 20 controls.

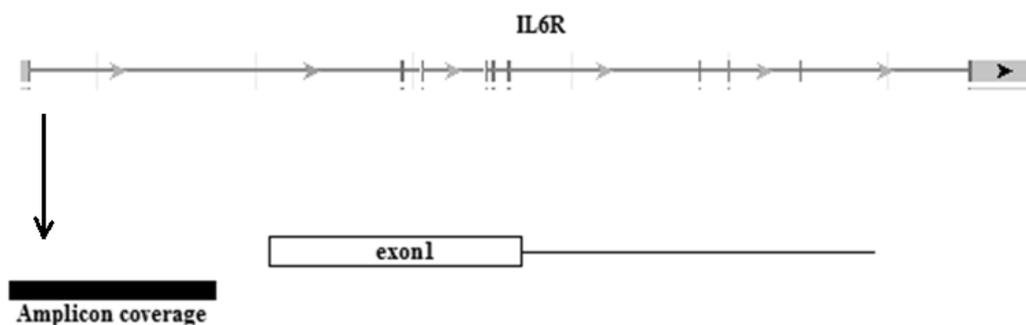
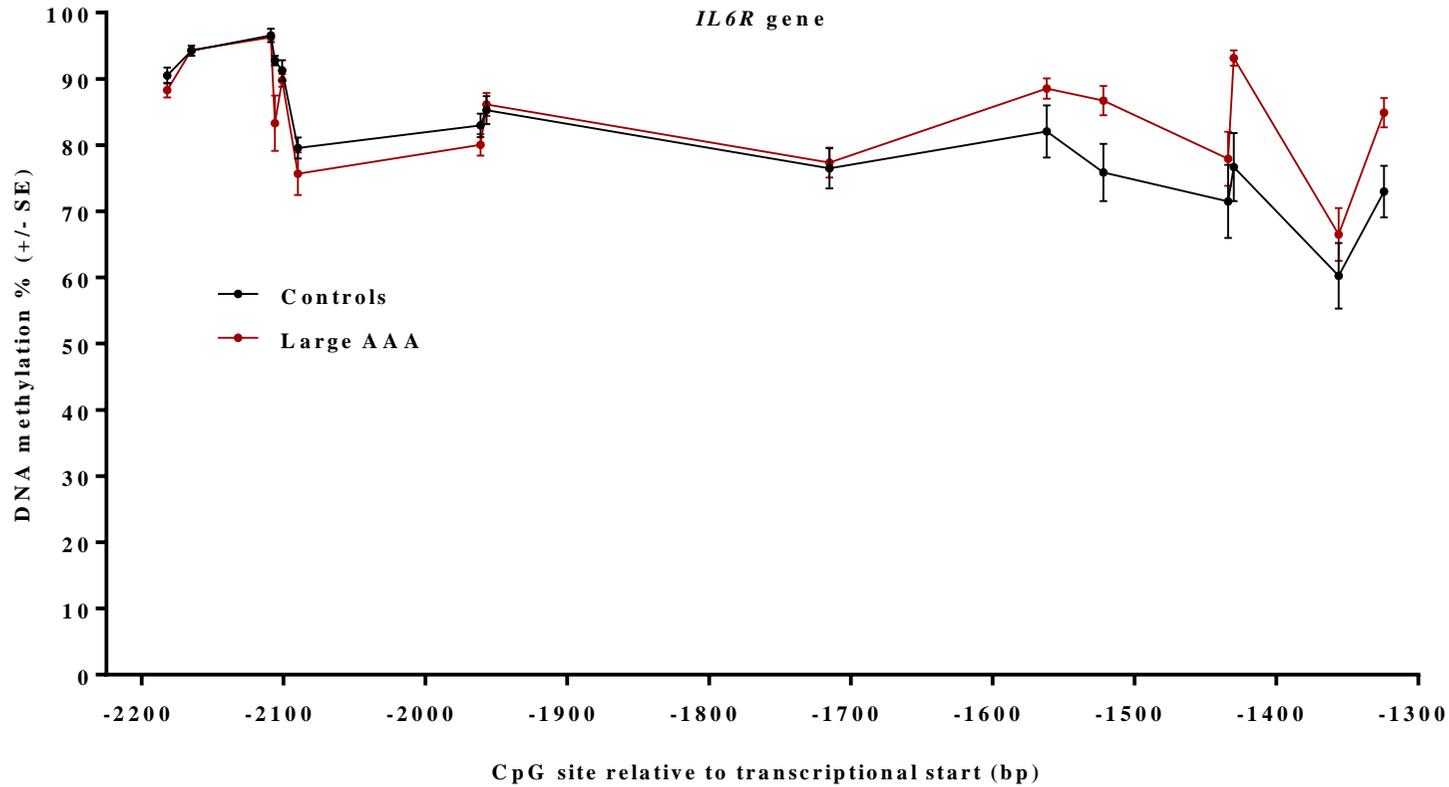


Figure 29 - Schematic representation of the sequenced region of the *IL6R* gene in PBMC and VSMC DNA.

The targeted CpG Island consists of 15 CpGs at NG\_012087.1: 2781-3729bp. The DNA methylation status of this region is shown in Figure 52.



**Figure 52 - DNA methylation status of the *IL6R* gene promoter in VSMC DNA:** DNA methylation status of 15 sequenced CpG sites in the *IL6R* gene promoter sequenced in 24 AAA and 20 controls. 2 CpG sites were hypermethylated in AAA: (-1430 (P=0.0113), -1324 (p=0.0176)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

Figure 52 shows that changes in methylation were seen in the *IL6R* gene promoter at the same CpGs as in PBMC assay (NG\_012087.1:g.3570\_3676|gom). Mean methylation was  $89\% \pm 5.83$  in AAA and  $74.85\% \pm 2.62$  in controls. The mean increase in DNA methylation in those with AAA at these two significantly differentially methylated CpGs was  $14.2\% \pm 3.22$  ( $P=0.0145$ ,  $Q=0.076$ ).

Descriptive statistics of the differential methylation observed in *IL6R* after bisulphite sequencing of VSMC DNA in 24 AAA and 20 controls are shown in Table 24.

Gene and location	Region (CpG site)	Relation to transcription start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>IL6R</i> Chr: 1	3570	-1430	76.7	93.17	16.47	0.0113	0.0743	3.11
NG_012087.1	3676	-1324	73	84.92	11.92	0.0176	0.0773	LS

**Table 24 – Descriptive statistics of the differentially methylated CpGs in *IL6R* after VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported. LS = lost significance.

### 5.4.3. *SERPINB9* gene

*SERPINB9* was not sequenced in the previous Chapter in PBMC DNA, as it had not been identified as relevant to AAA at the time the work was conducted. After the study by Ryer *et al.*, (2015)<sup>143</sup> was published, which indicated differential methylation in the *SERPINB9* gene body, the same CpG island was targeted in this project for VSMC sequencing.

A schematic representation of the section of *SERPINB9* gene that was isolated and sequenced is shown in Figure 53. This shows the amplicon sequencing coverage spanning a CpG island in the gene body prior to exon 6 and ending in exon 7.

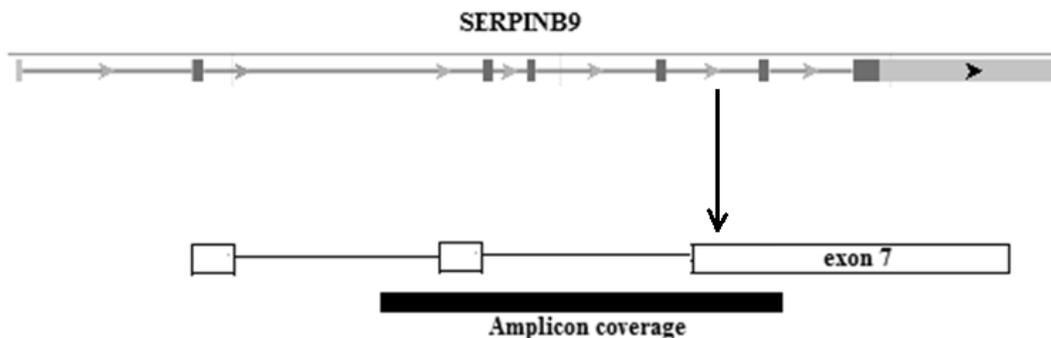
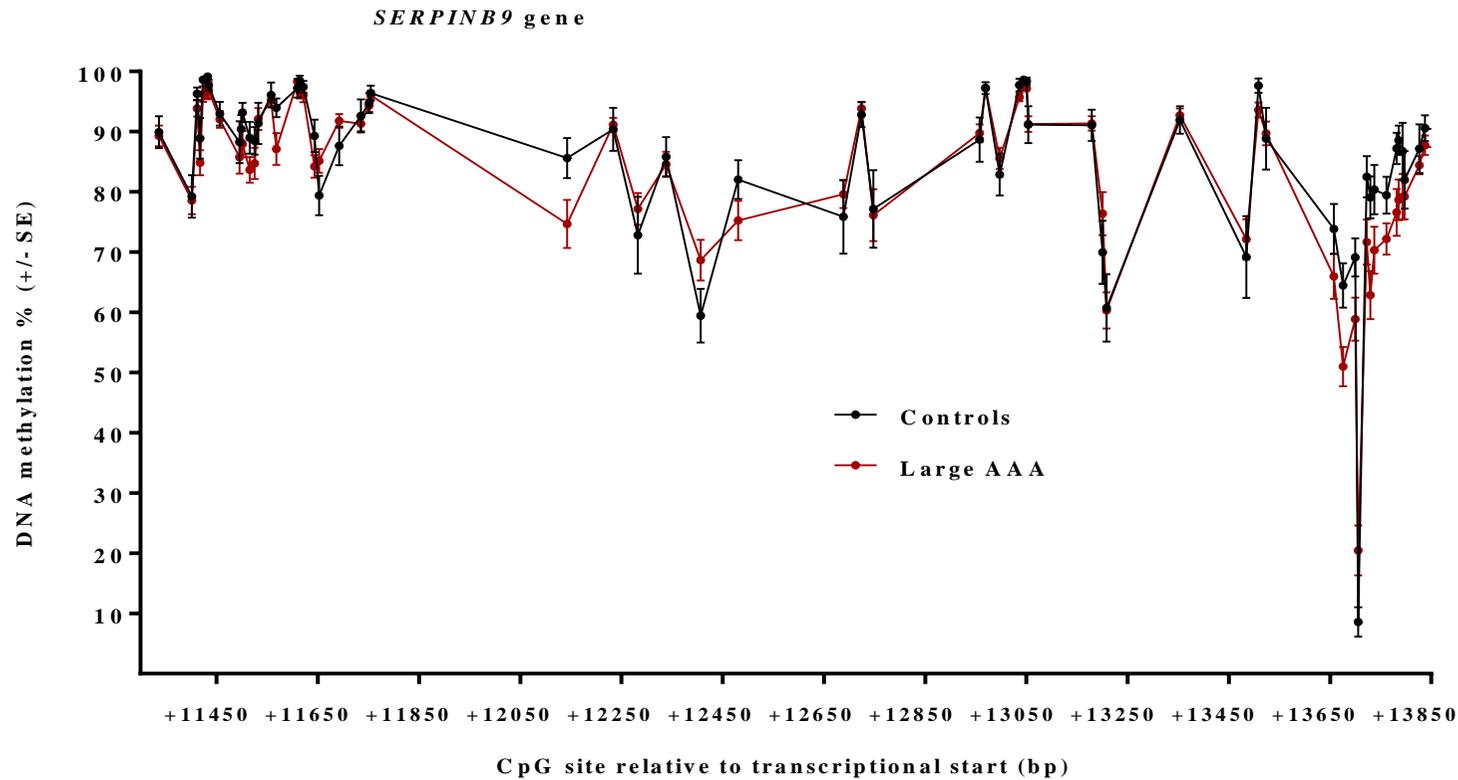


Figure 53 - Schematic representation of the sequenced region of the *SERPINB9* gene in VSMC DNA.

A total of 4 amplicons were designed to isolate the region of interest which consisted of 77 CpGs. The amplicons cover the genomic region of NC\_000006.12: 2889425-2892079.

The DNA methylation status of the sequenced CpG Island is shown in Figure 54.



**Figure 54 - DNA methylation status of the *SERPINB9* gene in VSMC DNA:** DNA methylation status of 77 sequenced CpG sites in a gene body CpG Island, sequenced in 24 AAA and 20 controls. 1 CpG was hypermethylated in AAA (+13636 (P=0.0098) and 5 were hypo-methylated in AAA (+13606 (P=0.0063), +13653 (P=0.0058), +13660 (P=0.0017), +13668 (P=0.0064), +13712 (0.0012), +13716 (P=0.0010)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

In the *SERPINB9* gene, seven CpGs located in the 3' untranslated region downstream of exon 7 were found differentially methylated in AAA compared to controls. Overall hypo-methylation of the region was observed (NC\_000006.12:g.2889674\_2889564|lom) where six CpGs were hypo-methylated (P=0.0037, Q=0.032), and 1 CpG was hypermethylated (P=0.0098, Q=0.0478) (NC\_000006.12:g.2889644|gom). This is in contrast to the overall AAA hypermethylation reported by Ryer *et al.*, (2016)<sup>143</sup> which was conducted on PBMC DNA.

Descriptive statistics of the differential methylation observed in *SERPINB9* after bisulphite sequencing of VSMC DNA in 24 AAA and 20 controls are shown in Table 25.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>SERPINB9</i> Chr: 6 NC_000006.12	2889674	+13606	64.47	51	-13.47	0.0063	0.0375	LS
	2889644	+13636	8.579	20.46	11.88	0.0098	0.0478	3.29
	2889627	+13653	82.53	71.67	-10.86	0.0058	0.0375	LS
	2889620	+13660	79.11	62.88	-16.23	0.0017	0.0273	LS
	2889612	+13668	80.37	70.33	-10.04	0.0064	0.0375	LS
	2889568	+13712	87.21	76.58	-10.63	0.0012	0.0273	0.33
	2889564	+13716	88.58	78.67	-9.91	0.001	0.0273	LS

**Table 25 – Descriptive statistics of the differentially methylated CpGs in *SERPINB9* after VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported. LS = lost significance.

#### 5.4.4. *SMYD2* gene

*SMYD2* is another gene which was not included in the previous sequencing assay, as it was only recently discovered in proximity to AAA genomic risk loci in the most recent GWAS meta-analysis.

A schematic representation of the section of the *SMYD2* gene which was isolated and sequenced is shown in Figure 55. This shows the amplicon sequencing coverage spanning the gene promoter CpG Island, past the transcriptional start, and ending in the intron downstream from the 3' end of exon 1.

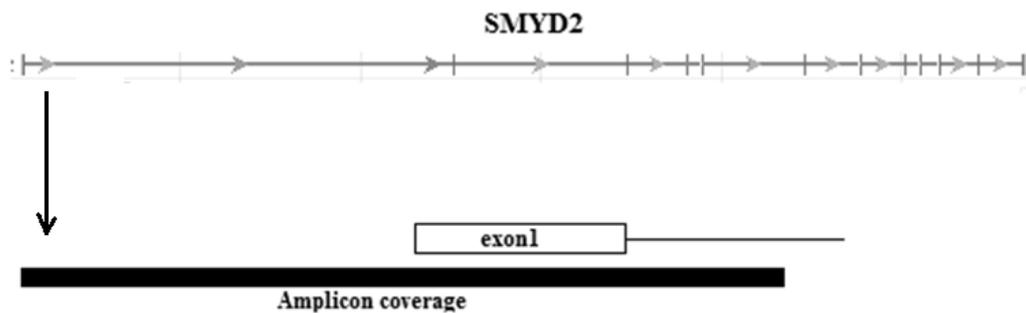
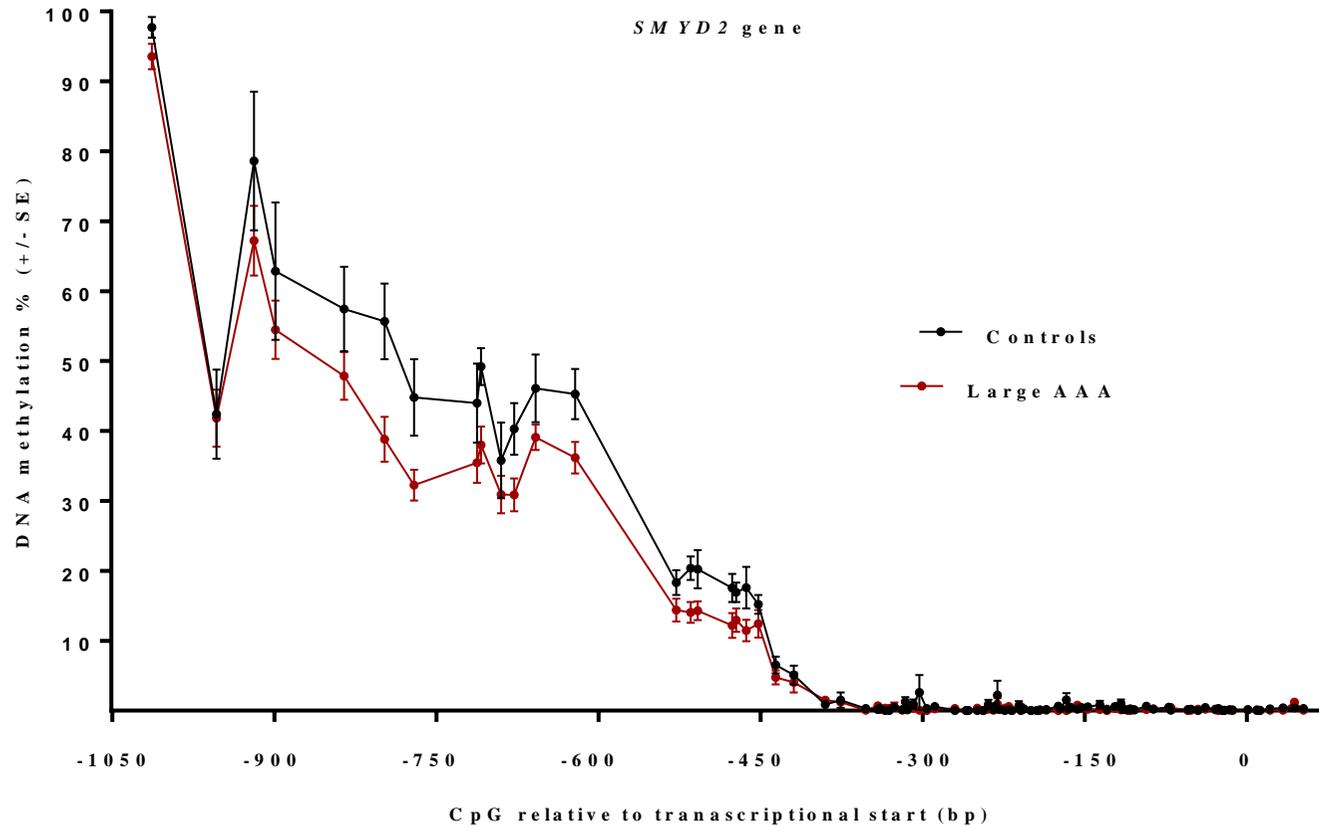


Figure 55 - Schematic representation of the sequenced region of the *SMYD2* gene in VSMC DNA.

A total of 3 amplicons were used to isolate and sequence 141 CpGs in the *SMYD2* gene at NC\_000001.11: 214280141-214281733.

DNA methylation status of the CpGs sequenced in *SMYD2* is shown in Figure 56.



**Figure 56 - DNA methylation status of the *SMYD2* gene in VSMC DNA:** DNA methylation status of 141 CpG sites, sequenced in 24 AAA and 20 controls. 4 CpGs were hypo-methylated in AAA compared to controls (-821 (P=0.0004), -792 (P=0.0014), -726 (P=0.0016), -633 (P=0.0014)). Bars represent standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

It was observed in the *SMYD2* gene that four CpGs were significantly hypo-methylated in those with AAA compared to controls. The specific sites of differential methylation were in the gene promoter upstream of the transcriptional start site (NC\_000001.11:g214280412\_214280600|lom).

Mean control methylation was  $48.75\% \pm 5.02$ , mean methylation in the cases was  $37.06\% \pm 2.95$ , and the mean overall decrease in methylation in AAA compared to controls was  $11.69\% \pm 8.2$  ( $P=0.0012$ ,  $Q=0.027$ ).

Descriptive statistics of the CpGs that were differentially methylated in the *SMYD2* gene promoter in 24 AAA compared to 20 controls are shown in Table 26.

Gene and location	Region (CpG site)	Relative to transcription Start (bp)	Meth % (controls)	Meth % (cases)	Meth % difference	P value	Q value	Age adjusted odds ratio
<i>SMYD2</i>	214280412	-821	55.67	39.5	-16.17	0.0004	0.02	LS
Chr: 1	214280441	-792	44.8	32.91	-11.89	0.0014	0.0299	0.09
NC_000001.11	214280507	-726	49.22	38.78	-10.44	0.0016	0.0299	0.06
	214280600	-633	45.3	37.04	-8.26	0.0014	0.0299	0.03

**Table 26 – Descriptive statistics of the differentially methylated CpGs in *SMYD2* after VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values, Q values and age adjusted odds ratios are reported. LS = lost significance.

### 5.4.5. *DAB2IP* gene

The *DAB2IP* gene was a new addition to this Chapter and was not sequenced in the PBMC assay from Chapter 4. A schematic representation of the *DAB2IP* gene regions that were sequenced in this study is shown in Figure 57. This shows the amplicon sequencing coverage spanning the two identified gene promoters. The first sequenced CpG Island is located in an intron downstream from the first exon, whilst the other is the opposite end of the gene, prior to and including exon 9.

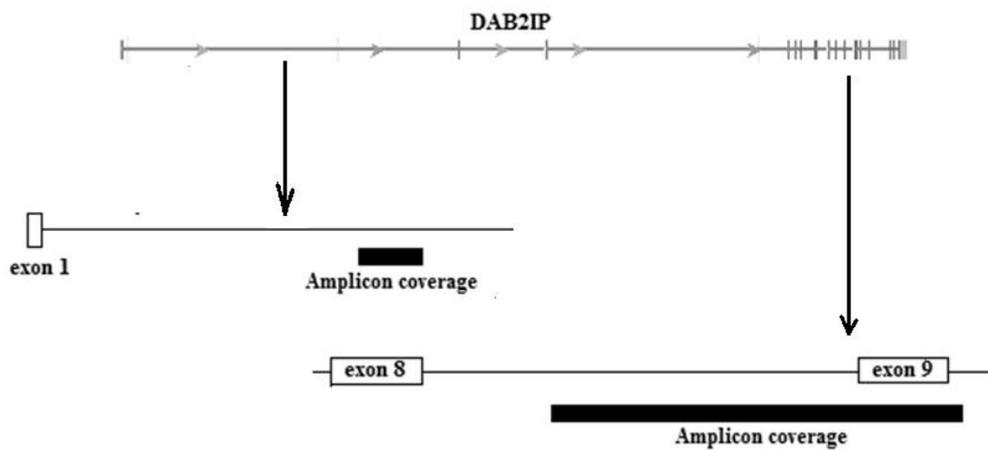


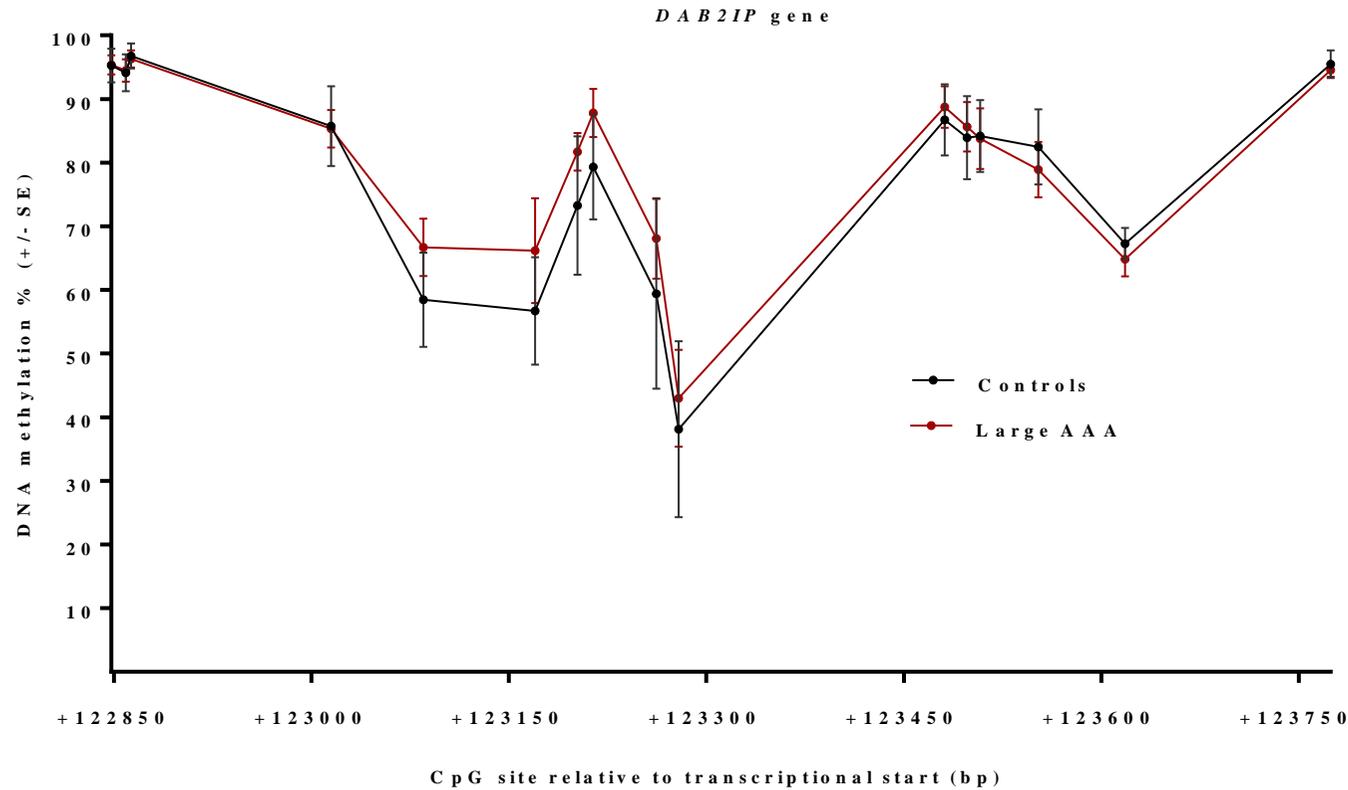
Figure 57 - Schematic representation of the sequenced region of the *DAB2IP* gene in VSMC DNA.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

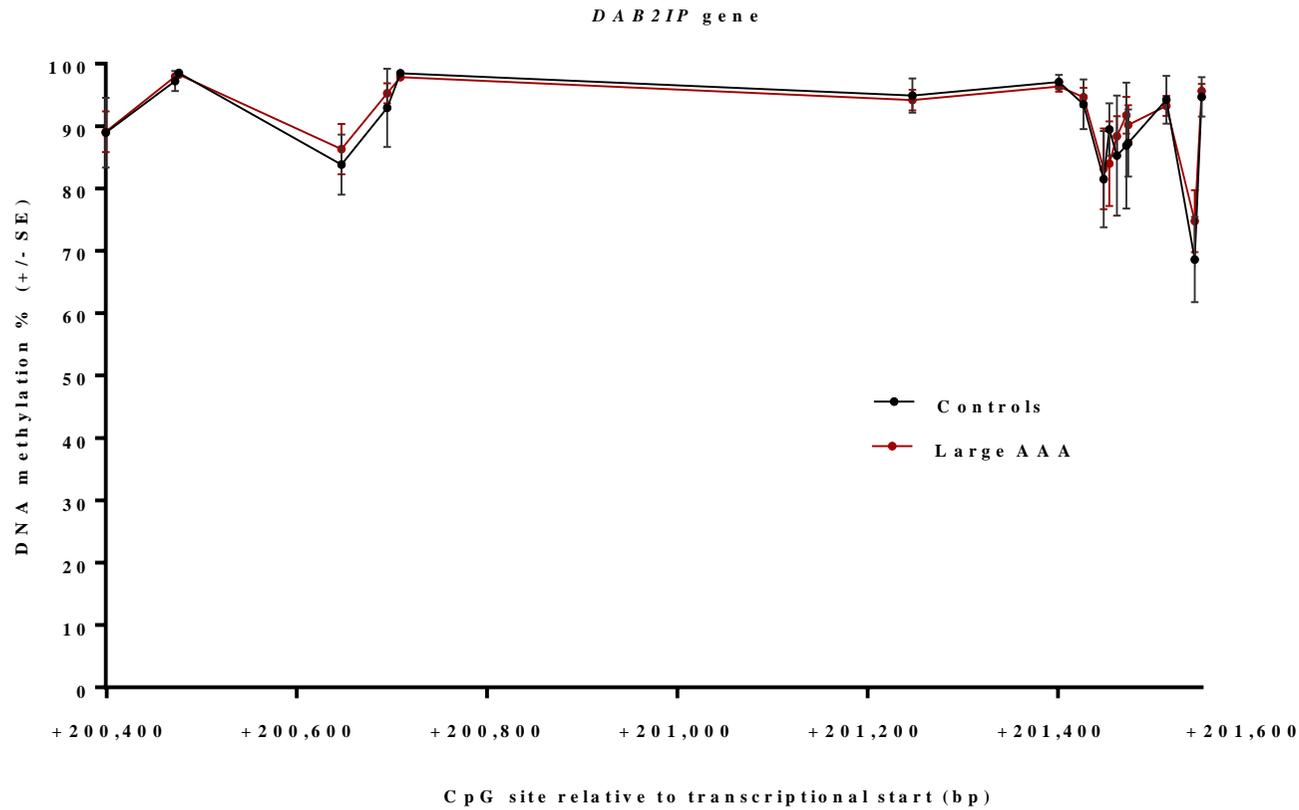
A total of 33 CpG sites were sequenced in the *DAB2IP* gene at the two separate locations with four sets of bisulphite PCR primers.

The first promoter near exon 1 was amplified and isolated with two amplicons that span 16 individual CpGs at NC\_000009.12: 121689899-121691062. The methylation status of these sites are shown in Figure 58, where there was no statistically significant differential methylation in the VSMC DNA of 24 AAA and 20 controls.

The second promoter, located near exon 9, was targeted with two amplicons that cover 17 individual CpGs at NC\_000009.12: 121767430-121768692, where there were no statistically significant changes in the methylation of any CpG between 24 AAA and 20 controls (Figure 59).



**Figure 58 - DNA methylation status in one of two *DAB2IP* gene promoters in VSMC DNA:** DNA methylation status of 16 sequenced CpG sites in *DAB2IP* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.



**Figure 59 - DNA methylation status of the second *DAB2IP* gene promoter in VSMC DNA:** DNA methylation status of 17 sequenced CpG sites in *DAB2IP* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

### 5.4.6. *LDLR* gene

After the positive results were observed in the *LDLR* gene from the previous Chapter in PBMC DNA taken from those with AAA compared to controls, sequencing coverage was increased for this assay by total of 80 CpG sites.

Amplicon coverage for sequencing is illustrated in Figure 60, which shows the previously sequenced region has been extended downstream further into the intronic region after exon 1 of the gene.

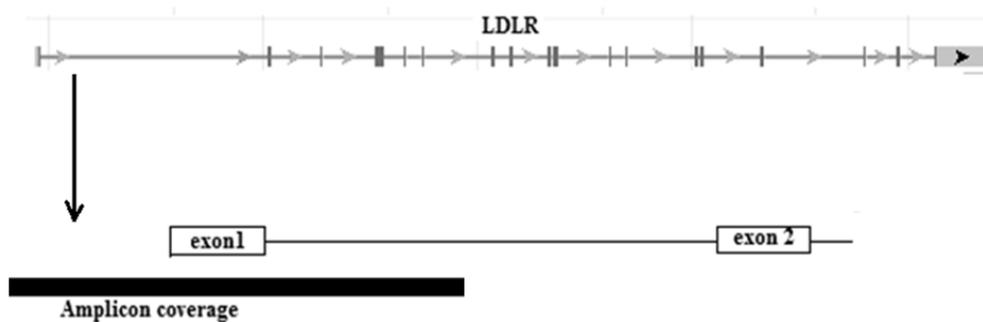
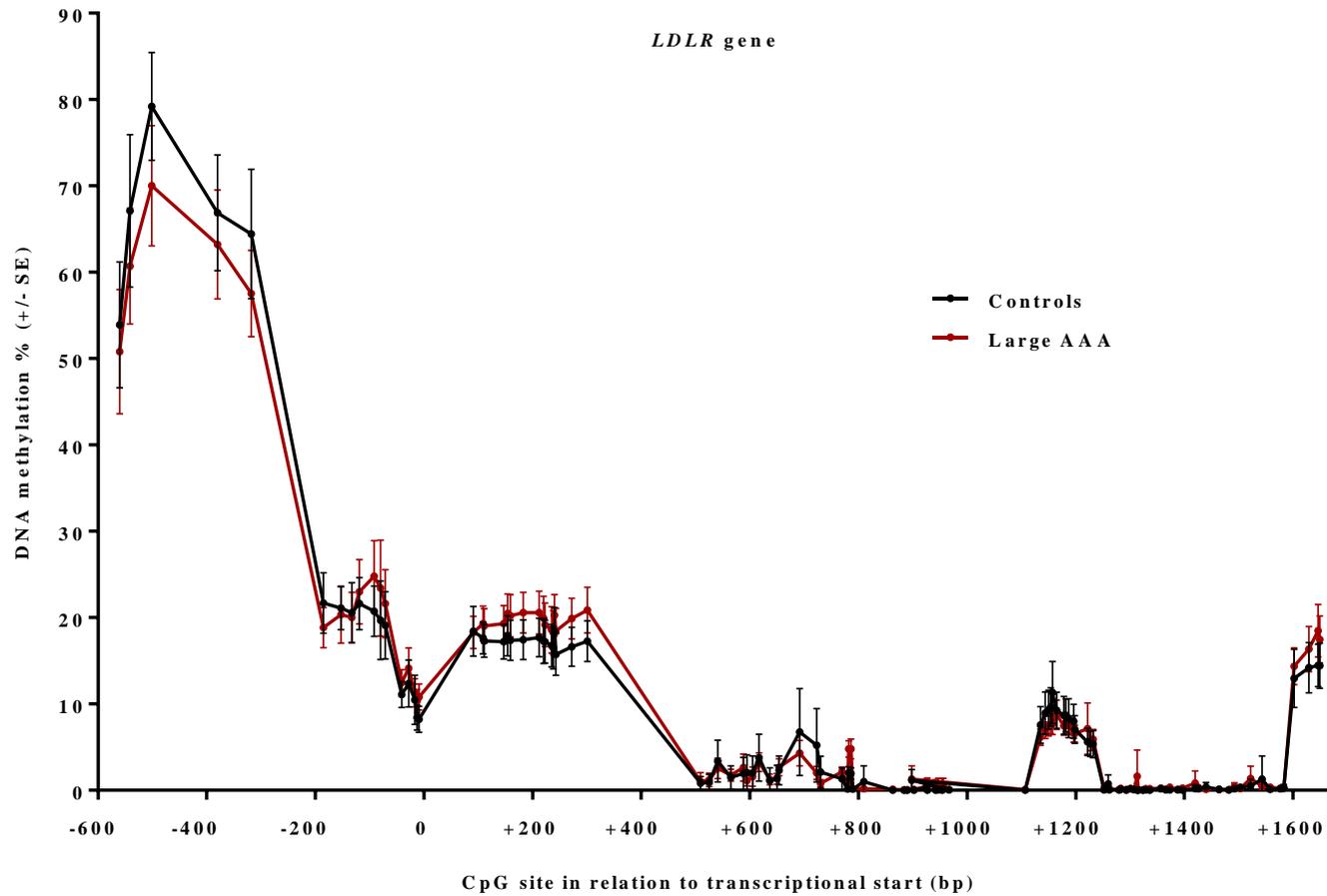


Figure 60 - Schematic representation of the sequenced region of the *LDLR* gene in VSMC DNA.

In total for this experiment, 4 amplicons were used to isolate VSMC DNA from the samples at NG\_009060.1: 4376-6785, covering a total number of 113 CpGs.

DNA methylation status of the sequenced *LDLR* gene region is displayed in Figure 61, where in contrast to the PBMC DNA sequencing assay, no statistically significant regions of methylation were present between 24 AAA compared to 20 controls.



**Figure 61 - DNA methylation status of the *LDLR* gene promoter in VSMC DNA:** DNA methylation status of 113 sequenced CpG sites in *LDLR* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

### 5.4.7. *MMP9* gene

Figure 41 is reproduced from the previous Chapter and displays the amplicon coverage of the sequenced *MMP9* gene promoter.

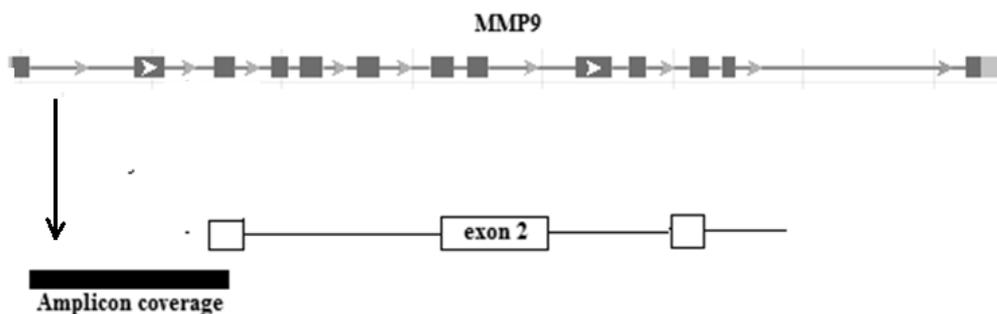
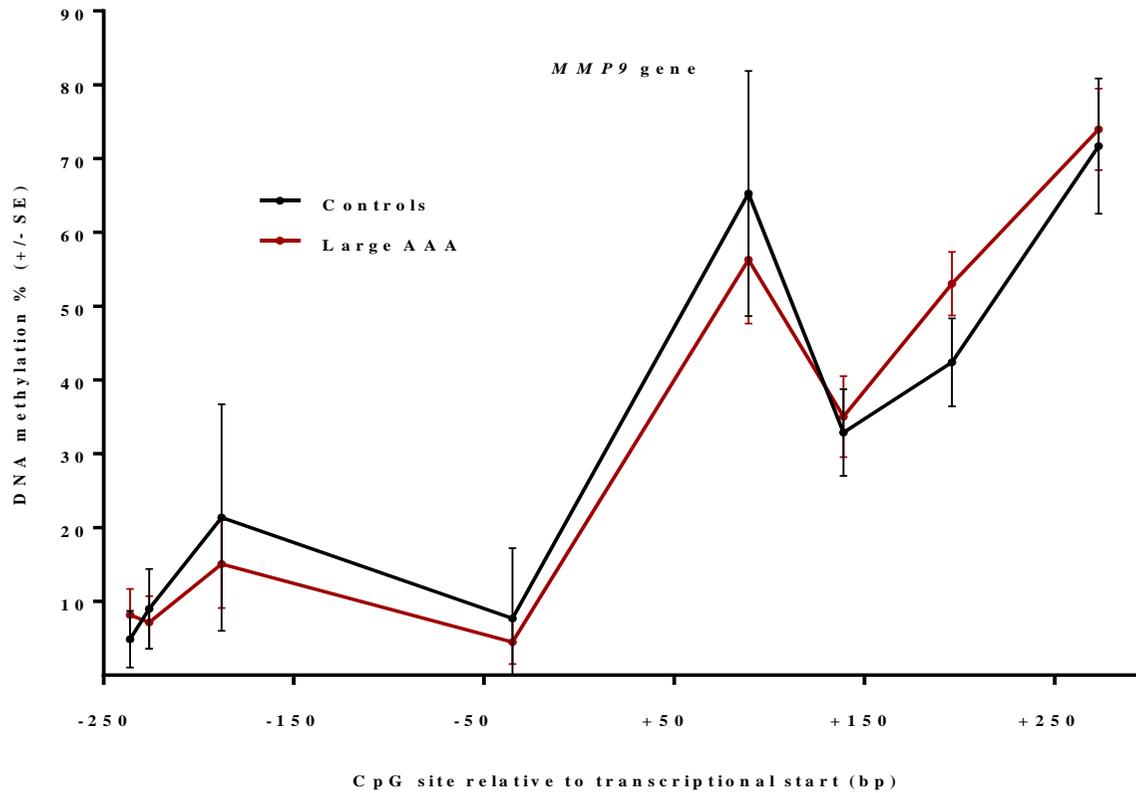


Figure 41 - Schematic representation of the sequenced region of the *MMP9* gene in PBMC and VSMC DNA.

In total, 15 CpGs were targeted at NG\_011468.1: 3980-5600 with 2 sets of bisulphite specific PCR primers. However, during data analysis, it was evident that there was some bad quality data for one of the amplicons (NG\_011468.1: 3980-4579). Hence, Figure 62 shows the methylation status of 8 CpGs at NG\_011468.1: 4640-5600.

There were no statistically significant differences in methylation between 24 AAA and 20 controls.



**Figure 62 - DNA methylation status of the *MMP9* gene promoter in VSMC DNA:** DNA methylation status of 8 sequenced CpG sites in *MMP9* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

#### 5.4.8. *LRP1* gene

The same regions, as well as an additional amplicon, were sequenced in the *LRP1* gene in this experiment following the PBMC DNA sequencing assay, which is reported in the previous Chapter. There was an increase in CpG coverage of 13 at NG\_016444.1: 3875-4240 upstream of the previously targeted region. A schematic illustration of the amplicon coverage for *LRP1* gene promoter sequencing in this Chapter is shown in Figure 63.

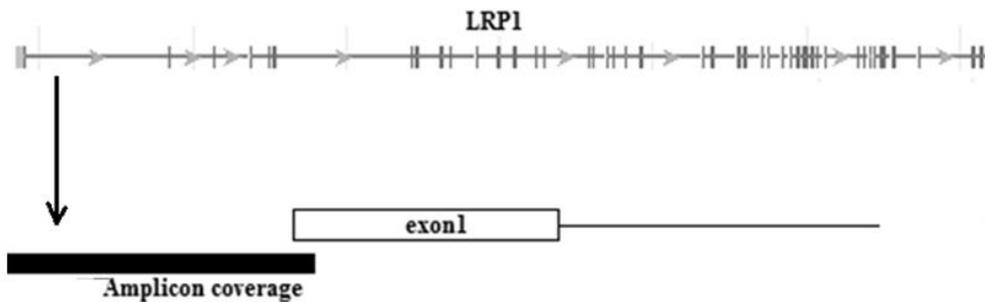
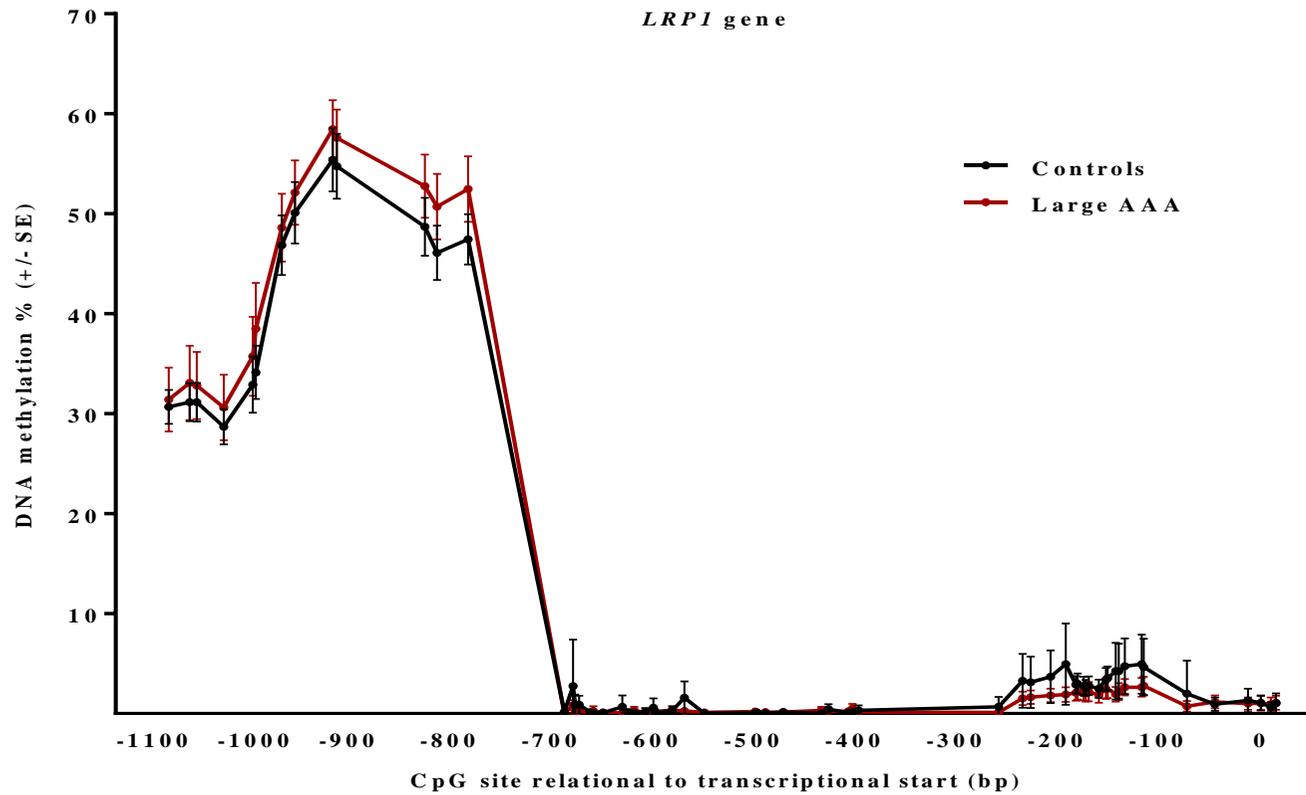


Figure 63 - Schematic representation of the sequenced region of the *LRP1* gene in VSMC DNA.

Total sequencing coverage was 62 CpG sites with 3 amplicons spanning NG\_016444.1 3875-5036. The methylation status of the sequenced regions are displayed in Figure 64.

There were no significant levels of differential methylation at any of the sites in the *LRP1* gene promoter, similarly to the results obtained from the previous Chapter.



**Figure 64 - DNA methylation status of the *LRP1* gene promoter in VSMC DNA:** DNA methylation status of 62 sequenced CpG sites in *LRP1* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

### 5.4.9. *SORT1* gene

The same *SORT1* CpG Island that was sequenced in the previous Chapter was also assessed in this Chapter. The amplicon spans 24 CpGs in the *SORT1* gene promoter at NG\_028280.1: 4363-4804bp, and Figure 31 has been reproduced from Chapter 4 showing a visual representation of the sequencing coverage in the *SORT1* gene.

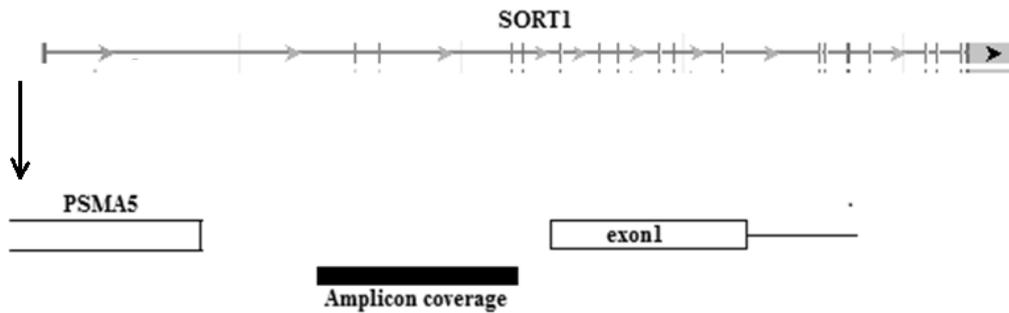
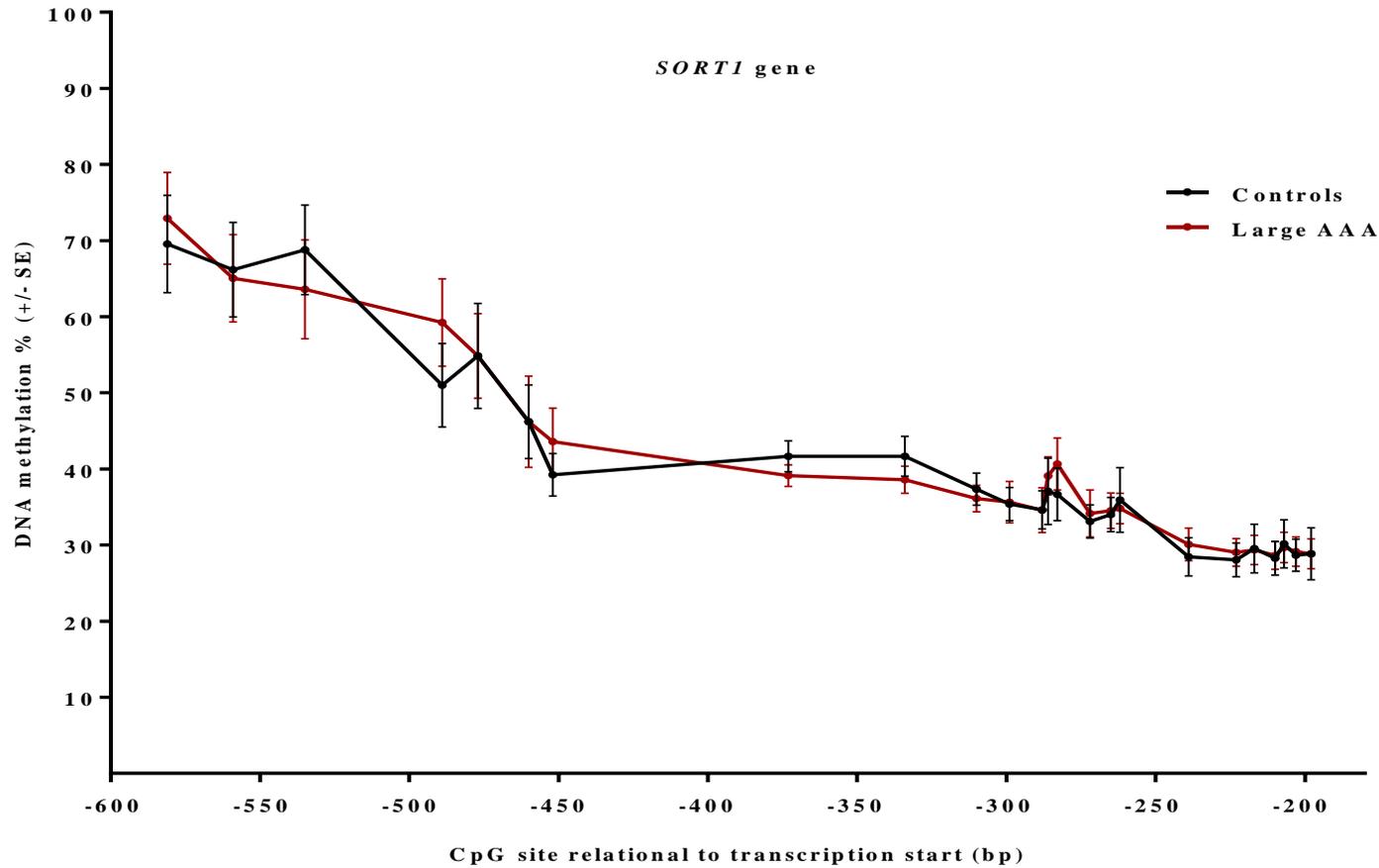


Figure 31 - Schematic representation of the sequenced region of the *SORT1* gene in PBMC and VSMC DNA.

The methylation status of the sequenced *SORT1* CpG Island is displayed in Figure 65, where the methylation values between 24 AAA and 20 controls were not statistically different.



**Figure 65 - DNA methylation status of the *SORT1* gene promoter in VSMC DNA:** DNA methylation status of 24 sequenced CpG sites in *SORT1* in 24 AAA and 20 controls. No differential methylation was observed. Bars represents standard error. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

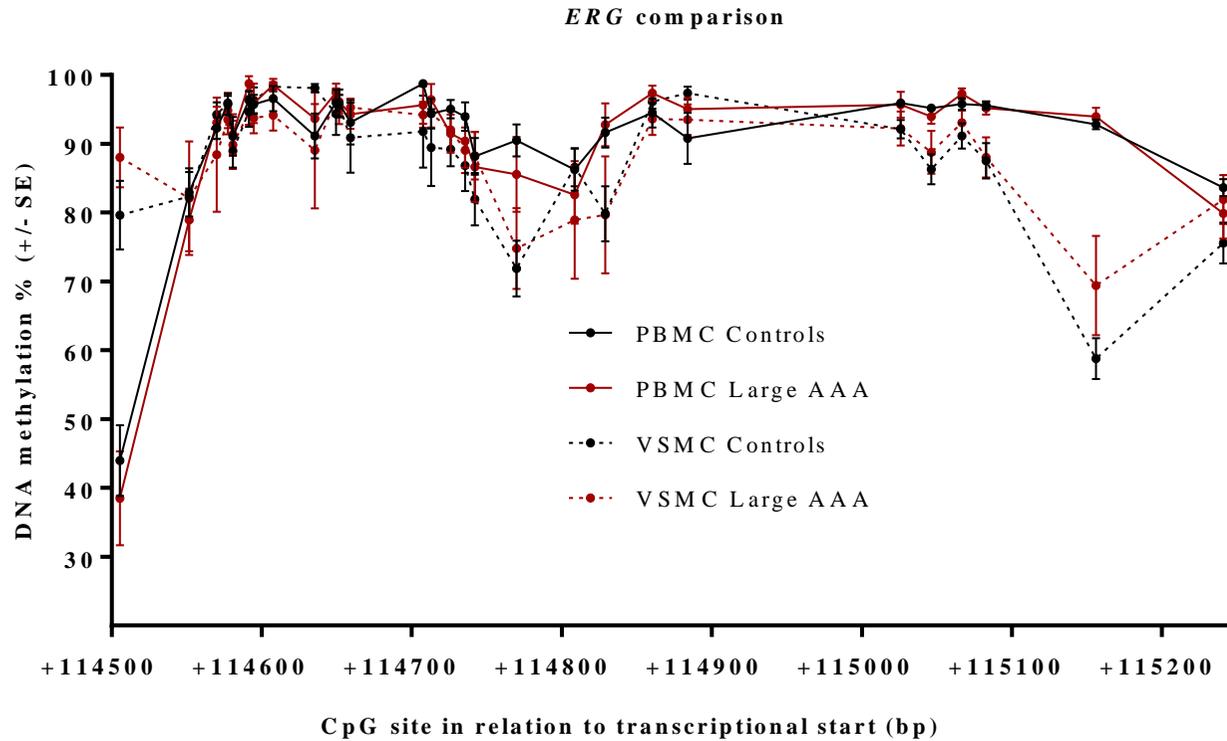
## 5.5. Comparing DNA methylation profiles in PBMC and VSMC DNA

As previously discussed, it is known that different cell types have independent epigenetic profiles. Essentially, epigenetic manipulation allows for extreme phenotypic diversity and it was considered likely before Chapter 4 and 5 were conducted to assess PBMC and VSMC DNA methylation profiles, that there may be differences between the two cell types. An important factor about these potential changes is that when there are clear differences between the methylation statuses of cells, the regions signify potential targets for further study, as it is possible that these specific regions are functionally active in gene activation/repression and subsequent phenotypic diversity.

In addition to what I already found from the sequencing assays in Chapter 4 and 5, I wanted to see whether DNA methylation patterns in the same genes would vary by cell type, and specifically whether these occurred at where I observed inter cell type differences between AAA and controls.

For every gene that was sequenced in both PBMC and VSMC DNA in this project (one of the *ERG* promoters, *IL6R*, *LDLR*, *LRP1* and *SORT1*), there is a Figure comparing the epigenetic profiles of the two sets of cells at each individual CpG site. These methylation values represent mean methylation of each CpG in each different cell type in AAA and controls. Figure 66, 67, 68, 69 and 70 represent *ERG*, *IL6R*, *LDLR*, *LRP1* and *SORT1* respectively.

5.5.1. *ERG* comparison



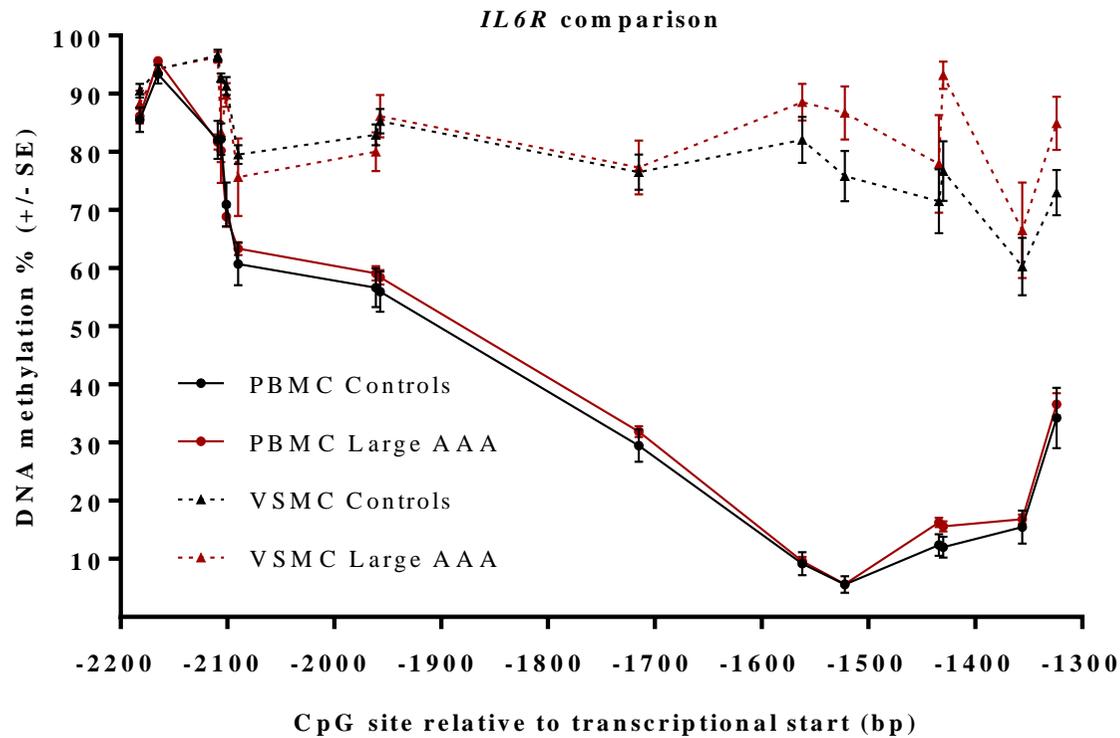
**Figure 66 – Comparison of PBMC and VSMC DNA methylation status in *ERG*:** Mean methylation of cases and controls were compared for each separate cell type. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

After comparing the methylation status of each CpG that was sequenced in the *ERG* gene promoter in PBMCs and VSMCs, it was observed that 3 were differentially methylated between the controls, and 4 were differentially methylated between AAAs. See table 27 for descriptive statistics.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (PBMC controls)	Meth % (VSMC controls)	Meth % Difference (controls)	P value (controls)	Q value (controls)	Meth % (PBMC cases)	Meth % (VSMC cases)	Meth % Difference (cases)	P value (cases)	Q value (cases)
<i>ERG</i> Chr: 21 NG_029732.1	282905	+114564	43.97	79.61	-35.64	<0.0001	<0.0001	38.5	88	-49.5	<0.0001	<0.0001
	283151	+114810	90.5	71.89	18.61	<0.0001	<0.0001	85.55	74.79	10.75	<0.0001	0.0003
	283206	+114865	ns	ns	ns	ns	ns	92.76	79.67	13.09	<0.0001	<0.0001
	283510	+115169	92.81	58.8	34.01	<0.0001	<0.0001	93.94	69.42	24.52	<0.0001	<0.0001

**Table 27 – Descriptive statistics of the differentially methylated CpGs in *ERG* after comparison of PBMC and VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values and Q values are reported. Ns = not significant.

5.5.2. *IL6R* comparison

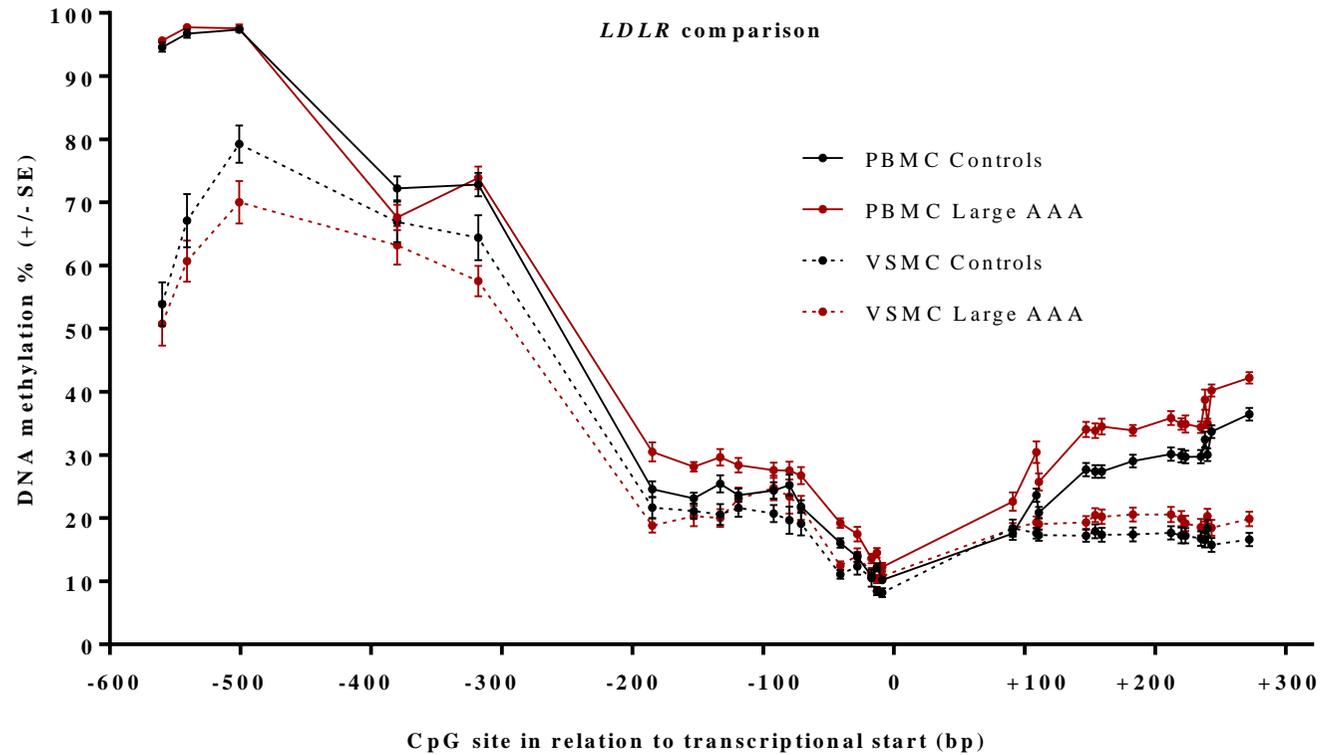
**Figure 67 – Comparison of PBMC and VSMC DNA methylation status in *IL6R*:** Mean methylation of cases and controls were compared for each separate cell type. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

After comparing the methylation status of each CpG that was sequenced in the *IL6R* gene promoter in PBMCs and VSMCs, it was observed that 13 were differentially methylated between the controls, and 12 were differentially methylated between AAAs. See table 28 for descriptive statistics.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (PBMC controls)	Meth % (VSMC controls)	Meth % Difference (controls)	P value (controls)	Q value (controls)	Meth % (PBMC cases)	Meth % (VSMC cases)	Meth % Difference (cases)	P value (cases)	Q value (cases)
<i>IL6R</i> Chr: 1 NG_012087.1	2891	-2105	82.09	96.55	-14.46	<0.0001	<0.0001	81.67	96.29	-14.63	<0.0001	<0.0001
	2894	-2102	82.21	92.75	-10.54	0.0007	0.0001	ns	ns	ns	ns	ns
	2899	-2097	70.94	91.3	-20.36	<0.0001	<0.0001	68.85	89.79	-20.94	<0.0001	<0.0001
	2910	-2086	60.68	79.55	-18.87	<0.0001	<0.0001	63.33	75.67	-12.33	<0.0001	<0.0001
	3039	-1957	56.66	82.95	-26.29	<0.0001	<0.0001	59.13	80.04	-20.92	<0.0001	<0.0001
	3043	-1953	55.98	85.3	-29.32	<0.0001	<0.0001	58.4	86.13	-27.73	<0.0001	<0.0001
	3285	-1711	29.49	76.5	-47.01	<0.0001	<0.0001	31.85	77.33	-45.48	<0.0001	<0.0001
	3438	-1558	9.191	82.05	-72.86	<0.0001	<0.0001	9.604	88.54	-78.94	<0.0001	<0.0001
	3478	-1518	5.565	75.85	-70.28	<0.0001	<0.0001	5.646	86.71	-81.06	<0.0001	<0.0001
	3566	-1430	12.38	71.5	-59.12	<0.0001	<0.0001	16.27	77.92	-61.65	<0.0001	<0.0001
	3570	-1426	12	76.7	-64.7	<0.0001	<0.0001	15.58	93.17	-77.58	<0.0001	<0.0001
	3644	-1352	15.45	60.25	-44.8	<0.0001	<0.0001	16.81	66.5	-49.69	<0.0001	<0.0001
	3676	-1320	34.23	73	-38.77	<0.0001	<0.0001	36.6	84.92	-48.32	<0.0001	<0.0001

**Table 28 – Descriptive statistics of the differentially methylated CpGs in *IL6R* after comparison of PBMC and VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values and Q values are reported. Ns = not significant.

5.5.3. *LDLR* comparison

**Figure 68 – Comparison of PBMC and VSMC DNA methylation status in *LDLR*:** Mean methylation of cases and controls were compared for each separate cell type. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

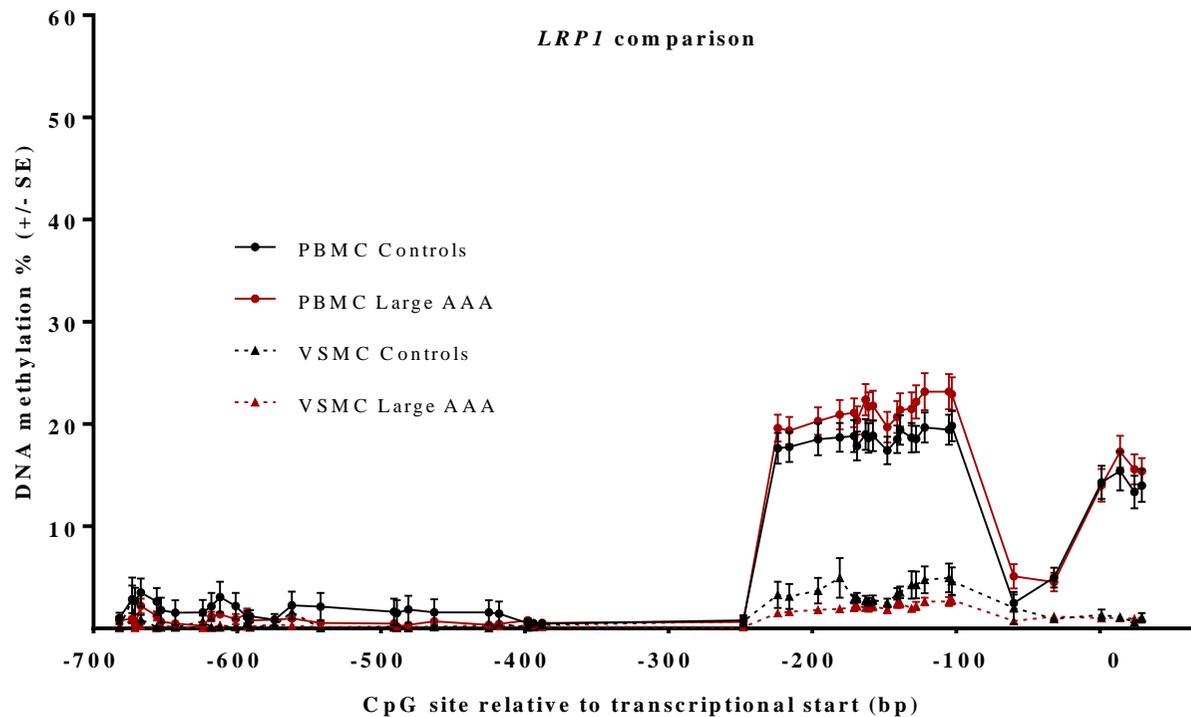
## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

After comparing the methylation status of each CpG that was sequenced in the *LDLR* gene promoter and transcriptional start site in PBMCs and VSMCs, it was observed that 19 were differentially methylated between the controls, and 17 were differentially methylated between AAAs. See table 29 for descriptive statistics.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (PBMC controls)	Meth % (VSMC controls)	Meth % Difference (controls)	P value (controls)	Q value (controls)	Meth % (PBMC cases)	Meth % (VSMC cases)	Meth % Difference (cases)	P value (cases)	Q value (cases)
<i>LDLR</i> Chr: 19 NG_009060.1	4420	-562	94.56	53.89	40.66	<0.0001	<0.0001	95.6	50.79	44.8	<0.0001	<0.0001
	4439	-543	96.71	67.11	29.61	<0.0001	<0.0001	97.74	60.71	37.03	<0.0001	<0.0001
	4479	-503	97.36	79.21	18.15	<0.0001	<0.0001	97.55	70	27.55	<0.0001	<0.0001
	4600	-382	72.23	66.89	5.338	0.0102	0.008203	ns	ns	ns	ns	ns
	4662	-320	72.84	64.42	8.416	<0.0001	<0.0001	73.88	57.54	16.34	<0.0001	<0.0001
	4900	-82	25.23	19.68	5.545	0.0068	0.0057	ns	ns	ns	ns	ns
	5089	+7	23.64	17.63	6.007	0.0034	0.0031	30.46	19.29	11.17	<0.0001	<0.0001
	5127	+145	27.69	17.21	10.48	<0.0001	<0.0001	34.02	19.29	14.73	<0.0001	<0.0001
	5134	+152	27.4	17.89	9.501	<0.0001	<0.0001	33.85	20.46	13.4	<0.0001	<0.0001
	5139	+157	27.4	17.37	10.03	<0.0001	<0.0001	34.5	20.21	14.29	<0.0001	<0.0001
	5163	+181	29.02	17.42	11.6	<0.0001	<0.0001	33.9	20.58	13.31	<0.0001	<0.0001
	5192	+210	30.13	17.68	12.44	<0.0001	<0.0001	35.83	20.58	15.25	<0.0001	<0.0001
	5200	+218	29.88	17.26	12.61	<0.0001	<0.0001	34.88	19.92	14.96	<0.0001	<0.0001
	5203	+221	29.71	17.21	12.5	<0.0001	<0.0001	34.88	19.17	15.71	<0.0001	<0.0001
	5215	+233	29.73	16.74	12.99	<0.0001	<0.0001	34.42	18.58	15.83	<0.0001	<0.0001
	5218	+236	32.42	16.53	15.89	<0.0001	<0.0001	38.77	18.25	20.52	<0.0001	<0.0001
	5220	+238	30.02	18.37	11.65	<0.0001	<0.0001	34.98	20.29	14.69	<0.0001	<0.0001
	5223	+241	33.69	15.74	17.95	<0.0001	<0.0001	40.19	18.46	21.73	<0.0001	<0.0001
	5252	+270	36.46	16.63	19.83	<0.0001	<0.0001	42.19	19.88	22.31	<0.0001	<0.0001

**Table 29 – Descriptive statistics of the differentially methylated CpGs in *LDLR* after comparison of PBMC and VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values and Q values are reported. Ns = not significant.

#### 5.5.4. *LRP1* comparison



**Figure 69 – Comparison of PBMC and VSMC DNA methylation status in *LRP1*:** Mean methylation of cases and controls were compared for each separate cell type.

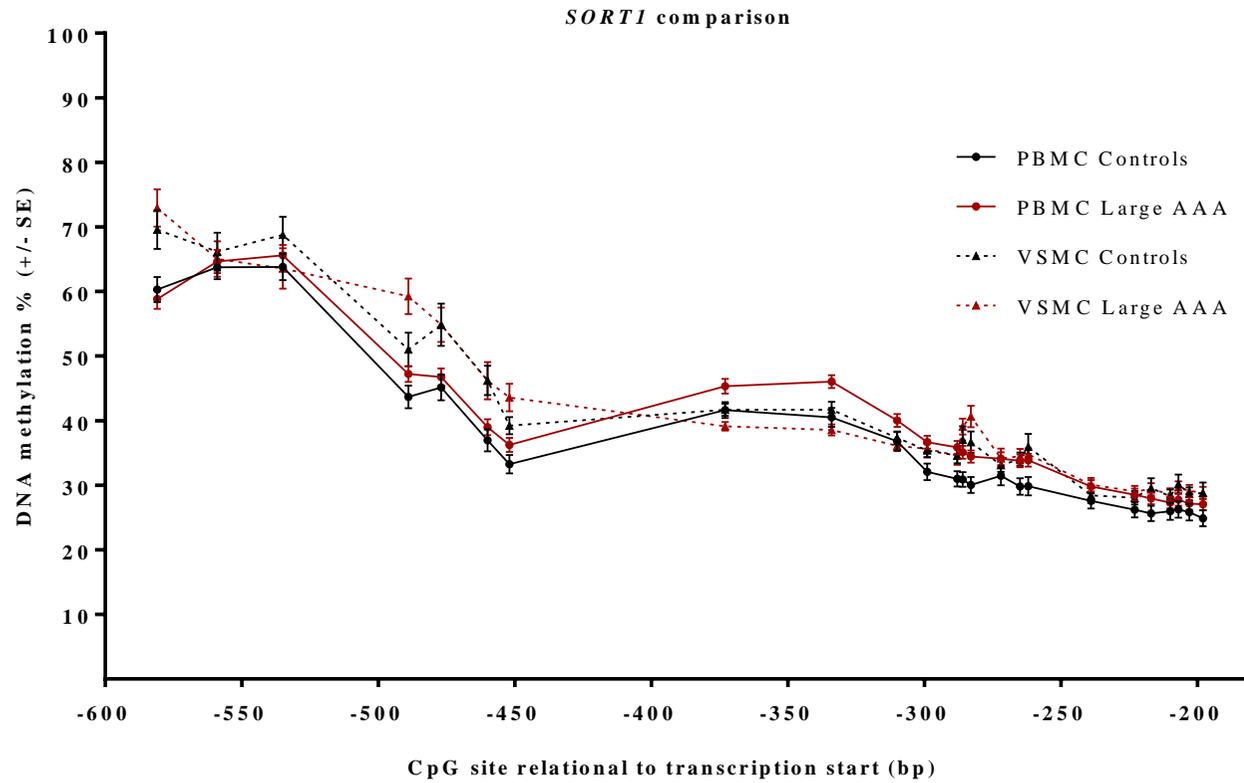
After comparing the methylation status of each CpG that was sequenced in the *LRP1* gene promoter in PBMCs and VSMCs, it was observed that 21 were differentially methylated between the controls, and 22 were differentially methylated between AAAs. See table 30 for descriptive statistics.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (PBMC controls)	Meth % (VSMC controls)	Meth % Difference (controls)	P value (controls)	Q value (controls)	Meth % (PBMC cases)	Meth % (VSMC cases)	Meth % Difference (cases)	P value (cases)	Q value (cases)
<i>LRPI</i> Chr: 12 NG_016444.1	4756	-239	17.64	3.3	14.34	<0.0001	<0.0001	19.61	1.522	18.09	<0.0001	<0.0001
	4764	-231	17.77	3.15	14.62	<0.0001	<0.0001	19.36	1.652	17.7	<0.0001	<0.0001
	4784	-211	18.51	3.7	14.81	<0.0001	<0.0001	20.3	1.826	18.48	<0.0001	<0.0001
	4799	-196	18.69	4.95	13.74	<0.0001	<0.0001	20.91	1.913	19	<0.0001	<0.0001
	4809	-186	18.82	2.9	15.92	<0.0001	<0.0001	21.11	2.087	19.02	<0.0001	<0.0001
	4811	-184	17.85	3	14.85	<0.0001	<0.0001	20.37	2.217	18.15	<0.0001	<0.0001
	4817	-178	18.97	2.75	16.22	<0.0001	<0.0001	22.4	2.043	20.36	<0.0001	<0.0001
	4819	-176	18.64	2.6	16.04	<0.0001	<0.0001	21.7	2.043	19.65	<0.0001	<0.0001
	4822	-173	18.87	2.8	16.07	<0.0001	<0.0001	21.8	2.217	19.59	<0.0001	<0.0001
	4832	-163	17.44	2.5	14.94	<0.0001	<0.0001	19.72	1.87	17.85	<0.0001	<0.0001
	4839	-156	18.53	3.4	15.13	<0.0001	<0.0001	20.69	2.478	18.21	<0.0001	<0.0001
	4841	-154	19.44	3.55	15.89	<0.0001	<0.0001	21.39	2.522	18.87	<0.0001	<0.0001
	4849	-146	18.67	4.25	14.42	<0.0001	<0.0001	21.48	1.957	19.52	<0.0001	<0.0001
	4852	-143	18.54	4.25	14.29	<0.0001	<0.0001	22.18	2.217	19.96	<0.0001	<0.0001
	4858	-137	19.67	4.75	14.92	<0.0001	<0.0001	23.16	2.652	20.51	<0.0001	<0.0001
	4875	-120	19.46	4.95	14.51	<0.0001	<0.0001	23.16	2.652	20.51	<0.0001	<0.0001
	4877	-118	19.82	4.65	15.17	<0.0001	<0.0001	22.91	2.783	20.12	<0.0001	<0.0001
	4920	-76	ns	ns	ns	ns	ns	5.095	0.7391	4.356	0.0032	0.004
	4981	-15	14.29	1.316	12.97	<0.0001	<0.0001	14	1	13	<0.0001	<0.0001
	4994	-2	15.42	1.105	14.32	<0.0001	<0.0001	17.29	1.043	16.24	<0.0001	<0.0001
5004	+8	13.37	0.6316	12.74	<0.0001	<0.0001	15.58	0.9565	14.62	<0.0001	<0.0001	
5009	+13	13.97	1.05	12.92	<0.0001	<0.0001	15.38	1.087	14.29	<0.0001	<0.0001	

**Table 30 – Descriptive statistics of the differentially methylated CpGs in *LRPI* after comparison of PBMC and VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values and Q values are reported. Ns = not significant.

5.5.5. *SORT1* comparison



**Figure 70 – Comparison of PBMC and VSMC DNA methylation status in *SORT1*:** Mean methylation of cases and controls were compared for each separate cell type. The solid line is to aid visualisation of the differences between AAA and controls and does not represent intermediate values.

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

After comparing the methylation status of each CpG that was sequenced in the *SORT1* gene promoter in PBMCs and VSMCs, it was observed that 3 were differentially methylated between the controls, and 8 were differentially methylated between AAAs. See table 31 for descriptive statistics.

Gene and location	Region (CpG site)	Relative to transcription start (bp)	Meth % (PBMC controls)	Meth % (VSMC controls)	Meth % Difference (controls)	P value (controls)	Q value (controls)	Meth % (PBMC cases)	Meth % (VSMC cases)	Meth % Difference (cases)	P value (cases)	Q value (cases)
<i>SORT1</i> Chr: 1 NG_028280.1	4399	-592	60.28	69.56	-9.281	0.0003	0.0019	58.81	72.92	-14.1	<0.0001	<0.0001
	4491	-500	ns	ns	ns	ns	ns	47.23	59.25	-12.02	<0.0001	<0.0001
	4503	-488	45.15	54.83	-9.683	0.0001	0.0019	46.75	54.83	-8.083	<0.0001	0.0004
	4520	-471	36.98	46.22	-9.247	0.0003	0.0019	38.98	46.21	-7.229	0.0004	0.0010
	4528	-463	ns	ns	ns	ns	ns	36.25	43.58	-7.333	0.0003	0.0001
	4607	-384	ns	ns	ns	ns	ns	45.31	39.13	6.188	0.0023	0.0047
	4646	-345	ns	ns	ns	ns	ns	46.06	38.58	7.479	0.0002	0.0009
	4697	-294	ns	ns	ns	ns	ns	34.44	40.67	-6.229	0.0022	0.0047

**Table 31 – Descriptive statistics of the differentially methylated CpGs in *SORT1* after comparison of PBMC and VSMC sequencing:** The candidate gene, genomic coordinates of the specific CpGs, mean methylation of AAAs and controls, differences in mean methylation, P values and Q values are reported. Ns = not significant.

## 5.6. Conclusion

This study is the first to assess DNA methylation in VSMCs of people with AAA. The results demonstrate further evidence that there is an epigenetic basis to AAA, and is proof of concept for further large scale analysis. It was identified that significant differentially methylated CpGs exist in *ERG*, *IL6R*, *SERPINB9* and *SMYD2* in VSMC DNA from people with AAA compared to controls.

*ERG* is a member of a transcription factor family and is required for vascular remodelling, *SERPINB9* is a serine protease inhibitor that inhibits the activity of Granzyme B, and *SMYD2* is a lysine methyltransferase which regulates activity of other genes. Dysregulation of any of these genes may cause irregular physiological events within the vasculature<sup>194-196</sup>, and considering aberrant methylation patterns have been identified at CpGs that regulate these genes, it is possible that functional dysregulation as a result of differential methylation could be a contributory factor in AAA pathobiology.

It is evident in the work from this Chapter, and previously in Chapter 4, how incredibly accurate the bisulphite sequencing methodology is when assessing DNA methylation, given the highly consistent results and low levels of variation between individual samples, especially when there is a larger sample size. This is very important because it then allows the identification of real, significant methylation differences between sample groups. The work in this chapter differs from Chapter 4 due to the origin of the DNA source used for methylation analysis. PBMC DNA is derived from the immune cells of the blood, whereas VSMC DNA is derived directly from the site of aneurysm in the aortic wall. Both of these DNA sources potentially offer valuable insight to the pathobiology of

## Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA

aneurysms given the distinguished role of blood leukocytes and vascular cell apoptosis during AAA formation and growth<sup>33</sup>.

Interestingly, the methylation profiles largely differed between DNA derived from PBMCs and VSMCs, reflecting exactly how differences in CpG methylation may control phenotypic plasticity. In PBMCs it was observed that large portions of the *SORT1* and *LDLR* gene promoters were hypermethylated in AAA compared to controls (Chapter 4), but this was not replicated in VSMC DNA (current Chapter). However there were differences in methylation between the cell types at the same CpGs in *LDLR* where there were differences between AAA and controls in PBMCs. In *IL6R*, CpGs were hypermethylated at the same location in both DNA sources in individuals with AAA, which is particularly interesting since there were also very strong differences at these CpGs between the different cell types. All other sites which were differentially methylated in VSMCs (*ERG*, *SERPINB9* and *SMYD2*) were new additions and were not sequenced in the PBMC assay.

There were limitations to this work, including limited sample size which was specifically hindered due to the difficulty of obtaining aortic tissues. This problem in turn had an effect on sample group demographics, and as a result it was not possible to control for gender, smoking and age as effectively as in the global methylation and PBMC sequencing assay. Furthermore, the small sample size had further impact on the results, including large deviation ranges of mean sample group methylation values, in addition to lower statistical power and confidence levels than the PBMC NGS assay which had more than double the sample size.

Hypoxia induced methylation changes have been studied extensively in a range of different cell types, identifying that significant lack of oxygen reaching cells can result in

## **Chapter 5: Vascular smooth muscle cell DNA methylation analysis in AAA**

aberrant gene function and adverse phenotypic effects<sup>197-199</sup>. This is particularly relevant to this work considering aortic tissues were detached from live and deceased human bodies and were subject to variable times of hypoxic conditions prior to primary cell culture. This may or may not have had a significant impact on the levels of DNA methylation in the vascular cells constituting the tunica media of the aortas due to the absence of literature on the subject specific to VSMCs. This factor would therefore represent an important potential future experiment.

Corroboration of the differential methylation observed in Chapter 4 (PBMC methylation) and Chapter 5 (VSMC methylation) will be addressed in the fourth and final results chapter (Chapter 6) of this thesis. The work in Chapter 5 concentrates on VSMC gene expression and blood plasma protein expression analysis to determine whether observed methylation changes are associated with gene functionality.

**6. Chapter 6 – Results section 4**

**Functional corroboration of differential methylation in AAA**

### 6.1. Introduction

This thesis has so far established that global PBMC DNA hypermethylation is associated with large AAA and increasing AAA size, which is independent of circulating HOMOCYSTEINE. In addition, gene specific CpG methylation changes were then identified in PBMC and VSMC DNA in individuals with AAA compared to controls. This work has provided compelling evidence that there is a previously unidentified epigenetic basis to AAA in a global and gene specific context. However, the functional effects of these methylation changes have not yet been addressed.

It is well known that DNA methylation is a key regulatory mechanism of gene functionality<sup>200, 201</sup>, which essentially allows for cell, tissue and organ differentiation, resulting in phenotypic diversity and specialised function<sup>202</sup>. When there are changes to normal DNA methylation patterns, particularly at CpG islands located in the regulatory regions of genes, adverse phenotypic events have been observed which can promote disease<sup>127, 131, 181, 193, 203, 204</sup>. Methylation aberrations are either a result of external environmental influence or due to genetic variation<sup>111, 205, 206</sup>. The methylation analysis performed during this PhD project has concentrated on potential inherited DNA methylation aberrations by controlling for non-genetic confounders of DNA methylation that are also risk factors of AAA (gender, age, smoking and ethnicity).

This chapter will assess the potential functional effects of differential methylation observed in genes from PBMC sequencing (Chapter 4; *IL6R*, *LDLR* and *SORT1*) and VSMC sequencing (Chapter 5; *ERG*, *IL6R*, *SERPINB9* and *SMYD2*).

**6.1.1 Aims for Chapter 6 – Results section 4**

1. To determine whether changes to CpG methylation in peripheral blood DNA (Chapter 4) are directly associated with blood plasma concentration of the respective gene product.
2. To determine whether changes to CpG methylation in vascular cell DNA (Chapter 5) are directly associated with mRNA expression of the same gene.
3. To establish whether protein expression in whole aortic tissue is evident in genes where associations between methylation and gene functionality are observed.

**6.1.2. Objectives for Chapter 6 – Results section 4**

1. Conduct ELISAs to establish blood plasma protein expression levels of differentially methylated genes from Chapter 4 in 20 AAA and 20 controls.
2. Conduct RT-qPCR gene expression analysis on mRNA from 24 AAA and 20 controls in differentially methylated genes from Chapter 5.
3. Conduct immunohistochemistry on 6 whole aortic tissue samples (3 AAA and 3 controls) where associations between methylation and gene expression are observed in Chapter 5.

**6.1.3. Hypotheses for Chapter 6 – Results section 4**

1. It is hypothesised that where methylation differences are seen in people with AAA compared to controls, differential gene and/or protein expression will also be present, and directly related to the level of CpG methylation.
2. It is hypothesised that where these relationships are observed, whole aortic tissue protein analysis will corroborate them.

## 6.2. Plasma protein analysis to corroborate PBMC methylation

In Chapter 4, bisulphite methylation sequencing was performed on DNA derived from the PBMCs of 48 AAA and 48 controls. In the current chapter, to see if the observed methylation changes in those with AAA are linked with gene function and potentially contribute towards AAA pathobiology, functional corroboration was needed.

Commonly this type of experiment would be performed with gene expression analysis (RT-qPCR) on mRNA extracted from the same buffy coats as the DNA source. However for the samples used in Chapter 4 (PBMC sequencing), this was not possible due to the way in which the blood samples were stored upon collection. To work around this issue, it was decided instead that blood plasma expression assays would be conducted to measure circulating blood concentrations of differentially methylated genes, which included *IL6R*, *SORT1* and *LDLR*. LDL lipid analysis was also conducted due to the role of the LDLR and SORT1 proteins in LDL lipid metabolism<sup>46, 207</sup>.

As described in Chapter 2, ELISAs were performed to acquire blood plasma concentrations from 40 samples (20 AAA and 20 controls; Table 32) for each separate assay. These were then compared to their respective DNA methylation values.

Given that the DNA samples used in this analysis were from the previous PBMC sequencing assay in Chapter 4, all were from Caucasian men over the age of 65 who had, at the time of collection, smoked for at least 10 years.

## Chapter 6: Functional corroboration of differential methylation in AAA

	20 large AAA (>5.5cm)	20 Controls (<25cm)
Median age (IQR)	74 (69-77)	68 (67-72)
Mean aortic diameter (cm (+/- SD))	6.67 (0.66)	1.91 (0.07)

**Table 32 - Demographic summary of samples which were used to assess blood plasma protein concentrations of SORT1, LDLR, LDL and IL6R.**

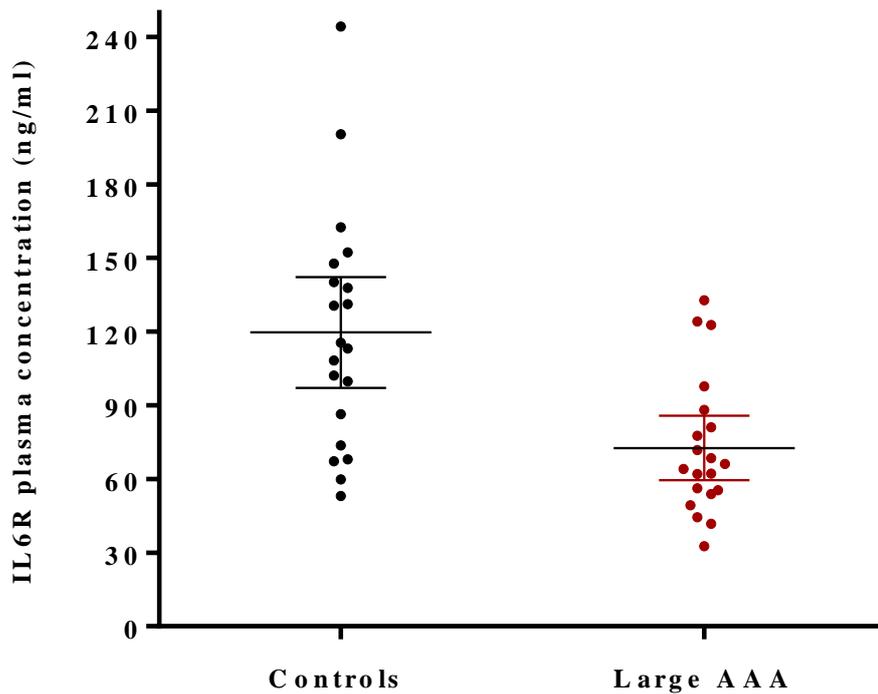
Below in Table 33 is a brief summary of the descriptive statistics from this analysis, and subsequent to that, each assay is described individually.

Protein	Blood plasma concentrations		
	Mean in controls (+/-SE)	Mean in AAA (+/-SE)	P value
IL6R	119.7 ng/ml ± 10.79	72.62 ng/ml ± 6.263	0.0007
LDLR	2.544 ng/ml ± 0.3407	2.436 ng/ml ± 0.358	0.82
SORT1	0.2956 ng/ml ± 0.05719	0.2631 ng/ml ± 0.02591	0.61
LDL	36.95 ug/ml ± 1.074).	36.41 ug/ml ± 2.36	0.834

**Table 33 - Descriptive statistics of the blood plasma assays to corroborate methylation in PBMCs:**  
The proteins which were assessed, mean methylation of controls, AAAs and P values are reported.

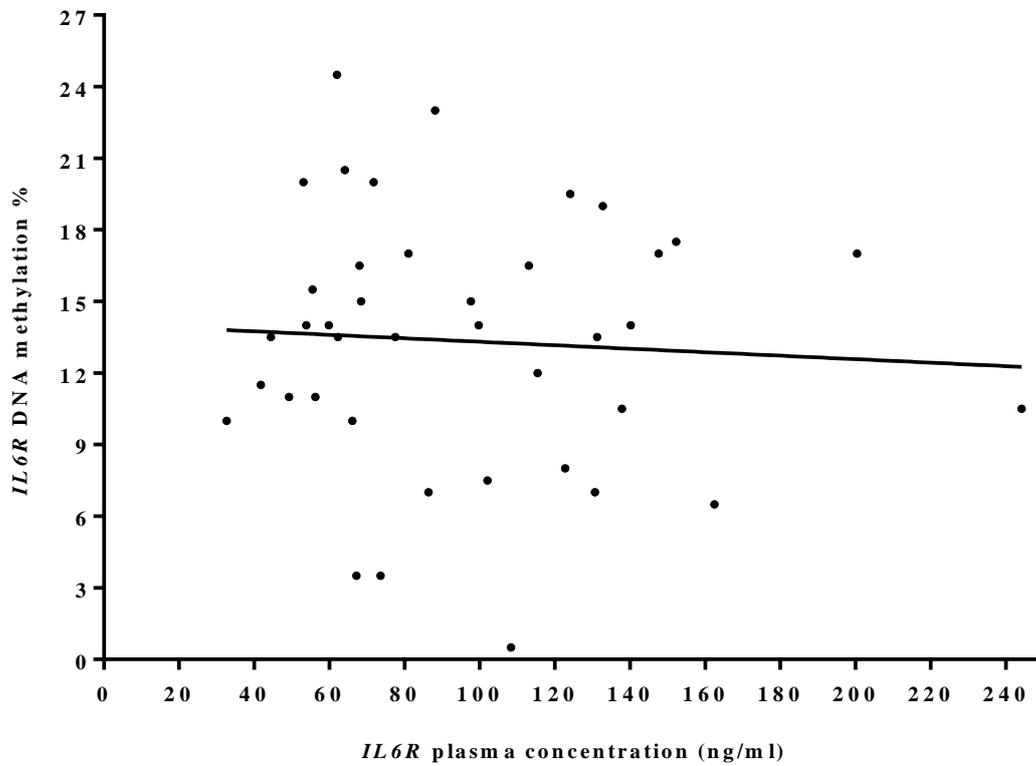
### 6.2.1. IL6R blood plasma analysis

Blood plasma IL6R levels were assessed in 20 AAA and 20 controls as previously described. There was a significant decrease ( $P=0.0007$ ) in those with AAA ( $72.62 \text{ ng/ml} \pm 6.263$ ) compared to controls ( $119.7 \text{ ng/ml} \pm 10.79$ ). These results are displayed in Figure 71.



**Figure 71 – IL6R blood plasma concentrations of 20 AAA and 20 controls:** There was a significant decrease ( $P=0.0007$ ) in those with AAA ( $72.62 \text{ ng/ml} \pm 6.263$ ) compared to controls ( $119.7 \text{ ng/ml} \pm 10.79$ ).

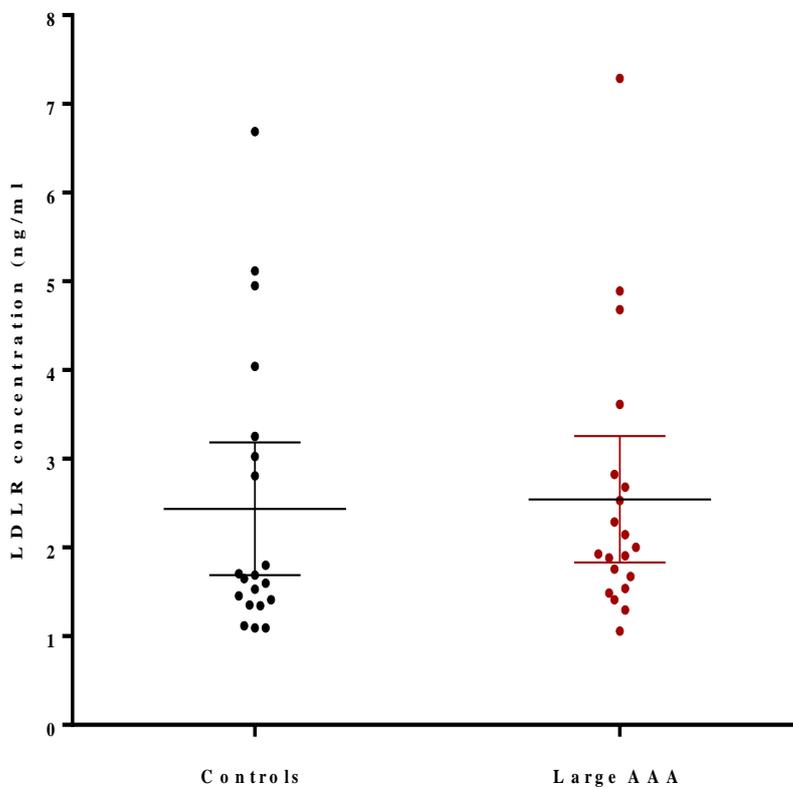
After performing linear regression analysis to determine the relationship between IL6R blood plasma concentration and DNA methylation in the *IL6R* gene promoter, it was evident that there was no correlation after statistical analysis ( $R^2=0.004$  and  $P=0.703$ , Figure 72). This suggests that the difference in circulating IL6R may not be a result of differential methylation in the *IL6R* gene promoter, and the two factors are not directly linked.



**Figure 72 - Linear relationship between IL6R blood plasma and IL6R gene promoter methylation:**  
No correlation was observed ( $R^2=0.004$  and  $P=0.703$ ) in 40 samples (20 AAA and 20 controls).

### 6.2.2. LDLR blood plasma analysis

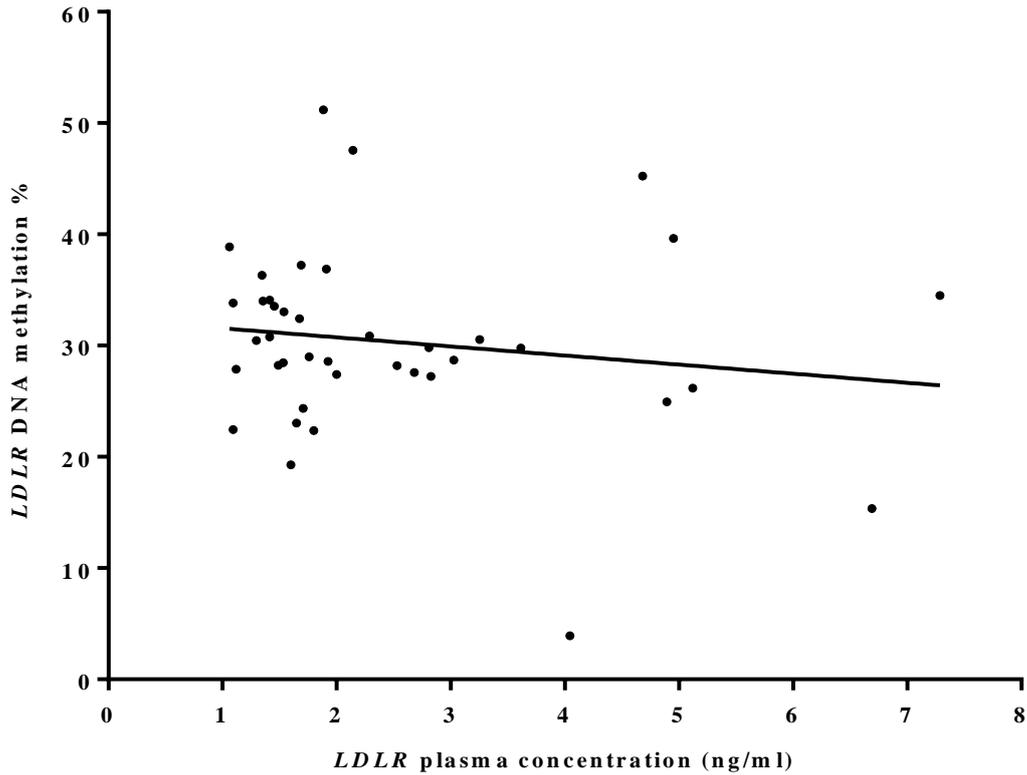
After assessing the circulating blood plasma LDLR concentrations of 20 AAA and 20 controls, it was identified that there was no significant difference ( $P=0.829$ ) between the two groups. Mean circulating LDLR in the controls was  $2.436 \text{ ng/ml} \pm 0.358$ , and mean circulating LDLR in those with AAA was  $2.544 \text{ ng/ml} \pm 0.3407$ . These results are displayed in Figure 73.



**Figure 73 – LDLR blood plasma concentrations of 20 AAA and 20 controls:** Mean circulating LDLR in the controls was  $2.436 \text{ ng/ml} \pm 0.358$ , and  $2.544 \text{ ng/ml} \pm 0.3407$  in the controls ( $P=0.82$ ).

## Chapter 6: Functional corroboration of differential methylation in AAA

When assessing the linear relationship between soluble LDLR and differential methylation in the *LDLR* gene promoter, there was no significant correlation ( $R^2=0.023$  and  $P=0.349$ ) (Figure 74).

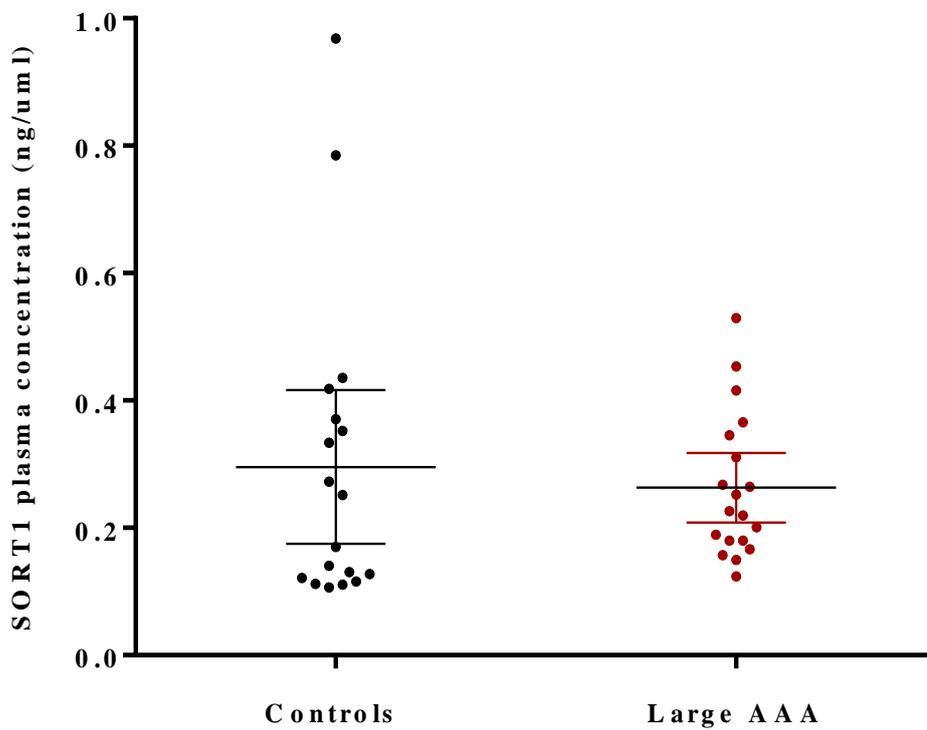


**Figure 74 - Linear relationship between LDLR blood plasma and *LDLR* gene promoter differential methylation:** No correlation was observed ( $R^2=0.023$  and  $P=0.349$ ) in 40 samples (20 AAA and 20 controls).

### 6.2.3. SORT1 blood plasma analysis

SORT1 circulating blood plasma analysis was conducted on the 20 AAA and 20 controls, and there were no significant differences ( $P=0.61$ ) between AAA ( $0.2631 \text{ ng/ml} \pm 0.02591$ ) and controls ( $0.2956 \text{ ng/ml} \pm 0.05719$ ).

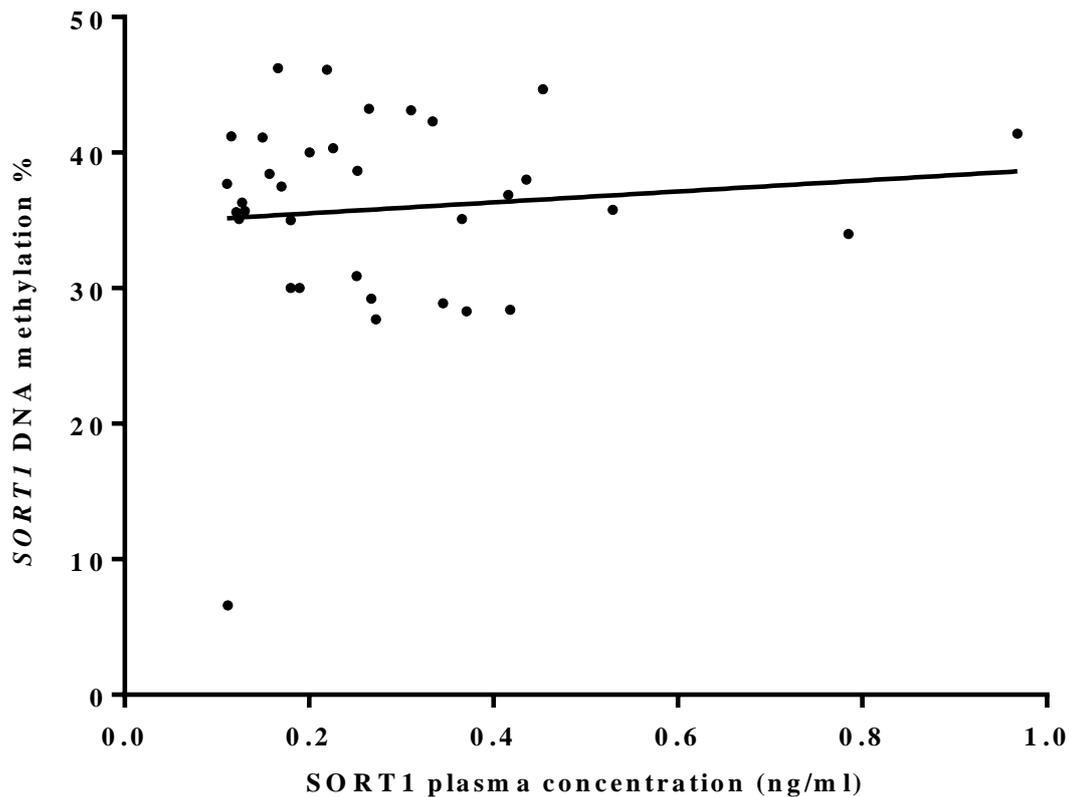
These results are displayed in Figure 75.



**Figure 75 – SORT1 blood plasma concentrations of 20 AAA and 20 controls:** There were no significant differences ( $P=0.61$ ) between AAA ( $0.2631 \text{ ng/ml} \pm 0.02591$ ) and controls ( $0.2956 \text{ ng/ml} \pm 0.05719$ ).

## Chapter 6: Functional corroboration of differential methylation in AAA

Linear regression analysis was conducted to assess the relationship between circulating plasma SORT1 levels, and differential methylation in the SORT1 gene promoter, and there was no statistically significant correlation ( $R^2=0.01$  and  $P= 0.562$ ) (Figure 76).



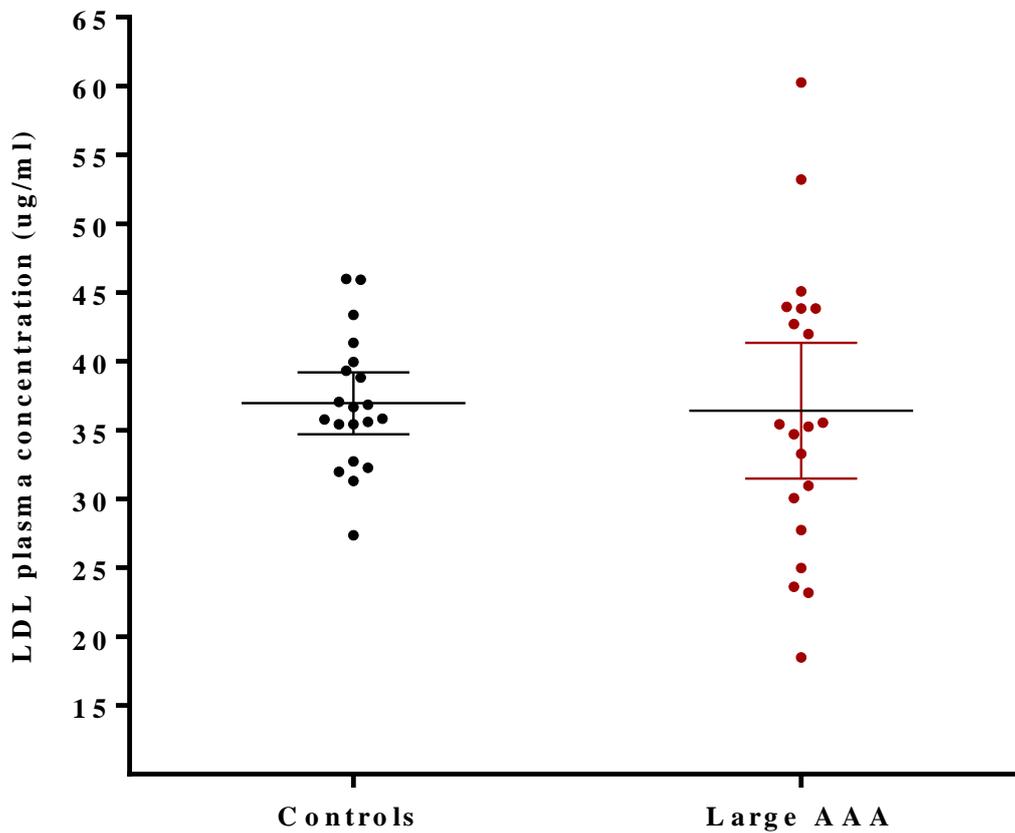
**Figure 76 - Linear relationship between SORT1 blood plasma and SORT1 gene promoter differential methylation:** No correlation was observed ( $R^2=0.01$  and  $P= 0.562$ ) in 40 samples (20 AAA and 20 controls).

### 6.2.4. LDL blood plasma analysis

The results from the circulating LDL assay revealed there was no significant difference in the blood plasma concentration of LDL ( $P=0.834$ ) between 20 AAA ( $36.41 \text{ ug/ml} \pm 2.36$ ) and 20 controls ( $36.95 \text{ ug/ml} \pm 1.074$ ).

Although there were no differences between the two groups it is still very clear that the deviation in LDL lipid profile of those with AAA is much larger than healthy controls.

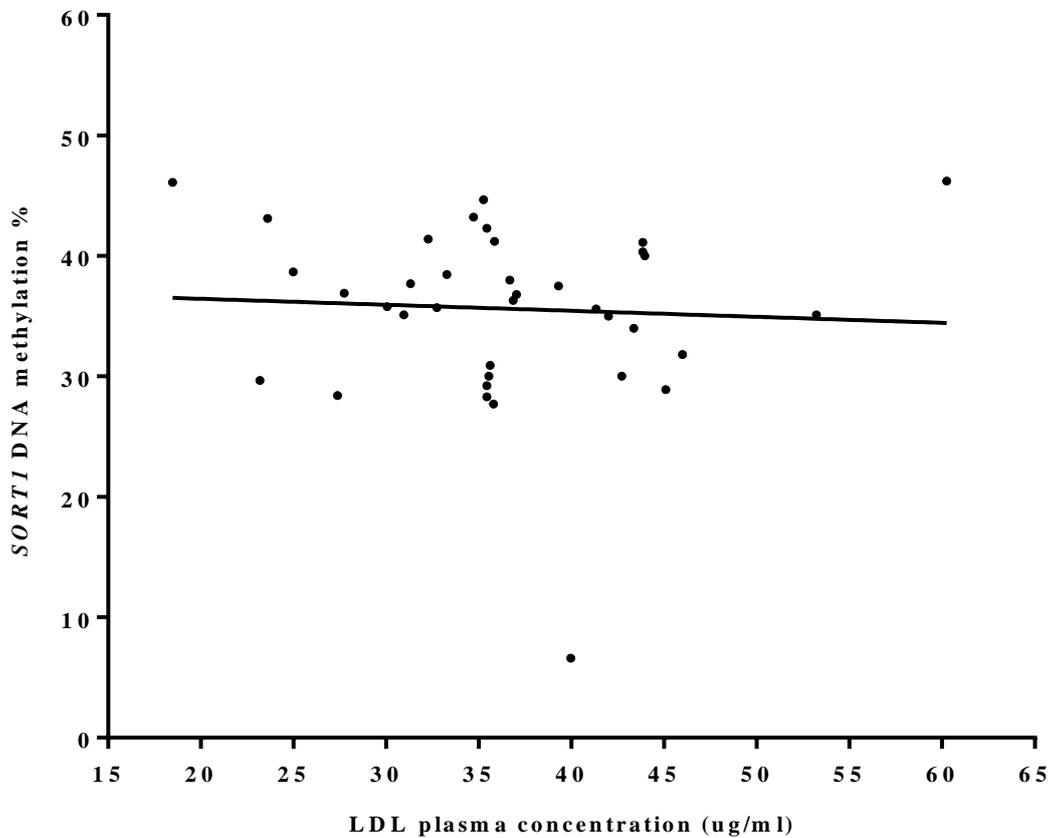
The results from this analysis are shown in Figure 77.



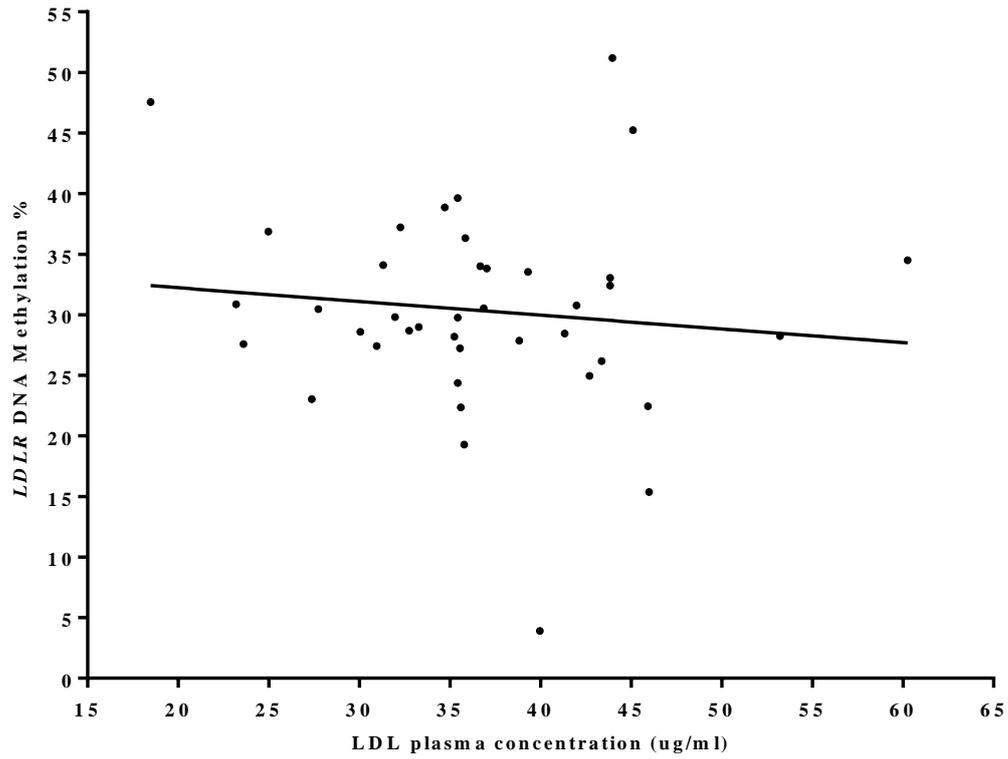
**Figure 77 – LDL blood plasma concentrations of 20 AAA and 20 controls:** There was no significant difference in the blood plasma concentration of LDL ( $P=0.834$ ) between 20 AAA ( $36.41 \text{ ug/ml} \pm 2.36$ ) and 20 controls ( $36.95 \text{ ug/ml} \pm 1.074$ ).

## Chapter 6: Functional corroboration of differential methylation in AAA

Linear regression was then performed to assess the relationship between LDL blood plasma concentrations, and differential methylation in the *SORT1* and *LDLR* gene promoters at previously identified differentially methylated sites. There was no significant relationship between plasma LDL and *SORT1* gene promoter methylation ( $R^2=0.003$  and  $P= 0.737$ ; Figure 78), or plasma LDL and *LDLR* gene promoter methylation ( $R^2=0.012$  and  $P=0.496$ ; Figure 79).



**Figure 78 - Linear relationship between LDL blood plasma and *SORT1* gene promoter differential methylation:** No correlation was observed ( $R^2=0.003$  and  $P= 0.737$ ) in 40 samples (20 AAA and 20 controls).



**Figure 79 - Linear relationship between LDL blood plasma and *LDLR* gene promoter differential methylation:** No correlation was observed ( $R^2=0.012$  and  $P=0.496$ ) in 40 samples (20 AAA and 20 controls).

### 6.3. Gene expression analysis to corroborate VSMC methylation

In genes where differentially methylated regions were identified in Chapter 5 (VSMC sequencing; *ERG*, *IL6R*, *SERPINB9* and *SMYD2*), mRNA gene expression analysis was performed in the same samples (Table 20, reproduced from Chapter 5) to assess the relationship between DNA methylation and gene expression. This analysis was conducted as previously described in Chapter 2.

	24 Large AAA (>5.5cm)	20 Cadaveric controls (Exact aortic diameters unknown)
No of males	24	16
No of females	0	4
Ethnicity	Caucasian	Caucasian
Smoker	22	15
Median age (IQR)	68 (65-73)	56 (49-62)
Mean aortic diameter (mm (+/- SD))	6.5 (1.2)	<3

**Table 20: Demographic summary of samples where vascular smooth muscle cell DNA was bisulphite treated and used for next generation sequencing.**

Figure 80 displays the results for this experiment, and Table 34 is a brief summary of the descriptive statistics.



## Chapter 6: Functional corroboration of differential methylation in AAA

Gene	Gene expression values (-delta Ct)		P value
	Mean in controls (+/-SE)	Mean in AAA (+/-SE)	
<i>SMYD2</i>	-6.55 (0.15)	-8.27 (0.25)	0.0000008
<i>SERPINB9</i>	-8.39 (0.27)	-10.61 (0.34)	0.00001
<i>IL6R</i>	-12.51 (0.62)	-13.04 (0.74)	0.588
<i>ERG</i>	-10.81 (0.73)	-9.397 (0.42)	0.096

**Table 34 - Descriptive statistics of the gene expression assays to corroborate methylation in VSMCs:**  
The genes which were assessed, mean methylation of controls, AAAs and P values are reported.

There was a significant overall reduction in the expression of *SMYD2* ( $P < 0.0001$ ) in AAA (-8.28 +/- 0.25) compared to controls (-6.55 +/- 0.15).

There was also a significant reduction in expression of *SERPINB9* ( $P < 0.0001$ ) in people with AAA (-10.61 +/- 0.34) compared to controls (-8.39 +/- 0.27).

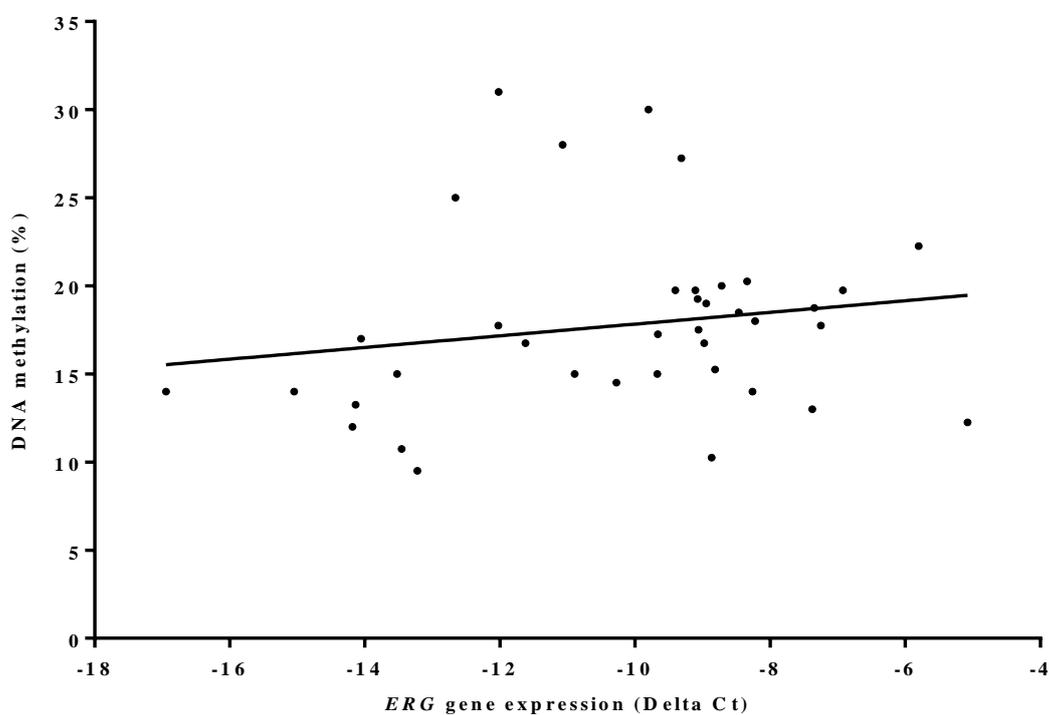
However, there were no differences in expression in *IL6R* ( $P = 0.588$ ) and *ERG* ( $P = 0.096$ ) mRNA from individuals with AAA compared to controls.

### 6.3.1. Gene expression and DNA methylation analysis

Differentially methylated CpG sites in each of the 4 genes were compared to their respective gene expression levels for each individual sample where expression and methylation values were acquired.

#### 7.3.1.1 *ERG* expression and DNA methylation

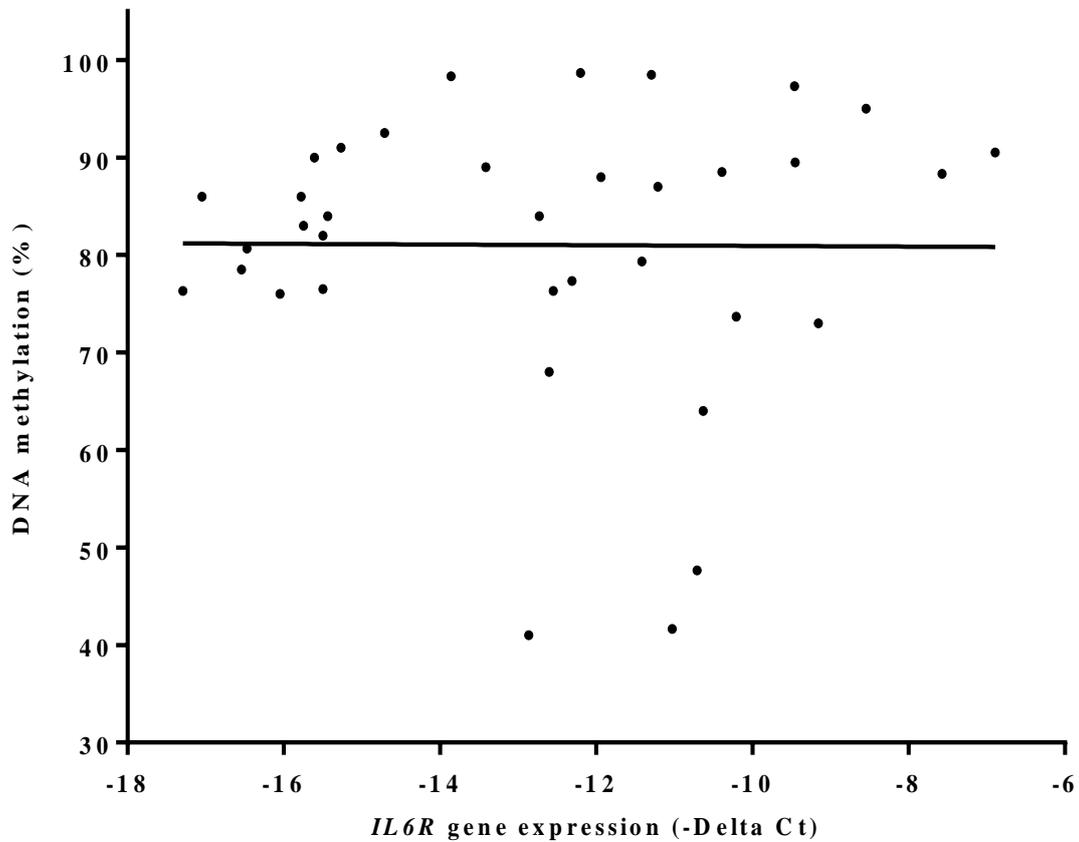
There were no statistically significant differences in gene expression between AAA and controls in *ERG* ( $P=0.096$ ), and there was also no relationship between the differential methylation of the *ERG* gene promoter and *ERG* gene expression ( $R^2=0.03$  and  $P=0.2955$ ) (Figure 81).



**Figure 81 - Linear regression between gene expression and mean DNA hypermethylation in *ERG*:** This graph shows the mean methylation values of the 13 consecutively hypermethylated CpGs (NG\_029732.1: 5389-5507) which were not associated with *IL6R* gene expression ( $P=0.31$ ,  $n=38$ ).

7.3.1.2. *IL6R* expression and DNA methylation

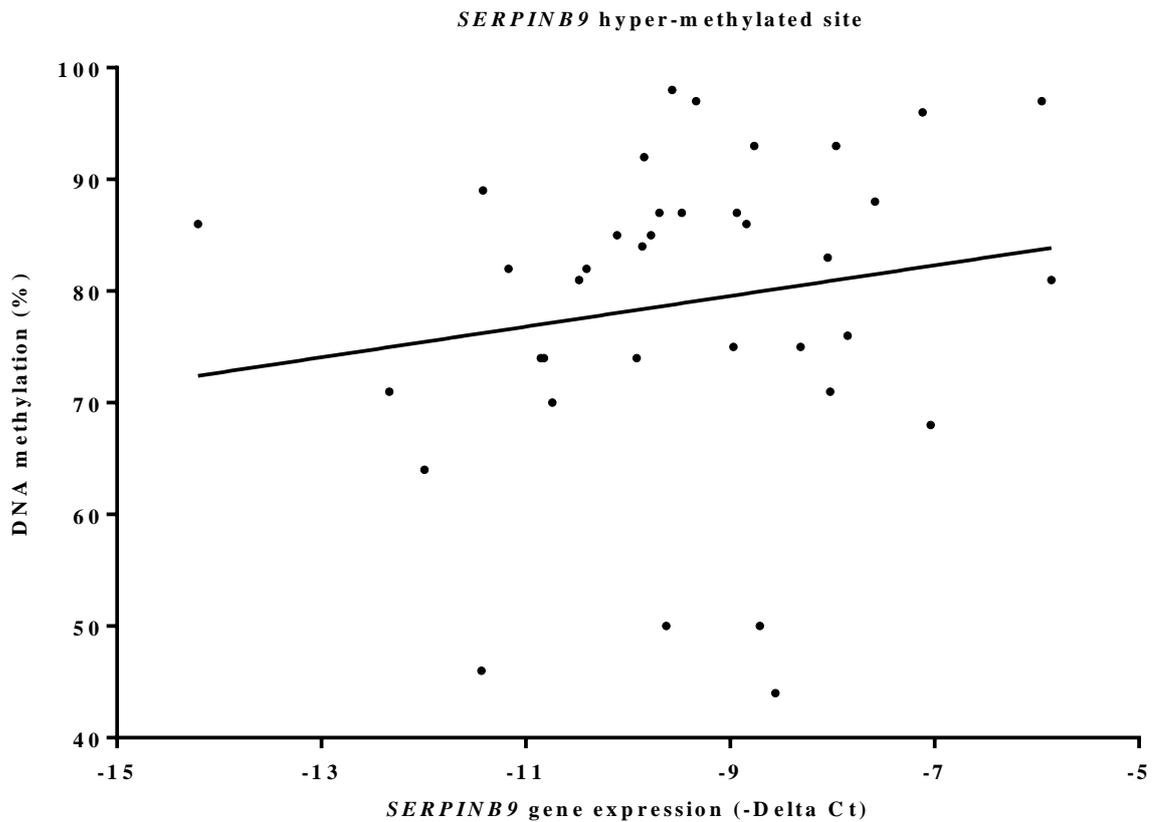
When conducting linear regression between *IL6R* gene expression and differentially methylated CpGs in the *IL6R* gene promoter, no statistically significant relationship was seen ( $R^2=0.00004$  and  $P=0.96$ ) (Figure 82).



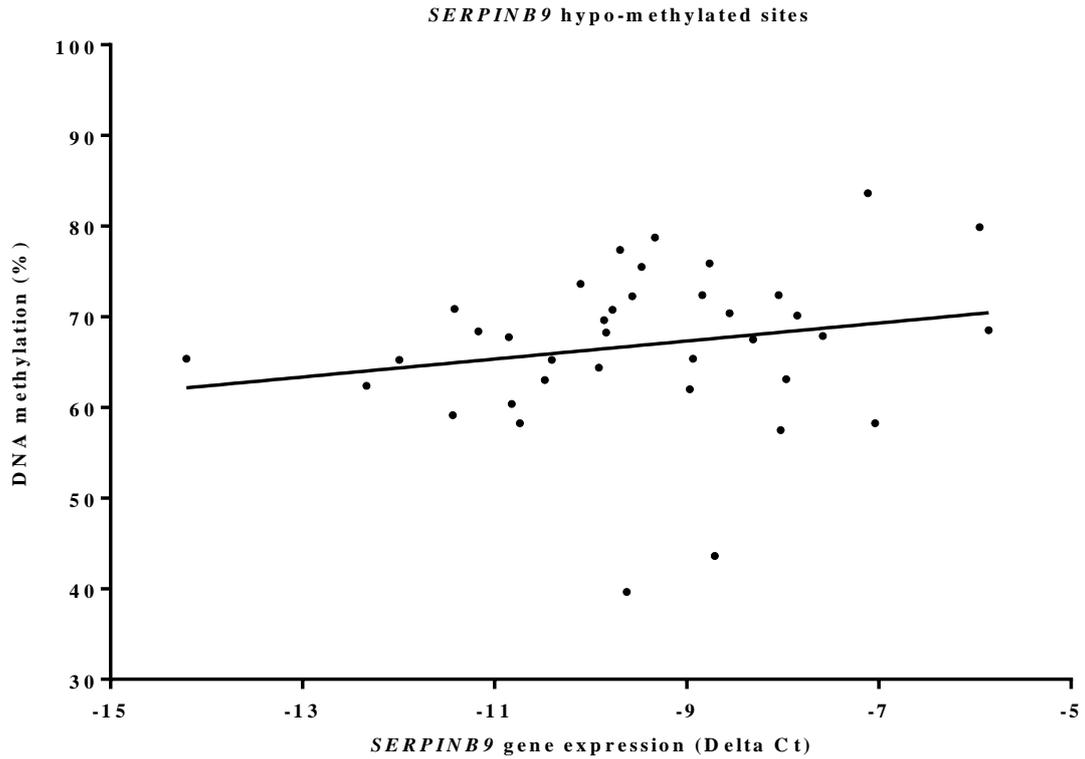
**Figure 82 - Linear regression between gene expression and mean DNA hypermethylation in *IL6R*:** This graph shows the methylation values of the two hypermethylated CpGs (3570 and 3676), which were not associated with *IL6R* gene expression ( $P=0.31$ ,  $n=38$ ).

7.3.1.3. *SERPINB9* expression and DNA methylation

In the *SERPINB9* gene, where significant expression and methylation differences were observed, neither the hyper nor hypo-methylated CpG sites were associated with *SERPINB9* expression levels (P=0.31 and P=0.243 and respectively) (Figure 83 and 84).



**Figure 83 - Linear regression between gene expression and mean DNA hypermethylation in *SERPINB9*:** This graph shows the methylation values at the single hypermethylated CpG 2889644 (P=0.31, n=38), which was not associated with *SERPINB9* gene expression.



**Figure 84 - Linear regression between gene expression and mean DNA hypo-methylation in *SERPINB9*.** : This graph shows the methylation values of the Hypo-methylated CpGs; 2889674, 2889627, 2889620, 2889612, 2889568, 2889564 (P=0.243, n=37), which were not associated with *SERPINB9* gene expression.

7.3.1.4. *SMYD2* expression and DNA methylation

The mean DNA methylation percentage of differentially methylated CpGs (214280412, 214280441, 214280507 and 214280600) in the *SMYD2* gene promoter was significantly associated with *SMYD2* gene expression ( $P=0.0383$ ,  $R^2 = 0.17$ , (Figure 85)).

This significant, positive association between differential gene promoter DNA methylation and differential gene expression of the same gene signifies a potential causal relationship between the two factors, given that methylation is a common regulator of gene expression. However, this cannot be stated without further work.

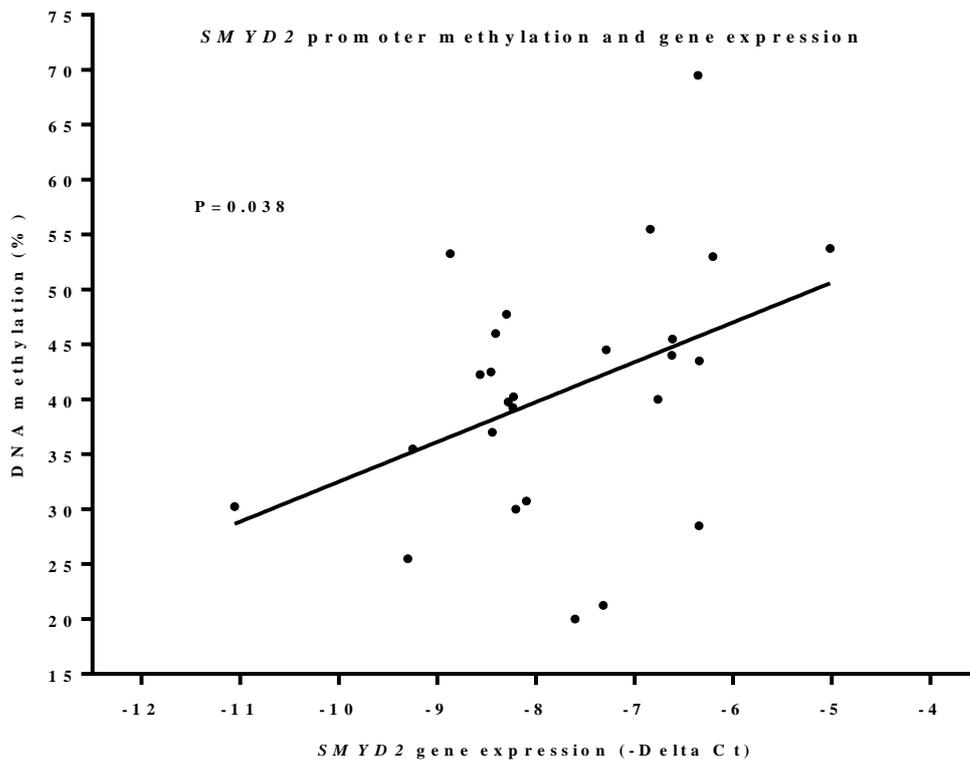


Figure 85 - Linear relationship between gene expression and mean DNA methylation in *SMYD2*:

Mean methylation status of CpGs (NC\_000001.11: 214280412, 214280441, 214280507 and 214280600) in the *SMYD2* promoter were significantly associated with *SMYD2* gene expression (n=26 where sufficient sequencing coverage and expression data was acquired).

#### 6.4. Immunohistochemistry to assess *SMYD2* in aortic tissue

It was described in the previous section that differential methylation in the *SMYD2* gene promoter was directly related to gene expression, and there were no such relationship shown in *ERG*, *IL6R* or *SERPINB9*. Specifically, DNA hypo-methylation was seen in the same VSMC samples where there was a lower mRNA expression of *SMYD2*. This represents a potentially significant concept, because differential gene activity as a consequence of differential promoter methylation may reveal a new insight to the pathobiology of the disease. To further corroborate this relationship, and to ensure that the differential gene activity between AAA and controls was applicable to the site of aneurysm in the aortic wall, histological staining of the Smyd2 protein was conducted in whole aortic tissues from 6 of the same samples used in methylation and gene expression analysis (3 AAA and 3 controls).

For each of the 6 samples, 3 individual stains were conducted (total = 18 slides) on tissue slices from the same aortic sections. One was for Smyd2 staining, one was for a Smyd2 control with no primary antibody, and one was for smooth muscle actin (SMA) staining. The controls were checked for the presence of Smyd2, and all were absent of brown colouring, which represents the presence of primary antibody. The SMA slides were then visualised and where the densest sections of smooth muscle fibres were seen, the Smyd2 slides were visualised and photographs were taken. Results are shown in Figure 86 which shows a comparison of the Smyd2 protein expression in VSMCs in abdominal aortic tissues of those with AAA compared to controls.

SMYD2 antibody staining

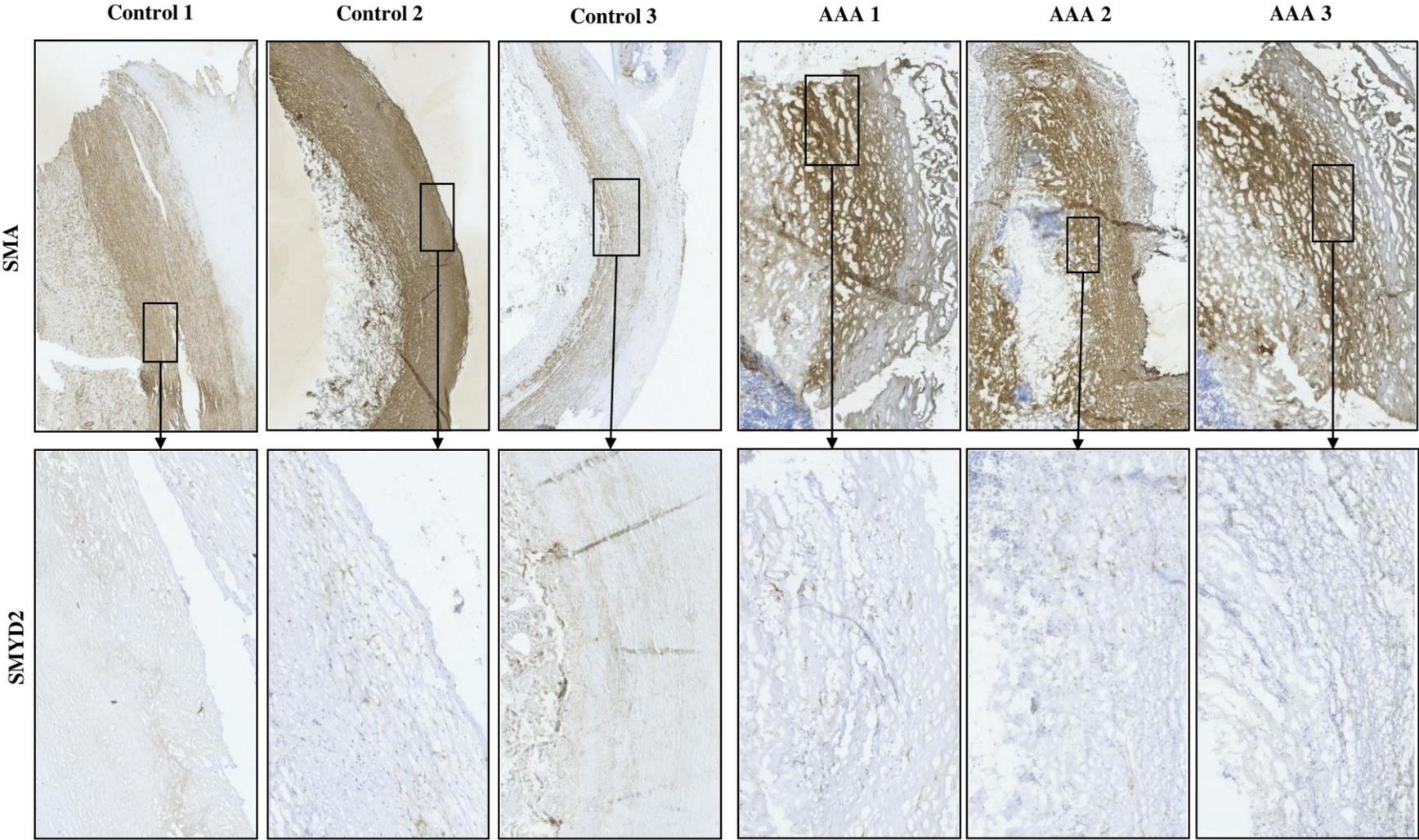


Figure 86 - Immunohistochemical staining for smooth muscle actin (SMA) and Smyd2: Staining of aneurysmal (n=3) and non-aneurysmal (n=3) abdominal aortas was conducted. Brown = presence of Smyd2antibody.

## Chapter 6: Functional corroboration of differential methylation in AAA

The adopted histological technique for this assay is hard to quantify, and is not as objective as other methodologies such as RT-qPCR for establishing gene expression for example. However, Figure 86 subjectively shows a reduced abundance of the Smyd2 protein (brown colouring) in the tissues of those with AAA compared to controls. Whilst a statistical analysis is not available, it appears that the same pattern of downregulation (from the mRNA gene expression analysis) is evident in AAA compared to controls, which corroborates the findings from the previous section. This demonstrates that transcript expression differences are also reflected from a translational protein perspective.

### 6.5. Conclusion

This chapter is the final results chapter of this thesis, in which the potential functional effects of differential methylation observed in genes from PBMC sequencing (Chapter 4, *IL6R*, *LDLR* and *SORT1*) and VSMC sequencing (Chapter 5, *ERG*, *IL6R*, *SERPINB9* and *SMYD2*) was conducted. Circulating *IL6R*, *SERPINB9* gene expression and *SMYD2* gene expression was decreased in those with AAA, and there was a linear relationship between *SMYD2* gene expression and *SMYD2* promoter methylation. Further to this, the expression pattern of the Ssmyd2 protein was assessed in aortic tissues using immunohistochemistry and visualised at the most VSMC dense regions. This analysis highlighted a reduced abundance of Ssmyd2 in the abdominal aortas of AAA compared to controls and further corroborated the *SMYD2* gene methylation/expression relationship. However, many limitations are applicable to this chapter overall, which especially relate to the PBMC experiments. Direct gene expression analysis could not be conducted in PBMC mRNA to corroborate PBMC DNA methylation changes. Due to the manner in which the UKAGS samples are stored after collection (patient blood samples are

## Chapter 6: Functional corroboration of differential methylation in AAA

centrifuged and frozen), it was not possible to obtain mRNA for candidate gene expression analysis, as freezing at -20 °C is not a viable storage option to prevent the degradation of mRNA over long term periods of time. Therefore gene expression analysis could not be conducted. As an alternative, blood plasma protein assays were performed to functionally corroborate methylation changes, however LDLR, SORT1, LDL and IL6R blood plasma concentrations were not linked with DNA methylation in any gene. It is possible that with the conduct of a more appropriate methodology such as RT-qPCR, there would be more certainty of the results.

Although DNA methylation status of the genes sequenced from the PBMC assay were not associated with gene functionality, the results obtained during PBMC sequencing (Chapter 4) are still of significant value, given there are clear signs of consistent hypermethylation in the regulatory regions of *SORT1*, *LDLR* and *IL6R*. This represents proof of concept and justification for more complete, large scale analysis of these loci as potential targets in AAA pathobiology. It would be suggested that this future analysis was conducted on samples where direct gene expression analysis could be performed, which would help alleviate any doubt regarding the results gained from the current study. This is important due to the fact that ELISAs to assess circulating protein/lipid levels are not as accurate or representative of gene function as direct mRNA gene expression analysis conducted via RT-qPCR.

Gene expression and blood plasma protein analysis for corroboration essentially does not establish a causal role of methylation on gene functionality. It is a novel concept in this study that differential DNA methylation is correlated with AAA, however there is also an important limitation because the study did not test if the correlation has any causality to AAA development, it only measured associations. Mechanistic studies such as luciferase

## Chapter 6: Functional corroboration of differential methylation in AAA

reporter assays would be ideal in the future to establish the functional role of the observed associations of this study.

In summary, the linear relationship that was observed between *SMYD2* methylation and gene transcript expression, in addition to the corroboration of protein expression in abdominal aortic tissues in this Chapter, suggests that hypo-methylation in the *SMYD2* gene promoter could play a role in decreased *SMYD2* expression and consequent irregular cardio-physiology. The value and implications of all the findings from this thesis will be discussed in the next chapter.

**7. Chapter 7 - Discussion**

## 7.1. Summary of the potential role of DNA methylation in AAA

DNA methylation patterns are long term inherited signatures that are passed through generations, and can be permanently or temporarily affected by both genetic and environmental factors<sup>208, 38</sup>. Changes in DNA methylation are involved in other cardiovascular diseases such as hypertension and atherosclerosis, both of which may be contributing factors to the development of AAA<sup>102, 140</sup>. Additionally, aberrant DNA methylation changes have been associated with age and smoking<sup>159, 160</sup>, as well as implicated in MMP induced vascular remodelling and inflammation, which are important pathogenic mechanisms in AAA development<sup>27</sup>. Variations in DNA sequence at polymorphic loci can result in variable patterns of DNA methylation. These genomic sites are known as methylation quantitative trait loci (meQTL), and are indicative of a direct genetic influence on levels of DNA methylation<sup>120, 126, 121</sup>. There is some knowledge of meQTL in disease, but this understanding is limited and currently there is no comprehensive database highlighting specific genotype-methylation associations.

It therefore is likely that there are unreported disease specific meQTL, and known polymorphic risk loci identified in previous AAA GWASs represent potentially insightful targets for further genetic and methylation analysis. The identification of adverse epigenetic modifications directly linked to patients with AAA could offer a more comprehensive understanding of AAA pathobiology, and an alternative research avenue in the search for a potential future treatment strategy<sup>15, 16</sup>.

Through an extensive review of the literature, which is presented in Chapter 1, the majority of factors associated with AAA are also associated with DNA methylation, yet there has been very limited investigation. The only study to date was conducted by Ryer *et al.*,<sup>143</sup> which was published mid-way through this PhD project, who conducted a methylation study on AAA in

peripheral blood DNA from 20 AAA patients (11 smokers and 9 non-smokers) and 21 control samples (10 smokers and 11 non-smokers) using Illumina 450k bead chip arrays. They identified differentially methylated regions in *ADCY10P1*, *CNN2*, *KLHL35*, and *SERPINB9*, but only *CNN2* and *SERPINB9* were differentially expressed. As a result of the limited investigation in to the epigenetic basis of AAA, in addition to the clear rationale for further study, this PhD project was designed to investigate the role of DNA methylation in AAA through adopting hypothesis based aims and objectives.

## **7.2. Summary of findings**

This thesis has demonstrated an original contribution to knowledge at each experimental stage, providing further insight to the epigenetic basis of AAA. Aberrant DNA methylation patterns have been identified which may contribute towards adverse gene function and the development of AAA. Various approaches were adopted in order to assess these relationships, and this is the first work to assess DNA methylation in AAA using NGS.

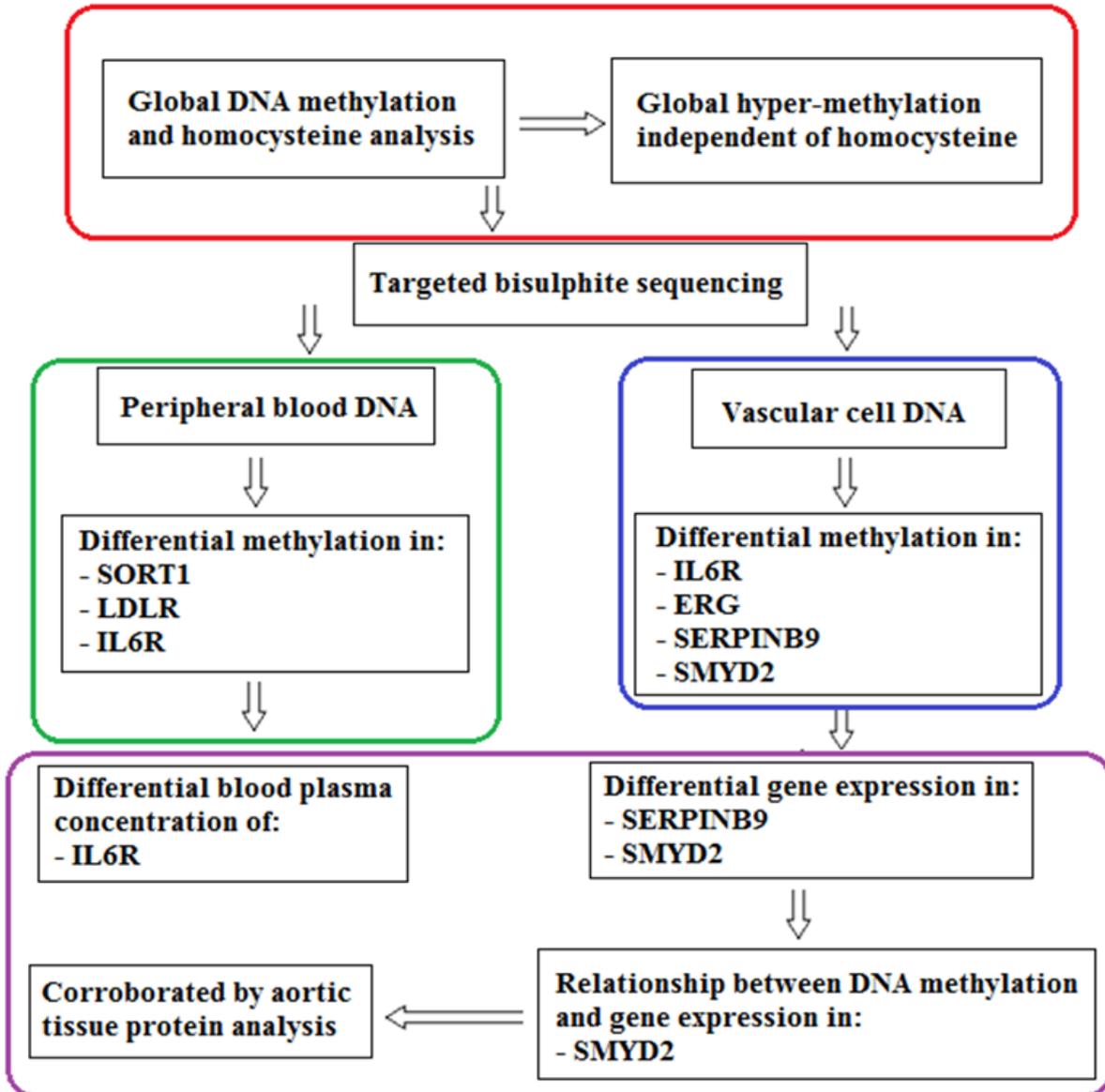
Firstly, global DNA hypermethylation (45 small AAA, 48 large AAA and 92 controls) was identified in those with a large AAA, and a linear association between AAA and aortic diameter was illustrated independently of circulating homocysteine (70 AAA and 67 controls). In addition, these results were independent of age, sex, ethnicity and smoking, likely highlighting the role of inflammation on global DNA methylation.

Regulatory regions of candidate genes known to associate with AAA at genomic risk loci were then sequenced using bisulphite NGS, and methylation patterns were found to differ in PBMC (48 AAA and 48 controls) and VSMC (24 AAA and 20 controls) DNA. Blood plasma protein analysis (20 AAA and 20 controls) and RT-qPCR gene expression analysis (24 AAA and 20 controls) revealed a decrease in circulating IL6R, *SERPINB9* gene expression and *SMYD2* gene expression in those with AAA.

Finally, a linear relationship was observed between *SMYD2* promoter methylation and *SMYD2* gene expression, suggesting that hypo-methylation in the *SMYD2* gene promoter could play a role in decreased *SMYD2* expression and AAA development. The decreased pattern of gene expression in those with AAA was further corroborated with the conduct of histological analysis in abdominal aortic tissues from 3 AAA and 3 controls.

Figure 87 is a flow chart summary of the results from the entire PhD project, and shows which Chapter the results apply to.

Table 35 is a timeline containing all of the work that was conducted during this PhD, and also includes each significant milestone which is directly related to the success of the project.



Chapter 3 - Results section 1

Chapter 4 - Results section 2

Chapter 5 - Results section 3

Chapter 6 - Results section 4

**Figure 87 – Summary flow chart of the results from this PhD project:** DNA was isolated from peripheral blood and aortic tissues of those with AAA and controls. Global DNA hypermethylation was identified in peripheral blood DNA, and gene specific differential methylation was identified in both sample resources. Changes in gene and plasma protein expression were identified, and gene functionality was associated with methylation in *SMYD2*. The chart displays the main positive findings from each experiment and chapter in which the results are described.

Month	Year 1 (2014-2015)												Year 2 (2015-2016)												Year 3 (2016-2017)											
	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	
Re-design project and academic committee approval	■	■																																		
Global methylation and homocysteine analysis		■	■	■	■																															
Write and publish first review article	■	■	■	■	■	■	■	■																												
PBMC bisulphite sequencing assay							■	■	■	■	■																									
Write probation report and attend review									■	■	■																									
Learn command line and bioinformatics											■	■	■	■	■																					
Present work at vascular society conference													■																							
5 day advanced bioinformatics course in Cambridge													■																							
Present work at SARS conference															■																					
3 day bioinformatics course in Leicester															■	■																				
Bioinformatics and statistical analysis of PBMC data															■	■	■	■	■	■	■	■	■													
Write and publish second review article																	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		
VSMC bisulphite sequencing assay																				■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		
Departmental seminar 30 minute presentation																				■																
Write and publish third review article																					■	■	■	■	■	■	■	■	■	■	■	■	■	■		
Write second year progress report and attend review																					■	■	■	■	■	■	■	■	■	■	■	■	■	■		
Present work at IMAD conference																						■														
Bioinformatics and statistical analysis of VSMC data																							■	■	■	■	■	■	■	■	■	■	■	■		
Functional corroboration of differential methylation																								■	■	■	■	■	■	■	■	■	■	■		
Write PhD results as paper and submit to JAHA																																				
Write thesis	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		
Write postdoc fellowship and job applications																												■	■	■	■	■	■	■		
Submit thesis																																		■		
Address reviewers comments and resubmit paper to JAHA																																		■		

**Table 35 – Timeline of PhD project:** All significant milestones which are directly related to the success of the PhD project, along with month and year in which they were conducted and completed are displayed.

### 7.3. Discussion of value and implications of findings

#### 7.3.1. Global DNA methylation and homocysteine

DNA methylation and histone modifications are the main epigenetic mechanisms responsible for the regulation of functionally dynamic gene expression pathways of protein coding genes, and are essential for genome stability<sup>200</sup>. Changes to normal CpG methylation patterns in genes, or in the regulatory regions of these genes, can contribute towards various traits and diseases<sup>102, 103, 125-127</sup>. However global methylation in non-protein coding regions, which comprises the majority of the genome is also functionally important<sup>209</sup>. Gene specific DNA methylation patterns are located in euchromatin, whilst global DNA methylation is more commonly associated with retaining genome stability in heterochromatic regions, which contain long interspersed repetitive and transposable elements<sup>116, 210</sup>. The majority of the genome is wound tightly in to the heterochromatic transcriptionally inactive state, which is also where the majority of methylated DNA is located. Methylation and histone modifications work in close concert to maintain epigenetic regulation<sup>209</sup>, and chromatin structure is controlled by histone modifications such as acetylation and deacetylation enzymes, resulting in the winding and unwinding of chromatin between the two different states<sup>115</sup>.

The importance of global methylation can be significant in disease, considering that some cancers are characterised by loss of global methylation which essentially destabilises the genome and subsequently impacts cell function and proliferation<sup>211-213</sup>. Although in contrast to this, and more relevant to this work, global DNA hypermethylation is commonly a hallmark of chronic inflammation and has been observed as a potential pathological marker of disease<sup>102, 147-149, 154, 155</sup>. It has previously been shown that an increase of pro-inflammatory cytokines can result in the up-regulation of DNA

methyltransferases<sup>214</sup>, highlighting one mechanism which could go some way to explaining the results in Chapter 3 of this thesis (global DNA hypermethylation is associated with large AAA and increasing AAA size). These results were independent of smoking, sex, age and ethnicity, highlighting the potential role of aberrant global methylation patterns in AAA as a result of chronic inflammation.

Homocysteine levels have been previously associated with global methylation<sup>189</sup> and is a proposed biomarker for cardiovascular disease<sup>166</sup>. It has also been suggested to play a role in AAA<sup>164</sup>. However, the differences seen in homocysteine between patients with AAA and controls observed in this study were not clinically different (hyperhomocysteinemia is described above 15  $\mu\text{mol/L}$ ), and supports the idea that homocysteine is not an appropriate biomarker for AAA as it is for cardiovascular risk, which is also suggested by Lindqvist *et al.*,<sup>172</sup>. In this project there was no linear relationship between homocysteine and global DNA methylation, suggesting that global DNA hypermethylation may be associated with AAA independently of circulating blood homocysteine in PBMC DNA. Similarly to other inflammatory diseases, increased global DNA methylation may be a factor in the pathobiology of AAA, and represents a hallmark of inflammatory disease. These results were used as a foundation and rationale for further in depth analysis of the role of DNA methylation in AAA.

### 7.3.2. PBMC DNA bisulphite sequencing

Significant levels of DNA hypermethylation were observed in the *IL6R* gene promoter in those with AAA compared to controls from PBMC and VSMC DNA (Chapter 4 and 5) and signified a potentially promising investigation in to the functional effects.

Subsequently, analysis of circulating IL6R from those whose PBMC DNA was

hypermethylated in the *IL6R* gene promoter were lower than those without AAA (Chapter 6), which is consistent with previous literature.

Two forms of Interleukin-6 (IL-6) signalling have been described; IL-6 binds the membrane bound *IL6R* which is expressed in leukocytes, and IL-6 binds to the circulating soluble *IL6R*. It has previously been shown that the expression of IL-6 and subsequent signalling are greater in AAA than in non-aneurysmal aortic tissue. An increase in circulating *IL6R* levels are in turn associated with the rs2228145 sequence variant which is protective against AAA and CAD<sup>80, 215-218</sup>. Essentially, reduced circulating *IL6R* is associated with increased inflammation and contributes towards AAA pathobiology<sup>218</sup>. However, the decreased circulating *IL6R* identified during this PhD was not associated with *IL6R* CpG promoter hypermethylation after linear regression analysis, suggesting that there may be no direct functional impact of differential methylation in the *IL6R* promoter.

Similarly, in VSMC DNA, there was no relationship between expression of the *IL6R* gene and *IL6R* CpG promoter hypermethylation. Although there was no relationship between *IL6R* gene function and methylation in PBMC or VSMC DNA, it would still be worth including *IL6R* in a large-scale replication experiment due to the clear levels of observed differential methylation in both cell types, which could turn out to have alternative functional effects than downstream *cis* regulation.

After PBMC DNA bisulphite sequencing (Chapter 4), the promoters of *SORT1* and *LDLR* were considerably and consistently hypermethylated in those with AAA. LDL metabolism is implicated as an important mediator of AAA development. *LDLR* rs6511720 is one of the most important LDL cholesterol variants in studies of lipid traits<sup>45</sup>. Importantly, a relationship between AAA and LDL cholesterol is not presently

established, even though there are clear implications of *LDLR* genetic polymorphisms in AAA, which is also shared with CAD<sup>219</sup>.

In line with this, there were no differences between circulating LDL levels in AAA and controls in this PhD project (Chapter 6). Previous individual reports in *LDLR* and LDL lipid profiles are inconsistent in AAA studies<sup>26, 45, 75</sup>, and it is probable that *LDLR* variation has an effect on another biological pathway in AAA development. This concept was also suggested by Jones *et al.*,<sup>83</sup> who revealed a central role for *MMP9* in AAA, and interactions between *ERG*, *IL6R* and *LDLR* as modifiers of *MMP9*. It is therefore unsurprising that no differences in circulating *LDLR* or LDL were observed in this PhD.

*SORT1* on the other hand, which is another important receptor involved in LDL lipid profile is independently associated with AAA (rs602633) and has been identified as differentially expressed in AAA<sup>83</sup>. *SORT1* is actually an eQTL<sup>83, 137</sup>, and as such, expression patterns may be regulated by differential DNA methylation, which is exactly what was observed in the PBMC DNA of those with AAA during this PhD. Circulating *SORT1* analysis is likely not a valuable measure of *SORT1* gene functionality as a result of differential *SORT1* promoter hypermethylation, and because of this it is still viable that the methylation changes observed in *SORT1* in this project are functionally active in AAA pathobiology. However, to assess this concept, further analysis is required where *SORT1* promoter methylation is compared directly to levels of *SORT1* gene function.

### **7.3.3. VSMC DNA bisulphite sequencing**

VSMCs are the main constituents of the tunica media, where cellular apoptosis occurs as a hallmark of AAA during aneurysm development and growth<sup>33</sup>. After the VSMC DNA bisulphite sequencing assay (Chapter 5), it was observed that differential methylation was present in *SMYD2*, *SERPINB9* and *ERG*, in addition to *IL6R* which has already been

discussed. Functional corroboration of these loci was conducted (Chapter 6) and no differences were seen in gene expression levels between AAA and controls in *ERG* or *IL6R*.

*MMP9* undoubtedly plays a major and central role in the development and progression of AAA<sup>18, 29, 36, 146</sup>, and was considered an exciting prospective candidate gene to study in this PhD. However no methylation changes were observed in AAA at the *MMP9* gene locus in PBMC or VSMC DNA. Although it is very interesting to note that the network analyses conducted in Jones *et al.*,<sup>83</sup> GWAS meta-analysis revealed a central role for *MMP9* in AAA via interaction between *ERG*, *IL6R* and *LDLR* as modifiers of *MMP9*. There was a direct interaction with *ERG*, which similarly to *SORT1* is a known eQTL<sup>136</sup>. Secondary interactions between *MMP9* and both *SMYD2* and *LDLR* were also demonstrated. These results suggest that these genes could act in concert, either synergistically or antagonistically, to affect the AAA phenotype. All of these genes implicated in the possible regulation of, and interaction with *MMP9* were differentially methylated in this PhD project and are therefore valuable targets for further research relating to the interactions between these genes.

There was a significant decrease in the expression of *SERPINB9*, but this was not associated directly with changes in DNA methylation after linear regression analysis was performed. Critically, when the Ryer *et al.*,<sup>143</sup> study identified differential methylation in *SERPINB9* in PBMC DNA, they did not conduct such analysis between the differentially methylated regions in the *SERPINB9* gene body and gene expression, but still concluded that DNA methylation could be related to gene function. The work conducted in my PhD has identified a similar pattern to the Ryer *et al.*, study, however it cannot be stated that methylation and expression in *SERPINB9* are linked. My work goes some way to deny

this insinuation, as no such direct relationship was seen. However, regardless of DNA methylation status, the differential expression of *SERPINB9* in AAA could still be important, as reduced *SERPINB9* activity has been previously associated with atherosclerotic disease progression and is inversely related to the extent of apoptosis within the intima<sup>196</sup>.

Potentially the most important results obtained from this project are that a significant level of DNA hypo-methylation was seen in the *SMYD2* gene in VSMCs (NC\_000001.11:g2891814\_2892002|lom), which was directly correlated with gene transcript expression from the same VSMC samples. In addition, the reduced gene expression of *SMYD2* in AAA was further corroborated in aortic tissues using immunohistochemistry. These results are particularly important when considering the role of *SMYD2* in inflammation and cardio-physiology, as previous studies have identified that the *SMYD2* protein is involved in cardio-protection and suppression of inflammation. Protein levels of *SMYD2* were decreased in cardiomyocytes after cellular apoptosis and after myocardial infarction. In addition, *SMYD2* deletion in cardiomyocytes *in vivo* promoted apoptotic cell death upon myocardial infarction<sup>195</sup>. It has also been reported that *SMYD2* is a negative regulator for macrophage activation. Elevated *SMYD2* expression suppresses the production of pro-inflammatory cytokines, including IL-6 and TNF. In addition, macrophages with high *SMYD2* expression promote regulatory T cell differentiation as a result of increased TGF- $\beta$  production and decreased IL-6 secretion<sup>220</sup>. Overall, previous literature demonstrates that a reduced activity of the *SMYD2* gene, which has been observed in this PhD, is linked with adverse cardio-physiology and an increase in inflammation, both of which are key hallmarks of AAA pathobiology.

Looking more closely at the relationship between *SMYD2* gene activity and gene promoter DNA methylation in this project, it is clear that the association is opposite to the traditional notion that promoter hypermethylation inhibits expression of genes located directly downstream from these CpG islands. For example, previous studies investigating the role of DNA methylation in certain traits/diseases have identified and highlighted that DNA methylation at CpG sites in gene promoters can inhibit the binding of transcriptional silencer proteins and RNA polymerases essential for transcriptional activation<sup>221-223</sup>. Specifically, promoter CpG methylation can be involved in adding stability to the repression of transcription when it is located at the start sites of mammalian genes<sup>224</sup>. Methylated CpG dinucleotides are bound by methylated DNA binding proteins, and the affinity of these proteins for a given promoter influences the extent to which methylated DNA binding proteins inhibit transcription factor binding<sup>225</sup>.

However, it is now known that DNA methylation is a much more dynamic process than traditionally thought. Gene activation and repression are tightly regulated to the correct cell type and correct developmental stage. These regulations are controlled, as previously described, by an array of transcription factors, histone modifying enzymes, chromatin remodelers, and DNA methylation<sup>226</sup>. It is therefore becoming better known that the commonly accepted paradigm previously discussed is not definitively correct, as multiple pieces of evidence suggest that methylation at CpG sites can also induce gene transcription, and the individual functional mechanism of each binding complex is dependent on the identity of certain transcriptional proteins which form the binding complex, whether it be an inducer or repressor, which can in turn recruit or inhibit the action of RNA polymerases respectively.<sup>203, 227</sup> It is evident then, that in some cases, methylation is required for activation of transcription and is therefore positively correlated with gene expression<sup>228</sup>, as opposed to repression. This explanation would be

more appropriate to the findings in this project, considering that *SMYD2* gene expression was decreased in those with AAA who also had lower levels of gene promoter methylation. Due to the factors discussed in this section, it must be stated that the exact mechanisms by which methylation controls gene expression is still not fully understood and is not always the same. This work therefore requires large scale replication and subsequent mechanistic functional analysis to determine the role of *SMYD2* promoter methylation on gene function, as the work conducted in this project did not establish cause and effect.

### 7.4. Limitations and future work

Limitations and future work which are specific to the experiments conducted during this PhD project are acknowledged in the conclusion sections of each results Chapter (Chapter 3, 4, 5 and 6). However there are limitations of the project that also apply more broadly and should be addressed here.

It was stated as a limitation in the conclusion of Chapter 6 that the methylation status of the *SMYD2* gene promoter and its relationship with *SMYD2* gene expression did not address causality, and therefore needs further work before any major conclusions can be drawn from the results. In addition, it is not possible to conclude whether methylation changes are inherited and play a role in development of the disease, or whether these methylation changes occur as a result of the disease process, e.g. they may be induced by inflammation, oxidative stress, or other factors involved in the pathobiology of AAA. Moving forward, this could be studied in a range of ways, considering DNA methylation is very dynamic and can be inhibited, induced or reversed<sup>101, 103, 112, 113</sup>. Specifically, *in vitro* or *in vivo* assays could be set up where *SMYD2* promoter methylation is

manipulated and the effects on gene expression are directly measured. This work could be based around techniques which are already developed.

Pathogenic regions that are hyper or hypo-methylated can be selectively targeted in the same way epigenetic therapies have been found to be effective in the treatment of cancer<sup>100, 103, 176</sup>. It was observed that a methyl supplementation donor (SAM) inhibits cancer associated skeletal metastasis by suppressing gene transcription of its selected target by inducing DNA hypermethylation<sup>176</sup>. Methylation can also be reversed with the use of 5-aza-2'-deoxycytidine, which in one study mitigated atherosclerosis in knockout mice deficient in the *LDLR* gene<sup>75</sup>. 5-aza-2'-deoxycytidine is a de-methylating compound and useful in the application of increasing the expression of down-regulated genes. In relation to the work in this thesis, SAM or 5-aza-2'-deoxycytidine can be used as methylation moderators in cells to establish the causal role of methylation on gene function and disease. The CRISPR/Cas9 genome editing system has been repurposed for targeted de/remethylation which enables direct study of functional relevance of precise epigenetic modifications and gene regulation<sup>229</sup>. By deactivating the nuclease activity of the normal CRISPR/Cas9 delivery system, epigenetic machinery such as DNA methyltransferases, SAM, demethylases and 5-aza-2'-deoxycytidine can be used to induce or reverse CpG site specific epigenetic modifications, enabling the most advanced functional epigenetic studies to be conducted.

The relatively limited amount of genomic coverage that using a targeted candidate gene approach provides compared to the use of other techniques that cover the whole genome (whole genome bisulphite sequencing, human 450k microarray and methylated immunoprecipitation) is another limitation. Although whole genome bisulphite NGS is regarded the best available methodology for methylation analysis, it is also generally

expensive to yield an acceptable sample size, and was beyond the scope of this PhD. This is why a targeted approach was adopted for this project, but future work with a larger research budget should concentrate on identifying whole genome methylation and transcriptome data in a case-control cohort. This will enable the elucidation of the complex interactions of methylation and gene regulation, which is especially important because methylation differences were observed in this project, but were not related to expression of the same gene. It is possible that these methylation differences are still functionally significant, given that differential methylation patterns may have functional impact in other areas of the transcriptome. Reduced representation bisulphite sequencing (RRBS) would be a particularly good way in which to address this issue, where restriction enzymes are used to cut DNA at CG rich areas<sup>230</sup>, allowing the isolation of the ~1% of the genome which is comprised of CpGs. By subsequently performing multiplexed bisulphite sequencing on an Illumina Hi-seq platform, the full CpG methylomes of multiple samples could then be established simultaneously.

The work conducted during this PhD was performed on samples selected from either the UKAGS (PBMCs), or from the local vascular surgery AAA research programme based at the NIHR Leicester Biomedical Research Centre (VSMCs). Considering these resources have not been used historically in previous large scale GWASs, there was no genomic data available for the conduct of meQTL analysis in this project, and it would have been too large of a task to genotype each sample considering the limited timeframe of the project. This is a significant limitation with respect to determining the potential causal role of genetic variation in AAA on levels of DNA methylation in the genes targeted in this PhD. Therefore replication of the positive results demonstrated during this project should be conducted on a mass scale in samples which have full GWAS datasets. Statistical

approaches, such as Mendelian Randomisation can then determine the relationship between polymorphisms associated with AAA and differential methylation.

Another limitation of the work presented in this thesis is the lack of a positive control group for atherosclerotic disease. Atherosclerosis represents an important risk factor for AAA, and studies have shown that co-morbidity of CAD and AAA is between around 25% and 55%<sup>43, 50</sup>. However there is recent epidemiological, clinical, and biological evidence that suggests that the two pathologies are more distinct than traditionally thought. For instance diabetes mellitus, hypercholesterolemia, and obesity are high risk for atherosclerosis development but are not as pronounced in AAA, whereas smoking, sex, and ethnicity are particularly high risk for AAA but less so for atherosclerosis<sup>231</sup>. In addition, genetic and epigenetic studies have identified independent risk loci involved in AAA susceptibility that are not associated with other cardiovascular diseases<sup>83, 102, 143, 231</sup>. A review article was produced during this PhD illustrating these concepts<sup>231</sup>. Further investigation is therefore required to address the extent to which the genetic and epigenetic basis of AAA and atherosclerosis is shared. It would be advantageous to include CAD control groups (which act as positive controls for cardiovascular disease) in future AAA research to decipher exactly whether positive results are unique to AAA.

Finally, linking all of the points raised previously, large scale analysis of the role of DNA methylation in AAA should be conducted to address the limitations in the current work. Significant increases in sample size will increase power and confidence of observed associations and should be conducted preferably using unbiased RRBS on all available samples in the Leicester cohort from previous GWASs. The assay should also include an atherosclerotic control group with available GWAS data, and where any positive associations are made between genetic variation and/or DNA methylation and the

presence of AAA and/or CAD, direct functional and mechanistic analysis should be conducted using direct mRNA gene expression, immunohistochemistry tissue protein expression, and luciferase gene promoter reporter assays. The conduct of such a study will reveal real significant insight in to the epigenetic basis and general aetiology of AAA and CAD.

## Appendix I – Reference sequence for PBMC NGS bioinformatics

### >LRP1

GGTACTAAGGTGGGCTCCATCCC**CG**AGCCCCA**CGCG**GG**CG**GACAAGCTC**CG****CG**TG  
 TCCCCT**CG**GGTGTCCCTGTTTACTC**CG**AGCC**CG**GGAG**CG**AGGTGGGGG**CG**GGTCC**C**  
**GCG**GTCCCTCCCCAACCC**CG**CCCCCTCCTT**CG**CAGGCCCAATCCCAGTC**CG**GTCTTC  
 CAAAGTCTCAAATCAAAGCTCAGCCTTCCCTGCACCTTCCC**CGCG**GACTGG**CG**GTC  
 CCTGTCCCCACCC**CG**GGCAGAGGAGGCACCTTCAGGGTTCCCCTAGAAAAT**CG**AGCC  
 T**CG**GCCCAGACCACTGAGTCC**CG**GA**CG**CCCC**CG**GAGGGAGCCTGAAATCCTAGAGTA  
 TGACACTGAGTTTCAAAGGGGAGC**CG**CTCAGCTCTC**CG**CCCACTGCCAATCCCACC  
 CT**CG**ACCCCTTCTGCTCTCTGGAGCACCAGGGAGGAGGGGCCAGCCAGGGGAGAGG  
 G**CG**CCCCAGGCCCAAAGTACTAGCCAG**CG**AGTCCC**CG**GGACCCTCACTCTTGGAC**CG**CT  
 TTCCAGAACTC**CG**AACCCTGT**CGCG**GGTC**CGCG**T**CG**CCCTCCCC**CG**GGCAG**CGCG**TC  
 AAAT**CG****CG**CATG**CG**CACTCACTGGATTGC**CGCG**TAGCTCTTCTCCCCCACCCACC  
 AACCTTTTTTCTTTCCC**CG**CCCCTTCCCTCCCTCCCTCCTCAACC**CG**TCCCCTCCCTCTC  
 CCCATCAGCCCCCCCCT**CG**GCACTTCAGTC**CG**GGGAACAG**CG**GTG**CG**AGCTCCAGG  
 CCCATGCACTGAGGAGG

### >ERG

GTCATCACAGAGAGCCTCTTTCCCCTGGAACCTGTGCTCACTGGAGTTTTCTTCCTC  
 TCCTGACCTCATTAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCCTGGAGCTCCT  
 GT**CG**GACAGCTCCAAGTCCAGCTGCATCACCTGGGAAGGCACCAA**CG**GGGAGTTCAA  
 GATGA**CG**GATCC**CG**A**CG**AGGTGGCC**CG****CG**CTGGGGAGAG**CG**GAAGAGCAAACCCA  
 ACATGAACTA**CG**AATAAGCTCAGC**CGCG**CCCTC**CG**TACTACTATGACAAGAATCA  
 TGACCAAGGTCCATGGGAAG**CG**CTA**CG**CCTACAAGTT**CG**ACTTCCA**CG**GGAT**CG**CCC  
 AGGCCCTCCAGCCCCACCCCC**CG**GAGTCATCTCTGTACAAGTACCCCTCAGACCTCC**C**  
**G**TACATGGGCTCCTATCA**CG**CCCACCCACAGAAGATGAACTTTGTGG**CG**CCCCACCC  
 TCCAGCCCTCCC**CG**TGACATCTTCCAGTTTTTTTGTGCTGCCCAAACCCATACTGGAAT  
 TCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCACCAGCCATATGCCTTCTC  
 ATCTGGGCACTTACTACTAAAGACCTGG**CG**GAGGCTTTTCCCATCAG**CG**TGCATTAC  
 CAGCCCAT**CG**CCACAAACTCTAT**CG**GAGAACATGAATCAAAAGTGCCTCAAGAGGA  
 ATGAAAAAAGCTTTACTGGGGCTGGGGAAGGAAGC**CG**GGGAAGAGATCCAAAGACT  
 CTTGGGAGGGAGTTACTGAAGTCTTACTACAGAAATGAGGAGGATGCTAAAAATGTC  
 A**CG**AATATGGACATATCATCTGTGGACTGACCTTGTAAGACAGTGTATGTAGAAG  
 CATGAAGTCTTAAGGACAAAGTGCCAAAGAAAGTGG

## &gt;MMP-9

GGCCTCCCAAAGTGCTAAGATTACAGGAATGAGCCACCATACTGGCCCTGAATCTT  
GGGTCTTGGCCTTAGTAATTA AAAACCAATCACCACCATC **CGTTGCG**GACTTACAACCT  
ACAGTGTTCTAAACATTTTATATGTTTGATCTCATTTAATCCTCACATCAATTTAGGG  
ACAAAGAGCCCCCACC **CG**TTTTTTTTTTTACAGCTGAGGAAACTTCAAAGTG  
GTAAGACATTTGCC **CG**AGGTCCTGAAGGAAGAGAGTAAAGCCATGTCTGCTGTTTC  
TAGAGGCTGCTACTGTCCCCTTTACTGCCCTGAAGATTCAGCCTG **CG**GAAGACAGGG  
GGTTGCCCCAGTGGAATTCCCAGCCTTGCCTAGCAGAGCCATTCCTTC **CG**CCCCCA  
GATGAAGCAGGGAGAGGAAGCTGAGTCAAAGAAGGCTGTCAGGGAGGGAAAAAGA  
GGACAGAGCCTGGAGTGTGGGGAGGGGTTTGGGGAGGATATCTGACCTGGGAGGGG  
GTGTTGCAAAGGCCAAGGATGGGCCAGGGGGATCATTAGTTTCAGAAAGAAGTCT  
CAGGGAGTCTTCCATCACTTTCCCTTGGCTGACCACTGGAGGCTTTCAGACCAAGGG  
ATGGGGGATCCCTCCAGCTTCATCCCCCTCCCTCCCTTTCATACAGTTCCCACAAGCT  
CTGCAGTTTGCAAACCCTACCCCTCCCTGAGGGCCTG **CG**GTTTCCTG **CG**GGTCTGG  
GGTCTTGCTGACTTGGCAGTGGAGACTG **CG**GGCAGTGGAGAGAGGAGGAGGTGGT  
GTAAGCCCTTTCTCATGCTGGTGCTGCCACACACACACACACACACACACACACA  
CACACACACACACACACCCTGACCCCTGAGTCAGCACTTGCCTGTCAAGGAGGGGGTG  
GGTTCACAGGAG **CG**CCTCCTTAAAGCCCCACAACAGCAGCTGCAGTCAGACACCTC  
TGCCCTCACCATGAGCCTCTGGCAGCCCTGGTCCTGGTGCTCCTGGTGC

## &gt;CDKN2B

AAGTATAATTTTTTTTTGTCTTATGTGTGCCAGGTTGCCACTCTCAATCT **CG**AACTAGT  
TTTTTCTCTTTTAAGGGTTGTATCCATAATGCAAAAATGGAAAGAATTA AAAAGCAC  
A **CG**CAAAACATGATTCT **CG**GGATTTTTCTCTATTTTTATGGTTGACTAATTCAAACAG  
AAAGACACATCCAAGAGAAAATTGCTAAGTTTGATAACAAGTTATGAACTTGTGAAG  
CCAAGTACTGCCTGGGGATGAATTTAACTTGTATGACAGGTGCAGAGCTGT **CG**CTT  
TCAGACATCTTAAGAAACA **CG**GAGTTATTTTGAATGACTTTCTCT **CG**GTCACAAGGG  
AGCCACCAA **CG**TCTCCACAGTGAAACCAACTGGCTGGCTGAAGGAACAGAAATCCTC  
TGCTC **CG**CCTACTGGGGATTAGGAGCTGAGGGCAGTGGTGAACATTCCCAAAATATT  
AGCCTTGGCTTTACTGGACATCCAG **CG**AGCAGTGCAGCCAGCATTCTGG **CG**GCTCC  
CTGGCCAGTCTCTGG **CG**CATG **CG**TCCTAGCATCTTTGGGCAGGCTTCCC **CG**CCCT **CG**  
TGA **CGCGT** **CG**GCC **CG**GGCCTGGCCTCC **CG** **CG**ATCACAG **CG**GACAGGGGG **CG**GAGC  
CTAAGGGGGTGGGGAGA **CG** **CG**GCCCCTTGGCCAGCTGAAAA **CG**GAATTCTTTGC  
**G**GCTGGCTCCCCACTCTGCCAGAG **CG**AGG **CG**GGGCAGTGAAGACTC **CGCGA** **CGCGT**C  
**CG**CACCCTG **CG**GCCAGAG **CG**GCTTTGAGCT **CG**GCTG **CG**TC **CGCG**CTAGG **CG**CTTTTTC

CCAGAAGCAATCCAGGCGCGCCCGCTGGTTCTTGAGCGCCAGGAAAAGCCCGGAGC  
 TAAACGACCGGCCTCGGCCACTGCACGGGGCCCAAGCCGCAGAAGGACGACGGG  
 AGGGTAATGAAGCTGAGCCCAGGTCTCCTAGGAAGGAGAGAGTG

## &gt;LDLR

CTGTTTGTTCAACTCCTTCTCCTAAGGGGAGAAATCAATATTTACGTCCAGACTCCAG  
 GTATCCGTACAATTGATTTTTTTCAGATGTTTATACTCAGCCAAAGCGGGATCCCACAA  
 AACAAAAATATTTTTTTGGCTGTACTTTTGTGAAGATTTTATTTAAATTCCTGATTGA  
 TCAGTGTCTATTAGGTGATTTGGAATAACAATGTAAAAACAATATACAAAGAAAGGA  
 AGCTAAAAATCTATACACAATTCCTAGAAAGGAAAAGGCAAATATAGAAAGTGGCG  
 GAAGTTCCCAACATTTTTAGTGTTTTCTTTTTGAGGCAGAGAGGACAATGGCATTAGG  
 CTATTGGAGGATCTTGAAAGGCTGTTGTTATCCTTCTGTGGACAACAACAGCAAAAT  
 GTTAACAGTTAAACATCGAGAAATTCAGGAGGATCTTTCAGAAGATGCGTTTCCAA  
 TTTTGAGGGGGCGTCAGCTCTTCACCGGAGACCCAAATACAACAAATCAAGTCCT  
 GCCCTGGACACTTTCAAGGACTGGAGTGGGAATCAGAGCTTCACGGGTTAAAAA  
 GCCGATGTACATCGGCCTCGAAACTCCTCCTTGCAGTGAGGTGAAGACATTT  
 GAAAATCACCCACTGCAAACTCCTCCCCCTGCTAGAAACCTCACATTGAAATGCTG  
 TAAATGACGTGGGCCCCAGTGCAATCGCGGGAAGCCAGGGTTTCCAGCTAGGACAC  
 AGCAGGTCTGTATCGGGTGGGACACTGCCTGGCAGAGGCTGCGAGCATGGGGCC  
 CTGGGGCTGGAAATTGCTGGACCGTCCCTTGCTCCTCGCGCGGGGGACTGC  
 AGGTAAGGCTTGCTCCAGGCGCCAGAATAGGTTGAGAGGGAGCCCCCGGGGGCCC  
 TTGGGAATTTATTTTTTTGGGTACAAATAATCACTCCATCCCTGGGAGACTTGTGGGG  
 TAATGGCA

## &gt;GDF7

GGGTTCCCTGCCCTGTACTGTGTTCCCATGGCAACCTGAGCTTACTACCCTGCTGGC  
 ATCACCCCTCTCCGTGCAGCCCCCTGCAGGGCGAGCTTGTCTGATTCCCTCTGTC  
 CCCAGTGCCAGCACCGGGCCTGGGACTCAGGGAGCTCTGAAAGGATGGGGAACTA  
 GATTTTGGGCTCAAAGAGAAGAACCCAAGCGGTAGGAAAGGAGAGCAGAGTCC  
 CCACTGCGCCCTGCACTGAACCAGTCAGTCCCAGGGCTCCCGCTGTCCTTGGCTA  
 GTTCTTACCCTCCAGAGGGCCTCGGTTTCTCATCTGTAAATCGGGGTCCACCTC  
 TCCTGCAGGGCGTGCAGGGGCCAGCCCGGATAAGCCACCGAGGGCGCTGGGGAG  
 ACCTGCACCAGGTCCCTGGCCTAGGAGGCGGCGCCTCCCCCGCCTCTCCCGCGG  
 GCGCCCCTGACTGCGAGCTGTGCGTGTGGGGCGCGGGGGTCTGGCTCACCTGGAGG  
 TGAAGCACCCTGGGACTGCATGACGGCGGGGACTCCGGCGGGGGCGCGC  
 AGTCAAGGACCCTGCAGTGCCCGCGTCACCGCGTCCCCTCCCTGCCCTCCCGCTGCC

CTGGCCTGCCCCCTGGTGGCAGCGGCCGCGCAGTTCCGCGCCTGTGGCCAGAGGCCGGG  
 GCGGAAGCGCGGCTGGGCGGGTCCCGCACCGGTCCCGTCCGCGCCTGCGCTGCTCGGAG  
 CTTCGCGAGCACCCAGGCTGAGCCGCGCCCGCCGCAACCACCTAGAAGAAGCCCAG  
 CCTGGCCCTGGCACA

## &gt;SPAG17

TATTTTAGATATTGATGAGTGCAGGGTGGGAAAAAAGGGGAAAAAAAAGTCTACTT  
 ACCAATCAATAAAAGGCAGTTTAAAGGACTGGAGCTGTTTGGGCAGATAAGTACATT  
 AATCCACCTACAGTTAGATAAGGGGTCACGTGGAAGGGTGATTGGACTTGTTTAGT  
 AAGGCCCCAGAGGCCGAGGCTGAAACAGTGCAGAGGGTAGAGGTCAAGGAGAGAC  
 AAATTTGGATTACATATAAGGAAGCATTGGGGGCAGATTAGAGCGGTTGGGAAAAAA  
 CGGCATGGGCTGCCTCCTGAAGGGGTCAGTATTCCAATCTTGTTACTCAGGATCGG  
 AAGAAATCTTGATGGAGGAGTCTGGATGGGATTCCAGAGATTTGAGTTTCTAGAC  
 AACCGCCTCAGACCGTAGAAGAACGTAGCCCGACCTGCGAGGCCAGCCTCGTCCTGCC  
 CTCCTCCAGTTGTGCAGCTTTCCTGCTGAAAAGTCCTAAAATTTGACCGGGGGACTGG  
 GAACCCCGCCCCAAGCTGCGCGCGCTCTGAAACTCCGTTGCTGTGGAGA CGGAGG  
 CGGTTGACCCAAGGATTCGAAACGGTGTGGGGGAGGGGAGCCGACCGTGGTTCAGG  
 AACTGGGCGTGGCTTCGGTTACTGTGGTAGCGGGGACGCTTAGGCAGGGCCTGCGCC  
 CAGTTTAGAGGCCAATGCTTCTCCCGTCCTTGCATGGCACCCAAGAAGGAGAAAGG  
 AGGAACTGTGAACA

## &gt;IL6R

TTCTTCTGCCAGGCTGGAGTGCAGTGGTGCCATCTCGGTTCACTGCAACCTCCGCC  
 TCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACCGGTAC  
 GCACCACCATGCCCGGCTAATTTTTGTATTTTTAGTAGAGATGTGGTTTTACCATGTT  
 GGTCAGGCTGGTCTCGAACTCCTGACCTCATGATCTGCCACCTTGGCCTCCCAAAGT  
 GCTGGAATTACAGGTGTGAGGCACCACGTCCGCGCCTGGCCAGGATTAATTTTTTCA  
 TGTGAACCACAACACTAAAACCTTTGTATATTGCTGCGCTTATCTTACAAATCAGGAA  
 ATAGGGGTTACAGAGAAGTTAAATAGCTTGCTTAAGTCACACAGCCAGCAAGTGGTAA  
 AGCTGGGATCTGCATGACTACAGAGCACACAGGCTCCCCACCAGTAGATTAGAGAG  
 GGGGAGGTCTGCTTTGGTGCAGAGACAGGTGGGATCTGTGATGCCCGTTCTTGGTTT  
 CAACATTCTTTAAACTCTCCTATCTGAGCCTAGGACTTCTGCAGCCTTCTGTTTAA  
 GGCTGGCAGCTCACAGTCCCTCTCTGGTTATTTTCAGGTCTGTGTGTGCCACAGAGAG  
 GAAGGGGGCAACCACAGTGGGAACCGCTTCCAGCCCTGCTGCAACCCCTTTTGAAA  
 TAGCAGGCGAGAGGGCTGGTGGCCTCCACTGTGCTTCTGTTCTGTCTGTGGGCAT  
 GGCTAAAGCAAACAAGCTCACCCACACCAGCTCCCATCGTGGCGGTGGATCAGTGCAT

GTGCTTGGTTCTGTCTGGCACTGCTGATGGGTGGTTCCTACAGAAAGATCCAGATGC  
A **CG**GGGTCTGAGGACTGTTTCAGAAGCTCTGTC **CG**TCCATGCCTGGCCCTGTCCACTT  
GGGAGTTCCATTAGGCCATGCCTGTTCCA

>**SORT1**

GTATGGAAAGACTATTCTTCAGGCTTGCAGGAGTCA **CG**GATGTCCTGATGTCTGTCTC  
**CG**TAGGATGTGGCCACATCAGTT **CG**CATTACCTTCTTGCAGCAAGGCAGCTGCTTA  
GCAGACAATGGG **CG**CTCCAGTGAG **CG**GGGTTTATAAAAACC **CG**AAGCCC **CG**GTTTCAT  
GATGGAGCCCCCTTTCCAGCTGAGCAAGCTCAGGGATTTCTGGGTAGGTTTTTCCA  
GGCTCTGCCTCAC **CG**AAGGAATTTTAGGAGTGTCTCTGGGGAACAGGAGGGA **CG**TAA  
CCCAGCCCCAACTTGAGGG **CG**CTAGAGGTG **CG**GCAAGGGGT **CGCGA** **CG**CCAGGAGC  
C **CG**GGGCT **CGG** **CG**GAAGGTATGAGAAGCTCCTA **CG**TGAACTCCACAAGC **CG**GGCC  
**CG**GGAGA **CGC** **CG**GG **CG**AGG **CG**GGGTTGACCTCAGCAGTCTCTGCCC **CG**TTCCAGCCA  
ATCAGTCC **CG**CATCTTAGCATC **CG**AATCCAGGACCCC **CG**AAGC **CG**GAGG **CGAC** **CGCGA**  
GCCAATGAGGAGTGGGC **CG**GGGAAGAGGGACAGG **CG**GCCAGC

## Appendix II – Reference sequence for VSMC NGS bioinformatics

## &gt;LRP1

GCCCAGAAGGGGGCAGTGACCAAAAGCA **CG**TTCCTGGCCCCTGGGAAC **CG**CCTGG  
**CG**CCTGCCTTCTGCAAAGTATCATTCC **CG**TGTGGGCTGAGCCTGGGGACAAAGGTC **C**  
**GGCG**CTCAGCAGACACCTCCTTGAACCC **CG**AGGCCTGGAAA **CG**AAGGTCCAGAGTC  
 CCTCATCTTATTCCAAATCCTGG **CG**CT **CG**GAATCTAAAACCTAACTTGATGGCAAGC  
 CAGTCAGAAGGTATTAATAAAAAAAAAAAAAAAAAAAAAAAAAACAGAAAAATAAATGAGC  
 CC **CG**ACTTCTTGGG **CG**AAAGGGGGTGGCCTTTCCTCAACCCACC **CG**CTGGTGACTC  
 ACCTCCCTCAAACCAGGGCCTGCCCTCC **CGCG**CCTAGGGCTGGG **CG**AGCAGGG **CG**  
 GG **CG**GGTACTAAGGTGGGCTCCATCCC **CG**AGCCCCA **CGCG**GG **CG**GACAAGCTC **CG**G  
**CG**TGTCCCCT **CG**GGTGTCCCTGTTTACTC **CG**AGCC **CG**GGAG **CG**AGGTGGGGG **CG**GGT  
 CCT **CGCG**GTCCCTCCCCAACCC **CG**CCCCCTCCTT **CG**CAGGCCCAATCCAGTCC **CG**GT  
 CTCCAAAGTCTCAAATCAAAGCTCAGCCTTTCCTGCACCTTCCC **CGCG**GACTGGC  
**CG**TCCCTGTCCCCACCC **CG**GGCAGAGGAGGCACCTTCAGGGTTCCTTAGAAAAT **CG**  
 AGCCT **CG**GCCCAGACCACTGAGTCC **CG**GAC **CG**CCCC **CG**GAGGGAGCCTGAAATCCTAG  
 AGTATGACACTGAGTTTCAAAGGGGAGC **CG**CTCAGCTCTC **CG**CCCACTGCCCAATCC  
 CACCCT **CG**ACCCCTTCTGCTCTCTGGAGCACCAGGGAGGAGGGCCCAGCCAGGGGAG  
 AGGG **CG**CCCCAGGCCCCAACTAGCCAG **CG**AGTCCC **CG**GGACCCTCACTCTTGACC  
**CG**CTTTCCAGAACTC **CG**AACCCTGT **CGCG**GGTC **CGCG**T **CG**CCCTCCCC **CG**GGCAG **CGC**  
**CG**TCAAAT **CG** **CG**CATG **CG**CACTCACTGGATTGC **CGCG**TAGCTCTTTCTCCCCCACCC  
 ACCAACCTTTTTTTCTTTCCC **CG**CCCCTTCCCTCCCTCCCTCCTCAACC **CG**TCCCCTCCC  
 TCTCCCCATCAGCCCCCCCCT **CG**GCACTTCAGTC **CG**GGGAACAG **CG**GTG **CG**AGTCC  
 CAGGCCCATGCACTGAGGAGG

## &gt;ERG1

GTCATCACAGAGAGCCTCTTTCCCCTGGAACCTGTGCTCACTGGAGTTTTCTTCCTC  
 TCCTGACCTCATTAGGCAGTGGCCAGATCCAGCTTTGGCAGTTCCTCCTGGAGTCTCT  
 GT **CG**GACAGCTCCA ACTCCAGCTGCATCACCTGGGAAGGCACCAA **CG**GGGAGTTCAA  
 GATGA **CG**GATCC **CG**A **CG**AGGTGGCC **CG** **CG**CTGGGGAGAG **CG**GAAGAGCAAACCCA  
 ACATGAACTA **CG**ATAAGCTCAGC **CGCG**CCCTC **CG**TACTACTATGACAAGAACATCA  
 TGACCAAGGTCCATGGGAAG **CG**CTA **CG**CCTACAAGTT **CG**ACTTCCA **CG**GGAT **CG**CCC  
 AGGCCCTCCAGCCCCACCCCC **CG**GAGTCATCTCTGTACAAGTACCCCTCAGACCTCC  
**CG**TACATGGGCTCCTATCA **CG**CCCACCCACAGAAGATGAACTTTGTGG **CG**CCCCACCC  
 TCCAGCCCTCCC **CG**TGACATCTTCCAGTTTTTTTTGCTGCCCAAACCCATACTGGAAT

TCACCAACTGGGGGTATATACCCCAACACTAGGCTCCCCACCAGCCATATGCCTTCTC  
 ATCTGGGCACTTACTACTAAAGACCTGGCGGAGGCTTTTCCCATCAGCGTGCATTAC  
 CAGCCCATCGCCACAACTCTATCGGAGAACATGAATCAAAGTGCCTCAAGAGGA  
 ATGAAAAAAGCTTTACTGGGGCTGGGGAAGGAAGCGGGGAAGAGATCCAAAGACT  
 CTTGGGAGGGAGTTACTGAAGTCTTACTACAGAAATGAGGAGGATGCTAAAAATGTC  
 ACGAATATGGACATATCATCTGTGGACTGACCTTGTAAGACAGTGTATGTAGAAG  
 CATGAAGTCTTAAGGACAAAGTGCCAAAGAAAGTGG

## &gt;ERG2

AAAACCTTCTGGAAGGGGCTTAGCCCAGAGCCCAGGAGCGGACAAGCTGAGCCTCCT  
 GGCTGCACCCTTTGGAGGCTGCTGGTTGCAGACGCGTCTCTCCAAGGGCGGCTG  
 TCACCGCGTGCCTCGGTCAGACCAGGCAGCGGTGCCCTCGGCATCCCAGCGTCCCGG  
 CGGCTGTATCCGGCTCCTCCCCACGACCGAGGTCCCGGGCACCGCGCCTCTCCCTCCG  
 GACCGCTCTGCGCGACGTCGTCGCCAGTTCTCATCAGCATCGAGGGCAGTCAGCGGTCA  
 TTTTATAAAAGTCTTAGTGTCAACCGTTTTCACTTTTACTATTTTCAATGTTTGCAA  
 CGGTTTTCAATGAGGCTGGAAAAATATCACTTCT

## &gt;MMP-9

ATAATCATGGCTCACTGTATCCTTGACCTTCTTTCTGGGCTCAAGCAATCCTCCCACC  
 TCGGCCTCCCAAAGTGCTAAGATTACAGGAATGAGCCACCATACTGGCCCTGAATC  
 TTGGGTCTTGGCCTTAGTAATTAACCAATCACCACCATCCGTTGCGGACTTACAAC  
 CTACAGTGTCTAAACATTTTATATGTTTGATCTCATTTAATCCTCACATCAATTTAGG  
 GACAAAGAGCCCCCACCCTCGTTTTTTTTTTTACAGCTGAGGAAACACTTCAAAGT  
 GGTAAGACATTTGCCAGAGTCTGAAGGAAGAGAGTAAAGCCATGTCTGCTGTTTT  
 CTAGAGGCTGCTACTGTCCCCTTTACTGCCCTGAAGATTCAGCCTGCGGAAGACAGG  
 GGGTTGCCCCAGTGGAATTCGCCAGCCTTGCCCTAGCAGAGCCATTCCCTCCGCCCC  
 AGATGAAGCAGGGAGAGGAAGCTGAGTCAAAGAAGGCTGTCAGGGAGGGAAAAAG  
 AGGACAGAGCCTGGAGTGTGGGGAGGGGTTTGGGGAGGATATCTGACCTGGGAGGG  
 GGTGTTGCAAAAGGCCAAGGATGGGCCAGGGGGATCATTAGTTTCAGAAAGAAGTC  
 TCAGGGAGTCTTCCATCACTTTCCCTTGGCTGACCACTGGAGGCTTTCAGACCAAGGG  
 ATGGGGGATCCCTCCAGCTTCATCCCCTCCCTCCCTTTCATACAGTTCCCACAAGCT  
 CTGCAGTTTGCAAAACCCTACCCCTCCCCTGAGGGCCTGCGGTTTCCTGCGGGTCTGG  
 GGTCTTGCCCTGACTTGGCAGTGGAGACTGCGGGCAGTGGAGAGAGGAGGAGGTGGT  
 GTAAGCCCTTTCATGCTGGTGCTGCCACACACACACACACACACACACACACACA  
 CACACACACACACACACCCTGACCCCTGAGTCAGCACTTGCCCTGTCAAGGAGGGGTG  
 GGGTCACAGGAGCGCCTCCTTAAAGCCCCACAACAGCAGCTGCAGTCAGACACCTC

TGCCCTCACCATGAGCCTCTGGCAGCCCCTGGTCCTGGTGCTCCTGGTGCTGGGCTGC  
 TGCTTTGCTGCCCCCAGACAGCGCCAGTCCACCCTTGTGCTCTTCCTGGAGACCTGA  
 GAACCAATCTCACCGACAGGCAGCTGGCAGAGGTGGGCAAACACCTAGTCTAGAGT  
 TGGGGAGGGCTGTCCGTGAGGGTGTGAGTGTCCAGAGAGGATGCAGGGCCTCAG  
 AGGAGATGCTTTAGGGGTGTGTTGGTGGTGTATGGGCGTATCTGAAGAACAGAGGTGT  
 CCAGGGTTAGGCAGTGGGGGGTCTTGTGGAGGCTTTGAGCAGTGATGGCCAGAAATG  
 GGCAATGGGGCTTTCCTAGGTGGGAAATGGGAAATGGTTTGGGGTGGGGGAGGCAT  
 TGGAGGGTTCTGGGGTAAGCATAGGCTGGGAGTGAACAGGGGCAAACCTTATGCAG  
 CTGTGGGGTAGAAATGGGCTAGAGGCATCCAGGGGTGAGAAGGAGCTGAGGATGTC  
 TAAGGAGGGGAGATCCCTGGGTGGTCAGAAAGCACTGGTGTCTGGAAAGCATTTAAT  
 GCTTTATTAATGTTAGTCCCTGC

**>LDLR**

CTGTTTGTTCAACTCCTTCTCCTAAGGGGAGAAATCAATATTTACGTCCAGACTCCAG  
 GTATCCGTACAATTGATTTTTTTCAGATGTTTATACTCAGCCAAAGGCGGGATCCCACAA  
 AACAAAAAATATTTTTTTGGCTGTACTTTTGTGAAGATTTTATTTAAATTCCTGATTGA  
 TCAGTGTCTATTAGGTGATTTGGAATAACAATGTAAAAACAATATACAACGAAAGGA  
 AGCTAAAAATCTATACACAATTCTAGAAAGGAAAAGGCAAATATAGAAAGTGGCG  
 GAAGTTCCCAACATTTTTTAGTGTTTTTCTTTTGAGGCAGAGAGGACAATGGCATTAGG  
 CTATTGGAGGATCTTGAAAGGCTGTTGTTATCCTTCTGTGGACAACAACAGCAAAT  
 GTTAACAGTTAAACATCGAGAAATTTTCAGGAGGATCTTTCAGAAGATGCGTTTCCAA  
 TTTTGAGGGGGCGTCAGCTCTTCACCGGAGACCCAAATACAACAAATCAAGTCCCT  
 GCCCTGGCGACACTTTCGAAGGACTGGAGTGGGAATCAGAGCTTCACGGTTAAAAA  
 GCATGTCACATCGGCCGTTCGAAACTCCTCCTCTTGCAGTGAGGTGAAGACATTT  
 GAAAATCACCCCACTGCAAACCTCCTCCCCCTGCTAGAAACCTCACATTGAAATGCTG  
 TAAATGACGTGGGCCCCAGTGCAATCGCGGGAAGCCAGGGTTTCCAGCTAGGACAC  
 AGCAGGTCTGATCCGGTGGGACACTGCCTGGCAGAGGCTGCGAGCATGGGGCC  
 CTGGGGCTGGAAATTGCGCTGGACCGTCCCTTGCTCCTCGCGCGGGGGACTGC  
 AGGTAAGGCTTGCTCCAGGCGCCAGAATAGGTTGAGAGGGAGCCCCCGGGGGGCC  
 TTGGGAATTTATTTTTTTGGGTACAAATAATCACTCCATCCCTGGGAGACTTGTGGGG  
 TAATGGCACGGGGTCTTCCCAAACGGCTGGAGGGGGCGCTGGAGGGGGCGCTGA  
 GGGGAGCGCGAGGGTGGGAGGAGTCTGAGGGATTAAAGGGAAAAGGGCACCGCT  
 GTCCCCCAAGTCTCCACAGGGTGGAGGACCGCATCTTCTTTGAGACGAGTCTAGCT  
 CTGTCCCAGGATGGAGTGCAGTGGCACATCTCAGCTCACTGCAACCTCCCGCTC  
 CCGGGTTTAAGCGAGTCTCCTCTCTCAGCCTCCCGAATAGCTGGGATTACAGGCC  
 CAACCACCACGCCCGCCTAATTTTTGTATTTTTAGTAGAGAAGGGTTTTACCATTTT

GGCCAGGCTGGTCTCGAACCCCGACCTCAGGTGATCTGCCAAAAGTGCTGGGATTA  
 CAGGCGTCAGCCACCGCGCCCGGC CGGGACCCTCTCTTCTAACTCGGAGCTGGGTGT  
 GGGGACCTCCAGTCTAAACAAGGGATCACTCCACCCC CGCCTTAAGTCCTTCTG  
 GGGGCGAGGGCGACTGGAGACC CGGATGTCCAGCCTGGAGGTCACCGCGGGCTCAG  
 GGGTCCCGATCCGCTTTGCGCGACCCAGGGCGCCACTGCCATCCTGAGTTGGGTGC  
 AGTCCCGGGATTC CGCGCGTGCTC CGGGACCGGGGGCCACCCCTCCCGCCCCTGCC  
 CC CGCCCCTTTGGCCCGCCCC CGAATTCCATTGGGTGTAGTCCAACAGGCCACCCTC  
 GAGCCACTCCCCTTGTCCAATGTGAGG CGGTGGAGG CGGAGG CGGG CGT CGGGAGG  
 ACGGGGCTTGTGTA CGAG CGGGG CGGGGCTGG CGCG GAAGTCTGAGCCTCACCTTGT  
 CCGGGGCGAGG CGGATGCAGGGGAGGCCTGG CGTTCCTC CGCG GTTCTGTACAAA  
 GGCGACGACAAGTCC CGGGTCCC CGGAGC CGCCTC CGCGACATACA CGAGT CGCCCT  
 CCGTTATCCTGGGCCCTCCTGG CGAAGTCCC CGGTTTC CGCTGTGCTCTGTGG CGACA  
 CCTC CGTCCCCACCTTGTCTGGGGGG CGCCCTCGCCCCACCAGCCC CGATCAAGTTC  
 ACAGAGGGGCCCC CGGCCACCCTCAAGGCCT CGGTTCTTA CGAGGTTGAAA CGTTG  
 CCTCAGAATCTCCC CGCCCCTCCTTGGTCTGCAGC CGAGATCTTCAGCCA CGGTGGGG  
 CAGCTATCCCC CGGGAC CGACCCCTGGGGTGGCCT CGCTTCTTCAGAGGCTGTGAA  
 TGGCTT CGGTTCACTGTCCAAG CGGCG ATTTTCTCTGGGTGAAATGGATTAGATT  
 TTAGATTTCCACAAGAGGCTGGTTAGTGCATGATCCTGAGTTAGAGCTTTTTAGGTGG  
 CTTTAAATTAGTTGCAGAGAGACAGCCT CGCCCTAGACAACAGCTACATGGCCCTTT  
 CCCTCCTGAG

**>IL6R**

TTCTTCTTGCCAGGCTGGAGTGCAGTGGTGCCATCTCGGTTCACTGCAACCTCCGCC  
 TCCTGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGACTACCGGTAC  
 GCACCACCATGCC CGGCTAATTTTGTATTTTATAGTAGAGATGTGGTTTTACCATGTT  
 GGTCAAGGCTGGTCTCGAACTCCTGACCTCATGATCTGCCACCTTGGCCTCCCAAAGT  
 GCTGGAATTACAGGTGTGAGGCACCA CGTCC CGGCCTGGCCAGGATTAATTTTTTCA  
 TGTGAACCACAACACTAAAACCTTTGTATATTGCTG CGCTTATCTTACAAATCAGGAA  
 ATAGGGGTTCAAGAGAAGTTAAATAGCTTGCTTAAGTCACACAGCCAGCAAGTGGTAA  
 AGCTGGGATCTGCATGACTACAGAGCACACAGGCTCCCCACCAGTAGATTAGAGAG  
 GGGGAGGTCTGCTTTGGTGCAGAGACAGGTGGGATCTGTGATGCC CGTTCCTTGGTTT  
 CAACATTCTTTAAACTCTCCTATCTGAGCCTAGGACTTTCTGCAGCCTTCTGTTTTA  
 GGCTGGCAGCTCACAGTCCCTCTCTGGTTATTTTCAGGTCTGTGTGTGCCACAGAGAG  
 GAAGGGGGCAACCACAGTGGGAAC CGCTTTCCAGCCCTGCTGCAACCCCTTTTGAAA  
 TAGCAGG CGAGAGGGCTGGTGGCCTCCACTGTGCTCTTCTGTTCTGTGTGGGCAT  
 GGCTAAAGCAAACAAGCTCACCCACACCAGCTCCCAT CGTG CGGTGGATCAGTGCAT

GTGCTTGGTTCTGTCTGGCACTGCTGATGGGTGGTTCCTACAGAAAGATCCAGATGC  
 ACGGGGTCTGAGGACTGTTTCAGAAGCTCTGTC CGTCCATGCCTGGCCCTGTCCACTT  
 GGGAGTTCCATTAGGCCATGCCTGTTCCA

## &gt;SORT1

GTATGGAAAGACTATTCTTCAGGCTTGCAGGAGTCA CGGATGTCCTGATGTCTGTCTC  
 CGTAGGATGTGGCCACATCAGTT CGCATTACCTTCTTGCAGCAAGGCAGCTGCTTA  
 GCAGACAATGGG CGCTCCAGTGAG CGGGGTTTATAAAAACC CGAAGCCC CGGTTCAT  
 GATGGAGCCCCCTTTTCCAGCTGAGCAAGCTCAGGGATTTCTGGGTAGGTTTTTCCA  
 GGCTCTGCCTCAC CGAAGGAATTTTAGGAGTGTCTCTGGGGAACAGGAGGGA CGTAA  
 CCCAGCCCCAACTTGAGGG CGCTAGAGGTG CGGCAAGGGGT CGCGA CGCCAGGAGC  
 CCGGGGCT CGG CGGGAAGGTATGAGAAGCTCCTA CGTGAACTCCACAAGC CGGGCC  
 CGGGAGA CGC CGGG CGAGG CGGGTTGACCTCAGCAGTCTCTGCCCG TTCCAGCCA  
 ATCAGTCC CGCATCTTAGCATC CGAATCCAGGACCCC CGAAGC CGGAGG CGAC CGCGA  
 GCCAATGAGGAGTGGGC CGGGGAAGAGGGACAGG CGGCCAGC

## &gt;SMYD2

GAGGTCCTATTTTGAAGTAGTGGTTTCTGCTAAATGATCT CGT CACAATTTTAGTCTT  
 TACTGCATGTACATCAATTCCAGTCTTTACATAATCATGTGGAACA CGGATTATATTT  
 TCCTTTCTCATATGAAAAGTAATAG CGTGCATGTGTATAGGATTTA CGTTTTCAGGCTT  
 TGTGGGCTTTAAATCCTTTGACTCCTAACCCAAGTGACTTTTCTAAGATCACAGTT CG  
 TGGGTGGTCAAGCATATTA AAACTACTGCAA ACTCTAG CGCAGTGTCTGGACTG  
 GCCACAGCAT CGCATTATCCCCAACAGCCTAAATTCCTTGAAGGCAAGGTTCTCTCT  
 GGCTTCTCTGCAA CGTT CGCAGCATGCCTCAGAGACT CGGTGAGCTGGGC CGGCCA  
 ATGCATTTCTCTTC CGTCCATGTAATTCCTAGAGGAGCAGAGGGCAA ACTA CGTT  
 CCCATTAAAGCCACAAGGTTTAAAACCTCTAACCTTGGAAAAGCACACTTCAACCC  
 TCTGCACACCAA ACTTCTCTACTGTGGTTTCCCCTCTGC CGCTTTCTCCTTGG CGTTCC  
 C CGATCACTGCCTCTAGGGTCTTTACAAGGGACAG CGAA CGTAAGGTTT CGGAGCTG  
 GCTT CGCCCCCTTCTATTTAC CGGGGGCTGGTCATCCTT CGGGCCAGGCTGACTGTCT  
 AGGGGTGGCCCT CGGATACCTGGTCCC CGCAGGTGACCCCTCCCTCCAC CGAGG  
 AGGAGGG CGCGTCC CGCT CGCTC CGTGGAGA CGC CGCG GAGG CGCACA CGCCTCAC  
 GCCTCCT CGCAGAAGCCACAGGGGCTG CGGCTCTTTGT CGCG GCCACAC CGGAAACG  
 GCC CGGAG CGCGCGCG GGG CGGCC CGAGGGT CGCGG CGCTAGG CGGGGC CGCG  
 GGCC CGTCCCCCTCCC CGCG GGGG CGCGCG AGGCC CGGG CGGGGC CGCCCTCCCTTC  
 CGGGGAG CGGGGAGC CGCG CGCGTC CGCGGG CGGCTCCACCC CGCCCC CGC  
 AGCTCTAGGTGA CGCG TCTCCAATAACAGCT CGCG GGAGC CGCAGCT CGGGCACAG

CCGGCGGC CGCG CCC CGCCG CCACCATGAGGGC CGAGGGCCT CGGCG GCCTGGAGC  
 GCTTCTGCAGCC CGGGCAAAGGC CGGGGGCTG CGGGCTCTGCAGCCCTTCCAGGTGG  
 GGGACTTGCTGTTCTCCTGCC CGGCCTATGCCTA CGTGCTCA CGGTCAA CGAG CGGG  
 GCAACCACTG CGAGTACTGCTTACCAGGTAGGG CGGCGGCGGCGGCGGCGGCGG  
 CGGGAGC CGGGGG CGCCGAG CGGGAGGCTTGGACGGCGGCGCCAGGAAGTG CGTGC  
 GGCTCGCGGGGTCTAG CGCCG GCCCTTGGGG CGGGGTGGGGGG CGGGGAGGGGAG  
 GAGGCCCGCGCTG CGGCCCGCGCTGCAGCCCCA CGCCAGG CGCCACGTGCGAGA  
 GGAAACAGGAC CGGTGCCCGCTGCA CGCCCTGCACTGCA CGTTTTTCAGAGTTGCTC  
 ACGCTGGATG CGCAAACC CGGGCGTG CAGGTTCTGCCCTCAAATGTGGGCTC

>SERPINB9

GGTTTCCATCACAGGTGCTGAAAACACTCTTGGAGGCAACTGCAGCTCTTTTGAAAG  
 TTTAATTGCGTCTTTCAATAATCCTTCCCCCAGGAGGAGCAAAGGCCAGTGCAGAT  
 GATGTATCAGGAGGCCA CGTTTAAGCT CGCCCA CGTGGG CGAGGTG CGCGCG CAGCT  
 GCTGGAGCTGCCCTA CGCCAGGAAGGAGCTGAGCCTGCTGGTGCTGCTGCCTGA CGA  
 CGGCGTGAGCTCAGCA CGGTAAGACC CGGGCTG CGGGAAGAACCAGGGACACCT  
 TTG CGGGCAGAACT CGAGTGCCACTTCCACCTCTCATATTCACCTTCTGAGTTGG CGA  
 TG CGGCAGA CGCACACTGTGCAGGCACTTGG CGTTGGGAT CGAACTTTTGTTCAAGG  
 CTGACTTTTCCCAATATTGTCTG CGTGATCTGCCAAACTACACATCTCAGTCCTCTGTT  
 TTTTCAT CGCTAGAAGGAGTGGG CGGCG GTTAAAATGCCTTTTAAAATAAATCAGGT  
 AGGAGGAAGCTTATAAAAATCCTGCTAGTTTGATGAATGGCAGACTTTTAACTCTAA  
 AAGCATAAATGACTGCT CGGGGTATTGCCTCCTAGAGTTATGATGAGGATAAAATG  
 TGGGTAATTCAGGTAACAAGCTTACCACATTGCCTGGCAACTAGTAAATGCTCTACA  
 AATGTGATCCATTATTACAATCATCATTCTTAGCATTATGCTTTTAAAGTCTAGGAAC  
 CTGAAATAGAGAATGAGGAAGTCTTAACTTTGACCTAACAAAGGCTGGGAGGGGGCA  
 TCCAGGGAGGTGTGAAGTCCAGTTCTCAGAGGGTGGGACCAGAGGCAGCCTCTGAG  
 GGCACATCCAGGCACT CGGCTTCTTTCAGCATGGATCTCCAGGGTGAGAGAGTCAA  
 CCATGGCAAATGGCAGGCAGTGGGACCCCATGACTTAAGAGGTCTGTGTG CGCAGCA  
 GCCATTGGCTCTCCCTCTGCCACCTTCTGAAACTATGCTC CGGAAACTATGCAGAA  
 TCTCTCTGGGGCAGATAATCCTCTGGTAACCTCTCTGCTC CGGAAACTATGCAGAATC  
 TCACCTGGGACAGATAATCCTGTTGGAGTGGCTCACACTGAGAGCCCTTT CGTTTCCA  
 ATCCTTTTCTTTACTAGCAATGGTTTGTGCTTTAAAAAGACTATGTTTGAAAAAAA  
 AACACAGT CGCCTGGGCATTGACACTTGTACAATAAAGATGTCTCTGGAGTTGTT  
 GGATAAGGAGAGGACAGCCACAGCCTTCTGCCTCTGACTGCTTGGTGTCTCTCCCC  
 CTTGTCAAGGAAGTGGCACATGGGACATAGGACAAGTAAGAGTGCTCACAAGCTTCT  
 GGCAGATGAAGGGGCCACTGATCCTATGTGAACAAACAACACCTA CGTGCAAGGGG

AGGAAGGGAAGTGTGAGTGCCTGTGCGCTTGTACATGTGCACTGAAGTCGGAATCTT  
 AAAGTCTAATTCTGGTCTTTCAGGTGGAAAAAGTCTCACTTTTGAGAACTCACAG  
 CCTGGACCAAGCCAGACTGTATGAAGAGTACTGAGGTTGAAGTTCTCCTTCCAAAAT  
 TAAACTACAAGAGGATTATGACATGGAATCTGTGCTTGGCATTGGAATTGTTG  
 ATGCCTTCCAACAGGGCAAGGCTGACTTGTGCAATGTCAGCGGAGAGAGACCTGT  
 GTCTGTCCAAGTTCTGTCACAAGAGTTTTGTGGAGGTGAATGAAGAAGGCACCGAGG  
 CAGCGGCAGCGTCTAGCTGCTTTGTAGTTGCAGAGTGCTGCATGGAATCTGGCCCCA  
 GGTTCTGTGCTGACCACCCTTTCCTTTTCTTCATCAGGCACAACAGAGCCAACAGCAT  
 TCTGTTCTGTGGCAGGTTCTCATCGCCATAAAGGGTGCACCTTACCGTGCCTCGGCCA  
 TTTCCCTCTTCTGTGTCCCAGATCCCCTACTACAGCTCCAAGAGGATGGGCCTAGAA  
 AGCCAAGTGCAAAGATGAGGGCAGATTCTTTACCTGTCTGCCCTCATGATTTGCCAG  
 CATGAATTCATGATGCTCCACACTCGCTTATGCTACTTAATCAGAATCTTGAGAAAAT  
 AGACCATAATGATTCCCTGTTGTATTAATAATTGCAGTCCAAATCCCATAGGATGGCA  
 AGCAAAGTTCTTCTAGAATTCCACATGCAATTCCTCTGGCGACCCTGTGCTTTCCTG  
 AACTGCGAATACATTCCTTAACCGCTGCCTCAGTGGTAATAAATGGTGCTAGATA  
 TTGCTACTATTTTATAGATTTCTGGTGCTTAGCCTTATAAAAAAGGTTGTAAAATGT  
 ACATTTATATTTTATCTTTTTTTTTTTTTTTTTTTCTGAGACGAGTCTGGCTCTCTGT  
 CG  
 CCCAGGCTGGAGTGCAGTGGCTCGATCTCGGCTCACTGCAAGCTCCCGCCTCCCGGGT  
 TCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCGCCAC  
 CACCGCCCGGCTAATTTTTTGTATTTTAGTAGAGACGGGGTTTCACCGTGTAGCCAG  
 GATGGTGTGATCTCCTGACCTCGTGATCCACC CGCCTCGGCCTCCCAAAGTGCTGGG  
 ATTACAGGCTTGAGCCAC

**>DAB2IP1**

GCCTCAGTCTGGTAATGGAATTACATGAGAGCTTTAGCTCTGGAGCAGCCTGTCTGCT  
 GTGTTGAACCCAGCTCTGCGCTTCTGACCGTGTGGTGCTGTGCAAATTAGTTGCC  
 TTTCTGAACTTCTGTTCTCCCATCTGCACACTGGGTTGTGAGAATTAAGAGAAGTCA  
 TGGGAAGCTCTAATACAGTGCCTGAAGCTTAGACAGGGCTTACTAAACACATCTGAC  
 TAATTCTCAGCAGCGTGTGGAAGAGCCCTCATTAGTATTAATAGCAGAGCTGGGTT  
 TTGACCCAGACTTATAGACTCCAAA CGTGGTTTTCTTGCCACTCATGCAGTGCACAG  
 AGCACAGGATAGGGGAGTTTTGCTGGGAATGAGCACTAAAGTCAGCAAGCCC CGG  
 GAGGACTGGCTGGGACCTCAGTGATAGGACGTGCAGGCAGGCGGAGATAGTATAGA  
 CCCTGCTGCAGGGGTGTAACCTCCTGGTGAATCGGCTTCCTGCAGCCA CGAGGCC  
 CTGGGTATCCAACCTCCCACCCCTTAGCCAGGGCTGGCTGAGAGCTTAAAGAGCTGA  
 TGAGGTGTTTGTAGCCATCCAGACCTCTGGGACAGGCTTATTATGCTAAGGCAAGTG  
 TCAGCTGGAGGATCTATGGGACAATGAGTCACTCCCTGGCTGCTCTGGGTTGTCACA

TCCCAAGCTGAGACCTTGGGAATGCGTAACTGGTTTCTGTTCTGATGCCCGCCTAA  
 TCCTGGGCACTGGCAGGTACCAGGAGGGCTCTGCTGACGTCCTCAGCTCCCCCTCCA  
 GCTTCACCTTTACCCTCTTCAAATCTGGGGCTCCCCTCTTTTTCTCCTCGTTAGACTCTCA  
 GTGTTCCCTTAGGGTTCCCCTTTACTTGGGTGTCCTCTGCCCTGGGGAAGGGATTACAA  
 GCCCAGGCTCAGGTGGTGGTGTATAACTTGGAGGCTGGAGGGACCTGTGATAACTCC  
 AGGGGCTCCCAGGGCAGCCAGACTCCCCTGTCCTACCTCTACACCTCTGCTTATG  
 CTGTGAGTCTGCCTGGAATGCCTTTATGTGGCTATCCTGCCCCAGCCAATCTCAGCAA  
 GCCCACCTGCTGTCCCACCTTCAATGTTTCTAAAATGCTGTCAAAAGCTGCCTCCTCC  
 AGCAAGTCTGCCCAGATTGCTCCCAAGTGAAATCCTTTTAAAAAGCCCCACAGAGCA  
 CTTTATCTTTTTCTTTTTATATCATTA

**>DAB2IP2**

TTGTTGGGAGAGGAGACAGGAAAGCCTCCAGGGTGACATGATCCCCAACTGCATGT  
 CAGGGTGAGGCGTTCATTGTCTCAAGAGGAGCAGAGGCTGGGCATCCCAGGGAGGG  
 CCATGCACTGTGCCAGGTCAGGAGAAGCGGGCGATGAGACCTAAGAGGACAGCAGG  
 GACCTGAAGGCCAGACAGGGCTGTGGGAGCCACCAGAGGGTTTTAAACAAGGGAGA  
 GGATGGCCTTCTCTGTTCTGGATCCTGAAAAGCTCACCTGACAGCAGTGGGCAGA  
 TGGGTTGGAGGGAGCAGGATGGTAGCCAGGAGACGTGAAAGGAGAGCATAGCTGTG  
 GTCCAGCAACAATAGCAGAAAATTCCAGGGAGGGGAGCGGTACAGGAGGCAGT  
 GGCTGCAAGGGACAAGGTTAGGGGTGATTAGAAGGCAAATCAGTGGGCCTGCCA  
 TTTGGGAATGAGGAAGAGGGAGGAGTTGACAGCTGCCAGAGATTCTGTCTCAGGTGT  
 ATCTGTATGGGTAGTGGGGTCTCCACTTAGGGTAGGGAGCTCAGGAGGAGGGCATGG  
 GTTTGGGTGTTGATGATGACTTCAGGTGTGGCAGGCTGAGTAGGAGGGACTCCCAGG  
 GGTGGTGCCTGCAGATATATTAATACATGCCTGAAAGCTCATTGGGAGACCTGGGTT  
 GAGCAGGGATTGAGCATCATCTGCATACCAGGGCAGTGAAACTTGGAGGAAAAGCG  
 GAAGGACACTGTAAACAGGCAGTGTATGGGGGACACAGAACATTCCAGGTCCTGAG  
 AATAAAGGTTAGAAAAGGATCGCTGGACTTACCATGAGCAAGGAGGGCCCTGGTG  
 ACCTGGGTGAGAGCATATTCAGCTGACTTGGGGAGTGAATGGAGTGGGAGCCTGCCA  
 TAGTGGGGAAAAGCGGGTGGTGGTGTCCCAGTGGAGGGAGTTCCTGGGCAGGGCCTG  
 CCTGGGCCAGTAGTGCTCACCAACTGCCCTCTCTCCAGATTTGGCAGCAAGGAGG  
 AATACATGTCCTTCATGAACCAGTTCCTAGAGCATGAGTGGACCAACATGCAGCGCT  
 TCCTGCTGGAGATCTCCAACCCCGAGACCCTCTCCAATACAGCCGCTTCGAGGGCT  
 ACATCGACCTGGGCAGCGAGCTCTCCAGCCTGCACTACTGCTCTGGGAGGCCTCA  
 GCCAGCTGGAGCAGGTGCCTGTTGCTGGGGCGGAGGTGGGGCCAAAAGCTGCCA  
 TCAGGCTTTTAGTGTTCCCCCTCCAGAGTAACCATGGAGGGCAGAGAGTTTGCCCA  
 AGTGGCATGATCAGGCCAGGTCCTGTCAGA

**References**

1. Abraha I, Romagnoli C, Montedori A, Cirocchi R. Thoracic stent graft versus surgery for thoracic aneurysm. *Cochrane Database Syst Rev.* 2009;Cd006796
2. Lu H, Rateri DL, Bruemmer D, Cassis LA, Daugherty A. Novel mechanisms of abdominal aortic aneurysms. *Current atherosclerosis reports.* 2012;14:402-412
3. Hager J, Lanne T, Carlsson P, Lundgren F. Lower prevalence than expected when screening 70-year-old men for abdominal aortic aneurysm. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery.* 2013;46:453-459
4. Svensjo S, Bjorck M, Wanhainen A. Current prevalence of abdominal aortic aneurysm in 70-year-old women. *The British journal of surgery.* 2013;100:367-372
5. Glover MJ, Kim LG, Sweeting MJ, Thompson SG, Buxton MJ. Cost-effectiveness of the national health service abdominal aortic aneurysm screening programme in England. *The British journal of surgery.* 2014;101:976-982
6. Iakoubova OA, Tong CH, Rowland CM, Luke MM, Garcia VE, Catanese JJ, Moomiaie RM, Sotonyi P, Ascady G, Nikas D, Dedelias P, Tranquilli M, Elefteriades JA. Genetic variants in *fbn-1* and risk for thoracic aortic aneurysm and dissection. *PloS one.* 2014;9:e91437
7. Earnshaw JJ. Triumphs and tribulations in a new national screening programme for abdominal aortic aneurysm. *Acta chirurgica Belgica.* 2012;112:108-110
8. Wilmink AB, Quick CR. Epidemiology and potential for prevention of abdominal aortic aneurysm. *The British journal of surgery.* 1998;85:155-162
9. Collaborators R, Bown MJ, Sweeting MJ, Brown LC, Powell JT, Thompson SG. Surveillance intervals for small abdominal aortic aneurysms: A meta-analysis. *JAMA : the journal of the American Medical Association.* 2013;309:806-813
10. Powell JT, Greenhalgh RM, Ruckley CV, Fowkes FG. The UK small aneurysm trial. *Ann N Y Acad Sci.* 1996;800:249-251

11. Schermerhorn ML, Cronenwett JL. The uk small aneurysm trial. *J Vasc Surg.* 2001;33:443
12. Cosford PA, Leng GC. Screening for abdominal aortic aneurysm. *Cochrane Database Syst Rev.* 2007:CD002945
13. Campbell WB. Mortality statistics for elective aortic aneurysms. *European journal of vascular surgery.* 1991;5:111-113
14. Participants UKSAT. Long-term outcomes of immediate repair compared with surveillance of small abdominal aortic aneurysms. *The New England journal of medicine.* 2002;346:1445-1452
15. Valdespino V, Valdespino PM. Potential of epigenetic therapies in the management of solid tumors. *Cancer Manag Res.* 2015;7:241-251
16. Napoli C, Grimaldi V, De Pascale MR, Sommese L, Infante T, Soricelli A. Novel epigenetic-based therapies useful in cardiovascular medicine. *World J Cardiol.* 2016;8:211-219
17. Longo GM, Buda SJ, Fiotta N, Xiong W, Griener T, Shapiro S, Baxter BT. Mmp-12 has a role in abdominal aortic aneurysms in mice. *Surgery.* 2005;137:457-462
18. Pearce WH, Shively VP. Abdominal aortic aneurysm as a complex multifactorial disease: Interactions of polymorphisms of inflammatory genes, features of autoimmunity, and current status of mmcs. *Annals of the New York Academy of Sciences.* 2006;1085:117-132
19. Samadzadeh KM, Chun KC, Nguyen AT, Baker PM, Bains S, Lee ES. Monocyte activity is linked with abdominal aortic aneurysm diameter. *The Journal of surgical research.* 2014;190:328-334
20. Sakalihasan N, Delvenne P, Nusgens BV, Limet R, Lapière CM. Activated forms of mmp2 and mmp9 in abdominal aortic aneurysms. *Journal of Vascular Surgery.* 1996;24:127-133
21. Newby AC. Metalloproteinase expression in monocytes and macrophages and its relationship to atherosclerotic plaque instability. *Arteriosclerosis, thrombosis, and vascular biology.* 2008;28:2108-2114
22. Nagase H, Visse R, Murphy G. Structure and function of matrix metalloproteinases and timp. *Cardiovascular research.* 2006;69:562-573

23. Bumdelger B, Kokubo H, Kamata R, Fujii M, Ishida M, Ishida T, Yoshizumi M. Induction of timp1 in smooth muscle cells during development of abdominal aortic aneurysms. *Hiroshima journal of medical sciences*. 2013;62:63-67
24. Xue M, Le NT, Jackson CJ. Targeting matrix metalloproteases to improve cutaneous wound healing. *Expert opinion on therapeutic targets*. 2006;10:143-155
25. Parks WC, Wilson CL, Lopez-Boado YS. Matrix metalloproteinases as modulators of inflammation and innate immunity. *Nat Rev Immunol*. 2004;4:617-629
26. Golledge J, Norman PE. Atherosclerosis and abdominal aortic aneurysm: Cause, response, or common risk factors? *Arteriosclerosis, thrombosis, and vascular biology*. 2010;30:1075-1077
27. Shimizu K, Mitchell RN, Libby P. Inflammation and cellular immune responses in abdominal aortic aneurysms. *Arteriosclerosis, thrombosis, and vascular biology*. 2006;26:987-994
28. Golledge J, Karan M, Moran CS, Muller J, Clancy P, Dear AE, Norman PE. Reduced expansion rate of abdominal aortic aneurysms in patients with diabetes may be related to aberrant monocyte-matrix interactions. *European heart journal*. 2008;29:665-672
29. St-Pierre Y, Van Themsche C, Esteve PO. Emerging features in the regulation of mmp-9 gene expression for the development of novel molecular targets and therapeutic strategies. *Current drug targets. Inflammation and allergy*. 2003;2:206-215
30. Chatzizisis YS, Coskun AU, Jonas M, Edelman ER, Feldman CL, Stone PH. Role of endothelial shear stress in the natural history of coronary atherosclerosis and vascular remodeling: Molecular, cellular, and vascular behavior. *Journal of the American College of Cardiology*. 2007;49:2379-2393
31. Dobrin PB, Mrkvicka R. Failure of elastin or collagen as possible critical connective tissue alterations underlying aneurysmal dilatation. *Cardiovascular surgery (London, England)*. 1994;2:484-488

32. Wassef M, Baxter BT, Chisholm RL, Dalman RL, Fillinger MF, Heinecke J, Humphrey JD, Kuivaniemi H, Parks WC, Pearce WH, Platsoucas CD, Sukhova GK, Thompson RW, Tilson MD, Zarins CK. Pathogenesis of abdominal aortic aneurysms: A multidisciplinary research program supported by the national heart, lung, and blood institute. *Journal of Vascular Surgery*. 2001;34:730-738
33. Thompson RW, Liao S, Curci JA. Vascular smooth muscle cell apoptosis in abdominal aortic aneurysms. *Coronary artery disease*. 1997;8:623-631
34. Alcorn HG, Wolfson SK, Jr., Sutton-Tyrrell K, Kuller LH, O'Leary D. Risk factors for abdominal aortic aneurysms in older adults enrolled in the cardiovascular health study. *Arterioscler Thromb Vasc Biol*. 1996;16:963-970
35. Sidloff D, Stather P, Dattani N, Bown M, Thompson J, Sayers R, Choke E. Aneurysm global epidemiology study: Public health measures can further reduce abdominal aortic aneurysm mortality. *Circulation*. 2014;129:747-753
36. Bergoeing MP, Arif B, Hackmann AE, Ennis TL, Thompson RW, Curci JA. Cigarette smoking increases aortic dilatation without affecting matrix metalloproteinase-9 and -12 expression in a modified mouse model of aneurysm formation. *J Vasc Surg*. 2007;45:1217-1227
37. Kuivaniemi H, Shibamura H, Arthur C, Berguer R, Cole CW, Juvonen T, Kline RA, Limet R, Mackean G, Norrgard O, Pals G, Powell JT, Rainio P, Sakalihan N, van Vlijmen-van Keulen C, Verloes A, Tromp G. Familial abdominal aortic aneurysms: Collection of 233 multiplex families. *J Vasc Surg*. 2003;37:340-345
38. Cornuz J, Sidoti Pinto C, Tevaearai H, Egger M. Risk factors for asymptomatic abdominal aortic aneurysm: Systematic review and meta-analysis of population-based screening studies. *European journal of public health*. 2004;14:343-349
39. Sweeting MJ, Thompson SG, Brown LC, Powell JT. Meta-analysis of individual patient data to examine factors affecting growth and rupture of small abdominal aortic aneurysms. *The British journal of surgery*. 2012;99:655-665

40. Norman PE, Curci JA. Understanding the effects of tobacco smoke on the pathogenesis of aortic aneurysm. *Arteriosclerosis, thrombosis, and vascular biology*. 2013;33:1473-1477
41. Wahlgren CM, Larsson E, Magnusson PK, Hultgren R, Swedenborg J. Genetic and environmental contributions to abdominal aortic aneurysm development in a twin population. *J Vasc Surg*. 2010;51:3-7; discussion 7
42. Joergensen TM, Christensen K, Lindholt JS, Larsen LA, Green A, Houliand K. Editor's choice - high heritability of liability to abdominal aortic aneurysms: A population based twin study. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2016;52:41-46
43. Kent KC, Zwolak RM, Egorova NN, Riles TS, Manganaro A, Moskowitz AJ, Gelijns AC, Greco G. Analysis of risk factors for abdominal aortic aneurysm in a cohort of more than 3 million individuals. *J Vasc Surg*. 2010;52:539-548
44. Palazzuoli A, Gallotta M, Guerrieri G, Quatrini I, Franci B, Campagna MS, Neri E, Benvenuti A, Sassi C, Nuti R. Prevalence of risk factors, coronary and systemic atherosclerosis in abdominal aortic aneurysm: Comparison with high cardiovascular risk population. *Vascular health and risk management*. 2008;4:877-883
45. Bradley DT, Hughes AE, Badger SA, Jones GT, Harrison SC, Wright BJ, Bumpstead S, Baas AF, Gretarsdottir S, Burnand K, Child AH, Clough RE, Cockerill G, Hafez H, Scott DJ, Ariens RA, Johnson A, Sohrabi S, Smith A, Thompson MM, van Bockxmeer FM, Waltham M, Matthiasson SE, Thorleifsson G, Thorsteinsdottir U, Blankensteijn JD, Teijink JA, Wijmenga C, de Graaf J, Kiemeney LA, Wild JB, Edkins S, Gwilliam R, Hunt SE, Potter S, Lindholt JS, Golledge J, Norman PE, van Rij A, Powell JT, Eriksson P, Stefansson K, Thompson JR, Humphries SE, Sayers RD, Deloukas P, Samani NJ, Bown MJ. A variant in *ldlr* is associated with abdominal aortic aneurysm. *Circulation. Cardiovascular genetics*. 2013;6:498-504
46. Herijgers N, Van Eck M, Groot PH, Hoogerbrugge PM, Van Berkel TJ. Low density lipoprotein receptor of macrophages

- facilitates atherosclerotic lesion formation in c57bl/6 mice. *Arteriosclerosis, thrombosis, and vascular biology*. 2000;20:1961-1967
47. Peshkova IO, Schaefer G, Koltsova EK. Atherosclerosis and aortic aneurysm - is inflammation a common denominator? *The FEBS journal*. 2016;283:1636-1652
  48. Elkalioubie A, Haulon S, Duhamel A, Rosa M, Rauch A, Staels B, Susen S, Van Belle E, Dupont A. Meta-analysis of abdominal aortic aneurysm in patients with coronary artery disease. *Am J Cardiol*. 2015;116:1451-1456
  49. Lee AJ, Fowkes FG, Carson MN, Leng GC, Allan PL. Smoking, atherosclerosis and risk of abdominal aortic aneurysm. *European heart journal*. 1997;18:671-676
  50. Ito S, Akutsu K, Tamori Y, Sakamoto S, Yoshimuta T, Hashimoto H, Takeshita S. Differences in atherosclerotic profiles between patients with thoracic and abdominal aortic aneurysms. *The American Journal of Cardiology*. 2008;101:696-699
  51. Johnsen SH, Forsdahl SH, Singh K, Jacobsen BK. Atherosclerosis in abdominal aortic aneurysms: A causal event or a process running in parallel? The tromso study. *Arteriosclerosis, thrombosis, and vascular biology*. 2010;30:1263-1268
  52. De Rango P, Farchioni L, Fiorucci B, Lenti M. Diabetes and abdominal aortic aneurysms. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2014;47:243-261
  53. Lederle FA. The strange relationship between diabetes and abdominal aortic aneurysm. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2012;43:254-256
  54. Torsney E, Pirianov G, Cockerill GW. Diabetes as a negative risk factor for abdominal aortic aneurysm - does the disease aetiology or the treatment provide the mechanism of protection? *Current vascular pharmacology*. 2013;11:293-298
  55. Nathan DM, Cleary PA, Backlund JY, Genuth SM, Lachin JM, Orchard TJ, Raskin P, Zinman B. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. *The New England journal of medicine*. 2005;353:2643-2653

56. Shantikumar S, Ajjan R, Porter KE, Scott DJ. Diabetes and the abdominal aortic aneurysm. *European journal of vascular and endovascular surgery : the official journal of the European Society for Vascular Surgery*. 2010;39:200-207
57. Chait A, Bornfeldt KE. Diabetes and atherosclerosis: Is there a role for hyperglycemia? *Journal of lipid research*. 2009;50 Suppl:S335-339
58. Miyama N, Dua MM, Yeung JJ, Schultz GM, Asagami T, Sho E, Sho M, Dalman RL. Hyperglycemia limits experimental aortic aneurysm progression. *J Vasc Surg*. 2010;52:975-983
59. Budoff MJ, Yang TP, Shavelle RM, Lamont DH, Brundage BH. Ethnic differences in coronary atherosclerosis. *Journal of the American College of Cardiology*. 2002;39:408-412
60. Kuivaniemi H, Ryer EJ, Elmore JR, Hinterseher I, Smelser DT, Tromp G. Update on abdominal aortic aneurysm research: From clinical to genetic studies. *Scientifica (Cairo)*. 2014;2014:564734
61. Blanchard JF, Armenian HK, Friesen PP. Risk factors for abdominal aortic aneurysm: Results of a case-control study. *American journal of epidemiology*. 2000;151:575-583
62. Zeina AR, Barmeir E, Zaid G, Odeh M. Coronary artery disease among hypertensive patients undergoing coronary computed tomography angiography. *Journal of cardiovascular medicine (Hagerstown, Md.)*. 2009;10:252-256
63. Bartnik M, Ryden L, Ferrari R, Malmberg K, Pyorala K, Simoons M, Standl E, Soler-Soler J, Ohrvik J. The prevalence of abnormal glucose regulation in patients with coronary artery disease across europe. The euro heart survey on diabetes and the heart. *European heart journal*. 2004;25:1880-1890
64. Sidloff D, Choke E, Stather P, Bown M, Thompson J, Sayers R. Mortality from thoracic aortic diseases and associations with cardiovascular risk factors. *Circulation*. 2014;130:2287-2294
65. Goldfinger JZ, Halperin JL, Marin ML, Stewart AS, Eagle KA, Fuster V. Thoracic aortic aneurysm and dissection. *Journal of the American College of Cardiology*. 2014;64:1725-1739
66. Pape LA, Tsai TT, Isselbacher EM, Oh JK, O'Gara P T, Evangelista A, Fattori R, Meinhardt G, Trimarchi S, Bossone E, Suzuki T, Cooper JV, Froehlich JB, Nienaber CA, Eagle KA.

- Aortic diameter  $\geq$  5.5 cm is not a good predictor of type a aortic dissection: Observations from the international registry of acute aortic dissection (irad). *Circulation*. 2007;116:1120-1127
67. Ruddy JM, Jones JA, Ikonomidis JS. Pathophysiology of thoracic aortic aneurysm (taa): Is it not one uniform aorta? Role of embryologic origin. *Progress in Cardiovascular Diseases*. 2013;56:68-73
  68. Bown MJ. Genomic insights into abdominal aortic aneurysms. *Annals of the Royal College of Surgeons of England*. 2014;96:405-414
  69. Isselbacher EM. Thoracic and abdominal aortic aneurysms. *Circulation*. 2005;111:816-828
  70. Olsson C, Thelin S, Stahle E, Ekblom A, Granath F. Thoracic aortic aneurysm and dissection: Increasing prevalence and improved outcomes reported in a nationwide population-based study of more than 14,000 cases from 1987 to 2002. *Circulation*. 2006;114:2611-2618
  71. Davies RR, Goldstein LJ, Coady MA, Tittle SL, Rizzo JA, Kopf GS, Elefteriades JA. Yearly rupture or dissection rates for thoracic aortic aneurysms: Simple prediction based on size. *The Annals of thoracic surgery*. 2002;73:17-27; discussion 27-18
  72. Upchurch GR, Jr., Schaub TA. Abdominal aortic aneurysm. *American family physician*. 2006;73:1198-1204
  73. Milewicz DM, Chen H, Park E-S, Petty EM, Zaghi H, Pai GS, Willing M, Patel V. Reduced penetrance and variable expressivity of familial thoracic aortic aneurysms/dissections. *The American Journal of Cardiology*. 1998;82:474-479
  74. Saratzis A, Bown MJ, Wild B, Nightingale P, Smith J, Johnson C, Melas N, Kitas GD. Association between seven single nucleotide polymorphisms involved in inflammation and proteolysis and abdominal aortic aneurysm. *J Vasc Surg*. 2015;61:1120-1128.e1121
  75. Saratzis A, Bown MJ. The genetic basis for aortic aneurysmal disease. *Heart (British Cardiac Society)*. 2014;100:916-922
  76. Saratzis A, Bown M, Wild B, Sayers RD, Nightingale P, Smith J, Johnson C, Kitas G. C-reactive protein polymorphism rs3091244

- is associated with abdominal aortic aneurysm. *J Vasc Surg.* 2014;60:1332-1339
77. Gretarsdottir S, Baas AF, Thorleifsson G, Holm H, den Heijer M, de Vries JP, Kranendonk SE, Zeebregts CJ, van Sterkenburg SM, Geelkerken RH, van Rij AM, Williams MJ, Boll AP, Kostic JP, Jonasdottir A, Jonasdottir A, Walters GB, Masson G, Sulem P, Saemundsdottir J, Mouy M, Magnusson KP, Tromp G, Elmore JR, Sakalihasan N, Limet R, Defraigne JO, Ferrell RE, Ronkainen A, Ruigrok YM, Wijmenga C, Grobbee DE, Shah SH, Granger CB, Quyyumi AA, Vaccarino V, Patel RS, Zafari AM, Levey AI, Austin H, Girelli D, Pignatti PF, Olivieri O, Martinelli N, Malerba G, Trabetti E, Becker LC, Becker DM, Reilly MP, Rader DJ, Mueller T, Dieplinger B, Haltmayer M, Urbonavicius S, Lindblad B, Gottsater A, Gaetani E, Pola R, Wells P, Rodger M, Forgie M, Langlois N, Corral J, Vicente V, Fontcuberta J, Espana F, Grarup N, Jorgensen T, Witte DR, Hansen T, Pedersen O, Aben KK, de Graaf J, Holewijn S, Folkersen L, Franco-Cereceda A, Eriksson P, Collier DA, Stefansson H, Steinhorsdottir V, Rafnar T, Valdimarsson EM, Magnadottir HB, Sveinbjornsdottir S, Olafsson I, Magnusson MK, Palmason R, Haraldsdottir V, Andersen K, Onundarson PT, Thorgeirsson G, Kiemenev LA, Powell JT, Carey DJ, Kuivaniemi H, Lindholt JS, Jones GT, Kong A, Blankensteijn JD, Matthiasson SE, Thorsteinsdottir U, Stefansson K. Genome-wide association study identifies a sequence variant within the *dab2ip* gene conferring susceptibility to abdominal aortic aneurysm. *Nature genetics.* 2010;42:692-697
78. Bown MJ, Jones GT, Harrison SC, Wright BJ, Bumpstead S, Baas AF, Gretarsdottir S, Badger SA, Bradley DT, Burnand K, Child AH, Clough RE, Cockerill G, Hafez H, Scott DJ, Futers S, Johnson A, Sohrabi S, Smith A, Thompson MM, van Bockxmeer FM, Waltham M, Matthiasson SE, Thorleifsson G, Thorsteinsdottir U, Blankensteijn JD, Teijink JA, Wijmenga C, de Graaf J, Kiemenev LA, Assimes TL, McPherson R, Folkersen L, Franco-Cereceda A, Palmen J, Smith AJ, Sylvius N, Wild JB, Refstrup M, Edkins S, Gwilliam R, Hunt SE, Potter S, Lindholt JS, Frikke-Schmidt R, Tybjaerg-Hansen A, Hughes AE, Golledge J, Norman PE, van Rij A, Powell JT, Eriksson P, Stefansson K,

- Thompson JR, Humphries SE, Sayers RD, Deloukas P, Samani NJ. Abdominal aortic aneurysm is associated with a variant in low-density lipoprotein receptor-related protein 1. *American journal of human genetics*. 2011;89:619-627
79. Jones GT, Bown MJ, Gretarsdottir S, Romaine SP, Helgadottir A, Yu G, Tromp G, Norman PE, Jin C, Baas AF, Blankensteijn JD, Kullo IJ, Phillips LV, Williams MJ, Topless R, Merriman TR, Vasudevan TM, Lewis DR, Blair RD, Hill AA, Sayers RD, Powell JT, Deloukas P, Thorleifsson G, Matthiasson SE, Thorsteinsdottir U, Golledge J, Ariens RA, Johnson A, Sohrabi S, Scott DJ, Carey DJ, Erdman R, Elmore JR, Kuivaniemi H, Samani NJ, Stefansson K, van Rij AM. A sequence variant associated with sortilin-1 (sort1) on 1p13.3 is independently associated with abdominal aortic aneurysm. *Human molecular genetics*. 2013;22:2941-2947
80. Harrison SC, Smith AJ, Jones GT, Swerdlow DI, Rampuri R, Bown MJ, Folkersen L, Baas AF, de Borst GJ, Blankensteijn JD, Price JF, van der Graaf Y, McLachlan S, Agu O, Hofman A, Uitterlinden AG, Franco-Cereceda A, Ruigrok YM, van't Hof FN, Powell JT, van Rij AM, Casas JP, Eriksson P, Holmes MV, Asselbergs FW, Hingorani AD, Humphries SE. Interleukin-6 receptor pathways in abdominal aortic aneurysm. *European heart journal*. 2013;34:3707-3716
81. Bown MJ, Braund PS, Thompson J, London NJ, Samani NJ, Sayers RD. Association between the coronary artery disease risk locus on chromosome 9p21.3 and abdominal aortic aneurysm. *Circulation. Cardiovascular genetics*. 2008;1:39-42
82. Leeper NJ, Raiesdana A, Kojima Y, Kundu RK, Cheng H, Maegdefessel L, Toh R, Ahn GO, Ali ZA, Anderson DR, Miller CL, Roberts SC, Spin JM, de Almeida PE, Wu JC, Xu B, Cheng K, Quertermous M, Kundu S, Kortekaas KE, Berzin E, Downing KP, Dalman RL, Tsao PS, Schadt EE, Owens GK, Quertermous T. Loss of cdkn2b promotes p53-dependent smooth muscle cell apoptosis and aneurysm formation. *Arteriosclerosis, thrombosis, and vascular biology*. 2013;33:e1-e10
83. Jones GT, Tromp G, Kuivaniemi H, Gretarsdottir S, Baas AF, Giusti B, Strauss E, van 't Hof FN, Webb T, Erdman R, Ritchie

- MD, Elmore JR, Verma A, Pendergrass S, Kullo IJ, Ye Z, Peissig PL, Gottesman O, Verma SS, Malinowski J, Rasmussen-Torvik LJ, Borthwick K, Smelser DT, Crosslin DR, de Andrade M, Ryer EJ, McCarty CA, Bottinger EP, Pacheco JA, Crawford DC, Carrell DS, Gerhard GS, Franklin DP, Carey DJ, Phillips VL, Williams MJ, Wei W, Blair R, Hill AA, Vasudevan TM, Lewis DR, Thomson IA, Krysa J, Hill GB, Roake J, Merriman TR, Oszkinis G, Galora S, Saracini C, Abbate R, Pulli R, Pratesi C, Saratzis A, Verissimo A, Bumpstead SJ, Badger SA, Clough RE, Cockerill GW, Hafez H, Scott DJ, Futers TS, Romaine SP, Bridge K, Griffin KJ, Bailey MA, Smith A, Thompson MM, van Bockxmeer F, Matthiasson SE, Thorleifsson G, Thorsteinsdottir U, Blankensteijn JD, Teijink JA, Wijmenga C, de Graaf J, Kiemeny LA, Lindholt JS, Hughes AE, Bradley DT, Stirrups K, Golledge J, Norman PE, Powell JT, Humphries SE, Hamby SE, Goodall AH, Nelson CP, Sakalihasan N, Courtois A, Ferrell RE, Eriksson P, Folkersen L, Franco-Cereceda A, Eicher JD, Johnson AD, Betsholtz C, Ruusalepp A, Franzén O, Schadt E, Björkegren JL, Lipovich L, Drolet AM, Verhoeven E, Zeebregts CJ, Geelkerken RH, van Sambeek MR, van Sterkenburg SM, de Vries J-PP, Stefansson K, Thompson JR, de Bakker PI, Deloukas P, Sayers RD, Harrison S, van Rij AM, Samani NJ, Bown MJ. Meta-analysis of genome-wide association studies for abdominal aortic aneurysm identifies four new disease-specific risk loci. *Circulation research*. 2016
84. Andelfinger G, Loeys B, Dietz H. A decade of discovery in the genetic understanding of thoracic aortic disease. *Canadian Journal of Cardiology*. 2016;32:13-25
85. Pomianowski P, Eleftheriades JA. The genetics and genomics of thoracic aortic disease. *Annals of cardiothoracic surgery*. 2013;2:271-279
86. Collod-Beroud G, Le Bourdelles S, Ades L, Ala-Kokko L, Booms P, Boxer M, Child A, Comeglio P, De Paepe A, Hyland JC, Holman K, Kaitila I, Loeys B, Matyas G, Nuytinck L, Peltonen L, Rantamaki T, Robinson P, Steinmann B, Junien C, Beroud C, Boileau C. Update of the umd-fbn1 mutation database

- and creation of an *fbn1* polymorphism database. *Human mutation*. 2003;22:199-208
87. Jones JA, Spinale FG, Ikonomidis JS. Transforming growth factor-beta signaling in thoracic aortic aneurysm development: A paradox in pathogenesis. *Journal of vascular research*. 2009;46:119-137
88. Holm TM, Habashi JP, Doyle JJ, Bedja D, Chen Y, van Erp C, Lindsay ME, Kim D, Schoenhoff F, Cohn RD, Loeys BL, Thomas CJ, Patnaik S, Marugan JJ, Judge DP, Dietz HC. Noncanonical *tgfbeta* signaling contributes to aortic aneurysm progression in marfan syndrome mice. *Science (New York, N.Y.)*. 2011;332:358-361
89. Loeys BL, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, Meyers J, Leitch CC, Katsanis N, Sharifi N, Xu FL, Myers LA, Spevak PJ, Cameron DE, De Backer J, Hellemans J, Chen Y, Davis EC, Webb CL, Kress W, Coucke P, Rifkin DB, De Paepe AM, Dietz HC. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in *tgfbr1* or *tgfbr2*. *Nature genetics*. 2005;37:275-281
90. Tran-Fadulu V, Pannu H, Kim DH, Vick GW, 3rd, Lonsford CM, Lafont AL, Boccalandro C, Smart S, Peterson KL, Hain JZ, Willing MC, Coselli JS, LeMaire SA, Ahn C, Byers PH, Milewicz DM. Analysis of multigenerational families with thoracic aortic aneurysms and dissections due to *tgfbr1* or *tgfbr2* mutations. *Journal of medical genetics*. 2009;46:607-613
91. Micha D, Guo DC, Hilhorst-Hofstee Y, van Kooten F, Atmaja D, Overwater E, Cayami FK, Regalado ES, van Uffelen R, Venselaar H, Faradz SM, Vriend G, Weiss MM, Sistermans EA, Maugeri A, Milewicz DM, Pals G, van Dijk FS. *Smad2* mutations are associated with arterial aneurysms and dissections. *Human mutation*. 2015;36:1145-1149
92. Schepers D, Doyle AJ, Oswald G, Sparks E, Myers L, Willems PJ, Mansour S, Simpson MA, Frysira H, Maat-Kievit A, Van Minkelen R, Hoogeboom JM, Mortier GR, Titheradge H, Brueton L, Starr L, Stark Z, Ockeloen C, Lourenco CM, Blair E, Hobson E, Hurst J, Maystadt I, Destree A, Girisha KM, Miller M, Dietz HC, Loeys B, Van Laer L. The *smad*-binding domain of *ski*: A

- hotspot for de novo mutations causing shprintzen-goldberg syndrome. *European journal of human genetics : EJHG*. 2015;23:224-228
93. Morisaki T, Morisaki H. Genetics of hereditary large vessel diseases. *Journal of human genetics*. 2016;61:21-26
94. Keramati AR, Sadeghpour A, Farahani MM, Chandok G, Mani A. The non-syndromic familial thoracic aortic aneurysms and dissections maps to 15q21 locus. *BMC medical genetics*. 2010;11:143
95. LeMaire SA, McDonald ML, Guo DC, Russell L, Miller CC, 3rd, Johnson RJ, Bekheirnia MR, Franco LM, Nguyen M, Pyeritz RE, Bavaria JE, Devereux R, Maslen C, Holmes KW, Eagle K, Body SC, Seidman C, Seidman JG, Isselbacher EM, Bray M, Coselli JS, Estrera AL, Safi HJ, Belmont JW, Leal SM, Milewicz DM. Genome-wide association study identifies a susceptibility locus for thoracic aortic aneurysms and aortic dissections spanning *fbn1* at 15q21.1. *Nature genetics*. 2011;43:996-1000
96. Guo DC, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, Johnson RJ, Kim DH, Pannu H, Willing MC, Sparks E, Pyeritz RE, Singh MN, Dalman RL, Grotta JC, Marian AJ, Boerwinkle EA, Frazier LQ, LeMaire SA, Coselli JS, Estrera AL, Safi HJ, Veeraraghavan S, Muzny DM, Wheeler DA, Willerson JT, Yu RK, Shete SS, Scherer SE, Raman CS, Buja LM, Milewicz DM. Mutations in smooth muscle alpha-actin (*acta2*) cause coronary artery disease, stroke, and moyamoya disease, along with thoracic aortic disease. *American journal of human genetics*. 2009;84:617-627
97. Guo DC, Pannu H, Tran-Fadulu V, Papke CL, Yu RK, Avidan N, Bourgeois S, Estrera AL, Safi HJ, Sparks E, Amor D, Ades L, McConnell V, Willoughby CE, Abuelo D, Willing M, Lewis RA, Kim DH, Scherer S, Tung PP, Ahn C, Buja LM, Raman CS, Shete SS, Milewicz DM. Mutations in smooth muscle alpha-actin (*acta2*) lead to thoracic aortic aneurysms and dissections. *Nature genetics*. 2007;39:1488-1493
98. Zhu L, Vranckx R, Khau Van Kien P, Lalande A, Boisset N, Mathieu F, Wegman M, Glancy L, Gasc JM, Brunotte F, Bruneval P, Wolf JE, Michel JB, Jeunemaitre X. Mutations in

- myosin heavy chain 11 cause a syndrome associating thoracic aortic aneurysm/aortic dissection and patent ductus arteriosus. *Nature genetics*. 2006;38:343-349
99. Renard M, Callewaert B, Baetens M, Campens L, MacDermot K, Fryns J-P, Bonduelle M, Dietz HC, Gaspar IM, Cavaco D, Stattin E-L, Schrandt-Stumpel C, Coucke P, Loeys B, De Paepe A, De Backer J. Novel myh11 and acta2 mutations reveal a role for enhanced tgfb $\beta$  signaling in ftaad. *International Journal of Cardiology*. 2013;165:314-321
  100. Verma M, Banerjee HN. Epigenetic inhibitors. *Methods in molecular biology (Clifton, N.J.)*. 2015;1238:469-485
  101. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004;429:457-463
  102. Zaina S, Heyn H, Carmona FJ, Varol N, Sayols S, Condom E, Ramirez-Ruz J, Gomez A, Goncalves I, Moran S, Esteller M. A DNA methylation map of human atherosclerosis. *Circulation. Cardiovascular genetics*. 2014
  103. Wongtrakongate P. Epigenetic therapy of cancer stem and progenitor cells by targeting DNA methylation machineries. *World journal of stem cells*. 2015;7:137-148
  104. Clark C, Palta P, Joyce CJ, Scott C, Grundberg E, Deloukas P, Palotie A, Coffey AJ. A comparison of the whole genome approach of medip-seq to the targeted approach of the infinium humanmethylation450 beadchip((r)) for methylome profiling. *PloS one*. 2012;7:e50233
  105. Michels KB, Binder AM, Dedeurwaerder S, Epstein CB, Grealley JM, Gut I, Houseman EA, Izzi B, Kelsey KT, Meissner A, Milosavljevic A, Siegmund KD, Bock C, Irizarry RA. Recommendations for the design and analysis of epigenome-wide association studies. *Nature methods*. 2013;10:949-955
  106. Jeltsch A. Reading and writing DNA methylation. *Nature structural & molecular biology*. 2008;15:1003-1004
  107. Bird A. DNA methylation patterns and epigenetic memory. *Genes & development*. 2002;16:6-21
  108. Goldberg AD, Allis CD, Bernstein E. Epigenetics: A landscape takes shape. *Cell*. 2007;128:635-638

109. Tsaprouni LG, Yang TP, Bell J, Dick KJ, Kanoni S, Nisbet J, Vinuela A, Grundberg E, Nelson CP, Meduri E, Buil A, Cambien F, Hengstenberg C, Erdmann J, Schunkert H, Goodall AH, Ouwehand WH, Dermitzakis E, Spector TD, Samani NJ, Deloukas P. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. *Epigenetics : official journal of the DNA Methylation Society*. 2014;9:1382-1396
110. Glier MB, Green TJ, Devlin AM. Methyl nutrients, DNA methylation, and cardiovascular disease. *Molecular nutrition & food research*. 2014;58:172-182
111. Jaenisch R, Bird A. Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature genetics*. 2003;33 Suppl:245-254
112. Kawasaki H, Taira K. Induction of DNA methylation and gene silencing by short interfering rnas in human cells. *Nature*. 2004;431:211-217
113. Cao Q, Wang X, Jia L, Mondal AK, Diallo A, Hawkins GA, Das SK, Parks JS, Yu L, Shi H, Shi H, Xue B. Inhibiting DNA methylation by 5-aza-2'-deoxycytidine ameliorates atherosclerosis through suppressing macrophage inflammation. *Endocrinology*. 2014;155:4925-4938
114. Yang X, Han H, De Carvalho DD, Lay FD, Jones PA, Liang G. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer cell*. 2014;26:577-590
115. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell research*. 2011;21:381-395
116. Conaway JW. Introduction to theme "chromatin, epigenetics, and transcription". *Annual review of biochemistry*. 2012;81:61-64
117. Robertson AK, Geiman TM, Sankpal UT, Hager GL, Robertson KD. Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases dnmt1 and dnmt3a in vitro. *Biochemical and biophysical research communications*. 2004;322:110-118
118. Chuang LS, Iannelli H, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-pcna complex as a target for p21waf1. *Science (New York, N.Y.)*. 1997;277:1996-2000

119. Kroeze LI, van der Reijden BA, Jansen JH. 5-hydroxymethylcytosine: An epigenetic mark frequently deregulated in cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*.
120. Drong AW, Nicholson G, Hedman AK, Meduri E, Grundberg E, Small KS, Shin SY, Bell JT, Karpe F, Soranzo N, Spector TD, McCarthy MI, Deloukas P, Rantalainen M, Lindgren CM. The presence of methylation quantitative trait loci indicates a direct genetic influence on the level of DNA methylation in adipose tissue. *PLoS one*. 2013;8:e55923
121. Zhi D, Aslibekyan S, Irvin MR, Claas SA, Borecki IB, Ordovas JM, Absher DM, Arnett DK. Snps located at cpg sites modulate genome-epigenome interaction. *Epigenetics : official journal of the DNA Methylation Society*. 2013;8:802-806
122. Rushton MD, Reynard LN, Young DA, Shepherd C, Aubourg G, Gee F, Darlay R, Deehan D, Cordell HJ, Loughlin J. Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. *Human molecular genetics*. 2015;24:7432-7444
123. Zhang H, Wang F, Kranzler HR, Yang C, Xu H, Wang Z, Zhao H, Gelernter J. Identification of methylation quantitative trait loci (mqtls) influencing promoter DNA methylation of alcohol dependence risk genes. *Human genetics*. 2014;133:1093-1104
124. Kitamoto T, Kitamoto A, Ogawa Y, Honda Y, Imajo K, Saito S, Yoneda M, Nakamura T, Nakajima A, Hotta K. Targeted-bisulfite sequence analysis of the methylation of cpg islands in genes encoding pnpla3, samm50, and parvb of patients with non-alcoholic fatty liver disease. *Journal of hepatology*. 2015;63:494-502
125. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P. Disease variants alter transcription factor levels and methylation of their binding sites. 2017;49:131-138
126. Richmond RC, Sharp GC, Ward ME, Fraser A, Lyttleton O, McArdle WL, Ring SM, Gaunt TR, Lawlor DA, Davey Smith G, Relton CL. DNA methylation and bmi: Investigating identified methylation sites at hif3a in a causal framework. *Diabetes*. 2016;65:1231-1244

127. Yoo S, Takikawa S, Geraghty P, Argmann C, Campbell J, Lin L, Huang T, Tu Z, Foronjy RF, Spira A, Schadt EE, Powell CA, Zhu J. Integrative analysis of DNA methylation and gene expression data identifies *epas1* as a key regulator of copd. *PLoS Genet.* 2015;11:e1004898
128. van Eijk KR, de Jong S, Boks MP, Langeveld T, Colas F, Veldink JH, de Kovel CG, Janson E, Strengman E, Langfelder P, Kahn RS, van den Berg LH, Horvath S, Ophoff RA. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. *BMC Genomics.* 2012;13:636
129. Heyn H, Sayols S, Moutinho C, Vidal E, Sanchez-Mut JV, Stefansson OA, Nadal E, Moran S, Eyfjord JE, Gonzalez-Suarez E, Pujana MA, Esteller M. Linkage of DNA methylation quantitative trait loci to human cancer risk. *Cell reports.* 2014;7:331-338
130. McClay JL, Shabalin AA, Dozmorov MG, Adkins DE, Kumar G, Nerella S, Clark SL, Bergen SE, Hultman CM, Magnusson PK, Sullivan PF, Aberg KA, van den Oord EJ. High density methylation qtl analysis in human blood via next-generation sequencing of the methylated genomic DNA fraction. *Genome biology.* 2015;16:291
131. Robertson KD. DNA methylation and human disease. *Nature reviews. Genetics.* 2005;6:597-610
132. Peirce JL. How replicable are mrna expression qtl? *Mamm Genome.* 2006;17
133. Huang GJ. High resolution mapping of expression qtls in heterogeneous stock mice in multiple tissues. *Genome Res.* 2009;19
134. Fehrmann RS. Trans-eqtls reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the hla. *PLoS Genet.* 2011;7
135. Ding J. Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eqtl signals. *American journal of human genetics.* 2010;87
136. Bryois J, Buil A, Evans DM, Kemp JP, Montgomery SB, Conrad DF, Ho KM, Ring S, Hurles M, Deloukas P, Davey Smith G,

- Dermitzakis ET. Cis and trans effects of human genomic variants on gene expression. *PLoS Genet.* 2014;10:e1004461
137. Brown CD, Mangravite LM, Engelhardt BE. Integrative modeling of eqtls and cis-regulatory elements suggests mechanisms underlying cell type specificity of eqtls. *PLoS Genet.* 2013;9:e1003649
138. Marian AJ, Belmont J. Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circulation research.* 2011;108:1252-1269
139. Lund G, Zaina S. Atherosclerosis: An epigenetic balancing act that goes wrong. *Current atherosclerosis reports.* 2011;13:208-214
140. Wang X, Falkner B, Zhu H, Shi H, Su S, Xu X, Sharma AK, Dong Y, Treiber F, Gutin B, Harshfield G, Snieder H. A genome-wide methylation study on essential hypertension in young african american males. *PloS one.* 2013;8:e53938
141. Hur K, Han TS, Jung EJ, Yu J, Lee HJ, Kim WH, Goel A, Yang HK. Up-regulated expression of sulfatases (sulf1 and sulf2) as prognostic and metastasis predictive markers in human gastric cancer. *The Journal of pathology.* 2012;228:88-98
142. Baccarelli A, Wright R, Bollati V, Litonjua A, Zanobetti A, Tarantini L, Sparrow D, Vokonas P, Schwartz J. Ischemic heart disease and stroke in relation to blood DNA methylation. *Epidemiology (Cambridge, Mass.).* 2010;21:819-828
143. Ryer EJ, Ronning KE, Erdman R, Schworer CM, Elmore JR, Peeler TC, Nevius CD, Lillvis JH, Garvin RP, Franklin DP, Kuivaniemi H, Tromp G. The potential role of DNA methylation in abdominal aortic aneurysms. *International journal of molecular sciences.* 2015;16:11259-11275
144. Krishna SM, Dear AE, Norman PE, Golledge J. Genetic and epigenetic mechanisms and their possible role in abdominal aortic aneurysm. *Atherosclerosis.* 2010;212:16-29
145. Diehm N, Dick F, Schaffner T, Schmidli J, Kalka C, Di Santo S, Voelzmann J, Baumgartner I. Novel insight into the pathobiology of abdominal aortic aneurysm and potential future treatment concepts. *Progress in cardiovascular diseases.* 2007;50:209-217

146. Ailawadi G, Knipp BS, Lu G, Roelofs KJ, Ford JW, Hannawa KK, Bishop K, Thanaporn P, Henke PK, Stanley JC, Upchurch GR, Jr. A nonintrinsic regional basis for increased infrarenal aortic mmp-9 expression and activity. *J Vasc Surg.* 2003;37:1059-1066
147. Zhang S, Zhong B, Chen M, Yang L, Yang G, Li Y, Wang H, Wang G, Li W, Cui J, Hoffman AR, Hu J. Epigenetic reprogramming reverses the malignant epigenotype of the mmp/timp axis genes in tumor cells. *International journal of cancer. Journal international du cancer.* 2014;134:1583-1594
148. Stenvinkel P, Karimi M, Johansson S, Axelsson J, Suliman M, Lindholm B, Heimbürger O, Barany P, Alvestrand A, Nordfors L, Qureshi AR, Ekstrom TJ, Schalling M. Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease? *Journal of internal medicine.* 2007;261:488-499
149. Fuggle NR, Howe FA, Allen RL, Sofat N. New insights into the impact of neuro-inflammation in rheumatoid arthritis. *Frontiers in neuroscience.* 2014;8:357
150. Pereira IT, Ramos EA, Costa ET, Camargo AA, Manica GC, Klassen LM, Chequin A, Braun-Prado K, Pedrosa Fde O, Souza EM, Costa FF, Klassen G. Fibronectin affects transient mmp2 gene expression through DNA demethylation changes in non-invasive breast cancer cell lines. *PloS one.* 2014;9:e105806
151. Yuan C, Zhang L, Gao Y, Peng D, Liu J, Cai Y. [DNA demethylation at the promoter region enhances the expression of mmp-9 in ectopic endometrial stromal cells of endometriosis]. *Xi bao yu fen zi mian yi xue za zhi = Chinese journal of cellular and molecular immunology.* 2014;30:1258-1261
152. Liu Y, Aryee MJ, Padyukov L, Fallin MD, Hesselberg E, Runarsson A, Reinius L, Acevedo N, Taub M, Ronninger M, Shchetynsky K, Scheynius A, Kere J, Alfredsson L, Klareskog L, Ekstrom TJ, Feinberg AP. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature biotechnology.* 2013;31:142-147

153. Cutolo M, Paolino S, Pizzorni C. Possible contribution of chronic inflammation in the induction of cancer in rheumatic diseases. *Clinical and experimental rheumatology*. 2014;32:839-847
154. Liyanage VR, Jarmasz JS, Murugesan N, Del Bigio MR, Rastegar M, Davie JR. DNA modifications: Function and applications in normal and disease states. *Biology (Basel)*. 2014;3:670-723
155. Kim M, Long TI, Arakawa K, Wang R, Yu MC, Laird PW. DNA methylation as a biomarker for cardiovascular disease risk. *PloS one*. 2010;5:e9692
156. Gunawardhana LP, Gibson PG, Simpson JL, Benton MC, Lea RA, Baines KJ. Characteristic DNA methylation profiles in peripheral blood monocytes are associated with inflammatory phenotypes of asthma. *Epigenetics : official journal of the DNA Methylation Society*. 2014;9:1302-1316
157. Nordon IM, Hinchliffe RJ, Loftus IM, Thompson MM. Pathophysiology and epidemiology of abdominal aortic aneurysms. *Nature reviews. Cardiology*. 2011;8:92-102
158. Wan ES, Qiu W, Baccarelli A, Carey VJ, Bacherman H, Rennard SI, Agusti A, Anderson W, Lomas DA, Demeo DL. Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome. *Human molecular genetics*. 2012;21:3073-3082
159. Lee KW, Pausova Z. Cigarette smoking and DNA methylation. *Frontiers in genetics*. 2013;4:132
160. Horvath S. DNA methylation age of human tissues and cell types. *Genome biology*. 2013;14:R115
161. Zykovich A, Hubbard A, Flynn JM, Tarnopolsky M, Fraga MF, Kerksick C, Ogborn D, MacNeil L, Mooney SD, Melov S. Genome-wide DNA methylation changes with age in disease-free human skeletal muscle. *Aging Cell*. 2014;13:360-366
162. Johnson AA, Akman K, Calimport SR, Wuttke D, Stolzing A, de Magalhaes JP. The role of DNA methylation in aging, rejuvenation, and age-related disease. *Rejuvenation research*. 2012;15:483-494
163. Guay SP, Legare C, Houde AA, Mathieu P, Bosse Y, Bouchard L. Acetylsalicylic acid, aging and coronary artery disease are

- associated with abca1 DNA methylation in men. *Clinical epigenetics*. 2014;6:14
164. Krishna SM, Dear A, Craig JM, Norman PE, Golledge J. The potential role of homocysteine mediated DNA methylation and associated epigenetic changes in abdominal aortic aneurysm formation. *Atherosclerosis*. 2013;228:295-305
  165. Mandaviya PR, Stolk L, Heil SG. Homocysteine and DNA methylation: A review of animal and human literature. *Molecular genetics and metabolism*. 2014;113:243-252
  166. Antoniadis C, Antonopoulos AS, Tousoulis D, Marinou K, Stefanadis C. Homocysteine and coronary atherosclerosis: From folate fortification to the recent clinical trials. *European heart journal*. 2009;30:6-15
  167. Biselli PM, Sanches de Alvarenga MP, Abbud-Filho M, Ferreira-Baptista MA, Galbiatti AL, Goto MT, Cardoso MA, Eberlin MN, Haddad R, Goloni-Bertollo EM, Pavarino-Bertelli EC. Effect of folate, vitamin b6, and vitamin b12 intake and mthfr c677t polymorphism on homocysteine concentrations of renal transplant recipients. *Transplantation proceedings*. 2007;39:3163-3165
  168. Saratzis A, Bown MJ, Wild B, Nightingale P, Smith J, Johnson C, Melas N, Kitas GD. Association between seven single nucleotide polymorphisms involved in inflammation and proteolysis and abdominal aortic aneurysm. *J Vasc Surg*. 2014
  169. Cao H, Hu X, Zhang Q, Li J, Wang J, Shao Y, Liu B, Xin S. Homocysteine level and risk of abdominal aortic aneurysm: A meta-analysis. *PloS one*. 2014;9:e85831
  170. Peeters AC, van Landeghem BA, Graafsma SJ, Kranendonk SE, Hermus AR, Blom HJ, den Heijer M. Low vitamin b6, and not plasma homocysteine concentration, as risk factor for abdominal aortic aneurysm: A retrospective case-control study. *J Vasc Surg*. 2007;45:701-705
  171. Liu Z, Luo H, Zhang L, Huang Y, Liu B, Ma K, Feng J, Xie J, Zheng J, Hu J, Zhan S, Zhu Y, Xu Q, Kong W, Wang X. Hyperhomocysteinemia exaggerates adventitial inflammation and angiotensin ii-induced abdominal aortic aneurysm in mice. *Circulation research*. 2012;111:1261-1273

172. Lindqvist M, Hellstrom A, Henriksson AE. Abdominal aortic aneurysm and the association with serum levels of homocysteine, vitamins b6, b12 and folate. *American journal of cardiovascular disease*. 2012;2:318-322
173. Alcazar O, Achberger S, Aldrich W, Hu Z, Negrotto S, Sauntharajah Y, Triozzi P. Epigenetic regulation by decitabine of melanoma differentiation in vitro and in vivo. *International journal of cancer. Journal international du cancer*. 2012;131:18-29
174. Chouliaras L, van den Hove DL, Kenis G, Draanen M, Hof PR, van Os J, Steinbusch HW, Schmitz C, Rutten BP. Histone deacetylase 2 in the mouse hippocampus: Attenuation of age-related increase by caloric restriction. *Current Alzheimer research*. 2013;10:868-876
175. Martin SL, Hardy TM, Tollefsbol TO. Medicinal chemistry of the epigenetic diet and caloric restriction. *Current medicinal chemistry*. 2013;20:4050-4059
176. Shukeir N, Stefanska B, Parashar S, Chik F, Arakelian A, Szyf M, Rabbani SA. Pharmacological methyl group donors block skeletal metastasis in vitro and in vivo. *British journal of pharmacology*. 2015
177. Patterson K, Molloy L, Qu W, Clark S. DNA methylation: Bisulphite modification and analysis. *Journal of visualized experiments : JoVE*. 2011:3170
178. Kurdyukov S, Bullock M. DNA methylation analysis: Choosing the right method. *Biology*. 2016;5:3
179. Eckert KA, Kunkel TA. DNA polymerase fidelity and the polymerase chain reaction. *PCR methods and applications*. 1991;1:17-24
180. Lee PY, Costumbrado J, Hsu CY, Kim YH. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of visualized experiments : JoVE*. 2012
181. Lou S, Lee HM, Qin H, Li JW, Gao Z, Liu X, Chan LL, Kl Lam V, So WY, Wang Y, Lok S, Wang J, Ma RC, Tsui SK, Chan JC, Chan TF, Yip KY. Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter

- and gene body methylation in transcriptional regulation. *Genome biology*. 2014;15
182. Mueller O, Hahnenberger K, Dittmann M, Yee H, Dubrow R, Nagle R, Ilsley D. A microfluidic system for high-speed reproducible DNA sizing and quantitation. *Electrophoresis*. 2000;21:128-134
  183. Paegel BM, Blazej RG, Mathies RA. Microfluidic devices for DNA sequencing: Sample preparation and electrophoretic analysis. *Curr Opin Biotechnol*. 2003;14:42-50
  184. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics (Oxford, England)*. 2014;30:2114-2120
  185. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and samtools. *Bioinformatics (Oxford, England)*. 2009;25:2078-2079
  186. Beasley TM, Erickson S, Allison DB. Rank-based inverse normal transformations are increasingly used, but are they merited? *Behavior genetics*. 2009;39:580-595
  187. Khoury LE, Posthumus M, Collins M, van der Merwe W, Handley C, Cook J, Raleigh SM. Eln and fbn2 gene variants as risk factors for two sports-related musculoskeletal injuries. *International journal of sports medicine*. 2015;36:333-337
  188. Wong ML, Medrano JF. Real-time pcr for mrna quantitation. *BioTechniques*. 2005;39:75-85
  189. Pushpakumar S, Kundu S, Narayanan N, Sen U. DNA hypermethylation in hyperhomocysteinemia contributes to abnormal extracellular matrix metabolism in the kidney. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*. 2015;29:4713-4725
  190. Li LC, Dahiya R. Methprimer: Designing primers for methylation pcers. *Bioinformatics (Oxford, England)*. 2002;18:1427-1431
  191. Patterson K, Molloy L, Qu W, Clark S. DNA methylation: Bisulphite modification and analysis. *Journal of visualized experiments : JoVE*. 2011
  192. Adalsteinsson BT, Gudnason H, Aspelund T, Harris TB, Launer LJ, Eiriksdottir G, Smith AV, Gudnason V. Heterogeneity in

- white blood cells has potential to confound DNA methylation measurements. *PloS one*. 2012;7:e46705
193. Wang P, Shen C, Diao L, Yang Z, Fan F, Wang C, Liu X, Sun X, Dong Z, Zhu H, Ma X, Cao Q, Zhao X, Ma D, Zou Y, Hu K, Sun A, Ge J. Aberrant hypermethylation of aldehyde dehydrogenase 2 promoter upstream sequence in rats with experimental myocardial infarction. *BioMed research international*. 2015;2015:503692
  194. Birdsey Graeme M, Shah Aarti V, Dufton N, Reynolds Louise E, Osuna Almagro L, Yang Y, Aspalter Irene M, Khan Samia T, Mason Justin C, Dejana E, Göttgens B, Hovalala-Dilke K, Gerhardt H, Adams Ralf H, Randi Anna M. The endothelial transcription factor erg promotes vascular stability and growth through wnt/ $\beta$ -catenin signaling. *Developmental Cell*. 2015;32:82-96
  195. Sajjad A, Novoyatleva T, Vergarajauregui S, Troidl C, Schermuly RT, Tucker HO, Engel FB. Lysine methyltransferase smyd2 suppresses p53-dependent cardiomyocyte apoptosis. *Biochimica et biophysica acta*. 2014;1843:2556-2562
  196. Hendel A, Cooper D, Abraham T, Zhao H, Allard MF, Granville DJ. Proteinase inhibitor 9 is reduced in human atherosclerotic lesion development. *Cardiovascular pathology : the official journal of the Society for Cardiovascular Pathology*. 2012;21:28-38
  197. Robinson CM, Neary R, Levendale A, Watson CJ, Baugh JA. Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with thy-1 promoter methylation and the development of a pro-fibrotic phenotype. *Respiratory research*. 2012;13:74
  198. Lu Y, Chu A, Turker MS, Glazer PM. Hypoxia-induced epigenetic regulation and silencing of the brca1 promoter. *Molecular and cellular biology*. 2011;31:3339-3350
  199. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, Ghesquiere B, Van Dyck L, Boeckx B, Schoonjans L, Hermans E, Amant F, Kristensen VN, Koh KP, Mazzone M, Coleman ML, Carell T, Carmeliet P, Lambrechts D. Tumour

- hypoxia causes DNA hypermethylation by reducing tet activity. *Nature*. 2016;537:63-68
200. Bestor TH. The DNA methyltransferases of mammals. *Human molecular genetics*. 2000;9:2395-2402
201. Esteller M. CpG island hypermethylation and tumor suppressor genes: A booming present, a brighter future. *Oncogene*. 2002;21
202. Ghosh S. Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues. *Epigenetics: official journal of the DNA Methylation Society*. 2010;5
203. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MSI, Kawaji H, Lassmann T, Harbers M, Forrest AR, Bajic VB. Effects of cytosine methylation on transcription factor binding sites. *BMC Genomics*. 2014;15:119
204. Baylin SB. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol*. 2005;2 Suppl 1:S4-11
205. Bird A. DNA methylation patterns and epigenetic memory. *Genes & development*. 2002;16
206. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature*. 2000;403
207. Patel KM, Strong A, Tohyama J, Jin X, Morales CR, Billheimer J, Millar J, Kruth H, Rader DJ. Macrophage sortilin promotes ldl uptake, foam cell formation, and atherosclerosis. *Circulation research*. 2015;116:789-796
208. Aslibekyan S, Claas SA, Arnett DK. Clinical applications of epigenetics in cardiovascular disease: The long road ahead. *Translational research : the journal of laboratory and clinical medicine*. 2014
209. Saksouk N, Simboeck E, Dejardin J. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics Chromatin*. 2015;8:3
210. Vaissiere T, Sawan C, Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. *Mutation research*. 2008;659:40-48
211. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, Valsesia A, Ye Z, Kuan S, Edsall LE, Camargo AA, Stevenson BJ, Ecker JR, Bafna V, Strausberg RL, Simpson AJ,

- Ren B. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* 2012;22
212. Liu Z, Liu S, Xie Z, Blum W, Perrotti D, Paschka P, Klisovic R, Byrd J, Chan KK, Marcucci G. Characterization of in vitro and in vivo hypomethylating effects of decitabine in acute myeloid leukemia by a rapid, specific and sensitive lc-ms/ms method. *Nucleic acids research.* 2007;35:e31
213. Zelic R, Fiano V, Grasso C, Zugna D, Pettersson A, Gillio-Tos A, Merletti F, Richiardi L. Global DNA hypomethylation in prostate cancer development and progression: A systematic review. *Prostate Cancer Prostatic Dis.* 2015;18:1-12
214. Winfield J, Esbitt A, Seutter SF, Desai B, Abdo M, Vasconez M, Laidlaw W, Green K, Shamseddin SM, Borghaei RC. Effect of inflammatory cytokines on DNA methylation and demethylation. *The FASEB Journal.* 2016;30:1053.1053-1053.1053
215. Liao M, Xu J, Clair AJ, Ehrman B, Graham LM, Eagleton MJ. Local and systemic alterations in signal transducers and activators of transcription (stat) associated with human abdominal aortic aneurysms. *The Journal of surgical research.* 2012;176:321-328
216. Swerdlow DI, Holmes MV, Kuchenbaecker KB, Engmann JE, Shah T, Sofat R, Guo Y, Chung C, Peasey A, Pfister R, Mooijaart SP, Ireland HA, Leusink M, Langenberg C, Li KW, Palmén J, Howard P, Cooper JA, Drenos F, Hardy J, Nalls MA, Li YR, Lowe G, Stewart M, Bielinski SJ, Peto J, Timpson NJ, Gallacher J, Dunlop M, Houlston R, Tomlinson I, Tzoulaki I, Luan J, Boer JM, Forouhi NG, Onland-Moret NC, van der Schouw YT, Schnabel RB, Hubacek JA, Kubinova R, Baceviciene M, Tamosiunas A, Pajak A, Topor-Madry R, Malyutina S, Baldassarre D, Sennblad B, Tremoli E, de Faire U, Ferrucci L, Bandenelli S, Tanaka T, Meschia JF, Singleton A, Navis G, Mateo Leach I, Bakker SJ, Gansevoort RT, Ford I, Epstein SE, Burnett MS, Devaney JM, Jukema JW, Westendorp RG, Jan de Borst G, van der Graaf Y, de Jong PA, Mailand-van der Zee AH, Klungel OH, de Boer A, Doevendans PA, Stephens JW, Eaton CB, Robinson JG, Manson JE, Fowkes FG, Frayling TM, Price JF, Whincup PH, Morris RW, Lawlor DA, Smith GD, Ben-

- Shlomo Y, Redline S, Lange LA, Kumari M, Wareham NJ, Verschuren WM, Benjamin EJ, Whittaker JC, Hamsten A, Dudbridge F, Delaney JA, Wong A, Kuh D, Hardy R, Castillo BA, Connolly JJ, van der Harst P, Brunner EJ, Marmot MG, Wassel CL, Humphries SE, Talmud PJ, Kivimaki M, Asselbergs FW, Voevoda M, Bobak M, Pikhart H, Wilson JG, Hakonarson H, Reiner AP, Keating BJ, Sattar N, Hingorani AD, Casas JP. The interleukin-6 receptor as a target for prevention of coronary heart disease: A mendelian randomisation analysis. *Lancet (London, England)*. 2012;379:1214-1224
217. Lindeman JH, Abdul-Hussien H, Schaapherder AF, Van Bockel JH, Von der Thusen JH, Roelen DL, Kleemann R. Enhanced expression and activation of pro-inflammatory transcription factors distinguish aneurysmal from atherosclerotic aorta: Il-6- and il-8-dominated inflammatory responses prevail in the human aneurysm. *Clinical science (London, England : 1979)*. 2008;114:687-697
218. Ferreira RC, Freitag DF, Cutler AJ, Howson JMM, Rainbow DB, Smyth DJ, Kaptoge S, Clarke P, Boreham C, Coulson RM, Pekalski ML, Chen W-M, Onengut-Gumuscu S, Rich SS, Butterworth AS, Malarstig A, Danesh J, Todd JA. Functional il6r358ala allele impairs classical il-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genetics*. 2013;9:e1003444
219. Martinelli N, Girelli D, Lunghi B, Pinotti M, Marchetti G, Malerba G, Pignatti PF, Corrocher R, Olivieri O, Bernardi F. Polymorphisms at Ildr locus may be associated with coronary artery disease through modulation of coagulation factor viii activity and independently from lipid profile. *Blood*. 2010;116:5688-5697
220. Xu G, Liu G, Xiong S, Liu H, Chen X, Zheng B. The histone methyltransferase smyd2 is a negative regulator of macrophage activation by suppressing interleukin 6 (il-6) and tumor necrosis factor alpha (tnf-alpha) production. *J Biol Chem*. 2015;290:5414-5423
221. Maurano Matthew T, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos John A. Role of DNA methylation in

- modulating transcription factor occupancy. *Cell reports*.12:1184-1195
222. Banati F, Koroknai A, Salamon D, Takacs M, Minarovits-Kormuta S, Wolf H, Niller HH, Minarovits J. CpG-methylation silences the activity of the rna polymerase iii transcribed eber-1 promoter of epstein-barr virus. *FEBS Lett*. 2008;582:705-709
223. Park JL, Lee YS, Song MJ, Hong SH, Ahn JH, Seo EH, Shin SP, Lee SJ, Johnson BH, Stampfer MR, Kim HP, Kim SY, Lee YS. Epigenetic regulation of rna polymerase iii transcription in early breast tumorigenesis. *Oncogene*. 2017;36:6793-6804
224. Jones PA. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*. 2012;13:484
225. Palacios D, Summerbell D, Rigby PW, Boyes J. Interplay between DNA methylation and transcription factor availability: Implications for developmental activation of the mouse myogenin gene. *Molecular and cellular biology*. 2010;30:3805-3815
226. Marchal C, Miotto B. Emerging concept in DNA methylation: Role of transcription factors in shaping DNA methylation patterns. *Journal of cellular physiology*. 2015;230:743-751
227. Bahar Halpern K, Vana T, Walker MD. Paradoxical role of DNA methylation in activation of foxa2 gene expression during endoderm development. *J Biol Chem*. 2014;289:23882-23892
228. Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, Fitzgerald P, Vinson C. CpG methylation of half-cre sequences creates c/ebpalpha binding sites that activate some tissue-specific genes. *Proc Natl Acad Sci U S A*. 2010;107:20311-20316
229. Vojta A, Dobrinic P, Tadic V, Bockor L, Korac P, Julg B, Klasic M, Zoldos V. Repurposing the crispr-cas9 system for targeted DNA methylation. *Nucleic acids research*. 2016;44:5615-5628
230. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic acids research*. 2005;33:5868-5877
231. Toghiani BJ, Saratzis A, Bown MJ. Abdominal aortic aneurysm—an independent disease to atherosclerosis? *Cardiovascular Pathology*. 2017;27:71-75