Epigenetic Approaches to the Study of Macrophages in Atherosclerosis

Michael Edward Reschen

St Cross College

Nuffield Department of Medicine

University of Oxford

A thesis submitted for

THE DEGREE OF DOCTOR OF PHILOSOPHY (Medical Sciences)

Trinity term 2015

Abstract

Epigenetic Approaches to the Study of Macrophages in Atherosclerosis

Michael Edward Reschen

St Cross College

THE DEGREE OF DOCTOR OF PHILOSOPHY

(Medical Sciences)

Trinity 2015

Coronary artery disease (CAD) is caused by atherosclerosis, a chronic inflammatory response to modified lipoproteins. A key pathophysiological event is the lipid-induced transformation of macrophages into lipid-laden foam cells and their accumulation in atherosclerotic plaques. Heritable CAD risk is associated with common genetic variants at over 40 genomic loci; the underlying causal mechanisms remain largely unknown and could affect transcriptional regulation in foam cells. Epigenetic and gene expression changes were measured in primary human macrophages before and after exposure to atherogenic, oxidized low-density lipoprotein—with resultant foam cell formation. This unbiased approach involved open chromatin mapping with formaldehyde-assisted isolation of regulatory elements with enhancer and transcription factor mapping using chromatin immuno-precipitation. Foam cell formation was associated with changes in a subset of open chromatin and enhancer sites that were strongly correlated with expression of nearby genes. OxLDL-regulated

enhancers were enriched for several transcription factors—including C/EBP-beta—that have no previously documented role in foam cell formation. OxLDL exposure up-regulated C/EBP-beta expression and increased C/EBP-beta binding across the genome, most prominently around genes involved in inflammatory response pathways. Variants at CAD-associated loci were enriched in the subset of oxLDL-regulated open chromatin sites. These included rs72664324 in an oxLDL-induced super-enhancer at the PPAP2B locus. OxLDL increased C/EBP-beta binding at rs72664324. C/EBP-beta binding, enhancer activity and oxLDL-induced upregulation of PPAP2B were stronger with the protective A allele of rs72664324. The PPAP2B protein product LPP3 was expressed in foam cells in human atherosclerotic plaques and was upregulated by oxLDL exposure in macrophages, so increasing the degradation of pro-inflammatory mediators. I also found several other CAD risk candidate genes were regulated by oxLDL: Phosphatase and actin regulator 1 (PHACTR1) and macrophage inducible Ca^{2+} dependent C-type lectin (Mincle). This led us to find a novel expression-quantitative-trait locus for PHACTR1 in macrophages and define new glycolipid ligands for Mincle. Our results demonstrate a genetic mechanism contributing to CAD risk at the PPAP2B locus and highlight the value of integrating gene expression and epigenetic changes to study disease processes involving pathogenic environmental stimuli.

I Table of Contents

I Table of Contents 4

II List of Tables 10

III List of Figures 12

IV Declaration of Authorship 17

V Acknowledgements 18

VI Abbreviations 19

1 General Introduction 21

1.1 Coronary artery disease 21

1.2 Epidemiological Importance of CAD 21

1.3 Pathogenesis of CAD 22

1.3.1 Macrophages and foam cells in atherosclerosis 24

1.4 Risk factors for CAD 32

1.4.1 Environmental Risk factors 33

1.4.2 Genetic risk of CAD 37

1.4.3 GWAS and Disease Mechanisms 40

1.4.4 Interplay between genetic and environmental factors 43

1.5 Conclusions and present study aims 44

2 Materials and Methods 49

2.1 Ethical Approval 49

2.2 General Methods 49

2.2.1 Reagents 49
2.2.2 Buffers and mediums 49
2.2.3 PCR 49
2.2.4 Cloning 50
2.3 Cell culture 51
2.4 Preparation of oxLDL 52
2.5 FAIRE 53
2.6 ChIP 53
2.7 Library preparation and sequencing 54
2.8 FAIRE and ChIP-seq analysis 55
2.9 Data Availability 56
2.10 Annotation of FAIRE and CHIP-seq sites 56
2.11 Display in genome browser 57
2.12 Transcription factor analysis 57
2.13 SNP enrichment analysis 58
2.14 Non-coding element enrichment 59
2.15 Gene expression 59
2.15.1 RNA extraction 59
2.15.2 Microarrays 60
2.15.3 RTqPCR 60
2.15.4 eQTL 62
2.16 Genotyping 63
2.17 Electrophoretic mobility shift assays (EMSA) 64
2.18 Luciferase reporter assays 65
2.18.1 Enhancers 65
2.18.2 NFAT7 Mincle Reporter Assay 67
2.19 Western blotting 68

2.20 Measurement of LPP3 phosphatase activity 69

2.21 Measurement of LPP3 substrates and products 70

2.22 Proteomic quantitation of LPP3 70

2.23 Immunohistochemistry 71

2.24 5' RACE 72

2.25 PHACTR1 cloning 72

2.26 Tissue bank, primary cell and cell line panel expression analysis 73

2.27 Confocal microscopy 74

3 Macrophage-derived Foam Cell Formation 75

3.1.1 Introduction 75

3.1.2 OxLDL-induced *in vitro* foam cell formation 76

3.1.3 Whole-genome approaches to gene expression in foam cells 78

3.2 Results 81

3.2.1 Foam cell formation 81

3.2.2 OxLDL regulated genes and CAD genetic risk 96

3.3 Discussion 98

4 Epigenomic Effects of Atherogenic Lipid on Macrophages 102

4.1 Introduction 102

4.1.1 Epigenomics and Atherosclerosis 103

4.1.2 Integrating Epigenomics with Genome-wide association studies 105

4.1.3 Epigenomic techniques 106

4.2 Results 109

4.2.1 OxLDL induces major changes in chromatin accessibility in macrophages 109

4.2.2 Macrophage enhancers undergo major oxLDL-induced changes 123

4.2.3 Dynamic enhancer sites identify C/EBP-beta as a novel oxLDL-regulated transcription factor in macrophages 130

4.2.4 CAD-associated genetic variants are enriched in regulatory DNA sites altered by oxLDL 135

4.3 Discussion 140

5 PPAP2B: A coronary artery disease risk variant influencing the macrophage response to lipid 145

5.1 Introduction 145

5.1.1 GWAS and PPAP2B 145

5.1.2 PPAP2B 146

5.1.3 Role of LPP3 in vascular function 148

5.1.4 Role of LPP3 enzymatic activity in atherosclerosis 149

5.1.5 PPAP2B expression and CAD risk genotype 151

5.2 Results 152

5.2.1 LPP3 expression in atherosclerosis and arteritis 152

5.2.2 OxLDL-induced upregulation of PPAP2B/LPP3 and enzymatic activity in macrophages 154

5.2.3 CAD-associated variant rs72664324 demonstrates allelic differences in oxLDL-induced enhancer activity and PPAP2B expression 158

5.2.4 CAD SNPs at the PPAP2B locus and transcriptional regulation in endothelial cells 169

5.3 Discussion 177

6 Regulation by genetic and atherogenic stimuli of coronary artery risk gene, phosphatase and actin regulating protein 1 (PHACTR1) 181

6.1 Introduction 181

6.1.1 Genome-wide association studies and the PHACTR1 genetic risk locus 181

6.1.2 Mechanism of genetic risk for CAD at the PHACTR1 locus 182

6.1.3 CAD risk SNP association with PHACTR1 expression 183

6.1.4 PHACTR1 gene and protein 183
6.1.5 PHACTR1 actin binding 184
6.1.6 PHACTR1 and Protein Phosphatase 1 Binding 184
6.1.7 Expression pattern and subcellular localization of PHACTR1 185
6.1.8 Function of PHACTR1 187
6.2 Results 188
6.2.1 *PHACTR1* expression is increased in human atherosclerotic lesions 188
6.2.2 *PHACTR1* is expressed as several transcripts from two separate promoters in vascular and immune cells. 190
6.2.3 PHACTR1 transcripts display tissue-specific differential expression 196
6.2.4 PHACTR1 protein is expressed in primary human macrophages and endothelial cells 200
6.2.5 PHACTR1 isoforms differ in their subcellular distribution 203
6.2.6 PHACTR1 expression is regulated by oxLDL and inflammatory stimuli in macrophages and endothelial cells 209
6.2.7 Genetic regulation of PHACTR1 expression by CAD SNP rs9349379 214
6.3 Discussion 220
7 Mincle: A receptor for lipid species on macrophages 227
7.1 Introduction 227
7.2 Results 231
7.2.1 Immunohistochemistry of Mincle 231
7.2.2 Mincle expression in macrophages is upregulated by oxLDL 233
7.2.3 Genomic landscape at the Mincle locus 233
7.2.4 NFAT reporter assay response to oxLDL 235
7.2.5 Human and mouse Mincle are receptors for several MTB strains 238
7.2.6 Signaling through MINCLE is triggered by an extensive repertoire of trehalose dimycolate related compounds 242

7.3 Discussion 252

7.3.1 Regulation of Mincle by C/EBP beta in human macrophages and foam cells 254

7.3.2 Conclusions 256

8 General Discussion 257

8.1 Lipid-induced epigenomic changes in macrophages 258

8.2 Integrating the macrophage-foam cell epigenome with CAD heritable risk 260

8.3 Mincle 263

8.4 Conclusions 264

9 Appendix 265

9.1 Published work 265

9.2 Example Scripts 268

9.2.1 Heat map of intersections between different samples of a genomic feature 268

9.2.2 Heat map of two different genomic feature sets 270

9.2.3 Correlating differential gene expression with dynamic genomic features. 272

9.2.4 Correlating gene expression with super-enhancer status 274

9.3 Appendix of transcription factor motif enrichment in dynamic chromatin sites. 282

9.4 PHACTR1 long and intermediate transcript alignment 295

10 Bibliography 298

II List of Tables

Table 1 CAD risk variants identified by GWAS studies of CAD phenotypes. 44

Table 2 Microarray studies of gene expression in macrophages and foam cells. 79

Table 3 Differentially expressed genes in foam cells compared to macrophages. 87

Table 4 Comparison of gene expression in macrophages and foam cells by microarray and RTqPCR for 6 different genes.....90

Table 5 Gene ontology enrichment for genes that are upregulated by oxLDL using DAVID.....94

Table 6 Gene ontology enrichment for genes that are down-regulated by oxLDL using DAVID.....95

Table 7 Differentially expressed genes within a 50kb radius of CAD loci. 97

Table 8 Top 10 enriched transcription factor binding site motifs in top 3000 (ranked by average signal intensity) dynamic open chromatin sites.....120

Table 9 Enrichment of transcription factor motifs in dynamic chromatin/enhancer sites. 131

Table 10 Top 10 most enriched transcription factor motifs in the top 3000 dynamic C/EBP-beta binding sites (ranked by average signal strength).....135

Table 11 Number of CAD loci and individual SNPs overlapping regulatory elements in macrophages and foam cells. 137

Table 12 Top 10 transcription factor motifs with the greatest difference in affinity at the rs72664324 site between the risk G and protective A allele.....161

Table 13 Transcription factor motif affinities for rs72664324 analyzed separately for each alleles. 162

Table 14 Comparison of transcription factor binding affinities to the A and G alleles of rs17114036. 170

Table 15 Comparison of transcription factor binding affinities to the A and C alleles of rs56348932. 171

Table 16 Summary of mRNA expression pattern of PHACTR1 transcripts. 198

Table 17 Chemical formula and calculated molecular weights of compounds tested in Mincle luciferase assays. 250

III List of Figures

Figure 3-1 Isolation of LDL from human plasma. 82

Figure 3-2 Light microscopy of foam cell formation. 83

Figure 3-3 Fluorescence microscopy to assess cell viability 84

Figure 3-4 Effect of oxLDL-induced foam cell formation on gene expression in human macrophages 89

Figure 3-5 Gene ontology enrichment for biological processes for genes that are upregulated by oxLDL. 92

Figure 3-6 Gene ontology enrichment for biological processes for genes that are down-regulated by oxLDL. 93

Figure 4-1 Genomic distribution of open chromatin sites in macrophages and foam cells. 110

Figure 4-2 Comparison of overlapping open chromatin sites in different cell types. 112

Figure 4-3 Proportion of overlap between different types of open chromatin sites 113

Figure 4-4 OxLDL-induced changes in chromatin structure determined by FAIRE-seq at gene transcription start sites. 114

Figure 4-5 FAIRE-seq profile at TSS for different gene sets. 115

Figure 4-6 Representative chromatin profile at the *CD36* gene locus showing a dynamic chromatin cluster. 116

Figure 4-7 Chromatin signal at dynamic chromatin sites. 117

Figure 4-8 Width of dynamic chromatin sites (bp). 118

Figure 4-9 Dynamic chromatin sites are enriched for non-coding regulatory elements found in 9 other cell types. 119
Figure 4-10 Dynamic chromatin clusters are enriched for gene sets involved in foam cell formation and lipid handling. 120
Figure 4-11 Correlation between change in gene expression and change in FAIRE-seq signal at dynamic open chromatin sites. 122
Figure 4-12 Characterization of the dynamic chromatin and enhancer landscape in macrophages and foam cells. 124
Figure 4-13 OxLDL regulated enhancer activity is correlated with gene expression. 125
Figure 4-14 Super-enhancers in macrophages and foam cells 126
Figure 4-15 Comparison of super-enhancer in macrophages and foam cells with 86 other cell types and tissues. 128
Figure 4-16 OxLDL induces changes in both open chromatin and enhancer status at a subset of sites with important regulatory function. 130
Figure 4-17 C/EBP-beta binding is responsive to oxLDL. 134
Figure 4-18 Regional plot of linkage disquilibrium relating to CAD risk SNP rs17114036 at the PPAP2B gene locus. 136
Figure 4-19 CAD loci are enriched in oxLDL-regulated open chromatin sites. 138
Figure 4-20 Up-regulation of IL6R and PPAP2B in foam cells compared to macrophages as assessed by quantitative RT-PCR. 139
Figure 5-1 Immunohistochemistry for LPP3 in human vascular tissue and tonsil. 153
Figure 5-2 Western blotting demonstrates up-regulation of glycosylated and non-glycosylated LPP3 protein in macrophage-derived foam cells 155

Figure 5-3. The enzymatic activity of LPP3 per amount of cellular protein is significantly increased by oxLDL stimulation. 156
Figure 5-4 OxLDL alters the levels of LPP3 substrates and products in macrophages. 158
Figure 5-5 Chromatin profile at the PPAP2B locus where rs72664324 is in a dynamic chromatin, enhancer and C/EBP beta site (red box). 160
Figure 5-6 EMSA for rs72664324 alleles with cold probe competition assays. 164
Figure 5-7 Effect of rs72664324 alleles on transcription factor binding. 165
Figure 5-8 EMSA showing C/EBP-alpha does not bind rs72664324 166
Figure 5-9 CAD risk SNP rs72664324 at the PPAP2B locus regulates enhancer activity and oxLDL-induced expression of PPAP2B. 168
Figure 5-10 EMSA showing endothelial cell nuclear protein binding to the A and G allele of rs17114036. 172
Figure 5-11 EMSA showing the effect of competition assays on endothelial nuclear protein binding to the rs17114036 SNP alleles 173
Figure 5-12 EMSA showing endothelial cell nuclear protein binding to the A and C alleles of rs56348932. 174
Figure 5-13 CAD SNPs at the PPAP2B locus affect transcriptional regulation in endothelial cells. 175
Figure 6-1. PHACTR1 expressing cells are increased in atherosclerotic vessels and in inflammatory arteritis 189
Figure 6-2 The human PHACTR1 gene expresses multiple transcripts from alternative transcription start sites. 193
Figure 6-3 Aligned RACE clone sequences for PHACTR1 transcription start site. . 195

Figure 6-4 Genomic annotations supportive of transcriptional activity at the start site of the ENSEMBL predicted short PHACTR1 coding transcript. 196
Figure 6-5 PHACTR1 transcript expression patterns in primary cells, tissues and cell lines. 199
Figure 6-6 Region of PHACTR1 gene used to raise raise commercial rabbit polyclonal Anti-PHACTR1 antibody. 200
Figure 6-7 PHACTR1 protein consistent with the intermediate transcript is expressed in macrophages and endothelial cells but not vascular smooth muscle cells. ... 202
Figure 6-8 The subcellular localization of PHACTR1 isoforms and response to serum, cytochalasin D and latrunculin B. 208
Figure 6-9 PHACTR1 transcripts are differentially regulated by atherogenic and inflammatory stimuli. 211
Figure 6-10 The relationship between expression of the short and intermediate PHACTR1 transcripts. 213
Figure 6-11 Genotype of CAD risk SNP rs9349379 affects expression of the short PHACTR1 transcript in macrophages. 216
Figure 6-12 Electrophoretic mobility shift assay for macrophage nuclear protein binding at the rs9349379 SNP shows binding of MEF2 only to the A allele. ... 217
Figure 6-13 Transcriptional activity at the rs9349379 SNP region in assayed in human THP1 macrophages. 218
Figure 7-1 Immunohistochemistry staining for Mincle in human arteries. 232
Figure 7-2 Epigenomic lanscape at the Mincle gene locus 235
Figure 7-3 Schematic of the Mincle luciferase reporter assay 236
Figure 7-4 Human LDL and oxLDL to not cause Mincle activation. 238
Figure 7-5 Human and mouse Mincle bind MTB 240

Figure 7-6 Several different strains of MTB activate Mincle. 241
Figure 7-7 TDM and its synthetic analogue, TDB, activate Mincle. 245
Figure 7-8 TMM is a Mincle ligand. 246
Figure 7-9 Effect of additional carbon-carbon and sugar head moiety on Mincle binding. 247
Figure 7-10 Monomycolyl glycerol (MMG) is a ligand for Mincle but a Mycobacterial sulpholipid is not. 248
Figure 7-11 Structure of Trehalose-related compounds used to stimulate Mincle 249
Figure 8-1 Model showing the interplay between oxLDL, C/EBP-beta and rs72664324 on expression of PPAP2B. 262

IV Declaration of Authorship

I declare that all this work is my own except for where indicated.

V Acknowledgements

I dedicate this endeavor to Joseph.

I would like to thank my supervisor, Professor Chris O'Callaghan, for his tremendous support and infinite sagacity, keenly shared. I would like to thank Dr Gerton Lunter, Dr Kyle Gaulton, Dr Da Lin and the entire O'Callaghan lab, for their superb assistance, and forbearance.

I would like to thank our collaborators Prof Andrew Morris, Prof Susan Smyth, Dr Elizabeth Soilleux, Prof Del Besra, Prof Fredrik Karpe, and the Oxford Biobank (including their generous volunteers) for their kind assistance.

I thank my parents for providing a decent genetic start in life and positive environmental influences together with their unyielding support and advocacy.

Finally, I thank my nuclear family; Ellie and Joseph—my little angel.
Joseph has spent far too many weekends and evenings without his Daddy, a situation for which I aim to make amends…and Ellie has gone on, never complaining, remaining steadfastly positive—as is her nature—sparkling.

VI Abbreviations

AP1	Activating protein 1
CAD	Coronary artery disease
CD36	Cluster of differentiation 36 (thrombospondin receptor)
C/EBP-beta	CCAT enhancer binding protein beta
CEU	Northern Europeans from Utah
ChIP	Chromatin immuno-precipitation
DAG	Diacylglycerol
EMSA	Electrophoretic mobility shift assay
eQTL	Expression quantitative trait locus
FAIRE	Formaldehyde-assisted isolation of regulatory elements
GWAS	Genome-wide association study
HUVEC	Human umbilical vein endothelial cell
IL4	Interleukin 4
IL6R	Interleukin 6 receptor
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LPA	Lysophosphatidic acid
LPP3	Lipid Phosphohydrolase 3
LPS	Lipopolysaccharide

MAPK	Mitogen-activated protein kinase
M-CSF	Macrophage colony stimulating factor
Mincle	Macrophage inducible Ca^{2+}-dependent C-type lectin
MMP	Matrix metalloproteinase
NFAT7	Nuclear factor of activated T cells 7
NFE2L2	Nuclear factor (erythroid-derived 2)-like 2
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
OxLDL	Oxidized low-density lipoprotein
PHACTR1	Phosphatase and actin regulating protein 1
PMA	Phorbol 12-myristate 13-acetate
Pol2	Polymerase 2
PP1	Protein Phosphatase 1
PPAP2B	Phosphatidic acid phosphatase type 2B
PPARG	Peroxisome Proliferator-Activated Receptor Gamma
PPM	Primary peritoneal macrophage (mouse)
SNP	Single nucleotide polymorphism
TNF-alpha	Tumour necrosis factor alpha
TSS	Transcription start site

1 General Introduction

1.1 Coronary artery disease

Coronary artery disease (CAD) is the leading cause of death worldwide [1]. Inadequate blood supply to the myocardium causes the clinical manifestations of CAD [2]. Myocardial ischemia results in several clinical syndromes including stable angina, unstable angina, myocardial infarction, and ischaemic cardiomyopathy manifesting as heart failure [2]. Death can result from either ischaemia-related arrhythmia or pump failure. Narrowed coronary arteries result in the exertional chest pain characterizing stable angina when myocardial perfusion cannot increase to meet the myocardial oxygen demand [2]. Myocardial infarction occurs when a vessel is suddenly occluded by thrombus formation leading to myocyte death [2]. Current treatments for myocardial infarction focus on rapidly restoring blood flow by a combination of mechanical treatments – percutaneous coronary angioplasty – and drug treatments – anticoagulants including heparin ameliorate further coronary thrombosis [2]. Preventive treatments include anti-platelet agents to inhibit thrombosis and lipid-lowering agents to reduce the formation of lipid-laden plaques in coronary arteries [2].

1.2 Epidemiological Importance of CAD

In the UK, death rates from myocardial infarction have halved since 2002 but there are still 80,000 deaths each year from CAD [3]. CAD also causes substantial morbidity with 405,000 hospitalizations in England in 2010 and cardiovascular disease costs the UK healthcare system about £8.7 billion per year [3]. There are

numerous risk factors for CAD, which are themselves influenced by both environmental and genetic factors. Major risk factors for CAD include low-density lipoprotein (LDL) levels, hypertension, smoking, diabetes, age, gender and heredity [2]. LDL is both a risk factor and a direct mediator of atherosclerosis that accumulates within atherosclerotic plaques. One of the most successful preventive treatments for CAD are statins, which act principally by lowering circulating LDL levels [4, 5].

1.3 Pathogenesis of CAD

The main cause of coronary artery disease is atherosclerosis. The word atherosclerosis is derived from the Greek words ‘athera’ meaning ‘gruel’, and ‘sclerosis’ meaning ‘hardening’. This term aptly describes the fatty plaques that have afflicted humans since at least Egyptian times, as evidenced by atherosclerosis in mummified remains[6]. Atherosclerosis is a chronic process that begins in childhood and can develop to the point of being symptomatic usually in middle to late adulthood[7]. The first step is endothelial activation at sites of sheer stress that are characterized by low or turbulent blood flow[8]. Activated endothelium expresses increased cell adhesion molecules like E-selectin, P-selectin and VCAM1 that contribute to binding and recruitment of circulating immune cells, chiefly monocytes[8]. At sites of endothelial dysfunction there is also enhanced endothelial permeability allowing lipoproteins to bind proteoglycans in the subendothelial space through a charge interaction[9]. According to the ‘response to lipid retention hypothesis’, these lipoproteins become modified in the arterial wall and establish a chronic inflammatory reaction whereby immune cells react to the lipids, becoming activated and retained[9]. Ultimately, this

maladaptive cellular response to deposited lipid further exacerbates endothelial dysfunction, lipid deposition, and the recruitment and retention of inflammatory cells. The 'oxidative theory' of atherosclerosis is also embedded in this pathogenesis since it is primarily oxidative changes to lipids occurring in the arterial wall that render the lipid species pro-inflammatory[10] and recognizable by macrophage scavenger receptors. Vascular smooth muscle cells in the arterial wall migrate into the inflammatory area, proliferate and de-differentiate in response to cytokines including PDGF released from endothelial cells [11]. As the plaque develops it is increasingly filled with lipid-laden macrophages, termed foam cells, and may be covered by a fibrous cap. At its center there is often acellular lipid deposits, released by dying foam cells [12]. Cells within the plaque release a variety of pro-inflammatory cytokines, matrix metalloproteinases (MMPs), and cathepsins that collectively weaken the fibrous elements of the plaque potentially leading to plaque rupture [13]. Plaque rupture exposes the thrombogenic lipid-rich components of the inner plaque, for example, lysophosphatidic acid, to the circulation causing platelet adhesion, aggregation and thrombosis [14]. A large literature exists describing many of the receptors, adhesion molecules, lipid species, and signaling pathways involved in this process. Many mouse models exist demonstrating that knockout of dozens of individual genes either retard or enhance atherosclerosis in response to atherogenic diets [8].

Despite the wealth of research into the pathogenesis of coronary artery disease our understanding of the precise molecular events that govern the key pathogenic events remains limited. The interaction of atherogenic lipid with macrophages in the arterial wall is clearly central to atherosclerosis. In this study I examine the molecular events

occurring when macrophages become lipid-laden foam cells. I systematically elucidate the transcriptional response to atherogenic lipid and use an unbiased approach to map the regulatory pathways activated by oxLDL. Ultimately, controlling the macrophage response to atherogenic lipid is a potential therapeutic strategy that could evolve from these data. This study for the first time integrates new genetic risk data from genome-wide association studies with human macrophage-derived foam cell formation and I explore the function of risk variants and genes highlighted by this integrative analysis.

1.3.1 Macrophages and foam cells in atherosclerosis

1.3.1.1 Lipid retention in the arterial wall

Retention of circulating low-density lipoprotein in the subendothelial space is a crucial, causal event throughout the course of atherosclerosis[9, 15]. Several scientists have suggested it is the single most important inciting event in atherosclerosis[9, 15]. Although these authors have argued that subendothelial lipid retention is both necessary and sufficient for atherosclerosis it is clearly not sufficient—other mechanisms predisposing to lipid retention in certain areas and an ensuing inflammatory response are also required.

Penetration of circulating lipoproteins beyond the endothelium is an ever-present event but lipoproteins are also removed continuously into the circulation[16]. Lipoprotein retention does not require patches of complete endothelial denudation[15]. Lipoprotein retention increases as levels of circulating lipids increase and statin drugs abrogate atherosclerosis largely through enhanced clearance of LDL via the liver[15]. The most clinically significant penetrants are ApoB containing

lipoproteins—chiefly LDL. ApoB-containing lipoproteins interact with extracellular matrix in the subendothelial space. The relatively small size of LDL—70nm—allows it to easily penetrate in tact endothelium by a process known as endothelial cell transcytosis[17]. Once in the subendothelial space a binding site on positively charged ApoB interacts with extracellular negatively charged proteoglycan molecules—including heparin sulfate, chondroitin sulfate, dermatan sulfate, and keratin sulfate[17]. Once in the subendothelial space the lipoproteins may undergo a number of modifications that enhance proteoglycan binding and also uptake by macrophages through scavenger receptors. These modifications include hydrolysis of surface phospholipids by soluble phospholipase A2 that leads to smaller and more dense LDL particles with greater proteoglycan adherence and exposure of another proteoglycan binding site on ApoB[17, 18]. The other main type of lipoprotein modification is oxidation. An extensive body of literature exists demonstrating that once outside the antioxidant-containing plasma, lipoproteins are enzymatically oxidized making them recognizable by scavenger receptors on macrophages[19, 20]. Modified lipoproteins also stimulate release of chemo-attractants, such as monocyte chemotactic protein 1 (MCP-1), and expression of surface adhesion molecules by endothelial cells, which attract circulating monocytes[21, 22]. As this process progresses, a fatty streak becomes an established plaque and a vicious cycle is established whereby disruption of the subendothelial architecture by modified lipoproteins, macrophages, foam cells, altered proteoglycan composition, and proliferation of vascular smooth muscle cells leads to greater retention of circulation lipoproteins and less capacity for removing them[17, 23]. Numerous factors may promote lipid retention including circulating lipid levels, modifications of circulating lipoproteins (e.g. sialyenation), diabetes-induced changes in subendothelial matrix, and any cause of endothelial dysfunction

such as smoking[17, 24]. Ultimately the cells that have the potential to clear pathogenic lipid from the fatty streaks and plaques are macrophages—as they would for most infective stimuli. Given that this process does not occur satisfactorily as evidenced by the massive accumulation of foam cells in atherosclerotic plaques, research is needed to understand how modified lipoproteins hijack normal macrophage functioning and how macrophages might be reprogrammed towards a more helpful role.

1.3.1.2 Monocyte recruitment and foam cell formation

Cellular atherosclerotic plaques begin when monocytes are recruited from the circulation at sites of endothelial dysfunction and differentiate into macrophages in the subendothelial space[25]. A high circulating monocyte count has been correlated with the risk of coronary artery disease[26]. Studies of macrophage colony stimulating factor (M-CSF) and M-CSF receptor knockout mice provided firm evidence for the importance of monocytes and macrophages in atherosclerosis—reduction in circulating monocyte count and macrophage survival markedly abrogated atherosclerosis[27, 28]. In the first stage of monocyte recruitment, monocytes adhere via their P-selectin glycoprotein ligand 1 (PSGL-1) receptor to P-selectin expressed by activated endothelial cells[26, 29-31]. Firm adhesion of monocytes to the endothelium occurs through binding of constitutively expressed monocyte integrin Very Late antigen-4 (VLA-4) which binds to vascular cell adhesion molecule 1 (VCAM1) on activated endothelial cells[32]. Activation of integrin binding occurs partly in response to monocyte CXC chemokines (CXCL1/2/3/4/8, CCL5) binding to heparin sulphate expressed on the surface of activated endothelial cells[21]. Endothelial cell sheer stress and subendothelial retention of lipoproteins are triggers to the endothelial cell activation required for monocyte adhesion[33, 34]. Endothelial

cells also express chemokines including CCL2 (chemokine (C-C motif) ligand 2, also known as MCP1) and CCL5 that contribute to atherosclerosis by both contributing to monocyte adhesion and triggering chemotaxis into the subendothelial space. Interestingly circulating activated platelets also contribute to monocyte recruitment by, for example, depositing CCL5 on the surface of inflamed endothelium[35]. A wide range of cytokines and chemotactic factors are produced by intimal cells of the developing plaque in response to the inflammatory effect of deposited modified lipoproteins. These include: monocyte chemo-attractant protein 1 (MCP-1), M-CSF, granulocyte macrophage colony stimulating factor (GM-CSF), migratory inflammatory protein 1 (MIP-1), TNF-alpha, transforming growth factor beta (TGF-beta), CCL5, Endothelin 1 (ET-1), and interleukins—IL6, IL8, IL10, IL12 [36].

Monocytes cross the endothelium by a process known as diapedesis—involving homotypic interactions between junctional adhesion molecule A (JAM-A) and platelet adhesion molecule (PECAM-1)[37, 38]. Once within the arterial wall monocytes differentiate into macrophages, a process dependent on the major regulator of myeloid differentiation, M-CSF—the experiments detailed later use M-CSF monocyte-derived macrophages in a model of foam cell formation[36]. Although plaque macrophages are thought to be derived largely from terminally differentiated circulating monocytes recent data indicates local macrophage proliferation is an important source of plaque macrophages [39, 40]. Although macrophages can egress atherosclerotic plaques this process does not significantly modify their numerical presence in atherosclerotic plaques[41]. Macrophages may also undergo apoptosis which can be helpful at an early stage in plaques when efferocytosis pathways (the mechanism by which debris from dying cells is cleared by other phagocytes) is not impaired[41]. Later in atherosclerotic plaques macrophage death is thought to be pro-atherogenic since it

leads to release of their lipid contents in an environment where effective efferocytosis is impaired[41]. Differentiation of monocytes into macrophages is associated with an expression pattern that establishes them as phagocytic cells readily able to uptake modified lipoproteins. The cell surface receptor, CD36, which accounts for the majority of oxLDL uptake by macrophages, is strongly upregulated by differentiation of monocytes into macrophages by M-CSF[42]. The transcription factor PPARG was also upregulated upon differentiation and is a major regulator of lipid processing pathways including lipid uptake[42].

In addition to CD36, other scavenger receptors including scavenger receptor–A (SR-A) are involved in oxLDL uptake. Additional potential mechanisms of uptake include pinocytosis and uptake via the LDL receptor[43]. The LDL receptor is part of what is termed the 'regulated lipid uptake pathway' since in response to cholesterol ester loading, macrophages appropriately downregulate the LDL receptor and upregulate genes associated with the processing and efflux of cholesterol including perilipin 2 and ATP-binding cassette, sub-family A (ABC1), member 1 (ABCA1), respectively [44, 45]. CD36 also directs assembly of Toll-like receptor 4 (TLR4) and TLR6 in response to oxLDL and leads to inflammasome activation and secretion of inflammatory cytokines including IL-1[46, 47].

Ultimately an imbalance in the uptake and efflux of low density lipoprotein leads to lipid loading of macrophages resulting in foam cell formation[36]. Once within macrophages, the lipoproteins are transported in perilipin 2 studded vesicles to lysosomes and cholesterol is stored predominantly as cholesterol esters. Some oxLDL is also known to remain in a relatively unprocessed state within macrophage vesicles[36]. Foam cell formation is also discussed as part of section 3 of this thesis.

1.3.1.3 Transcriptional responses in macrophages and foam cells

Of the roughly 1544 transcription factors in humans[48], only a handful have been shown to be utilized in the transcriptional response of macrophages to oxLDL. Activated pathways can be divided into lipid response pathways, oxidative stress response pathways and cell vitality pathways.

One of the key transcription factors is Peroxisome proliferator-activated receptor gamma (PPARG). In fact PPARG upregulates genes involved in the uptake, processing and excretion of cholesterol by macrophages[49]. PPARG is a nuclear receptor that is activated in response to a variety of lipid-based ligands including fatty acids—and translocates from the cytoplasm to the nucleus on ligand binding [50]. PPARG induces secondary transcriptional responses by upregulating LXR-alpha, which then mediates some of its downstream effects, for example, on the upregulation of cholesterol exporter ABCA1 gene [49].

Overall, mouse models with PPARG agonists and knockouts suggest PPARG expression in atherosclerosis is protective, likely dependent on the balance between its control of lipid uptake, processing and efflux from macrophages[51].

Liver X receptor alpha (LXRα) is a nuclear receptor expressed by hepatocytes and macrophages that has oxysterol (oxidized derivatives of cholesterol) as its natural ligand[52, 53]. The functions of LXRα are viewed as atheroprotective since its activation leads to upregulation of ABCA1 and ATP-binding cassette sub-family G member 1 (ABCG1) involved in cholesterol efflux from foam cells [53]. The genomic binding profile of LXRα was analyzed by ChIP-seq in THP1 tumour macrophages treated with and without synthetic LXRα agonist and with or without oxLDL [54]. OxLDL was found to differentially regulate binding at 690 genomic sites although the

foam cell specific sites had much weaker signal and were not enriched for transcription factor binding motifs raising the possibility of stochastic effects [54]. For foam cell specific LXRα sites gene ontology analysis suggested enrichment for the "regulation of cholesterol storage" pathway and "organic acid biosynthetic process" pathway whilst the authors additionally noted proximity to 14 cell death-associated genes [54]. These enriched pathways were different to those elicited by the synthetic LXRα agonist suggesting that foam cell formation in some way alters the physiological response to LXRα activation. In resting macrophages there were low levels of LXRα expression and lack of genomic binding indicating a lack of constitutive activity [54]. A notable example of foam cell-specific binding compared to ligand-induced binding was at the Poly (ADP-ribose) polymerase 1 (PARP1) gene locus—although the role of LXRα in PARP1 expression remains unclear [54].

The Activating protein 1 (AP1) transcription factor is a heterodimer composed variously of c-FOS, c-JUN, ATF and JDP proteins. OxLDL has been shown to modulate AP1 activity in human macrophages in association with MMP9 upregulation [55].

NF-κB (nuclear factor kappa-light-chain-enhancer of activated B cells) is a transcription factor composed of a complex of proteins that has a key role in mediating inflammatory responses, for example, the response of macrophages to endotoxin. NF-κB pathways exacerbate oxLDL-induced foam cell formation and oxLDL increased NF-κB activation [56].

OxLDL induces an oxidative stress response causing activation of Nuclear factor (erythroid-derived 2)-like 2 (NFE2L2) [57]. Upon nuclear accumulation, NFE2L2 binds to anti-oxidant response element DNA binding sites resulting in the upregulation of a set of enzymes that abrogate the cellular damage caused by reactive

oxygen species[57]. In macrophages, however, activation of NFE2L2 also results in upregulation of CD36[58]. A constituent of oxLDL, 4-hydroxynonenal is one of the major stimuli to NFE2L2 nuclear accumulation[58].

There is conflicting data on whether NFE2L2 is atheroprotective or atherogenic. In the LDL receptor knockout mouse model, simultaneous NFE2L2 knockout resulted in enhanced foam cell formation, and increased production of inflammatory cytokines MCP1 and IL6[59]. Knockout of NFE2L2 in the ApoE mouse knockout model of atherosclerosis resulted in reduced atherosclerosis[60]. The beneficial effect of NFE2L2 in mediating a protective response to oxidative stress was outweighed by its atherogenic effect on stimulating CD36 expression[60].

Given that there are hundreds of other transcription factors, additional approaches are required to study which of these also have a role in the response to oxLDL in macrophages. The downstream effects of the transcriptional responses have been assessed to a limited extent using microarray studies to measure the steady state expression levels of thousands of mRNAs. In chapter 3 I discuss in-depth the outcomes and limitations of these existing studies.

1.3.1.4 Foam cells and advanced plaque lesions

The most devastating consequence of atherosclerosis is rupture of an atherosclerotic plaque leading to thrombosis formation, occlusion of the blood vessel and downstream ischaemia. Macrophages and foam cells play a critical role in the destabilization of atheromatous plaques. Patients with acute coronary syndromes caused by plaque rupture have plaques containing significantly more macrophages than patients with stable angina [61]. At the center of plaques is a lipid-rich core composed of aggregates of foam cells. Death of foam cells either by necrosis or

apoptosis may occur in response to DNA damage caused by oxLDL, or ER stress resulting from a prolonged unfolded protein stress response from excessive accumulation of free cholesterol [13, 62]. Macrophage and foam cell death leads to the release of pro-inflammatory oxidized lipids and cellular debris[62]. Inflammatory macrophages release TNF-alpha that contributes to apoptosis of vascular smooth muscle cells leading to reduced extracellular matrix production and weakening of the fibrous cap of the plaque [63, 64]. Apoptosed macrophages have been visibly associated with areas of plaque rupture[65]. The process by which apoptotic cell clearance takes place—efferocytosis—is also impaired by the effect of oxLDL on phagocytic cells in the plaque; oxLDL constituents compete with apoptotic cell debris for receptors on phagocytic cells involved in phagocytosis [66]. Macrophages may contribute to plaque instability by secreting activated MMP9 [67]. Foam cell formation also triggers increased expression of the chemo-attractant and mitogenic cytokines IL8 and MCP-1—these attract neutrophils, T cells, vascular smooth muscle and enhancer proliferation of vascular smooth muscle cells [68]. Limited studies of the H3K4Me3 histone mark in macrophages (a marker of transcription elongation and enhancer activity) show that oxLDL triggered increased H3K4Me3 marks at the promoters of pro-inflammatory cytokines induced by oxLDL including TNFα, IL-6, IL-18, MCP-1 [69, 70]. This shows that oxLDL-induced changes in gene expression are associated with epigenetic changes.

1.4 Risk factors for CAD

Modifiable risk factors (e.g. smoking status) and non-modifiable risk factors (e.g. age), play a major and additive role in the risk of coronary artery disease[71].

Furthermore, increasing evidence demonstrates the link between the traditional environmental risk factors and recently described genetic risk factors.

1.4.1 Environmental Risk factors

Cigarette smoking has an adverse effect across all elements of atherosclerosis and accounts for over 10% of cardiovascular deaths [72]. At the physiological level smoking induces a state of endothelial dysfunction in which there is abnormal regulation of vascular tone and haemostasis [72-74]. At the cellular level reactive oxygen species within smoke activate NADPH oxidase, deplete endothelial nitric oxide, cause oxidative changes to proteins and activate NF-κB mediated expression of adhesive surface molecules including E-selectin and P-selectin thereby increasing thrombogenicity [75-77].

Smoking also has a detrimental effect on atherosclerotic plaque stability, increasing the likelihood of plaque rupture and thrombosis [72]. Plaques from smokers contain significantly more oxidized-low density lipoprotein than non-smokers [78]. They also contain significantly more foam cells, suggesting a more inflammatory plaque phenotype[79]. Smoking increases the expression of several matrix metalloproteinases, the enzymes that degrade extracellular matrix proteins during tissue growth and repair. For example, plaque macrophages from smokers express more MMP12 (Macrophage Elastase) than non-smokers[80] —MMP12 is elastolytic and associated with plaque progression[80, 81]. As well as increasing expression, smoke constituents, such as acrolein, increased activity of MMP9 in human macrophages [82]. MMP9 is a gelatinase most closely studied in cancer for its role in facilitating angiogenesis[83]. Smokers also have a thinner fibrous covering to plaques in association with downregulation of the key collagen-forming enzyme N-proly 4-

hydroxylase in the arterial wall[84]. Smoking is associated with higher circulating IL6 levels and this cytokine has been shown to downregulate N-prolyl 4-hydroxylase in a mitogen-activated kinase (MAPK) and c-Jun dependent pathway in human aortic smooth muscle cells[85]. Coupled with greater plaque instability smoking increases thrombogenic tissue factor expression by endothelial cells, monocytes, and smooth muscle cells as well as enhancing platelet aggregation and interfering with the balance of fibrinolysis [72]. Presently there are no therapeutic strategies that directly ameliorate the adverse consequences of smoking on atherosclerosis.

Hypertension is a common and treatable risk factor for coronary heart disease—nearly 30% of Americans are hypertensive [86, 87]. Large population-based studies suggest that hypertension causes as much as 47% of ischemic heart disease [88]. The continuous change in cardiovascular risk with blood pressure has meant that the definition of clinically relevant hypertension and its treatment continues to vary. Nevertheless, many studies show that therapeutic lowering of blood pressure in hypertensive patients is highly beneficial in reducing coronary events (typically by 13%) but does not entirely ameliorate cardiovascular risk [87, 89]. Hypertension leads to frank changes in the structure of the arterial wall including fibro-elastic hyperplasia [90].

Endothelial dysfunction is throught to be the primary mechanism by which hypertension contributes to atherosclerosis [90, 91]. However, it remains unclear whether hypertension is a direct cause of endothelial dysfunction or whether endothelial dysfunction results in hypertension. The mechanism of endothelial dysfunction in hypertension primarily involves depletion of endothelial nitric oxide and oxidative stress [92]. The mechanism by which this occurs remains unknown but

once endothelial dysfunction develops other common pathways in atherosclerosis are facilitated such as the trapping of LDL in the sub-intimal space and recruitment of immune cells. This sequence of events is known as the 'response to injury' theory of atherosclerosis [93].

Type 1 and Type 2 diabetes are independent risk factors for CAD [94-96]and the risk of MI was increased from 1.5-5.5 and 1.5-2% in patients with diabetes compared to those without [97]. Intensive lowering of blood glucose reduces cardiovascular events [98]. Numerous pathways underlie the deleterious effects of hyperglycemia on atherosclerosis. One of the main such pathways is the production of advanced glycation end products (AGE) that arise from the intracellular effects of hyperglycemia causing the generation of reactive dicarbonyls that react with the amino groups of proteins [99]. These altered proteins may lead to abnormal crosslinking of the extracellular matrix or interact directly with the receptor for AGE (RAGE) which leads to NFKB-mediated inflammatory signaling. Several of the other mechanisms include activation of protein kinase C, an imbalance in the hexosamine pathway associated with TGF-beta production, and increased activity of 12- and 15-lipoxygenase are enzymes that leads to inflammatory lipid based signaling [99]. Microarray studies of gene expression in macrophages have shown that oxLDL stimulated foam cell formation leads to greater expression of inflammatory genes in the presence of hyperglycemia [100].

Chronic kidney disease, manifested by a declining glomerular filtration rate or proteinuria, is a risk factor for coronary artery disease [101, 102]. Even moderate reductions in glomerular filtration rate to between 30 and 44mls/minute are associated

with a 2 fold increased risk of cardiovascular events. In dialysis patients the risk of cardiovascular death is increased massively—10 to 20 fold [103]. Many potential mechanisms have been postulated to underpin the effect of CKD, most culminating in a perturbation of oxidative stress, lipid handling, vascular calcification, and endothelial dysfunction, brought about by elevated levels of uremic toxins including indoxyl sulphate and P-cresyl [104]. A number of observations suggest that the risk is mediated in part through macrophages and foam cells. In uremic patients macrophages express higher levels of CD36, the principle scavenger receptor responsible for unregulated macrophage lipid uptake [105]. In a mouse atherosclerosis model uni-nephrectomy was associated with greater lesional macrophage content and macrophages demonstrated increase migratory activity in response to MCP-1[106]. Reduction in renal mass also leads to endothelial dysfunction manifested by higher expression of VCAM1, monocyte adhesion and infiltration [107, 108].

In total over 88 uremic toxins have been identified that may in part cause these effects [104]. The prototypical uremic toxin indoxyl sulphate is derived from the action of enteric flora on tryptan, forming indole, that is then converted in the liver to indoxyl and finally conjugated with a sulphate moiety [104]. Circulating Indoxyl sulphate levels are up to 50 times higher in dialysis patients and poorly removed by dialysis due to high protein binding [109]. Indoxyl sulphate increases endothelial leukocyte adhesion molecules (e.g. ICAM-1) via NFKB activation and also causes endothelial cell senescence, impairing wound healing [110]. Indoxyl sulphate increases ROS in endothelial cells and macrophages with resultant reduction in nitric oxide availability [104]. Indoxyl sulphate increases monocyte lineage inflammatory markers and mediators including MAC-1 (Integrin, Alpha M (Complement Component 3 Receptor 3 Subunit))surface expression and IL6 production [104, 111]. Indoxyl sulphate also

contributes to proliferation and dedifferentiation of vascular smooth muscle cells increasing their propensity towards an osteoblastic phenotype and potentially vascular calcification [104].

Circulating levels of lipoproteins LDL and HDL, as well as triglycerides, are well characterized risk factors for coronary artery disease and death from coronary artery disease. The most robust effect is that seen with elevated levels of LDL—levels above 4.13 as compared with levels below 3.35 conferred a 6 fold increase in death from cardiovascular disease[4]. Randomized controlled trials of LDL lowering using 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors have confirmed the central importance of lipoproteins as a risk factor since they significantly reduce the risk or coronary heart disease [112]. LDL is a direct mediator as well as risk factor for coronary artery disease[113]. In this thesis I study the effects of modified LDL on macrophages in a model of foam cell formation.

In addition to the major risk factors discussed above there are several other risk factors including advancing age, obesity, and lack of exercise. Broadly these risk factors have effects overlapping those described above—including endothelial dysfunction, perturbations in circulating lipid levels and oxidative stress.

1.4.2 Genetic risk of CAD

1.4.2.1 Genetic risk factors identified by family studies, linkage studies

CAD is a complex disease with a strong heritable component[114]. A seminal study comparing monozygotic and dizygotic twins demonstrated that if a monozygotic twin had died from CAD aged under 55 then the other twin had a hazard ratio of 8.1 of themselves dying of CAD—for dizygotic twins the corresponding hazard ratio was only 3.8[115]. Overall, over 50% of the risk of myocardial infarction is accounted for

by genetic risk[114, 116, 117]. The genetic architecture of risk is divided into rare, single gene disorders—constituting the minority of the overall risk—and common, minor genetic variants that account for most of the heritable risk[114]. After determining the overall contribution of changes in DNA sequence between individuals to the risk of CAD the next step is identifying the locations of these genetic variations. Linkage studies compare the segregation of genetic markers across generations of large families consisting of members with and without a given trait or disease. After finding markers associated with the phenotype more detailed mapping of what may be a several megabase region can be undertaken to localize the pathogenic variants. These studies have little statistical power to detect small effect sizes but have been successful in identifying genes associated with monogenic CAD disease. A linkage study in Icelandic individuals mapped a coronary artery disease risk haplotype to the Arachidonate 5-Lipoxygenase-Activating Protein gene and showed that individuals with the risk haplotype produced more of the inflammatory mediator, leukotriene-B4[118]. However, only a handful of genes have been identified using this approach; in the last 10 years large scale association studies have shed much greater light on the genetic architecture of CAD.

1.4.2.2 Genetic Factors identified by GWAS

A major development in the study of the heritable component of complex diseases has been the application of genome-wide association studies (GWAS) to identify regions of the genome that contain genetic variants that mediate heritable risk. GWAS look for statistically significant differences in the genetic make-up of thousands of individuals with a trait or disease and individuals who do not manifest the disease—typically by a combination of genotyping thousands of single nucleotide

polymorphisms and imputation. These studies were facilitated by efforts such as the HapMap project and 1000 genomes project that systematically annotated human genetic variation. This allowed the production of microarray-based DNA chips for cost-effective genotyping of many individuals[119, 120]. The microarrays sample a subset of known genetic variants, most commonly single nucleotide polymorphisms (SNPs), and the SNPs tested represent marker SNPs for a region. Imputation is the process by which the genotype at other SNPs can be accurately inferred based on a deeper knowledge of the haplotypes from sequencing many additional SNPs in a large number of individuals[121, 122]. Many more SNPs than the actual physically genotyped SNP can therefore be tested for disease association.

GWAS have identified 46 genomic loci harboring genetic variants that influence CAD risk in persons of European ancestry[123](genome-wide significance adj P $<0.5x10^{-8}$ **Table 1**). The candidate genes at these loci (based on proximity) are predominantly not known to be associated with traditional risk factors such as hypertension of LDL cholesterol. An example is the risk signal identified for rs646776 at the 1p13.3 locus. Individuals with the risk allele T, which has a population frequency of 0.81, have a 1.19 odds ratio of having early onset myocardial infarction [124]. The contribution of each locus to the overall risk is therefore typically small, but the gene pathways mediating risk at each individual locus may contain useful biological information yielding potential therapeutic targets[123]. The common and relatively subtle effect of each CAD risk locus is illustrated by noting that over 50% of the population have at least 50% of the CAD associated variants[125]. The majority of CAD risk variants do not alter the sequence of protein coding genes; GWAS studies themselves do not identify the mechanism by which a

risk locus operates. Furthermore, the particular SNP identified by a GWAS study may only be in linkage disequilibrium (LD) with the causative SNP. Presently the risk loci identified by GWAS studies explain less than 10% of CAD heritability. Several hypothesis have been proposed to explain the 'missing heritability' including the overwhelming problem of multiple testing correction necessary when testing thousands of variants, which then limits the number of 'significant' loci[126]. Other explanations include the possible existence of rare variants with strong effects yet to be discovered or that part of the heritability is based on epigenetic changes such as heritable DNA methylation rather than sequence changes[126]. It is likely that in the future more risk loci will be discovered through advancements in, for example, imputation based on the increased availability of a large number of highly detailed genomes.

1.4.3 GWAS and Disease Mechanisms

A major bottleneck in harnessing the new knowledge of the genetic architecture of CAD for therapeutic purposes is not knowing how a genetic variant mechanistically affects the disease process. For most CAD loci the mechanism by which CAD-associated variants affect the disease is unknown. Indeed, for most CAD loci, it is not even clear in which cell type the risk variants exert their effects. To gain further insight into how CAD variants may act several studies have used network analysis and gene set enrichment analysis to detect pathways that may be affected. A network analysis of CAD SNPs annotated to genes on the basis of proximity or coinciding eQTLs found the top networks pertained to: lipid metabolism, cellular movement, tissue morphology (including size and area of atherosclerotic lesions, quantity of

leukocytes, macrophages and smooth muscle cells), and immune migration and adhesion [123]. A similar study mapped SNPs to genes based purely on eQTLs in adipose, liver, blood and endothelial cells—termed SNP set enrichment analysis (SSEA) [127]. Measures of known canonical pathway enrichment and data driven pathway enrichment suggested that CAD variants may affect processes pertaining to lipid metabolism, immune response and inflammation, coagulation, and vascular wall function [127]. This study however, used whole blood eQTLs and not data specific to monocytes or macrophages. Overall, these studies are supportive of a role for macrophages and foam cells since all of the suggested pathways are known to be active in these cells.

Several studies have used maps of genomic features such as enhancers to annotate genetic variants with regulatory potential[128, 129]. These studies have shown statistical enrichment for disease-associated variants in regulatory DNA, most commonly enhancers. The fact that these enrichment patterns are often cell-type specific suggests a link between cell type and regulatory activity of the variant. For coronary artery disease relatively little evidence of enrichment of CAD variants for regulatory DNA has been found. This indicates a need to map regulatory DNA in a greater number of human cell types, prioritizing those cell types most closely associated with atherosclerosis—including macrophages and foam cells.

Another approach towards establishing a disease mechanism is to look for effects of the disease variant on gene expression—so called expression quantitative traits (eQTL). As with regulatory DNA, eQTLs are often cell type specific and therefore rely on having measured gene expression and genotype in the relevant purified cell types which is not always technically feasible.

STAGE identified 22 CAD gene eQTLs across 7 different CAD relevant tissues and most were operative in only a single tissue or at most 2 tissues[130]. However, the composite nature of the tissues, including for example whole blood, limits the utility of these findings.

Overall, both of these techniques provide a rational strategy for probing the disease mechanism which then requires more in-depth molecular, animal and therapeutic studies to validate the mechanism and ultimately its therapeutic utility. To date no therapeutic strategies have evolved directly from CAD GWAS studies.

For some of the CAD risk loci detailed studies have been performed that employed the strategies above. At the Sortilin 1 gene locus the minor allele A of rs12740374 was found to create a C/EBPA transcription factor binding site, associated with increased transcription of SORT1 and lower circulating LDL levels[131]. At another locus a mechanism was partially elucidated that suggests a risk allele operates not through classical lipoprotein pathways but by altering inflammatory responses[132]. At the 9p21 locus the CAD risk alleles of rs10811656 and rs10757278 were found to disrupt STAT1 binding at an enhancer site in lymphocytes. Other experiments suggested that STAT1 binding in lymphoblastoid cells was important for expression of a long antisense non-coding RNA termed ANRIL[133]. However, evidence for a direct genetic condition linking these two phenomena and relevance to atherosclerosis was unclear[132]. At the 6q23.2 locus the CAD risk allele of rs12190287 was found to disrupt binding of transcription factor AP1 leading to altered TCF21 expression in human coronary artery vascular smooth muscle cells[134]. TCF21 is a transcription factor important in regulating cell differentiation during vascular development. In a subsequent publication the same authors also showed that the same CAD risk allele of

the same SNP increased miR-224 binding in the 3'UTR of TCF21, reducing its expression[135].

1.4.4 Interplay between genetic and environmental factors.

Most CAD risk factors share an environmental and genetic origin. Circulating LDL levels, for example, are influenced by both dietary intake and a plethora of common and Mendelian genetic factors. Indeed 60-80% of the variability in LDL level is known to be accounted for by genetic factors[125]. It is possible that adding an environmental effect such as statin treatment may be less effective in individuals with an adverse LDL genetic profile. Another exciting avenue of contemporary genetics research is the effect of environmental stimuli on the activity of genetic variants. It has been hypothesized that certain genetic variants may only come into play in individuals when an individual is challenged with a given stimulus. A landmark study recently identified thousands of eQTLs in primary human cells that were only present when cells were stimulated with an inflammatory stimulus[136]. The activity of the eQTLs even changed with he duration of the stimulus. This study suggests the genetic risk of certain variants may effectively lay dormant until in this case cells were challenged with a pathogenic stimulus. The mechanisms behind this effect could involved not only activation of regulatory pathways, where transcription factor binding is directly impaired, but also by epigenetic changes that lead to silencing or activation of a disease variant-containing regulatory DNA site. Presently there have been extensive studies using epigenetic data to annotate regulatory DNA but these have primarily focused on cells and tissues in a resting 'healthy' state[137].

1.5 Conclusions and present study aims

The aim of the present study was to understand the macrophage transcriptional response to oxLDL—the major CAD risk factor and direct pathogenic stimulus—at the level of gene expression and at the epigenomic level coupled with an analysis of the upstream regulatory pathways connecting oxLDL to events proximal to DNA. A second major aim was to integrate this data with existing GWAS-derived genetic risk variants to determine which variants might affect macrophages and foam cells.

Table 1 CAD risk variants identified by GWAS studies of CAD phenotypes.

(RF= risk factors)

Chromosome	**Position**	**Reported variant**	**Candidate gene**	**Gene name**	**Influences Classic RF**
1	154395946	rs4845625	IL6R	Interleukin 6 receptor	
1	222793632	rs17465637	MIA	Melanoma inhibitory activity	
1	55496039	rs11206510	PCSK9	Proprotein convertase subtilisin/Kexin type 9	LDL cholesterol
1	56945666	rs17114036	PPAP2B	Phosphatidic acid phosphatase type 2B	
1	109817192	rs599839	SORT1	Sortilin 1	LDL cholest

					erol
2	44072576	rs6544713	ABCG5	ATP-Binding Cassette, Sub-Family G (WHITE), member 5	LDL cholesterol
2	21286057	rs515135	APOB	Apolipoprotein B	LDL cholesterol
2	85760395	rs1561198	VAMP5	Vehicle-Associated Membrane Protein 5	
2	203639395	rs6725887	WDR12	WD Repeat Domain 12	
2	145801461	rs2252641	ZEB2	Zinc Finger E-Box Binding Homeobox 2	
3	138052754	rs2306374	MRAS	Muscle RAS Oncogene Homolog	
4	148364920	rs1878406	EDNRA	Endothelin Receptor Type A	
4	156626078	rs7692387	GUCY1A3	Guanylate Cyclase 1, Soluble Alpha 3	Hypertension
5	131667353	rs273909	SLC22A4	Solute Carrier Family 22 (Organic Cation/Zwitterion Transporter), Member 4	
5	131867452	rs2706399	IL5	Interleukin 5	
6	35003153	rs17609940	ANKS1A	Ankyrin Repeat And Sterile Alpha Motif Domain Containing 1A	HDL cholesterol
6	39166496	rs109477	KCNK5	Potassium Channel, Two Pore Domain	

		89		Subfamily K, Member 5	
6	160848743	rs3798220	LPA	Lipoprotein, Lp(A)	LDL cholesterol
6	12911965	rs12526453	PHACTR1	Phosphatase And Actin Regulator 1	
6	161127125	rs4252120	PLG	Plasminogen	
6	134214525	rs12190287	TCF21	Transcription factor 21	
7	107243559	rs10953541	BCAP29	B-Cell Receptor-Associated Protein 29	
7	19031935	rs2023938	HDAC9	Histone Deacetylase 9	
8	19803093	rs264	LPL	Lipoprotein Lipase	
8	126475770	rs2954029	TRIB1	Tribbles Pseudokinas 1	Triglycerides
9	136141870	rs579459	ABO	ABO blood group	
9	22081397	rs11556924	ZC3HC1	Zinc Finger, C3HC-Type Containing 1	
9	22125503	rs1333049	CDKN2B-AS1	Cyclin Dependent Kinase Inhibitor 2B Antisense RNA 1	
10	44740776	rs1746048	CXCL12	Chemokine (C-X-C Motif) ligand 12	
10	30300787	rs3739998	KIAA1462	KIAA1462	

10	91002804	rs1412444	LIPA	Lipase A, Lysosomal Acid, Cholesterol Esterase	
10	104638480	rs12413409	NT5C2	5'-Nucleotidase, Cytosolic II	
11	103660567	rs974819	PDGFD	Platelet Derived Growth Factor D	
11	116648917	rs964184	ZNF259	ZPR1 Zinc Finger	Triglycerides
12	111884358	rs3184504	SH2B3	SH2B Adapter Protein 3	Hypertension
13	110960685	rs4773144	COL4A1	Collagen, Typ1 IV, Alpha 1	
13	28962686	rs9319428	FLT1	Fms-Related Tyrosine Kinase 1	
14	100111955	rs2895811	HHIPL1	Hedgehog Interactin Protein-Like 1	
15	79075746	rs3825807	ADAMTS7	ADAM Metallopeptidase With Thrombospondin Type 1 Motif, 7	
15	91416550	rs17514846	FES	FES Proto-Oncogene, Tyrosine Kinase	Hypertension
17	17531709	rs12936587	RAI1	Retinoic Acid Induced 1	
17	2095954	rs216172	SMG6	SMG6 Nonsense Mediated mRNA Decay Factor	
17	46967061	rs46522	UBE2Z	Ubiquitin-Conjugating Enzyme E2Z	
19	11159096	rs1122608	LDLR	Low Density Lipoprotein Receptor	LDL cholesterol

19	45395619	rs2075650	APOE	Apolipoprotein E	LDL cholesterol
21	35593827	rs9982601	KCNE2	Potassium Channel, Voltage Gated Subfamily E Regulatory Beta Subunit 2	

2 Materials and Methods

2.1 Ethical Approval

Ethical approval for the study was obtained from the NHS Research Ethics Committee (South Central-Hampshire B, reference 13/SC/0392) and all participants provided informed consent.

2.2 General Methods

2.2.1 Reagents

General chemical reagents are supplied by Sigma (Gillingham, UK) or Fisher Scientific (Rochester, NY, USA).

2.2.2 Buffers and mediums

LB medium: 1% Tryptone, 0.5% Yeast Extract, 1% NaCl in dH2O, autoclaved. TBE buffer (1x): 89mM Tris-borate pH8.3, 2mM EDTA. TAE buffer (1x): 40mM Tris-acetate pH8.3, 1mM EDTA

2.2.3 PCR

For cloning PCR reactions were performed in 50ul using Q5 polymerase according to the manufacturer's protocol (New England Biolabs (NEB), MA, USA). Annealing

temperatures for primer combinations were determined using the NEB annealing temperature webtool (NEB, MA, USA). PCR products were purified using the Qiagen PCR purification kit (Qiagen, CA, USA).

2.2.4 Cloning

Cloning was performed using restriction enzyme methodology unless otherwise indicated. PCR products and plasmids were cut using restriction enzymes from NEB or from Fermentas (York, UK) according to the manufacturer's recommended conditions. Cut products were purified by running on a 1% agarose gel with ethidium, and appropriate bands were excised and purified using the Qiagen Gel Extraction kit (Qiagen). Ligation reactions were performed according to the manufacturer's protocol using T4 ligase (NEB).

Competent E. coli were transformed and plated onto agar plates containing ampicillin. After appropriate incubation at 37C colonies were picked, grown in LB and plasmid DNA was isolated using either the Qiagen Midiprep kit or the GeneJET plasmid miniprep kit (Fermentas).

For plasmid Sanger sequencing Big Dye (Invitrogen) was used and samples were submitted to the Zoology Sequencing Department (University of Oxford). Sequence traces were analyzed using ChromasPro and correctly sequenced plasmids were deposited in the O'Callaghan Laboratory Plasmid archive and electronic database after assignment of a unique plasmid ID.

2.3 Cell culture

$CD14^+$ monocytes were isolated from healthy human volunteers by centrifugation of peripheral blood over Ficoll-Paque PLUS (GE Healthcare LifeSciences, Piscataway, NJ) followed by extraction with magnetic beads conjugated to anti-CD14 antibody (Miltenyi Biotec, Bergisch Gladbach, Germany). Monocyte purity was assessed by flow cytometry using anti-CD14 antibody (AbD Serotec, Raleigh, NC) and was ≥ 95%. Cells were maintained in RPMI 1640 medium with 10% fetal calf serum, 4 mM L-glutamine, 50 units/ml penicillin and 50 μg /ml streptomycin (Sigma, St Louis, MO), supplemented with 50 ng/ml macrophage colony stimulating factor (eBioscience, San Diego, CA). After 7 days these macrophages were treated with either control buffer (1 mM ethylenediaminetetraacetic acid (EDTA), 25 μm $CuCl_2$, phosphate-buffered saline (PBS)) or 50 μg/ml oxLDL for 48hrs. Oil red O staining of intracellular lipids and measurement of cellular cholesterol and cholesterol ester content by mass spectrometry confirmed foam cell formation in oxLDL-treated cells. Viability of cells was confirmed to be > 98% using the Invitrogen LIVE/DEAD fluorescent microscopy kit (Life Technologies, Carlsbad, CA). THP1 cells (ATCC) and NFAT7 reporter cells (a Jurkat cell clone with an NFAT responsive luciferase gene) were maintained in RPMI 1640 with 10% fetal calf serum, 4 mM L-glutamine, 50 units/ml penicillin and 50 μg/ml streptomycin. THP1 cells were treated with 50 ng/ml phorbol myristate acetate (PMA, Sigma) for 48 hours to obtain adherent macrophage cells. Primary human coronary artery vascular smooth muscle cells (Invitrogen, Carlsbad, CA) were cultured in Medium 231 (Invitrogen) and Smooth Muscle Growth Supplement (Invitrogen). Primary human aortic endothelial cells (Invitrogen) were cultured in Medium 200 (Invitrogen) with Low Serum Growth

Supplement. T cells ($CD3^+$) and B cells ($CD19^+$) were isolated from the peripheral blood of healthy human volunteers. Blood was centrifuged over Ficoll-Paque PLUS (GE Healthcare LifeSciences, Buckinghamshire, UK) to isolate PBMCs. Subsets of cells were isolated using positive selection with magnetic bead-conjugated antibodies (Miltenyi Biotec, Bergisch Gladback, Germany). Individual populations were confirmed to be >95% pure using flow cytometry for the markers used for positive selection.

To assess PHACTR1 expression in response to stimuli, primary human macrophages were treated with 50 ug/ml oxLDL or 10 ng/ml LPS for 24 hours and assayed in biological triplicates. HAEC were similarly treated with 50ug/ml oxLDL for 24 hours or 10ng/ml TNF-alpha for the time periods indicated. Triplicates were assessed and Student's t-test applied with significance defined as a P value < 0.05.

2.4 Preparation of oxLDL

LDL (d 1.019–1.063 g/ml) was freshly isolated from human plasma by ultracentrifugation using a discontinuous potassium bromide gradient [138]. Precautions were taken to prevent endotoxin contamination and maintain sterility. LDL was extensively dialyzed against PBS in sterile gamma-irradiated cartridges (Pierce, Rockford, Il) and protein concentration was measured with the BCA method (Pierce, Rockford, Il). LDL was oxidized by incubation with 25 μm $CuCl_2$ at 37° C for 18 hours. Oxidation was confirmed by thiobarbituric reactive substances assay (Caymen Chemical, Ann Arbor, MI). Oxidation was terminated by addition of 1 mM

EDTA and storage at 4C. OxLDL was used within 2 weeks of production. For endotoxin testing, samples were first heated to 75°C for 15 minutes to remove the plasma inhibitor and then assayed using the gel clot method according to the manufacturer's instructions (Associates of Cape Cod, East Falmouth, MA). Levels were < 0.1 EU/ml.

2.5 FAIRE

FAIRE was performed as described with modifications [139]. Five million primary human macrophages or foam cells were cross-linked, lyzed and sonicated (30 pulses of 15 seconds at maximum intensity using a BioRuptor (Diagenode, Denville, NJ)). Nucleosome-depleted DNA was extracted using 4 phenol-chloroform extractions and purified by ethanol precipitation. Three independent biological replicates were produced using cells from three healthy donors. Donors were homozygous for the reference allele of rs17114036/rs72664324.

2.6 ChIP

ChIP was performed using the Invitrogen Magnify kit (Invitrogen) according to the manufacturer's instructions using 200,000 primary human macrophages or foam cells. Cells were cross-linked with formaldehyde for 10 minutes on ice, lyzed and sonicated (32 pulses of 15 seconds, maximum intensity, Diagenode Bioruptor). H3K27ac-enriched DNA was immunoprecipitated using rabbit polyclonal anti-histone H3 (acetyl K27) antibody (ab4729, Abcam, Cambridge, MA). ChIP was performed in

technical duplicates (pooled—total 400,000 cells) and two independent biological replicates were produced from two donors. Donors were homozygous for the reference allele of rs17114036/rs72664324. C/EBP-beta enriched DNA was immunoprecipitated using rabbit polyclonal anti-C/EBP-beta antibody (SC-150X, Santa Cruz, Santa Cruz, CA). ChIP was performed in technical duplicates (pooled total 400,000) cells and 4 biological replicates were produced from 4 donors (2 homozygotes for the reference allele and 2 for the non-reference allele of rs17114036/rs72664324).

2.7 Library preparation and sequencing

Libraries were generated from gel-purified ~200 bp DNA fragments. After adapter ligation and PCR-based amplification, samples were sequenced on the Illumina HiSeq 2000 or 2500 platform (Illumina, San Diego, CA). 50 bp paired-end reads were mapped against the UCSC hg19 reference genome using STAMPY v1.021 [140]. For FAIRE and H3K27ac ChIP samples, reads were filtered (MAPQ minimum 15) in SAMTOOLS yielding at least 54 and 49 million reads respectively per sample [141]. For C/EBP-beta ChIP, paired end reads were mapped against the hs37d5 genome and filtered for a minimum MAPQ score of 4 (removes all non-uniquely mapping reads) and further filtered to remove duplicates (using the markduplicates tool in PICARD available at http://picard.sourceforge.net./). For all subsequent analysis one read of each proper pair was retained along with all unpaired reads passing the requisite MAPQ threshold. Reads mapping to chrM, random contigs, unplaced contigs and ENCODE blacklisted regions were removed from subsequent analysis. To confirm reproducibility of FAIRE data, peaks were called on individual FAIRE samples using

Fseq V1.84 (default parameters) [142]. FAIRE peaks between replicates were intersected and shown to exceed ENCODE FAIRE-seq guideline standards (ENCODE and modENCODE Guidelines for Experiments Generating CHIP, DNase, FAIRE, and DNA Methylation Genome Wide Location Data Version 2.0 [143]) before pooling for further analysis. ChIP samples peaks were called using MACS1.4.2 for H3k27ac and MACS2 for C/EBP-beta (versus input control DNA) and confirmed to exceed ENCODE guideline standards[144].

2.8 FAIRE and ChIP-seq analysis

For FAIRE, dynamic chromatin sites between macrophages and foam cells were identified from pooled biological replicates using Diffreps V1.55 (windows 200 bp, step 20 bp, G-test, FDR < 2.5%) [145]. Clusters of dynamic chromatin sites were identified using the Diffreps HotSpot algorithm ($p < 0.05$). FAIRE-seq peaks were also determined for macrophages and foam cells individually using Fseq on pooled replicates (default settings, with threshold 8.5 for macrophages and 8 for foam cells). Peaks were merged if within 140 bp and filtered out if < 50 bp or > 5 kb in width. For ChIP dynamic H3K27ac enriched sites were identified with Diffreps using separate biological replicates (windows 500 bp, step 50 bp, FDR < 2.5%). ChIP-seq peaks were determined for each cell type individually using MACS1.4.2 on pooled biological replicates versus input control DNA with default parameters. Super enhancer sites were determined using the ROSE algorithm according to the originally established method [146]. For C/EBP-beta ChIP-seq dynamic sites were identified from pooled biological replicates using Diffreps V1.55 (windows 200 bp, step 20 bp, G-test, FDR <2.5%, filtered on minimum 50 reads in dynamic site). Quantile

normalized read counts in dynamic C/EBP-beta binding sites were determined for each genotype in order to compare signal at the rs72664324 dynamic site.

2.9 Data Availability

The high throughput sequencing and microarray data discussed in this thesis have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Super Series accession number GSE54975.

2.10 Annotation of FAIRE and CHIP-seq sites

Genomic features were annotated using HOMER (V4.1), Diffreps and CEAS (cis regulatory element annotation system) [145, 147, 148]. For average signal profiling at transcription start sites using CEAS, wig files of pooled biological replicates were produced using the Java genomics toolkit (https://github.com/timpalpant/java-genomics-toolkit) and subsets of ~200 genes were interrogated (for differentially expressed genes the top two quartiles of upregulated/downregulated genes based on fold change were used). To compare the FAIRE-seq signal between macrophages and foam cell promoters in different subsets of genes, the FAIRE-seq signal at all promoters was first quantified using HOMER (+ /- 1 kb transcriptional start site) and then quantile normalized. Signal at the promoters of all differentially expressed genes was then compared (paired Student's t-test, significance $p < 0.05$) and displayed as boxplots. For correlating dynamic chromatin or dynamic enhancer sites with gene expression, the dynamic sites were annotated with the nearest differentially expressed gene. For correlating dynamic chromatin sites that were specifically within dynamic enhancers, the sites were annotated to the nearest gene and expression compared using a paired Student's t-test, significance $p < 0.05$. For assigning dynamic enhancer status to

dynamic chromatin sites, the dynamic chromatin sites were expanded by 300 bp in each direction to identify co-localization with histone marks of enhancer status that are on adjacent nucleosomes; finally, the two datasets were intersected using bedtools [149]. For genes with more than one Illumina microarray probe the differential expression data were extracted from the differential probe set with the most significant change between the two conditions. Gene set enrichment analysis of various dynamic chromatin/enhancer sites was performed using GREAT V2.0.2 with default parameters [150].

2.11 Display in genome browser

For display of FAIRE-seq, H3K27ac or C/EBP-beta signal in the UCSC genome browser [151] wig files were generated using ngs.BaseAlignCounts function in the Java genomics toolkit (available at http://palpant.us/java-genomics-toolkit/). Wig files were then normalized using the wigmath.scale function (default parameters) and macrophage signal was subtracted from foam cell signal to generate a track of dynamic signal.

2.12 Transcription factor analysis

Transcription factor motif enrichment analysis was performed using SeqPos in the Cistrome Galaxy environment with the curated Cistrome motif database and *de novo* motif generator [152]. Transcription factor motif affinity analysis for binding at specific SNPs was performed using TRAP and the Transfac database [153].

2.13 SNP enrichment analysis

SNPs with genome-wide association ($p < 5x10^{-8}$) to any trait in European individuals present in the GWAS catalogue were identified. Index SNPs were pruned ($r^2 > 0.1$ in CEU samples) so that each 'locus' was only represented by one index SNP to avoid counting redundant loci. Each index SNP was then used to identify variants in 1000 Genomes Project (1KG) pilot 1 data in high LD ($r^2 > 0.8$) in CEU samples using HaploReg [154].

Thus, an associated 'locus' consists of an index SNP and the set of 1KG SNPs in high CEU LD. A background set was created containing all qualifying loci, binned based on the number of total variants (index + high LD SNPs) in the locus.

Published reports of CAD-associated variants with $p < 5x10^{-8}$ in European samples were identified. Where studies had reported different variants at the same locus the variant with the most significant p value was used [123, 124, 155-158] [159]. For each CAD-associated variant, loci were created that contained 1KG variants in high LD ($r^2 > 0.8$) in CEU samples [154].

The following steps to test for enrichment of trait-associated SNPs were designed and performed by Dr Kyle J. Gaulton:

Using the set of CAD-associated loci, the number of loci containing a variant overlapping a given annotation was calculated. 100,000 permutations were performed of the set of loci drawing from matching bins in the background set and recalculated the number of loci with a variant overlapping an annotation. The observed number of loci and the total number of loci and the expected number obtained via permutation

were compared using a binomial test (a Perl script to perform this analysis is available at https://bitbucket.org/kjgaulton/enloc/wiki/Home).

2.14 Non-coding element enrichment

The following steps to test for non-coding element enrichment were designed and performed by Dr Kyle J. Gaulton:

Non-coding element data for ChromHMM chromatin state, TFBS, and multi-species conservation from the UCSC genome browser were obtained. For chromatin state data, 'Enhancers', 'Promoters' and 'Insulators' identified as such in any cell type were pooled. The overlap of dynamic chromatin sites with each regulatory class was calculated, and then performed peak-shifting of the chromatin sites a random distance within a 10 kb window. A recalculation was performed using the overlap with the shifted sites, and a background distribution of expected overlap over 100 permutations was derived. Fold-enrichment values were calculated relative to the background mean and p values for each overlap were obtained directly via permutation.

2.15 Gene expression

2.15.1 RNA extraction

Cells were lyzed and total RNA extracted using the TRIZOL RNA PLUS extraction kit (Life Technologies) with on-column Purelink DNAse treatment (Life Technologies).

2.15.2 Microarrays

The integrity of the total RNA was analyzed on an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Total RNA was reverse transcribed, amplified and biotinylated using the Illumina TotalPrep-96 RNA Amplification Kit (Ambion, Austin, TX). Biotinylated cRNA was hybridized to a single Human HT-12 V4 BeadCHIP (Illumina) at 58°C for 18 hours. The BeadCHIP was scanned using the Illumina Iscanner and data pre-processed using the Illumina Bead Studio to correct for local background effects, remove outlier beads, to compute average bead signal and SD for each probe and gene and to calculate p values. Hierarchical clustering showed that macrophages samples and foam cells samples clustered together respectively. The LUMI pipeline was used for data analysis. Data were transformed (variance-stabilizing transformation or log 2 where indicated) and normalized (robust spline normalization). Differentially expressed genes (probe-centric) were determined using both a paired macrophage-foam cell design and unpaired design after removal of probes with absent expression. For differential expression an FDR-corrected p value threshold of 0.05 was deemed significant [160]. Heat maps of differentially expressed genes were generated using the heatmap.2 function in R based on the subset of the genes with adjusted p value < 0.0001, and for genes with more than one microarray probe, the most significant probe was used (using the unpaired design). Gene set enrichment analysis was performed using GREAT.

2.15.3 RTqPCR

For real time quantitative PCR (RTqPCR) 1 μg of DNaseI-treated total RNA was reverse transcribed using Bioscript (Bioline, London, UK) with random hexamers.

RTqPCR was performed in 20ul reaction volumes with SYBR Green reagents using a Step One Plus machine (Applied Biosystems, Foster City, CA) using technical triplicates and biological triplicates. Melt curve analysis demonstrated a single product for each primer pair.

Primers used:

PPAP2B

CO5547 TTCTGGCAGGATTTGCTCAA

CO5548 AGGGAGAGCGTCGTCTTAGTCTT

IL6R

CO4830 GCATTGCCATTGTTCTGAGGTT

CO4831 ACCAGCTGCCCCAAAGAGT

GAPDH

CO3744 TTGCCATCAATGACCCCTTCA

CO3745 CGCCCCACTTGATTTTGGA.

PHACTR1 intermediate transcript

CO5399 GGAAGAAGAAAAGCGAAAAGTTCA

CO5400 CTTCTCTGCTTTGCCTCATAGATATTT

PHACTR1 short transcript

CO4987 ATGTATCTGCAAGGGCCGAG

CO5122 CTCTTGATCTCTCTCTTCTCCTCCTG

LAT2

CO3885 AACTTCAGCAAAGGAAGCAGACA

CO3886 ACCGCCCCCAGTTGTAATACT

ANG

CO4071 AGAAGCGGGTGAGAAACAAAAC

CO4072 GCCCATCACCATCTCTTCCA

IMPA2

CO3901 ACAGCCCGACGTGGATCAT

CO3902 TCTTGTCGAACAGCAAATCCAA

MEF2C

CO3905 TCAGCCATTTCAACAACATATGG

CO3906 GCCAGCCAGTTACTGAACCAA

2.15.4 eQTL

For PHACTR1, expression levels were determined by RT-qPCR with SYBR green reagents (Invitrogen) using either cDNA from macrophages or $CD14^{+}$-depleted PBMCs from genotyped healthy individuals from the Oxford Biobank. RT-qPCR primers used were: CO5399/CO5400 (intermediate transcripts); CO4987/CO5122 (short transcript); CO3744/CO3745 (control GAPDH). For PPAP2B expression levels were measured using RTqPCR with SYBR green reagents and primers CO5547/5548 in macrophages and foam cells from genotyped individuals (rs72664324-G n = 10, rs72664324-G/A n = 2, rs72664324-A n = 6). For each sample triplicates were assayed.

For PHACTR1, a Kruskal-Wallis test was applied across the three possible genotypes to identify a significant genotype-expression relationship with significance defined as a P value <0.05. For PPAP2B fold change in expression was compared between homozygotes for the G allele of rs72664324 and pooled heterozygotes/homozygotes for the A allele. Student's t test was applied with a significance level of <0.05.

2.16 Genotyping

Genomic DNA was extracted from blood using Quick-gDNA minipreps (Zymo Research, Irvine, CA). Genotyping of the rs72664324 SNP was performed using a custom designed Taqman assay (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions.

Forward primer	AGGTGACCAGATATGCAAGTTGTC
Reverse primer	ACAGGGACTAGGACGAAGGAA

Allele specific MGB probes:

A allele	AGGAAATGAACCAATGTCT
G allele	AGGAAATGAACCGATGTCT

Genotyping of rs17114036 was performed using an inventoried Taqman assay (C__33268873_10, Applied Biosystems). For allelic expression analysis, genotyped individuals were recruited from a cohort of healthy participants of the Oxford Biobank. Homozygous individuals at rs72664324 and rs17114046 all had alleles consistent with the known high LD such that rs72664324-AA was always associated with rs17114036-GG. Participants were previously genotyped by the Oxford Biobank at rs9349379.

2.17 Electrophoretic mobility shift assays (EMSA)

EMSA was performed using nuclear extracts from primary human macrophages (rs72664324) or THP1s (other SNPs) treated with ^{32}P gamma-ATP end-labeled double-stranded DNA probes (PerkinElmer, Waltham, MA) as previously described [161]. The forward strand probe sequences were:

rs72664324-A	AGGAAATGAACCAATGTCTGTTCCT
rs72664324-G	AGGAAATGAACCGATGTCTGTTCCT
CEBP consensus	TGCAGATTGCGCAATCTGCA
NFYA consensus	CGTCTCCACCAATGGGAGGGCTGGGC
STAT5A consensus	AGATTTCTAGGAATTCAATCC
GR consensus	AGAGGATCTGTACAGGATGTTCTAGAT
rs9349379-A	GATCATATAAAAATAGCTTAAAATC
rs9349379-G	GATCATATAAAAGTAGCTTAAAATC
MEF2 consensus	GATCGCTCTAAAAATAACCCTGTCG
rs56348932-A	ACGGACCTCATTCAAGGGGAAAGGCA
rs56348932-C	ACGGACCTCATTCCAGGGGAAAGGCA
rs17114036-G	TCAGGCAGTGGTGACTACTTTTTCC
rs17114036-A	TCAGGCAGTGGTAACTACTTTTTCC
AP1 consensus	CGCTTGATGACTCAGCCGGAA
Mutant AP1 consensus	CGCTTGATGACTTGGCCGGAA

For standard EMSA, 5 μg of nuclear extract was incubated with 100 fmol labeled probe in a 10 ul binding reaction containing 1 μg poly(dI-dC). For competition assays unlabeled probe at 100-fold excess was added to the binding reaction before addition

of labeled probe. For super-shift assays the nuclear extract was pre-incubated with 1 μg antibody for 30 minutes on ice before probe was added. The following antibodies were used: rabbit polyclonal anti-C/EBP beta (sc-150, Santa Cruz), rabbit polyclonal anti-C/EBP alpha (sc-61, Santa Cruz), rabbit polyclonal ant-cJUN (sc-1694, Santa Cruz).

2.18 Luciferase reporter assays

2.18.1 Enhancers

To test enhancer activity at the rs72664324 locus in response to oxLDL, primary human macrophages were transfected with pGL4.23 reporter plasmids using nucleofection in biological triplicates as per the manufacturer's macrophage-optimized protocol (Amaxa, Lonza, Portsmouth, NH). After 24 hours cells were treated with oxLDL 50 μg/ml or control buffer for a further 24 hours. To test enhancer activity in THP1 cells, 1 million cells were added to each well of a 6 well plate and treated with 10ng/ml PMA for 48 hours. THP1 cells were then transfected in at least triplicate using Lipofectamine LTX PLUS with 2.5mcg of vector (for over-expression of C/EBP-beta the 2.5 ug was split 50:50 between C/EBP-beta over-expression plasmid and firefly luciferase vector) for 24 hours. To test enhancer responses in HUVECS, cells were grown to 70% confluence in 6 well plates and transfected using Promofectin-HUVEC (Promokine, Heidelberg, Germany) with 2.5 μg of vector for 48 hours, according to the manufacturer's protocol. To control for transfection efficiency in each replicate, firefly luciferase plasmids were co-transfected with pRL-SV40 encoding Renilla luciferase (1/20 DNA amount compared to Firefly). Cells were lyzed for luciferase assay using the Dual Luciferase Reporter Assay System

(Promega, Fitchburg, WI). The luciferase activity of each sample was normalized to Renilla luciferase activity and shown as relative light units (RLU). For each construct two individual clones were tested independently and representative data are shown. Data were analyzed using Student's t-test on at least triplicates. Plasmids used in luciferase assays shown below.

Plasmid name	Vector	Insert	Restriction enzymes	Comments on insert	Used for:
pOC1252	pCDNA3.1	Human C/EBP-beta	Bamhi-xhoi	Encodes LAP isoform	Over-expression in THP1
pOC167	pCDNA3.1	Empty	NA	Used as control	Over-expression in THP1
pOC1250	pGL4.23	rs72664324-A	Kpni-xhoi	See notes re. insert 4x tandem	Primary human macrophages Antisense orientation
pOC1251	pGL4.23	rs72664324-G	Kpni-xhoi	See notes re. insert 4x tandem	Primary human macrophages Antisense orientation
pOC1268	pGL4.23	rs72664324-A	nhei-xhoi	790bp	HUVECS Antisense orientation
pOC1269	pGL4.23	rs72664324-G	nhei-xhoi	790bp	HUVECS Antisense orientation

31/5/13midi1	pGL4.23	rs17114036-A	Kpni-xhoi	770bp	HUVECS Antisense orientation
31/5/13midi2	pGL4.23	rs17114036-G	Kpni-xhoi	770bp	HUVECS Antisense orientation
16/4/13midi5	pGL4.23	rs56348932- A	Kpni-xhoi	335bp	HUVECS Antisense orientation
16/4/13midi6	pGL4.23	rs56348932- C	Kpni-xhoi	335bp	HUVECS Antisense orientation
pOC1266	pGL4.23	Rs9349379 – G	Kpni-xhoi	223bp insert	THP1
pOC1267	pGL4.23	Rs9349379 – A	Kpni-xhoi	223bp insert	THP1

The inserts for pOC1250 and pOC1251 were generated by annealing primers containing 4 tandem repeats of GGAAATGAACCAATGTCT or GGAAATGAACCGATGTCT. All plasmids were verified by Sanger sequencing using BigDye (Life Technologies).

2.18.2 NFAT7 Mincle Reporter Assay

NFAT7 reporter cells are a stable Jurkat cell clone that express luciferase under the control of an NFAT7 response element. NFAT7 cells lack significant endogenous Mincle. NFAT7 cells were stably transfected using a lentiviral system so that they constitutively expressed FcRγ (plasmid pOC950) and either Mincle and GFP (plasmid pOC942) or GFP alone (plasmid pOC845). Lentiviral particles were produced by

transfecting 293T cells using Genejuice (Merck Millipore) with these plasmids and the packaging plasmids psPAX2 and pMD2.G (ratio of genejuice to plasmid DNA 3ul/1mcg plasmid DNA). The media containing viral particles was collected at 48 and 72 hours post-transfection and spun at 300g for 15 minutes to yield a viral particle-containing supernatant. Supernatants were ultracentrifuged at 21K rpm in a Beckman Ultracentrifuge for 2 hours at 4C. NFAT7 cells were transfected using the resuspended pellet and transfection efficiency was measured using flow cytometry to measure the proportion of cells expressing GFP, which was confirmed to be more than 90%. In the reporter assays compounds suspended in organic solvents were dried onto 96 well plates and 0.5M reporter cells were added per well with biological triplicate replicates (amount of compound indicated in results figures). Plate bound anti-Mincle monoclonal antibody (7D2) was used as a positive control (50ul per well of 7D2 anti-Mincle at 50mcg/ml). The anti-Mincle monoclonal (7D2) was previously generated and validated by Dr A Mistry in the O'Callaghan laboratory. After 6 hours cells were lyzed and luciferase activity assayed using the single luciferase reporter assay system (Promega) in a plate reading luminometer (Tecan, Männedorf, Switzerland). To control for transfection efficiency and cell number, GFP expression was measured in lysates and the luciferase activity was normalized to GFP expression. All assays were performed with biological triplicates and data are representative of two experiments.

2.19 Western blotting

Cells were lyzed in RIPA buffer with protease inhibitors and subjected to SDS-PAGE electrophoresis using 10 ug of protein per lane and PageRuler Plus pre-stained marker

lane. Endogenous PHACTR1 was detected using a rabbit polyclonal anti-PHACTR1 antibody (Novus Biologicals, Littleton, CO) at 1:1000 and secondary HRP-conjugated goat anti-rabbit IgG antibody (Vector labs, Burlingame, CA) at 1:25000. A second rabbit polyclonal anti-PHACTR1 antibody was used where indicated (Abcam ab84646). Chemiluminescent reactions were performed using ECL Advance (GE Healthcare LifeSciences) and imaged using either a BioRAD gel doc imaging station (BioRad, Hercules, CA) or Amersham hyperfilm ECL (GE Healthcare LifeSciences). *Western blotting for LPP3 was performed by Andrew Morris (Division of Cardiovascular Medicine, The Gill Heart Institute, University of Kentucky, Lexington, Kentucky, United States of America) using snap frozen cell pellets of primary human macrophages and foam cells provided by the author*. LPP3 was detected using an extensively characterized and validated rabbit polyclonal anti-human LPP3 antibody and fluorescently multiplexed with beta-actin staining. Images shown are monochrome images of multiplex western blots [162-164].

2.20 Measurement of LPP3 phosphatase activity

LPP3 phosphatase activity was measured by Andrew Morris (Division of Cardiovascular Medicine, The Gill Heart Institute, University of Kentucky, Lexington, Kentucky, United States of America) using snap frozen cell pellets of primary human macrophages and foam cells from 3 different donors provided by the author. LPP3 was immunoprecipitated from Triton X-100 extracted macrophage/foam cell proteins and phosphatase activity determined using heptadecanoyl lysophosphatidic acid and S1P as substrate and measuring heptadecanoyl monoacylglycerol and sphingosine

produced using HPLC electrospray ionization tandem mass spectrometry. The reagents and methods employed have been described in detail elsewhere [162] [165].

2.21 Measurement of LPP3 substrates and products

The following measurements were performed by Andrew Morris (Division of Cardiovascular Medicine, The Gill Heart Institute, University of Kentucky, Lexington, Kentucky, United States of America) using snap frozen cell pellets of primary human macrophages and foam cells from 3 different donors provided by the author. Cholesterol, cholesterol ester and a series of LPP3 substrates and their cognate dephosphorylation products were quantitated in macrophage and foam cell lipid extracts by electrospray ionization tandem mass spectrometry using methods that have been described previously [166, 167]

2.22 Proteomic quantitation of LPP3

The following measurements were performed by Andrew Morris (Division of Cardiovascular Medicine, The Gill Heart Institute, University of Kentucky, Lexington, Kentucky, United States of America) using snap frozen cell pellets of primary human macrophages and foam cells from 3 different donors provided by the author. LPP3-derived tryptic peptides were quantitated using capillary flow reverse phase HPLC and electrospray ionization tandem mass spectrometry using an AB Sciex 5600 Q-TOF mass spectrometer (AB SCIEX, Framingham, MA) operated in high-resolution selected ion monitoring mode. The following peptides were quantified--STIQNPYVAALYK, NGGSPALNNNPR, EILSPVDIIDR—and a mass-labelled derivative of the latter of these was used (New England Peptides, Gardner, MA) as an internal standard [168].

2.23 Immunohistochemistry

Immunohistochemical staining was performed and analyzed by Dr Elizabeth Soilleux and her team (Nuffield Department of Clinical and Laboratory Science, University of Oxford, John Radcliffe Hospital).

Immunohistochemical staining for LPP3 was undertaken using mouse monoclonal antibody clone 7H7D3 (Biosensis, Temecula, CA), which has previously been validated, including for immunohistochemistry [169]. Immunostaining for PHACTR1 and Mincle was performed using rabbit polyclonal anti-PHACTR1 antibody (ab84646, Abcam, Cambridge, UK) and monoclonal anti-Mincle antibody (7D2), respectively. Formalin-fixed paraffin-embedded tissue samples of non-atherosclerotic aorta (n = 5), and atherosclerotic aorta (n = 5) obtained with full ethical approval from the National Research and Ethics Service (Oxfordshire Research and Ethics Committee A: reference 04/Q1604/21), were immunostained using the Bond Max™ fully automated immunohistochemistry system (Leica Microsystems, Wetzlar, Germany). Heat induced antigen retrieval was performed using the EDTA-based Novocastra Bond Epitope Retrieval Solution 2 (pH 9.0) for 30 minutes and the NovolinkTM max polymer detection system (Leica Microsystems) was used, as per the manufacturer's instructions. Detection reagents only were used as a negative control. Slides were mounted in Aquatex mounting medium (Merck, Darmstadt, Germany). Stained sections were photographed with a Nikon DS-FI1 camera with a Nikon DS-L2 control unit (Nikon, Kingston-upon-Thames, UK) and an Olympus BX40 microscope (Olympus, Watford, UK).

2.24 5' RACE

RNA was extracted using the Trizol RNA Plus extraction kit with on-column DNAse treatment according to the manufacturer's instructions (Invitrogen). For 5' RACE, 5 ug of total RNA was reverse transcribed using 20 pmol of gene specific primer (CO5441 TTTGGCTGAAGGATCTTGGG) and then poly A tailed at the 3' cDNA end. First round PCR was performed using an inner gene specific primer (CO5439 CCATCCATGATGTCTGACGGTTGG) and oligo(dT) primer (with additional 5' sequence for round two PCR, CO1352) for 18 cycles using Bioscript (Bioline). Second round PCR was performed using gene specific primer (CO5400 CTTCTCTGCTTTGCCTCATAGATATTT) and a primer (CO3984 CTAGGAATTCTAGAGGTACCTCGAG) annealing to a 5 prime site created in the 1st round PCR. PCR products were gel extracted and ligated into Strataclone vector pSC-A-amp/kan using the strataclone PCR cloning kit (Agilent, Santa Clara, CA). Colonies were cultured, plasmid DNA extracted using a miniprep kit and sequenced using BigDye (Invitrogen) with both inner gene-specific primer CO5513 (CTGACGTGTGTTTGAACTTTTCGC) and also CO5400 (CTTCTCTGCTTTGCCTCATAGATATTT). Sequences were aligned to the hg19 reference genome using the BLAT program in the UCSC genome browser.

2.25 PHACTR1 cloning

The long PHACTR1 transcript was amplified from an Image clone (IRATp970E05120D) using primers CO4400 ACTGATGATCAATGGATTATCCCAAAATGGATTATTTTC, and CO4807

TCAGTCTCGAGTTAAGGTCGGTGAAACCTTGTTAAGTG and cloned into pCDNA3.1 with an N-terminal myc tag to produce plasmid pOC1259.

For other transcripts: RNA was extracted using the Trizol RNA Plus extraction kit with on-column DNAse treatment according to the manufacturer's instructions (Invitrogen). cDNA was produced by reverse transcription of 1 ug of RNA using Bioscript reverse transcriptase (Bioline, London, UK) with random hexamers (Qiagen, Venlow, Limburg, Netherlands).

The intermediate transcript was amplified from $CD14^+$ macrophage cDNA using primers CO4976 ACTGATGATCAATGCGTTCTGACTCCCTCGTCC and CO4807 and cloned into pCDNA3.1 with an N-terminal myc tag to produce plasmid pOC1256. The short PHACTR1 transcript (ENST00000379335) was amplified from macrophage cDNA using CO4806 AATGAGGATCCATGTATCTGCAAGGGCCGAGG and CO4807 and cloned into pCDNA3.1 with an N-terminal Myc tag to make plasmid pOC1257.

2.26 Tissue bank, primary cell and cell line panel expression analysis

An oligo(dT)-primed cDNA library set made from adult and fetal human tissue was screened for expression of PHACTR1 transcripts. PHACTR1 transcripts and beta-actin controls were amplified using BioTaq polymerase with 33 cycles and 25 cycles respectively. For cell lines the same primers were used but the number of

amplification cycles for PHACTR1 and beta-actin was 30 and 20 respectively. The cell line cDNA was the generous gift of Dr Da Lin (O'Callaghan Laboratory).

Products were run on a 1% agarose gel and stained using ethidium bromide. The primer pairs used were as follows: Long transcript: CO3948 TGCTCACAGACTCTTGGATGTTG CO3880 CTTCTCTGCTTTGCCTCATAGATATTT, Intermediate transcript CO3879 GGAAGAAGAAAAGCGAAAAGTTCA CO3876 GGCAGTAGCAGGACTGGTTTG, Short transcript CO4987 ATGTATCTGCAAGGGCCGAG CO3944 TGAATTCATTGAGCTCCTTTCG, Beta-actin CO1851 TCCAGCCTTCCTTCCTGGGCAT CO1852 GTCAAGAAAGGGTGTAACGCAACT

2.27 Confocal microscopy

U2OS cells were grown on coverslips and transfected with one of the three N-terminal myc-tagged PHACTR1 constructs or control vector (empty pCDNA3.1) using Genejuice (Merck). After 24 hours cells were fixed in methanol-free formaldehyde, permeabilized with 0.1% Triton X100 and blocked with 3% BSA/3% goat serum. Fixed cells were stained with a primary mouse monoclonal anti-myc tag antibody (Clone 9E10) and a secondary goat anti-mouse IgG antibody conjugated to AF488 (A-11001, Invitrogen). Actin and cell nuclei were stained with phalloidin-rhodamine (Invitrogen) and Vectashield containing DAPI (Vector Labs), respectively. Slides were visualized using a Zeis 510 Metahead confocal microscope, running Zeiss LSM510 v4.2 operating software, using a 63x oil immersion lens at room temperature.

3 Macrophage-derived Foam Cell Formation

3.1.1 Introduction

From the beginnings of atherosclerotic lesions as mere fatty streaks to the development of advanced lumen-encroaching atheromatous plaques, monocyte recruitment, retention and foam cell formation play a crucial role [170, 171]. The profound inhibition of atherosclerosis in mice severely deficient in monocytes through lack of monocyte colony stimulating factor (M-CSF) [27] highlights the central role of macrophages in atherosclerosis.

In atherosclerosis, LDL cholesterol is deposited in the arterial wall and undergoes modifications, such as oxidation, that in turn promote pro-inflammatory processes [172-174]. Monocytes are recruited to these sites and differentiate into macrophages, which have a variety of scavenger surface receptors for oxidized LDL (oxLDL) [172, 173]. The uptake of oxLDL by macrophages occurs via scavenger receptors, phagocytosis and macro-pinocytosis and is fundamental to the development of atherosclerotic lesions [173]. An imbalance in the uptake and degradative metabolism of oxLDL leads to the accumulation of lipid-laden vesicles in macrophages giving rise to a foam cell phenotype [172, 173].

Foam cells contribute to the pathogenesis of atherosclerosis by releasing pro-inflammatory mediators (including IL8 and MCP-1 TNFα, IL-6, IL-18 [69, 175]) that recruit additional cells and matrix metallo-proteinase 9 [67]that can destabilize plaques [172, 173]. In addition, foam cell apoptosis releases toxic lipids into the

necrotic core of the lesion[13, 176]. Plaque rupture can expose this thrombogenic mixture to luminal blood leading to thrombosis and blood vessel occlusion[172]. Several studies have documented changes in lipid response and oxidative stress response pathways during foam cell formation, but our understanding of how oxLDL influences transcriptional regulation in macrophages is incomplete[8, 57]. Reprogramming of the macrophage response to lipid is a plausible therapeutic strategy. In this chapter I use a whole genome approach to comparing gene expression in human macrophages and oxLDL-elicited foam cells. In subsequent chapters I integrate this expression data with epigenomic data to produce a composite picture of the effects of oxLDL on macrophages as a whole and I perform in-depth analysis of candidate CAD genes, PPAP2B and PHACTR1.

3.1.2 OxLDL-induced *in vitro* foam cell formation

LDL is one of 5 major classes of circulating lipoproteins and contains a mixture of apoB100 protein, cholesterol, phospholipids and triglycerides [177]. LDL is generated by the action of hepatic lipase on intermediate density lipoprotein. LDL is removed from the circulation by the liver, through binding of apoB100 to the LDL receptor on hepatocytes, and also by the LDLR on other cell types [177]. In atherosclerosis LDL entrapment in the arterial sub-intimal space occurs at sites of endothelial damage and once outwith the anti-oxidant environment of circulating plasma, LDL particles become modified by oxidation [113]. Ultracentrifugation of human plasma over a discontinuous density gradient yields orangey LDL in the fraction between d 1.019 and d 1.063[138]. To model endogenous oxidized lipoproteins, LDL is oxidized using a variety of techniques, including by the addition of copper [178-180]. Lipid laden macrophages resembling those of plaque foam cells are generated *in vitro* by

incubating human monocyte-derived macrophages with copper-oxidized LDL [179-182].

Foam cells generated in this way show phenotypic similarities with plaque foam cells including engorgement with neutral lipids, readily measured by staining with oil red O [179, 180]. *In vitro* studies of oxLDL induced foam cells have shown a complex induction of both pro and anti-inflammatory phenotypes[179]. On the one hand, oxLDL activates pathways associated with protection from atherosclerosis including PPARG signaling[183, 184]; on the other hand, oxLDL also acutely stimulates production of MMPs, cytokines and contributes to chronic inflammation through reduced mobility and altered apoptosis[25, 173, 185, 186]. Due to the difficulty in obtaining sufficient numbers of human macrophages many investigators have used tumour cell lines including the monocytic cell line THP1 [54, 187, 188]. The use of tumor lines is, however, less desirable since these cells are unlikely to accurately model endogenous macrophages and the pitfalls in using cancer cells have been documented[189]. Furthermore, differentiation into an adherent macrophage-like phenotype requires treatment with the powerful protein kinase activator phorbol 12-myristate 13-acetate (PMA).

Different classes of macrophages have been described, largely by arbitrarily defining their phenotype along the lines of the well-established classification of T cells. The two main classes of macrophages include the M1 macrophage, which is elicited in-vitro by adding interferon gamma—regarded as a pro-inflammatory macrophage—and M2 polarization triggered by IL4—considered an anti-inflammatory phenotype[182, 190, 191]. However, there is inconsistent application of this distinction in the literature, the cytokines used, relevance to endogenous macrophages and extrapolation from mouse models to human macrophages, particularly with

regards marker based classification. In atherosclerosis it is less clear how this taxonomization applies to foam cells since it remains unclear from which type of macrophages plaque foam cells are derived[182] and how their final phenotype fits in with this system [192]. Macrophages generated *in vitro* by the differentiation of monocytes using M-CSF alone are termed M0 macrophages and readily become foam cells after oxLDL stimulation [179, 182]. In the present study I derived foam cells from M-CSF differentiated macrophages.

3.1.3 Whole-genome approaches to gene expression in foam cells.

The 21st century development of high-throughput assays based on DNA probe based microarray technology allowed the simultaneous measurement of potentially the entire known transcriptome. In practice many of the microarray platforms that have been generated over the last 10-15 years have been relatively limited to measuring less than 1/3rd of known genes—and fewer in some cases. I identified all studies in the literature that assayed gene expression in foam cells using microarray technology (Table 2). Overall, I identified 18 studies that used a range of microarray formats with differing levels of sophistication, replicates, statistical analysis, and cell culture methods. I noted that across these studies there were nearly 6000 differentially expressed genes in total but only 305 genes were differentially expressed in more than 3 datasets. One of the most consistently upregulated genes was CD36, which was differentially expressed in 5 datasets. CD36 is one of the most important cell surface scavenger receptors in terms of oxLDL uptake and foam cell formation, being responsible for up to 70-90% of oxLDL uptake *in vitro*[173, 193]. The CD36 receptor

is known to be paradoxically upregulated by oxLDL unlike the LDLR receptor which is subject to appropriate repression by high intra-cellular lipid levels [173]. RNA-seq has potential advantages, primarily being able to potentially survey any transcript since it is not restricted to pre-designed probes. To date no RNA-seq study has been performed comparing gene expression between human macrophages and foam cells. Studies comparing the use of microarrays and RNA-seq for evaluating the transcriptome in macrophages have shown a very close correlation between methodologies[194]. In the present study I used the most sophisticated microarray platform and statistical tests available to compare gene expression in M-CSF differentiated macrophages treated with or without oxLDL.

Table 2 Microarray studies of gene expression in macrophages and foam cells.

Author	Date	Journal	Cells	Cytokine differentiation	Lipid stimulation	Number of genes/transcripts tested	Repository data
Andersson	2001	Pathobiology	THP-1	PMA	oxLDL		NA
Scholz	2005	Circulation	THP-1	PMA	oxLDL	14500	NA
Levula	2006	Scandinavian Journal of Clinical & Laboratory Investigation	Primary (blood)	?	oxLDL		NA
Svensson	2006	Atherosclerosis	Primary (blood)	GM-CSF	oxLDL		NA
Hägg	2006	Atherosclerosis	Primary (blood)	GM-CSF	oxLDL/mmLDL		GSE9874
Groeneweg	2006	J Lipid Res	Primary (BMD)	NA	oxLDL		NA
Hung	2006	FEBS Lett	THP-1	Not used	oxLDL/acLDL	9600	NA
Jiang	2007	European Journal	U937	Not used	oxLDL		NA

		of Pharmacology					
Dahl	2007	Circulation	Carotid endartrectomy tissue	NA	in vivo		NA
Cho	2007	Physiol Genomics	Primary (blood)	M-CSF	LDL/mmLDL/oxLDL	12978	GSE7138
Qiu	2007	Journal of Lipid research	THP-1	PMA	oxLDL	110	NA
Hagg	2008		Primary (blood)	GM-CSF	mmLDL		GSE9874
Yuan	2009	Heart and Circulatory Physiology	Primary (aorta)	NA	NA	10328	GSE14854
Ferreira	2009	Unpublished	THP-1	PMA	oxLDL		GSE12434
Wiesner	2010	Circulation Research	RAW267.1, J774, PPM	Thioglycollate (PPM)	mmLDL		NA
Eijgelaar	2010	Physiol Genomics	Primary (tissue)	NA	NA		GSE70126
Pritchard	2010	Unpublished	PPM	Thioglycollate	acLDL	28853	GSE19926
Sung	2011	Molecular Biology Reports	THP-1	PMA	oxLDL		NA
Hirose	2011	Lipids in health and disease	Primary (blood)	M-CSF	oxLDL		NA
Wang	2011	Biochemical and Biophysical Research Communications	U937	Not used	oxLDL		NA
Present study:							
Reschen	2015	Plos Genetics	Primary (blood)	M-CSF	oxLDL	31,000 (47000 probes)	GSE54975

3.2 Results

3.2.1 Foam cell formation

Primary human monocytes were isolated from peripheral blood from 5 healthy volunteers of European descent, by positive selection of $CD14^+$ cells. To differentiate these cells into macrophages they were cultured in R10 with 50ng/ml M-CSF for 7 days. LDL was isolated from the blood of healthy European volunteers by ultracentrifugation with a discontinuous potassium bromide density gradient. The fraction between d 1.019 and d 1.063, visible as an orange band was aspirated by puncturing the side of the ultracentrifuge bottle (Figure 3-1). The LDL was then extensively dialyzed against PBS to remove potassium bromide. The LDL was then adjusted to a protein concentration of 500mcg/ml and extensively oxidized by incubation with 25um Copper sulphate for 18 hours. The copper was then chelated with EDTA and the fresh oxLDL stored at 4 degrees. After 7 days of cell culture half of each participant's macrophages were exposed to oxLDL (50mcg/ml, copper oxidized) and half were exposed to control buffer (PBS with identical amounts of Copper sulphate and EDTA) for 48 hours to generate foam cells or control macrophages, respectively. Foam cell formation was confirmed by light microscopy of macrophages and oxLDL treated macrophages using oil red O staining to detect neutral lipids (Figure 3-2). There was extensive cytoplasmic oil red O staining indicating lipid accumulation and foam cell formation in oxLDL treated cells. OxLDL treatment did not have an adverse effect on cell survival (Figure 3-3). Mass spectrometry of cellular lipids also indicated marked cholesterol and cholesterol ester accumulation in oxLDL treated macrophages (personal communication – Andrew

Morris). Foam cell experiments measuring gene expression and chromatin changes were performed with foam cells resulting from 48 hours oxLDL exposure.

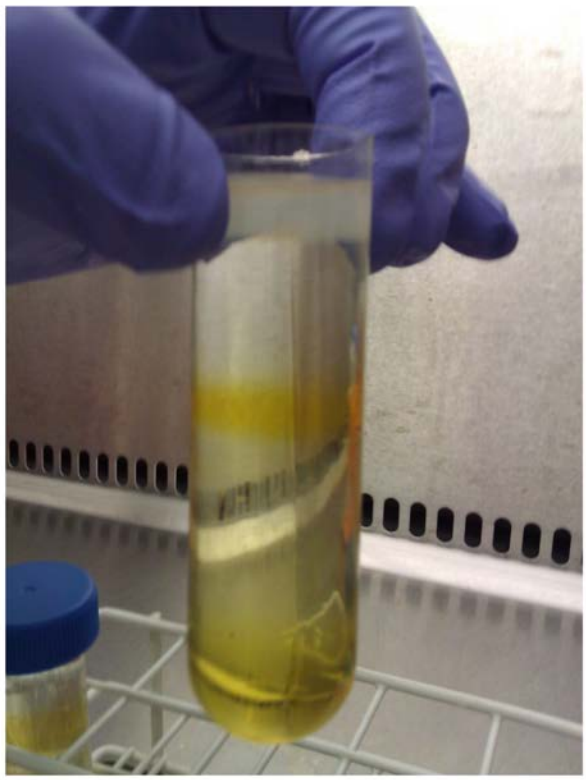


Figure 3-1 Isolation of LDL from human plasma.

Photograph shows a demonstration of the isolation of LDL from human plasma by ultracentrifugation of plasma using a discontinuous density technique. LDL migrates to between the 1.019 and 1.063 density layers and is visible in the middle of the tube as a thick orange band. In this specimen chylomicrons are visible at the top of the tube indicating a non-fasting sample. To make LDL for experimental purposes an alternative centrifuge tube was used with a rubber lid seal to avoid contamination.

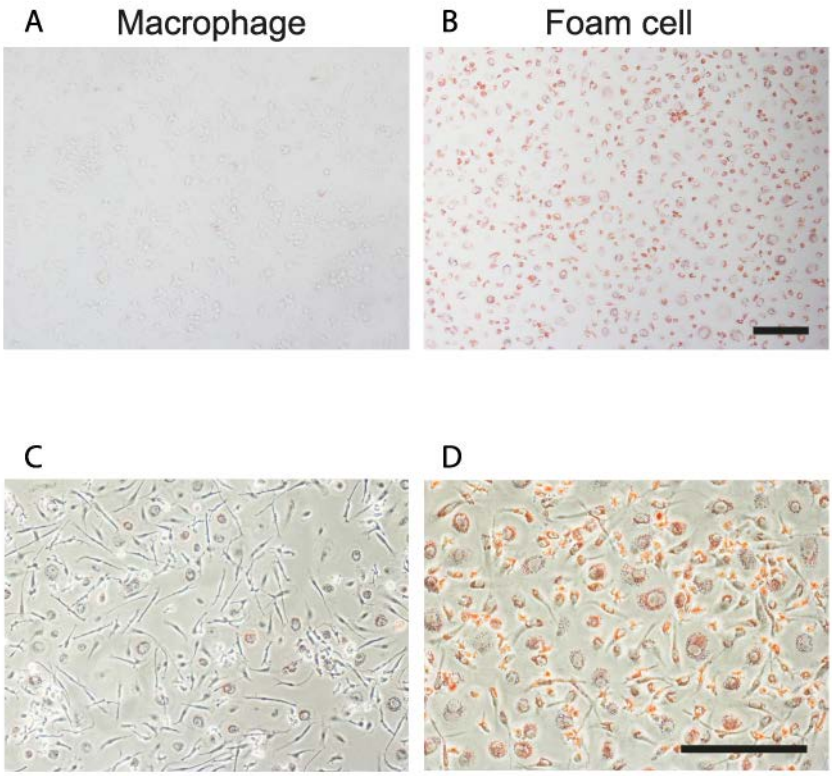
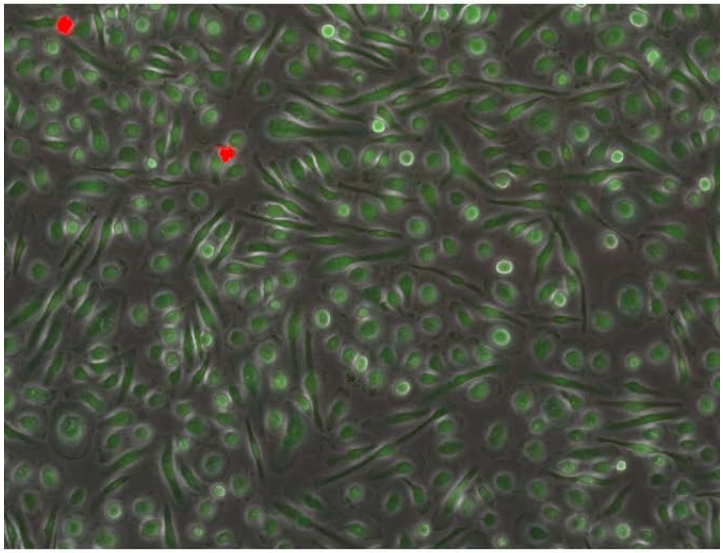


Figure 3-2 Light microscopy of foam cell formation.

Light microscopy of primary human macrophages (A, C) and oxLDL treated macrophages (B, D) stained with oil red O and visualized with (C, D) and without (A, B) phase contrast demonstrates extensive red-colored intracellular lipid in oxLDL-treated cells which confirms foam cell formation (scale bar indicates 100 um).

Macrophages



Foam cells

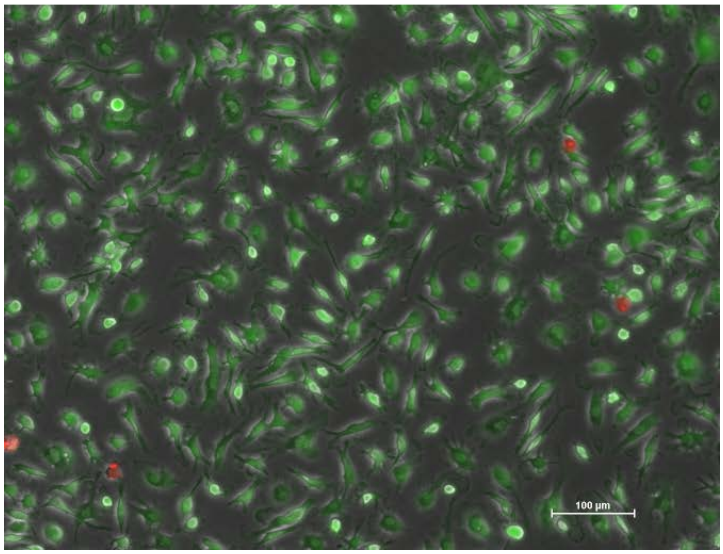


Figure 3-3 Fluorescence microscopy to assess cell viability

Live cells with an intact membrane stain green, non-viable cells stain red.

I then used the Illumina microarray platform to measure whole genome expression in macrophages and foam cells. Five sets of paired macrophage and foam cell samples from different individuals (Caucasian/European) were produced and expression measured using the HT-12 Illumina microarrays. The data were then processed using an established pipeline—LUMI—in the R software environment. Raw data were normalized (robust spline normalization) and then transformed (variance-stabilizing transformation) before performing differential expression analysis using LIMMA software. Differential expression was regarded as significant if the FDR adjusted P value was less than 0.05. Data analysis was performed using both a paired and unpaired approach. For the paired analysis I identified 1,283 and 1,376 genes significantly up- and down-regulated in oxLDL-induced foam cells. For the unpaired analysis I identified 416 up-regulated genes and 453 down-regulated genes. Unless indicated I subsequently used the unpaired approach list of differentially expressed genes for further analysis since all other types of whole genome data did not use the same individuals or pair-wise data analysis (Table 3). To help confirm the validity of the microarray findings I performed RTqPCR for 6 different genes that were differentially expressed on the microarray. In each case, there was also a similar significant change in mRNA when the same samples were assessed by RTqPCR (Table 4). Hierarchical clustering analysis of the expression levels of differentially expressed genes (with adjusted P value < 0.0001) showed the relationship of the samples very clearly (Figure 3-4). The first branch point in the dendrogram shows that the samples cluster distinctly into two segments based on whether the samples are from macrophages or foam cells. The next branch points show the similarity between individual samples in the two main segments. Each sample within a segment clusters

within way in both segments—this demonstrates the pairing of samples between macrophages and foam cells.

Among the most differentially expressed genes were several with previously known involvement in foam cell formation, including the low-density lipoprotein receptor (*LDLR)* and perilipin 2 (*PLIN2)* [44, 45](Table 3). I observed enrichment among all differentially expressed genes for gene ontology terms pertaining to lipid and sterol handling (Figure 3-5). Pathway analysis of the differentially regulated genes further highlighted pathways involved in lipid handling and inflammation (Figure 3-6). The list of differentially expressed genes includes many with no known role in foam cell formation. For example, CD93 is a cell surface glycoprotein shed by macrophages that is involved in phagocytosis [195, 196]. Circulating levels are associated with CAD and in the present study CD93 was strongly down-regulated by oxLDL[197, 198]. Angiogenin (ANG) is up-regulated by oxLDL; this finding was confirmed by RTqPCR. This is an important finding because ANG is a potent inducer of angiogenesis—neovascularization of atherosclerotic plaques is involved in destabilization and plaque rupture [13, 199].Circulating levels of angiogenin are also positively correlated with the severity of 3 vessel coronary disease [200].

As an external reference I identified two studies using broadly similar methodology for which the full datasets were available and reanalyzed the data using the same industry standard approach applied to our data [179, 181]. 668 of the 2659 differentially expressed genes in our study were also differentially expressed in another study[179] despite the other study having a number of methodological differences. 165 genes were similarly differentially expressed in the study by Hagg et al. [181]. Overall 68 genes were differentially regulated in our study and the studies by Cho and Hagg et al. [179, 181]. This core list of genes included members of major

known pathways involved in foam cell formation including haemoxygenase 1 and Glutamate-Cysteine Ligase, Modifier Subunit (GCLM) involved in the antioxidant response [201], and ATP-Binding Cassette, Sub-Family A (ABC1), Member 1 (ABCA1), a membrane associated transporter that mediates cholesterol efflux from macrophages. Important differences with our study include differential expression of CD36 and LDLR in our study but not in both of the others. Given that differential expression of these genes has been established many times over in studies of foam cell formation I believe our study most closely approximates *in vivo* foam cell formation [202, 203].

Table 3 Differentially expressed genes in foam cells compared to macrophages.

	Up-regulated genes			Down-regulated genes		
Rank	Gene symbol	Gene	Fold change (log2)	Gene symbol	Gene	Fold change (log2)
1	*PLIN2*	Perilipin 2	2.70	*LDLR*	Low Density Lipoprotein Receptor	-3.09
2	*AKR1C4*	Aldo-Keto Reductase Family 1, Member C4	2.50	*F13A1*	Coagulation Factor XIII, A1 Polypeptide	-2.81
3	*AKR1C3*	Aldo-Keto Reductase Family 1, Member C3	2.45	*SC4MOL(MSMO1)*	Methylsterol Monooxygenase 1	-2.57
4	*ABCG1*	ATP-Binding Cassette, Sub-Family G (WHITE), Member 1	2.44	*SQLE*	Squalene Epoxidase	-2.46
5	*PDK4*	Pyruvate Dehydrogenase Kinase, Isozyme 4	1.95	*INSIG1*	Insulin Induced Gene 1	-2.15
6	*PPAP2B*	Phosphatidic Acid Phosphatase	1.65	*FCGBP*	Fc Fragment Of IgG Binding Protein	-2.01

		Type 2B				
7	*LOC644496*	Pseudo gene similar to cathepsin L1 preproprotein	1.54	*CD93*	CD93 Molecule	-2.00
8	*TMEM158*	Transmembrane Protein 158	1.51	*ALOX5AP*	Arachidonate 5-Lipoxygenase-Activating Protein	-1.91
9	*FABP4*	Fatty Acid Binding Protein 4, Adipocyte	1.50	*DHCR7*	7-Dehydrocholesterol Reductase	-1.83
10	*SLC7A11*	Solute Carrier Family 7	1.45	*FADS1*	Fatty acid desaturase 1	-1.81

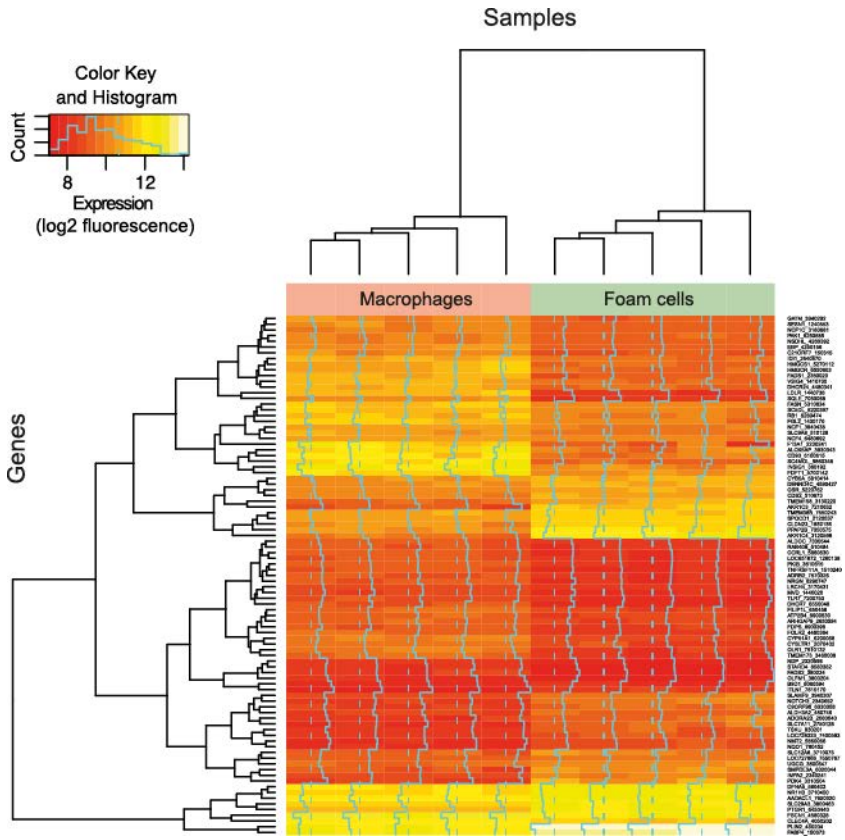


Figure 3-4 Effect of oxLDL-induced foam cell formation on gene expression in human macrophages.

Heat map showing genes differentially expressed between macrophages and foam cells (subset with adjusted $p < 0.0001$). Each column in each half of the main panel represents one of the 5 donors. Each row in the figure represents one gene. Dendrograms indicate clustering of genes by expression profile (left axis) and by sample (top axis). Dashed vertical lines indicate the median expression signal and solid vertical lines indicate expression for that gene and sample relative to the median. The inset indicates the expression distribution for the data and the color key for the heat map.

Table 4 Comparison of gene expression in macrophages and foam cells by microarray and RTqPCR for 6 different genes.

A complete list of differentially expressed microarray probes for each gene is shown.

Gene	Illumina probe	Microarray Fold Change (log2)	Microarray Adj P value <0.05	RTqPCR Fold Change (log2)	RTqPCR P value
PPAP2B (Phosphatidic Acid Phosphatase Type 2B)	7050575 6220097 3710678	1.65 1.63 0.40	1.17E-05 5.24E-05 0.014	2.17	0.008
PHACTR1 (And Actin Regulator 1)	3170139	0.95	0.001	2.56	
LAT2 (Linker For Activation Of T Cells Family, Member 2)	3060092 2810162 3290692	-0.82 -0.66 -0.35	0.0008 0.006 0.021	-1.50	0.003
ANG (Angiogenin, Ribonuclease, RNase A Family, 5)	3930392	0.49	0.0492	1.00	0.002
IMPA2 (Inositol(Myo)-1(Or 4)-Monophosphatase 2)	2340241	1.12	7.05E-05	1.27	0.03
MEF2C (Myocyte Enhancer Factor 2C)	4040162	-1.10	0.001	-1.69	0.001

To gain further insights into the biological process affected by oxLDL we performed gene ontology enrichment analysis using two different software platforms. The list of differentially expressed genes was split into genes that were up-regulated or down-regulated by oxLDL and then analyzed for gene ontology enrichment (Figure 3-5, Figure 3-6, Table 5, Table 6). For up-regulated genes the most significantly enriched biological process found by the GREAT software platform was for 'response to oxygen-containing compound' which encompasses genes such as super oxidize

dismutase 2 (SOD2)(Figure 3-5). SOD2 is an enzyme involved in the reducing the deleterious effects of reactive oxygen species and is cardio-protective in ischemia-reperfusion injury. The gene ontology analysis using the DAVID[204] software identified the second most enriched cluster of ontology terms as pertaining to oxioreductase activity—'Nicotinamide adenine dinucleotide phosphate'(Table 5). These findings highlight the appropriate cellular response of macrophages to the oxidizing stimulus—oxLDL. For down regulated genes there was robust enrichment for genes sets involved in the synthesis of cholesterol and sterol—in effect the cells are appropriately shutting down their production of cholesterol due to the severe cholesterol loading imposed by the oxLDL stimulus (Figure 3-6). The DAVID tool, which clusters similar gene ontology terms, showed that the most enriched cluster for downregulated genes was also for sterol biosynthesis (Table 6). Interestingly the GREAT software analysis also suggested that macrophages are de-differentiated and that the immune response is suppressed in response to oxLDL(Figure 3-6).

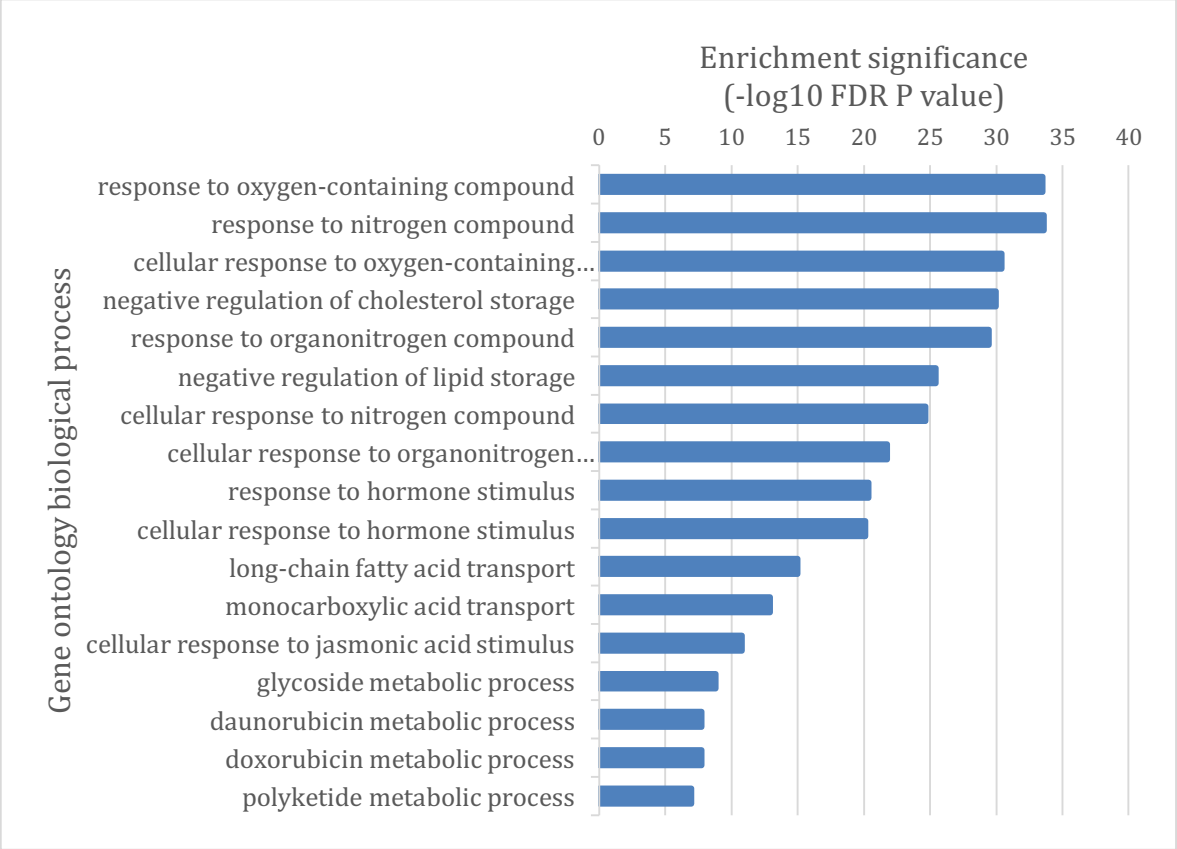


Figure 3-5 Gene ontology enrichment for biological processes for genes that are upregulated by oxLDL.

Enrichment statistic calculated by GREAT (see methods) using default settings and all oxLDL upregulated genes as input.

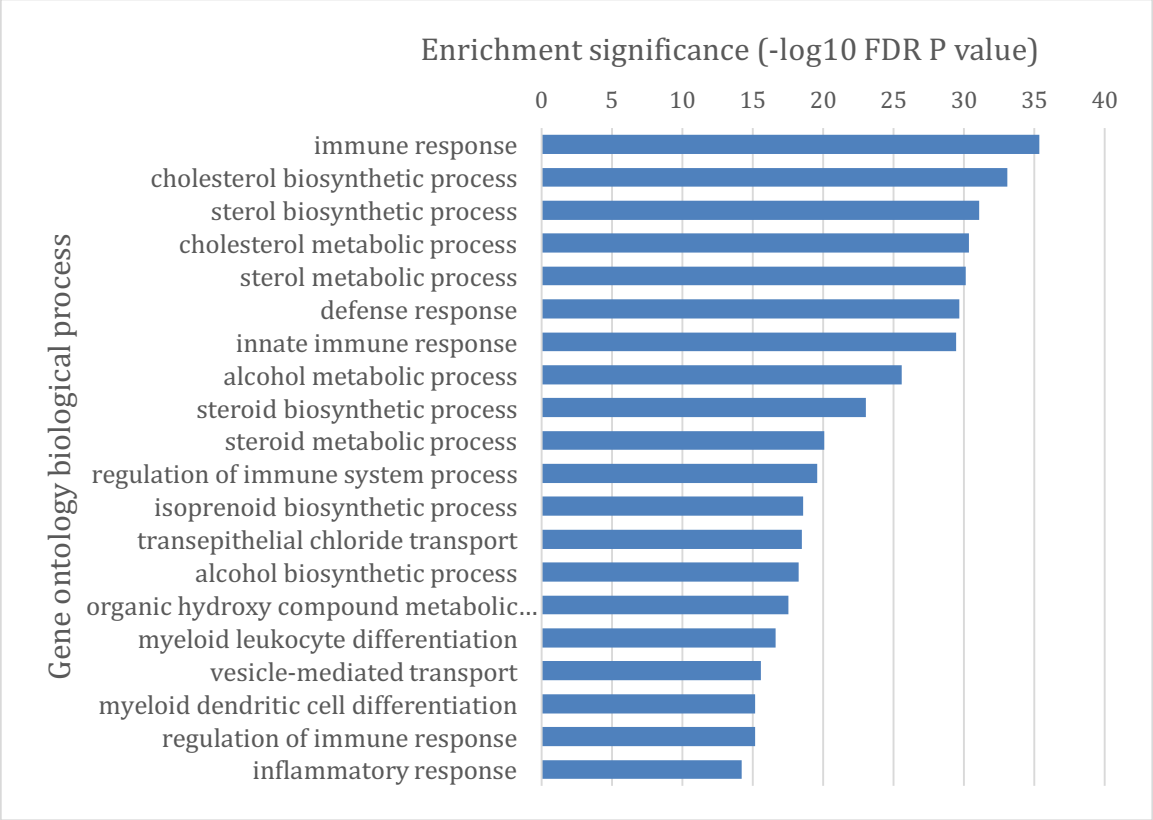


Figure 3-6 Gene ontology enrichment for biological processes for genes that are down-regulated by oxLDL.

Enrichment statistic calculated by GREAT (see methods) using default settings and all oxLDL upregulated genes as input.

Table 5 Gene ontology enrichment for genes that are upregulated by oxLDL using DAVID

The list of genes upregulated by oxLDL was processed using an alternative enrichment tool (DAVID https://david.ncifcrf.gov/) using default settings. The program clusters similar gene ontology terms that are composed of overlapping gene sets. The table below shows the most significant gene ontology term for each of the top 10 enriched clusters of gene ontology terms.

Cluster	Term	Genes hit	Fold enrichment	P value (-log10)
1	Lipid metabolism	14	3.71	5.20
2	Nicotinamide adenine dinucleotide phosphate (oxioreductase)	13	3.45	4.35
3	Lysosome	12	3.18	3.88
4	Regulation of cholesterol storage	5	1.33	4.27
5	Lipid transport	13	3.45	4.31
6	Membrane	160	42.44	5.25
7	Glutathione metabolic process	6	1.59	3.61
8	Aldo/keto reductase	5	1.33	3.48
9	Organophosphate metabolic process	13	3.45	3.01
10	Glutathione metabolic process	6	1.59	3.61

Table 6 Gene ontology enrichment for genes that are down-regulated by oxLDL using DAVID

The list of genes down-regulated by oxLDL was processed using an alternative enrichment tool (DAVID https://david.ncifcrf.gov/) using default settings. The program clusters similar gene ontology terms that are composed of overlapping gene sets. The table below shows the most significant gene ontology term for each of the top 10 enriched clusters of gene ontology terms.

Cluster	Term	Genes hit	Fold enrichment	P value (-log10)
1	Sterol biosynthesis	15	3.65	17.12
2	SH3 (SRC Homology 3) domain	15	3.65	3.89
3	Endocytosis	17	4.14	4.22
4	Response to wounding	27	6.57	3.54
5	Plasma membrane part	75	18.25	3.93
6	Apoptosis	29	7.06	3.38
7	Isoprene biosynthesis	3	0.73	1.96
8	Fatty acid biosynthetic process	10	2.43	4.04
9	Endosomal part	7	1.70	2.79
10	Ubiquitin like modifier conjugation	26	6.33	3.24

3.2.2 OxLDL regulated genes and CAD genetic risk

I explored the relationship between gene expression changes during foam cell formation and CAD genetic risk variants identified by GWAS. For this I identified genomic intervals containing all SNPs in high linkage disequilibrium (LD) ($r^2 > 0.8$) with known CAD-associated variants and then identified all genes within 50kb of these CAD risk locus intervals. These 132 genes at CAD risk loci were enriched for processes relevant to atherosclerosis, including lipid handling and foam cell formation (data not shown). I identified 17 of these genes as being regulated by oxLDL (microarray paired data analysis identifies 16 genes, and Interleukin-6 receptor (IL6R) was added as differentially expressed by RT-qPCR)(Table 7). These genes are of particular interest since they suggest that the GWAS CAD mechanisms could in part influence genes involved in foam cell formation. Within this list of GWAS candidate genes, several have no known role in foam cell formation. For example, Phosphatidic Acid Phosphatase Type 2B (PPAP2B) is not known to be expressed in human macrophages or foam cells but has a potentially relevant role since it dephosphorylates receptor active lipid species including lysophosphatidic acid [162]. Similarly, the Phosphatase and actin regulator 1 (PHACTR1) gene encodes an identically named protein that has a role in regulating the activity of a critical cellular phosphatase, protein phosphatase 1, but has no known role in foam cell formation [205].

Table 7 Differentially expressed genes within a 50kb radius of CAD loci.

Gene Symbol	Gene	Fold change (log2, foam cell relative to macrophage)
ANKS1A	ankyrin repeat and sterile alpha motif domain containing 1A	-0.325
COL4A1	collagen, type IV, alpha 1	0.510
CUX2	cut-like homeobox 2	-0.238
FAM109A	family with sequence similarity 109, member A	-0.445
FAM117B	family with sequence similarity 117, member B	-0.523
FES	FES proto-oncogene, tyrosine kinase	-0.311
HDAC9	histone deacetylase 9	-0.396
IL6R	interleukin 6 receptor	0.251
LDLR	low density lipoprotein receptor	-3.090
PHACTR1	phosphatase and actin regulator 1	0.949
PPAP2B	Phosphatidic acid phosphatase type 2B	1.65
PSRC1	proline/serine-rich coiled-coil 1	-0.229
SLC22A4	solute carrier family 22 (organic cation/zwitterion transporter), member 4	0.285
SLC22A5	solute carrier family 22 (organic cation/carnitine transporter), member 5	0.710
TMEM150A	transmembrane protein 150A	0.225
VAMP5	vesicle-associated membrane protein 5	-0.863
WDR12	WD repeat domain 12	0.310

3.3 Discussion

Treatment of cultured human M-CSF-induced macrophages with oxLDL causes phenotypic changes resembling those of foam cells pervading human atherosclerotic plaques [36, 179]. To understand these changes at the molecular level it is necessary to study the effect of oxLDL on gene expression. In the last 10 years, the development of whole genome approaches to surveying gene expression led to a number of studies comparing gene expression in macrophages and foam cells. Although these studies have provided useful insights a number of limitations required us to conduct our own expression studies. Firstly, most studies have been performed using either non-human cells or tumour cells that have some resemblance to human macrophages[206-212]. Several studies have also been published that include only abbreviated information with regards differentially expressed genes and lack publically deposited data[206, 207, 213]. Statistical approaches were evolving during the production of these studies and several studies omitted formal statistical tests of differential gene expression[182, 210]. Some studies used only two or three biological replicates[208, 212]. In all cases there was variation in the methodology of cell isolation, culture, cytokines, type of modified lipoprotein and data platform, statistical analysis and reporting. I therefore performed my own genome-wide expression analysis as part of a comprehensive assessment—including epigenomic analysis—so that there was consistency across all our experiments. Our expression study uses the most advanced platform of any study comparing macrophages and foam cells together with the industry standard statistical methodology and an appropriate number of replicates [160]. The present study also utilized 5 biological replicates, primary cells and was analyzed using what is widely

regarded as the most robust way of determining differential gene expression. As validation of the experimental model I showed that foam cells had abundant neutral lipid as measured by oil red O staining and that there was no overt effect on cell vitality by assaying the number of live and dead cells.

Comparison of gene expression in macrophages and foam cells identified a large number of differentially expressed genes, which in effect are CAD candidate genes. Analysis of enrichment for gene sets suggest the primal effect was on cellular lipid handling and many of the differentially expressed genes have well known roles in foam cell formation. Upregulated genes included sets of genes involved in processing lipids whereas the list of downregulated genes suggested a robust shutdown of any de novo cholesterol or sterol synthesis. Although not contained within the top 10 differentially expressed genes, other oxLDL regulated genes included haem-oxygenase 1 (HMOX1) which has an established role in the protective response to oxidant stress imposed by oxLDL stimulation [201]. Gene ontology enrichment analysis of upregulated genes highlighted the vital role of the cellular response oxidant stress.

Interestingly the list of differentially expressed genes included the CAD GWAS candidate genes PPAP2B and PHACTR1 which were also upregulated by oxLDL in the two external datasets used for comparison [179, 181]. The original published analysis by the authors of the other two studies did not list these genes as differentially expressed due to analytical limitations. This suggests that useful information could be extracted from existing microarray studies by re-analyzing them with the latest statistical software. In subsequent chapters, I investigate the role of PPAP2B and PHACTR1 in the genetic risk of CAD and their cellular roles in atherosclerosis.

Previous microarray studies have observed that foam cells are 'not simply a more pro-inflammatory version of a macrophage'[179]. This view is predicated on the relative paucity of classic inflammatory genes, like IL-1 that are upregulated by oxLDL[179]. Similarly, I also found that relatively few typical pro-inflammatory genes were upregulated by oxLDL. Indeed, oxLDL reduced the expression of the pro-inflammatory gene IL1B. However, I do not believe our gene expression data imply that foam cells are not pro-inflammatory. Firstly, there is a distinction between acute and chronic inflammation. Foam cells are part of a process in which there is chronic inflammation. The properties of chronic inflammation in relation to foam cells include cellular accumulation, abnormal mobility, deranged apoptosis and accumulation of partially processed lipids [36, 62]. For example, enhanced CD36 signaling is associated with prolonged residence of macrophages within plaques[214]. In addition, changes in gene expression after induction of foam cell formation do not necessarily reflect the release by foam cells of pro-inflammatory cytokines.

Comparison of our data to known expression profiles of lesional macrophages suggests that oxLDL induced foam cell formation is associated with more of an M2 phenotype. NOD-like receptor family, pyrin domain containing 3 (NLRP3) plaque expression is associated with an M1 profile whereas NLRP3 was down-regulated by oxLDL[214]. Nevertheless, analysis of surface markers and mRNA in plaque foam cells have not shown a conclusive polarization towards either the M1 or M2 phenotype[192].

Limited studies employing laser capture dissection of macrophages from mouse atherosclerotic plaques have measured gene expression *in vivo* but no studies exist in humans [215]. Advancements in the use of low cell numbers for RNA based

experiments may permit an adequate yield using human plaque that would allow comparison of gene expression profiles with the cellular model employed in the present study.

Overall I believe that our transcriptomic study provides a robust gene expression signature of foam cells and that individual differentially expressed genes require further investigation, prioritizing genes such as PPAP2B that are also implicated by genetic risk studies.

4 Epigenomic Effects of Atherogenic Lipid on Macrophages

4.1 Introduction

Development and maintenance of cell type identity depends on the binding of hundreds of different transcription factors to thousands of *cis*-regulatory elements [216]. Active *cis*-regulatory regions are typically characterized by nucleosome depletion, a necessary condition for DNA binding by many transcription factors[217-219]. A variety of techniques have emerged which allow mapping of this accessible open chromatin, such as DNAseI hypersensitivity or formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) [139, 220]. In parallel, chromatin immunoprecipitation (ChIP-seq) approaches have been developed to identify genomic sites where histones in nucleosomes that flank open chromatin have undergone post-translation modifications that indicate the function, such as enhancer function, of the open chromatin [143, 221]. The identification of cis-regulatory elements is useful in the study of functional genetic risk variants. Overlap of a risk variant with a functional regulatory element in a particular cell type indicates that the variant might exert its effect in that cell type by altering the activity of the regulatory element[220, 222, 223]. Several large-scale studies have demonstrated that transcriptional regulatory sites identified via high-throughput assays in specific human cell types are enriched for disease-associated GWAS variants, highlighting the importance of pinpointing cell type regulatory elements[128, 224].

4.1.1 Epigenomics and Atherosclerosis

Epigenetics refers to the study of functionally relevant changes to genomic DNA that do not alter its base sequence[225]. Studies employing a whole genome approach to epigenetics are termed epigenomic studies. Epigenetic data is most useful in the study of how gene expression is regulated across cell types and cellular conditions. Both reversible and stable epigenetic changes may be brought about by environmental stimuli and these epigenetic changes typically involve chromatin modifications. These changes facilitate or suppress the activity of regulatory elements such as enhancers and promoters which act together to activate or suppress gene transcription[225]. Promoters are crucial regulatory DNA elements in proximity to the transcription start site of genes where Polymerase 2 binds, in association with other transcription factors to initiate transcription[226]. Enhancers are distal regulatory elements which may be 100s of kbs away from gene start sites but which interact spatially and functionally to enhance transcription [227]. Enhancers are the most variable of regulatory elements and largely responsible for the fine tuning of gene expression across cellular states and in response to environmental stimuli[228]. Both enhancers and promoters are characterized by an area of open chromatin known as euchromatin; non-active regulatory DNA is characterized by closed chromatin known as heterochromatin[217, 229]. The study of both chromatin landscapes and modifiers of chromatin landscapes is a major area of contemporary research.

The chromatin state at regulatory elements is in part regulated by enzymatic modifications to histone tails including, for example, acetylation and de-acetylation [227]. Active enhancers are characterized by changes to histones in the neighboring

nucleosomes including acetylation of lysine 27 in H3 histones. This change is brought about by histone acetylase enzymes and together with other histone modifications forms what is termed the 'histone code' [230]. External cellular influences may activate certain transcription factors that cause a change in the balance of histone modifications including acetylation and deacetylation, which in turn governs the overall accessibility of an open chromatin site. The most efficient of these are termed 'pioneer' transcription factors since they can bind relatively closed chromatin—examples of pioneer factors include activating protein 1 (AP1) and CCAAT enhancer binding proteins [219]. On the other hand many transcription factors, including the glucocorticoid receptor (GR), a classic nuclear factor, bind at pre-established open chromatin site that remain largely unaltered by GR binding [231]. In breast cancer a pioneer transcription factor, PBX1, modifies the open chromatin landscape and effectively lays the ground for the oncogenic effects of excessive estrogen receptor transcription factor DNA binding [232].

Foam cells and atherosclerotic plaques appear so abnormal on macroscopic examination that it can readily be hypothesized that epigenetic changes may occur that both contribute to and reflect the diseased state. Numerous studies have identified epigenetic changes in atherosclerosis. DNA methylation involves addition of a methyl group to cytosine residues in a CpG dinucleotide context and is associated with suppression of gene expression[225]. Coronary artery disease and cardiovascular death are both associated with a global increase in DNA methylation of peripheral blood cells [233, 234]. *In vitro* experiments have demonstrated that pathogenic stimuli such as shear stress also alter methylation of DNA in endothelial cells. Drugs that inhibit histone deacetylases, including trichstatin A have been shown to exacerbate

atherosclerosis in animal models[235]. *In vitro* studies have shown that atherogenic stimuli such as shear stress affect the epigenetic landscape and regulate gene expression[236]. Short term exposure of human monocytes to oxLDL induced a persistent inflammatory and pro-atherogenic state that was associated with chromatin modification at the promoters of several genes involved in inflammation and atherosclerosis[69]. Taken together, these studies suggest that atherosclerosis involves epigenetic changes and mandate a systematic study of epigenetic changes in response to atherogenic stimuli.

4.1.2 Integrating Epigenomics with Genome-wide association studies

Genome-wide association studies identify regions of the genome where genetic variants, most commonly SNPs, affect the odds of having a given trait or disease[237]. Mostly GWAS studies cannot resolve the causative SNP among many in high linkage disequilibrium. This is a barrier to performing mechanistic and functional studies exploring the cellular effect of disease-associated genetic variation[237]. In some cases, a combination of fine mapping and imputation has refined the probability of a given SNP being the causal variant. For example, at the MTNR1B locus associated with fasting glycaemia, there is a 0.999 probability of rs10830963 being the causal variant [238]. For the majority of known trait-associated loci there is insufficient genetic evidence to confirm which of many variants in linkage disequilibrium are likely to be causal. Another successful method in refining the list of potentially causal genetic variants is to filter out those which do not alter a coding sequence or are within a DNA regulatory element. The widespread application

of epigenomic techniques to map DNA regulatory elements across multiple cell types has afforded the opportunity to systematically annotate disease-associated SNPs with regulatory function. Indeed, numerous studies have identified enrichment of disease-associated SNPs for enhancers and studies of individual SNPs have shown they alter enhancer function by altering transcription factor binding[128, 129, 131, 239]. A large scale genetic and epigenetic fine mapping study for 21 autoimmune disease found that 90% of causal variants were in non-coding DNA and 60% mapped to enhancers[240]. Enrichment for regulatory DNA for a given trait has also been found to be cell and tissue specific, further refining the site of action of risk variants[128]. For example, variants associated with type 2 diabetes risk are enriched for enhancers in pancreatic islets[222]. A variant within an islet specific enhancer was shown to disrupt binding of the pancreatic specific transcription factor NeuroD1 and abrogate enhancer activity[222].

A similar approach of using epigenomic maps to pin-point functional variants could be useful in coronary artery disease. However, open chromatin and enhancer maps were not existent for two of the key cell types associated with atherosclerosis, macrophages and macrophage-derived foam cells. In this chapter I describe the production of epigenomic maps in these cells and use them to identify regulatory DNA associated CAD variants.

4.1.3 Epigenomic techniques

Several techniques have been developed which allow a genome-wide assessment of open chromatin status or histone modifications. Formaldehyde assisted isolation of regulatory elements with high throughput sequencing (FAIRE-seq) is a method for

enriching DNA that is in ‘open chromatin’ and then subjecting the enriched DNA to next-generation sequencing [139, 220, 241]. Open chromatin is characterized by displacement of nucleosomes and increased accessibility of DNA for the binding of sequence specific transcription factors. In a FAIRE-seq experiment formaldehyde is applied to cells or tissue and forms reversible protein-DNA crosslinks between histones and DNA. This occurs preferentially at “closed chromatin” sites where the DNA is tightly wrapped around histones[139]. The sample is then sonicated to break up the genomic DNA into, typically, 250bp pieces and then subjected to several phenol-chloroform extractions[139]. DNA that is unbound by histones—open chromatin DNA—preferentially localizes to the aqueous phase and can then be subjected to next generation sequencing[139]. Numerous studies have successfully used FAIRE-seq to identify regulatory DNA across a range of cell types and overall there is a close correlation with open chromatin maps derived using alternative techniques such as DNAse hypersensitivity assays[217, 232, 242]. FAIRE-seq has a number of potential advantages compared to DNAse-seq including requiring fewer cell numbers, and rapid fixing of cells to instantly freeze and capture their cellular state rather than subjecting cells to various manipulations prior to harvesting the genomic DNA—as is the case for DNAse seq[139, 243]. The principal limitation of FAIRE-seq is that the enrichment process generates relatively noisy data with many reads mapping to sites that are not in open chromatin sites. This causes difficulties in the downstream analysis of genuine open chromatin sites and in the comparison between replicates [139]. To overcome these problems multiple replicates can be assessed for enrichment using RTqPCR with amplicons in known open and closed chromatin sites and multiple replicates can be combined to help boost genuine signal and cancel out noise.

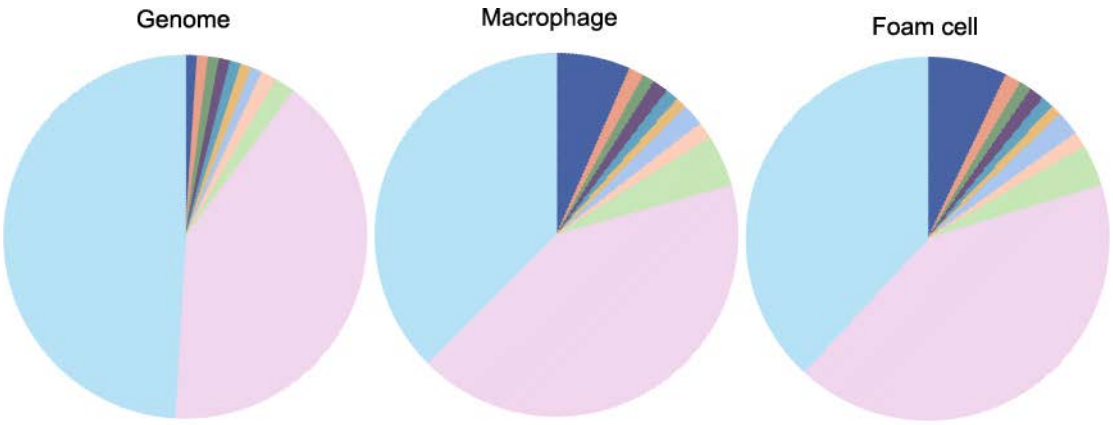
Post-translational histone modifications, including acetylation of lysine residue 27, correlate with regulatory DNA activity. H3K27ac marks identify active enhancers and promoters[221]. Chromatin immunoprecipitation with high throughput sequencing involves similar cross-linking steps to FAIRE-seq but enrichment of H3k27ac-bound DNA is performed using immunoprecipitation with an anti-H3k27ac antibody. The H3K27ac mark is particularly useful since it has been shown to vary not only with cell ontogeny but also with subtler changes in cellular state induced by cytokines and growth factors[244, 245]. No genome-wide studies have evaluated the effect of oxLDL on the enhancer landscape in atherosclerosis related cell types. Finally, ChIP-seq can be performed to identify specific transcription factor binding sites. Multiple layers of epigenomic data can then be overlaid and compared between different cell types and cell states to identify the regulatory DNA sites that govern mRNA expression and hence cellular phenotypes—across cells which share the same underlying genomic DNA sequence.

In this study I employ FAIRE-seq and ChIP-seq assays to show that exposure of primary human macrophages to oxLDL causes changes in chromatin accessibility and histone modification at a subset of regulatory sites. I then demonstrate that these changes are correlated with local expression of genes involved in foam cell and atherosclerotic processes and also binding of key transcription factors such as C/EBP-beta

4.2 Results

4.2.1 OxLDL induces major changes in chromatin accessibility in macrophages

To understand the transcriptional pathways mediating the effect of oxLDL on gene expression, I identified candidate regulatory DNA elements that underwent oxLDL-induced changes in chromatin structure. I began by generating genome-wide maps of open chromatin in primary human macrophages before and after oxLDL exposure, with consequent foam cell formation (3 biological replicates were generated for each condition). Formaldehyde-assisted isolation of regulatory elements with high-throughput sequencing (FAIRE-seq) identified 130,491 and 123,400 nucleosome-depleted open chromatin sites in macrophages and foam cells respectively (Data available at http://www.ncbi.nlm.nih.gov/geo/ under Super Series GSE54975S5) [139, 217, 246]. The overall number of sites identified by our analysis is similar to the 100,000 to 225,000 open chromatin sites found in other cell types [217]. Sites at promoters, defined as ≤ 1kb upstream of the Refseq transcriptional start site (TSS), accounted for 6.1% of open chromatin sites in macrophages and 7.1% in foam cells (Figure 4-1).



Feature (bp)	Genome (%)	Macrophages (%)	Foam cells (%)
Promoter (<1000)	1.1	6.6	7.1
Promoter (1000-2000)	0.7	1.3	1.4
Promoter (2000-3000)	0.6	1.1	1.1
Downstream (<1000)	0.9	1.4	1.3
Downstream (1000-2000)	0.7	1.2	1.1
Downstream (2000-3000)	0.6	1	1
5'UTR	0.4	1.9	2.1
3'UTR	1.4	1.5	1.4
Coding exon	1.9	4.6	3.7
Intron	41.6	42	41.6
Distal Intergenic	50	37.4	38.1

Figure 4-1 Genomic distribution of open chromatin sites in macrophages and foam cells.

FAIRE-seq derived open chromatin sites in macrophages and foam cells were annotated using CEAS and the background genome distribution derived from the percentage of genome covered by RefSeq annotations.

To gain an overview of the similarities and differences of macrophage and foam cell open chromatin sites I compared them with other cell types for which FAIRE-seq data was available. I downloaded the ENCODE FAIRE-seq data and identified open chromatin sites using F-seq with a liberal threshold such that over 125,000 peaks were called per cell type. The signal strength of the open chromatin sites was used to rank them in descending order. The first 25,000 open chromatin sites contain peaks with the greatest signal strength—most open chromatin—and the last group of 25,000 peaks (between 100,000th and 125000th peaks) contain the weakest peak for each cell type. All foam cell open chromatin sites were compared against all the other cell types to produce a bubble plot showing that the greatest overlap is for stronger sites and that foam cells sites overlap most with macrophage sites (Figure 4-2). For some cell types, such as islets, however, the actual number of genuine FAIRE-seq derived open chromatin sites is less than 25,000 and the small number of overlaps in weaker islet sites may be stochastic. I then annotated individual open chromatin sites to RefSeq genomic annotations using the CEAS software tool. I then calculated the proportion of open chromatin sites for the various cell types that overlapped foam cell open chromatin sites (Figure 4-3). Overall across the different categories macrophage open chromatin sites showed the greatest proportion of overlap with foam cells. In terms of genomic annotations, promoter sites and 5'UTR sites had by far the greatest similarity across the different cell types. By comparison only about 10% of intronic open chromatin sites were shared with any given cell type (apart from the macrophage/foam cell comparison. Across the two different comparison strategies used above, the next closest cell type in terms of overlapping open chromatin sites with macrophages and foam cells is GM18507, a Nigerian lymphoblastoid transformed B-cell line.

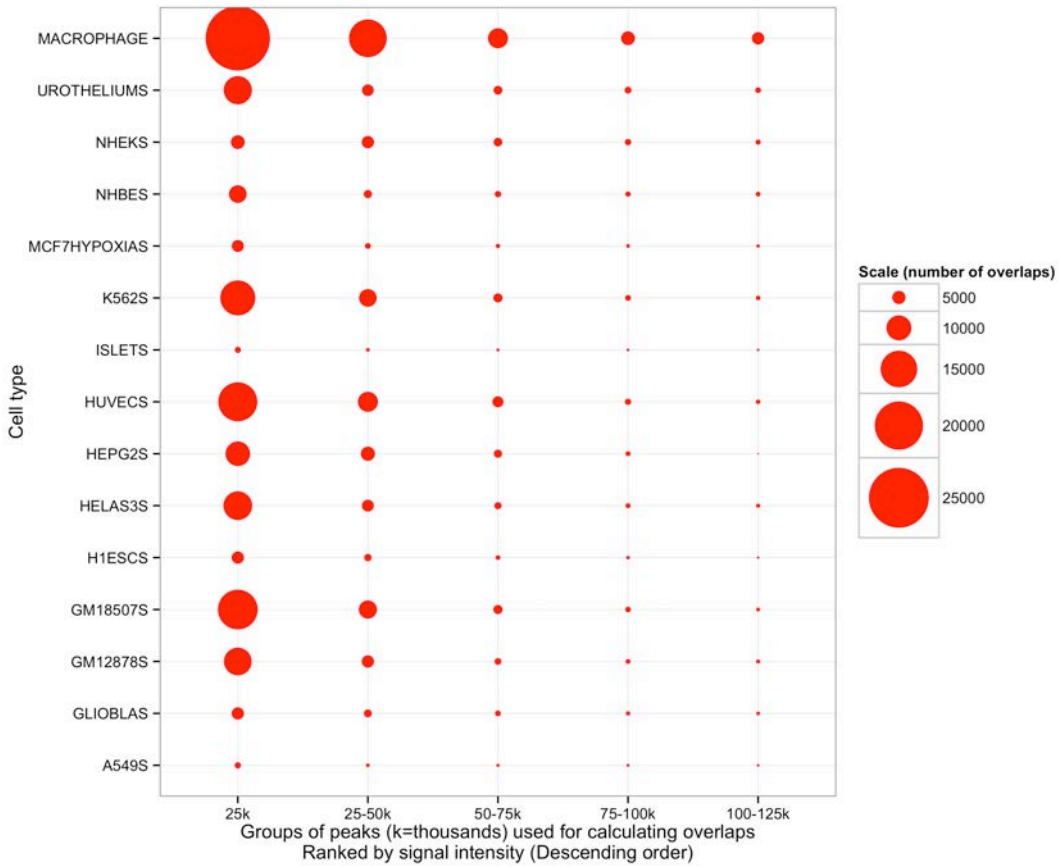


Figure 4-2 Comparison of overlapping open chromatin sites in different cell types.

Open chromatin sites for several ENCODE cell types and macrophages were identified and ranked by signal strength in descending order. The sites were then overlapped with foam cell sites to show the number of overlapping open chromatin sites for each signal strength category (Up to 125,000 peaks were called using a liberal Fseq peak calling threshold and no additional filtering of peaks was performed).

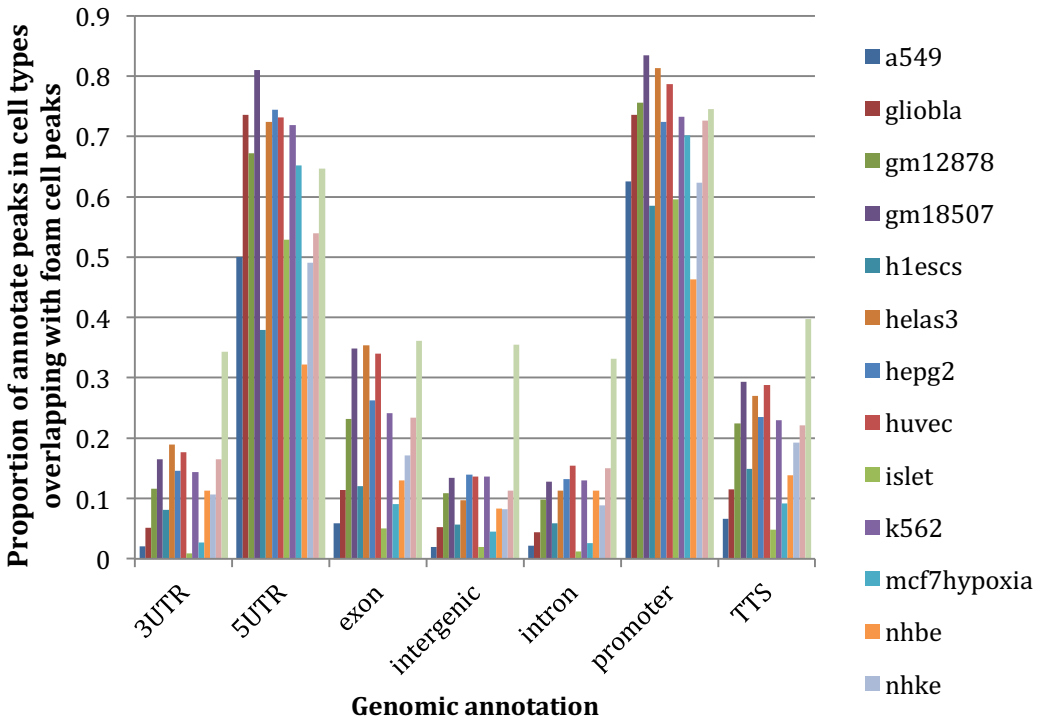


Figure 4-3 Proportion of overlap between different types of open chromatin sites

Open chromatin sites were annotated with RefSeq genomic annotations and the proportion of sites overlapping foam cell open chromatin sites was calculated.

The promoters of genes with a low level of expression had low FAIRE signal consistent with a relatively nucleosome-bound, closed chromatin profile. By contrast, the promoters of maximally expressed genes had a high FAIRE signal consistent with nucleosome-depletion (Figure 4-4, Figure 4-5) [217]. Genes whose expression was altered by oxLDL exposure had an intermediate chromatin profile at their promoters. OxLDL significantly increased nucleosome depletion at the promoters of up-regulated genes, ($p = 1.08 \times 10^{-27}$) and reduced it at the promoters of down-regulated genes ($p = 5.42 \times 10^{-21}$) indicating its capacity to trigger chromatin remodeling (Figure 4-4, Figure 4-5).

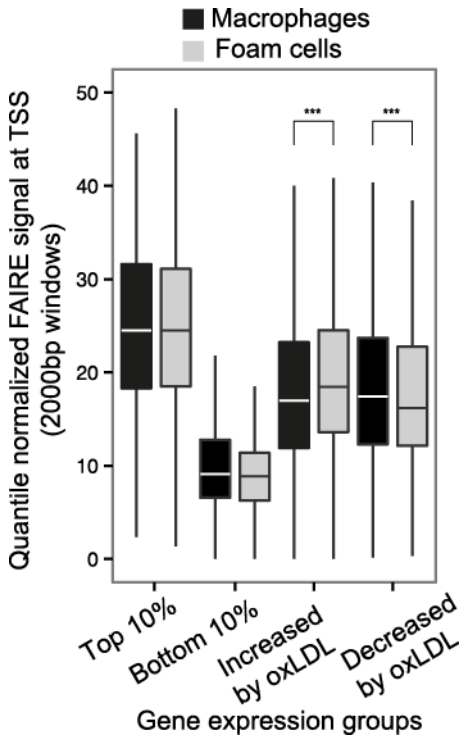


Figure 4-4 OxLDL-induced changes in chromatin structure determined by FAIRE-seq at gene transcription start sites.

Normalized FAIRE-seq signal in 2000bp windows centered on the TSS for gene sets grouped by their expression characteristics. OxLDL-induced changes in open chromatin were associated with concordant changes in gene expression (higher FAIRE signal indicates more open chromatin, *** $p < 0.0005$).

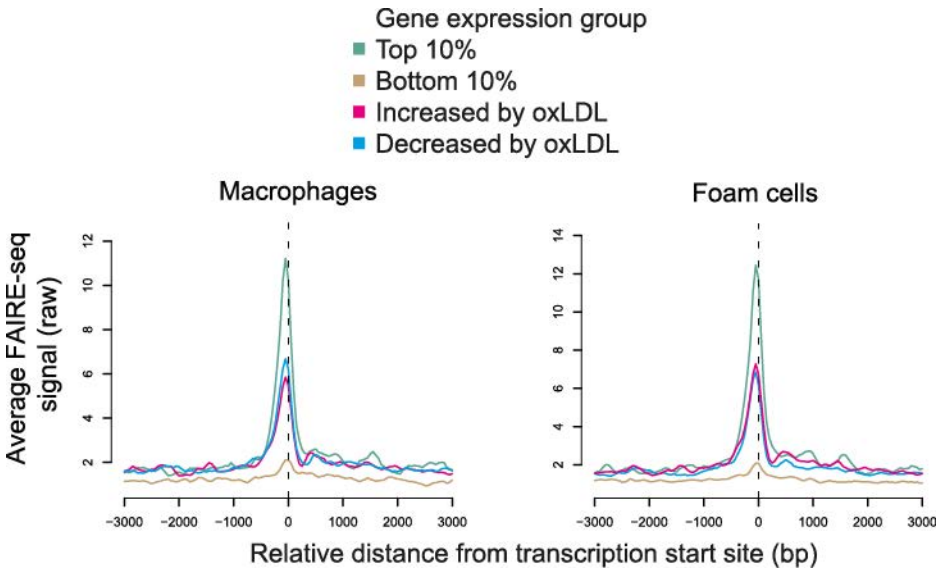


Figure 4-5 FAIRE-seq profile at TSS for different gene sets.

Profile view of FAIRE-seq signal around TSSs showing sharply delineated promoter open chromatin profiles and their relationship with expression.

All genomic sites with significant oxLDL-induced changes in chromatin profile were identified and termed dynamic chromatin sites (Figure 4-6). OxLDL changed the chromatin profile at around 10% of all open chromatin sites (13,516 sites of which 12,754 (94%) were non-promoter sites (defined as 1kb from a known RefSeq TSS). 7,276 (54%) of these 13,516 sites had a more nucleosome-depleted profile and 6,240 had a more nucleosome-bound profile in foam cells (Figure 4-7). The size distribution of the dynamic chromatin sites demonstrates that changes at most non-promoter sites involve displacement of a single nucleosome, whereas those at promoter sites may involve displacement of one or two nucleosomes (Figure 4-8). Overall, only a fraction of macrophage open chromatin sites—the majority distal to known promoter regions—are changed in response to oxLDL.

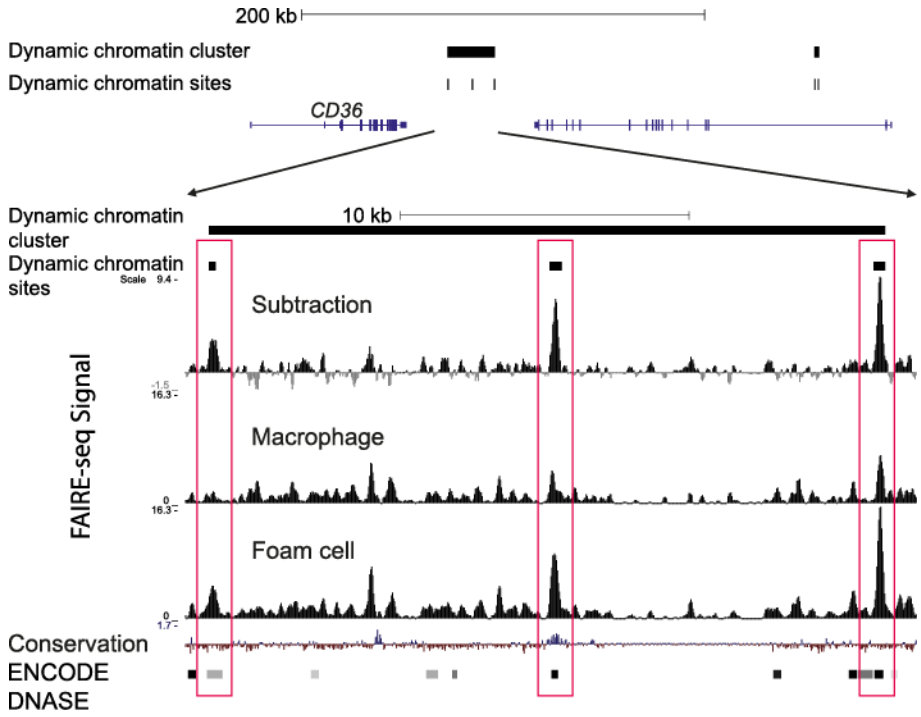


Figure 4-6 Representative chromatin profile at the *CD36* gene locus showing a dynamic chromatin cluster.

The dynamic chromatin cluster comprises three dynamic chromatin sites each with higher signal in foam cells. CD36 is upregulated in foam cells compared to macrophages. The lower panel close-up view shows the relevant FAIRE-seq signals for macrophages, foam cells and the normalized dynamic signal (foam cell minus macrophage signal). Dynamic sites are highlighted in red boxes and are known DNAse hypersensitivity sites in other cell types in the ENCODE database; one site shows conservation in vertebrates.

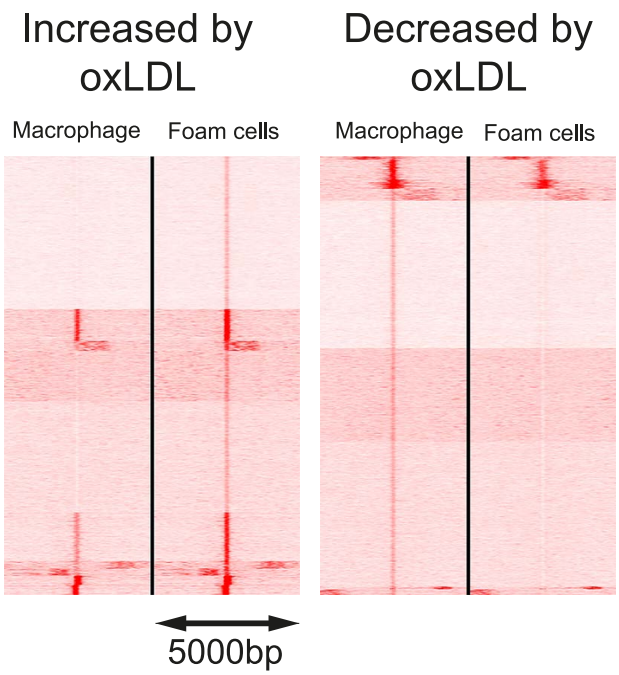


Figure 4-7 Chromatin signal at dynamic chromatin sites.

Heat map of open chromatin profile at all dynamic chromatin sites, split into panels showing sites with increased or decreased FAIRE-seq signal. Each line represents a 5kb window centered on an individual site. Sites are clustered into groups with similar signal patterns.

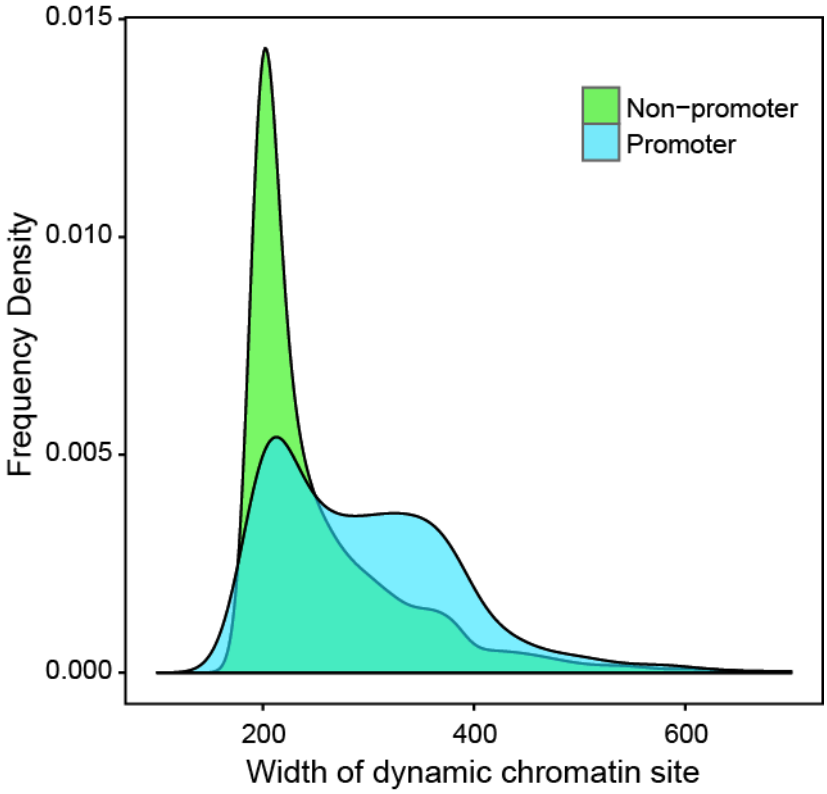


Figure 4-8 Width of dynamic chromatin sites (bp).

The frequency distributions of widths of dynamic chromatin sites suggest most are not more than one nucleosome in width with a relative increase in wider sites at promoters.

To externally validate our dynamic chromatin sites and show they represent true regulatory DNA sites we checked for typical characteristics of regulatory DNA. First, we tested the extent to which these sites overlapped regulatory elements identified by the ENCODE project, compared to control DNA sequences randomly chosen from within 10kb of each site (see Methods). Dynamic sites were most prominently enriched for overlap with enhancer and TFBS elements ($p < 0.01$) (Figure 4-9). Second, I determined whether the dynamic chromatin sites were evenly distributed across the genome. I observed significant clustering ($p < 0.05$) of dynamic sites

suggesting a non-random distribution of these sites. Gene set enrichment analysis of the clustered sites identified processes relevant to foam cell formation and cellular lipid handling (Figure 4-10). Further, I performed motif analysis using the genomic sequence within dynamic chromatin sites. I identified several motifs enriched in these sequences, including JunD/AP1, NFE2L2 and EGR1 (Table 8). These results suggest dynamic chromatin genuine regulatory DNA sites.

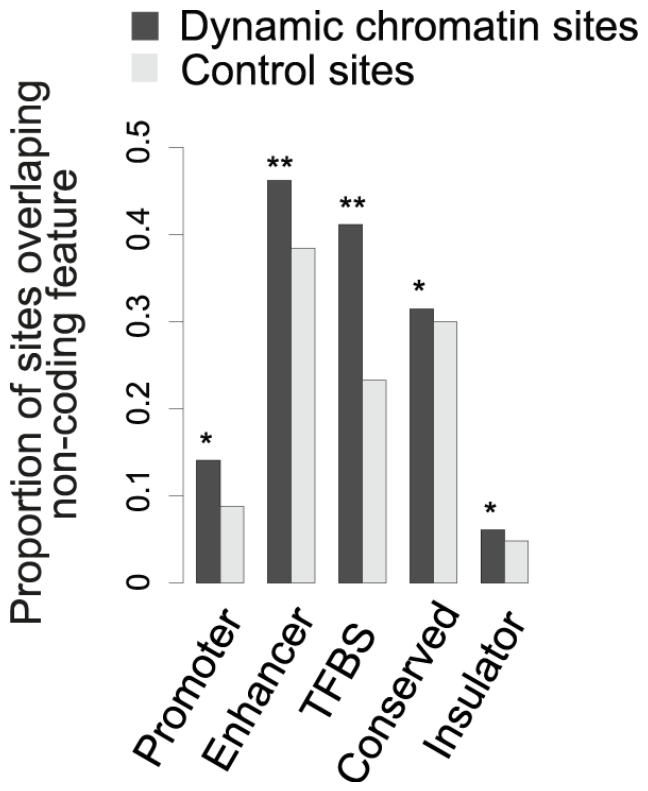


Figure 4-9 Dynamic chromatin sites are enriched for non-coding regulatory elements found in 9 other cell types.

We validated our FAIRE-seq derived dynamic chromatin sites by overlapping it with existing ENCODE data and showing that dynamic sites were enriched for known

regulatory elements (* p < 0.05, ** p < 0.005, TFBS—transcription factor binding site, see Methods for derivation of control sites)

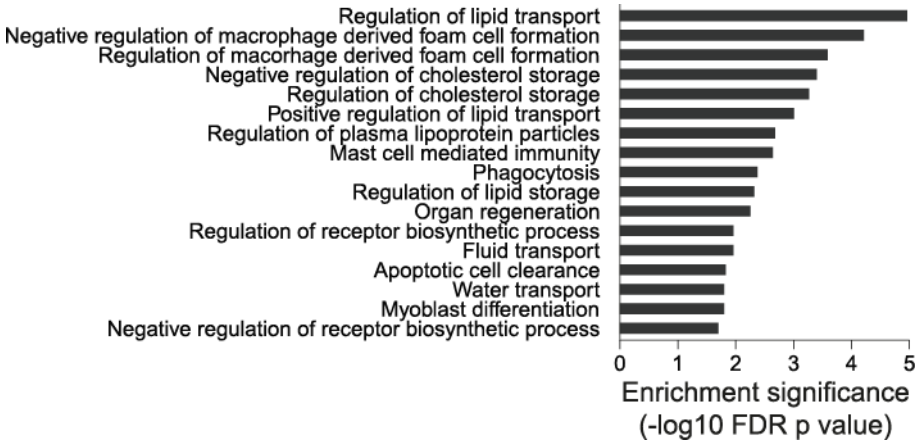


Figure 4-10 Dynamic chromatin clusters are enriched for gene sets involved in foam cell formation and lipid handling.

Table 8 Top 10 enriched transcription factor binding site motifs in top 3000 (ranked by average signal intensity) dynamic open chromatin sites.

Motif enrichment finding was performed using the SeqPos tool that uses a genomic background to test for enrichment. Similar motifs were clustered using the SeqPos algorithm to avoid reporting redundant motifs and the top motif from each of the top 10 clusters is shown below.

Rank	Matrix	Trnscription factor	DNA binding domain	Hits	P value (-10*log10)	Motif
1	MC00456	JUND	Leucine Zipper Family	1213	690.776	
2	MS00288	TFE3	Helix-Loop-Helix Family	1569	690.776	
3	MC00339	NFE2L2	Leucine Zipper Family	1318	690.776	
4	MC00319	EGR2	BetaBetaAlpha-zinc finger Family	2289	690.776	
5	MA0067	PAX2	Homeodomain Family	1668	690.776	
6	MC00163	Creb1	Leucine Zipper Family	1640	690.776	
7	MC00337	Ikzf3	BetaBetaAlpha-zinc finger Family	976	688.093	
8	MS00302	CEBPG	Leucine Zipper Family	522	687.539	
9	MS00223	SP3	BetaBetaAlpha-zinc finger Family	2377	671.922	
10	denovo16	EGR1	BetaBetaAlpha-zinc finger Family	2970	631.408	

At promoter sites I demonstrated that FAIRE-seq signal intensity was associated with gene expression in macrophages and foam cells. To compare changes in FAIRE-seq signal at dynamic chromatin sites with changes in gene expression I annotated each dynamic chromatin site to the nearest differentially expressed gene. I then compared the expression change and mean change in FAIRE-seq signal for chromatin peaks annotated to each gene. Overall there was a weak positive correlation between changes in FAIRE-seq signal and change in gene expression in macrophages and foam cells (Figure 4-11). The positive relationship was statistically significant for genes annotated with chromatin sites that on average had higher signal in response to oxLDL ($P = 0.042$) but was not significant for down-regulated dynamic chromatin sites ($P = 0.58$). This suggests that activation of open chromatin sites by oxLDL is more closely related to gene expression than suppression of open chromatin sites. A limitation of this approach is that open chromatin sites are comprised of different types of regulatory elements—in the next section I integrate the open chromatin data with the enhancer landscape in macrophages and foam cells.

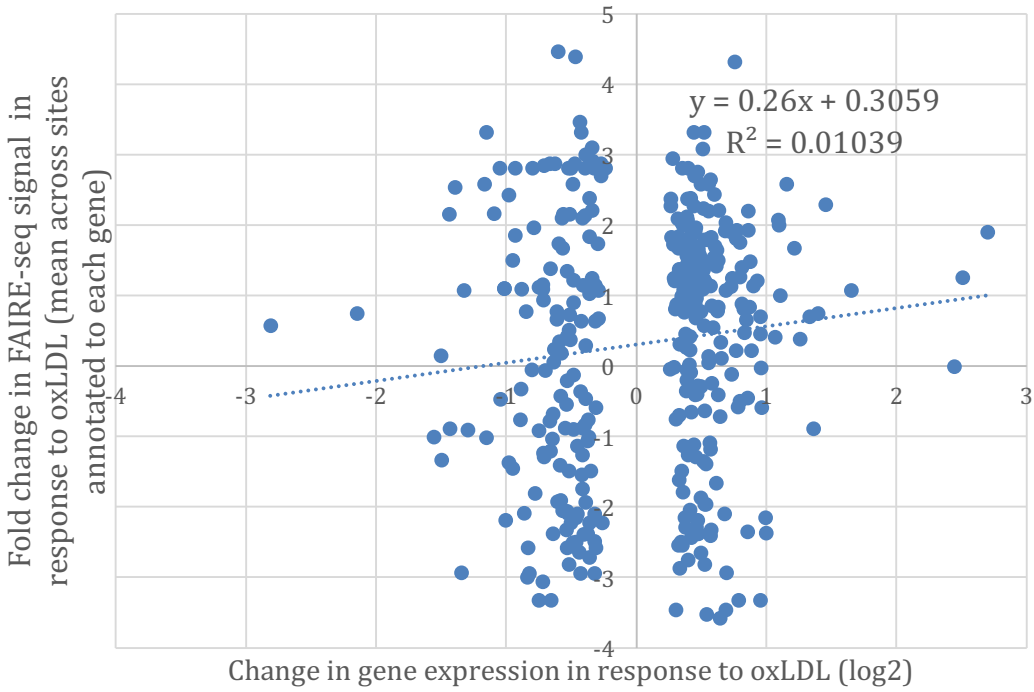


Figure 4-11 Correlation between change in gene expression and change in FAIRE-seq signal at dynamic open chromatin sites.

All dynamic open chromatin sites were annotated to the nearest differentially expressed gene. The mean fold change in FAIRE-seq signal for dynamic chromatin sites annotated with the same gene was calculated and plotted vs the log2 fold change in gene expression in response to oxLDL.

4.2.2 Macrophage enhancers undergo major oxLDL-induced changes

As the majority of dynamic chromatin sites were distal to known promoters, I hypothesized that many of these sites represented enhancer elements involved in the macrophage response to oxLDL and thus in foam cell identity. We, therefore, mapped the enhancer landscape in macrophages and foam cells using ChIP-seq for the histone modification H3K27ac, which is known to mark active enhancers [221, 244]. I observed significant oxLDL-induced changes in H3K27ac signal at 39,194 sites, termed 'dynamic enhancer sites', 95% of which were distal to known promoter regions (Figure 4-12). I found a strong positive correlation between the change in enhancer signal and the change in expression of the nearest differentially expressed gene ($r^2 = 0.45$; $p < 2.2x10^{-16}$) (Figure 4-13A/B). I identified 'super-enhancer' elements, which are spatially clustered enhancers that, in other contexts, are relevant to the expression of cell-type specific genes [247]. I identified 1,081 and 1,098 super enhancers in macrophages and foam cells, respectively (Figure 4-14). Genes overlapping or adjacent to super enhancers had significantly higher expression than those associated with other enhancers (Figure 4-14). Of the 1098 foam cell super-enhancers, 213 super-enhancers were only found in foam cells and genes nearest to or overlapping them had significantly higher expression in foam cells compared to macrophages ($p = 4.61x10^{-10}$). The analysis of enhancers indicates that oxLDL-induced changes in gene expression are mediated in large part through changes in local enhancer activity.

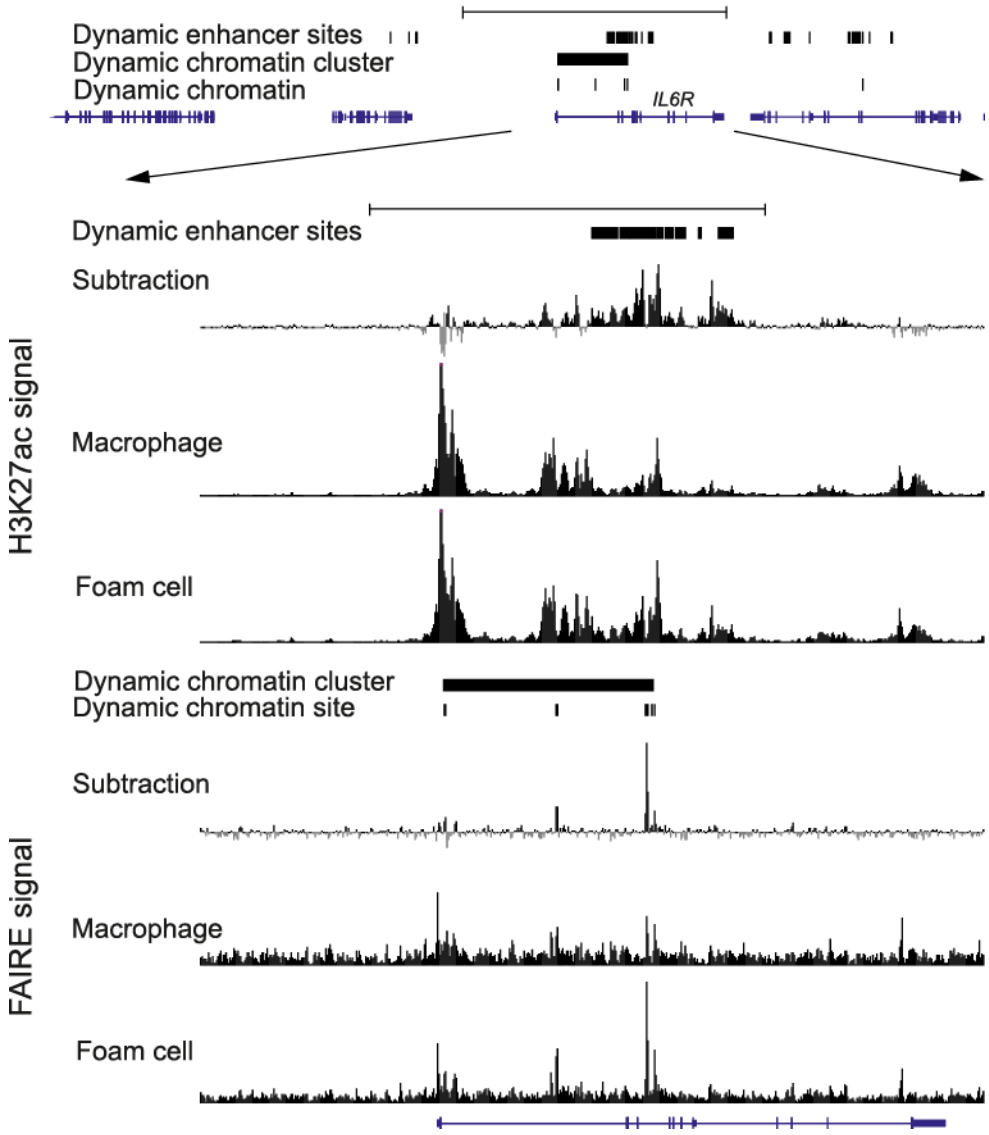


Figure 4-12 Characterization of the dynamic chromatin and enhancer landscape in macrophages and foam cells.

Representative chromatin profile at the IL6R locus (a genetic CAD risk locus) showing a dynamic chromatin cluster comprising four dynamic chromatin sites and a large dynamic enhancer region. The lower panel close-up view shows the relevant FAIRE-seq and enhancer signal for macrophages, foam cells and the normalized dynamic signal (foam cell minus macrophage signal).

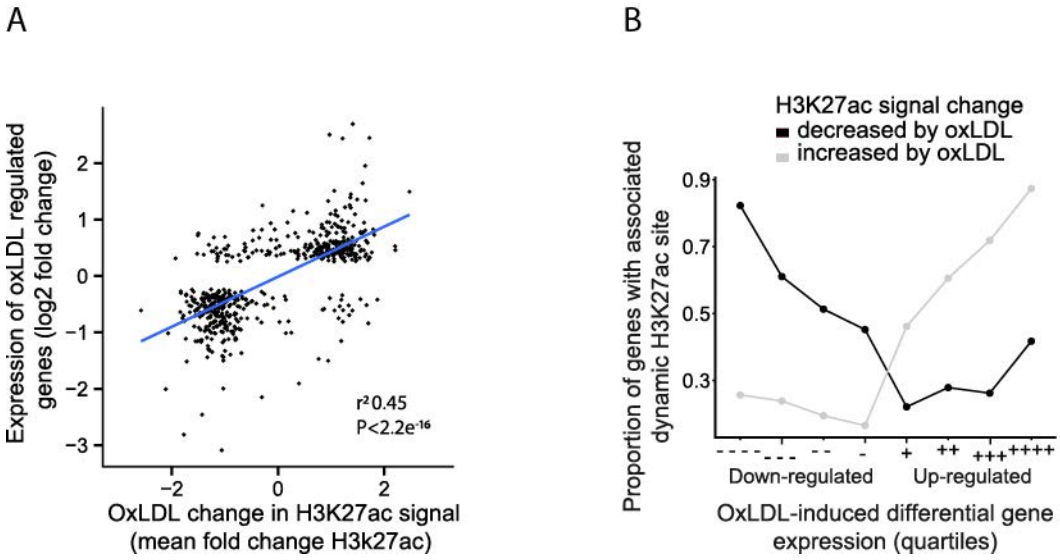


Figure 4-13 OxLDL regulated enhancer activity is correlated with gene expression.

(A) Correlation between fold change in gene expression for differentially expressed genes and mean fold change in enhancer signal for dynamic sites annotated to their nearest gene (r^2 = 0.45, Pearson r = 0.67). (B) The proportion of the genes in each quartile that are associated with a dynamic enhancer site is plotted. Differentially expressed genes were split into oxLDL up and down regulated genes and further split into quartiles based on fold change in expression (-/+ and ----/++++ indicate least and most differential quartiles, respectively). The proportion of differentially expressed genes in each quartile that are associated with a directionally concordant change in a dynamic enhancer site is correlated with the magnitude of gene expression change.

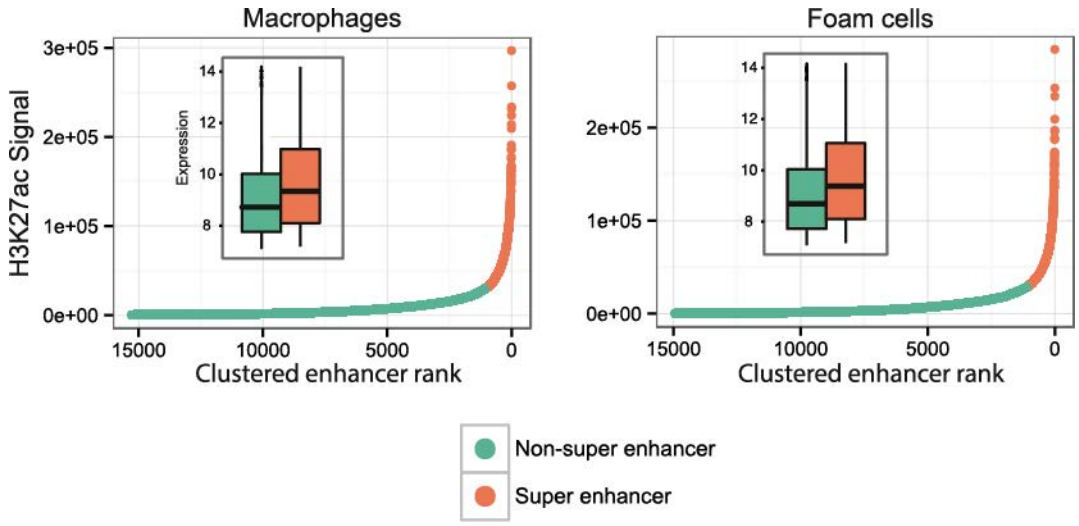


Figure 4-14 Super-enhancers in macrophages and foam cells.

Enhancers in macrophages or foam cells were aggregated if within 12.5kb and plotted according to H3k27ac signal rank (X axis indicates the ranks of these aggregated enhancers described as 'clustered enhancers'. The inflection point of the curve describing the relationship between clustered enhancer rank and overall enhancer signal was arbitrarily used (using the protocol established in the literature—see methods) as the cutoff point for describing a clustered enhancer as having or not having 'super' status. The insets show gene expression (log2, normalized) for genes overlapping or adjacent to either enhancers within super enhancer domains or other enhancers (Welch two sample t-test: macrophage gene expression enhancers vs super enhancers p = 2.8x10-29, foam cell gene expression enhancers vs super enhancers p = 1.9x10-33)

The method used to identify super enhancers in macrophages and foam cells was identical to that used to identify super enhancers in 86 different human cell and tissue types[146]. I was therefore able to compare the landscape of super enhancer in

macrophages and foam cells with these other datasets as well as looking for patterns of transcription factor binding common to super enhancers across all the datasets. Hierarchical clustering showed that the macrophage and foam cell samples were most closely related with each other and the next closest cell type was unsurprisingly the cells from which they are derived—$CD14^+$ monocytes (Figure 4-15). The transcription factors PU.1 and CEBPB, both known to be important in myeloid development clustered adjacent to each other and showed greatest similarity in binding patterns among $CD14^+$, macrophage and foam cell super enhancers [248, 249]. Interestingly visual inspection of the heat map also indicated a striking correlation between super enhancers in the H1 embryonic stem cell and the transcription factors NANONG and POU5F1, both of which are required for conferring embryonic stem cell status.

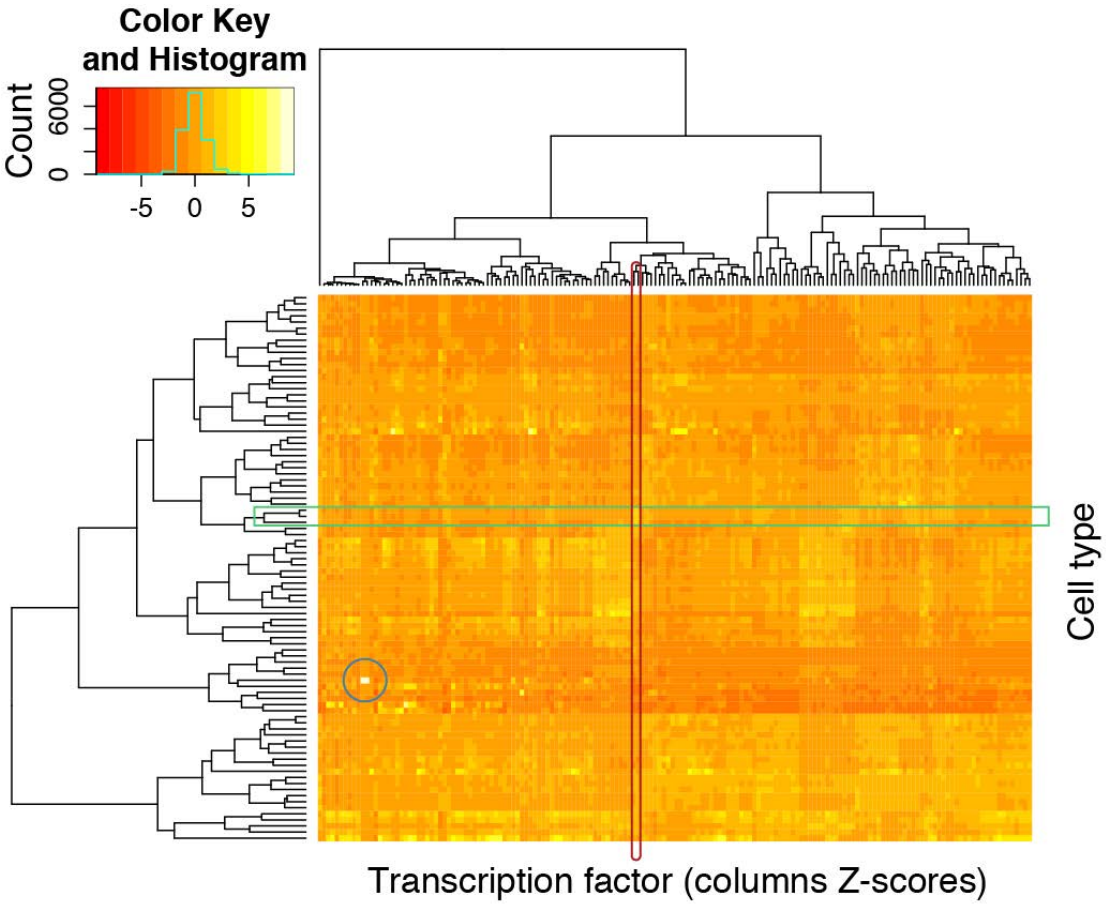


Figure 4-15 Comparison of super-enhancer in macrophages and foam cells with 86 other cell types and tissues.

Heat map showing hierarchical clustering of super enhancers in a range of cell types and ENCODE ChIP-seq derived transcription factor binding sites by overlapping each dataset and calculating the Jaccard statistic using Bedtools. Super-enhancer dendrograms on the Y axis; transcription factor ChIP-seq dendrograms on the X axis. The data has been scaled column-wise such that each column is a z-score, so that transcription factors comparisons are effectively normalized). Horizontal box indicates rows corresponding to macrophages, foam cells and CD14 cells in descending order. The vertical red box corresponds to CEBPB and PU1 transcription factors. The circle indicates the transcription factors NANONG and POU5F1 at the centre for the H1 embryonic stem cell type.

I next identified sites that had significant oxLDL-induced changes in both chromatin signal and enhancer signal. 1743 dynamic chromatin sites overlapped one or more of 2004 dynamic enhancer sites. Of these 2004 enhancer sites there was concordance in the direction of change in enhancer and open chromatin signals in 1817 sites (90.6%). These 1743 dynamic chromatin sites were, in turn, significantly associated with a directionally concordant change in expression of the nearest genes and chromatin signal (oxLDL-induced chromatin sites $p = 6.9x10^{-27}$, oxLDL-suppressed chromatin sites $p = 0.0004$, Figure 4-16A). I also observed enrichment among genes nearest to these 1,743 dynamic sites for gene sets related to atherosclerotic vascular disease (Figure 4-16B). These results demonstrate that a discrete number of sites with changes in both chromatin accessibility and enhancer signal are associated with disease-relevant changes in gene expression.

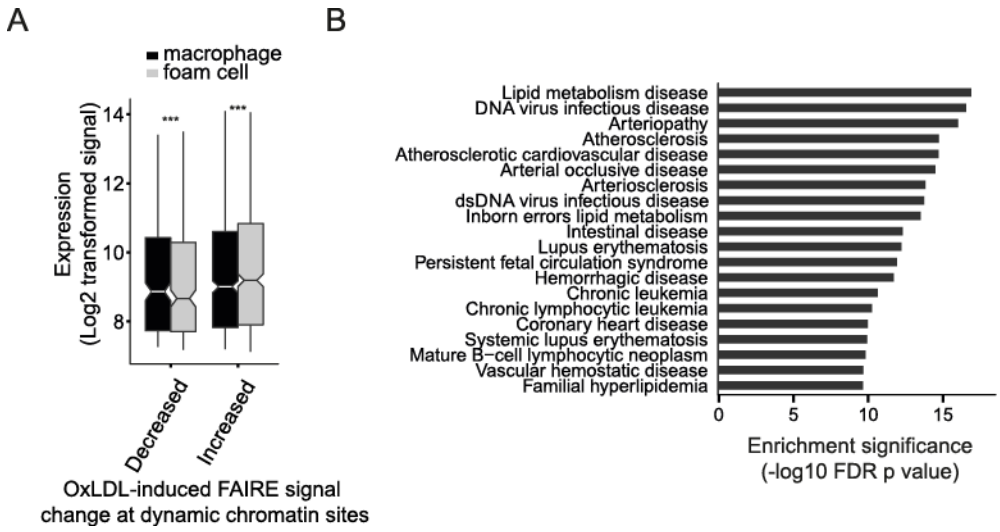


Figure 4-16 OxLDL induces changes in both open chromatin and enhancer status at a subset of sites with important regulatory function.

(A) The oxLDL regulated change in open chromatin—at sites which also have changes in enhancer signal—is correlated with expression of the nearest gene (*** $p < 0.0005$). (B) The refined subset of sites with dynamic open chromatin and dynamic enhancer status shows enrichment for proximity to disease-associated gene sets linked with atherosclerosis.

4.2.3 Dynamic enhancer sites identify C/EBP-beta as a novel oxLDL-regulated transcription factor in macrophages

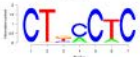
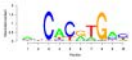
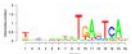
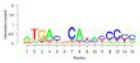
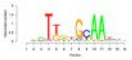
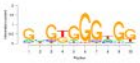
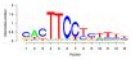
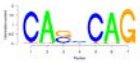
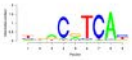
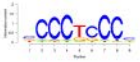
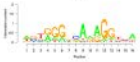
I used the dynamic chromatin and enhancer sites to identify potential upstream transcriptional regulators driving oxLDL-induced changes in macrophage chromatin and gene expression. I performed motif enrichment analysis using the genomic sequence of the 1,743 sites with changes in both chromatin and enhancer signal (Table 9). I observed enrichment for AP1-related factors, such as JunD, which bind enhancer elements, and NFE2L2, which is involved in the anti-oxidant response[57, 219]. I also observed enrichment for factors regulated directly by cellular lipids,

including Liver X receptor alpha (LXRA) and Peroxisome proliferator-activated receptor alpha (PPARG)[53]. Notably, I observed enrichment for several predicted CCAAT/enhancer binding protein (C/EBP) transcription factor motifs including C/EBP-beta, which is a known myeloid transcription factor, but has no known role in foam cell formation[249]. I also performed the motif enrichment after separating the 1743 dynamic sites into sites with oxLDL-induced FARIE-seq signal (1270) and the remaining sites with oxLDL-suppressed signal (see appendix 9.3 for results). This analysis indicated that most of the transcription factor enrichment—including C/EBP-beta—was driven by oxLDL-induced sites. By contrast in oxLDL-suppressed sites there was enrichment for E26 transformation-specific transcription factor motifs—ETS factors can be associated with suppression of gene expression. Given the overall enrichment for C/EBP-beta motifs in dynamic enhancer sites (driven by oxLDL induced chromatin sites), in addition to up-regulation of the *CEBPB* gene itself after oxLDL exposure (microarray data – log2 fold increase 0.35, adjusted p = 0.049), I hypothesized that C/EBP-beta regulatory activity in macrophages would be stimulated by oxLDL exposure.

Table 9 Enrichment of transcription factor motifs in dynamic chromatin/enhancer sites.

The subset of open chromatin sites with dynamic open chromatin and enhancer signal was analyzed for transcription factor motif enrichment and the top 25 motifs are shown with the p values for enrichment. Motifs were identified using SeqPos by comparison with known motifs from the Cistrome database and by identification of *de novo* motifs. SeqPos clusters similar motifs and the transcription factors listed are

representative of each cluster. The SeqPos program uses genomic background and assesses sites for central motif enrichment.

Rank	Transcription factor	P value	Motif	Rank	Transcription factor	P value	Motif
1	JUND	8x10-70		13	*de novo* 1	6x10-21	
2	NFE2L2	8x10-70		14	UBP1	2x10-20	
3	TFE3	8x10-70		15	KLF12	2x10-19	
4	*de novo* 17	8x10-70		16	LXRa	4x10-19	
5	JUND	2x10-63		17	CNTN2::CREB1	2x10-18	
6	CEBPB	5x10-59		18	EGR2	3x10-18	
7	PAX2	5x10-59		19	CREB3L2	3x10-17	
8	PU.1	2x10-30		20	*de novo* 4	3x10-16	
9	JDP2	4x10-40		21	SMAD4	7x10-16	
10	MAFA	1x10-36		22	AHR	7x10-16	
11	CREB1	5x10-35		23	ZBTB7B	8x10-16	
12	PPARA	5x10-24		24	PBX3	7x10-14	
				25	PPARG	9x10-9	

We, therefore, performed ChIP-seq for C/EBP-beta binding in macrophages before and after oxLDL-induced foam cell formation. I observed 10,212 genomic sites with significantly different C/EBP-beta binding signal between cell types, almost all of which (97%, 9,907 sites) had increased signal in foam cells (Figure 4-17A). Dynamic C/EBP-beta sites were enriched for proximity to genes relating to innate immune functions (Figure 4-17B), suggesting that C/EBP-beta may be involved in inflammatory processes triggered by oxLDL during foam cell development. As

expected, within dynamic C/EBP-beta binding sites the most enriched transcription factor motif was for C/EBP-beta itself (Table 10).

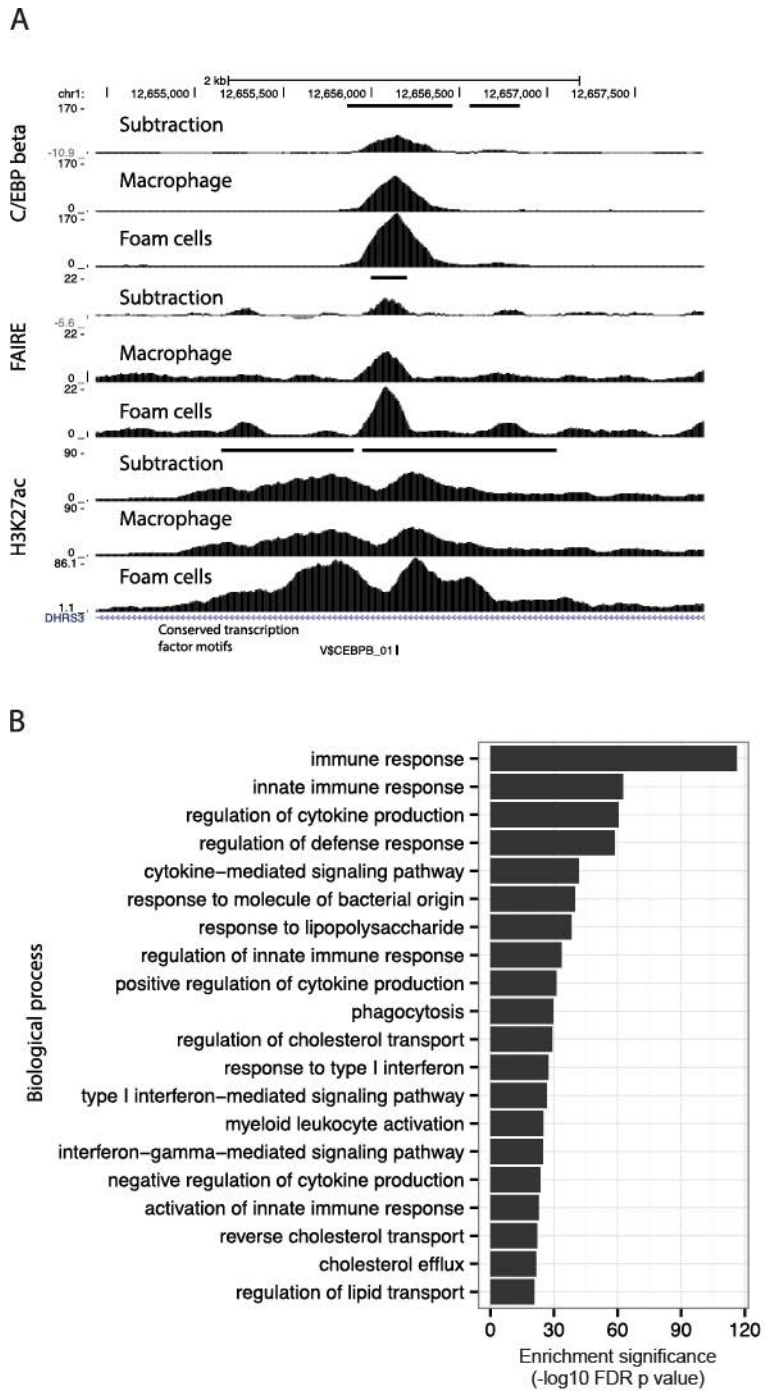
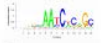
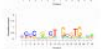


Figure 4-17 C/EBP-beta binding is responsive to oxLDL.

(A) Track view of a dynamic open chromatin and enhancer site which also overlaps a dynamic C/EBP-beta binding site with increased C/EBP-beta binding in foam cells (horizontal black bars indicate dynamic sites). (B) Enrichment of gene sets associated with different biological processes for dynamic C/EBP-beta binding sites.

Table 10 Top 10 most enriched transcription factor motifs in the top 3000 dynamic C/EBP-beta binding sites (ranked by average signal strength).

Cluster rank	Matrix	Transcription factor	DNA binding domain	Hits	P value (-log10)	Motif
1	MC00121	CEBPB	Leucine Zipper Family	2978	690.776	
2	denovo4	C-EBPbeta	Leucine Zipper Family	2523	690.776	
3	MC00293	Gfi1b	BetaBetaAlpha-zinc finger Family	494	690.776	
4	M00514	ATF4	Leucine Zipper Family	2988	690.776	
5	MC00456	JUND	Leucine Zipper Family	1378	690.776	
6	UP00031	ZBTB3	BetaBetaAlpha-zinc finger Family	1690	690.776	
7	MS00313	Creb5	Leucine Zipper Family	2634	690.776	
8	M00139	AHR	Helix-Loop-Helix Family	2849	690.776	
9	MC00457	ELF1	Ets Domain Family	2978	690.776	
10	denovo8	PU.1	Ets Domain Family	2466	690.776	

4.2.4 CAD-associated genetic variants are enriched in regulatory DNA sites altered by oxLDL

Disease-associated genetic variants may alter transcriptional regulation by affecting transcription factor binding to regulatory DNA elements [131]. Therefore, I sought to profile the relationship between the oxLDL-induced changes I had identified in macrophages and genetic variants associated with CAD risk. I identified 45 independent CAD-associated genomic loci and catalogued the reported index SNP and all SNPs in high LD ($r^2 > 0.8$) [119](Figure 4-18). I then intersected macrophage and foam cell epigenomic data with the resulting set of variants at CAD-associated loci (Table 11). At 22 of the 45 CAD loci, one or more variants lay within an open chromatin or enhancer site in macrophages and/or foam cells. We tested for enrichment of variants at CAD-associated loci in dynamic sites compared to the expected overlap derived from a background set of matched non-CAD GWAS loci

(see Methods). We observed significant enrichment of CAD-associated loci in dynamic chromatin sites (fold = 4.26; binomial p = 0.0027) (Figure 4-19), an effect that was stronger when considering sites with both dynamic chromatin and enhancer signal (fold = 6.66; p = 0.036). Conversely, we observed no significant enrichment when considering chromatin sites in macrophages and foam cells alone (p = 0.19, p = 0.31). These results suggest that variants at CAD-associated loci are specifically enriched in the subset of macrophage regulatory sites altered by oxLDL exposure.

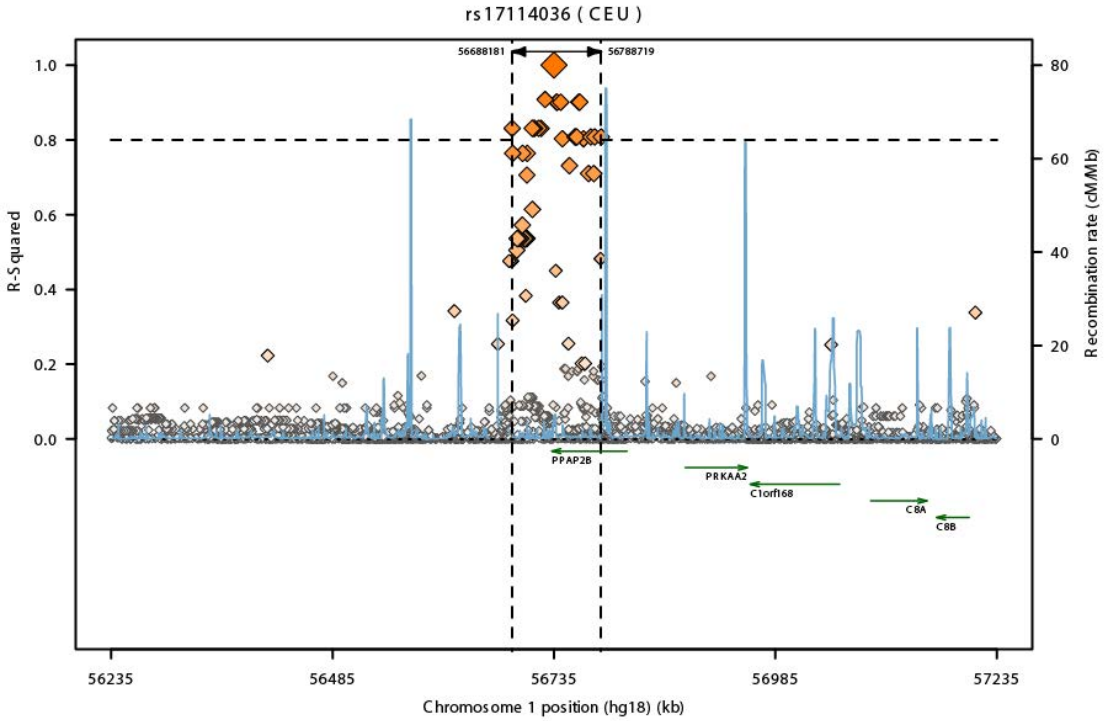


Figure 4-18 Regional plot of linkage disquilibrium relating to CAD risk SNP rs17114036 at the PPAP2B gene locus.

Orange diamonds indicate SNPs in linkage disequilibrium with rs17114036 (the upper most diamond). Vertical dashed lines indicate region bounded by SNPs with LD r^2 > 0.8.

Table 11 Number of CAD loci and individual SNPs overlapping regulatory elements in macrophages and foam cells.

For each type of genomic annotation (e.g. open chromatin site in macrophages) we overlapped the positions of CAD associated SNPs (All SNPs in high linkage disequilibrium with a reported CAD SNP). Shown in the table is the overall number of SNPs in a particular genomic annotation and the number of loci that have at least one SNP in a genomic annotation (Locus defined as the GWAS CAD reported SNP and all SNPs in high linkage disequilibrium ($R^2 > 0.8$)).

Feature type		**Number of CAD loci**	**Number of SNPs**
Open chromatin (FAIRE-seq)	Macrophage	12	26
	Foam cell	12	26
Enhancer (H3k27ac)	Macrophage	18	109
	Foam cell	19	114
Dynamic chromatin site (FAIRE-seq)		6	9
Dynamic enhancer site (H3k27ac)		9	38

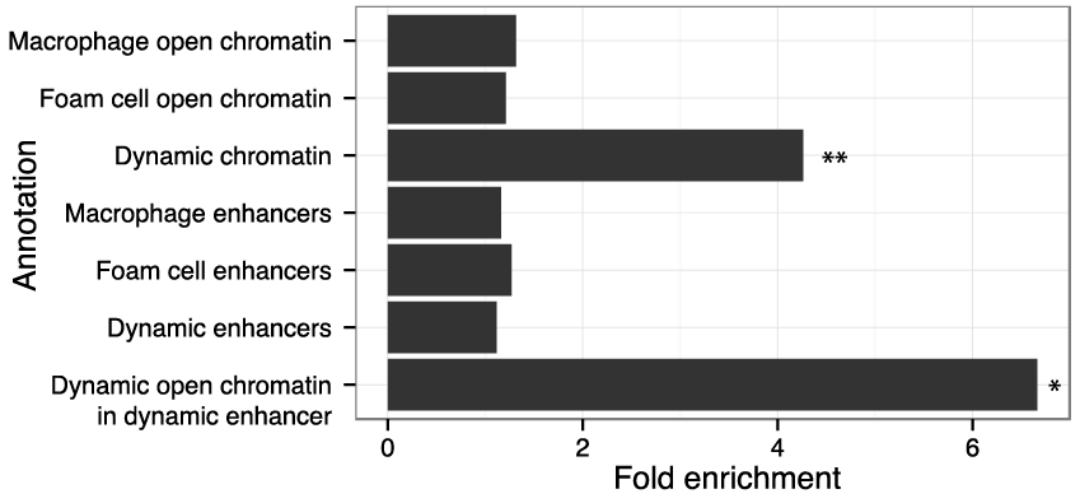


Figure 4-19 CAD loci are enriched in oxLDL-regulated open chromatin sites.

Fold enrichment of CAD loci with variants overlapping different types of regulatory features compared to background loci (* $p < 0.05$, ** $p < 0.005$).

CAD-associated variants at the Interleukin-6 receptor (*IL6R)* and Phosphatidic acid phosphatase type 2B (*PPAP2B)* loci were within overlapping dynamic chromatin and enhancer sites. At both loci the CAD-associated variants (rs7549338 and rs7553796 at *IL6R;* rs72664324 at *PPAP2B*) overlap single dynamic sites that had increased signals after oxLDL exposure; further, both *IL6R* and *PPAP2B* had significantly up-regulated expression in response to oxLDL (Figure 4-20A/B). *IL6R* encodes the receptor for the pro-inflammatory cytokine interleukin-6 and circulating levels of a soluble form of IL6R have been associated with coronary artery disease[250, 251]. *PPAP2B* encodes lipid phosphate phosphohydrolase 3 (LPP3), an enzyme that metabolizes and so deactivates pro-inflammatory mediators [162]. The mechanism through which the *PPAP2B* locus influences CAD risk is unknown and the expression pattern and regulation of *PPAP2B* by oxLDL has not been studied previously. These results demonstrate that dynamic sites can highlight candidate causal variants and genes at

CAD-associated loci. The functions ascribed to IL6R and PPAP2B suggests that they have opposing effects on the overall inflammatory state in response to oxLDL since IL6R may increase inflammation and PPAP2B decrease inflammation. In terms of the GWAS effect one would therefore hypothesize that the risk variants would increase IL6R production and decrease PPAP2B production—we test for the effects of CAD variants on PPAP2b expression in the following section. Indeed, the IL6 locus protective variant associated with CAD and diabetes has been associated with reduced membrane expression of IL6R on lymphocytes suggesting that protection may be associate with abrogation of IL6R signaling[251].

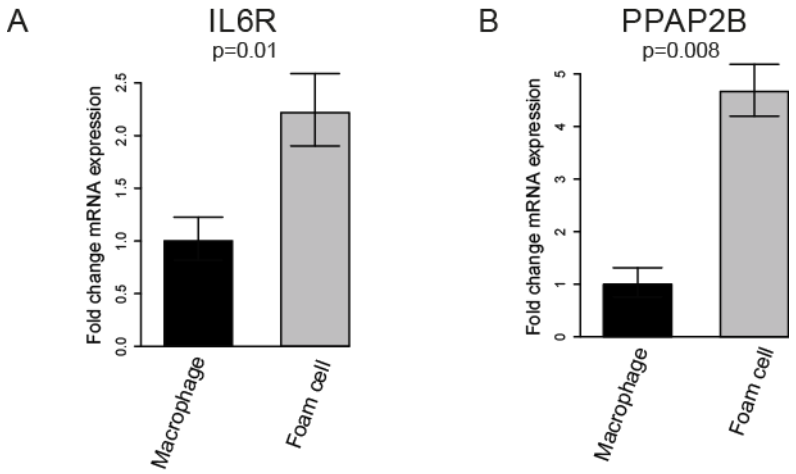


Figure 4-20 Up-regulation of IL6R and PPAP2B in foam cells compared to macrophages as assessed by quantitative RT-PCR.

(A) OxLDL induces expression of IL6R at the mRNA level (n = 3). (B) OxLDL induces expression of PPAP2B at the mRNA level (n = 3, homozygous major genotype rs72664324, rs17114036)

4.3 Discussion

Both environmental and genetic factors influence an individual's risk of developing coronary artery disease. I have used epigenetic techniques for mapping chromatin structure at cis-regulatory elements to study the interacting effects of an environmental stimulus and risk-associated genetic variants. Overwhelming evidence indicates that LDL is the major environmentally influenced contributor to the development of atherosclerosis. In humans circulating levels of LDL are influenced by diet and are tightly correlated with CAD risk; lowering LDL lowers risk[4, 5, 252]. In various animal models, diets that elevate LDL accelerate atherosclerosis[253]. The most striking cellular effect of lipid in atherosclerosis results from the interaction of oxLDL with macrophages. Progressive oxLDL uptake results in the formation of lipid-laden foam cells[172, 254]. Foam cells contribute to the inflammatory nature of the atherosclerotic lesions in various ways, as they have reduced motility, release inflammatory cytokines, chemokines, degradative enzymes, reactive oxygen species and ultimately undergo apoptosis with release of a cocktail of further pro-inflammatory cellular contents[13, 172, 254].

I mapped the effects of oxLDL on chromatin remodeling and gene expression in macrophages and integrated these data with the results of GWAS in CAD to identify candidate functional variants. These results provide a detailed map of the epigenetic changes arising from exposure of macrophages to oxLDL and provide a rational basis for studying approaches to ameliorate foam cell formation.

Open chromatin mapping with FAIRE-seq identifies the locations of all types of cis-regulatory elements and can be used to study changes in chromatin structure in cells subjected to a variety of stimuli or undergoing ontogenetic change [220, 232, 255-257]. However, as FAIRE-seq does not directly provide information about the function of DNA elements, I simultaneously undertook ChIP-seq for H3K27ac which marks active enhancers[221, 245]. Identifying the subset of open chromatin sites with active enhancer status is important because they play a central role in gene regulation, particularly the control of cell-type specific gene expression[258]. H3K27ac also marks promoter sites, which are easily distinguishable by their genomic context.

FAIRE-seq derived open chromatin sites that vary between macrophages and foam cells contain regulatory DNA sequences that play a role in oxLDL-induced foam cell formation. I found that the number of open chromatin sites that were modified by oxLDL was substantially greater than the number of differentially expressed genes. This is likely to reflect the combinatorial effects of sets of these dynamic chromatin sites and clusters of spatially linked sites were enriched around genes whose expression was altered by oxLDL. Although I detected a significant change in open chromatin structure at oxLDL-regulated promoters, the fold change in signal was smaller than that seen at non-promoter sites. This is consistent with data indicating that promoter chromatin structure is largely set at an early stage during cell ontogeny as 70% of promoter sites remained constant during adipogenesis compared to only 25–40% of non-promoter sites[257].

OxLDL also induced widespread changes in the H3K27ac enhancer signal and a close correlation was demonstrated between differential gene expression and the mean change in enhancer signal across multiple dynamic H3K27ac sites annotated to their

nearest gene. Other studies have found that not all sites of H3K27ac signal are associated with an open chromatin site[259]. I found that at a subset of 1743 non-promoter sites there were oxLDL-induced changes in both the open chromatin site and the surrounding enhancer marks and that the changes were directionally concordant at over 90% of these sites. The value of combining H3K27ac and FAIRE-seq data in this way is evident from the increased strength of the relationship of these dually identified sites with nearby gene expression[260]. Similarly, I found a robust relationship between the oxLDL-induced change in FAIRE signal at these sites and expression of the nearest gene. The importance of this set of sites was underlined by their proximity to gene-sets strongly enriched for genes known to be involved in coronary artery disease. As a subset of enhancers for further study, these sites represent a valuable prioritized set, since the dynamic open chromatin sites pinpoint the precise locations where altered transcription factor binding must occur.

Given the strong association of these dynamic enhancers with gene expression, identifying the transcription factors that bind them would very valuable in defining the upstream regulatory pathways influenced by oxLDL. Within the 1,743 sites I identified a distinct signature of transcription factor binding using motif analysis. As validation of our approach, I found enrichment for binding sites for NRF2 (NFE2L2) and PPARG, which are both known to be associated with foam cell formation. NRF2 mediates the oxidative stress response to oxLDL and PPARG is a nuclear receptor that binds constituents of oxLDL[261, 262]. The list of enriched transcription factors included C/EBP-beta and AP1, which are both known to act as pioneer factors and so can bind relatively closed chromatin, causing chromatin remodeling and thus allowing further transcription factors to bind[219]. Using ChIP-seq I confirmed oxLDL increased C/EBP-beta binding at multiple genomic sites. In keeping with these results

showing an effect of oxLDL on C/EBP-beta, the saturated fatty acid palmitate has been shown to induce inflammatory changes in murine macrophages by a C/EBP-beta-dependent mechanism[263]. In adipocytes C/EBP-beta causes chromatin remodeling which opens up sites for PPARG binding[259]; our data raise the possibility of a similar process in foam cell formation with oxLDL exposure resulting in C/EBP-beta opening up chromatin for other transcription factors[259].

C/EBP-beta is a member of the CCAAT/enhancer binding protein transcription factors that bind DNA either as a dimer or as heterodimer with other CEBP proteins[264, 265]. C/EBP-beta itself exists as three different isoforms encoded by an intronless gene. It is predominantly expressed in liver and myeloid cells[266]. The shortest isoform—termed LIP ("liver enriched inhibitory protein")—is thought to act as a dominant negative whereas the LAP (liver enriched activating protein) isoform has been shown to be the major active isoform. An additional isoform—LAP*—is longer than LAP and possesses similar activity. C/EBP-beta is essential for myeloid differentiation and is known to increase the expression of several inflammatory genes including IL6[249]. It remains unclear how oxLDL itself causes activation of C/EBP-beta binding and our CHIP-seq data is not able to distinguish binding of the different isoforms. The likely mechanism involves phosphorylation of certain amino acid residues. C/EBP-beta activity is required for differentiation of adipocytes and this is dependent on phosphorylation of threonine residue 188 through the extracellular signal-regulated kinase pathway [267]. Similarly, in fibroblasts threonine 188 phosphorylation by MAPK (Mitogen-activated protein kinases) was important for transcriptional activity [268]. Since oxLDL is known to activate the MAPK pathway this represents a plausible way in which oxLDL would lead to increased phosphorylation of residues in C/EBP-beta that promote its DNA binding [269].

Profiling the co-localization of regulatory sites with genetic risk variants identified by GWAS has been useful in the identification of cell types in which disease risk variants operate [128, 224]. Given the established role of macrophages and foam cells in atherosclerosis it was initially surprising that we found no enrichment for CAD risk loci in either context alone. However, we found that those sites that underwent oxLDL-induced changes in chromatin accessibility were enriched for CAD-associated loci. This demonstrates that the response to oxLDL, with concomitant foam cell formation, induces chromatin changes at the sites of a subset of CAD variants and provides a cellular context for risk variants at these sites to operate in altering disease risk. Our data show that the regulatory capacity of key SNPs in dynamic sites is actuated in macrophages by an environmentally influenced stimulus, oxLDL, so demonstrating the mechanism for a direct interplay between environmental and genetic risk factors.

5 PPAP2B: A coronary artery disease risk variant influencing the macrophage response to lipid

5.1 Introduction

5.1.1 GWAS and PPAP2B

The Phosphatidic phosphatase 2B gene (PPAP2B) encodes lipid phosphate phosphohydrolase (LPP3), an enzyme that regulates glycerolipid metabolism [162]. Genomewide-association studies have implicated PPAP2B in the genetic risk of atherosclerosis. These studies have identified SNPs in the last intron of PPAP2B that have a statistically significant effect on the risk of coronary artery disease [123, 155, 270]. The reported SNP rs17114036[155] has G and A alleles; the risk allele A, has a frequency of 0.91 [119]. Thus the minor allele, G of rs17114036 is associated with protection from coronary artery disease. The precise mechanism by which these SNPs influence the function of PPAP2B or CAD risk is unknown. Neither the GWAS-reported SNP nor SNPs in high LD alter the coding sequence of PPAP2B[123]. The SNPs may act by altering transcriptional regulation through mechanisms that could involve altered transcription factor binding or miRNA binding, as has been reported for other CAD risk loci[131, 135].

5.1.2 PPAP2B

The PPAP2B gene spans about 85kb on the short arm of chromosome 1, runs in the antisense direction and comprises 6 exons. Northern blotting for mRNA expression in a human tissue panel demonstrated ubiquitous expression[271] . PPAP2B expression was up-regulated by epidermal growth factor (EGF) in HeLa cells[271]. Differential protein level expression of the PPAP2B-encoded LPP3 mouse homologue was demonstrated in mouse tissues with highest expression in kidney, lung and brain and lesser expression in liver, heart spleen and skeletal muscle[162]. In human umbilical vein endothelial cells (HUVEC) LPP3 was upregulated at the protein level by stimulation with Vascular endothelial growth factor ($VEGF^{165}$)[272]. Differentiation of Human multipotent adipose-derived stem cells into mature adipocytes was also associated with a reduction in PPAP2B expression[273]. Comparison of PPAP2B expression between dendritic cells with a TH1 or TH2 type phenotype showed that PPAP2B expression was significantly higher in IL3-induced TH2-type dendritic cells[274]. LPP enzymes metabolize bioactive signaling phospholipids that are derived through complex metabolic pathways from membrane phospholipids. LPP3 is a glycosylated enzyme[162] with 6 putative transmembrane domains that hydrolyzes lysophosphosphatidic acid, ceramide-1-phosphate and sphingosine-1-phoshate yielding monoacylglycerol, ceramide and sphingosine, respectively[271, 275, 276]. LPP3 also hydrolyzes the phosphorylated form of extracellular Fingolimod, an immunomodulatory drug used for multiple sclerosis treatment[277]. Endogenous LPP3 is located predominantly at the cell periphery in Swiss 3T3 cells with lesser

staining in the Golgi[162]. The mouse homologue of PPAP2B was down regulated upon differentiation of pre-adipocytes to adipocytes and was shown to dephosphorylate extracellular LPA, indicating ecto-enzyme activity at the cell membrane[278]. LPP3 localizes to the basolateral membrane in stably transfected Madin-Darby Canine Kidney Epithelial Cells although there was a discrepancy with its enzymatic activity that remained low in the basolateral segment[279]. Human LPP3 possesses an RGD integrin binding motif and it is possible that baso-laterally localized LPP3 may be involved in cell-cell interactions[279, 280]. In Chinese hamster ovary cells endogenous LPP3 co-localized with sphingosine kinase in intracellular vesicles and shifted to a more perinuclear distribution upon stimulation with PMA[281]. Flow cytometry using a poly-clonal anti-LPP3 antibody reportedly showed cell surface staining with intact endothelial cells suggesting that LPP3 is a cell surface antigen[272]. LPP3 has also been shown to localize to caveolin-1 associated lipid rafts that might be involved in segregating and compartmentalizing its enzymatic activity[162, 282]. Both recombinant and endogenous LPP3 and LPP2 are capable of forming catalytically active hetero- and homo- oligomers[283].

The mouse homologue of LPP3 is required for embryogenesis and lack of LPP3 was lethal by day 10.5 of embryogenesis although the precise molecular mechanism underlying this effect remains unknown[163]. Embryos lacking LPP3 showed marked defects in extra-embryonic vasculature development and in embryonic axis formation[163]. Lack of LPP3 was associated with alterations in extracellular lysophosphatidic acid (LPA), and intracellular diacylglycerol (DAG). The mouse phenotype resembled that of axin deficiency[163]. Axin is a regulator of the wnt signaling pathway involved in control of apoptosis[284]. Activation of the wnt

pathway causes accumulation of β-catenin, which binds to TCF/LEF repressor transcription factors preventing their inhibitory effect on gene expression[284]. Lack of LPP3 resulted in increased β-catenin levels, and reduced DNA binding activity TCF transcription factors at a TCF response element. [163]. The converse effect occurred when LPP3 was overexpressed but LPP3 phosphatase activity was not required indicating a role of LPP3 beyond it enzymatic activity[163].

Malignant ascites from patients with ovarian cancer has high levels of LPA that is produced by ovarian cancer cells and contributes to cellular proliferation. Overexpression of LPP3 in ovarian cancer cell lines reduced extracellular LPA and was associated with reduced proliferation [285]. Injection of ovarian cancer cell lines overexpressing human LPP3 into mice resulted in fewer tumours and reduced tumour growth compared to injection of untransfected cells[285]. Over expression of LPP3 also reduced intracellular S1P concentration in serum-starved HEK293T cells and increased apoptosis[281]. By contrast, overexpression of LPP3 in bowel cancer cell line SW480 increased cellular proliferation and knockdown of LPP3 in glioblastoma cells reduced proliferation[286]. Overexpression of LPP3 in SW480 cells was associated with increased B-catenin stability and CYCLIN-D1 synthesis, thereby increasing expression of genes associated with cellular proliferation[286].

5.1.3 Role of LPP3 in vascular function

LPP3 may have roles in vascular biology including endothelial development and effects on plaque levels of bioactive signaling lipids. In a model of capillary formation

using HUVECs, an anti-LPP3 antibody was able to block VEGF-induced capillary morphogenesis[272], suggesting a role for LPP3 in angiogenesis. Knockdown of LPP3 also reduced migration of endothelial cells and this was associated with alterations in the ß-catenin signaling pathway[169]. In mice with an endothelial cell and haemopoietic selective LPP3 knockout there was increased endothelial permeability and capillary leak when the knockout was activated in adult mice. [287]. The constitutive selective knockout of LPP3 was lethal *in utero*[287]. In a mouse model of atherosclerosis PPAP2B expression was increased in diseased mouse aortic endothelial cells, foam cells, and atherosclerotic plaque[288]. A mouse model of neointimal hyperplasia showed increased PPAP2B expression in vascular smooth muscle cells (VSMC) after vascular injury[165]. Mice with smooth muscle cell selective PPAP2B knockout were viable and fertile but demonstrated exaggerated neointimal hyperplasia[165]. There is a paucity of data available on the expression pattern of PPAP2B in atherosclerosis in humans.

5.1.4 Role of LPP3 enzymatic activity in atherosclerosis

Increasing evidence implicates the substrates of LPP3 enzymatic activity as having a role in atherosclerosis. LPA is an obligate intermediary of triglyceride and glycerophospholipid synthesis and an extracellular ligand for 6 different LPA receptors [289, 290]. There is an accumulation of LPA in plaques in both human atherosclerotic disease and murine models of atherosclerosis[291-294]. LPA has pro-atherogenic effects on most of the cells involved in atherosclerosis, promoting vascular smooth muscle proliferation, endothelial cell adhesion molecule expression,

and stimulating oxLDL uptake by macrophages[295]. In addition, LPA is a highly thrombogenic mediator and its release upon atherosclerotic plaque rupture can contribute to thrombotic occlusion of the artery[292]. Strategies that modulate LPA levels in plaques could, among other beneficial effects, reduce thrombogenicity. Inhibition of LPA signaling in a mouse model of an inducible, selective LPP3 knockout in endothelial and haemopoietic cells, mitigated the deleterious effects produce by LPP3 deficiency[287]. LPP3 expression is reduced in animals and human at sites of low shear stress where atherosclerotic plaques form. LPP3 was required for endothelial layer integrity under athero-protective flow conditions by suppressing LPA signaling through LPA receptor 1 (LPAR1) and LPAR2 receptors in a cellular model [296]. Overall, LPP3 is likely to play a significant role in maintaining endothelial integrity through mechanisms including limiting of LPA signaling and possibly integrin binding.

S1P is a receptor active sphingolipid which signals via 5 G-protein linked cell surface receptors resulting in diverse effects including immune cell trafficking and angiogenesis[297]. Regulated expression of LPP3 reduces levels of S1P in lymphoid tissue and is required for egression of lymphocytes, that follow the S1P concentration gradient, into the circulation[298]. Recent studies have sought to characterize the precise role of S1P in atherosclerosis using murine models in which S1P receptors have been knocked out. Mice deficient for S1P2R or S1P3R and ApoE have reduced numbers of macrophages and foam cells in plaques[299]. Bone marrow chimeras indicate that this effect in S1P2R deficient mice is dependent on haemopoietic cells and S1P2R deficiency reduced macrophage inflammatory response. Nevertheless, the precise role of S1P in different aspects of atherosclerosis remains to be determined.

5.1.5 PPAP2B expression and CAD risk genotype

Evidence for transcriptional regulatory activity of a SNP can be investigated by comparing the levels of gene expression between individuals of differing genotypes. An eQTL study of PPAP2B expression in human endothelial cells showed that the risk allele of the CAD-associated SNP, rs17114036, was associated with reduced expression of PPAP2B [288]. A further eQTL exists in human blood monocytes whereby the minor allele of rs4634932 is associated with increased expression of PPAP2B [300]. rs4634932 is in high linkage disequilibrium with the CAD SNP rs17114036 (r^2 0.82, 1000 genomes), whose minor allele is associated with protection from CAD. eQTL studies of other human tissues have not found a significant association between genotype at rs17114036 and PPAP2B expression[301]. To date no studies have assessed individual risk SNPs at the PPAP2B gene locus for their effects on transcription factor binding or transcriptional activity in, for example, reporter gene assays. The mechanism governing the eQTL effect is unknown and a shared effect in GWAS CAD studies for rs17114036 may be coincidental.

In the present study, I characterized the expression pattern and enzymatic activity of PPAP2B in response to atherogenic lipid and the gene regulatory effects of CAD-associated SNPs. I found that foam cells are the predominant source of PPAP2B in plaque and that disease associated SNP rs72664324 regulated PPAP2B expression in response to atherogenic lipid through altered C/EBP-beta transcription factor binding.

5.2 Results

5.2.1 LPP3 expression in atherosclerosis and arteritis

I studied the *in vivo* expression pattern of LPP3 using immunohistochemistry staining of LPP3 in human tissues (Figure 5-1). In healthy arteries LPP3 was expressed mainly by endothelial cells and occasional lymphocytes and macrophages (Figure 5-1A). By contrast, in atherosclerotic plaques there were large numbers of foam cells containing abundant LPP3. In foam cells the LPP3 appeared to be distributed in a coarse granular pattern suggestive of entrapment within cytoplasmic organelles (Figure 5-1B/C). Atherosclerotic plaques are often associated with lymphoid aggregates in the adventitial space. These lymphocytes almost uniformly expressed high levels of LPP3 (Figure 5-1D,E). This led us to examine lymphoid tissue in tonsils for LPP3 expression. LPP3 was expressed only by lymphocytes in germinal centers that are predominantly naive B cells. The surrounding lymphocytes that comprise a mixture of T and B cells were relatively devoid of LPP3 expression (Figure 5-1G). These findings indicate that there is differential expression of LPP3 across lymphocyte subsets. LPP3 expression was largely absent from the vascular smooth muscle containing tunica media (Figure 5-1F). Giant cell arteritis is a condition in which there is narrowing of the luminal space caused by circumferential infiltration of macrophages, lymphocytes and multinucleated giant cells. Infiltrating macrophages and giant cells expressed LPP3 whereas only some lymphocytes expressed

LPP3(Figure 5-1H, I). Overall these data show for the first time that the predominant source of LPP3 in atherosclerotic plaques is from foam cells.

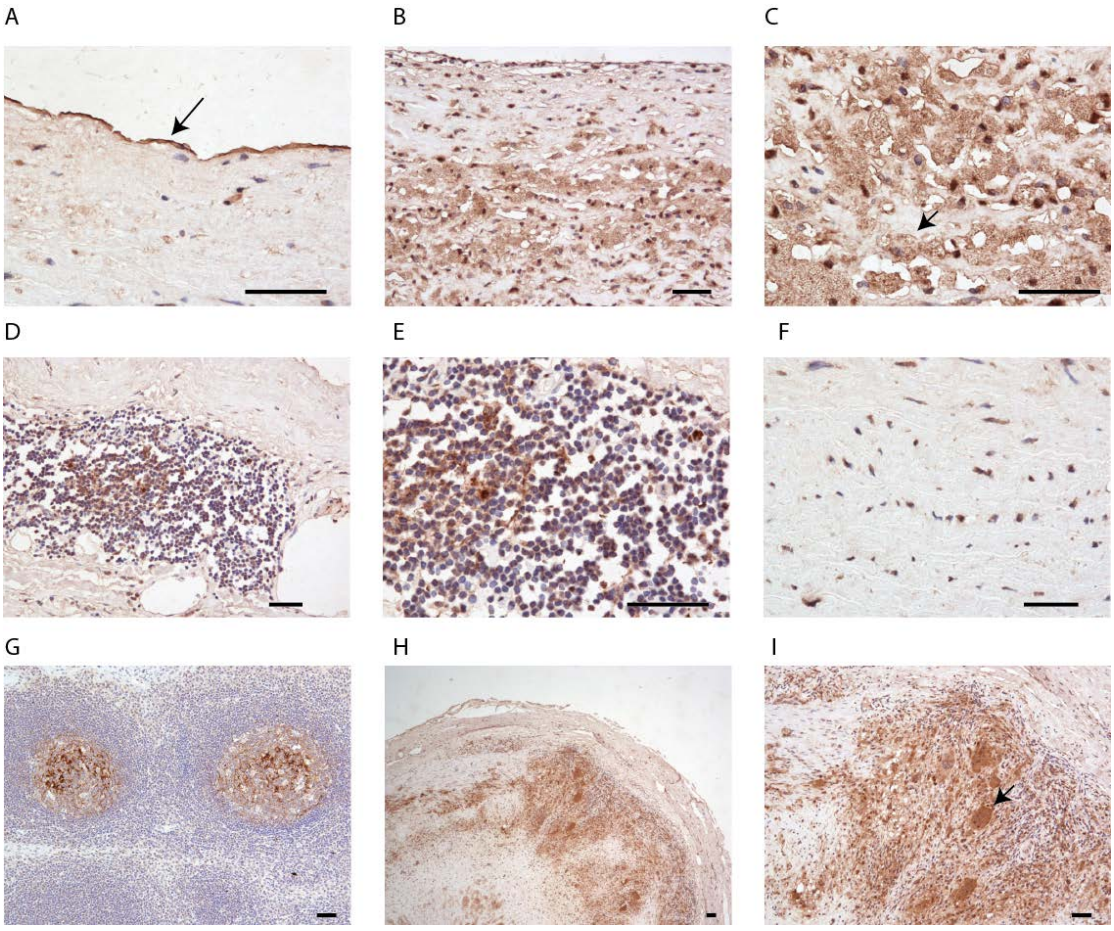


Figure 5-1 Immunohistochemistry for LPP3 in human vascular tissue and tonsil.

(A) LPP3 is expressed by arterial endothelial cells (arrowed) in a healthy vessel. (B, C) LPP3 expressing foam cells are abundant in an atherosclerotic plaque (arrowed). LPP3 is distributed in a granular pattern in the cytoplasm suggesting localization to a cellular organelle (D, E) Lymphocytes in plaque associated adventitial aggregates express LPP3. (F) Smooth muscle cell layer showing minimal LPP3 expression. (G) B-lymphocytes in tonsillar germinal centers express. Lymphocytes outwith the germinal center to do not express LPP3. (H, I) Transverse section of an artery affected

by giant cell arteritis shows abundant LPP3 expressing macrophages, lymphocytes and multinucleated giant cells (arrowed). Scale bars indicate 50um, sections representative of 5 individuals. Photos courtesy of Dr E Soilleux.

5.2.2 OxLDL-induced upregulation of PPAP2B/LPP3 and enzymatic activity in macrophages

Our whole genome expression analysis of oxLDL treated macrophages indicated that PPAP2B was the 7th most upregulated gene at the mRNA level (Table 3). I next sought to confirm this finding at the protein level by performing western blotting for LPP3 expression in macrophages and foam cells from 3 different individuals. Western blotting demonstrated the presence of both the known glycosylated and non-glycosylated forms of LPP3 (Figure 5-2). Across all three individuals LPP3 expression of both glycosylated and non-glycosylated protein was increased by oxLDL-induced foam cell formation (Western blotting performed by A Morris, University of Kentucky).

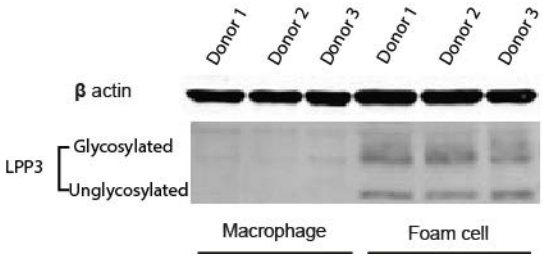


Figure 5-2 Western blotting demonstrates up-regulation of glycosylated and non-glycosylated LPP3 protein in macrophage-derived foam cells.

Actin loading control shown above (Western blotting performed by A. Morris using macrophage and foam cell extracts generated by M Reschen, data from 3 donors shown).

We next compared the overall phosphatase activity of LPP3 in macrophages and foam cells. We used a highly robust assay that has been established in the literature [162, 165]. First, immunoprecipitation for LPP3 is performed to exclude the activity of other phosphatases with effects on LPA and S1P. Next, mass-spectrometry based measurement of the hydrolysis of labelled LPA and S1P is performed [162, 165]. There was significantly increased enzymatic activity of LPP3 per amount of total protein in foam cells compared to macrophages (Figure 5-3A, B).

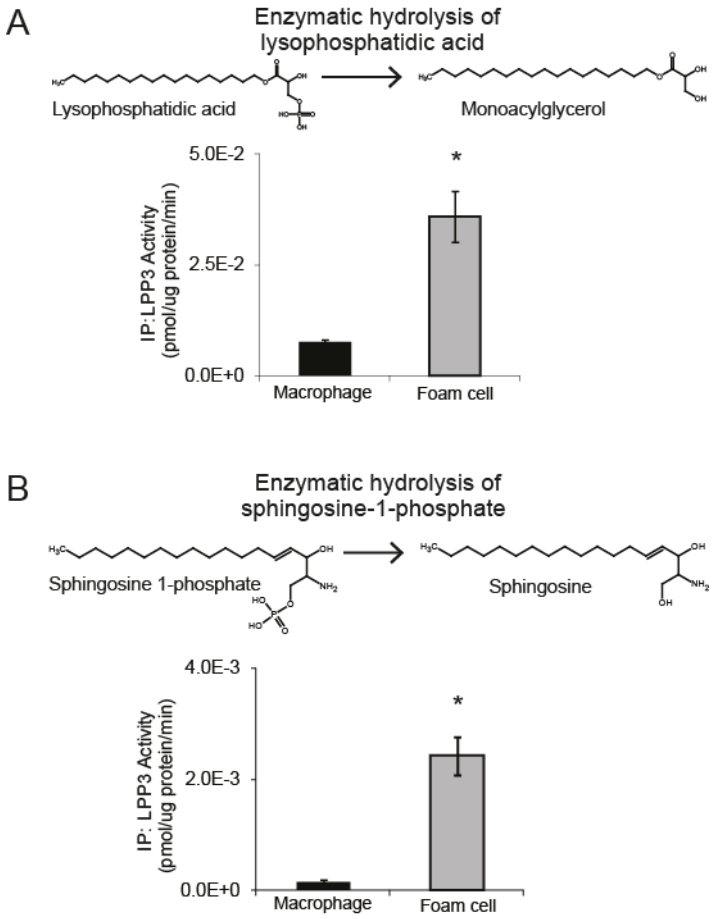


Figure 5-3. The enzymatic activity of LPP3 per amount of cellular protein is significantly increased by oxLDL stimulation.

(A) LPP3 hydrolysis of LPA to MAG is increased (P=0.008, student's T test, n=3, error bars indicate standard error). (B) LPP3 mediated hydrolysis of S1P to Sphingosine is significantly increased (P=0.01, student's T test, n=3, error bars indicate standard error, assay and figures produced by A Morris using macrophages/foam cells provided by the author).

OxLDL stimulation may lead to changes in the cellular levels of LPP3 substrates and products and the upregulation of LPP3 enzymatic activity is likely to be part of a protective response. We found that LPA levels were increased in foam cells and there was a trend towards increased levels of its LPP3-hydrolyzed product, MAG (Figure

5-4). S1P levels were not increased by oxLDL but sphingosine levels were increased (Figure 5-4). Overall, these data indicated that foam cell formation is associated with abnormalities in the levels of some of the substrates and products of LPP3 enzymatic activity and further studies employing knockdown or inhibition of LPP3 would be useful to define its precise role in regulating their levels. It is interesting to note that overall both substrates and products of LPP3 were increased in foam cells. This suggests that the enzymatic activity and increased expression of LPP3 in foam cells is not enough to neutralize the cellular loading effects of oxLDL on S1P and LPA. This suggests that oxLDL likely still overwhelms the ability of cells to normalize LPA and S1P levels. Additional experiments using LPP3 knockout or knockdown would be required to clarify the exact role of LPP3 in regulating the levels of these substrates in foam cells.

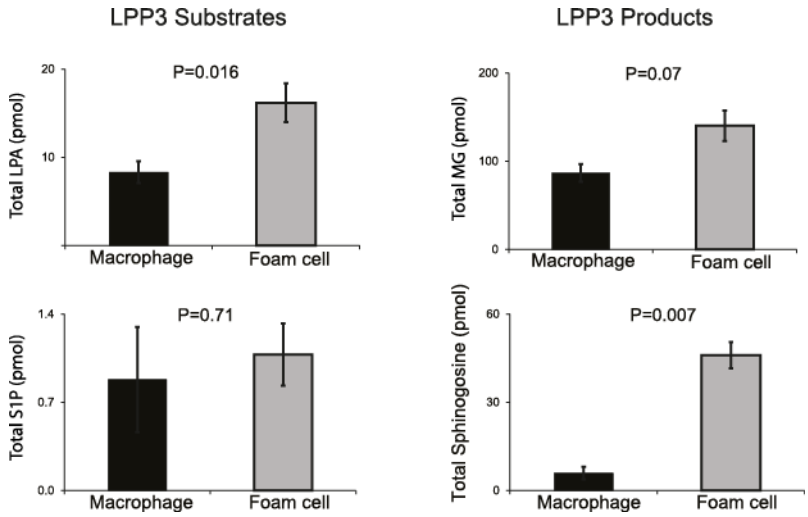


Figure 5-4 OxLDL alters the levels of LPP3 substrates and products in macrophages.

Mass spectrometry based assays in lipid extracts from macrophages and oxLDL-induced foam cells showed increased levels of LPA and sphingosine in foam cells with a trend towards increased MG levels (n=3 biological replicates, Student's t test P values shown in the figure, error bars indicate standard deviation, assays and figures produced by A Morris using macrophages and foam cells supplied by the author).

5.2.3 CAD-associated variant rs72664324 demonstrates allelic differences in oxLDL-induced enhancer activity and PPAP2B expression

I sought to identify and characterize candidate causal CAD-associated variants at the *PPAP2B* locus. As discussed earlier, a single CAD-associated variant, rs72664324, overlapped a dynamic chromatin and enhancer site, which in turn lies in an oxLDL-induced 'super-enhancer' cluster (Figure 5-5). None of the other variants in high LD

with the GWAS reported SNP rs17114036 overlapped an open chromatin site, although several were within regions with enhancer histone marks. Our ChIP-seq data indicated that rs72664324 also lies within a dynamic C/EBP-beta binding site that shows increased C/EBP-beta binding in foam cells. I tested the rs72664324 SNP for transcriptional regulatory potential. The ChIP-seq data was produced in individuals homozygous for either the A or G allele of rs72664324—therefore it was not possible to perform allele-specific analysis which would have required heterozygotes. To compare the binding between genotypes I quantile normalized the signal across all C/EBP-beta ChIP-seq samples for dynamic C/EBP-beta binding sites. At the rs72664324 site I observed greater C/EBP-beta signal in individuals homozygous for the A allele compared to the G allele (3.7 fold) in macrophages and in foam cells (1.3 fold). Overall, oxLDL increased the C/EBP-beta signal by an average of 2.6-fold at this site, consistent with the direction of the oxLDL-induced dynamic chromatin and enhancer site. Due to the small number of replicates (2 for each genotype) and the problems with comparing different individuals the allele specific results did not achieve statistical significance.

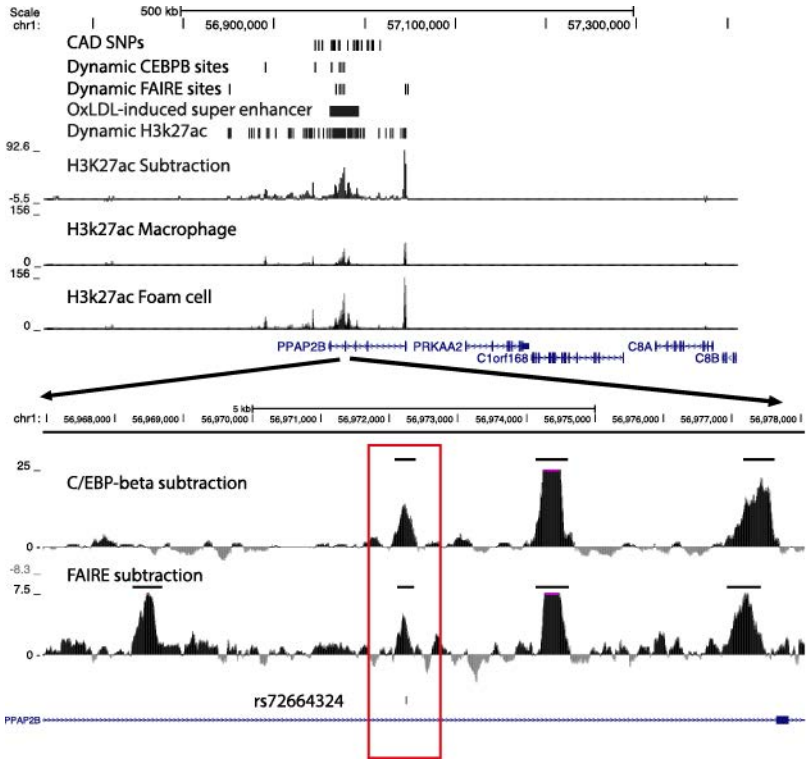


Figure 5-5 Chromatin profile at the PPAP2B locus where rs72664324 is in a dynamic chromatin, enhancer and C/EBP beta site (red box).

The region is part of an oxLDL-induced super enhancer.

I compared the predicted affinity for binding of transcription factors to the A and G alleles of rs7264324 using the Transcription Factor Affinity Prediction Webtool and the Transfac transcription factor motif database[153]. The C/EBP transcription factor motif showed the greatest difference in affinity with significant affinity only for the A allele (Table 12). Additionally, motif analysis predicted binding for both alleles of rs72664324 to several factors including as STAT5A, GR and NFY (Table 13).

Table 12 Top 10 transcription factor motifs with the greatest difference in affinity at the rs72664324 site between the risk G and protective A allele.

The sequences tested were identical to the EMSA probes (rs72664324-A AGGAAATGAACCAATGTCTGTTCCT, rs72664324-G AGGAAATGAACCGATGTCTGTTCCT)

Rank	Difference log(p) for two sequences	A allele p value	G allele p value	Matrix ID (Transfac)	Transcription factor
1	1.71	0.0137	0.693	M00190	CEBP
2	1.48	0.0263	0.794	M00116	CEBPA
3	0.978	0.0452	0.429	M00621	CEBPDELTA
4	0.874	0.0102	0.0762	M01147	DMRT2
5	0.828	0.104	0.698	M01334	NKX11
6	0.822	0.0471	0.312	M00309	ACAAT
7	0.821	0.0498	0.33	M00775	NFY
8	0.816	0.0902	0.591	M00109	CEBPB
9	-0.814	0.936	0.144	M00975	RFX
10	-0.802	0.199	0.0314	M00792	SMAD

Table 13 Transcription factor motif affinities for rs72664324 analyzed separately for each alleles.

(Matrices with P value < 0.05 shown).

Rank	P Value	Matrix ID	Name
Alternative A allele			
1	<0.000625	M00493	STAT5A
2	<0.000625	M00921	GR
3	<0.000625	M00500	STAT6
4	<0.000625	M00496	STAT1
5	<0.000625	M00498	STAT4
6	0.000625	M00494	STAT6
7	0.00853	M00747	IRF1
8	0.00932	M01250	E2F1
9	0.0102	M01147	DMRT2
10	0.0113	M00181	E2
11	0.0116	M00928	E2
12	0.0123	M00107	E2
13	0.0137	M00190	CEBP
14	0.0245	M01123	NANOG
15	0.0249	M01252	E2F6
16	0.0263	M00116	CEBPA
17	0.0282	M01281	NFAT1
18	0.036	M00185	NFY
19	0.0449	M00463	POU3F2
20	0.0452	M00621	CEBPDELTA
21	0.0471	M00309	ACAAT
22	0.0473	M00750	HMGIY
23	0.0498	M00775	NFY
Reference G allele			
1	<0.000938	M00493	STAT5A
2	<0.000938	M00921	GR
3	<0.000938	M00500	STAT6
4	<0.000938	M00496	STAT1
5	<0.000938	M00498	STAT4
6	0.000938	M00494	STAT6
7	0.00287	M00181	E2
8	0.0053	M00928	$E2
9	0.00563	M00107	E2
10	0.00852	M00747	IRF1
11	0.00931	M01250	E2F1
12	0.0203	M01123	NANOG
13	0.0249	M01252	E2F6
14	0.0282	M01281	NFAT1
15	0.0314	M00792	SMAD
16	0.0443	M00070	TAL1BETAITF2
17	0.0475	M00750	HMGIY

Electrophoretic mobility shift assays (EMSAs) were used to compare nuclear protein binding to DNA probes comprising a 25bp sequence centered on the rs72664324 SNP. There was strong protein binding to the A allele but almost total absence of binding to the G allele using primary human macrophage nuclear extracts (Figure 5-6). I used 4 different cold probes containing the consensus binding sequences for CEBP, NFY, STAT5A and GR to compete off protein binding and determine a likely candidate transcription factor. Only the CEBP consensus probe competed off binding to the A allele (Figure 5-6). Since our ChIP-seq data indicated that C/EBP-beta bound to the site *in vivo*, I used super-shift assays with anti-C/EBP-beta antibody to confirm that the protein bound to the A allele of rs72664324 was in fact C/EBP-beta (Figure 5-7). A further control experiment with an anti-CEBP-alpha antibody showed no C/EBP-alpha binding to the site (Figure 5-8). Overall, the EMSA studies indicate that C/EBP-beta present in macrophage nuclear extract is capable of binding to rs72664324 *in vitro.*

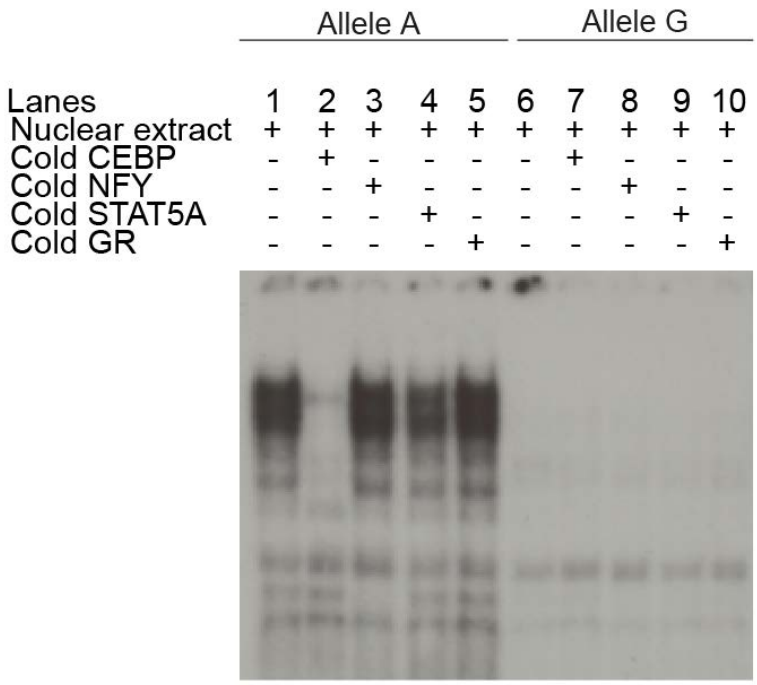


Figure 5-6 EMSA for rs72664324 alleles with cold probe competition assays.

Only the C/EBP consensus sequence competes off binding to the A allele. Cold probes for NFY, STAT5A and GR did not compete off binding. No protein binding to the G allele was detectable.

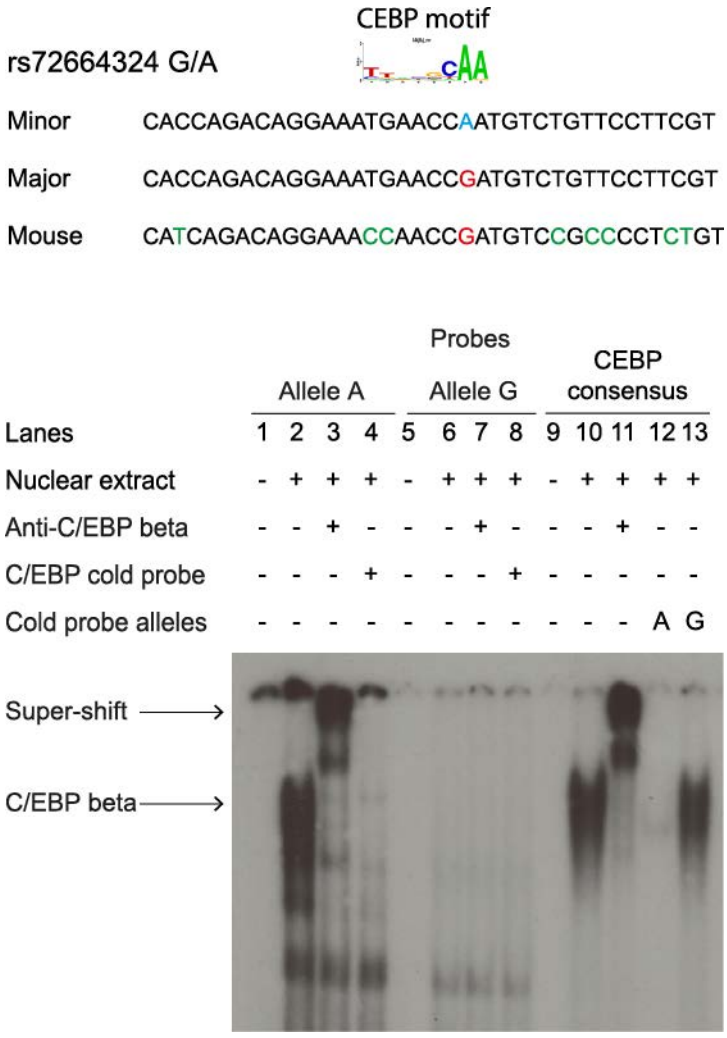


Figure 5-7 Effect of rs72664324 alleles on transcription factor binding.

Upper panel shows a ccomparison of the human rs72664324 alleles and the corresponding mouse sequence with the CEBP motif aligned above. Lower panel shows an EMSA demonstrating that only the A allele at rs72664324 binds nuclear protein, which is shown to be C/EBP-beta by a super-shift in the presence of anti-C/EBP-beta antibody. Also, only a cold probe with the A allele competes off binding to a C/EBP consensus probe.

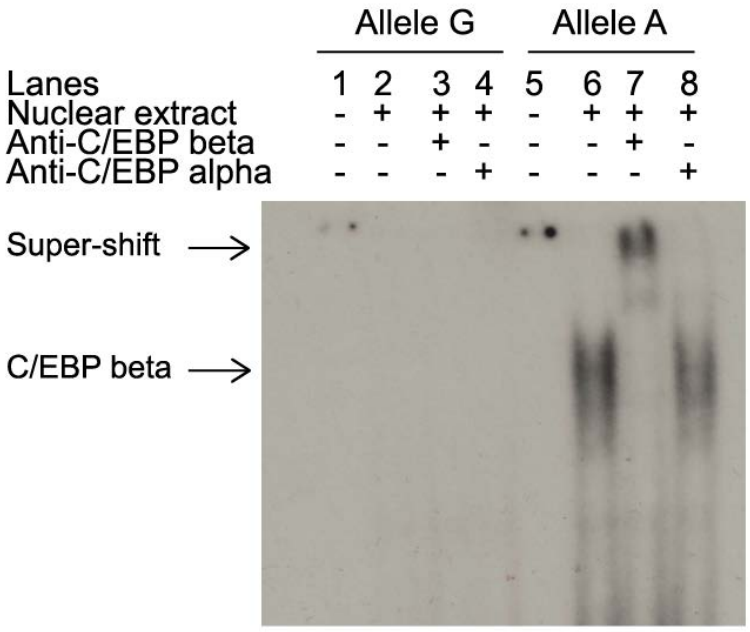


Figure 5-8 EMSA showing C/EBP-alpha does not bind rs72664324.

The A allele binds C/EBP-beta as demonstrated by a super-shift with anti-C/EBP beta antibody. The anti-C/EBP-alpha antibody produces no super-shift.

I next tested the effects of rs72664324 on oxLDL-induced enhancer activity using luciferase reporter assays in primary human macrophages and foam cells. I observed increased enhancer activity with the A allele compared to the G allele in both macrophages (1.73 fold, p = 0.0005) and foam cells (2.25 fold, p = 0.025) (
Figure 5-9). Furthermore, I observed a greater oxLDL-induced increase in enhancer activity at the rs72664324 site with the A allele (1.43 fold, p=0.03) compared to the G allele (1.098, p = 0.78) (
Figure 5-9). Overexpression of C/EBP-beta increased transcriptional enhancer activity of the A allele significantly more than that of the G allele (p = 0.012) (
Figure 5-9). These results suggest that, compared to the G allele, the A allele preferentially increases enhancer activity upon oxLDL induction through increased C/EBP-beta binding.

I obtained primary macrophages from healthy individuals with different rs72664324 genotypes. I observed significantly higher oxLDL-induced PPAP2B expression in macrophages from individuals with at least one copy of the A allele compared to those with only the G allele (p = 0.0013) (
Figure 5-9). The A allele of rs72664324 is in LD with the protective allele of the reported index SNP at this locus, so higher PPAP2B expression is associated with reduced CAD risk

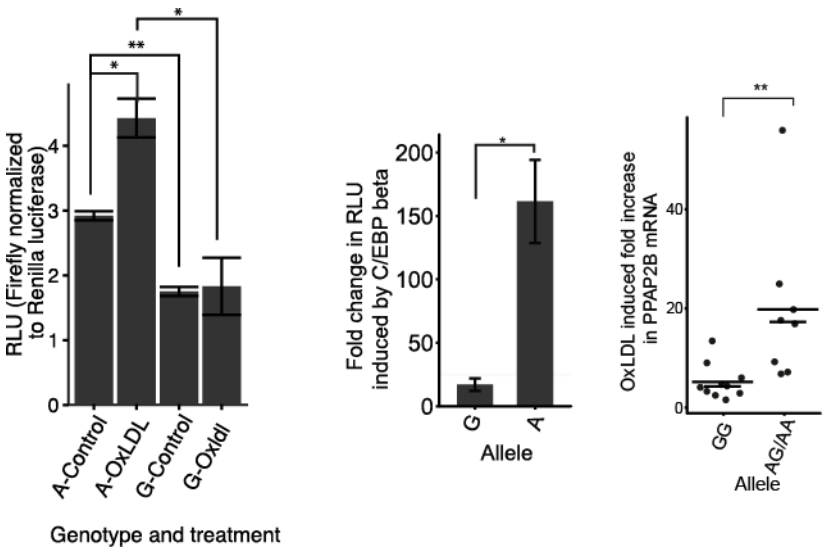


Figure 5-9 CAD risk SNP rs72664324 at the PPAP2B locus regulates enhancer activity and oxLDL-induced expression of PPAP2B.

(left panel) Luciferase reporter assays in primary human macrophages and foam cells with a reporter element containing rs72664324 with either allele (* $p < 0.05$, ** $p < 0.005$). (middle panel) Effect of C/EBP-beta overexpression on luciferase reporter activity with either the A or G allele relative to empty vector (* $p < 0.05$). (right panel) Induction of PPAP2B expression by oxLDL in primary human macrophages from individuals with allele G or at least one copy of the A allele (long lines indicate mean, short lines indicate median, Mann-Whitney U test, $p = 0.0013$, GG n = 10, GA n = 2, AA n = 6).

5.2.4 CAD SNPs at the PPAP2B locus and transcriptional regulation in endothelial cells

An eQTL for PPAP2B expression exists in human aortic endothelial cells whereby the risk allele of rs17114036 is associated with reduced expression of PPAP2B [301]. Two of the CAD associated SNPs at the PPAP2B risk locus are in endothelial cell open chromatin sites [216]. These include the GWAS reported SNP rs17114036, and rs56348932 which is in high linkage disequilibrium (r^2 0.93). Transcription factor affinity analysis suggests that rs17114036 may alter binding of the AP1 transcription factor complex (Table 14) and that rs56348932 may alter binding of a range of transcription factors (Table 15) [153]. ChIP-seq data indicate that c-FOS binds at the rs17114036 SNP[216]. c-FOS is a transcription factor that only binds to DNA as a hetero-dimer with c-Jun, forming the AP1 transcription factor complex. There is no known ChIP-seq derived transcription factor binding at rs56348932.

Table 14 Comparison of transcription factor binding affinities to the A and G alleles of rs17114036.

MAFB and AP1 motifs show a greater affinity for the rs17114036 site containing the G allele rather than the A allele.

Rank	Difference log (p) for two sequences	A allele p value	G allele p value	Matrix ID	Transcription factor
1	-2	0.137	<0.00137	M01227	MAFB
2	-1.78	0.24	0.00403	M00173	AP1
3	-1.72	0.25	0.00479	M00188	AP1
4	-1.3	0.17	0.00855	M00174	AP1
5	-1.21	0.187	0.0116	M00172	AP1FJ
6	-1.11	0.541	0.0423	M00482	PITX2
7	0.963	0.0408	0.375	M01392	HOXA6
8	0.914	0.0292	0.24	M01425	HNF1B
9	0.894	0.0872	0.683	M01111	RBPJK
10	0.872	0.0782	0.583	M01388	DLX5

Table 15 Comparison of transcription factor binding affinities to the A and C alleles of rs56348932.

Several different transcription factors are predicted to have greater affinity for either allele.

Rank	Difference log(p) for two sequences	A allele P value	C allele P value	Matrix ID	Transcription factor
1	1.49	0.0116	0.363	M00671	TCF4
2	1.25	0.02	0.358	M00328	PAX8
3	-1.25	0.146	0.00818	M01305	TEF
4	-1.23	0.119	0.00698	M01109	SZF11
5	1.17	0.0165	0.247	M01661	HBP1
6	1.17	0.038	0.569	M00794	TTF1
7	0.892	0.0906	0.707	M00097	PAX6
8	-0.836	0.36	0.0525	M00704	TEF1
9	0.807	0.111	0.713	M01286	SOX
10	0.795	0.113	0.705	M01372	NKX22

Given this analysis, I used EMSA with HUVEC nuclear extract to test for differential transcription factor binding at these two SNPs. EMSA demonstrated transcription factor binding to both the A and G allele of rs17114036 (Figure 5-10). The binding was stronger with allele G and the band migrated slightly further in the EMSA. I

demonstrated a super-shift with anti-c-Fos antibody with the G allele probe and there was no super-shift with the A allele. The transcription factor c-Fos binds to DNA only as a heterodimer with c-Jun and together the dimer binds to the AP1 DNA activation element [302]. A consensus cold probe for the AP1 DNA activation element was able to compete off binding of the protein to the G allele but had no effect on the protein bound to the A allele (Figure 5-11). Taken together, these findings indicate AP1 is capable of binding to the G allele of rs17114036 but not the A allele. The identity of the protein binding to the A allele remains unknown.

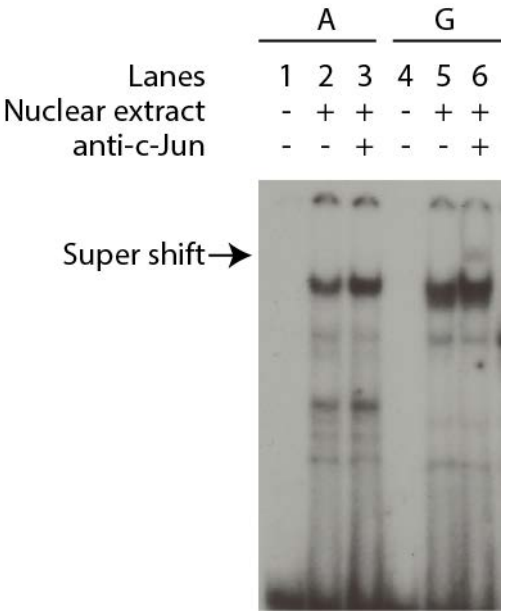


Figure 5-10 EMSA showing endothelial cell nuclear protein binding to the A and G allele of rs17114036.

Anti-c-Jun antibody causes a supershift with the G allele (arrowed) but not the A allele indicating that c-Jun can only bind to the G allele. The protein binding the the A allele migrates a shorter distance in the EMSA indicating that it is a different transcription factor to that binding the G allele.

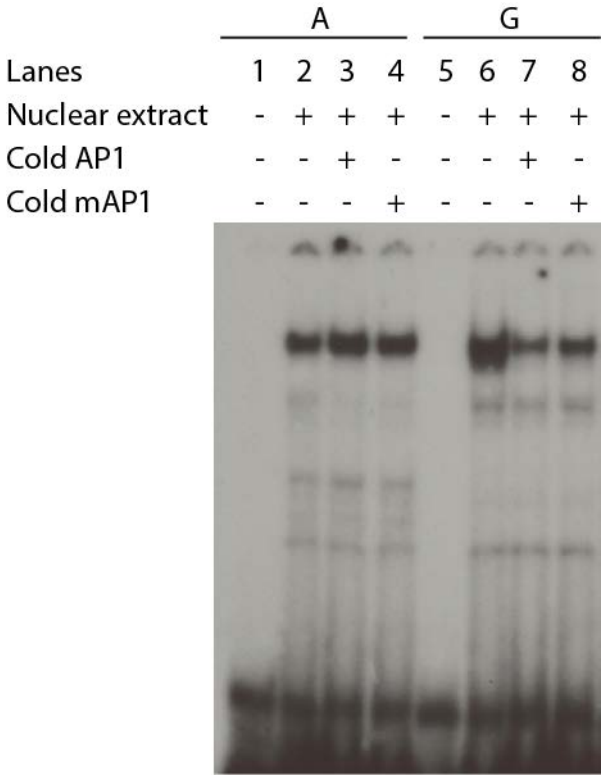


Figure 5-11 EMSA showing the effect of competition assays on endothelial nuclear protein binding to the rs17114036 SNP alleles.

Two different AP1 consensus sequence cold probes are able to compete off most of the protein binding to the G allele but have no effect on the protein binding to the G allele. (AP1 cold probe CGCTTGATGACTCAGCCGGAA, mAP1 cold probe sequence CGCTTGATGACTTGGCCGGAA (mAP1 contains an arbitrarily mutated AP1 consus sequence that still has a high bioinformatically derived affinity for AP1 according to TRAP (see methods) with little affinity for other transcription factors (data not show)).

With the rs56348932 SNP there was relatively symmetrical nuclear protein binding with both the A and C alleles but an additional band near the top of the gel was present with the A allele (Figure 5-12). Further studies using a selection of cold probes and antibodies would be required to identify this factor.

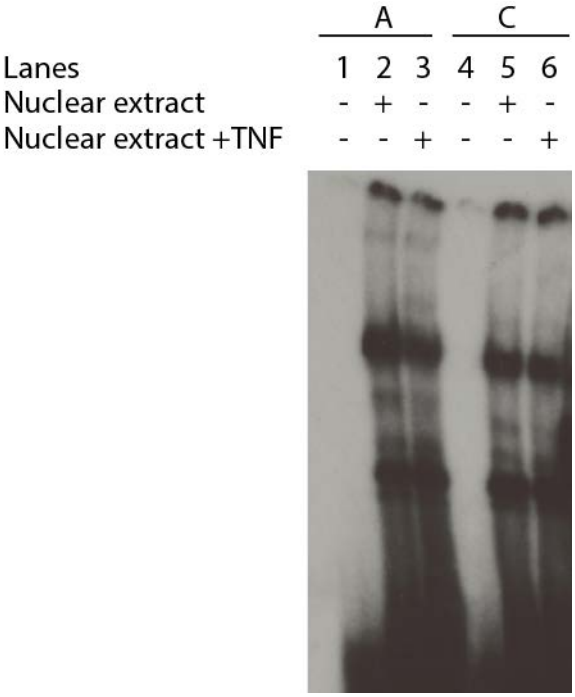


Figure 5-12 EMSA showing endothelial cell nuclear protein binding to the A and C alleles of rs56348932.

Protein binding is evident with the A allele but no binding is seen with the C allele. Symmetrical bands migrating further to the mid portions of the gel likely represent non-specific protein binding. Nuclear extract was extracted from HUVECS treated with and without TNF-alpha (10ng/ml).

EMSA studies indicated that both of the CAD risk SNPs in open chromatin sites in HUVECs showed allele specific difference in transcription factor binding. I therefore undertook luciferase assays in HUVECS using plasmids encoding the open chromatin sites and varying only by the SNP alleles (Figure 5-13). As a control I also tested the rs72664324 SNP which shows transcriptional regulatory activities in macrophages.

Although the plasmids effectively constitute open chromatin, the greatest activity overall was seen with the two sites that are native open chromatin sites in endothelial cells and contain rs17114036 or rs56348932. In both cases there was a significant increase in luciferase activity with the allele that showed greater nuclear protein binding in the EMSA. In the case of rs17114036 there was 6.5 greater activity with the A allele that binds AP1 selectively compared to the G allele (P=0.034). At the rs56348932 contain site there was 2.8 fold increased activity with the A allele compared to the C allele (P=0.02). For the control site containing rs72664324 there was no significant difference between each allele. These findings indicate that rs17114036 and rs56348932 could account in part for the rs17114936 eQTL by altering transcription factor binding and enhancer activity.

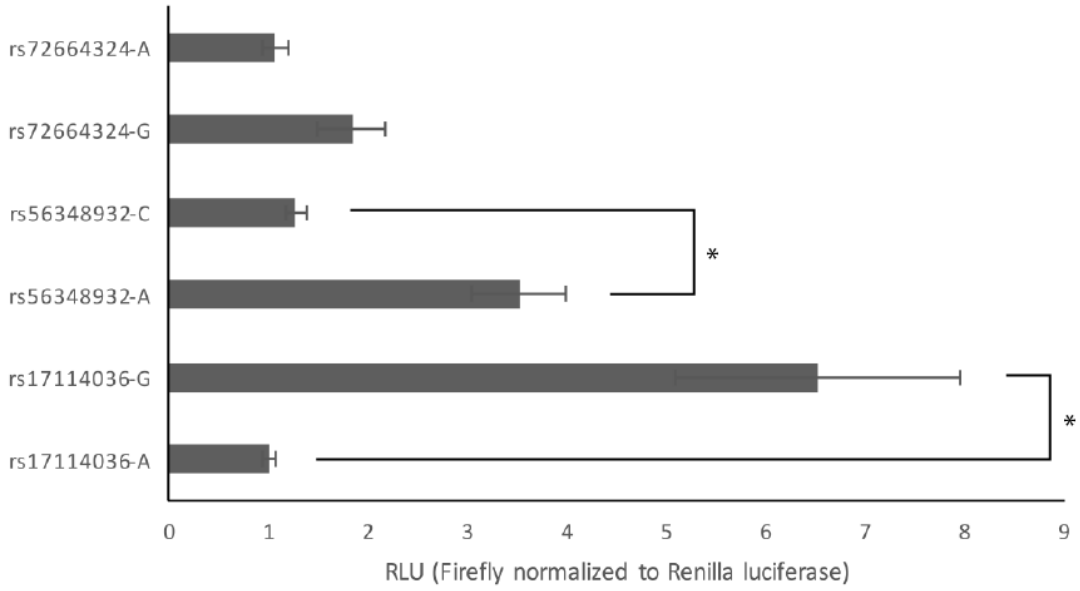


Figure 5-13 CAD SNPs at the PPAP2B locus affect transcriptional regulation in endothelial cells.

Luciferase reporter assay showing transcriptional activity of putative regulatory DNA sites containing containing 3 CAD associated SNPs. Rs17114036 and rs56348932 are

in open chromatin sites in endothelial cells. For comparison the assay was also performed with the rs72664324 SNP which shows regulatory activity in macrophages. Transcriptional activity was highest with the G allele of rs17114036 with 6.5 fold increase in activity over the A allele. The A allele of rs56348932 showed 2.8 fold more activity than with the C allele (error bars indicate standard error, 3 biological replicates, * $P < 0.05$, Student's t test).

5.3 Discussion

The *PPAP2B* CAD risk locus was of particular interest because oxLDL induced more open chromatin, greater enhancer activity and up-regulated PPAP2B expression in macrophages. Our finding that PPAP2B and its protein product, LPP3 were up-regulated by oxLDL complements a growing body of evidence implicating dysregulation of its substrates, LPA and S1P in atherosclerosis. LPP3 hydrolyzes LPA and S1P to their non-receptor active forms [162, 271]. LPP3 is known to have a role in vascular development and endothelial integrity, but its role in macrophages and foam cells remains unknown [296, 303]. I found that *PPAP2B* was one of the most up-regulated genes in foam cells, was abundant in human plaque foam cells and its specific enzymatic activity towards LPA and S1P in macrophages was increased by oxLDL. Our immunohistochemical studies of human atherosclerotic plaque indicated that the predominant source of LPP3 was in foam cells. In keeping with an accumulation of LPA in atherosclerotic plaque, we found that foam cells contained significantly more LPA than macrophages. Taken together our findings suggest that by degrading LPA, LPP3 could have a role as a negative regulator of pro-inflammatory LPA signaling during the macrophage response to oxLDL and in foam cells.

The *PPAP2B* locus is well validated as a CAD risk locus and our data demonstrate that the risk allele reduced the induction of PPAP2B by oxLDL in macrophages [123, 155, 270]. This suggests a plausible model whereby the risk allele lowers transcription of PPAP2B in response to oxLDL exposure in macrophages. This in turn would lower

levels of LPP3 in foam cells, resulting in higher levels of LPA in plaque and consequently increased pro-inflammatory signaling, vascular smooth muscle proliferation, retention of foam cells in lesions and thrombogenicity. Our data demonstrate that an effect on PPAP2B transcription may in part be mediated through rs72664324, a SNP that is in high linkage disequilibrium with the reported GWAS SNP (rs17114036). The risk allele at rs72664324 reduces C/EBP-beta binding at this site and so reduces the enhancer activity of the site. The overall mechanism I propose is in keeping with that shown for two other CAD loci where rs12740374 affected SORT1 transcription and SNPs at the 9p21 locus altered STAT1 binding and interferon gamma signaling [131, 132].

Further evidence suggests that rs72664324 is a causal SNP. None of the other 20 SNPs in high linkage disequilibrium were in the coding sequence or the 3'UTR, where altered miRNA binding is an additional mechanism by which CAD SNPs might act [135]. The chromatin structure at the reported SNP, rs17114036, remained closed before and after oxLDL exposure in macrophages. Moreover, no other SNP in high linkage disequilibrium at the locus was within an open chromatin site before or after oxLDL treatment. Interestingly, I found that the DNA sequence at the dynamic chromatin site containing rs72664324 has been highly conserved across vertebrate species. Active enhancers have been associated with the production of small RNA transcripts, enhancer RNA. Using CAGE-seq to measure enhancer transcripts across the genome in hundreds of primary cells and tissues, the only enhancer RNA that overlapped any of the 21 SNPs in high linkage disequilibrium with the reported risk SNP was that from the rs72664324 site [304]. This enhancer was only detected in monocytes, the precursors of macrophages [304]. Finally, I show that the presence of the protective A allele at the rs72664324 site enhances the upregulation of PPAP2B

expression that is triggered by oxLDL in primary macrophages. Nevertheless, I cannot exclude that other linked SNPs may be important, especially since our primary open chromatin data and enhancer mapping was performed in individuals homozygous for the major allele of the CAD reported SNP.

Our data indicate that a CAD risk variant operates to alter the response of macrophages to oxLDL exposure by altering binding of C/EBP beta to an enhancer site regulating PPAP2B expression. This will influence inflammatory and other aspects of the atherosclerotic disease process. Targeting the activity of PPAP2B-encoded LPP3 in macrophages and foam cells is a plausible therapeutic strategy. Future studies could directly address the role of *PPAP2B* in the response to oxLDL using murine models in which *PPAP2B* is selectively knocked out from myeloid cells. Overall, our study establishes a link between CAD genetic susceptibility, the macrophage response to atherogenic lipid and receptor active lipid signaling. This study demonstrates the utility of chromatin and enhancer mapping in primary human cells before and after a pathogenic environmental stimulus and this approach may have applications in the study of other diseases.

In human endothelial cells an eQTL for PPAP2B expression coincides with the GWAS risk variant rs17114036[288]. To investigate a plausible mechanism by which variants might affect expression of PPAP2B in endothelial cells I used a similar approach to that in macrophages whereby I focused on variants that were in regulatory DNA as defined by the presence of open chromatin. I found that the G allele increased transcriptional activity by increasing the binding of AP1 and the G allele shows increased expression in the eQTL study[288]. The second risk SNP rs56348932 in open chromatin in high linkage with rs17114036 showed increased

transcriptional activity with the A allele. The A allele also showed additional transcription factor binding in the EMSA assay. However, given the high linkage disequilibrium between rs17114036 and rs56348932 the increased transcriptional activity with the A allele of rs56348932 is unexpected since the A allele is associated with reduced transcription of PPAP2B. Overall these findings suggest that the altered binding of AP1 to rs17114036 is responsible for the eQTL in endothelial cells. In endothelial cells AP1 binding activity is increased by shear stress [305]. Shear stress occurs at sites of turbulent arterial flow and contributes to the early development of atherosclerotic plaques. Given that PPAP2B is essential for vascular development it is possible that individuals with the protective allele at rs17114036 express higher levels of PPAP2B and in through AP1 in response to shear stress.

6 Regulation by genetic and atherogenic stimuli of coronary artery risk gene, phosphatase and actin regulating protein 1 (PHACTR1)

6.1 Introduction

GWAS have identified SNPs at the 6p24.1 locus that are strongly associated with the risk of CAD. This locus is one of the most robust and replicated CAD risk loci having been identified in multiple GWAS (see next section). The risk SNPs identified by these studies lie within an intron of the Phosphatase and actin regulating protein 1 (PHACTR1) gene and the next nearest gene is TBC1D7 which lies about 440kb away. Although PHACTR1 itself has not been studied in relation to atherosclerosis itself, the proximity to the GWAS risk SNPs identifies it as the likely candidate gene through which the genetic effect is exerted. Through a series of experiments beginning with studying expression in human atherosclerosis I found compelling evidence linking PHACTR1 to atherosclerosis and the genetic risk signal.

6.1.1 Genome-wide association studies and the PHACTR1 genetic risk locus

GWAS studies have shown that the PHACTR1 risk locus is associated with a variety of CAD phenotypes including coronary artery calcification[159], early myocardial infarction[124], and a composite CAD endpoint comprising myocardial infarction, coronary revascularization, ACS or angina[155, 270]. In addition to these studies in

Europeans, three studies have shown that one of the same risk SNPs (rs9349379) also mediates CAD risk in Lebanese, Han Chinese and South East Asians[270, 306, 307].

A genetic fine mapping studied demonstrated that the most likely causal variant at the locus is rs9349379[308]. The G allele was associated with an increased risk of CAD. The precise mechanism by which this common genetic variant influences the risk of CAD is largely unknown.

6.1.2 Mechanism of genetic risk for CAD at the PHACTR1 locus

Based on predictions of transcription factor binding motif disruption by either of the alleles of rs9349379, the DNA sequence at the SNP has been tested for Myocyte enhancer factor 2 (MEF2) transcription factor binding in an *in vitro* assay[308]. An electrophoretic mobility shift assay using nuclear extract from endothelial cells demonstrated binding of MEF2 only to a probe sequence containing the A allele and not to a probe containing the G allele[308]. This suggests that the risk allele of rs9349379 could affect transcriptional regulation by altering binding of one of the family of MEF2 transcription factors. The MEF2 family comprises 4 transcription factors that influence the development and differentiation of a wide range of cell types[309]. However, there is no experimental evidence that a MEF2 factor binds *in vivo* at the location of the SNP and no evidence that the region possesses transcriptional regulatory activity such as enhancer activity, which is typically studied using reporter gene assay systems.

6.1.3 CAD risk SNP association with PHACTR1 expression

Several studies have identified an association between expression of PHACTR1 mRNA and genotype at CAD risk SNPs. The identification of such an expression quantitative trait locus (eQTL) for a GWAS identified risk SNP supports the premise that the mechanism of genetic risk at a given locus may operate by altering expression of a particular gene. An eQTL for PHACTR1 has been demonstrated in composite right coronary artery tissue whereby the risk allele, G, of rs9349379 is associated with reduced expression of PHACTR1[308]. An eQTL study of GWAS CAD associated genes using PBMCs failed to detect significant overall expression of PHACTR1 and was therefore unable to measure an eQTL effect. An eQTL study in endothelial cells did not find a significant association with variants at the rs12526453 SNP, which is in LD with rs9349379 (R^2 with rs9349379 = 0.32). Thus the presently available eQTL studies do not offer a clear mechanism that would account for the increased risk of atherosclerosis with the risk genotype of the GWAS CAD SNP.

6.1.4 PHACTR1 gene and protein

The PHACTR family comprises 4 genes that encode proteins which interact directly with both actin and protein phosphatase 1 (PP1) [205]. PHACTR1 was first identified and cloned from a rat brain cDNA library using a yeast two-hybrid system with PP1 as bait[205]. In humans the PHACTR1 gene is on chromosome 6 and a transcript with a 1743 open reading frame has been cloned from human brain[310]. This transcript encodes a 580 amino acid protein that possesses 4 highly conserved RPEL domains[205]. The RPEL domain consists of repeats of the amino acid sequence RPxxxEL and has been shown to bind to unpolymerized G Actin [311].

6.1.5 PHACTR1 actin binding

Both mouse and rat orthologues of PHACTR1 have been shown to bind actin in a RPEL dependent manner[205, 312]. PHACTR1 also has two nuclear localization signal (NLS) motifs at both ends of the protein[312]. Nuclear localization signals on cargo proteins bind to importins. Importins are adapter proteins that facilitate transport of cargo proteins between the cytoplasmic and nuclear compartments, through the nuclear pore complex [313]. In the mouse ortholog, PHACTR1's NLS motifs have been shown to facilitate importin-dependent translocation into the nuclear in response to serum stimulation. Although both NLS sites facilitate nuclear importation, the N-terminal motif was shown to be more important than the C-terminal motif in terms of facilitating importation[312]. Serum stimulation induced translocation by causing Rho-dependent remodeling of actin from unpolymerized G actin to filamentous F actin. Since the RPEL domains in PHACTR1 only bind G-actin, serum stimulation effectively releases PHACTR1 from being held in the cytoplasm, allowing nuclear importation to proceed[312, 314].

6.1.6 PHACTR1 and Protein Phosphatase 1 Binding

Protein Phosphatase 1 (PP1) is a ubiquitously expressed serine/threonine phosphatase that is present in both the nucleus and cytoplasm [315]. PP1 enzymatic activity is crucial in regulating signaling cascades by reversing phosphorylation at serine and threonine residues in substrate proteins. The enzymatic specificity of PP1 is achieved through the binding of unrelated accessory proteins, known as phosphatase binding proteins (PIPs)[316]. Almost 200 PIPs have been identified that bind via PP1-docking

motifs to multiple PP1 surface grooves [316]. Disordered PP1 activity is implicated in a number of diseases that involve the introduction of exogenous PIPs, such as Hepatitis B virus protein Hbx, or affect expression or functional endogenous PIPs [317, 318].

Screening of a rat brain cDNA library using a yeast two-hybrid system with catalytic subunit of PP1α as bait identified PHACTR1 as a PIP[205]. Immunoprecipitation using ectopically expressed PHACTR1 confirmed endogenous PP1 binding[205]. Additional *in vitro* PP1 enzymatic activity assays demonstrated that recombinant PHACTR1 inhibited PP1 activity[205]. Rat PHACTR1 was expressed with a series of deletions and removal of the last 10 amino acids was found to abolish PP1 binding suggesting that the C-terminal end was where binding occurs [205].

6.1.7 Expression pattern and subcellular localization of PHACTR1

Northern blotting with a rat tissue cDNA library demonstrated expression of a 5kb transcript in brain, lung, heart, kidney and testis [205]. Using ENSEMBL predicted mouse transcripts for RPEL containing genes, three transcripts pertaining to PHACTR1 were cloned from mouse kidney tissue. The shortest transcript, about 500bp, termed RPEL-CS was expressed in kidney and to a lesser extent in mouse heart lung and brain[319]. The longer transcript of about 2kb was present only in kidney tissue and alternative splicing produced two transcripts differing by 200bp due to variable inclusion of exon 5(termed RPEL-CL1 and RPEL-CL2) [319]. PHACTR1 mRNA was expressed in several breast cancer cell lines including MDA-MB-231, SCP2, SUM159PT and SUM149PT but was not expressed in MCF7 cells[320]. TGF-β upregulated PHACTR1 expression in MDA-MB231, SCP2 and SUM159PT cells.

PHACTR1 mRNA was not detected at a significant level using a Taqman RTqPCR assay in human PBMCs[321]. HUVECs express PHACTR1 at the mRNA level and expression is upregulated 6 fold by VEGF-A_{165} [322]. In human tissues PHACTR1 mRNA was detected using RTqPCR in heart, aorta, primary endothelial cells and right coronary artery tissue[308].

Western blotting for PHACTR1 in rat tissues demonstrated expression only in brain[205]. PHACTR1 is expressed in several breast cancer cell lines including MDA-MB231, SCP2 and SUM159PT and is upregulated by TGF-β in these cells[320]. Western blotting showed PHACTR1 was expressed in HUVECs with a molecular weight of about 55kDa[322].

Using immunohistochemistry with rat brain, PHACTR1 was shown to be expressed in the striatum and olfactory tubercle, with lesser expression in the cortex and septum[205]. Staining was confined to the cytoplasm and fluorescent microscopy of N-GFP tagged ectopically expressed PHACTR1 in hippocampal neurons also showed absence of PHACTR1 from the nucleus[205]. N-terminal EGFP tagged mouse PHACTR1 transcripts cloned from mouse kidney were expressed ectopically in the mouse myoblast cell line, C2C12[319]. The long forms of mouse PHACTR1 (RPEL-CL1, RPEL-CL2) were expressed throughout the cells whereas in 50% of cells the short form (RPEL-CS) was excluded from the nucleus[319]. In mouse NIH3T3 fibroblasts flag tagged PHACTR1 was predominantly cytoplasmic in serum starved cells and predominantly nuclear in serum stimulated cells[312]. Similar findings were elicited for endogenous PHACTR1 in CHL-1 human melanoma cells [312].

6.1.8 Function of PHACTR1

The precise role of PHACTR1 remains unknown but likely involves it capacity to bind actin and PP1. PHACTR1 may affect actin structure itself since siRNA-mediated PHACTR1 knockdown in HUVECs reduced F-actin filament numbers and repartitioning as well as increasing cell protrusion dynamics[322]. Other studies using a combination of overexpression and siRNA-mediated knockdown of PHACTR1 have suggested roles in cell motility, vascular morphogenesis and tumor invasion[312, 319, 322]. The role of PHACTR1 in atherosclerosis itself is unknown.

I hypothesized that PHACTR1 would be expressed in atherosclerosis related cells and tissues and regulated by atherogenic stimuli. I tested whether expression of PHACTR1 was associated with genotype at the CAD SNP rs9349379.

6.2 Results

6.2.1 *PHACTR1* expression is increased in human atherosclerotic lesions

To understand the role of PHACTR1 in CAD we used immunohistochemistry to evaluate human PHACTR1 protein expression *in vivo* in atherosclerotic lesions in arteries from patients with atherosclerosis (n=5) and in normal arteries from healthy controls (n=5). In normal arteries PHACTR1 was expressed in endothelial cells and in scantily present immune cells, but was absent from vascular smooth muscle cells (Figure 6-1A). Atherosclerotic plaque lesions contained an abundance of infiltrating macrophages and foam cells. These cells strongly expressed PHACTR1 unlike VSMCs where no expression was seen. (Figure 6-1B, C). In the peri-plaque adventitial areas of atherosclerotic vessels we detected lymphocyte aggregates, which expressed high levels of PHACTR1 (Figure 6-1D). We therefore assessed secondary lymphoid tissue itself and found that PHACTR1 was strongly expressed, but predominantly in the cytoplasm, rather than the nucleus (Figure 6-1E). In the other cell types mentioned above, PHACTR1 staining was seen in both cytoplasmic and nuclear locations. Given the inflammatory nature of atherosclerosis, we also studied arteries from patients with the autoimmune vascular inflammatory disorder of giant cell inflammatory arteritis (n=5). In arteries affected by giant cell arteritis there was extensive infiltration by PHACTR1 expressing macrophages, lymphocytes and giant cell cells in the sub-intimal space (Figure 6-1F). These cells all strongly expressed PHACTR1 in the nucleus and cytoplasm. Overall, these data highlight the marked expression of PHACTR1 in immune cells in atherosclerotic vessels.

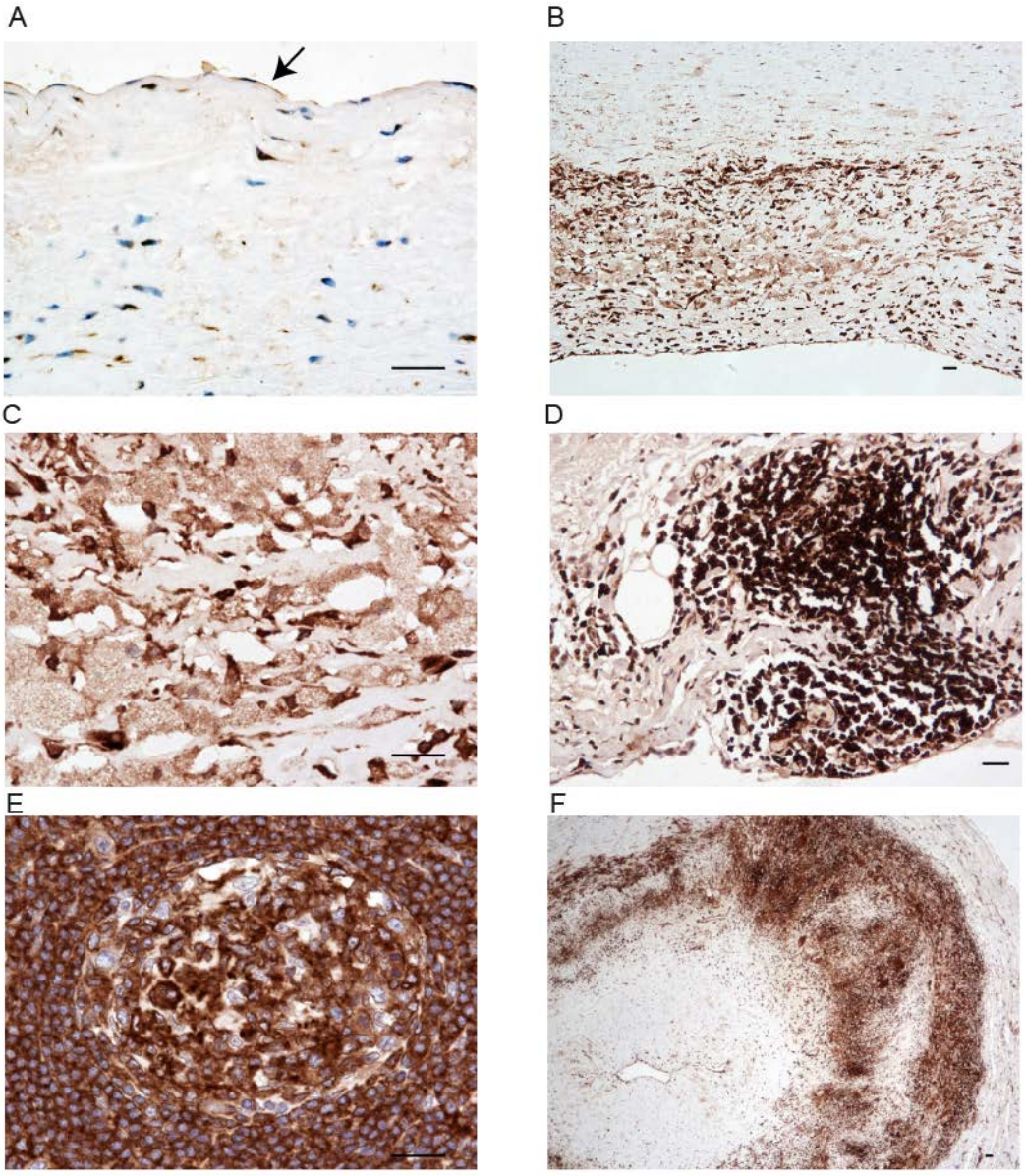


Figure 6-1. PHACTR1 expressing cells are increased in atherosclerotic vessels and in inflammatory arteritis.

(A) Healthy aorta demonstrates PHACTR1-positive endothelium (arrowed) and occasional positive sub-intimal macrophages. (B) Atherosclerotic plaque showing endothelial positivity and positivity in the sub-intimal space, comprising predominantly foam cells (C). Higher power view of foam cells in atherosclerotic

plaque showing cytoplasmic and nuclear staining for PHACTR1. (D) Lymphocytic aggregate in aortic adventitia of atherosclerotic vessel showing strong positivity for PHACTR1 in the cytoplasm and nuclei. (E) Germinal follicle in tonsillar secondary lymphoid tissue showing cytoplasmic staining, but absence of staining in the nuclei of most cells. (F) Temporal artery biopsy in giant cell arteritis showing dense infiltrate of positively staining macrophages and giant cells. (Staining for human PHACTR1, using HRP/DAB, section representative of data from 5 individuals per condition, scale bars indicate 50um. IHC was performed by Dr E. Soilleux, Consultant Histopathologist at the John Radcliffe Hospital, Oxford University NHS Hospitals Trust)

6.2.2 *PHACTR1* is expressed as several transcripts from two separate promoters in vascular and immune cells.

Given that PHACTR1 was abundantly expressed in atherosclerotic vessels *in vivo* I next sought to establish the expression of individual PHACTR1 transcripts in the major cell types comprising atherosclerotic plaques. Initial analysis of the automatically generated ENSEMBL transcript database suggested he presence of multiple PHACTR1 transcripts [323].

To facilitate our experimental studies, I further analyzed a range of genomic data to gain insights into the structure of the most likely underlying expressed transcripts. I compared this analysis to the structure of the experimentally derived RefSeq (NM_030948.2) transcript for PHACTR1 that was cloned from human neuronal tissue (I hereon refer to this transcript as the ‘long’ transcript).

Our analysis of expressed human sequence tag data, chromatin state mapping based on ChIP-seq data and cap analysis of gene expression with high throughput sequencing (CAGE-seq) suggested the presence of a transcriptional start site (TSS) around 240 kb downstream from that of the Refseq transcript (NM_030948.2)(Figure 6-2). Additionally there appeared to be a further TSS close to the 3' end of this transcript[216, 324].

I next mapped the TSS of PHACTR1 transcripts in primary human macrophages, aortic endothelial cells (HAEC), monocytes and peripheral blood mononuclear cells (PBMC) depleted of $CD14^+$ monocytes using Rapid Amplification of cDNA Ends (RACE) (Figure 6-2A, Figure 6-3). This demonstrated that the known TSS of the human RefSeq transcript (NM_030948.2) was not used, but instead the RACE clones confirmed a TSS 240 kb downstream (Figure 6-2A). I then designed primers to amplify transcripts between the RACE TSS and the known 3' end of the RefSeq transcript. A 1674 bp open reading frame was cloned from primary human macrophages and arterial endothelial cells. This transcript encodes a 557 amino acid PHACTR1 isoform that matches the predicted NCBI Reference Sequence: XP_011512694.1 (Figure 6-2A, B, D). On the basis of its nucleotide length, I hereon refer to this transcript as the 'intermediate transcript' (see appendix 9.4 for an alignment of the complete long and intermediate transcript coding sequences). The open reading frame of the intermediate transcript begins partway through exon 3 of NM_030948.2 and contains all subsequent downstream exons. Additionally, the intermediate transcript includes an extra exon (exon 6) that is not present in NM_030948.2 (Figure 6-2B/C). In addition to the intermediate transcript, several smaller transcripts were cloned that variously skipped exons 5-9, but invariably retained the first 2 exons and the final 5 exons. In $CD14^+$-depleted PBMCs, the

transcriptional start site was around 600 bp upstream of that seen in both primary arterial endothelial cells and $CD14^+$ cells, but this did not alter the open reading frame (Figure 6-3).

The combined genomic analysis supported the presence of a distal TSS that matched the beginning of ENSEMBL transcript ENST00000379335 (Figure 6-2A/B). Additional analysis of chromatin mapping data from our published study of primary human macrophages demonstrated an open chromatin site at the TSS that was flanked by H3K27ac marked histones consistent with promoter activity (Figure 6-4)[246]. Using this information, I cloned a 435 bp transcript from macrophages that encoded a 144 amino acid protein. This 'short' transcript shares its last four 4 exons with the longer transcripts, but begins with an extended version of exon 10 that starts 5' to the exon 10 start position seen in the longer transcripts (Figure 6-2B/C).

Overall, in atherosclerosis-related cells PHACTR1 transcripts are expressed from two distinct promoter regions transcribing two transcripts that share the final 4 exons. These data have particular implications for expression based studies of PHACTR1 since primers, microarray probes or antibodies would need to be designed to take into account the unique or overlapping coding sequences. The antibodies used in this study for detection of PHACTR1 protein by immunohistochemistry and western blotting were raised against a peptide mapping to all of the transcripts (Figure 6-6).

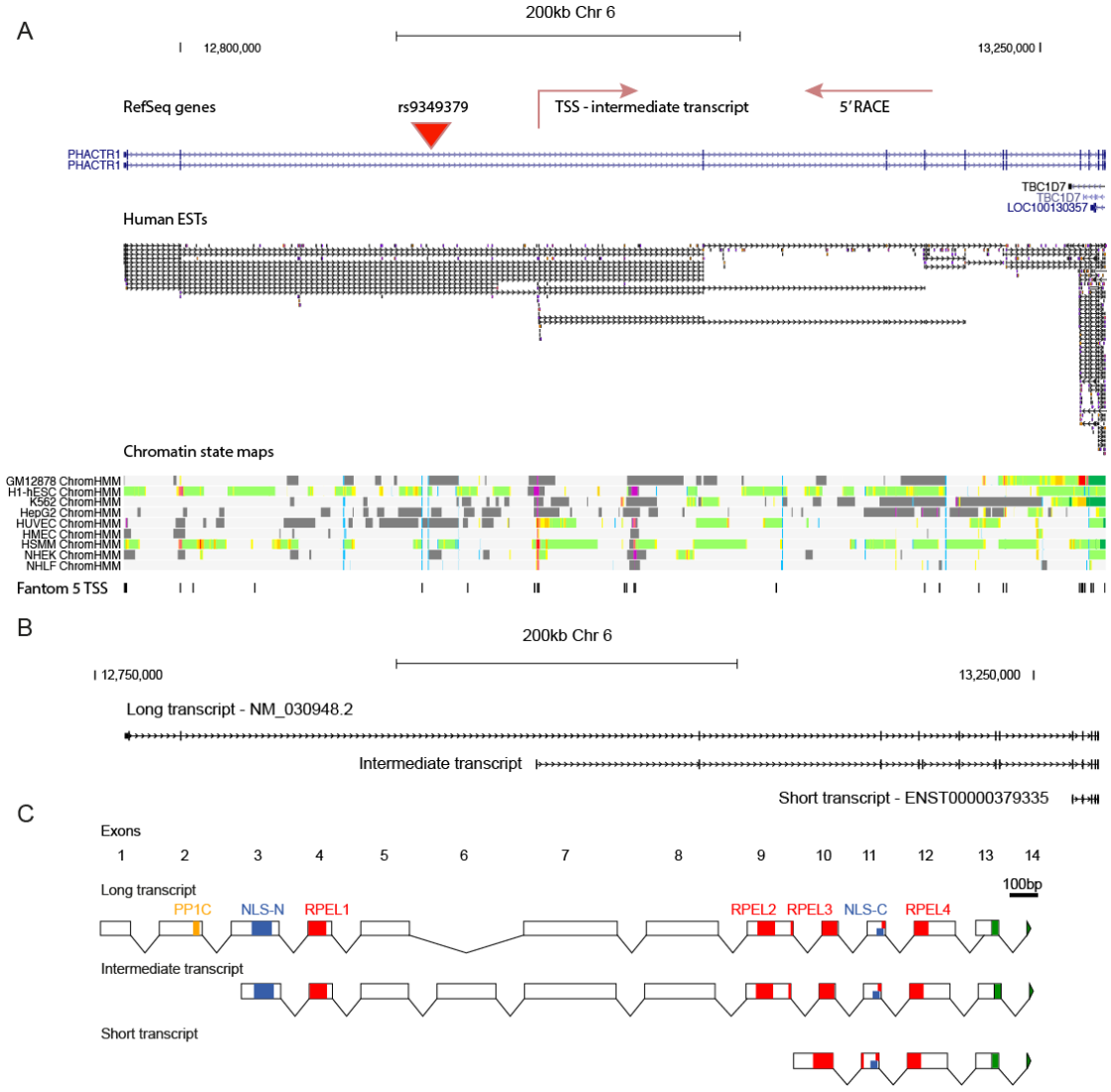


Figure 6-2 The human PHACTR1 gene expresses multiple transcripts from alternative transcription start sites.

(A) The transcript structure for PHACTR1 is shown for RefSeq curated transcripts showing 5 different transcripts in blue. Shown below is human expressed sequence tag data, which is suggestive of a transcription start site halfway along the gene and a further possible site at the 3' end of the gene. These data were compared to chromatin state data that uses multiple ChIP-seq markers to classify the genome into different types of regulatory element or transcription state. Chromatin state maps are shown for 9 cell types with the chromatin state colored (red - promoter, orange - enhancer,

yellow - weak enhancer, blue - insulator, green - transcribed, dark grey - not transcribed). Red regions depicting active promoter elements are seen half way along the gene coinciding with a TSS suggested by EST data. In one cell type a promoter is also seen at the 3' end. These sites coincide with precise TSS identified by CAGE-seq in a variety of human cell types. Taken together these data suggest three main TSS in three main promoter regions. The position of coronary artery disease risk variant, rs9349379, is indicated by the red triangle (B) The transcript structure for the RefSeq transcript is shown uppermost and is aligned to the intermediate transcript that was cloned from macrophages and endothelial cells. The short transcript cloned from macrophages is also shown below. Vertical lines indicate exons. (C) Coding structure of PHACTR1 transcripts with exons drawn to scale and motifs annotated [325]. Motif locations were predicted using the Eukaryotic Linear Motif resource. The region of experimentally proven PP1 binding is shown in green. (D) Sequence alignment of the coding sequence of the long and intermediate PHACTR1 transcripts.

```
CLUSTAL 2.1 multiple sequence alignment


macrophage                TTTTTTTTTTTTTTTTCTT-------------------------------- 19
aortic_endothelial        ------------------------------------------------
PBMC_depleted_of_CD14     TTTTTTTTTTTTTTTTCTAGATTGGGGACGGCCGAGGCAGGTCGGGGGTC 50


macrophage                ------------------TTTTTTTTTTTTTTTTTTTTAAAAAGCCA-- 49
aortic_endothelial        --------------------TTTTTTTTTTTTTTTTGT------------ 18
PBMC_depleted_of_CD14     TGGTTTGGATCCCCGGGTTACTTTCTGTTTCATTTTGAAATGTTCCCTGT 100
                                             *** * *** :**** :

macrophage                ---GGAAACAG------ACCCAG--------AGAGGCGG--ATGCATGTG 80
aortic_endothelial        ---GGAAACAG------ACCCAG--------AGAGGCGG--ATGCATGTG 49
PBMC_depleted_of_CD14     TTCTGCAACCGGTTGCCTCCAATGTCAAGTAAAACGCAGGACTCCTTGGA 150
                              *.***.*      :**.*         *.* **.*  .* *:** .

macrophage                GGAGCAAAGAG-----CTTTGGGCAG-----CCTTTAAGCTGAGGAAGTG 120
aortic_endothelial        GGAGCAAAGAG-----CTTTGGGCAG-----CCTTTAAGCTGAGGAAGTG 89
PBMC_depleted_of_CD14     GACGCCAGCTGGGCTTCTAAGGACTGGGGCTCTGTGAAGCTGAGGAAGTG 200
                          *..**.*. :*     **::**.*:*     *  * **************
```

Figure 6-3 Aligned RACE clone sequences for PHACTR1 transcription start site.

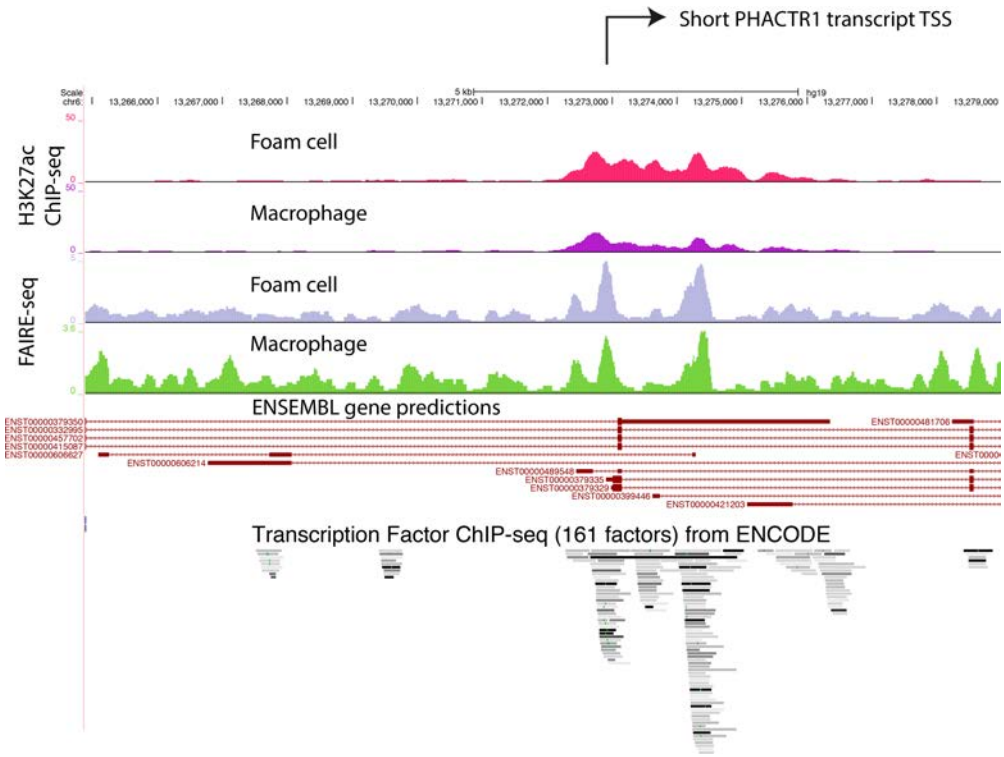


Figure 6-4 Genomic annotations supportive of transcriptional activity at the start site of the ENSEMBL predicted short PHACTR1 coding transcript.

Peaks in the FAIRE-seq signal show open chromatin overlying the putative start site and the zone has H3K27ac marks, that are consistent with promoter or enhancer activity. A condensed view of different ENCODE ChIP-seq derived transcription factor binding sites shows many binding events in this region, including POL2 binding.

6.2.3 PHACTR1 transcripts display tissue-specific differential expression

I studied the mRNA expression in pattern primary human atherosclerosis-related cell types in detail by leveraging our transcript structure data to design RT-PCR primers specific for each of the three transcripts described (Table 16). The long transcript was not expressed in any of the primary cell types tested (Figure 6-5A). The intermediate transcript was expressed in monocytes, macrophages, arterial endothelial cells and

vascular smooth muscle cells, but was barely detected in B cells and was absent from T cells (Figure 6-5A). The short transcript was expressed in primary human monocytes, macrophages, B cells, PBMC-depleted-of-$CD14^+$ cells, and to a lesser extent in T cells, but not in endothelial cells or vascular smooth muscle cells.

I subsequently screened a panel of adult and fetal tissues and the long transcript was only present in adult brain tissue (Figure 6-5B). The intermediate transcript was expressed in colon, pancreas and fetal heart (Figure 6-5B). The short transcript was expressed in adult spleen, PBMCs, colon, fetal spleen and at a low level in the thymus (Figure 6-5B).

Finally, I tested for PHACTR1 expression in a panel of 33 human cell lines that included myeloid, T and B cell lineages (Figure 6-5C). The long transcript was only present in the hypotriploid T47D breast cancer cell line. The intermediate transcript was expressed in T47D, HCA7 (colonic adenocarcinoma), SuSa (testicular tumor), U2OS (osteosarcoma), 293T (human embryonic kidney), NK92C (natural killer cells), Mo7e (megakaryoblastic), MDA435 (melanoma), U937 (histiocytic lymphoma) and HT1080 (fibrosarcoma). The short transcript was expressed in myeloid (THP1, U937, HL60, K562) and B cell tumor lines (Nalm6, Ramos, Raji, and weakly in Daudi) but was absent from T cell (Jurkat, Molt4) or NK cell lines (NK92C). Lower level expression of the short transcript was also seen in Susa cells and 786-0 clear cell renal cancer cells. Overall, these findings indicate that expression of the short transcript is restricted in primary and cell lines to myeloid and B-cells. The intermediate transcript was expressed in immune cells, endothelial cells and a variety of cancer cell lines. By

comparison, the long transcript had the most restricted pattern of expression that was limited to brain tissue and T47D cells.

Table 16 Summary of mRNA expression pattern of PHACTR1 transcripts.

Cell type	Long	Intermediate	Short
$CD14^+$ monocytes	No	Barely detectable	Yes
Macrophages	No	Yes	Yes
Arterial endothelial cells	No	Yes	No
Vascular smooth muscle cells	No	Yes	No
PBMC depleted of $CD14^+$ cells	No	Yes	Yes
B cells ($CD19^+$)	No	Barely detectable	Yes
T cells ($CD3^+$)	No	No	Barely detectable

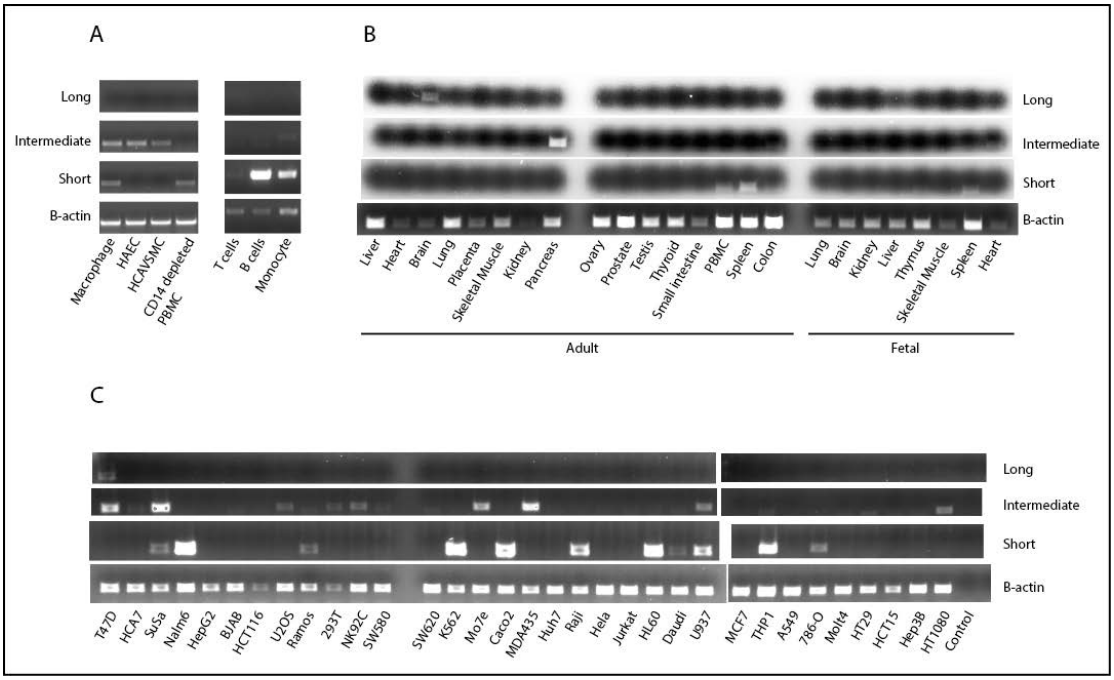


Figure 6-5 PHACTR1 transcript expression patterns in primary cells, tissues and cell lines.

(A) Expression in purified primary human cells showed that the long transcript was not expressed among the cells tested. By contrast, the intermediate transcript showed robust expression in human monocytes, macrophages, arterial endothelial cells (HAEC) and vascular smooth muscle cells (VSMC), but was not expressed by T cells and was barely detectable in B cells and PBMC-depleted of monocytes. (B) Analysis of a human tissue cDNA panel showed the long transcript exclusively expressed in adult brain. The intermediate transcript showed strong expression in the pancreas, but was also detectable in fetal heart. By contrast the short transcript was present in PBMC and spleen, including fetal spleen. (C) A panel of 33 cell lines was tested for PHACTR1 expression. Only one cell line, T47D, an epithelial breast cancer line expressed the long transcript. The intermediate and short transcripts were expressed more widely, but only concurrently in U937 and Susa cells. These findings demonstrate that the 3 transcripts differ markedly in their expression patterns,

indicating that tissue-specific regulatory influences can affect each transcript differently. Notably the short transcript was restricted to expression in immune cells, excluding T cells.

6.2.4 PHACTR1 protein is expressed in primary human macrophages and endothelial cells

To assess PHACTR1 protein expression in primary human cell types involved in atherosclerotic lesions, western blotting was undertaken using an antibody raised against a region of the protein shared by all PHACTR1 isoforms (Figure 6-6).

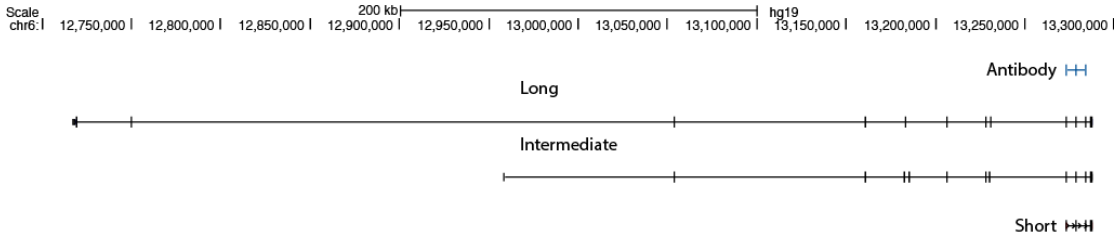


Figure 6-6 Region of PHACTR1 gene used to raise raise commercial rabbit polyclonal Anti-PHACTR1 antibody.

I also assayed ectopically expressed N-terminal myc tagged PHACTR1 clones to help confirm which endogenous transcripts were expressed at the protein level by comparing their electrophoretic migration (Figure 6-7). The isoform encoded by the long transcript, migrated with an apparent molecular weight of about 75 kDa (predicted 66 kDa, pI 6.51) similar to that of the rat brain isoform (Figure 6-7A)[205]. The intermediate isoform is 23 amino acids shorter, but migrated with an apparently

higher molecular weight, equivalent to about 80kDa (predicted 64 kDa, pI 6.81) (Figure 6-7A). The intermediate transcript codes for exon 6 which is missing from the long transcript. It is possible the apparent MW is higher for the intermediate transcript protein because of structural rigidity conferred by the proline rich exon 6 (14 of 69 residues in exon 6 are proline) or additional post-translational modifications. PHACTR1 protein was completely absent from primary vascular smooth muscle cells despite moderate level mRNA expression in these cells (Figure 6-7A). However, in macrophages, foam cells, aortic endothelial cells and oxLDL-exposed aortic endothelial cells, there was strong expression of the intermediate isoform (Figure 6-7A). I did not detect protein expression of the short transcript but detection of this protein was difficult even following overexpression in 293T cells (Figure 6-7 A/B). Although it was not possible to reliably assess the endogenous short isoform, these findings are consistent with our transcript studies and indicate that primary macrophages and aortic endothelial cells express the novel intermediate, but not the long protein isoform.

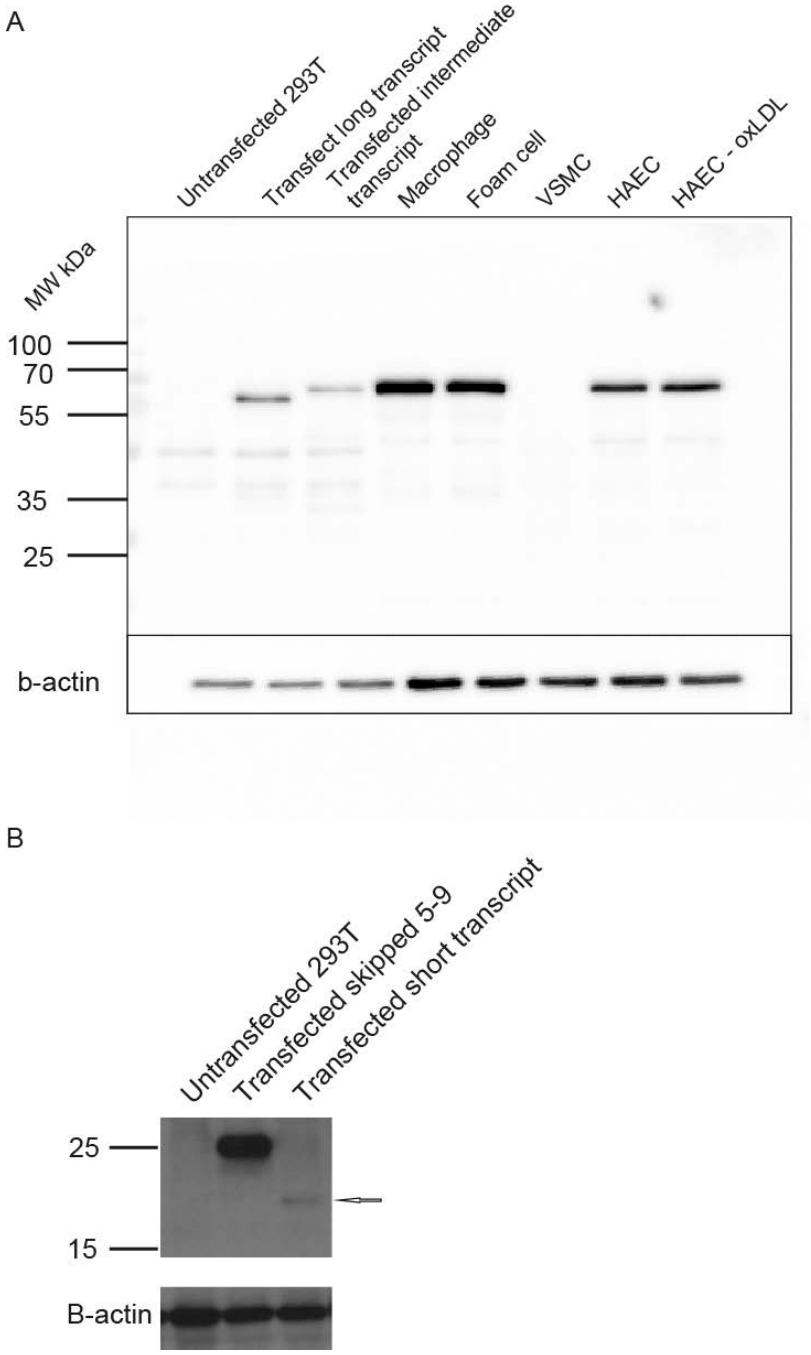


Figure 6-7 PHACTR1 protein consistent with the intermediate transcript is expressed in macrophages and endothelial cells but not vascular smooth muscle cells.

(A) PHACTR1 over-expression in 293T cells transfected with the long and intermediate isoforms compared to endogenous PHACTR1 in primary human cells. The shorter intermediate isoform has an apparently greater size than the long isoform

in 293T cells. Macrophages, foam cells and HAECs expressed PHACTR1 with a similar size to the intermediate isoform. PHACTR1 protein was absent from VSMC. A band corresponding to endogenous expression of the short isoform was not seen in any cell type. (B) Over-expression in 293T cells of the short transcript cloned from macrophages. A cloned form of the intermediate transcript skipping exons 5-9 is also shown for comparison.

6.2.5 PHACTR1 isoforms differ in their subcellular distribution

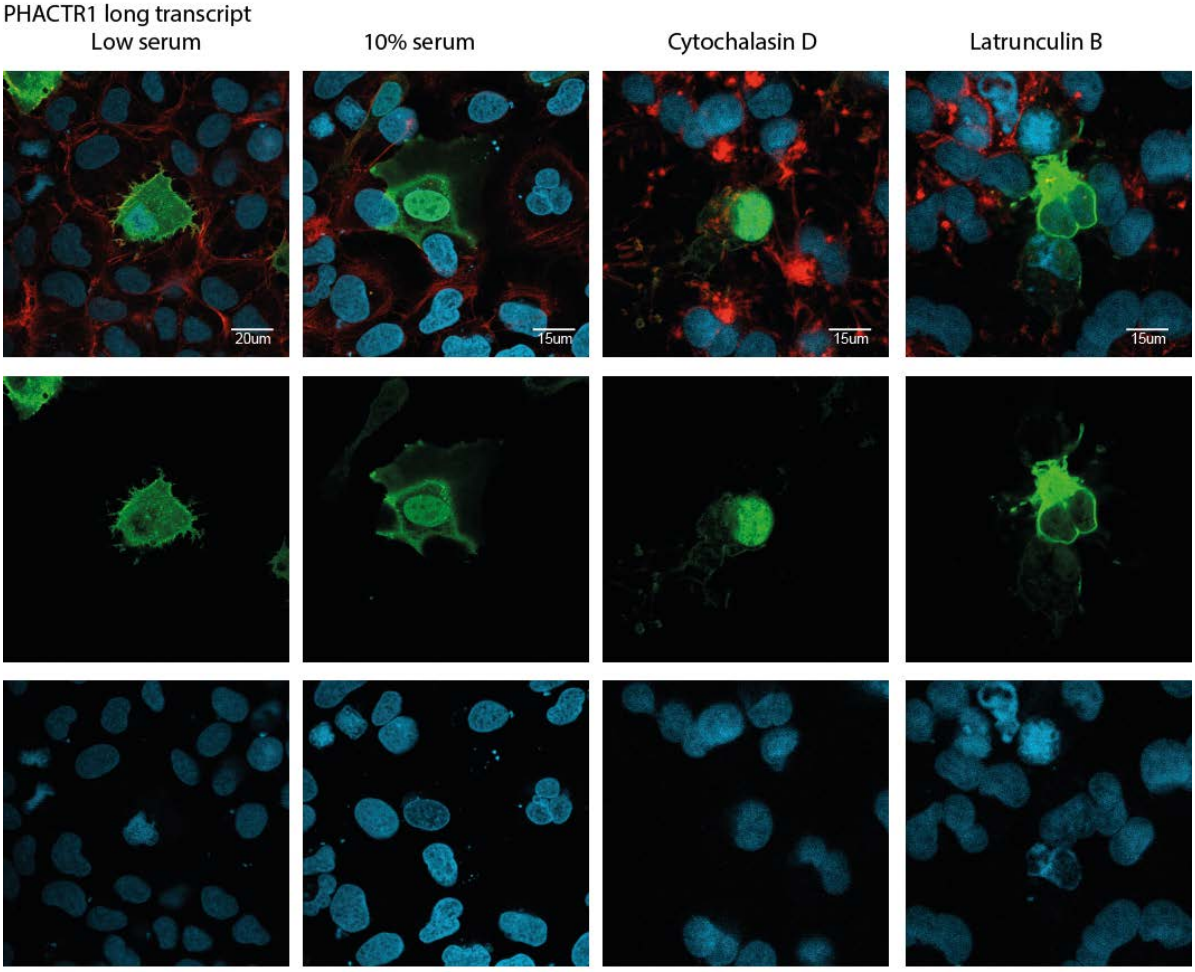
The amino acid sequence of the long human PHACTR1 isoform includes two nuclear localization signal (NLS) motifs (one at each of the N- and C-termini), 4 RPEL actin-binding domains, a predicted PP1-binding motif near the N-terminus, and a C-terminal region that is implicated in PP1 binding (Figure 6-2B)[205]. By comparison the intermediate form contains similar motifs apart from that it lacks the PP1-binding motif near the N-terminus. The short transcript contains only the two C-terminal RPEL domains and lacks the dominant nuclear localization signal motif, which is present at the N-terminus of the longer transcripts (Figure 6-2B)[312]. The long isoform of murine PHACTR1 undergoes cytoplasmic to nuclear translocation in response to a serum-induced transition from G- to F-actin[312]. Therefore, we hypothesized that the intermediate PHACTR1 isoform would undergo similar nuclear translocation in response to actin remodeling, but that the short isoform would not.

We expressed N-terminal myc-tagged PHACTR1 constructs encoding the three different human protein isoforms to assess their sub-cellular localization (Figure

6-8A, B, C, D). The long isoform was present predominantly in the cytoplasm in serum-starved cells, and accumulated in the nucleus in response to serum stimulation, as expected from pre-existing studies using the murine equivalent (Figure 6-8A). Inhibition of actin polymerization from G-actin to F-actin with latrunculin B caused PHACTR1 to remain in the cytoplasm. This is consistent with PHACTR1 binding to cytoplasmic G-actin and so being restrained from moving into the nucleus (Figure 6-8A). The mycotoxin, cytochalasin D, prevents G-actin from binding to RPEL domains in mouse PHACTR1 and this effectively releases PHACTR1 from the cytoplasm allowing enhanced nuclear accumulation [312, 326]. In keeping with this, cytochalasin D induced nuclear localization of the long human PHACTR1 isoform (Figure 6-8A). With the intermediate isoform there was somewhat more PHACTR1 in the nucleus in serum-treated cells, but there was also considerable accumulation in the peri-nuclear region (Figure 6-8B). As expected latrunculin B treatment was not associated with movement of the intermediate isoform into the nucleus. In response to cytochalasin D, the nuclear accumulation of the intermediate isoform did not increase as might have been expected given the nuclear accumulation of the long isoform in response to cytochalasin D. Instead there was concentration of the intermediate isoform of PHACTR1 in the peri-nuclear region. The short transcript was predominantly cytoplasmic with only a small amount present in the nucleus (Figure 6-8C). However, in contrast to the long and intermediate isoforms, serum failed to induce nuclear or peri-nuclear accumulation and the localization of the short isoform was unaffected by either latrunculin B or cytochalasin D.

A

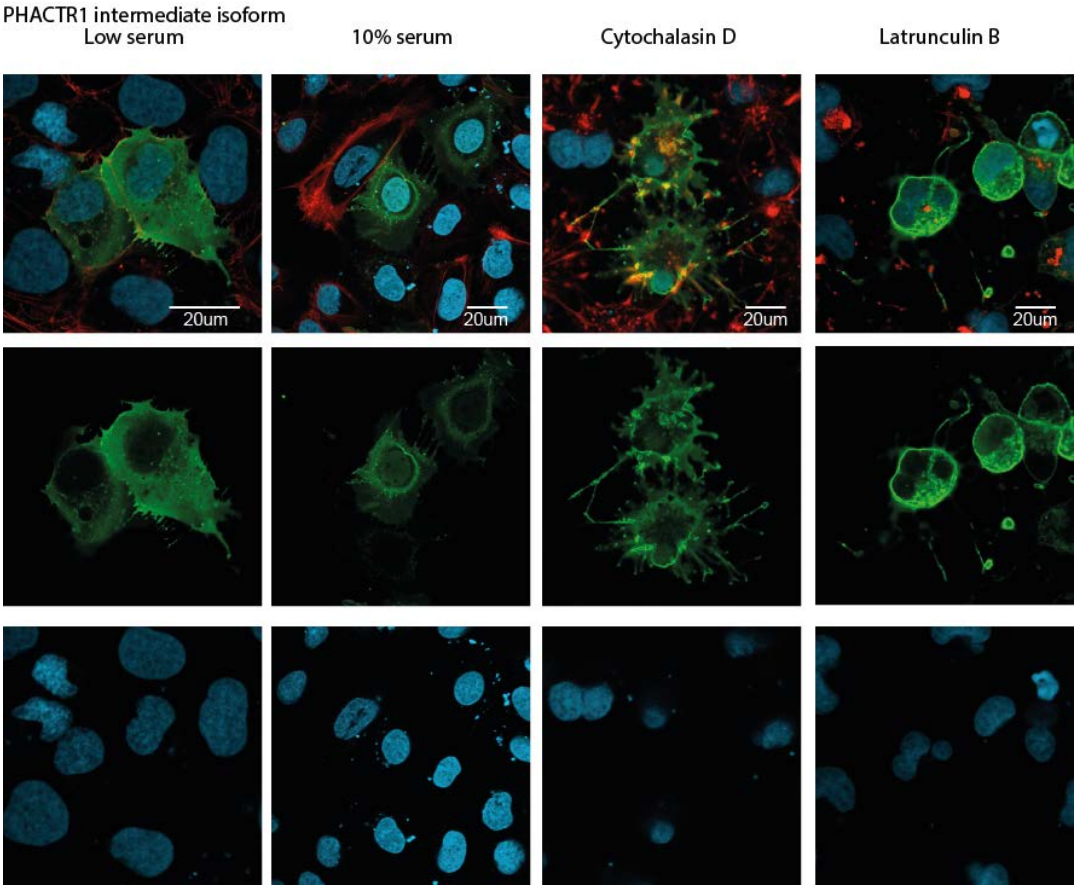
PHACTR1 long transcript



Blue = DAPI
Green = PHACTR1 (AF488)
Red = F-actin (Phalloidin-Rhodamine)

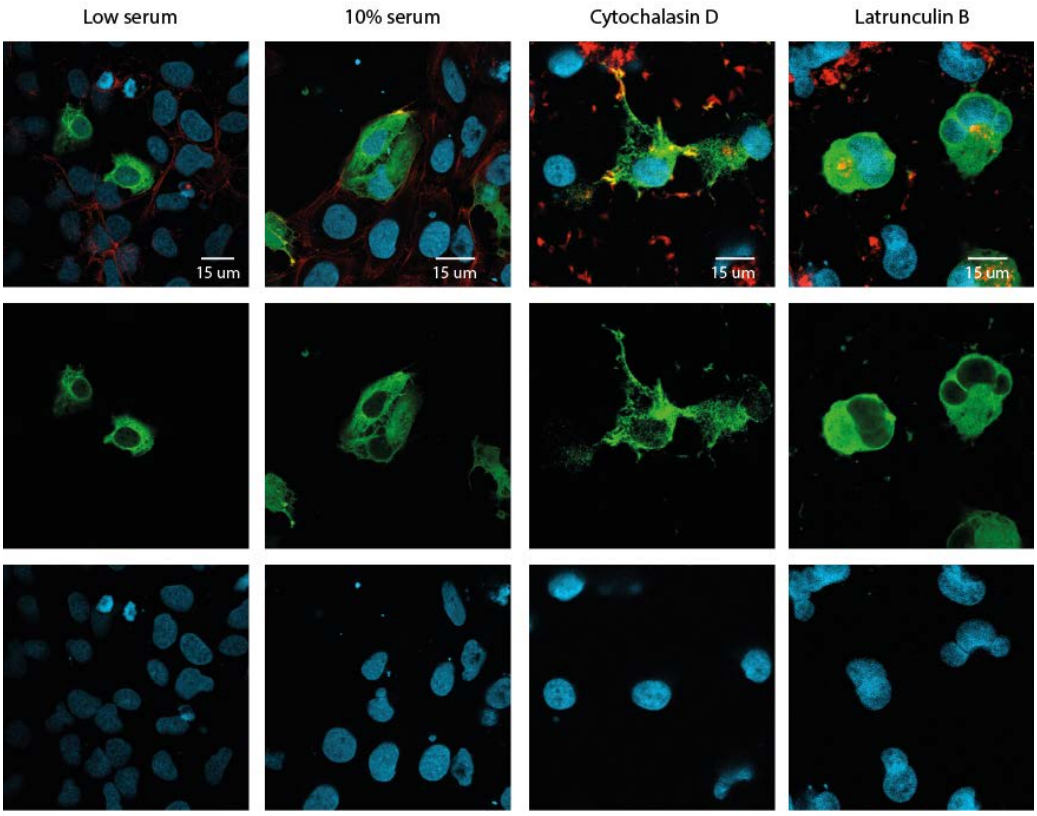
B

PHACTR1 intermediate isoform



C

PHACTR1 short isoform



D

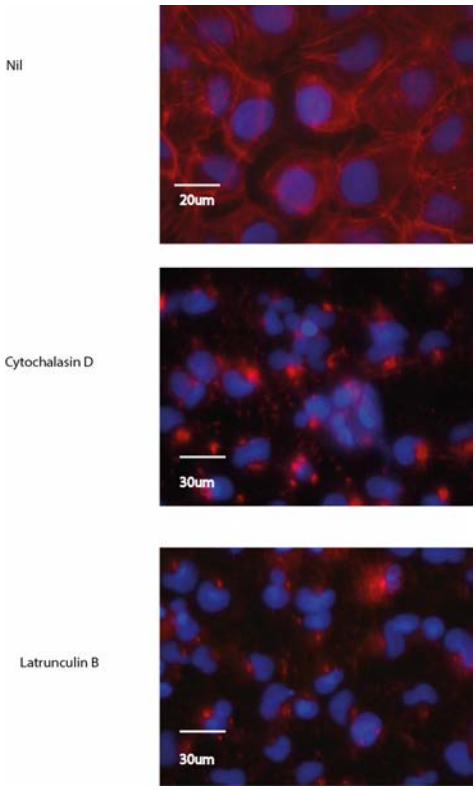


Figure 6-8 The subcellular localization of PHACTR1 isoforms and response to serum, cytochalasin D and latrunculin B.

(A) Long PHACTR1 isoform – addition of serum or cytochalasin D causes nuclear accumulation, which is prevented by latrunculin B. (B) Intermediate isoform – addition of serum caused only minor nuclear accumulation, but with perinuclear accumulation. Cytochalasin D did not increase nuclear accumulation. (C) Short isoform – remained predominantly cytoplasmic under all conditions. (D) As a control U2OS cells were transfected with empty pCDNA3 vector and no myc staining was observed, indicating that the antibody did not have a high level of non-specific reactivity (FITC channel).

6.2.6 PHACTR1 expression is regulated by oxLDL and inflammatory stimuli in macrophages and endothelial cells

We hypothesized that PHACTR1 would be regulated by atherogenic and inflammatory stimuli since our immunohistochemistry showed overall increased expression in atherosclerotic plaque and in arteritis lesions compared to in normal vessels. Lipopolysaccharide (LPS) treatment of primary human macrophages increased expression of the intermediate transcript more than 4-fold, but had a counter effect on the short transcript, which underwent more than a 10-fold reduction in expression (Figure 6-9A). In contrast, oxLDL exerted an opposing effect by increasing expression of the short transcript almost 6-fold and reducing expression of the intermediate transcript slightly, but significantly (Figure 6-9B). To set the expression changes in the context of global changes in gene expression, we analyzed gene expression data from our existing whole genome study of human macrophages [246]. Among 416 oxLDL-induced genes, PHACTR1 was the 36th most upregulated gene. To corroborate our findings, we re-processed and re-analyzed two further published microarray datasets in which human macrophages were treated with oxLDL. In both studies PHACTR1 was within the top 20 most upregulated genes out of over 500 differentially expressed genes [179, 181]. The microarray probes annealed to all transcripts (e.g. Illumina HT12 platform probe ILMN_1736982, sequence AGATAAGCCGTGGACCCGCCTCACCGCTGCAGACAAAGCTGCCATCCGAA, anneals to third from final exon) but our transcript-specific RT-qPCR data (Figure

6-9B) suggest that the upregulation can be attributed to an oxLDL-induced increase in expression of the short transcript. In human aortic endothelial cells, which only express the intermediate transcript, exposure to either TNFα or oxLDL caused increased expression of the intermediate PHACTR1 transcript (Figure 6-9C, D). Overall, these data suggest that in macrophages inflammatory pathways suppress the short transcript and upregulate the intermediate transcript, whereas lipid-responsive pathways upregulate the short transcript. The opposing regulation of the different transcripts in macrophages in response to oxLDL and LPS suggests that each promoter is responsive to different pathways. For example, the short transcript may have a promoter or enhancers that are activated by lipid-based signaling pathways such as by PPARG binding. By contrast LPS would be expected to activate NFKB based pathways suggesting NFKB binding at the intermediate transcript promoter or enhancer activates this transcript but NFKB binding relating to the short transcript suppresses its transcription. Further studies using ChIP, luciferase reporter assays and EMSA evaluating the promoters of each transcript, and enhancers identified by my H3k27ac ChIP-seq may elucidate the precise mechanism governing the differential effects of these stimuli on PHACTR1 transcripts.

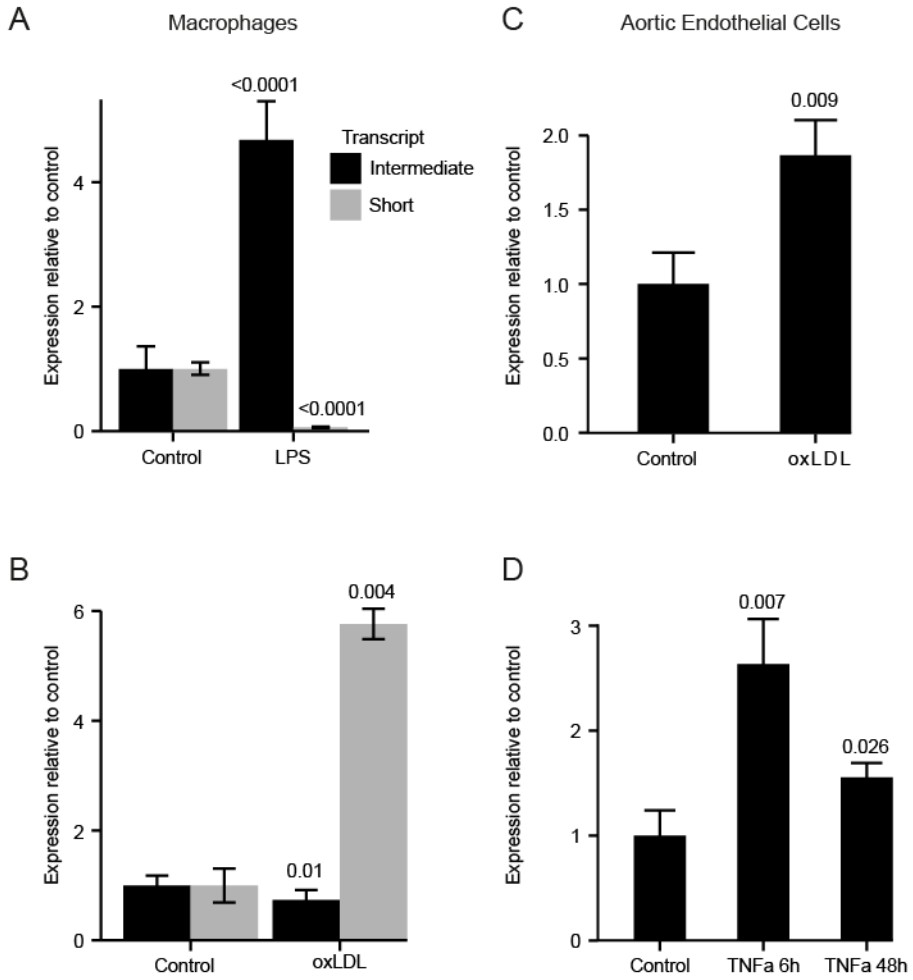


Figure 6-9 PHACTR1 transcripts are differentially regulated by atherogenic and inflammatory stimuli.

(A) Effect of LPS on mRNA expression in primary human macrophages. The intermediate transcript was upregulated, but the short transcript was suppressed. (B) Exposure of macrophages to oxLDL caused modest down-regulation of the intermediate transcript and strong upregulation of the short transcript. (C) The intermediate transcript was upregulated by oxLDL in aortic endothelial cells. (D) TNF-alpha caused upregulation of PHACTR1 at 6 and 48-hour time points in aortic endothelial cells. The short transcript was not expressed in aortic endothelial cells (n=3).

We next tested whether there was a specific correlation between expression of the intermediate and short transcripts in macrophages and foam cells. To explore the relationship between the two transcripts, we determined the level of expression of the intermediate and short transcripts in macrophages from 19 healthy individuals. We found a significant inverse correlation (R^2 0.25, P value 0.029) such that a higher level of the intermediate transcript was associated with reduced expression of the downstream short transcript (Figure 6-10A). We further examined whether this relationship was maintained in response to oxLDL. Treatment of macrophages with oxLDL reversed the resting inverse correlation, producing a significant positive correlation (R^2 0.43, P value 0.002) between expression of each transcript, such that a higher level of the intermediate transcript was associated with a higher level of the short transcript (Figure 6-10B). These findings indicate that the atherogenic stimulus, oxLDL, causes changes in the relationship between expression of each transcript and so disrupts the normal steady state relationship between the two PHACTR1 transcripts.

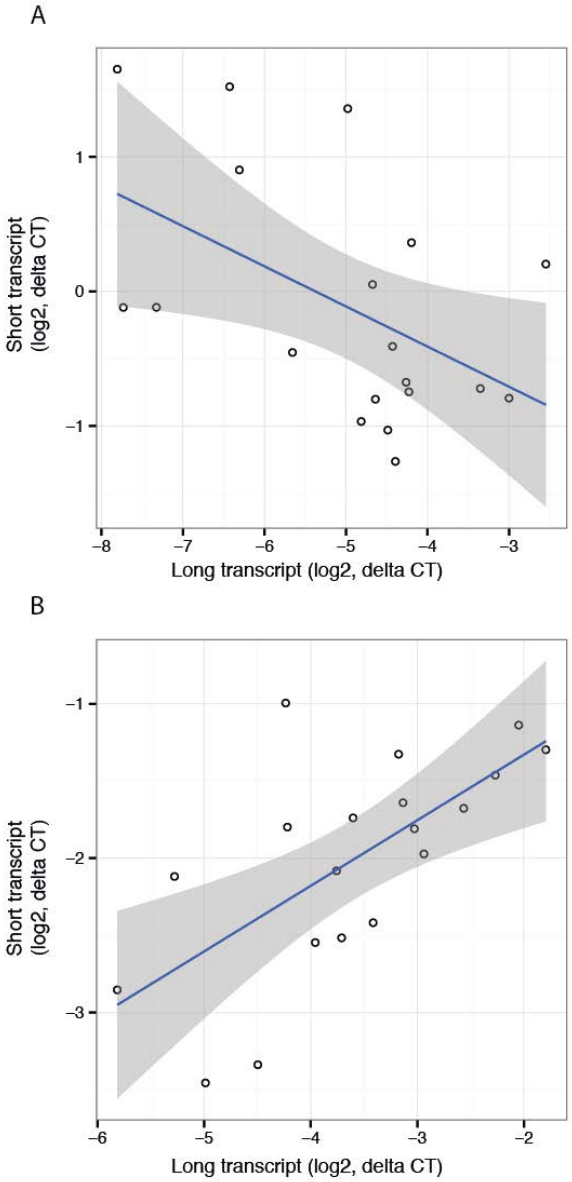


Figure 6-10 The relationship between expression of the short and intermediate PHACTR1 transcripts.

(A) Scatterplot showing mRNA level expression of PHACTR1 transcripts in human macrophages in 19 individuals. Expression of the short transcript is inversely related to the intermediate transcript (delta CT indicates difference between PHACTR1 CT and GAPDH CT). (B) Expression of PHACTR1 transcripts in oxLDL-induced foam cells from 19 individuals. OxLDL reversed the natural relationship causing a positive correlation between expression of both transcripts.

6.2.7 Genetic regulation of PHACTR1 expression by CAD SNP rs9349379

GWAS have identified the single nucleotide polymorphism, rs9349379, as associated with the risk of CAD[159, 270]. The mechanism by which alternative alleles of this SNP influence CAD risk is unknown, but since it is within an intronic region of PHACTR1, and is not in high LD with any coding SNPs, it may affect transcriptional regulation of PHACTR1 (Figure 6-2A). We tested for an association between the alleles of rs9349379 and expression of the short and intermediate transcripts in macrophages and $CD14^+$-depleted PBMCs. The SNP has either A or G alleles with G being the minor allele with a minor allele frequency of 0.3774 [119]. Transcript-specific RT-qPCR was used to quantify the intermediate and short transcripts in macrophages from individuals with different alleles at this SNP. This demonstrated a significant allelic effect on expression of the short transcript with the A allele associated with higher levels of the short transcript (Figure 6-11A). Macrophages from individuals homozygous for the protective A allele had 2.5-fold higher expression than macrophages from individuals homozygous for the G allele, which is the coronary artery disease risk-associated allele. There was a trend towards reduced expression of the intermediate transcript with the A allele, but this did not reach significance (Figure 6-11). In $CD14^+$-depleted PBMCs no association was detectable between genotype and expression of either the short or intermediate transcript (Figure 6-11C, D). Overall, these results demonstrate that rs9349379 is both a CAD risk-associated variant and an expression quantitative trait locus (eQTL) for PHACTR1 in macrophages, but not in $CD14^+$-depleted PBMC. Therefore, decreased expression of the short transcript in macrophages is associated with CAD. This genetic risk effect recapitulates the effect of endotoxin, which also caused suppression of the short

transcript (Figure 6-9A). The mechanism whereby rs9349379 is associated with changes in transcriptional regulation has not been extensively investigated. There are no known open chromatin sites, or enhancer marks in macrophages or endothelial cells at the rs9349379 location[216, 246]. Binding of transcription factor MEF2A/C to the A allele but not the G allele has been observed using *in vitro* electrophoretic mobility shift assays (EMSA) with nuclear extract from human umbilical vein endothelial cells [308]. Given our findings of an eQTL in macrophages, we undertook EMSA using nuclear extract from human macrophages and demonstrated differential MEF2 binding to an oligonucleotide probes containing each of the SNP alleles (Figure 6-12). The EMSA demonstrated protein binding, consistent with MEF2, to the A allele but not to the G allele. To assess the transcriptional regulatory activity of the genomic location and effects of the SNP alleles we performed luciferase assays in human THP1 macrophages using reporter plasmids containing a 223 genomic DNA segment varying only at the SNP. We did not detect significant enhancer activity with either allele and did not detect a difference between the alleles (Fold increase A/G = 18.5%, P=0.26, n=4 biological replicates)(Figure 6-13). Overall, these findings demonstrate that the CAD risk effect of genetic variants at rs9349379 may be operational in macrophages by affecting expression of the short PHACTR1 transcript but it remains unclear whether the SNP itself operates by affecting enhancer function through MEF2 binding.

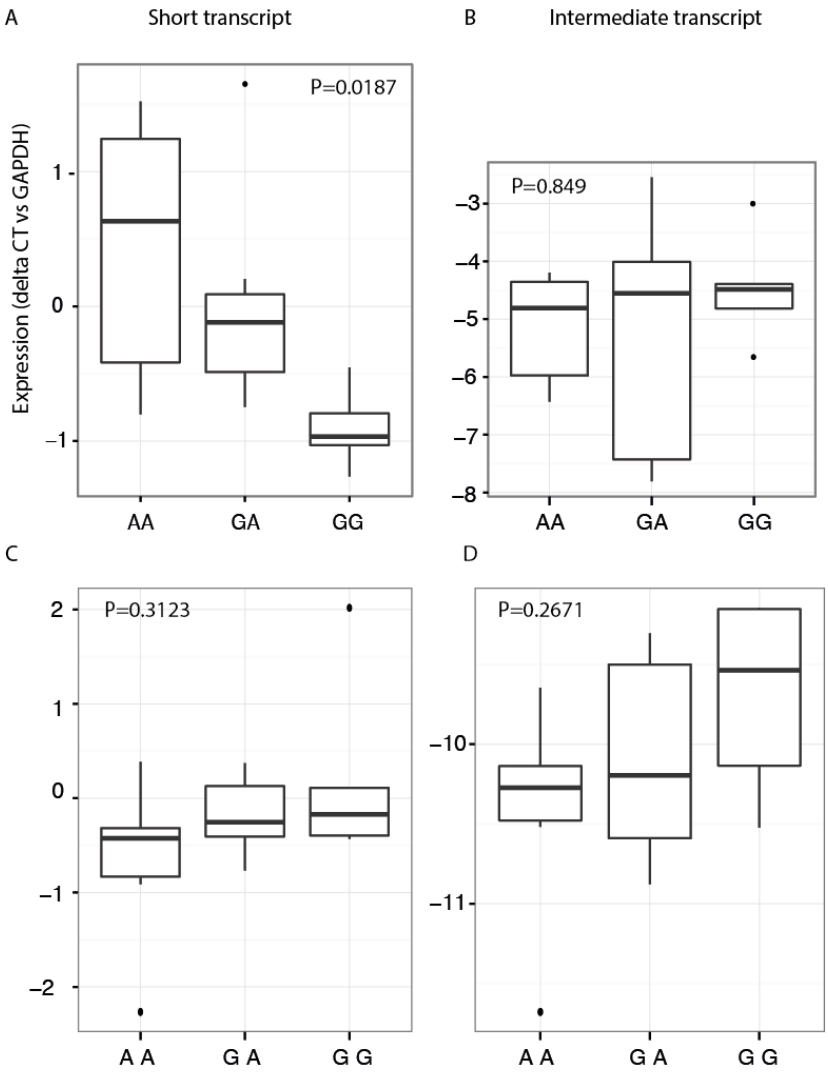


Figure 6-11 Genotype of CAD risk SNP rs9349379 affects expression of the short PHACTR1 transcript in macrophages.

PHACTR1 expression in primary human macrophages relative to GAPDH. (A) A significant genotype-expression level correlation was demonstrated in primary human macrophages (n=19). Individuals with the AA genotype had the highest level of expression of the short PHACTR1 transcript and subjects with the GG genotype had the lowest expression. (B) There was no significant correlation with the long transcript (n=19). (C, D) Primary human $CD14^{+}$-depleted PBMCs did not show a significant correlation between expression and genotype of rs9349379 (n=20).

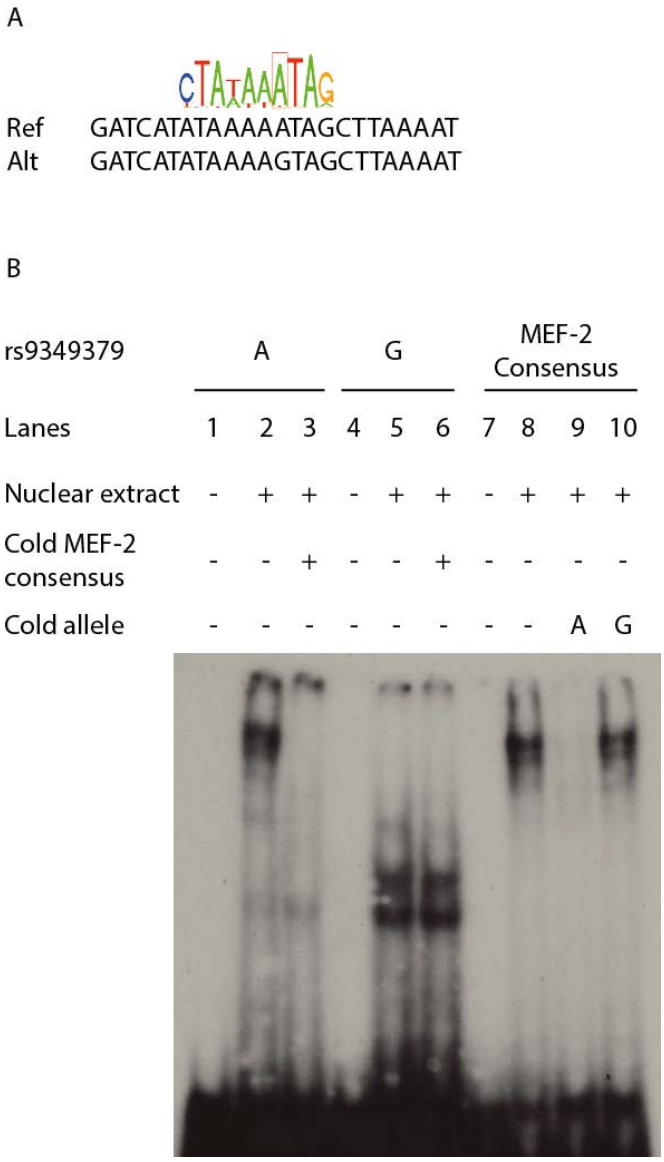


Figure 6-12 Electrophoretic mobility shift assay for macrophage nuclear protein binding at the rs9349379 SNP shows binding of MEF2 only to the A allele.

(A) Probes were designed to be 25bp in length with the SNP centrally positioned. EMSAs were performed using THP1 macrophage nuclear extract. The motif for MEF2 is shown aligned to the probe sequence and the SNP site is marked by a red box. (B) An upper band indicates nuclear protein binding with the A allele that is competed off with a cold probe comprising the MEF2 consensus sequence. With the

G allele this band is absent. A smaller band is likely non-specific as it would indicate a small protein, is not competed off by excess cold consensus probe and is not replicated by the consensus probe. Equally it is possible that an alternative protein binds to the G allele. When the consensus probe is radio-labelled it produces an identical band to the probe containing the endogenous sequence with the A allele. The cold probe with endogenous sequence at the SNP with the A allele is able to completely compete off binding to the MEF2 consensus probe, whereas a cold probe with the G allele has no effect. These findings indicate that nuclear extract from macrophages binds to the A allele, but not the G allele with features indicating that the bound species is MEF2.

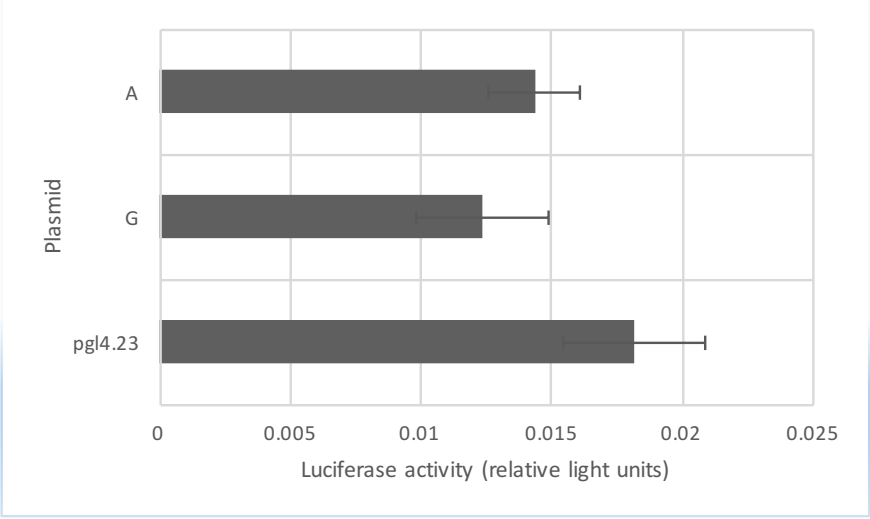


Figure 6-13 Transcriptional activity at the rs9349379 SNP region in assayed in human THP1 macrophages.

The two plasmids containing the SNP region differ only at a single nucleotide corresponding to the SNP alleles. There is a lack of enhancer activity with both alleles

when compared to the control empty plasmid pgl4.23. There is no significant difference between the two SNP alleles (n=4 biological replicates).

6.3 Discussion

For CAD, one of the most robustly described GWAS risk loci is at the *PHACTR1* gene locus where several studies have identified intronic SNPs associated with the risk of myocardial infarction, coronary stenosis and coronary calcification [159, 270, 306, 307]. The precise role of PHACTR1 in cellular physiology remains unknown but an overall understanding of PHACTR1 is beginning to emerge. PHACTR1 is a G-actin sensor that remains localized to the cytoplasm until cells receive a 'serum' type stimulus. Serum induces remodeling of actin into filamentous F actin that leads to the release of PHACTR1. Unbound PHACTR1 is then free to move into the nucleus using the importin pathway[327]. Upon nuclear localization it likely binds to nuclear protein phosphatase 1, thereby regulating its activity. Regulated protein phosphatase 1 activity may effect the phosphorylation status and activity of certain transcription factors. Although the effect of PHACTR1 nuclear translocation has not been extensively studied I believe this is a plausible model to explain a cellular role for PHACTR1. Thus far serum is the only physiological stimulus known to induce nuclear accumulation of PHACTR1. To understand better the role of PHACTR1 in atherosclerosis, we investigated PHACTR1 expression, regulation, subcellular localization and the relationship between allelic variation and expression.

Using immunohistochemistry, we demonstrated increased expression of PHACTR1 in atherosclerotic lesions *in vivo*. In human arteries the distribution was both nuclear and cytoplasmic in macrophages, foam cells and adventitial lymphocytes, whereas in

secondary lymphoid tissue lymphocytes, PHACTR1 was excluded from the nucleus. Certain inflammatory lipids accumulate in atherosclerotic plaques, including lysophosphatidic acid (LPA), which activates RhoA signaling causing G- to F-actin polymerization and so might promote PHACTR1 nuclear accumulation [295]. A limitation of the immunohistochemistry data is that the antibody we used cannot distinguish between the different isoforms and would potentially detect all of the isoforms I have studied in this manuscript.

Alternative pre-mRNA splicing and alternative transcription start sites allow individual genes to encode multiple isoforms. The previously reported PHACTR1 transcript was isolated from brain and encodes a 580 amino acid protein[205, 310]. We only detected the original PHACTR1 RefSeq transcript in brain and one tumor cell line. Previous molecular and functional studies employing this transcript may have limited relevance to understanding PHACTR1 beyond the nervous system. In cell types involved in atherosclerotic lesions, we found PHACTR1 was transcribed from a more distal TSS, giving rise to an uncharacterized transcript. This transcript was the prevalent protein isoform in human immune and endothelial cells and the encoded protein maintains the 4 previously characterized actin-binding RPEL domains and NLS regions. Previous investigators noted the absence of a PP1 binding motif in rat PHACTR1, but using the Eukaryotic Linear Motif resource (www.elm.eu.org) we identified a PP1C binding motif towards the N-terminus of the human long isoform, but not the intermediate isoform [205]. However, previous experimental work indicates that PP1 binding occurs near the C-terminus, which is shared by both isoforms[205]. Thus the role of the predicted PP1 binding motif in the long isoform remains unknown. Transcription of the intermediate transcript from the

alternative TSS is differentially regulated across multiple cell types and this expression is rarely associated with concordant expression of the short transcript in non-immune cells. Interestingly, although PHACTR1 mRNA was present in vascular smooth muscle cells, it was absent at the protein level suggesting additional post-transcriptional regulation. The actin-binding capability of PHACTR1 may be undesirable in these cells, which have specialized contractile functions.

In addition, we found transcription of a short transcript from an even more distal TSS. The encoded protein would only contain two RPEL motifs and lack the N-terminal NLS. The expression pattern of this transcript differed markedly from the longer forms and expression was only detected in immune cells, principally macrophages and B-lymphocytes. Macrophages were notable as the only cell type with high-level expression of the intermediate and short transcript. This limited distribution in immune cells is consistent with a role for the short transcript in immunity and inflammation. We were unable to detect protein expressed from the endogenous short transcript in the cell types studied despite using an antibody raised against peptides encompassing most of its coding region. This may indicate post-transcriptional effects, rapid turnover or technical limitations of the available antibodies.

The nuclear membrane separates nuclear transcription from cytoplasmic translation and proteins can move across the nuclear membrane via nuclear pore complexes [328]. Previous work demonstrated that the mouse long isoform translocates to the nucleus in response to serum-induced actin remodeling [312]. The RPEL actin-binding domains bind to G-actin [312, 314] and upon RhoA-dependent G-actin remodeling to filamentous or F-actin, PHACTR1 dissociates from actin and moves

into the nucleus [312]. PHACTR1 binds protein phosphatase 1 and inhibits its activity *in vitro* suggesting a role in regulating the activity of nuclear phosphatases, with possible consequences for transcriptional regulation[205]. We anticipated that the intermediate isoform would undergo serum-induced nuclear accumulation as the long isoform does, because both isoforms share identical nuclear localization motifs and RPEL domains. However, serum failed to induce the profound nuclear accumulation seen with the long isoform. It is possible that the use of the proline-rich exon 6 in the intermediate transcript may partially inhibit nuclear localization. The short transcript encodes a protein with a predicted size of less than 40 kDa and so potentially small enough to enter the nucleus passively[328]. This protein was mainly cytoplasmic and did not accumulate in the nucleus in response to either serum or cytochalasin D. This is consistent with the lack of the N-terminal nuclear localization signal, which is the dominant nuclear localization signal in the long mouse transcript [312].

Atherogenic stimuli can induce alternative transcripts of genes such as the HMGCR lipid metabolism gene [329]. Similarly, oxLDL exposure caused upregulation of the short transcript and suppression of the intermediate transcript in macrophages. Lipopolysaccharide caused upregulation of the intermediate transcript and suppression of the short transcript in macrophages. These findings demonstrate that PHACTR1 expression is responsive to inflammatory stimuli and to lipid or oxidative stress pathways activated by oxLDL. Since macrophages express both short and intermediate transcripts we examined the relationship between the two transcripts across different individuals. There was an inverse relationship between levels of the two transcripts, which was reversed following oxLDL treatment. As the transcripts are expressed from different TSSs with distinct promoter elements, this relationship

might arise from *trans* regulation of both promoters by a common pathway or *cis* regulation, perhaps by transcriptional interference.

Knockdown of PHACTR1 in endothelial cells impairs aspects of actin polymerization and other cellular functions involved in angiogenesis [322]. In aortic endothelial cells only the intermediate transcript was expressed and this was upregulated by both TNF-alpha and oxLDL. In previous studies with umbilical vein endothelial cells TNF-alpha had no effect on PHACTR1 expression [308, 322]. However, we found that PHACTR1 was upregulated by TNF-alpha in primary aortic arterial endothelial cells, emphasizing the importance of the choice of cell type for these experiments. These findings raise the possibility that oxLDL and TNF-alpha could influence angiogenesis or atherosclerosis by affecting PHACTR1 expression in arterial endothelial cells.

The rs9349379 SNP is robustly associated with CAD and a fine mapping study highlighted it as the likely causal variant in the region [308]. We have established a new association in macrophages between expression of the previously uncharacterized, short PHACTR1 transcript and genotype at rs9349379. An eQTL for PHACTR1 expression has recently been demonstrated in right coronary artery tissue, but the cellular origin of this effect is unclear due to the cellular heterogeneity of the samples[308]. In addition, the different transcripts were not distinguishable because the qPCR primers annealed to both the long and intermediate transcripts and the short transcript was not studied [308]. An eQTL study in endothelial cells using primers that would detect only the longer transcripts did not find a significant association with variants at the rs12526453 SNP, which is in LD with rs9349379 (R^2 with rs9349379 = 0.32)[330]. The genotype-expression correlation that we have identified may only

exist in macrophages, since we did not detect a correlation in $CD14^{+}$-depleted PBMCs and the short transcript was not expressed in endothelial cells or vascular smooth muscle cells. Two other large-scale eQTL studies in monocytes and macrophages used microarray technology and in both cases the probe detecting expression would have annealed to multiple PHACTR1 transcripts and so not yielded specific information about each transcript [136, 300]. An eQTL study in PBMCs of CAD candidate genes, including PHACTR1, failed to detect significant PHACTR1 expression, but the PHACTR1 Taqman assay used would only have measured the long transcript which we have demonstrated is not expressed in these cells [331].

Motif analysis predicts that the minor G risk allele abolishes binding of the transcription factor MEF2. Our microarray data indicate that MEF2A and C are expressed in macrophages and foam cells [246]. Using human umbilical endothelial cell nuclear extract the G allele was shown to disrupt MEF2A and B binding *in vitro* using an EMSA [308]. We found a similar effect *in vitro* using macrophage nuclear extract. To date, limited ENCODE MEF2A and MEF2C ChIP-seq data have not demonstrated *in vivo* binding at this site and MEF2 ChIP has not been performed in endothelial cells or macrophages. Whether MEF2 binding occurs *in vivo* is therefore unknown. The lack of enhancer marks or open chromatin at the SNP in macrophages and endothelial cells suggests that the site lacks regulatory activity [143, 246]. Our luciferase reporter assays suggested the SNP site does not function as an enhancer in human THP1 macrophages.

Suppression of the short PHACTR1 transcript by pro-inflammatory stimuli mirrors the effect of the CAD risk allele, which is associated with reduced expression of the

short transcript compared to that seen with the protective allele. This suggests that having the risk genotype is, in this sense, equivalent to persistent LPS stimulation. As the short transcript was only expressed in immune cells, these data imply that the short transcript could play a role in the macrophage inflammatory response in atherosclerotic plaques. Up-regulation of the short PHACTR1 transcript in response to oxLDL exposure may be a protective response that could be attenuated in individuals with the risk genotype.

Following up GWAS results in CAD to identify underlying mechanisms mediating genetic risk is a major challenge of contemporary atherosclerosis research, especially for loci containing poorly characterized genes. We establish *PHACTR1* as an important gene expressed in atherosclerotic lesions and as the candidate gene at the *PHACTR1* CAD locus by demonstrating genetic regulation of a novel transcript and regulation by atherogenic stimuli. Our data show that previous molecular studies of PHACTR1 involved a transcript not expressed in cells involved in atherosclerotic lesions. We have identified novel intermediate and short transcripts as the dominant forms expressed in atherosclerosis. Our genotype-expression analysis suggests macrophages as a site of action of the CAD associated genetic variant and macrophages were the dominant cell type expressing PHACTR1 in human atherosclerotic plaques.

7 Mincle: A receptor for lipid species on macrophages

7.1 Introduction

C-type lectin receptors (CLR) are a group of germ-line encoded pattern recognition receptors (PRR) that contain at least one C-type lectin domain (CTLD). CLRs variously recognize carbohydrate or lipid based pathogen-associated molecular patterns (PAMPs), or self-antigens termed damage associated molecular patterns (DAMPs)[332]. C-type lectins are predominantly expressed on myeloid immune cells and have roles in both pathogen immune responses and in inflammation associated with binding to endogenous ligands[333]. The C-type designation refers to the requirement for calcium ions for carbohydrate binding[332]. C-type lectins have been implicated in the pathogenesis of coronary artery disease. For example, lectin-type oxidized LDL receptor 1 (LOX1) binds to and internalizes oxLDL[334]. LOX-1 mice knockout models demonstrate reduced susceptibility to atherosclerosis s[335] and polymorphisms in the 3' untranslated region of the OLR1 gene are associated with angiographic CAD severity, though the precise mechanism is unknown[336].

Macrophage-inducible C-type lectin (MINCLE) is a member of the C-type lectin super-family that binds to trehalose di-mycolate (TDM)—a major mycobacterial cell wall glycolipid—and transduces activating signals[337, 338]. MINCLE is not known to have a direct role in atherosclerosis but has been implicated in obesity-induced inflammation and hypothesized to detect modified lipids in adipose tissue[339]. Our genome-wide expression study showed that MINCLE was upregulated by oxLDL in macrophages[246]. We hypothesized that MINCLE may recognize oxLDL-related

lipid species and profiled the binding characteristics of MINCLE to modified lipoproteins and a variety of systematically synthesized glycolipids.

Human MINCLE is a 219 single pass type II transmembrane protein encoded by the MINCLE gene on chromosome 12[340, 341]. It contains a single carbohydrate recognition domain (CRD). The transmembrane stalk region has a conserved, positively charged arginine residue that allows it to couple with the immune tyrosine-based activation motif (ITAM) containing protein,

Fc receptor common γ-chain (FcRγ), which has a negatively charged residue in its transmembrane region[342]. Stimulation of MINCLE with an activating antibody triggers FcRγ-dependent induction of pro-inflammatory cytokines from mouse macrophages[342]. The downstream activation pathway beyond the MINCLE-FcRγ complex involves phosphorylation of spleen tyrosine kinase (syk) and requires Caspase recruitment domain-containing protein 9 (CARD9)[342]. Mincle does not appear to be directly involved in phagocytosis[342].

The structures of the extracellular CRD domain of human and bovine MINCLE have been solved using crystallography [343, 344]. These studies have provided insights into the underlying mechanism by which MINCLE binds to its ligand TDM. TDM itself consists of trehalose, a disaccharide, and two mycolyl chains, which consist of branched acyl chains. Neither trehalose alone nor mycolate are sufficient to activate Mincle suggesting that Mincle is specific for the ester bond between trehalose and mycolic acid [337]. Crystallographic analysis of MINCLE has shown a hydrophobic groove extending from the carbohydrate binding area that is thought to accommodate a mycolyl chain of TDM [343, 344]. There is an additional area for binding of the second sugar moiety in trehalose. Surface plasmon resonance experiments have

suggested that only one acyl/mycolyl chain of at least 10 carbons in length is required for efficient binding [343]. In addition to TDM, several other related synthetic compounds have been shown to bind MINCLE; trehalose-6 6-dibehenate (TDB), a synthetic analogue of TDM, glycerol monomycolate (GMM) which has a single mycolyl chain and simple glycerol head[345] and trehalose monoesters with either a 22 and 26 length acyl chain[346]. Mincle is also a receptor for several fungal pathogens including Candida albicans[347, 348], Malassezia[349], and Fonsecaea pedrosoi[350]. For Malassezia the Mincle ligands are thought to be a glyceroglycolipid termed 44-1 and a mannosyl fatty acid linked to mannitol termed 44-2 [351].

Expression studies have focused primarily on blood cells where Mincle is predominantly expressed on macrophages, dendritic cells, neutrophils, B cells and to some extent on mast cells [341, 348, 352, 353]. Mincle expression is upregulated by inflammatory stimuli such as LPS, TDM, TLR ligand CpG [304, 341, 353] and exposure to certain pathogens including Candida[347]. Mouse Mincle mRNA expression was increased by exposure of macrophages to adipocytes as well as by stimulation with palmitate, a saturated fatty acid[354]. Mincle expression was also increased in human adipose tissue and circulating monocytes in people with obesity [354].

In addition to the study of Mincle-ligand interactions, functional studies using animal disease models have shown that Mincle is important in the response to tuberculosis, albeit in a complex way, and fungal infections[337, 347, 355, 356]. TDB has been developed as a vaccine adjuvant and refinement of synthetic Mincle ligands could

further aid the development of potent and safe vaccines [357-359]. As well as being important in the response to mycobacterial infection Mincle has been shown to bind to Sin3A-Associated Protein, 130kDa (SAP130), which is released by dying cells[342]. SAP130 triggers an inflammatory response through MINCLE, essentially causing sterile inflammation[342]. The exact nature of how SAP130 binds Mincle has not been described. Recent experiments in mouse models of obesity have suggested that MINCLE is involved in adipose chronic inflammation—postulated to occur through binding of MINCLE to an as yet unknown ligand produced or released by adipose cells[332, 339].

Mincle has been indirectly associated with the genetic risk of myocardial infarction. Several genome-wide association studies have identified common genetic variants at the 9p21.3 locus that strongly affect the risk of developing coronary artery disease [123]. The mechanism by which these variants mediate CAD risk remains unclear. A recent study analyzed whole genome expression levels in monocytes from patients with myocardial infarction who were homozygous for either the risk or protective GWAS SNPs at the 9p21.3 locus [360]. Mincle expression was significantly higher in patients with the risk genotype [360].

We studied the expression pattern of Mincle in human atherosclerosis, the ability of oxLDL to act as a ligand and further defined the binding characteristics of MINCLE using reporter assays.

7.2 Results

7.2.1 Immunohistochemistry of Mincle

Immunohistochemistry was performed by Lucy Martin in the laboratory of Dr E. Soilleux (for further details see her Masters thesis). We began by surveying Mincle expression in human tissues by performing IHC in vascular tissues including healthy vessels, atherosclerotic arteries and in arteritic lesions. In normal arteries Mincle was expressed in scantily present macrophages and dendritic cells within the intima and adventitia (Figure 7-1A). Aortic endothelial cells also expressed Mincle. In atheromatous plaque lesions there was an abundance of macrophage-derived foam cells expressing Mincle (Figure 7-1B) and in giant cell arteritis lesions Mincle was expressed within giant cells as well as in macrophages and dendritic cells (Figure 7-1C).

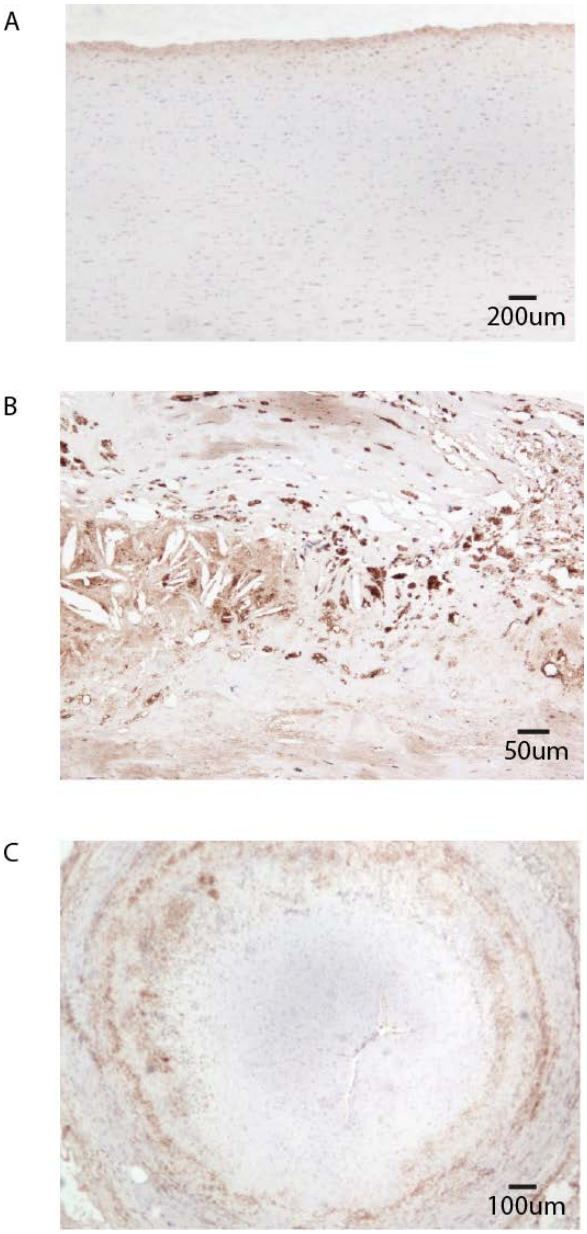


Figure 7-1 Immunohistochemistry staining for Mincle in human arteries.

(A) Normal human aorta showing endothelial positivity for Mincle but lack of Mincle expressing cells in the media. (B) In atheromatous plaque Mincle is strongly expressed in foam cells adjacent to an area of cholesterol clefts. (C) Transverse section of temporal artery affected by giant cell arteritis demonstrates Mincle positive macrophage and dendritic cells in the sub-intimal lymphoid aggregates, including in giant cells.

7.2.2 Mincle expression in macrophages is upregulated by oxLDL

Analysis of our microarray data of the effects of oxLDL on gene expression in macrophages showed that oxLDL-induced foam cell formation was associated with a 26% upregulation of Mincle (adj P = 0.019). By contrast other differentially expressed C-type lectins were downregulated by oxLDL (CLEC1A, CLEC4A, OLR1, CD93). We performed RTqPCR to confirm the microarray findings using cDNA from two additional individuals whose cDNA was not used for the microarrays. Mincle expression was increased 5.3 and 3.4 fold respectively. Additionally, Mincle was upregulated in the THP1 monocytic cell line upon induction with PMA of an adherent macrophage phenotype (18.2 fold, P=0.003) and was further upregulated by subsequent addition of oxLDL (2.2 fold, P=0.012).

7.2.3 Genomic landscape at the Mincle locus

We examined the effect of oxLDL on the open chromatin and enhancer landscape at the Mincle locus in human macrophages to identify putative regulatory DNA sites activated or suppressed by oxLDL (Figure 7-2). Additionally, since C/EBP-beta is required for Mincle upregulation in response to inflammatory stimuli, we looked for C/EBP-beta binding sites at the Mincle promoter or nearby enhancers[361]. In both macrophages and foam cells Mincle appeared to be located in a region of the genome where the FAIRE signal had a relatively low signal to noise ratio and no clear peak was evident near the transcription start site suggesting a lack of open chromatin at the

promoter. However, in the H3k27ac ChIP-seq dataset there was a clear peak in the signal around the transcription start site of Mincle, suggesting that it is an active promoter. Weak C/EBP-beta binding sites were identified at the promoter and in an intron of Mincle. None of the three datasets showed a significant change in signal at the TSS of Mincle in response to oxLDL. The nearest enhancer with high-level C/EBP-beta binding was detected at an intron of CLEC6A, about 70kb downstream from the Mincle TSS. This C/EBP-beta binding site was within a robust open chromatin and enhancer site. C/EBP-beta binding was significantly increased by oxLDL exposure although CLEC6A itself showed no change in expression upon oxLDL stimulation. It is possible that C/EBP-beta binding could regulate Mincle expression through this enhancer in response to oxLDL. C/EBP-beta is known to be critical for upregulation of MINCLE in response to mycobacterial ligands since C/EBP-beta knockout mouse macrophages failed to upregulate MINCLE in response to mycobacterial ligands[361]. It is possible that oxLDL also upregulates MINCLE expression through C/EBP-beta activation.

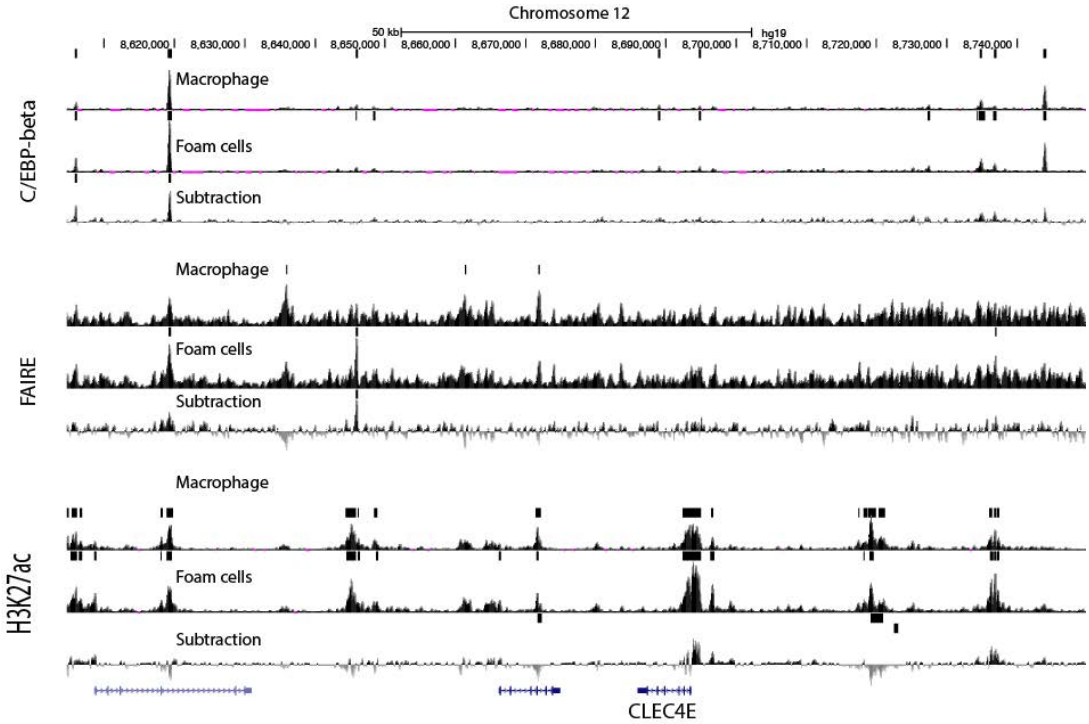


Figure 7-2 Epigenomic lanscape at the Mincle gene locus.

Signal tracks are displayed for FAIRE-seq signal, H3k27ac ChIP-seq signal and C/EBP-beta ChiP-seq signal in macrophages and foam cells. For each dataset a subtraction signal track was produced by subtracting the normalized macrophage signal from the foam cell signal. Black bars above each track identify significant peaks in signal. Genes are shown at the bottom of the figure where CLEC4E (Mincle) is shown running in the anti-sense direction. A striking oxLDL regulated C/EBP-beta binding site is seen to the left of the figure in an intron of CLEC6A. The promoter region of CLEC4E has H3K27ac enhancer marks and a weak C/EBP-beta binding site.

7.2.4 NFAT reporter assay response to oxLDL

To study whether lipid compounds were ligands for Mincle we used a cellular reporter assay that detects Mincle activation. A stably transfected clone of Jurkat cells with an NFAT-luciferase reporter element was stably co-transfected with human Mincle and FcRγ. As a control cells were transfected with a similar vector that lacked Mincle. In the system Mincle activation can be measured by assaying luciferase activity in lyzed cells using a luminometer and the transfection efficiency can be controlled for using the GFP reporter which is bi-cistronic with Mincle. A major advantage of the system includes leveraging the wide dynamic signal range of luciferase assays and the output quantitatively reflects the magnitude of ligand binding on a per cell basis (Figure 7-3).

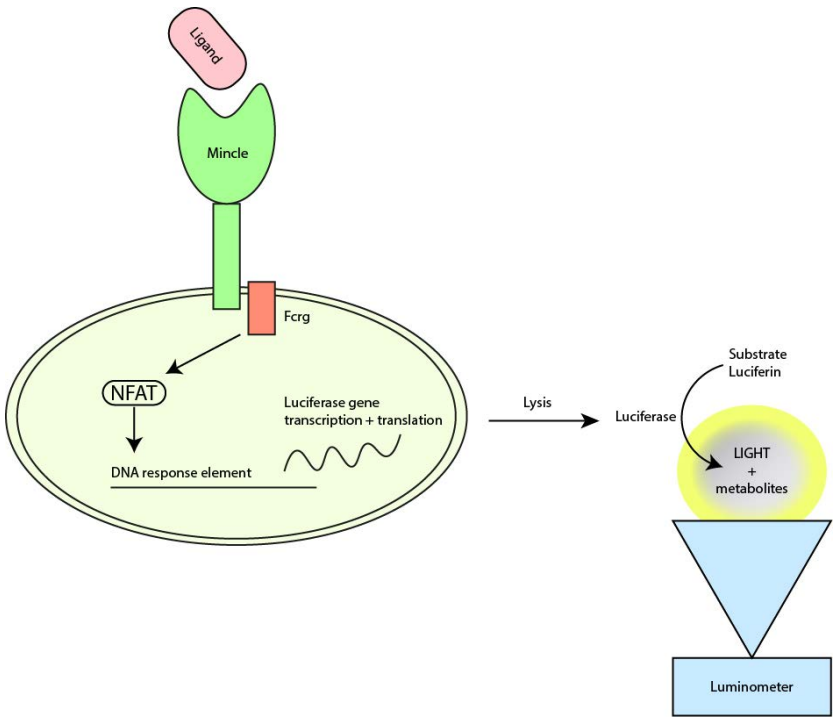


Figure 7-3 Schematic of the Mincle luciferase reporter assay.

Mincle and FcR γ were co-transfected into a Jurkat cell clone that has an NFAT DNA response element upstream of a luciferase gene. Binding of a ligand to Mincle

activates a signaling cascade which results in an increase in NFAT binding to its response element. This results in increased transcription of the luciferase gene and increased translation of the luciferase enzyme. After a defined time period the cells are lyzed and luciferin substrate is added. A luminometer is used to measure the light output with the signal being proportional to the binding of Mincle to a ligand.

To study whether atherogenic lipids themselves contain Mincle ligands, we stimulated reporter cells with LDL (50mcg/ml), oxLDL (50mc/ml) or an equivalent volume of control buffer or PBS. We found that these lipids did not lead to a significant increase in luciferase activity in Mincle expressing cells suggesting that LDL and oxLDL do not contain Mincle ligands (Figure 7-4). By contrast the anti-Mincle antibody produced a substantial increase in luciferase activity (3.2 fold, P=0.007, n=3).

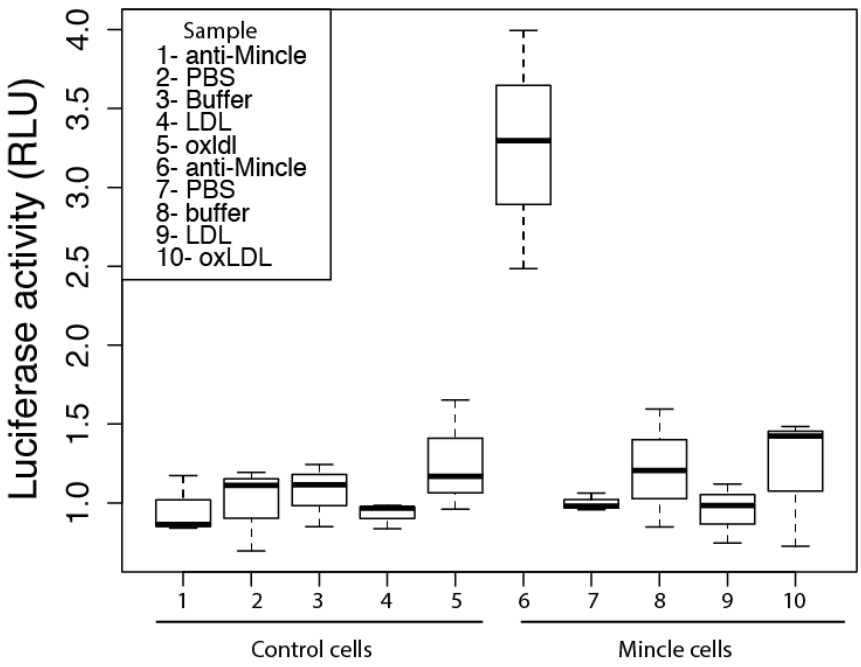


Figure 7-4 Human LDL and oxLDL to not cause Mincle activation.

Luciferase reporter assay showing Mincle transfected and control transfected NFAT reporter cells after stimulation with either anti-Mincle antibody, PBS, control buffer, LDL or oxLDL. Only the anti-Mincle antibody stimulates luciferase activity and this effect is specific to Mincle bearing reporter cells.

7.2.5 Human and mouse Mincle are receptors for several MTB strains

Although low-density lipoprotein and oxLDL did not trigger Mincle activation it remains conceivable that an endogenous ligand may exist among the milieu of lipid species present within atherosclerotic plaque, and in particular lipids that could be released from dying cells. Rather than screening a panel of lipid species known to be associated with atherosclerosis we continued our investigation by starting with the known ligands TDM and TDB, and then testing a series of compounds which differ iteratively and systematically in their head and tail regions. This approach allowed us to define more precisely the binding requirements for Mincle and could provide a

rationale basis for testing similar endogenous compounds for Mincle binding. It is possible that a greater knowledge of what MINCLE 'senses' will lead to the discovery of additional endogenous MINCLE ligands that promoter atherosclerosis through MINCLE based inflammatory signaling. In addition, there is some evidence that mycobacterial infection is itself pro-atherogenic since animal and human studies have found high levels of mycobacterial proteins in atherosclerotic plaque and vaccination of animals with this protein causes vascular wall changes[362]. It is not known whether specific MINCLE ligands e.g. TDM are themselves involved in atherosclerosis.

Heat killed MTB stimulated Mincle reporter cells in a dose-dependent manner affirming Mincle's critical role in recognizing ligands on Mycobacterium tuberculosis (Figure 7-5). At the upper dose of 10 million MTB cells incubated with 0.5 million reporter cells the stimulation was equivalent to that of our activating monoclonal anti-Mincle antibody which is specific for human Mincle (n=3, Fold increase over PBS = 3.3 P=0.0013, compared with fold change of 0.6 in the control cells, P=0.1479). We also found that MTB was capable of signaling through the mouse Mincle (whereas our monoclonal anti-Mincle antibody does not).

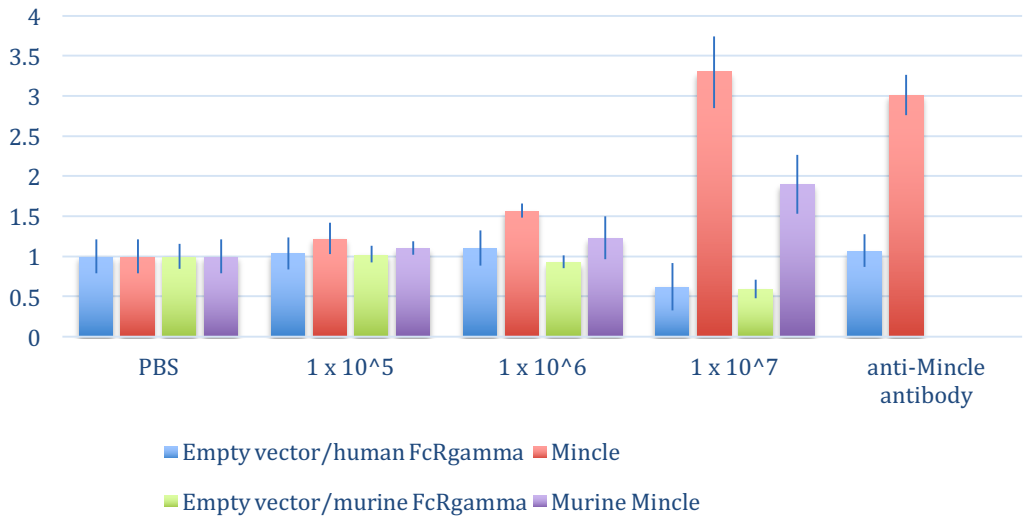


Figure 7-5 Human and mouse Mincle bind MTB.

Luciferase reporter assays showing dose dependent effect on Mincle activation by MTB with highest dose shown equivalent to activation by an activating anti-Mincle antibody. An equivalent reporter assay using mouse Mincle also shows a response to MTB (n=3, vertical bars indicate standard deviation).

We extended our observations of the effects of MTB on Mincle signaling by testing whether less pathogenic strains of Mycobacteria were as capable as MTB in stimulating Mincle signaling (Figure 7-6). We found that M. bovis, BCG and M. Smegmatis at a dose of $1x10^7$ heat killed organisms produced a similar 4-fold induction of reporter activity in Mincle bearing cells (10^7 Bovis $P<0.0001$, 10^7 smegmatis $P<0.0001$). The RD1 strain lacks a 9.5kb virulence factor region and elicited a similar response to smegmatis and bovis at the 10^6 dose. It is unclear why a higher dose did not elicit a greater response. Heat killed E. coli at a dose of 10^7 produced a small fold induction in luciferase activity in both control cells and Mincle bearing cells (control cells fold 1.24, P=0.066, Mincle cells fold 1.6, P=0.0008). This effect reached statistical significance in Mincle bearing cells. It is unclear whether E.

coli indeed possesses a ligand capable of activating Mincle or whether the large endotoxin dose could have triggered non-specific activation.

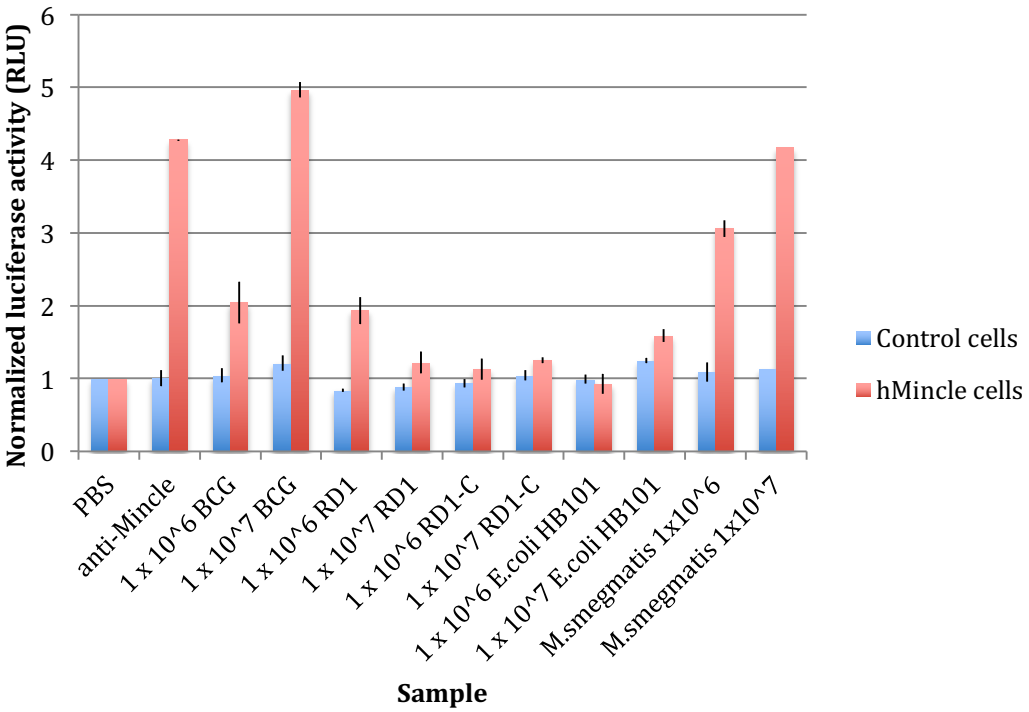


Figure 7-6 Several different strains of MTB activate Mincle.

Luciferase reporter assay shows that the upper dose of BCG and smegmatis produces a similar response to the antibody whereas RD1 produce a lesser response and E. coli does not produce a significant response. The upper dose of RD1 does not produce a significant response suggesting the possibly of cellular death caused by the preparation tested interfering with the assay (n=3, vertical error bars indicate standard deviation).

7.2.6 Signaling through MINCLE is triggered by an extensive repertoire of trehalose dimycolate related compounds

Given that the Mincle ligand within MTB is known to be TDM we systematically synthesized an iterative range of compounds deviating from TDM in chain length, number of chains, intra-chain bonds and alteration of the head trehalose moiety (compounds were the generous gift of Del Besra).

We began by further validating the assay system by demonstrating that TDM caused significant Mincle activation in the luciferase assay (fold increase at 1000pmol dose 2.7, P=0.0002)(Figure 7-7). The threshold for an effect was greater than 250pmol TDM per 0.5M reporter cells since no activation was seen with this dose. Escalating the dose to 10,000 pmol did not increase the response further possibly due to subtle toxicity to the cells. Trehalose 6,6'-dibehenate (TDB) is a synthetic analogue of TDM that is known to be a potent adjuvant and Mincle agonist. TDB possesses two acyl chains but these are not branched like mycolyl chains and are shorter (22 carbons in the acyl chains). TDB produced significant Mincle activation in a dose dependent manner (fold increase at 1000pmol dose 2.9, P= 0.0094)(Figure 7-7).

Trehalose-6-monomycolate (TMM) is a mycobacterial glycolipid that differs by only having one mycolyl chain. Interestingly it showed a strikingly similar efficacy to TDM, demonstrating that two mycolyl chains are not required for Mincle binding and activation (Figure 7-8). To determine the optimal chain length for Mincle stimulation we synthesized compounds with a trehalose head and either one or two acyl chains with lengths varying from 8 to 24 carbon atoms (Figure 7-11). Trehalose with two 16 carbon chains had the greatest activity. Two chain compounds with 12 or 8 showed no

significant increase in Mincle stimulation compared to the control cells, which lack Mincle (Figure 7-8). Two of the synthetic compounds with a single acyl chain (16 and 18 carbons) caused rapid cell death and therefore their activities could not be tested (Figure 7-8). It is unclear whether the cell death was caused by a leftover contaminant or whether the compounds are intrinsically lethal. Single acyl chain compounds with 12 and 20 carbons showed significant activity whereas the compound with 8 carbons showed no Mincle stimulation.

Glucose mono-mycolate GMM is a mycobacterial glycolipid produced when mycobacteria couple endogenous mycolic acid to exogenous glucose within infected tissues[363]. It therefore differs in that the carbohydrate moiety is composed of a single glucose unlike trehalose, which has two glucoses. We synthesized GMM compounds with acyl chain lengths varying from 8 to 24. A strong effect was seen in our reporter assay showing that trehalose itself is not necessary to activate Mincle as only one parent glucose molecule is present within GMM (Figure 7-9). We found that each of the GMM compounds was able to activate Mincle demonstrating that with a single carbohydrate moiety a short acyl chain of just 8 carbons is enough to produce a potent response.

We next tested whether kinks in the acyl chain—made by adding additional double carbon bonds—would alter Mincle binding and activation. We synthesized compounds with a trehalose carbohydrate moiety and either one or two carbon chains containing either oleic or linoleic carbon to carbon bonds. With a single oleic chain, we found that there was no Mincle stimulation. In the case of linoleic bonds, which

maintain a more linear structure to the acyl chain, both the single and double chain compounds stimulated Mincle.

Monomycolyl glycerol (MMG) is an endogenously occurring mycobacterial lipid which has been found to have greater stimulatory effects on dendritic cells than TDM [364]. A synthetic analogue (32 carbon chain length) was found to exert similar effects and use as an adjuvant in mice immunized with a TB subunit vaccine produced robust immune responses. MMG lacks a hydrophilic carbohydrate moiety since it is composed of just a glycerol head moiety and an acyl chain. MMG has recently been shown to bind to human Mincle but not mouse Mincle[345]. However, these studies utilized a reporter assay, which uses the proportion of GFP positive cells as an output measure and is therefore susceptible to saturation in the number of GFP positive cells. Our assay, which quantifies the average signal for each cell, demonstrated that MMG was capable of activating Mincle to a level that was similar to our activating antibody and to TDM(Figure 7-10). Finally, we tested a mycobacterial sulpholipid analogue that possesses a sulphate group attached to one of the sugar moieties and the lipid chains derive from neighbouring carbon atoms 3 and 4. This compound had no effect on Mincle stimulation indicating that the position of the acyl chain most likely caused steric hindrance in binding to Mincle (Figure 7-10).

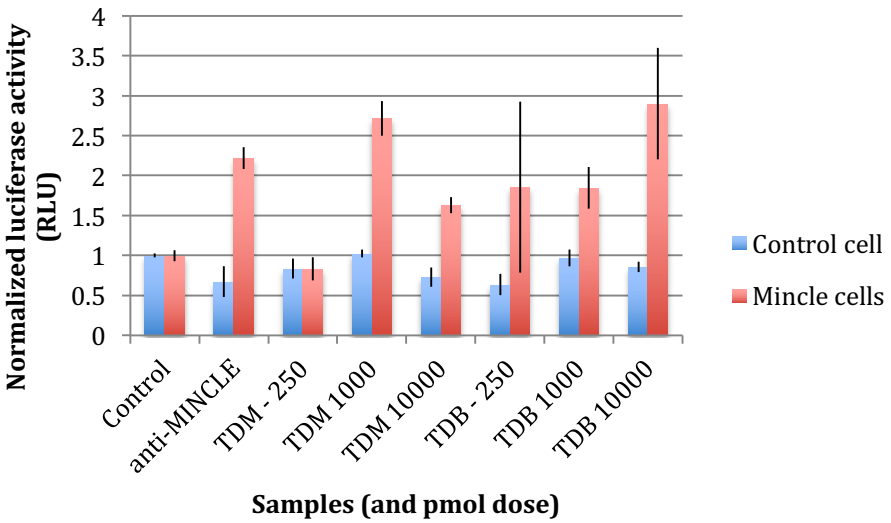


Figure 7-7 TDM and its synthetic analogue, TDB, activate Mincle.

Luciferase assay showing that 1000pmol doses of TDM and TDB are similar in potency to the anti-Mincle antibody (n=3 biological reps, error bars indicate standard deviation).

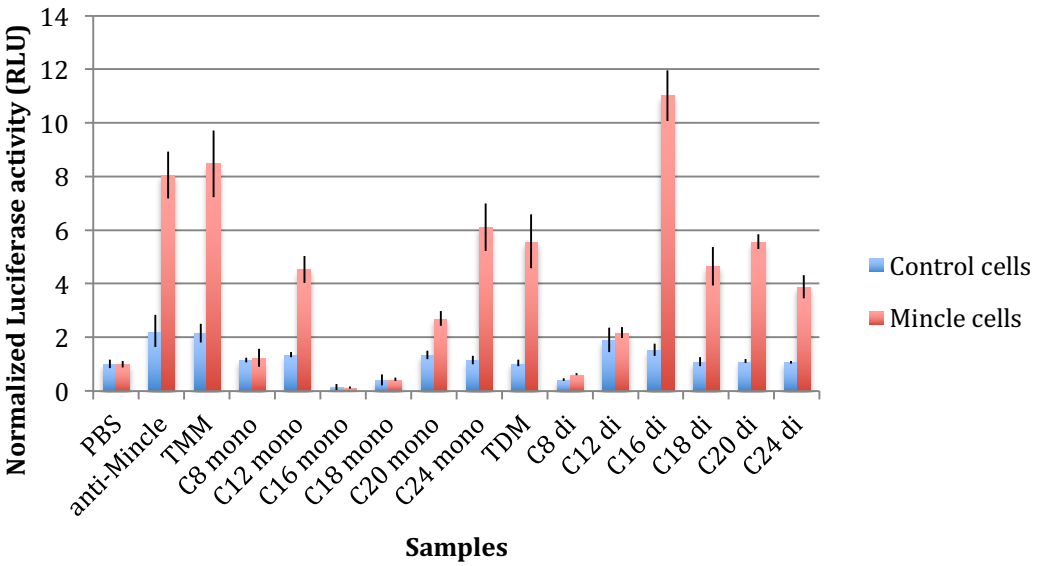


Figure 7-8 TMM is a Mincle ligand.

Luciferase reporter assay shows that compounds with only one acyl chain can bind to Mincle although compounds with only 8 carbons in the acyl chain failed to bind Mincle. Certain compounds induced cell death including C16 mono, C18 mono and trehalose di-C8 (n=3 biological reps, error bars show standard deviation).

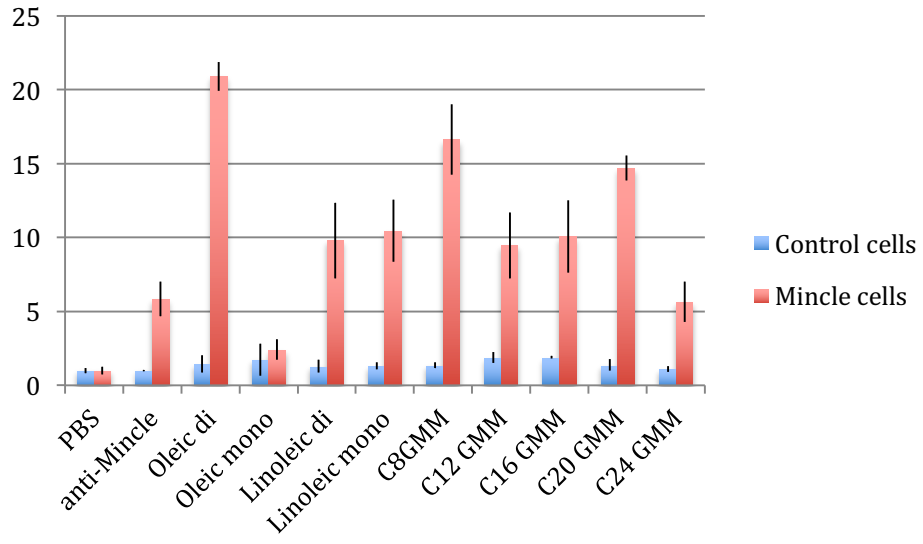


Figure 7-9 Effect of additional carbon-carbon and sugar head moiety on Mincle binding.

Introduction of additional carbon-to-carbon bonds within the acyl chains of trehalose compounds does not impair Mincle binding except for trehalose with one oleic acyl chain (Oleic mono). Oleic residues introduce a kink into the acyl chain which could cause steric hindrance to Mincle binding. A double carbohydrate moiety is not required for Mincle binding since a glucose carbohydrate head group attached to one acyl chain is capable of binding to Mincle (C8-C25 GMM) (n=3 biological replicates, error bars indicate standard deviation).

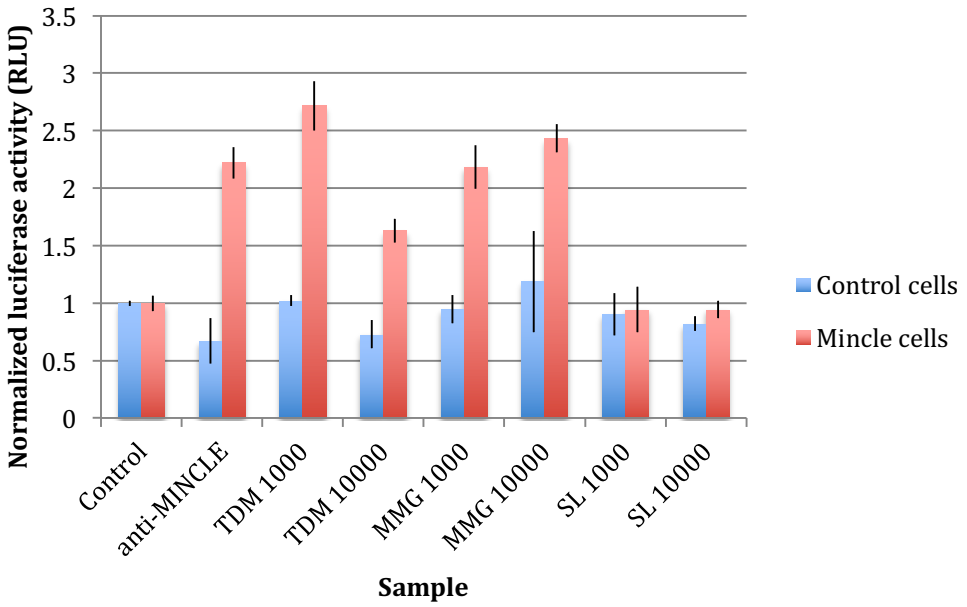


Figure 7-10 Monomycolyl glycerol (MMG) is a ligand for Mincle but a Mycobacterial sulpholipid is not.

MMG contains a simple glycerol head moiety instead of a carbohydrate. Sulpholipids contain a trehalose like head moiety but two acyl chain originate from the same glucose moiety and likely cause steric hindrance to Mincle binding.

Taken together our results demonstrate that Mincle binding is not dependent on the presence of two lipid chains or a paired parent sugar molecule. Mincle activation was possible with a chain length of as little as 8 carbon atoms. Sulpholipid compounds bearing the lipid chains on contiguous carbon atoms were not capable of Mincle activation. MMG was confirmed as a ligand for Mincle.

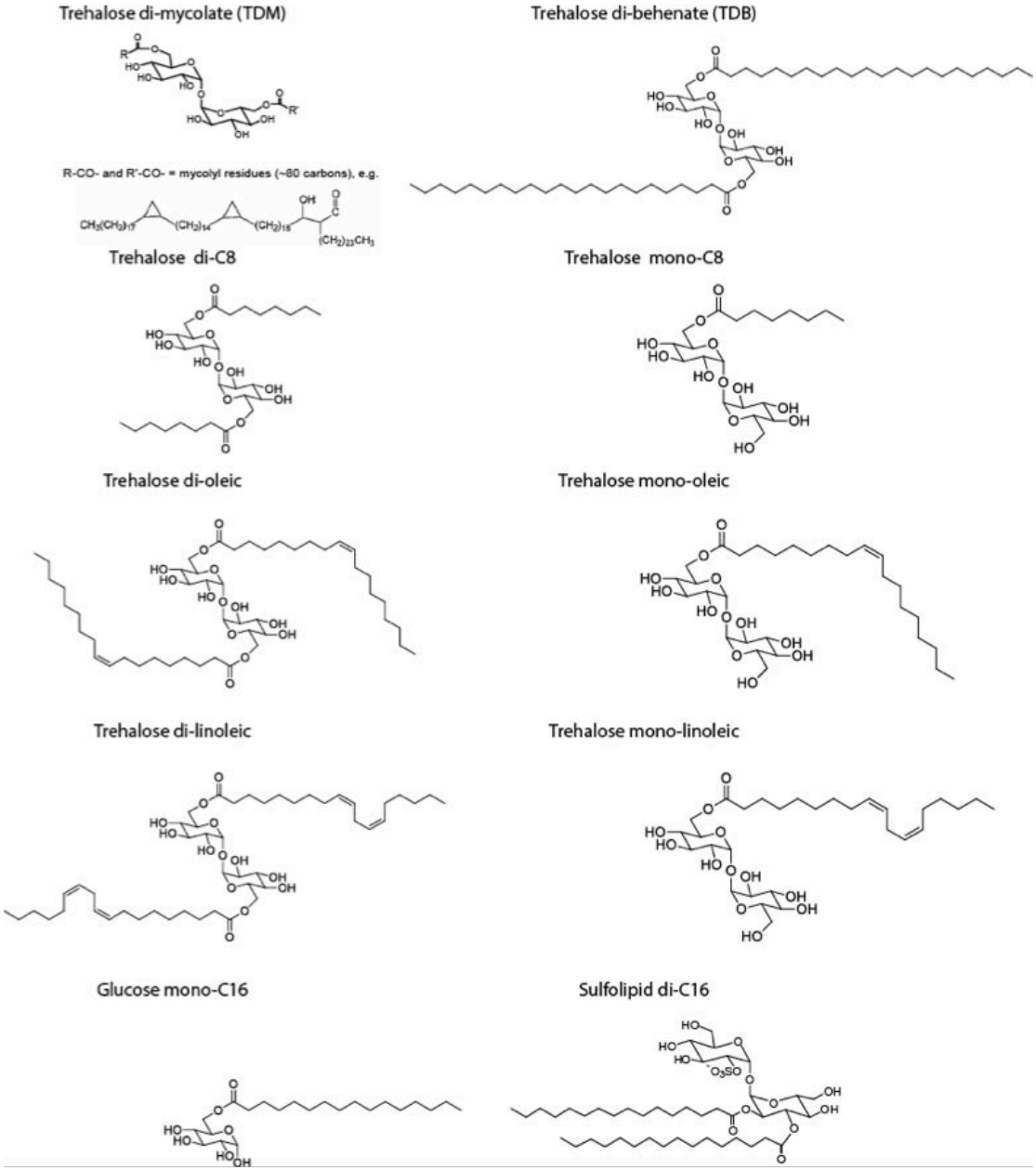


Figure 7-11 Structure of Trehalose-related compounds used to stimulate Mincle

Table 17 Chemical formula and calculated molecular weights of compounds tested in Mincle luciferase assays.

	Formula	Molecular weight (g/mmol)
TDM	$C_{172}H_{328}O_{15}$	2,636.47
TDB	$C_{56}H_{106}O_{13}$	987.43
Trehalose di-C8	$C_{28}H_{50}O_{13}$	594.6888
Trehalose di-C12	$C_{36}H_{66}O_{13}$	706.9014
Trehalose di-C16	$C_{44}H_{82}O_{13}$	819.1141
Trehalose di-C18	$C_{48}H_{90}O_{13}$	875.2204
Trehalose di-C20	$C_{52}H_{98}O_{13}$	931.3267
Trehalose di-C24	$C_{60}H_{114}O_{13}$	1043.5394
Trehalose mono-C8	$C_{20}H_{36}O_{12}$	468.4926
Trehalose mono-C12	$C_{24}H_{44}O_{12}$	524.5990
Trehalose mono-C16	$C_{28}H_{52}O_{12}$	580.7053
Trehalose mono-C18	$C_{30}H_{56}O_{12}$	608.7584
Trehalose mono-C20	$C_{32}H_{60}O_{12}$	636.8116
Trehalose mono-C24	$C_{36}H_{68}O_{12}$	692.9179
Sulpholipid	$C_{27}H_{78}O_{16}$	658.9
Trehalose di-oleic	$C_{48}H_{86}O_{13}$	871.1886
Trehalose mono-oleic	$C_{30}H_{54}O_{12}$	606.7426
Trehalose di-linoleic	$C_{48}H_{82}O_{13}$	867.1569
Trehalose mono-linoleic	$C_{30}H_{52}O_{12}$	604.7267

MMG	$C_{35}H_{70}O_5$	593.5121
Glucose C8	$C_{14}H_{26}O_7$	306.3520
Glucose C12	$C_{18}H_{34}O_7$	362.4584
Glucose C16	$C_{22}H_{42}O_7$	418.5647
Glucose C20	$C_{26}H_{50}O_7$	474.6710
Glucose C24	$C_{30}H_{58}O_7$	530.7773

7.3 Discussion

Innate immune cells including macrophages express a diverse range of pattern recognition receptors that allow a naive immune system to immediately respond to certain pathogens by recognizing conserved pathogen associated molecular patterns. Mincle is one of only two receptors for trehalose di-mycolate, a major cell wall component and PAMP that is unique to mycobacteria[365, 366]. As well as being involved in the host response to mycobacterial infection Mincle also has a role in vaccine development since TDM and its synthetic analogue TDB are potent adjuvants[367]. Damage associated molecular patterns are endogenous molecules that can bind innate immune receptors and initiate or perpetuate immune responses and inflammation. One such molecule is SAP130, a protein that forms part of the histone deacetylase corepressor complex. SAP130, an endogenous Mincle ligand, is released from dying cells and causes an inflammatory response that is abrogated by using an antibody to block Mincle-SAP130 binding[342]. We sought to establish whether Mincle might have a role in atherosclerosis and further define the ligand attributes necessary for Mincle binding.

We found that whereas Mincle expressing cells are scantily present in healthy human arteries, both atherosclerotic arteries and arteries affected by giant cell arteritis have an abundance of Mincle expressing macrophages. The abundance of Mincle in atherosclerotic lesions raises a number of questions as to whether it might be directly involved in disease development or progression. These questions include whether

Mincle binds oxidized lipoproteins, and whether Mincle induces inflammatory responses through binding to SAP130 released from dying cells within atherosclerotic plaques. Our mRNA data suggest that oxLDL does indeed upregulate Mincle expression in macrophages, albeit at a relatively modest level. Another study has found that Mincle was also upregulated by the fatty acid palmitate and in monocytes of obese individuals[354]. Palmitate-induced Mincle expression was in part dependent on TLR4 signaling[354].

We tested whether LDL or oxLDL contained a ligand for Mincle using a Mincle-luciferase reporter assay. Overall neither compound produced a significant response in the assay. Subsequent to our study it was reported that cholesterol itself is an endogenous ligand for Mincle although the precise nature of the interaction remains unknown[368]. If cholesterol is a ligand for Mincle then clearly Mincle would be a key activating receptor within atherosclerotic plaques. Mincle may also senses SAP130 released from dying cells within atherosclerotic plaques.

7.3.1 Regulation of Mincle by C/EBP beta in human macrophages and foam cells

The transcription factor C/EBP-beta has dual role in Mincle associated inflammatory responses. Firstly, TDB increased expression of C/EBP-beta in a Mincle dependent manner in mouse macrophages[361]. Secondly, C/EBP-beta was required for high level Mincle expression and induction of inflammatory responses in reaction to Mincle ligands [361]. In human macrophages we found that Mincle and C/EBP-beta were simultaneously upregulated by oxLDL and there was a low level C/EBP-beta ChIP-seq binding site near the transcription start site. Additionally, there was a robust enhancer site within 75kb of the Mincle gene with oxLDL increased C/EBP-beta binding.

A higher level of Mincle expression on monocytes of patients with myocardial infarction is associated with the genetic risk haplotype at the 9p21.3 locus[360]. Our finding of high levels of Mincle expressing macrophages and foam cells in atherosclerotic plaques, coupled with upregulation by atherogenic lipid suggest that the genetic risk could operate within lipid laden plaques by further increasing Mincle expression leading to enhanced inflammatory responses.

To refine our knowledge of the necessary structural requirements of ligands for Mincle binding, we tested an extensive range of compounds that differed iteratively and systematically from TDM. TDM comprises a trehalose carbohydrate moiety made

up from two glucose residues with a lipid chain attached to each glucose residue. We tested compounds with chain lengths as short as 8 carbon atoms and as long as 24 carbon atoms. The carbohydrate head ranged from the full trehalose moiety to a simple glycerol head. In almost all cases, these molecules were able to bind Mincle since they produced an effect comparable to both a stimulating antibody and TDM, in our Mincle luciferase assays. These data demonstrate the relative tolerance of the Mincle binding site for various permutations of TDM-related ligands and these compounds could all be tested as vaccine adjuvants. A limitation of our data is that three compounds caused rapid cellular death and we could not therefore derive meaningful results from these. A non-cellular binding study using recombinant Mincle protein could be used to test these compounds. Alternatively, refinement of the production process of these compounds may remove unknown adulterants that could be toxic to cells. Another limitation of a cell based experiment is that the overall signal to noise ratio of the assay varied over time likely due to changes in the population of Mincle transfected cells which did not have a positive selection marker. In the future, selection of individual clones and using plasmids with selection markers could allow greater consistency between experiments. Nevertheless, we performed adequate positive and negative controls with each experiment to determine the dynamic range of the assay and allow comparison of the effect of compounds between individual assays. A few of the ligands we have tested have also been tested in other published studies and these results are in keeping with our results. Trehalose compounds with one acyl chain with either 22 or 26 carbon atoms triggered Mincle dependent IL-6 release and NO production from mouse bone marrow macrophages [346].

MMG is a recently discovered mycobacterial lipid, which together with a synthetic analogue, has been shown to activate human dendritic cells [364].

The synthetic MMG analogue posses the simplest structure of the compounds we have tested, being composed of just an acyl chain and glycerol head moiety. Our finding of it being a ligand for human Mincle are similar to a recent study in which MMG was shown to bind human Mincle but not mouse Mincle [345]. Using our reporter assay we have also found that mouse Mincle does not bind MMG (personal communication - Anita Mistry). Further studies comparing the structure of human and mouse Mincle may reveal why MMG uniquely only binds human Mincle.

7.3.2 Conclusions

Numerous lines of evidence inculpate Mincle in the pathophysiology of atherosclerosis, including oxLDL-regulated expression, high level expression in atherosclerotic plaque, and a trans-eQTL for common genetic variants associated with CAD. Our experiments have refined our knowledge of the ligand binding abilities of Mincle which may aid the rational development of Mincle stimulating vaccine adjuvants as well as the search for further endogenous ligands.

8 General Discussion

In this study I investigated a fundamental aspect of atherosclerosis—foam cell formation in response to atherogenic lipid. We focused on how macrophages are re-programmed into foam cells at the epigenetic level. Preventing foam cell formation is already a key strategy in the treatment of atherosclerosis—namely the use of lipid lowering agents such as statins. Since the development of statins there has, however, been limited progress in developing new classes of athero-protective drugs. If macrophages or foam cells could be reprogrammed to respond less harmfully to atherogenic lipid, then atherosclerosis may be ameliorated. Simple inspection of macrophages down a microscope shows the profound effect of atherogenic lipid on their conformation. We know from developmental studies that marked cellular changes are accompanied by epigenetic changes—the alterations in proteins in proximity to DNA that both govern and reflect changes in the utilization of DNA for gene expression[221]. I hypothesized that the effect of oxLDL on macrophages would include epigenetic changes and that these alterations would both reflect the upstream pathways activated by oxLDL stimulation and would reveal the parts of the genome where a response to oxLDL controls altered gene expression. In this study I went beyond simply describing an 'epigenetic signature of disease'. I integrated it with the recently discovered CAD genetic risk variants to identify their role in the process of foam cell formation. This important work was recently published[246] and established for the first time a plausible mechanism by which the heritable CAD risk may in part be mediated through altered PPAP2B expression in response to oxLDL—

PPAP2B as an enzyme is clearly established as an immediate priority for the development of specific activators and inhibitors.

In addition to this work I provide a wealth of data that is freely available for use by other researchers who desire to know the epigenomic landscape and regulation of gene expression during foam cell formation, or who wish to study the role of C/EBP-beta in response to oxLDL.

8.1 Lipid-induced epigenomic changes in macrophages

These results show that epigenetic changes accompany the striking macroscopic changes associated with foam cell formation. Epigenetic changes were closely correlated with expression of the nearest differentially expressed gene.

FAIRE-seq as a tool has certain benefits and limitations in the study of foam cell formation. The most striking benefit is the ease with which cells can be effectively 'frozen' in time by cross-linking with formaldehyde. On the other, optimizing the signal to noise ratio required multiple titrations of crosslinking time and phenol-chloroform extractions, often yielding very small quantities of DNA for sequencing. We achieved a greater signal to noise ratio than ENCODE FAIRE-seq data viewable in the UCSC genome browser, partly by pooling biological triplicate samples. Nevertheless, there remained areas of the genome with relatively poor signal. Another disadvantage of FAIRE-seq is that the signal is not sharp enough to perform transcription factor foot-printing. Foot-printing has proved to be useful approach with DNAse-seq data but these experiments require far larger biological samples that are

not readily feasible with primary cells and require a depth of sequencing which is substantially costlier[369]. ChIP-seq for H3K27ac proved to be highly useful since it gave both an excellent signal and high reproducibility, eliminating the need to pool biological replicates for any of the analysis. Numerous lines of evidence have shown that H3K27ac ChIP-seq is an excellent surrogate for enhancer activity[245, 370]. Our enhancer and open chromatin maps provide a highly useful platform for researchers looking for the regulatory elements governing individual genes of interest.

To connect oxLDL with changes in open chromatin and enhancer activity we looked for what was different about the DNA sequences in the genome where oxLDL affects the open chromatin structure. We expected to find that the main enrichment would be for transcription factors with known roles such as PPARG and LXR-alpha[261]. The pattern of enrichment for transcription factor binding motifs did however show that AP1 was by some margin the most enriched factor, closely followed by C/EBP-beta. It may be that a refined approached to dividing the 1000s of dynamic chromatin sites would delineate patterns of enrichment for PPARG and LXR-alpha. Given that AP1 and C/EBP-beta are both pioneer factors whose binding can activate processes that lead to further 'chromatin opening' it is perhaps not surprising that they are most closely associated with dynamic sites[219, 371]. In both cases, my analysis suggests these factors warrant further study as transcription factors mediating the response to oxLDL. Based on these results I went on to identify C/EBP-beta binding sites using ChIP-seq and confirmed that oxLDL does indeed regulate C/EBP-beta binding. Inhibition of C/EBP-beta activity is a potential strategy to ameliorate foam cell formation and knockdown of C/EBP-beta may reduce inflammation since its binding sites were enriched for immune genes.

In my hands, foam cells existed for many days in culture after treatment with oxLDL. It would be interesting to perform a similar epigenomic analysis at a later time-point when the foam cells have been presented with the opportunity to process and expel the excess lipid. This analysis would determine to what extent the epigenetic changes we have seen are reversible if the pathogenic stimulus is removed and what capacity foam cells have for recovery to a normal state.

I have used a model of foam cell formation. I sought to use the highest grade of model possible by only using primary human cells and by isolating our own fresh LDL from human blood. A major concern in producing the LDL for these experiments is endotoxin contamination. The most commonly used assay for endotoxin detection in academic laboratories is the limulus test. Plasma contains an inhibitor of the limulus reaction that can provide falsely reassuring results. A close study of the literature revealed that an extra heating step was required to degrade this inhibitor without affecting the endotoxin concentrations. To our knowledge, no previous published work with human LDL and the limulus test has described using this method.

8.2 Integrating the macrophage-foam cell epigenome with CAD heritable risk

Genome-wide association studies have opened up the exploration of genetic risk for common, complex diseases by identifying relatively precisely areas of the genome harboring risk-associated polymorphisms and candidate genes. Though thousands of

risk SNPs for hundreds of human traits and diseases have been published, their biological effects are for the most part completely unknown[372]. Our study helps to unlock this bottleneck for CAD by attempting to identify which of the many variants might be functional in macrophages. Overall, of the 45 loci we found that 22 loci have SNPs in regulatory DNA identified by our open chromatin and enhancer maps. Given the possibility of chance associations we used a robust method to determine whether there was enrichment of CAD variants in these regulatory DNA sites. Finding enrichment is a major pointer to a particular cell type or tissue being the site of action for a set of variants. Surprisingly we did not find statistically significant enrichment for CAD SNPs compared to a background of other GWAS SNPs in either macrophages or foam cells. At face value this suggests macrophages or foam cells do not convey the majority of the GWAS CAD risk. For the first time we looked for enrichment of SNPs in the subset of regulatory elements that are involved in the pathogenic process of foam cell formation. Notably we did find statistically significant enrichment in this subset of sites. We suggest that there are SNPs that are in effect brought into play during the pathogenic transformation of healthy cells into diseased cells. Future studies can test this hypothesis in other cell types, for example, healthy endothelial cells and those undergoing shear stress. It is also possible that assaying a broader range of chromatin marks, using additional techniques for open chromatin mapping, and performing the studies in more individuals will identify additional regulatory elements harboring risk SNPs. The PHACTR1 analysis is a potential case in point. I find that PHACTR1 is expressed in macrophages, regulated by oxLDL and inflammatory stimuli, has an eQTL in macrophages; all pointers to a role in the genetic CAD risk at this locus. I did not find however, that any of the disease associated SNPs mapped to an open chromatin site or enhancer site in

macrophages and foam cells. The PPAP2B locus proved highly interesting since PPAP2B was a hit across all our datasets—oxLDL regulated expression, enhancer activity, open chromatin, C/EBP-beta binding and SNPs in regulatory DNA. From this data and additional functional testing of the rs72664324 SNP I propose a model whereby rs7264324 may in part account for the CAD risk at this locus (Figure 8-1).

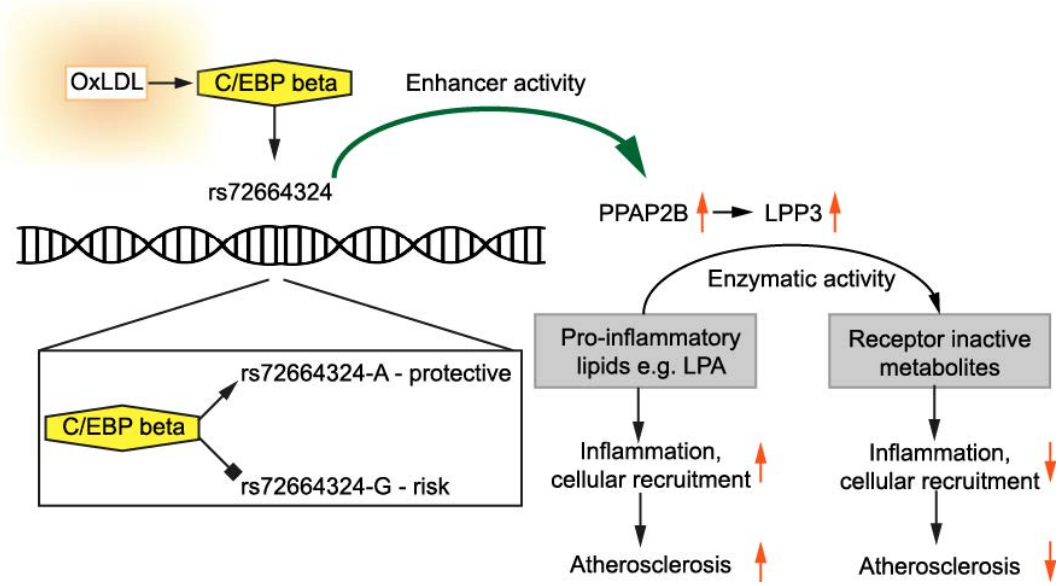


Figure 8-1 Model showing the interplay between oxLDL, C/EBP-beta and rs72664324 on expression of PPAP2B.

OxLDL-induced C/EBP beta binding is greater to the A allele and leads to increased induction of PPAP2B in response to oxLDL compared to the G allele. PPAP2B encodes LPP3 whose enzymatic activity reduces pro-inflammatory signaling mediated by LPA.

The IL6R CAD risk locus shares many similarities with the PPAP2B locus. IL6R is upregulated by oxLDL and some risk SNPs are in an oxLDL-induced, intronic enhancer that is also part of a super-enhancer. IL6R encodes the receptor for the pro-inflammatory cytokine interleukin-6 and circulating levels of a soluble form of IL6R

have been associated with coronary artery disease[250, 251]. Modulators of Il6R function are in development and future studies may also investigate the role of the subset of SNPs implicated by our study[373].

8.3 Mincle

Relatively few C-type lectins are known to have a direct role in atherosclerosis. The LOX1 receptor recognizes and internalizes oxLDL and mutations in its coding sequence are associated with CAD [334]. Mincle has been indirectly linked with atherosclerosis because macrophages carrying the risk haplotype at the 9p21.3 locus have higher Mincle expression—a trans eQTL—and macrophage Mincle expression is induced by an interaction with adipocytes[354, 360]. Our study supports a role for Mincle in atherosclerosis by demonstrating increased expression in atherosclerotic plaque and upregulation upon foam cell formation. Mincle is a major receptor for the MTB associated glycolipid, TDM but in our study oxLDL did not appear to contain endogenous ligands for Mincle. Emerging data suggest that cholesterol itself may be a ligand for Mincle[374]. Our study has helped to define the ligand repertoire of Mincle by finding a range of new synthetic derivatives of TDM that bind to Mincle. These will be helpful in future vaccine development as adjuvants.

8.4 Conclusions

Overall, this study establishes a link between CAD genetic susceptibility, the macrophage response to atherogenic lipid and receptor active lipid signaling. This study demonstrates the utility of chromatin and enhancer mapping in primary human cells before and after a pathogenic environmental stimulus and this approach may have applications in the study of other diseases.

Our high throughput sequencing and microarray datasets are freely available and provide a wealth of primary human cell data for further analysis. Further studies may involve development of myeloid specific mouse knockouts for PPAP2B, PHACTR1, MINCLE and CEBPB to assess their functional role in atherosclerosis. The development of specific inhibitors and activators of PPAP2B would also be a useful tool. Similarly, the interaction of PHACTR1 with actin or PP1 might be specifically targeted and the effects of PHACTR1 disruption on gene expression studied. Our enhancer and open chromatin data provide a platform for exploring the transcriptional pathways that that regulate the oxLDL mediated induction of these genes.

9 Appendix

9.1 Published work

A substantial part of the work in this thesis formed the basis for my first author paper in Plos Genetics. The abstract is shown below. The PDF version of the manuscript is attached as an addendum at the end of this thesis.

PLoS Genet. 2015 Apr 2;11(4):e1005061. doi: 10.1371/journal.pgen.1005061. eCollection 2015.

Lipid-induced epigenomic changes in human macrophages identify a coronary artery disease-associated variant that regulates PPAP2B Expression through Altered C/EBP-beta binding.

Reschen ME[1], Gaulton KJ[1], Lin D[1], Soilleux EJ[2], Morris AJ[3], Smyth SS[3], O'Callaghan CA[1].

Author information

Abstract

Genome-wide association studies (GWAS) have identified over 40 loci that affect risk of coronary artery disease (CAD) and the causal mechanisms at the majority of loci are unknown. Recent studies have suggested that many causal

GWAS variants influence disease through altered transcriptional regulation in disease-relevant cell types. We explored changes in transcriptional regulation during a key pathophysiological event in CAD, the environmental lipid-induced transformation of macrophages to lipid-laden foam cells. We used a combination of open chromatin mapping with formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) and enhancer and transcription factor mapping using chromatin immuno-precipitation (ChIP-seq) in primary human macrophages before and after exposure to atherogenic oxidized low-density lipoprotein (oxLDL), with resultant foam cell formation. OxLDL-induced foam cell formation was associated with changes in a subset of open chromatin and active enhancer sites that strongly correlated with expression changes of nearby genes. OxLDL-regulated enhancers were enriched for several transcription factors including C/EBP-beta, which has no previously documented role in foam cell formation. OxLDL exposure up-regulated C/EBP-beta expression and increased genomic binding events, most prominently around genes involved in inflammatory response pathways. Variants at CAD-associated loci were significantly and specifically enriched in the subset of chromatin sites altered by oxLDL exposure, including rs72664324 in an oxLDL-induced enhancer at the PPAP2B locus. OxLDL increased C/EBP beta binding to this site and C/EBP beta binding and enhancer activity were stronger with the protective A allele of rs72664324. In addition, expression of the PPAP2B protein product LPP3 was present in foam cells in human atherosclerotic plaques and oxLDL exposure up-regulated LPP3 in macrophages resulting in

increased degradation of pro-inflammatory mediators. Our results demonstrate a genetic mechanism contributing to CAD risk at the PPAP2B locus and highlight the value of studying epigenetic changes in disease processes involving pathogenic environmental stimuli.

PMID: 25835000

9.2 Example Scripts

9.2.1 Heat map of intersections between different samples of a genomic feature

This script takes a a directory of files with genomic coordinates in bed file format and computes the Jaccard statistic for all possible comparisons including comparison of a file against itself to give a Jaccard of 1 (I do not include an example of the heat map generated in this thesis—but it is in essence a more simple version of Figure 4-15). The data are output as a space delimited text file and a symmetrical, hierarchically clustered heat map is produced in R which illustrates the Jaccard statistic for all possible comparisons. This program is useful for getting an overall view of the similarity between a particular genomic feature in multiple cell types or conditions. Critically, the genomic coordinate files must first be sorted. This is typically performed at the Unix command line by first sorting by chromosome and then by start position of the genomic feature using: sort –k1,1 –k2,2n <file>. This script was tested with Ubuntu 12.04.

```
#! /usr/bin/env perl

use warnings; use strict; use File::Slurp; use Getopt::Long; use Statistics::R;

my $usage = "usage: supergenie.pl <arguments...>
options:
 --directory <path to bed file directory>
 --output    <path and name of output file>
 ";

die $usage unless $ARGV[1];
die $usage unless $ARGV[2];
```

```
my $directory;
my $output_file;


# get command line arguments
GetOptions(
    "directory=s"  => \$directory,
    "output=s"     => \$output_file,
);

# get file names in directory
my @files = read_dir "$directory";

# variable for results - in this case Jaccard statistic
my @overall_counts;

# get the Jaccard number for each file compared against the others

for my $file ( @files ) {
   push (@overall_counts, $file, " ");
   for my $file_bed (@files) {
      my $intersection_count = `bedtools Jaccard -a "$directory/$file" -b
"$directory/$file_bed"`;
      chomp $intersection_count;
      my ($Jaccard) = "$intersection_count" =~ m/\d+\s+\d+\s+(\d.*)$/;
      push (@overall_counts, $Jaccard, " ");
   }
   push (@overall_counts, "\n");
}

# output results to screen
#print "Sample @files", "\n", @overall_counts;
# output results to file. filename and path specified on the command line

open(my $out, ">$output_file") or die "error creating $output_file. $!";
print $out "Sample @files\n", @overall_counts;
close $out;

# output a heat map of the results. note the columns need to be renamed.

my $R = Statistics::R->new();
$R->startR;
$R->send("data1 = read.delim('$output_file', sep = ' ', header=T, row.names=1)");
$R->send("no_rows=nrow(data1)");
$R->send("no_cols=ncol(data1)");
$R->send("no_colsminone = no_cols - 1");
$R->send("data2=data1[1:no_rows, 1:no_colsminone]");
$R->send("new_col_names= c()");
foreach my $name (@files) {
   $R->send("new_col_names=c(new_col_names, '$name')");
}
```

```
$R->send('colnames(data2)=new_col_names');
$R->send('data3 = as.matrix(data2)');
$R->send('font_size= 2/ sqrt(no_rows)');
$R->send("pdf(file = '$output_file.pdf')");
$R->send("heat map (data3, symm=T, cexRow=font_size, cexCol=font_size)");
$R->send("dev.off()");
$R->stopR;

# typical command  heatmap_intersections.pl -directory
/Users/michaelreschen/Dropbox/chipseq/super_enhancers/rose_super_enhancers -
output /Users/michaelreschen/Documents/test_heatmap
```

9.2.2 Heat map of two different genomic feature sets

This script produces a slightly more advanced heat map derived from two different types of genomic feature. This was used to produce figure Figure 4-15. The main use is in for example comparing the ENCODE ChIP-seq data with another type of feature such as open chromatin to look for patterns of similarity between different samples and types of feature. The script was used in Ubuntu 12.04.

```
# sample command

# heatmap_intersections_two_features.pl -directory1 /path/enhancers.bed -directory2
/path/ChiPseq.bed -output /path/enhancers_vs_chip.txt


#! /usr/bin/env perl

# this program makes Jaccard heat map for comparing two feature types e.g super
enhancers and chip-seq

use warnings; use strict; use File::Slurp; use Getopt::Long; use Statistics::R;

my $usage = "usage: supergenie.pl <arguments...>
options:
 --directory1 <path to bed file directory e.g. enhancers>
 --directory2 <path to bed file directory e.g. chip>
 --output    <path and name of output file>
 ";
```

```
die $usage unless $ARGV[1];
die $usage unless $ARGV[2];
die $usage unless $ARGV[3];

my $directory_enhancer;
my $directory_chip;
my $output_file;

# get command line arguments

GetOptions(
    "directory1=s"  => \$directory_enhancer,
    "directory2=s"  => \$directory_chip,
    "output=s"      => \$output_file,
);

# get file names in directory
my @files_enhancer = read_dir "$directory_enhancer";
my @files_chip     = read_dir "$directory_chip";


# variable for results - in this case Jaccard statistic
my @overall_counts;

# get the Jaccard number for each file compared against the others

my $num_enh_files  = @files_enhancer;
my $num_chip_files = @files_chip;

my $total_bedtools = $num_enh_files * $num_chip_files;
my $counter = 0;

for my $file_enhancer ( @files_enhancer ) {
   push (@overall_counts, $file_enhancer, " ");
   for my $file_chip (@files_chip) {
      my $intersection_count = `bedtools Jaccard -a "$directory_enhancer/$file_enhancer"
-b "$directory_chip/$file_chip"`;
      chomp $intersection_count;
      my ($Jaccard) = "$intersection_count" =~ m/\d+\s+\d+\s+(\d.*)$/;
      push (@overall_counts, $Jaccard, " ");
      $counter = $counter + 1;
      my $remaining = $total_bedtools - $counter;
      print "remaining = ", $remaining, "\n";
   }
   push (@overall_counts, "\n");
}

# output results to screen
#print "Sample @files", "\n", @overall_counts;
```

```
# output results to file. filename and path specified on the command line


open(my $out, ">$output_file") or die "error creating $output_file. $!";
print $out "Sample @files_chip\n", @overall_counts;
close $out;

# output a heat map of the results. note the columns need to be renamed.

my $R = Statistics::R->new();
$R->startR;
$R->send("data1 = read.table('$output_file', header=T, row.names=1)");
$R->send('no_rows=nrow(data1)');
$R->send('no_cols=ncol(data1)');
$R->send('data2 = as.matrix(data1)');
$R->send('font_size_row = 2/ sqrt(no_rows)');
$R->send('font_size_col = 2/ sqrt(no_cols)');
$R->send("pdf(file = '$output_file.pdf')");
$R->send("heat map (data2, cexRow=font_size_row, cexCol=font_size_col)");
$R->send("dev.off()");
$R->stopR;
```

9.2.3 Correlating differential gene expression with dynamic genomic features.

Changes in cellular gene expression may occur in response to stimuli or development. There may be simultaneous changes in histone marks and transcription factor binding. How these relate to gene expression changes is often unclear. This tool was written to help detect correlations between changes in gene expression between two conditions and concurrent changes in histone marks or transcription factor binding events. This script was used to produce the underlying data points for Figure 4-13A

The user supplies three files.

1. A tab delimited file of gene expression with HUGO gene symbol in the first column and fold change (typically log base 2) in the second column.

2. A bed file of genomic sites showing dynamic signal between the two conditions used for gene expression testing, with fold change in signal in the 4th column.

3. A bed file of the genomic coordinate of genes with a HUGO gene symbol in the 4th column. Dynamic sites from file 2 are then annotated to the nearest gene. Potentially using TSS instead of whole gene body can help.

```
#!/usr/bin/perl -w

# perl program to correlate fold changes in chromatin marks with gene expression
change between two conditions.

use strict;
use Getopt::Long;
use List::Util qw(sum max min);

# print user instructions

my $usage = "usage: chromgenie.pl <arguments>
options:
 --gene_expression <path to gene expression data - gene symbol col1 and fold change
col2>
 --chrom_change <path to bed file with fold change in col4>
 --genes <path to bed file of genes each annotated with gene symbol in col4>
 ";

die $usage unless @ARGV;

my $gene_expr;
my $chrom_expr;
my $genes;

# get user input file names

GetOptions(
    "gene_expression=s"         => \$gene_expr,
    "chrom_change=s"            => \$chrom_expr,
    "genes=s"                   => \$genes,
);

my %expr;
my %hash_chrom_fold_for_gene;

open(EXPR, "< $gene_expr") or die "Can't open '$gene_expr': $!";
while(my $expr_line = <EXPR>) {
   chomp($expr_line);
   my @expra = split (/\t/, $expr_line);
   $expr{$expra[0]} = $expra[1];
}
close(EXPR);
```

```
my @anno_chrom = `closestBed -a $chrom_expr -b $genes -t first`;

foreach my $line (@anno_chrom) {
   chomp ($line);
   my @anno_chrom_split  = split(/\t/, $line);
   my @chrom_gene_fold = @anno_chrom_split[8,3];
   my $fold = $chrom_gene_fold[1];
   my $gene_name = $chrom_gene_fold[0];
   push (@{$hash_chrom_fold_for_gene{$gene_name}}, $fold);
}

my $gene_based_group_of_chrom_fold_change;
my @super_final_result;

for $gene_based_group_of_chrom_fold_change ( keys %hash_chrom_fold_for_gene ) {
   my @pre_average = @{
$hash_chrom_fold_for_gene{$gene_based_group_of_chrom_fold_change} };
   my $average   = sum (@pre_average) / @pre_average;
   my $maxi      = max (@pre_average);
   my $mini      = min (@pre_average);
   my $num_chrom = @pre_average;
   if (exists $expr{$gene_based_group_of_chrom_fold_change}) {
      push (@super_final_result, $gene_based_group_of_chrom_fold_change,
"\t",$expr{$gene_based_group_of_chrom_fold_change},"\t",$average,"\t", $maxi,"\t",
$mini,"\t", $num_chrom,"\n");
   }
}

my $outfile = "$chrom_expr.output";
open(my $out, ">$outfile") or die "error creating $outfile. $!";
print $out
"gene_name\tgene_change\tmean_chromatin_change\tmax_chrom_change\tmin_chrom
_change\tnum_of_chrom_sites\n",@super_final_result;
close $out;
```

9.2.4 Correlating gene expression with super-enhancer status

This script is used to expression of genes overlapping or near enhancers with super enhancer status compared to genes that are near enhancer without super enhancer status. This script was used to generate the box plots and statistics for Figure 4-14. The expectation is that genes which are overlapping or near super enhancer will have higher expression overall compared to those that are not near super enhancers. The

script requires several input files including a gene expression file with the gene name in the first column and numerical expression value in another column (the number of the column is input to the script as a command line argument—NB the first column is numbered column 1.)

Example gene expression file (tab-delimited, each gene should only appear once):

name	ID	foam_mean	macro_mean	difference	absolute_difference
LDLR	1440736	8.23336360167639	11.3233003347604	-3.08994	3.08994
F13A1	2230241	8.9964997225946	11.8093080177823	-2.81281	2.81281
PLIN2	1400446	12.3854596082269	9.6884521324441	2.69701	2.69701

The super-enhancer data is the output from the ROSE program with the header lines manually removed. Example 3 lines of tab-delimited file:

37_MACS_peak_25454_lociStitched	chr20	48753100	49192307	37	321117	379957.9757	1	1
3_MACS_peak_41758_lociStitched	chr9	137215311	137353426	3	128191	296974.873	2	1
5_MACS_peak_41706_lociStitched	chr9	134450101	134623161	5	169103	257894.012	3	1

22_MACS_peak_36297_lociStitched chr7 2641365 2885111 22 177665 254470.824 4 1

The user also supplies a file of gene coordinates in bed format. The gene name can be repeated if there are different transcripts as each gene hit will only be counted once. Example 3 lines of tab-delimited file:

chr19 58858171 58864865 NM_130786 A1BG

chr19 58863335 58866549 NR_015380 A1BG

chr10 52559168 52645435 NM_001198820 A1CF

Typical command used to run the script:

```
super_genie_with_R_v1.1.pl –genes /path/RefSeq.txt –enhancers /path/enhancers.txt –gene_expression /path/expression.txt –output macro –column_expr 3 –stats yes –nearest yes
```

The script was run on Ubuntu 12.04.

Script:

```
#!/usr/bin/env perl -w

# perl program to correlate super enhancer status with gene expression of genes overlapping the super enhancer
# the three inputs are:
# gene position file e.g. refseq
# output from ROSE program with all enhancers and their super status
# gene expression file
```

```
use strict;
use Getopt::Long;
use List::Util qw(sum max min);
use List::MoreUtils qw( uniq );
use Statistics::R;

my $usage = "usage: supergenie.pl <arguments...>
options:
 --gene_expression <path to gene expression data>
 --enhancers <path to enhancer file>
 --genes <path to bed file of gene start and end coordinates>
 --output <name to append to gene expression file name to make output filename>
 --column_expr <numerical value of column with gene expression data. NB first column
is zero>
 --nearest <type 'yes' if want to included nearest gene to enhancer, not just overlapping
genes>
 --stats <type 'yes' or leave blank. if yes then will draw boxplot and print basic stats>
 ";

die $usage unless @ARGV;

my $gene_expr;
my $sup_enhancer;
my $gene_bed;
my $output_name;
my $col_expr_value;
my $stats   = "empty";
my $nearest = "empty";

GetOptions(
    "gene_expression=s"  => \$gene_expr,
    "enhancers=s"       => \$sup_enhancer,
    "genes=s"           => \$gene_bed,
    "output=s"          => \$output_name,
    "column_expr=i"     => \$col_expr_value,
    "nearest:s"         => \$nearest,
    "stats:s"           => \$stats,
);


# now get the gene position data
# condense if there are multiple transcripts with the same gene name
# by making hash of arrays with gene symbol = key
# and for each the array is just the all the start and and positions for that gene
# then just take the smallest and largest number to define the max width of the gene.

# open the file first - path given at command line

open(GENE_POS, "< $gene_bed") or die "Can't open '$gene_bed': $!";

# then make hash of gene name and chromosome
```

```
my %gene_chr;
my %gene_pos;

while (my $gene = <GENE_POS>) {
   chomp $gene;
   my @gene = split(/\t/, $gene);
   $gene_chr{$gene[4]} = $gene[0];
   push (@{$gene_pos{$gene[4]}}, $gene[1], $gene[2]);
};

close(GENE_POS);

# then make a hash of gene positions, get min and max and write a final array of
amalgamated positions for each gene symbol

my @condensed_gene_bed;

foreach my $gene_symbol ( keys %gene_pos ) {
   my @positions = @{ $gene_pos{$gene_symbol} };
   my $maxi      = max (@positions);
   my $mini      = min (@positions);
   my $final = join("\t", $gene_chr{$gene_symbol}, $mini, $maxi, $gene_symbol, "\n");
   push(@condensed_gene_bed, $final);
};

my $outfile = "$gene_bed.output";
open(my $out, ">$outfile") or die "error creating $outfile. $!";
print $out "@condensed_gene_bed";
close $out;


# nb note the process above was technically not necessary as the script is just looking
for genes hit.
# so if a gene was hit 10 times it wouldn't have mattered as it is count once only.

# use bedtools to get all the genes which overlap a super enhancer or a non super
enhancer

# first get the ROSE super enhancer output and split into stitched enhancers with and
without super status



my @super_enh_bed;
my @not_super_enh_bed;

open(SUP_ENH, "< $sup_enhancer") or die "Can't open '$sup_enhancer': $!";

while (my $line = <SUP_ENH>) {
   chomp $line;
   my @line = split (/\t/, $line);
```

```
   if ($line[8] == 1) {
      my $super_enh_bed = join("\t", $line[1], $line[2], $line[3], "\n");
           push (@super_enh_bed, $super_enh_bed);
   } else {
      my $not_super_enh_bed = join("\t", $line[1], $line[2], $line[3], "\n");
           push (@not_super_enh_bed, $not_super_enh_bed);
   }
}

close (SUP_ENH);

my $outfile2 = "$gene_bed.output.not_sup_enh_bed";
open(my $out2, ">$outfile2") or die "error creating $outfile2. $!";
print $out2 "@not_super_enh_bed";
close $out2;

my $outfile3 = "$gene_bed.output.sup_enh_bed";
open(my $out3, ">$outfile3") or die "error creating $outfile3. $!";
print $out3 "@super_enh_bed";
close $out3;

# now get the gene expression data as a hash

my %gene_expr;
open(GENE_EXPR, "< $gene_expr") or die "Can't open '$gene_expr' $!";
while (my $line1 = <GENE_EXPR>) {
   chomp $line1;
   my @line1 = split (/\t/, $line1);
   $gene_expr{$line1[0]} = $line1[$col_expr_value];
}
close (GENE_EXPR);


# now use bedtools intersect to get all genes overlapping a stitched enhancer

my @sup_enh_intersection          =`bedtools intersect -wa -a $gene_bed.output -b
$gene_bed.output.sup_enh_bed`;
my @not_sup_enh_intersection      =`bedtools intersect -wa -a $gene_bed.output -b
$gene_bed.output.not_sup_enh_bed`;



my @gene_list_overlap_super;
my $name_gene;
my @final_result = ("gene\texpression\tstatus\n");


foreach my $line3 (@sup_enh_intersection) {
   chomp $line3;
   my @line3 = split (/\t/, $line3);
   $name_gene = $line3[3];
   push (@gene_list_overlap_super, $name_gene);
```

```
}

if ($nearest eq 'yes') {
   my @sup_enh_intersection_closest     = `bedtools closest -a
$gene_bed.output.sup_enh_bed -b $gene_bed.output`;

   foreach my $line_3 (@sup_enh_intersection_closest) {
      chomp $line_3;
      my @line_3 = split (/\t/, $line_3);
      $name_gene = $line_3[6];
      push (@gene_list_overlap_super, $name_gene);
   }
}

my @unique_super_gene = uniq @gene_list_overlap_super;

# now extract gene expression values for subset of genes associated with a super
enhancer

foreach my $unique_super_gene (@unique_super_gene) {
   if (exists $gene_expr{$unique_super_gene}) {
       push (@final_result, $unique_super_gene, "\t", $gene_expr{$unique_super_gene},
"\tsuper\n");
   }
}

# perform the same process for genes associated with stitched non-super-enhancers.

my @gene_list_not_overlap_super;

foreach my $line4 (@not_sup_enh_intersection) {
   chomp $line4;
   my @line4 = split (/\t/, $line4);
   my $name_gene = $line4[3];
   push (@gene_list_not_overlap_super, $name_gene);
}

if ($nearest eq 'yes') {
   my @not_sup_enh_intersection_closest = `bedtools closest -a
$gene_bed.output.not_sup_enh_bed -b $gene_bed.output`;

   foreach my $line_4 (@not_sup_enh_intersection_closest) {
      chomp $line_4;
      my @line_4 = split (/\t/, $line_4);
      my $name_gene = $line_4[6];
      push (@gene_list_not_overlap_super, $name_gene);
   }
}

my @unique_not_super_gene = uniq @gene_list_not_overlap_super;
```

```
# ensure the gene cannot be counted twice if it was in proximity to a super enhancer.
Super enhancers take priority to remove any gene from the non-super enhancer list
which has already been associated with a super enhancer.

my %hash;

$hash{$_}++ for (@unique_super_gene);

foreach my $unique_not_super_gene (@unique_not_super_gene) {
   if (exists $gene_expr{$unique_not_super_gene}) {
      unless (exists $hash{$unique_not_super_gene}) {
         push (@final_result, $unique_not_super_gene, "\t",
$gene_expr{$unique_not_super_gene}, "\tnot_super\n");
      }
  }
}

# produce the output file with gene, expression value and enhance status.

my $outfile4 = "$gene_bed.output.final_result.$output_name";
open(my $out4, ">$outfile4") or die "error creating $outfile4. $!";
print $out4 "@final_result";
close $out4;

# calculate t.test statistic for expression difference between genes assigned or not
assigned to super enhancers. Produce boxplot of the output.

if ($stats eq 'yes') {

   my $R = Statistics::R->new();
   $R->startR;
   $R->send('library(ggplot2)');
   $R->send("data1 = read.delim('$gene_bed.output.final_result.$output_name')");
   $R->send('ggplot(data1, aes(x=status, y=expression)) + geom_boxplot() +
theme_bw()');
   $R->send("ggsave('$gene_bed.output.final_result.$output_name.pdf', width=8,
height=8, units='cm')");
   $R->send('test = t.test(data1$expression ~ data1$status)');
   $R->send('print(test)');
   my $ret  = $R->read;
   print $ret;
   $R->send('print(test$p.value)');
   my $ret2 = $R->read;
   print "The exact P value is: ", $ret2;
   $R->stopR;
}
print "\nCompleted ok, goodluck!\n\n";
```

9.3 Appendix of transcription factor motif enrichment in dynamic chromatin sites.

In addition to the transcription factor enrichment motif finding in double dynamic enhancer chromatin sites (i.e. non-promoter dynamic chromatin site, within 300bp of a dynamic h3k27ac site) we repeated the analysis after separating the sites into those that had increased FAIRE-seq signal in response to oxLDL and those with less signal.

1270 sites were identified with oxLDL induced FAIRE-seq signal. These were submitted to SeqPos in the Cistrome computing environment using the Cistrome custom database of transcription factor motifs. Enriched transcription factor motifs were clustered based on a similarity cut-off of 0.6 and the output is shown below. Overall the enrichments are broadly similar to that when using all the double dynamic sites.

clusters	collapsed_id	factor	DNA binding domain	hits	cutoff	zscore	#NAME?	similarity to top	mean_position
1	MC00456	JUND	Leucine Zipper Family	919	4.43	-26.04	690.78		-0.20
1	denovo10		None	777	4.69	-25.27	690.78	0.91	-0.21
1	MC00371	JUNB	Leucine Zipper Family	737	4.67	-25.10	690.78	0.90	-0.21
1	MC00349	Jun	Leucine Zipper Family	946	6.21	-24.79	690.78	0.90	-0.20
1	MC00356	Tbx21	Transcription Factor T-Domain	996	6.05	-24.54	690.78	0.87	-0.21
1	denovo0		None	547	6.72	-24.45	690.78	0.94	-0.22
1	MC00351	FOSL1	Leucine Zipper Family	286	4.80	-24.34	690.78	0.88	-0.21
1	MC00330	FOS	Leucine Zipper Family	639	4.93	-24.02	690.78	0.91	-0.22
1	denovo21		None	900	4.99	-23.50	690.78	0.90	-0.18
1	MC00	JUN	Leucine Zipper	7	5.	-	690	0.90	-0.20

	321		Family	4	18	23.	.78		
				7		42			
				9		-			
	M004		Leucine Zipper	0	3.	21.	690		
1	95	BACH1	Family	3	19	60	.78	0.94	-0.17
				8		-			
	M004		Leucine Zipper	1	5.	21.	690		
1	90	BACH2	Family	6	15	50	.78	0.95	-0.17
				8		-			
	MS00		Leucine Zipper	0	3.	20.	690		
1	319	JDP2	Family	5	58	86	.78	0.93	-0.17
				7		-			
	MS00		Leucine Zipper	9	3.	20.	690		
1	336	NFE2	Family	9	09	71	.78	0.92	-0.17
				9		-			
	MS00		Leucine Zipper	3	2.	20.	690		
1	324	Jdp2	Family	6	56	56	.78	0.95	-0.16
				1					
				1		-			
	M005		Leucine Zipper	4	2.	10.	599		
1	14	ATF4	Family	0	24	65	.49	0.78	-0.08
				1					
				0		-			
	deno			5	5.	17.	690		
2	vo8		None	2	18	20	.78		-0.13
				9		-			
	deno			6	5.	15.	690		
2	vo14		None	2	20	27	.78	0.69	-0.12
				9		-			
	MA00		Homeodomain	3	5.	10.	585		
2	67	PAX2	Family	9	08	51	.53	0.77	-0.09
				9		-			
	MC00	NFE2L	Leucine Zipper	5	2.	16.	690		
3	339	2	Family	0	79	80	.78		-0.13
				1					
				1		-			
	M009		Leucine zipper	5	4.	16.	690		
3	83	MAF	Family	4	48	80	.78	0.92	-0.12
				1					
				1	-	-			
	MS00		Leucine Zipper	8	1.	9.3	466		
3	326	MAFF	Family	3	51	2	.08	0.87	-0.07
				4		-			
	MS00		Leucine Zipper	6	4.	8.3	380		
3	331	MAFK	Family	4	76	7	.82	0.84	-0.10
3	MS00	Mafb	Leucine Zipper	5	4.	-	312	0.79	-0.09

	333		Family	2	06	7.5	.68		
				1		3			
				5		-			
	MS00		Leucine Zipper	1	3.	6.0	211		
3	327	MAFG	Family	2	38	7	.37	0.81	-0.08
				1					
				1		-			
	MS00		Helix-Loop-Helix	0	3.	15.	690		
4	289	TFEB	Family	6	39	24	.78		-0.11
				1					
				0		-			
	MS00		Helix-Loop-Helix	9	3.	14.	690		
4	288	TFE3	Family	8	84	72	.78	0.99	-0.11
				9		-			
	deno			0	5.	14.	690		
4	vo4		None	7	76	61	.78	0.83	-0.12
				1					
				1		-			
	MS00		Helix-Loop-Helix	7	3.	13.	690		
4	808	TFEC	Family	3	40	86	.78	0.96	-0.10
				1					
				1		-			
	MC00		Helix-Loop-Helix	4	1.	13.	690		
4	248	USF1	Family	0	34	26	.78	0.96	-0.10
				8		-			
	MC00		Helix-Loop-Helix	0	3.	12.	690		
4	169	Usf2	Family	0	04	83	.78	0.92	-0.11
				4		-			
	MC00		Leucine Zipper	4	5.	11.	690		
4	473	ATF3	Family	2	61	46	.78	0.94	-0.14
				1					
				2	-	-			
	MC00		Helix-Loop-Helix	6	0.	11.	642		
4	461	Max	Family	0	48	04	.16	0.89	-0.08
				1					
				0		-			
	MC00		Helix-Loop-Helix	3	2.	10.	543		
4	124	MAX	Family	6	52	11	.23	0.93	-0.08
				1					
				2	-	-			
	MS00	SREBF	Helix-Loop-Helix	2	0.	10.	542		
4	281	2	Family	5	45	10	.47	0.94	-0.07
				1					
				1		-			
	MS00	BHLHE	Helix-Loop-Helix	8	0.	9.6	499		
4	251	41	Family	2	01	7	.32	0.94	-0.07

				8		-9			
	MC00		Helix-Loop-Helix	7	3.	.4	476		
4	394	MYC	Family	0	74	3	.66	0.88	-0.08
				9		-9			
	MS00	BHLHB	Helix-Loop-Helix	5	2.	.3	466		
4	247	2	Family	9	68	3	.62	0.94	-0.08
				9		-9			
	MS00	MLXIP	Helix-Loop-Helix	8	3.	.1	445		
4	265	L	Family	5	01	0	.43	0.94	-0.08
				1					
				2	-	-			
	MS00		Helix-Loop-Helix	5	0.	9.0	444		
4	282	Srebf1	Family	3	69	9	.06	0.93	-0.07
				7		-9			
	MC00		Helix-Loop-Helix	0	5.	.0	442		
4	362	Arntl	Family	0	06	6	.10	0.84	-0.09
				8		-8			
	MS00	BHLHB	Helix-Loop-Helix	8	2.	.9	431		
4	248	3	Family	6	58	5	.62	0.93	-0.08
				8		-8			
	MS00		Helix-Loop-Helix	6	2.	.8	425		
4	252	Bhlhb2	Family	4	78	8	.50	0.95	-0.08
				1					
				2		-8			
	MC00		Helix-Loop-Helix	1	1.	.8	424		
4	373	Mxi1	Family	2	66	7	.07	0.88	-0.07
				1					
				2		-8			
	MS00		Helix-Loop-Helix	0	0.	.8	423		
4	270	Mlx	Family	3	42	6	.16	0.93	-0.07
				1					
				2		-8			
	M002	SREBF	Helix-Loop-Helix	0	1.	.8	421		
4	20	1	Family	7	50	4	.86	0.93	-0.07
				1					
				2		-8			
	MC00		Helix-Loop-Helix	6	0.	.7	411		
4	532	Mycn	Family	4	77	3	.60	0.89	-0.06
				1					
				1		-8			
	MS00		Helix-Loop-Helix	5	0.	.4	391		
4	266	MLX	Family	9	77	9	.04	0.95	-0.07
				8		-7			
	MC00		Helix-Loop-Helix	6	3.	.8	340		
4	407	Myc	Family	9	63	8	.12	0.86	-0.07
4	MS00	MNT	Helix-Loop-Helix	1	0.	-	297	0.94	-0.06

	267		Family	2 6 9 8	32	7.3 2 - 	.40		
4	MS00 260	HEY2	Helix-Loop-Helix Family	6 3 2	3. 15	5.6 6 -	187 .21	0.87	-0.06
4	MS00 258	HEY1	Helix-Loop-Helix Family	9 5 2	5. 43	5.1 8 -	160 .24	0.88	-0.09
4	MC00 254	E2F6	Transcription Factor Family	5 9 2	6. 07	4.9 0 -	145 .24	0.71	-0.09
4	MC00 262	ARNT	Helix-Loop-Helix Family	8 2 7	6. 30	4.8 9 -	145 .22	0.73	-0.09
4	MS00 253	CLOCK	Helix-Loop-Helix Family	8 7 4	3. 67	4.4 6 -	123 .88	0.84	-0.05
4	MC00 224	EPAS1	Helix-Loop-Helix Family	8 9 7	5. 34	3.1 8 -	72. 25	0.76	-0.05
5	MC00 154	Jund	Leucine Zipper Family	4 8 8	5. 17	13. 72 -	690 .78		-0.12
5	MC00 411	BATF	Leucine zipper Family	1 0 6	4. 74	10. 16 -	548 .23	0.99	-0.09
5	MC00 508	Batf	Leucine zipper Family	6 3 5	5. 34	9.3 5 -	469 .07	0.97	-0.09
5	MC00 227	IRF4	Interferon Regulatory Factor	1 4 5	6. 14	8.3 0 -	374 .81	0.99	-0.10
5	MC00 190	Irf4	Interferon Regulatory Factor	2 4 4	5. 91	5.3 8 -	171 .24	0.94	-0.07
6	MC00 121	Cebpb	Leucine Zipper Family	3 8 6	6. 71	12. 90 -	690 .78		-0.15
6	MC00 122	Cebpa	Leucine Zipper Family	0 4 8	5. 96	12. 87 -	690 .78	0.98	-0.13
6	MS00 300	CEBPE	Leucine Zipper Family	0 1	4. 53	12. 00	690 .78	0.89	-0.11
6	MC00	CEBPA	Leucine Zipper	7	4.	-	690	0.95	-0.11

	273		Family	0.55	76	11.57	.78		
6	MS00299	CEBPD	Leucine Zipper Family	2.27	5.34	-11.19	659.06	0.89	-0.12
6	MS00302	CEBPG	Leucine Zipper Family	1.59	3.31	-11.16	656.18	0.87	-0.11
6	MS00293	ATF4	Leucine Zipper Family	3.06	2.70	-10.82	618.10	0.85	-0.09
6	MC00272	CEBPB	Leucine Zipper Family	1.29	5.73	-10.31	563.53	0.95	-0.11
6	MS00337	NFIL3	Leucine Zipper Family	6.310	3.72	-9.20	454.31	0.89	-0.08
6	MS00317	HLF	Leucine Zipper Family	2.511	3.03	-7.29	294.93	0.85	-0.06
6	MS00295	Atf4	Leucine Zipper Family	5.35	0.28	-7.26	292.63	0.82	-0.06
6	M00621	CEBPD	Leucine Zipper Family	3.87	6.13	-7.24	291.52	0.89	-0.09
6	MS00315	DBP	Leucine Zipper Family	9.17	2.99	-5.89	200.86	0.79	-0.06
6	MS00316	Dbp	Leucine Zipper Family	7.510	2.18	-3.87	97.97	0.78	-0.05
7	MS00320	JDP2	Leucine Zipper Family	3.612	0.28	-9.26	459.79		-0.07
7	MS00313	Creb5	Leucine Zipper Family	6.412	2.21	-8.65	405.03	0.98	-0.06
7	MS00294	ATF7	Leucine Zipper Family	2.22	2.01	-8.47	389.43	0.95	-0.06

7	UP00020	ATF1	Leucine zipper Family	6 4 6 1 0	2. 60	-8.29	374.38	0.90	-0.09
7	MS00325	Jdp2	Leucine Zipper Family	9 1 6	-0. 53	-8.28	373.34	0.99	-0.07
7	MC00163	Creb1	Leucine Zipper Family	1 7 2	4. 39	-7.95	345.81	0.83	-0.09
7	MS00296	BATF3	Leucine Zipper Family	9 5 3	0. 73	-6.75	256.13	0.89	-0.11
7	M00041	ATF2::JUN	Leucine zipper Family	4 2 1 2	5. 29	-4.22	112.94	0.75	-0.07
8	denovo12		None	3 6 1 1	2. 71	-8.41	384.10		-0.06
8	MC00281	PBX3	Homeodomain Family	4 7 4	3. 26	-4.93	147.04	0.66	-0.04
9	denovo18		None	0 3 7	6. 65	-7.42	304.70		-0.10
9	MC00295	Esr1	Hormone-nuclear Receptor Family	7 8 6	3. 08	-3.34	77.71	0.62	-0.04
9	MC00335	ESR1	Hormone-nuclear Receptor Family	1 1 7	4. 44	-3.15	71.25	0.62	-0.05
10	MS00657	NR2F1	Hormone-nuclear Receptor Family	1 7 4	6. 26	-7.37	300.71		-0.07
10	M01282	PPARA	Hormone-nuclear Receptor Family	2 0 1 1	7. 25	-7.26	292.94	0.96	-0.10
10	MC00524	Nr1d1	Hormone-nuclear Receptor Family	8 8 4	3. 77	-5.79	194.41	0.82	-0.05
10	M01268	NR1H4	Hormone-nuclear Receptor Family	0 2	6. 84	-5.36	170.08	0.94	-0.08

10	MC00186	Nr1d2	Hormone-nuclear Receptor Family	8963	4.86	-8	149.76	0.90	-0.05
10	M01722	RORB	Hormone-nuclear Receptor Family	2911	5.42	-3.80	95.33	0.87	-0.07
10	MC00471	Rxra	Hormone-nuclear Receptor Family	3312	8.30	-3.59	87.12	0.95	-0.10
11	MC00427	ETV1	Ets Domain Family	2411	1.21	-6.90	267.03		-0.05
11	MC00506	Spi1	Ets Domain Family	6511	1.13	-6.60	246.01	0.84	-0.05
11	MC00180	Sfpi1	Ets Domain Family	6511	1.13	-6.60	246.01	0.84	-0.05
11	MC00395	Sfpi1	Ets Domain Family	6511	1.13	-6.59	245.36	0.84	-0.05
11	MC00151	SPI1	Ets Domain Family	2089	-0.92	-6.45	236.30	0.81	-0.05
11	UP00085	Sfpi1	Ets Domain Family	2211	4.99	-6.44	235.20	0.89	-0.06
11	MC00385	Ets1	Ets Domain Family	6251	3.70	-5.71	189.80	0.94	-0.05
11	MS00039	SPIB	Ets Domain Family	6061	4.43	-5.62	184.38	0.79	-0.07
11	MC00367	GABPA	Ets Domain Family	5012	-0.10	-5.11	156.32	0.92	-0.04
11	MC00513	SFPI1	Ets Domain Family	7706	4.05	-4.66	133.50	0.83	-0.06

11	MC00160	ETV3	Ets Domain Family	126.75	1.82	-4.59	130.22	0.89	-0.04
11	MS00041	Spic	Ets Domain Family	75.76	3.56	-4.59	130.16	0.79	-0.06
11	MC00168	ERG	Ets Domain Family	100.91	2.66	-4.43	122.58	0.92	-0.04
11	MC00457	ELF1	Ets Domain Family	116.59	1.48	-4.31	117.07	0.92	-0.04
11	MC00337	Ikzf3	BetaBetaAlpha-zinc finger Family	94.81	4.67	-4.26	114.93	0.89	-0.04
11	MS00021	ETV6	Ets Domain Family	121.58	2.46	-4.16	110.65	0.88	-0.04
11	MS00040	SPIC	Ets Domain Family	58.37	4.51	-3.66	89.69	0.76	-0.04
11	M00341	GABPB1	Ets Domain Family	74.21	3.45	-3.49	83.10	0.90	-0.04
11	MC00206	ELK1	Ets Domain Family	126.51	-1.04	-3.38	79.35	0.92	-0.03
11	MS00022	Elf5	Ets Domain Family	100.06	4.68	-3.13	70.35	0.88	-0.04
12	M00647	NR1H3	Hormone-nuclear Receptor Family	65.51	6.99	-6.59	245.18		-0.23
12	MC00448	Nr1h2	Hormone-nuclear Receptor Family	74.46	7.84	-6.37	230.65	0.86	-0.14
12	MS00700	Rarb	Hormone-nuclear Receptor Family	16.18	0.28	-5.12	156.95	0.82	-0.06
12	MS00	Rarg	Hormone-nuclear	9	-	-	139	0.81	-0.05

	705		Receptor Family	17	2.03	4.78	.39		
12	MS00684	RARG	Hormone-nuclear Receptor Family	11 84 12	1.03	-3.88	98.44	0.82	-0.04
13	M01709	MAFA	Leucine Zipper Family	63 7	3.62	-6.38	231.58		-0.05
14	MC00319	EGR2	BetaBetaAlpha-zinc finger Family	96 4	3.50	-6.21	220.46		-0.06
14	MC00430	Egr2	BetaBetaAlpha-zinc finger Family	65 12	5.01	-5.50	177.86	0.95	-0.07
14	MC00320	EGR1	BetaBetaAlpha-zinc finger Family	14 1	1.11	-5.45	175.15	0.96	-0.05
14	denovo13		None	58 4	7.71	-5.28	165.64	0.80	-0.12
15	MS00343	FOXB1	Forkhead Domain Family	17 10	4.30	-6.16	217.44		-0.09
16	denovo2		None	79 5	3.43	-5.99	206.48		-0.05
16	MC00178	Nkx2-5	Homeodomain Family	19 6	5.42	-4.57	129.25	0.65	-0.06
17	M00694	E4F1	Helix-Loop-Helix Family	75 12	3.05	-5.74	191.43		-0.06
17	M00179	ATF2	Leucine Zipper Family	69 3	1.33	-5.32	167.53	0.82	-0.04
18	MC00130	Pparg	Hormone-nuclear Receptor Family	54 4	6.43	-5.49	177.47		-0.09
18	MC00134	Ppara	Hormone-nuclear Receptor Family	38	6.16	-5.36	170.12	0.95	-0.08

18	MC00467	RXRA	Hormone-nuclear Receptor Family	29191	6.65	-4.88	144.26	0.98	-0.09
18	MC00142	PPARG	Hormone-nuclear Receptor Family	96661	4.08	-4.48	124.85	0.96	-0.05
19	M00139	AHR	Helix-Loop-Helix Family	12531	-3.81	-5.38	171.11		-0.04
20	M00115	CNTN2::CREB1	Leucine zipper Family	11505	-4.052	-5.30	166.89		-0.05
21	denovo5		None	59310	6.45	-5.23	162.74		-0.06
21	M00468	KLF12	BetaBetaAlpha-zinc finger Family	62110	4.53	-5.10	155.89	0.77	-0.05
22	MS00311	Creb3l2	Leucine Zipper Family	11000	-0.051	-4.96	148.79		-0.05
22	MS00255	HES5	Helix-Loop-Helix Family	11405	-1.77	-4.60	130.81	0.76	-0.04
22	M01009	HES1	Helix-Loop-Helix Family	98121	4.99	-3.49	83.14	0.62	-0.05
22	MS00256	HES7	Helix-Loop-Helix Family	67111	-6.00	-3.43	81.16	0.74	-0.03
23	M01175	ZBTB7B	BetaBetaAlpha-zinc finger Family	78112	2.03	-4.32	117.79		-0.04
23	M00649	MAZ	BetaBetaAlpha-zinc finger Family	69122	0.57	-3.35	78.25	0.96	-0.03
24	M00733	SMAD4	MH1 Domain Family	12	0.09	-4.1	109.49		-0.04

				6		4			
				6					
				1					
				2	-	-			
	MS00		Hormone-nuclear	5	5.	3.7	91.		
25	694	RXRG	Receptor Family	9	11	1	82		-0.04
				1					
			CP2 Transcription	1		-			
	M006		Factor Domain	8	2.	3.4	83.		
26	44	UBP1	Family	6	92	9	25		-0.03
				9		-			
	M007		Helix-Loop-Helix	3	4.	3.1	71.		
26	12	MYOG	Family	0	50	7	92	0.84	-0.04
				1					
				2		-			
	M003		Homeodomain	5	1.	3.2	75.		
27	73	PAX4	Family	3	33	8	63		-0.03
				3		-			
	M009		Hormone-nuclear	6	6.	3.1	69.		
28	61	VDR	Receptor Family	3	20	1	67		-0.06

The analysis was repeated using the subset of double dynamic chromatin sites which had oxLDL-suppressed FAIRE-seq signal (473 sites) and the corresponding data output is shown below. In contrast to the upregulated sites there was little enrichment for transcription factor motifs in these sites apart from for ETS transcription factors.

clusters	collapsed_id	factor	DNA binding domain	hits	cut off	zscore	-log10 P value	similarity to top	mean_position
				3		-			
	MC003			4	4.8	5.5			
1	85	Ets1	Ets Domain Family	3	09	77	182.167		-0.087
				2		-			
	MC004			3	5.0	4.4			
1	27	ETV1	Ets Domain Family	9	38	02	121.342	0.939	-0.088
				3		-			
	MC003		BetaBetaAlpha-zinc	7	4.6	4.3			
1	37	Ikzf3	finger Family	3	03	32	118.144	0.911	-0.07
				1		-			
	M0067			8	4.8	4.0			
1	8	ETV7	Ets Domain Family	8	86	62	106.228	0.875	-0.094
				1		-			
	MC001			8	5.5	3.9			
1	68	ERG	Ets Domain Family	3	09	27	100.537	0.928	-0.093
				4		-			
	MC001			2	3.0	3.6			
1	60	ETV3	Ets Domain Family	8	42	01	87.498	0.878	-0.058
				3		-			
	MC003	GABP		7	2.4	3.4			
1	67	A	Ets Domain Family	9	86	6	82.167	0.906	-0.06

				4		-			
	MS000			6	2.0	3.3			
1	21	ETV6	Ets Domain Family	1	47	77	79.101	0.89	-0.053
				2		-			
	denov			7	6.5	3.3			
1	o0		None	6	75	69	78.822	0.689	-0.068
				3		-			
	denov			6	5.1	3.1			
1	o22		None	5	32	96	72.687	0.813	-0.058
				3		-			
	MC004			8	2.5	3.1			
1	57	ELF1	Ets Domain Family	9	51	17	69.971	0.907	-0.055
				1		-			
	denov			5	6.5	5.3			
2	o17		None	4	27	52	169.514		-0.126
				1		-			
	denov			9	6.4	4.9			
3	o15		None	2	69	25	146.791		-0.106
				1		-			
	denov			9	6.6	4.2			
4	o16		None	9	85	33	113.703		-0.094
				3		-			
	denov			7	5.0	4.2			
5	o18		None	7	21	11	112.742		-0.068
				1		-			
	denov			4	6.4	3.7			
5	o1		None	4	33	67	93.998	0.83	-0.101
				4		-			
	M0064		CP2 Transcription	4	2.9	4.0			
6	4	UBP1	Factor Domain Family	1	29	56	105.971		-0.061
				4		-			
	MC003	Myod		4	2.9	3.4			
6	69	1	Helix-Loop-Helix Family	0	16	03	80.074	0.91	-0.055
				4		-			
	M0128	NEUR		2	4.1	3.2			
6	8	OD1	Helix-Loop-Helix Family	5	26	67	75.163	0.775	-0.054
				3		-			
	M0170	MAF		7	4.9	3.7			
7	9	A	Leucine Zipper Family	8	81	57	93.626		-0.063
				3		-			
	denov			9	5.3	3.7			
8	o7		None	7	59	48	93.264		-0.061
				4		-			
	denov			7	2.4	3.1			
9	o5		None	0	91	6	71.443		-0.051

9.4 PHACTR1 long and intermediate transcript alignment

Sequence alignment of the coding sequence of the long and intermediate PHACTR1 transcripts.

```
long            ATGGATTATCCCAAAATGGATTATTTTCTGGATGTAGAGTCTGCTCACAGACTCTTGGAT 60
intermediate    ------------------------------------------------------------
long            GTTGAGTCAGCTCAAAGATTCTTCTACAGTCAAGGAGCTCAAGCTCGCCGGGCGACCCTG 120
intermediate    ------------------------------------------------------------
long            CTCCTGCCTCCCACATTAATGGCGGCATCCTCGGAGGATGATATAGACCGGCGGCCCATC 180
intermediate    ------------------------------------------------------------
long            CGGAGAGTGCGCTCCAAGAGCGACACGCCGTACCTCGCAGAGGCCAGGATCTCCTTTAAC 240
intermediate    ------------------------------------------------------------
long            CTGGGGGCAGCTGAGGAAGTGGAGAGGCTGGCGGCGATGCGTTCTGACTCCCTCGTCCCA 300
intermediate    ----------------------------------ATGCGTTCTGACTCCCTCGTCCCA 24
                                                   ************************
long            GGCACCCACACCCCACCCATCCGCAGGAGAAGTAAGTTTGCCAACCTGGGAAGGATTTTC 360
intermediate    GGCACCCACACCCCACCCATCCGCAGGAGAAGTAAGTTTGCCAACCTGGGAAGGATTTTC 84
                ************************************************************
long            AAGCCTTGGAAATGGAGGAAGAAGAAAAGCGAAAAGTTCAAACACACGTCAGCAGCCCTG 420
intermediate    AAGCCTTGGAAATGGAGGAAGAAGAAAAGCGAAAAGTTCAAACACACGTCAGCAGCCCTG 144
                ************************************************************
long            GAAAGGAAAATATCTATGAGGCAAAGCAGAGAAGAGCTGATAAAGCGAGGAGTCCTGAAG 480
intermediate    GAAAGGAAAATATCTATGAGGCAAAGCAGAGAAGAGCTGATAAAGCGAGGAGTCCTGAAG 204
                ************************************************************
long            GAAATCTATGATAAAGATGGGGAACTCTCTATATCCAATGAAGAGGACTCCCTAGAAAAT 540
intermediate    GAAATCTATGATAAAGATGGGGAACTCTCTATATCCAATGAAGAGGACTCCCTAGAAAAT 264
                ************************************************************
long            GGGCAGTCCCTGAGCTCCAGCCAGCTGTCTCTGCCTGCCCTGTCCGAAATGGAGCCAGTC 600
intermediate    GGGCAGTCCCTGAGCTCCAGCCAGCTGTCTCTGCCTGCCCTGTCCGAAATGGAGCCAGTC 324
                ************************************************************
long            CCAATGCCCAGGGATCCCTGCTCATATGAGGTGCTCCAACCGTCAGACATCATGGATGGG 660
intermediate    CCAATGCCCAGGGATCCCTGCTCATATGAGGTGCTCCAACCGTCAGACATCATGGATGGG 384
                ************************************************************
long            CCAG------------------------------------------------------- 664
intermediate    CCAGTGTCTGAAGAGAGTCCCTCTGCCAGTGAGTCTGGAGTCCTCCTGTCCCAAGATCCT 444
                ****
long            ------------------------------------------------------------
intermediate    TCAGCCAAACCAGTCCTGCTACTGCCCCCCAAAAAACCTGCTGCTTTCCCTGGAGACCAT 504
long            ------------------------------------------------------------
intermediate    GAAGAGACCCCAGTGAAGCAGCTGCCCCTTCTCAAGCAGCCCCCGGCCCTGCCTCCCAAA 564
long            --------------------------------ATCCTGGCGCCCCCTGTGAAATTGCCTTGT 693
intermediate    CCCACTACCAGGATTGCCAACCACTTAACAGATCCTGGCGCCCCTGTGAAATTGCCTTGT 624
                                               ****************************
long            CTGCCAGTGAAACTGTCGCCTCCGCTACCTCCAAAGAAAGTCATGATCTGTATGCCCGTG 753
intermediate    CTGCCAGTGAAACTGTCGCCTCCGCTACCTCCAAAGAAAGTCATGATCTGTATGCCCGTG 684
                ************************************************************
long            GGGGGGCCAGACCTCTCACTGGTGTCCTACACAGCCCAGAAGAGTGGCCAGCAGGGTGTG 813
intermediate    GGGGGGCCAGACCTCTCACTGGTGTCCTACACAGCCCAGAAGAGTGGCCAGCAGGGTGTG 744
                ************************************************************
long            GCCCAGCACCACCACACTGTCCTGCCCTCCCAGATCCAGCACCAGCTGCAGTACGGCAGC 873
intermediate    GCCCAGCACCACCACACTGTCCTGCCCTCCCAGATCCAGCACCAGCTGCAGTACGGCAGC 804
                ************************************************************
```

```
long            CACGGCCAGCACCTCCCCTCCACCACCGGCTCCCTCCCCATGCACCCCTCGGGCTGCAGA 933
intermediate    CACGGCCAGCACCTCCCCTCCACCACCGGCTCCCTCCCCATGCACCCCTCGGGCTGCAGA 864
                ************************************************************

long            ATGATAGACGAGCTCAACAAAACGCTGGCCATGACCATGCAGAGGCTGGAAAGCTCTGAG 993
intermediate    ATGATAGACGAGCTCAACAAAACGCTGGCCATGACCATGCAGAGGCTGGAAAGCTCTGAG 924
                ************************************************************

long            CAGCGGGTCCCCTGTTCCACTTCTTACCACAGCTCTGGGTTGCACTCGGGTGATGGGGTC 1053
intermediate    CAGCGGGTCCCCTGTTCCACTTCTTACCACAGCTCTGGGTTGCACTCGGGTGATGGGGTC 984
                ************************************************************

long            ACCAAAGCAGGACCTATGGGCCTTCCAGAAATAAGACAAGTGCCAACTGTTGTGATTGAA 1113
intermediate    ACCAAAGCAGGACCTATGGGCCTTCCAGAAATAAGACAAGTGCCAACTGTTGTGATTGAA 1044
                ************************************************************

long            TGTGATGACAATAAAGAAAATGTGCCTCATGAGTCAGACTACGAAGACTCTTCTTGCCTG 1173
intermediate    TGTGATGACAATAAAGAAAATGTGCCTCATGAGTCAGACTACGAAGACTCTTCTTGCCTG 1104
                ************************************************************

long            TATACAAGAGAAGAGGAGGAAGAGGAGGAGGACGAAGACGACGACAGCTCATTATACACC 1233
intermediate    TATACAAGAGAAGAGGAGGAAGAGGAGGAGGACGAAGACGACGACAGCTCATTATACACC 1164
                ************************************************************

long            AGCTCCCTGGCCATGAAGGTCTGCAGGAAGGACTCCTTAGCCATCAAACTCAGCAACAGG 1293
intermediate    AGCTCCCTGGCCATGAAGGTCTGCAGGAAGGACTCCTTAGCCATCAAACTCAGCAACAGG 1224
                ************************************************************

long            CCCTCCAAGCGAGAGCTGGAAGAAAAGAACATCCTTCCCAGGCAGACGGATGAGGAGCGG 1353
intermediate    CCCTCCAAGCGAGAGCTGGAAGAAAAGAACATCCTTCCCAGGCAGACGGATGAGGAGCGG 1284
                ************************************************************

long            CTGGAGCTGAGGCAACAGATTGGCACCAAGCTCACCAGGCGGCTGAGCCAGAGGCCAACT 1413
intermediate    CTGGAGCTGAGGCAACAGATTGGCACCAAGCTCACCAGGCGGCTGAGCCAGAGGCCAACT 1344
                ************************************************************

long            GCAGAGGAACTGGAACAGAGGAACATTTTGAAACCTCGGAATGAACAAGAGGAACAGGAG 1473
intermediate    GCAGAGGAACTGGAACAGAGGAACATTTTGAAACCTCGGAATGAACAAGAGGAACAGGAG 1404
                ************************************************************

long            GAGAAGAGAGAGATCAAGAGGAGGCTAACCCGAAAGCTCAGTCAAAGGCCCACGGTGGAA 1533
intermediate    GAGAAGAGAGAGATCAAGAGGAGGCTAACCCGAAAGCTCAGTCAAAGGCCCACGGTGGAA 1464
                ************************************************************

long            GAGCTTCGGGAAAGAAAGATCCTCATCCGCTTCAGTGACTACGTGGAGGTGGCTGACGCT 1593
intermediate    GAGCTTCGGGAAAGAAAGATCCTCATCCGCTTCAGTGACTACGTGGAGGTGGCTGACGCT 1524
                ************************************************************

long            CAGGACTATGACCGCAGGGCAGATAAGCCGTGGACCCGCCTCACCGCTGCAGACAAAGCT 1653
intermediate    CAGGACTATGACCGCAGGGCAGATAAGCCGTGGACCCGCCTCACCGCTGCAGACAAAGCT 1584
                ************************************************************

long            GCCATCCGAAAGGAGCTCAATGAATTCAAAAGCACTGAGATGGAAGTTCATGAATTGAGT 1713
intermediate    GCCATCCGAAAGGAGCTCAATGAATTCAAAAGCACTGAGATGGAAGTTCATGAATTGAGT 1644
                ************************************************************

long            AGACACTTAACAAGGTTTCACCGACCTTAA 1743
intermediate    AGACACTTAACAAGGTTTCACCGACCTTAA 1674
                ******************************
```

10 Bibliography

1. WHO, *Global status report on noncommunicable diseases 2010*. 2011, Geneva, Switzerland: World Health Organization.
2. Longo, D.L., et al., *Harrison's principles of internal medicine*. 2012, McGraw-Hill,: New York ; London.
3. Foundation., B.H., *Coronary heart disease statistics*. 2012, British Heart Foundation: London. p. v.
4. Pekkanen, J., et al., *Ten-year mortality from cardiovascular disease in relation to cholesterol level among men with and without preexisting cardiovascular disease.* N Engl J Med, 1990. **322**(24): p. 1700-7.
5. Scandinavian Simvastatin Survival Study, G., *Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S).* The Lancet, 1994. **344**(8934): p. 1383-1389.
6. Ruffer, M.A., A. Rietti, and A. Rietti, *Remarks on the histology and pathological anatomy of Egyptian mummies*. 1910, Alexandria: A. Moures. Various pagings.
7. McGill, H.C., Jr., et al., *Origin of atherosclerosis in childhood and adolescence.* Am J Clin Nutr, 2000. **72**(5 Suppl): p. 1307S-1315S.
8. Hopkins, P.N., *Molecular biology of atherosclerosis.* Physiol Rev, 2013. **93**(3): p. 1317-542.
9. Tabas, I., K.J. Williams, and J. Boren, *Subendothelial lipoprotein retention as the initiating process in atherosclerosis: update and therapeutic implications.* Circulation, 2007. **116**(16): p. 1832-44.
10. Jessup, W., L. Kritharides, and R. Stocker, *Lipid oxidation in atherogenesis: an overview.* Biochem Soc Trans, 2004. **32**(Pt 1): p. 134-8.
11. Rudijanto, A., *The role of vascular smooth muscle cells on the pathogenesis of atherosclerosis.* Acta Med Indones, 2007. **39**(2): p. 86-93.
12. Hegyi, L., et al., *Foam cell apoptosis and the development of the lipid core of human atherosclerosis.* J Pathol, 1996. **180**(4): p. 423-9.
13. Silvestre-Roig, C., et al., *Atherosclerotic plaque destabilization: mechanisms, models, and therapeutic strategies.* Circ Res, 2014. **114**(1): p. 214-26.
14. Siess, W., *Athero- and thrombogenic actions of lysophosphatidic acid and sphingosine-1-phosphate.* Biochim Biophys Acta, 2002. **1582**(1-3): p. 204-15.
15. Williams, K.J. and I. Tabas, *The response-to-retention hypothesis of early atherogenesis.* Arterioscler Thromb Vasc Biol, 1995. **15**(5): p. 551-61.
16. Nordestgaard, B.G. and L.B. Nielsen, *Atherosclerosis and arterial influx of lipoproteins.* Curr Opin Lipidol, 1994. **5**(4): p. 252-7.
17. Usman, A., et al., *From Lipid Retention to Immune-Mediate Inflammation and Associated Angiogenesis in the Pathogenesis of Atherosclerosis.* J Atheroscler Thromb, 2015. **22**(8): p. 739-49.
18. Flood, C., et al., *Molecular mechanism for changes in proteoglycan binding on compositional changes of the core and the surface of low-density lipoprotein-containing human apolipoprotein B100.* Arterioscler Thromb Vasc Biol, 2004. **24**(3): p. 564-70.

19. Yoshida, H. and R. Kisugi, *Mechanisms of LDL oxidation.* Clin Chim Acta, 2010. **411**(23-24): p. 1875-82.
20. Podrez, E.A., et al., *Macrophage scavenger receptor CD36 is the major receptor for LDL modified by monocyte-generated reactive nitrogen species.* The Journal of Clinical Investigation, 2000. **105**(8): p. 1095-1108.
21. Mestas, J. and K. Ley, *Monocyte-endothelial cell interactions in the development of atherosclerosis.* Trends Cardiovasc Med, 2008. **18**(6): p. 228-32.
22. Gerszten, R.E., et al., *MCP-1 and IL-8 trigger firm adhesion of monocytes to vascular endothelium under flow conditions.* Nature, 1999. **398**(6729): p. 718-23.
23. Camejo, G., et al., *Binding of low density lipoproteins by proteoglycans synthesized by proliferating and quiescent human arterial smooth muscle cells.* J Biol Chem, 1993. **268**(19): p. 14131-7.
24. Camejo, G., et al., *Interaction of low density lipoproteins with arterial proteoglycans. The role of charge and sialic acid content.* Atherosclerosis, 1985. **55**(1): p. 93-105.
25. Ley, K., Y.I. Miller, and C.C. Hedrick, *Monocyte and macrophage dynamics during atherogenesis.* Arterioscler Thromb Vasc Biol, 2011. **31**(7): p. 1506-16.
26. Sluiter, W., et al., *Leukocyte adhesion molecules on the vascular endothelium: their role in the pathogenesis of cardiovascular disease and the mechanisms underlying their expression.* J Cardiovasc Pharmacol, 1993. **22 Suppl 4**: p. S37-44.
27. Smith, J.D., et al., *Decreased atherosclerosis in mice deficient in both macrophage colony-stimulating factor (op) and apolipoprotein E.* Proc Natl Acad Sci U S A, 1995. **92**(18): p. 8264-8.
28. Rajavashisth, T., et al., *Heterozygous osteopetrotic (op) mutation reduces atherosclerosis in LDL receptor- deficient mice.* J Clin Invest, 1998. **101**(12): p. 2702-10.
29. Weyrich, A.S., et al., *Monocyte tethering by P-selectin regulates monocyte chemotactic protein-1 and tumor necrosis factor-alpha secretion. Signal integration and NF-kappa B translocation.* J Clin Invest, 1995. **95**(5): p. 2297-303.
30. Johnson-Tidey, R.R., et al., *Increase in the adhesion molecule P-selectin in endothelium overlying atherosclerotic plaques. Coexpression with intercellular adhesion molecule-1.* Am J Pathol, 1994. **144**(5): p. 952-61.
31. Gotsch, U., et al., *Expression of P-selectin on endothelial cells is upregulated by LPS and TNF-alpha in vivo.* Cell Adhes Commun, 1994. **2**(1): p. 7-14.
32. Imhof, B.A. and M. Aurrand-Lions, *Adhesion mechanisms regulating the migration of monocytes.* Nat Rev Immunol, 2004. **4**(6): p. 432-44.
33. World, C.J., G. Garin, and B. Berk, *Vascular shear stress and activation of inflammatory genes.* Curr Atheroscler Rep, 2006. **8**(3): p. 240-4.
34. Zand, T., et al., *Lipid deposition in rat aortas with intraluminal hemispherical plug stenosis. A morphological and biophysical study.* Am J Pathol, 1999. **155**(1): p. 85-92.
35. Huo, Y. and K.F. Ley, *Role of platelets in the development of atherosclerosis.* Trends Cardiovasc Med, 2004. **14**(1): p. 18-22.

36. Bobryshev, Y.V., *Monocyte recruitment and foam cell formation in atherosclerosis.* Micron, 2006. **37**(3): p. 208-22.
37. Ostermann, G., et al., *Involvement of JAM-A in mononuclear cell recruitment on inflamed or atherosclerotic endothelium: inhibition by soluble JAM-A.* Arterioscler Thromb Vasc Biol, 2005. **25**(4): p. 729-35.
38. Ostermann, G., et al., *JAM-1 is a ligand of the beta(2) integrin LFA-1 involved in transendothelial migration of leukocytes.* Nat Immunol, 2002. **3**(2): p. 151-8.
39. Sieweke, M.H. and J.E. Allen, *Beyond stem cells: self-renewal of differentiated macrophages.* Science, 2013. **342**(6161): p. 1242974.
40. Robbins, C.S., et al., *Local proliferation dominates lesional macrophage accumulation in atherosclerosis.* Nat Med, 2013. **19**(9): p. 1166-72.
41. Randolph, G.J., *Mechanisms That Regulate Macrophage Burden in Atherosclerosis.* Circulation Research, 2014. **114**(11): p. 1757-1771.
42. Dong, C., et al., *RNA sequencing and transcriptomal analysis of human monocyte to macrophage differentiation.* Gene, 2013. **519**(2): p. 279-87.
43. Kruth, H.S., et al., *Macropinocytosis is the endocytic pathway that mediates macrophage foam cell formation with native low density lipoprotein.* J Biol Chem, 2005. **280**(3): p. 2352-60.
44. Hansson, G.K. and A. Hermansson, *The immune system in atherosclerosis.* Nat Immunol, 2011. **12**(3): p. 204-12.
45. Buechler, C., et al., *Adipophilin is a sensitive marker for lipid loading in human blood monocytes.* Biochim Biophys Acta, 2001. **1532**(1-2): p. 97-104.
46. Stewart, C.R., et al., *CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer.* Nat Immunol, 2010. **11**(2): p. 155-61.
47. Sheedy, F.J., et al., *CD36 coordinates NLRP3 inflammasome activation by facilitating intracellular nucleation of soluble ligands into particulate ligands in sterile inflammation.* Nat Immunol, 2013. **14**(8): p. 812-20.
48. Zhang, H.M., et al., *AnimalTFDB: a comprehensive animal transcription factor database.* Nucleic Acids Res, 2012. **40**(Database issue): p. D144-9.
49. Chawla, A., et al., *A PPAR gamma-LXR-ABCA1 pathway in macrophages is involved in cholesterol efflux and atherogenesis.* Mol Cell, 2001. **7**(1): p. 161-71.
50. Ferre, P., *The biology of peroxisome proliferator-activated receptors: relationship with lipid metabolism and insulin sensitivity.* Diabetes, 2004. **53 Suppl 1**: p. S43-50.
51. Lee, C.H. and R.M. Evans, *Peroxisome proliferator-activated receptor-gamma in macrophage lipid homeostasis.* Trends Endocrinol Metab, 2002. **13**(8): p. 331-5.
52. Janowski, B.A., et al., *An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha.* Nature, 1996. **383**(6602): p. 728-31.
53. Nagy, Z.S., Z. Czimmerer, and L. Nagy, *Nuclear receptor mediated mechanisms of macrophage cholesterol metabolism.* Mol Cell Endocrinol, 2013. **368**(1-2): p. 85-98.
54. Feldmann, R., et al., *Genome-wide analysis of LXRalpha activation reveals new transcriptional networks in human atherosclerotic foam cells.* Nucleic Acids Res, 2013. **41**(6): p. 3518-31.

55. Radhika, A. and P.R. Sudhakaran, *Upregulation of macrophage-specific functions by oxidized LDL: lysosomal degradation-dependent and -independent pathways.* Mol Cell Biochem, 2013. **372**(1-2): p. 181-90.
56. Wang, Y., et al., *NF-kB activity-dependent P-selectin involved in ox-LDL-induced foam cell formation in U937 cell.* Biochem Biophys Res Commun, 2011. **411**(3): p. 543-8.
57. Kobayashi, E., T. Suzuki, and M. Yamamoto, *Roles nrf2 plays in myeloid cells and related disorders.* Oxid Med Cell Longev, 2013. **2013**: p. 529219.
58. Ishii, T., et al., *Role of Nrf2 in the regulation of CD36 and stress protein expression in murine macrophages: activation by oxidatively modified LDL and 4-hydroxynonenal.* Circ Res, 2004. **94**(5): p. 609-16.
59. Ruotsalainen, A.K., et al., *The absence of macrophage Nrf2 promotes early atherogenesis.* Cardiovasc Res, 2013. **98**(1): p. 107-15.
60. Sussan, T.E., et al., *Disruption of Nrf2, a key inducer of antioxidant defenses, attenuates ApoE-mediated atherosclerosis in mice.* PLoS One, 2008. **3**(11): p. e3791.
61. Moreno, P.R., et al., *Macrophage infiltration in acute coronary syndromes. Implications for plaque rupture.* Circulation, 1994. **90**(2): p. 775-8.
62. Galkina, E. and K. Ley, *Immune and inflammatory mechanisms of atherosclerosis (*).* Annu Rev Immunol, 2009. **27**: p. 165-97.
63. Boyle, J.J., P.L. Weissberg, and M.R. Bennett, *Tumor necrosis factor-alpha promotes macrophage-induced vascular smooth muscle cell apoptosis by direct and autocrine mechanisms.* Arterioscler Thromb Vasc Biol, 2003. **23**(9): p. 1553-8.
64. Clarke, M.C., et al., *Apoptosis of vascular smooth muscle cells induces features of plaque vulnerability in atherosclerosis.* Nat Med, 2006. **12**(9): p. 1075-80.
65. Kolodgie, F.D., et al., *Localization of apoptotic macrophages at the site of plaque rupture in sudden coronary death.* Am J Pathol, 2000. **157**(4): p. 1259-68.
66. Schrijvers, D.M., et al., *Phagocytosis of apoptotic cells by macrophages is impaired in atherosclerosis.* Arterioscler Thromb Vasc Biol, 2005. **25**(6): p. 1256-61.
67. Gough, P.J., et al., *Macrophage expression of active MMP-9 induces acute plaque disruption in apoE-deficient mice.* J Clin Invest, 2006. **116**(1): p. 59-69.
68. Wang, N., et al., *Interleukin 8 Is Induced by Cholesterol Loading of Macrophages and Expressed by Macrophage Foam Cells in Human Atheroma.* Journal of Biological Chemistry, 1996. **271**(15): p. 8837-8842.
69. Bekkering, S., et al., *Oxidized low-density lipoprotein induces long-term proinflammatory cytokine production and foam cell formation via epigenetic reprogramming of monocytes.* Arterioscler Thromb Vasc Biol, 2014. **34**(8): p. 1731-8.
70. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.
71. Wilson, P.W., *Established risk factors and coronary artery disease: the Framingham Study.* Am J Hypertens, 1994. **7**(7 Pt 2): p. 7S-12S.
72. Csordas, A. and D. Bernhard, *The biology behind the atherothrombotic effects of cigarette smoke.* Nat Rev Cardiol, 2013. **10**(4): p. 219-30.

73. McVeigh, G.E., et al., *Vascular abnormalities associated with long-term cigarette smoking identified by arterial waveform analysis.* Am J Med, 1997. **102**(3): p. 227-31.
74. Levine, P.H., *An acute effect of cigarette smoking on platelet function. A possible link between smoking and arterial thrombosis.* Circulation, 1973. **48**(3): p. 619-23.
75. Barua, R.S., et al., *Reactive oxygen species are involved in smoking-induced dysfunction of nitric oxide biosynthesis and upregulation of endothelial nitric oxide synthase: an in vitro demonstration in human coronary artery endothelial cells.* Circulation, 2003. **107**(18): p. 2342-7.
76. Collins, T., *Endothelial nuclear factor-kappa B and the initiation of the atherosclerotic lesion.* Lab Invest, 1993. **68**(5): p. 499-508.
77. Cacciola, R.R., F. Guarino, and R. Polosa, *Relevance of endothelial-haemostatic dysfunction in cigarette smoking.* Curr Med Chem, 2007. **14**(17): p. 1887-92.
78. Wissler, R.W., *New insights into the pathogenesis of atherosclerosis as revealed by PDAY. Pathobiological Determinants of Atherosclerosis in Youth.* Atherosclerosis, 1994. **108 Suppl**: p. S3-20.
79. Botti, T.P., et al., *A comparison of the quantitation of macrophage foam cell populations and the extent of apolipoprotein E deposition in developing atherosclerotic lesions in young people: high and low serum thiocyanate groups as an indication of smoking. PDAY Research Group. Pathobiological Determinants of Atherosclerosis in Youth.* Atherosclerosis, 1996. **124**(2): p. 191-202.
80. Kangavari, S., et al., *Smoking increases inflammation and metalloproteinase expression in human carotid atherosclerotic plaques.* J Cardiovasc Pharmacol Ther, 2004. **9**(4): p. 291-8.
81. Yamada, S., et al., *Matrix metalloproteinase 12 accelerates the initiation of atherosclerosis and stimulates the progression of fatty streaks to fibrous plaques in transgenic rabbits.* Am J Pathol, 2008. **172**(5): p. 1419-29.
82. O'Toole, T.E., et al., *Acrolein activates matrix metalloproteinases by increasing reactive oxygen species in macrophages.* Toxicol Appl Pharmacol, 2009. **236**(2): p. 194-201.
83. Bonnans, C., J. Chou, and Z. Werb, *Remodelling the extracellular matrix in development and disease.* Nat Rev Mol Cell Biol, 2014. **15**(12): p. 786-801.
84. Raveendran, M., et al., *Cigarette suppresses the expression of P4Halpha and vascular collagen production.* Biochem Biophys Res Commun, 2004. **323**(2): p. 592-8.
85. Zhang, K., et al., *Interleukin 6 destabilizes atherosclerotic plaques by downregulating prolyl-4-hydroxylase alpha1 via a mitogen-activated protein kinase and c-Jun pathway.* Arch Biochem Biophys, 2012. **528**(2): p. 127-33.
86. Egan, B.M., Y. Zhao, and R.N. Axon, *US trends in prevalence, awareness, treatment, and control of hypertension, 1988-2008.* JAMA, 2010. **303**(20): p. 2043-50.
87. Rapsomaniki, E., et al., *Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1.25 million people.* Lancet, 2014. **383**(9932): p. 1899-911.
88. Lawes, C.M., et al., *Global burden of blood-pressure-related disease, 2001.* Lancet, 2008. **371**(9623): p. 1513-8.

89. Blood Pressure Lowering Treatment Trialists, C., et al., *Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials.* BMJ, 2013. **347**: p. f5680.
90. Alexander, R.W., *Theodore Cooper Memorial Lecture. Hypertension and the pathogenesis of atherosclerosis. Oxidative stress and the mediation of arterial inflammatory response: a new perspective.* Hypertension, 1995. **25**(2): p. 155-61.
91. Benjamin, E.J., et al., *Clinical correlates and heritability of flow-mediated dilation in the community: the Framingham Heart Study.* Circulation, 2004. **109**(5): p. 613-9.
92. Thuillez, C. and V. Richard, *Targeting endothelial dysfunction in hypertensive subjects.* J Hum Hypertens, 2005. **19 Suppl 1**: p. S21-5.
93. Ross, R., J. Glomset, and L. Harker, *Response to injury and atherogenesis.* Am J Pathol, 1977. **86**(3): p. 675-84.
94. Wilson, P.W., *Diabetes mellitus and coronary heart disease.* Am J Kidney Dis, 1998. **32**(5 Suppl 3): p. S89-100.
95. Kannel, W.B. and D.L. McGee, *Diabetes and cardiovascular risk factors: the Framingham study.* Circulation, 1979. **59**(1): p. 8-13.
96. Singer, D.E., et al., *Association of HbA1c with prevalent cardiovascular disease in the original cohort of the Framingham Heart Study.* Diabetes, 1992. **41**(2): p. 202-8.
97. Almdal, T., et al., *The independent effect of type 2 diabetes mellitus on ischemic heart disease, stroke, and death: a population-based study of 13,000 men and women with 20 years of follow-up.* Arch Intern Med, 2004. **164**(13): p. 1422-6.
98. Stettler, C., et al., *Glycemic control and macrovascular disease in types 1 and 2 diabetes mellitus: Meta-analysis of randomized trials.* Am Heart J, 2006. **152**(1): p. 27-38.
99. Gleissner, C.A., et al., *Mechanisms by which diabetes increases cardiovascular disease.* Drug Discov Today Dis Mech, 2007. **4**(3): p. 131-140.
100. Shashkin, P.N., et al., *Insulin and glucose play a role in foam cell formation and function.* Cardiovasc Diabetol, 2006. **5**: p. 13.
101. Go, A.S., et al., *Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization.* N Engl J Med, 2004. **351**(13): p. 1296-305.
102. Solomon, S.D., et al., *Influence of albuminuria on cardiovascular risk in patients with stable coronary artery disease.* Circulation, 2007. **116**(23): p. 2687-93.
103. Foley, R.N., P.S. Parfrey, and M.J. Sarnak, *Epidemiology of cardiovascular disease in chronic renal disease.* J Am Soc Nephrol, 1998. **9**(12 Suppl): p. S16-23.
104. Ito, S. and M. Yoshida, *Protein-bound uremic toxins: new culprits of cardiovascular events in chronic kidney disease patients.* Toxins (Basel), 2014. **6**(2): p. 665-78.
105. Chmielewski, M., et al., *Expression of scavenger receptor CD36 in chronic renal failure patients.* Artif Organs, 2005. **29**(8): p. 608-14.

106. Suganuma, E., et al., *Antiatherogenic effects of angiotensin receptor antagonism in mild renal dysfunction.* J Am Soc Nephrol, 2006. **17**(2): p. 433-41.
107. Bro, S., et al., *Chronic renal failure accelerates atherogenesis in apolipoprotein E-deficient mice.* J Am Soc Nephrol, 2003. **14**(10): p. 2466-74.
108. Massy, Z.A., et al., *Uremia accelerates both atherosclerosis and arterial calcification in apolipoprotein E knockout mice.* J Am Soc Nephrol, 2005. **16**(1): p. 109-16.
109. Itoh, Y., et al., *Protein-bound uremic toxins in hemodialysis patients measured by liquid chromatography/tandem mass spectrometry and their effects on endothelial ROS production.* Anal Bioanal Chem, 2012. **403**(7): p. 1841-50.
110. Tumur, Z., et al., *Indoxyl sulfate upregulates expression of ICAM-1 and MCP-1 by oxidative stress-induced NF-kappaB activation.* Am J Nephrol, 2010. **31**(5): p. 435-41.
111. Adesso, S., et al., *The uremic toxin indoxyl sulphate enhances macrophage response to LPS.* PLoS One, 2013. **8**(9): p. e76778.
112. *Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S).* Lancet, 1994. **344**(8934): p. 1383-9.
113. Navab, M., et al., *Oxidized lipids as mediators of coronary heart disease.* Curr Opin Lipidol, 2002. **13**(4): p. 363-72.
114. Sayols-Baixeras, S., et al., *Pathogenesis of coronary artery disease: focus on genetic risk factors and identification of genetic variants.* Appl Clin Genet, 2014. **7**: p. 15-32.
115. Marenberg, M.E., et al., *Genetic susceptibility to death from coronary heart disease in a study of twins.* N Engl J Med, 1994. **330**(15): p. 1041-6.
116. Zdravkovic, S., et al., *Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins.* J Intern Med, 2002. **252**(3): p. 247-54.
117. Nora, J.J., et al., *Genetic--epidemiologic study of early-onset ischemic heart disease.* Circulation, 1980. **61**(3): p. 503-8.
118. Helgadottir, A., et al., *The gene encoding 5-lipoxygenase activating protein confers risk of myocardial infarction and stroke.* Nat Genet, 2004. **36**(3): p. 233-9.
119. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
120. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
121. Scheet, P. and M. Stephens, *A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase.* Am J Hum Genet, 2006. **78**(4): p. 629-44.
122. Marchini, J. and B. Howie, *Genotype imputation for genome-wide association studies.* Nat Rev Genet, 2010. **11**(7): p. 499-511.
123. Consortium, C.A.D., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nat Genet, 2013. **45**(1): p. 25-33.
124. Myocardial Infarction Genetics, C., et al., *Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants.* Nat Genet, 2009. **41**(3): p. 334-41.

125. Roberts, R., *Genetics of coronary artery disease.* Circ Res, 2014. **114**(12): p. 1890-903.
126. Bjorkegren, J.L., et al., *Genome-wide significant loci: how important are they? Systems genetics to understand heritability of coronary artery disease and other common complex disorders.* J Am Coll Cardiol, 2015. **65**(8): p. 830-45.
127. Makinen, V.P., et al., *Integrative genomics reveals novel molecular pathways and gene networks for coronary artery disease.* PLoS Genet, 2014. **10**(7): p. e1004502.
128. Maurano, M.T., et al., *Systematic Localization of Common Disease-Associated Variation in Regulatory DNA.* Science, 2012. **337**(6099): p. 1190-1195.
129. Karczewski, K.J., et al., *Systematic functional regulatory assessment of disease-associated variants.* Proc Natl Acad Sci U S A, 2013. **110**(23): p. 9607-12.
130. Foroughi Asl, H., et al., *Expression quantitative trait Loci acting across multiple tissues are enriched in inherited risk for coronary artery disease.* Circ Cardiovasc Genet, 2015. **8**(2): p. 305-15.
131. Musunuru, K., et al., *From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus.* Nature, 2010. **466**(7307): p. 714-9.
132. Harismendy, O., et al., *9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response.* Nature, 2011. **470**(7333): p. 264-8.
133. Holdt, L.M., et al., *Alu Elements in ANRIL Non-Coding RNA at Chromosome 9p21 Modulate Atherogenic Cell Functions through Trans-Regulation of Gene Networks.* PLoS Genet, 2013. **9**(7): p. e1003588.
134. Miller, C.L., et al., *Disease-Related Growth Factor and Embryonic Signaling Pathways Modulate an Enhancer of TCF21 Expression at the 6q23.2 Coronary Heart Disease Locus.* PLoS Genet, 2013. **9**(7): p. e1003652.
135. Miller, C.L., et al., *Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation.* PLoS Genet, 2014. **10**(3): p. e1004263.
136. Fairfax, B.P., et al., *Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression.* Science, 2014. **343**(6175): p. 1246949.
137. Hardison, R.C., *Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies.* J Biol Chem, 2012. **287**(37): p. 30932-40.
138. Havel, R.J., H.A. Eder, and J.H. Bragdon, *The distribution and chemical composition of ultracentrifugally separated lipoproteins in human serum.* The Journal of clinical investigation, 1955. **34**(9): p. 1345-53.
139. Simon, J.M., et al., *Using formaldehyde-assisted isolation of regulatory elements (FAIRE) to isolate active regulatory DNA.* Nat Protoc, 2012. **7**(2): p. 256-67.
140. Lunter, G. and M. Goodson, *Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads.* Genome Res, 2011. **21**(6): p. 936-9.
141. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
142. Boyle, A.P., et al., *F-Seq: a feature density estimator for high-throughput sequence tags.* Bioinformatics, 2008. **24**(21): p. 2537-8.

143. Consortium, E.P., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
144. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
145. Shen, L., et al., *diffReps: detecting differential chromatin modification sites from ChIP-seq data with biological replicates.* PLoS One, 2013. **8**(6): p. e65598.
146. Hnisz, D., et al., *Super-enhancers in the control of cell identity and disease.* Cell, 2013. **155**(4): p. 934-47.
147. Pham, T.-H., et al., *Dynamic epigenetic enhancer signatures reveal key transcription factors associated with monocytic differentiation states.* Blood, 2012. **119**(24): p. e161-e171.
148. Shin, H., et al., *CEAS: cis-regulatory element annotation system.* Bioinformatics, 2009. **25**(19): p. 2605-6.
149. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
150. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.
151. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
152. Liu, T., et al., *Cistrome: an integrative platform for transcriptional regulation studies.* Genome Biol, 2011. **12**(8): p. R83.
153. Thomas-Chollier, M., et al., *Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs.* Nat Protoc, 2011. **6**(12): p. 1860-9.
154. Ward, L.D. and M. Kellis, *HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.* Nucleic Acids Res, 2012. **40**(Database issue): p. D930-4.
155. Schunkert, H., et al., *Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.* Nature genetics, 2011. **43**(4): p. 333-8.
156. Samani, N.J., et al., *Genomewide association analysis of coronary artery disease.* N Engl J Med, 2007. **357**(5): p. 443-53.
157. Erdmann, J., et al., *Genome-wide association study identifies a new locus for coronary artery disease on chromosome 10p11.23.* Eur Heart J, 2011. **32**(2): p. 158-68.
158. Wild, P.S., et al., *A genome-wide association study identifies LIPA as a susceptibility gene for coronary artery disease.* Circ Cardiovasc Genet, 2011. **4**(4): p. 403-12.
159. O'Donnell, C.J., et al., *Genome-Wide Association Study for Coronary Artery Calcification With Follow-Up in Myocardial Infarction.* Circulation, 2011. **124**(25): p. 2855-2864.
160. Du, P., W.A. Kibbe, and S.M. Lin, *lumi: a pipeline for processing Illumina microarray.* Bioinformatics, 2008. **24**(13): p. 1547-8.
161. Lin, D., et al., *NF-kappaB regulates MICA gene transcription in endothelial cell through a genetically inhibitable control site.* J Biol Chem, 2012. **287**(6): p. 4299-310.

162. Sciorra, V.A. and A.J. Morris, *Sequential actions of phospholipase D and phosphatidic acid phosphohydrolase 2b generate diglyceride in mammalian cells.* Mol Biol Cell, 1999. **10**(11): p. 3863-76.
163. Escalante-Alcalde, D., et al., *The lipid phosphatase LPP3 regulates extra-embryonic vasculogenesis and axis patterning.* Development, 2003. **130**(19): p. 4623-37.
164. Lopez-Juarez, A., et al., *Expression of LPP3 in Bergmann glia is required for proper cerebellar sphingosine-1-phosphate metabolism/signaling and development.* Glia, 2011. **59**(4): p. 577-89.
165. Panchatcharam, M., et al., *Lipid phosphate phosphatase 3 negatively regulates smooth muscle cell phenotypic modulation to limit intimal hyperplasia.* Arterioscler Thromb Vasc Biol, 2013. **33**(1): p. 52-9.
166. Salous, A.K., et al., *Mechanism of rapid elimination of lysophosphatidic acid and related lipids from the circulation of mice.* J Lipid Res, 2013. **54**(10): p. 2775-84.
167. Federico, L., et al., *Autotaxin and its product lysophosphatidic acid suppress brown adipose differentiation and promote diet-induced obesity in mice.* Mol Endocrinol, 2012. **26**(5): p. 786-97.
168. Dillen, L., et al., *Comparison of triple quadrupole and high-resolution TOF-MS for quantification of peptides.* Bioanalysis, 2012. **4**(5): p. 565-79.
169. Humtsoe, J.O., et al., *Lipid phosphate phosphatase 3 stabilization of beta-catenin induces endothelial cell migration and formation of branching point structures.* Molecular and cellular biology, 2010. **30**(7): p. 1593-606.
170. Gerrity, R.G., *The role of the monocyte in atherogenesis: I. Transition of blood-borne monocytes into foam cells in fatty lesions.* Am J Pathol, 1981. **103**(2): p. 181-90.
171. Stary, H.C., et al., *A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association.* Arterioscler Thromb Vasc Biol, 1995. **15**(9): p. 1512-31.
172. Glass, C.K. and J.L. Witztum, *Atherosclerosis. the road ahead.* Cell, 2001. **104**(4): p. 503-16.
173. Moore, K.J., F.J. Sheedy, and E.A. Fisher, *Macrophages in atherosclerosis: a dynamic balance.* Nat Rev Immunol, 2013. **13**(10): p. 709-21.
174. Palinski, W., et al., *Low density lipoprotein undergoes oxidative modification in vivo.* Proc Natl Acad Sci U S A, 1989. **86**(4): p. 1372-6.
175. Wang, N., et al., *Interleukin 8 is induced by cholesterol loading of macrophages and expressed by macrophage foam cells in human atheroma.* J Biol Chem, 1996. **271**(15): p. 8837-42.
176. Ball, R.Y., et al., *Evidence that the death of macrophage foam cells contributes to the lipid core of atheroma.* Atherosclerosis, 1995. **114**(1): p. 45-54.
177. Kwiterovich, P.O., Jr., *The metabolic pathways of high-density lipoprotein, low-density lipoprotein, and triglycerides: a current review.* Am J Cardiol, 2000. **86**(12A): p. 5L-10L.
178. Quehenberger, O., et al., *Oxidation of human low density lipoprotein initiated by copper (II) chloride.* Basic Life Sci, 1988. **49**: p. 387-90.
179. Cho, H.J., et al., *Induction of dendritic cell-like phenotype in macrophages during foam cell formation.* Physiol Genomics, 2007. **29**(2): p. 149-60.

180. Miller, Y.I., et al., *Toll-like receptor 4-dependent and -independent cytokine secretion induced by minimally oxidized low-density lipoprotein in macrophages.* Arterioscler Thromb Vasc Biol, 2005. **25**(6): p. 1213-9.
181. Hagg, D.A., et al., *Expression profiling of macrophages from subjects with atherosclerosis to identify novel susceptibility genes.* Int J Mol Med, 2008. **21**(6): p. 697-704.
182. Hirose, K., et al., *Different responses to oxidized low-density lipoproteins in human polarized macrophages.* Lipids in health and disease, 2011. **10**: p. 1.
183. Nagy, L., et al., *Oxidized LDL regulates macrophage gene expression through ligand activation of PPARgamma.* Cell, 1998. **93**(2): p. 229-40.
184. Ivanova, E.A., et al., *Peroxisome proliferator-activated receptor (PPAR) gamma in cardiovascular disorders and cardiovascular surgery.* J Cardiol, 2015. **66**(4): p. 271-8.
185. van Tits, L.J., et al., *Oxidized LDL enhances pro-inflammatory responses of alternatively activated M2 macrophages: a crucial role for Kruppel-like factor 2.* Atherosclerosis, 2011. **214**(2): p. 345-9.
186. Kojima, C., et al., *MMP-9 inhibition by ACE inhibitor reduces oxidized LDL-mediated foam-cell formation.* J Atheroscler Thromb, 2010. **17**(1): p. 97-105.
187. Sung, H.J., et al., *N-acetyl cysteine suppresses the foam cell formation that is induced by oxidized low density lipoprotein via regulation of gene expression.* Mol Biol Rep, 2012. **39**(3): p. 3001-7.
188. Ferreira, V., et al., *Macrophage-specific inhibition of NF-kappaB activation reduces foam-cell formation.* Atherosclerosis, 2007. **192**(2): p. 283-90.
189. Qin, Z., *The use of THP-1 cells as a model for mimicking the function and regulation of monocytes and macrophages in the vasculature.* Atherosclerosis, 2012. **221**(1): p. 2-11.
190. Tugal, D., X. Liao, and M.K. Jain, *Transcriptional Control of Macrophage Polarization.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2013. **33**(6): p. 1135-1144.
191. Mosser, D.M. and J.P. Edwards, *Exploring the full spectrum of macrophage activation.* Nat Rev Immunol, 2008. **8**(12): p. 958-969.
192. Stoger, J.L., et al., *Distribution of macrophage polarization markers in human atherosclerosis.* Atherosclerosis, 2012. **225**(2): p. 461-8.
193. Park, Y.M., *CD36, a scavenger receptor implicated in atherosclerosis.* Exp Mol Med, 2014. **46**: p. e99.
194. Beyer, M., et al., *High-resolution transcriptome of human macrophages.* PLoS One, 2012. **7**(9): p. e45466.
195. Norsworthy, P.J., et al., *Murine CD93 (C1qRp) contributes to the removal of apoptotic cells in vivo but is not required for C1q-mediated enhancement of phagocytosis.* J Immunol, 2004. **172**(6): p. 3406-14.
196. Steinberger, P., et al., *Identification of human CD93 as the phagocytic C1q receptor (C1qRp) by expression cloning.* J Leukoc Biol, 2002. **71**(1): p. 133-40.
197. Malarstig, A., et al., *Plasma CD93 concentration is a potential novel biomarker for coronary artery disease.* J Intern Med, 2011. **270**(3): p. 229-36.
198. Youn, J.C., et al., *Soluble CD93 levels in patients with acute myocardial infarction and its implication on clinical outcome.* PLoS One, 2014. **9**(5): p. e96538.

199. Gao, X. and Z. Xu, *Mechanisms of action of angiogenin.* Acta Biochim Biophys Sin (Shanghai), 2008. **40**(7): p. 619-24.
200. Krecki, R., et al., *Relationship of serum angiogenin, adiponectin and resistin levels with biochemical risk factors and the angiographic severity of three-vessel coronary disease.* Cardiol J, 2010. **17**(6): p. 599-606.
201. Wang, L.J., et al., *Expression of heme oxygenase-1 in atherosclerotic lesions.* Am J Pathol, 1998. **152**(3): p. 711-20.
202. Goldstein, J.L. and M.S. Brown, *The LDL receptor.* Arterioscler Thromb Vasc Biol, 2009. **29**(4): p. 431-8.
203. Collot-Teixeira, S., et al., *CD36 and macrophages in atherosclerosis.* Cardiovasc Res, 2007. **75**(3): p. 468-77.
204. Huang da, W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nat Protoc, 2009. **4**(1): p. 44-57.
205. Allen, P.B., et al., *Phactrs 1-4: A family of protein phosphatase 1 and actin regulatory proteins.* Proc Natl Acad Sci U S A, 2004. **101**(18): p. 7187-92.
206. Sung, H.J., et al., *N-acetyl cysteine suppresses the foam cell formation that is induced by oxidized low density lipoprotein via regulation of gene expression.* Molecular Biology Reports, 2011.
207. Wang, Y., et al., *NF-kB activity-dependent P-selectin involved in ox-LDL-induced foam cell formation in U937 cell.* Biochemical and Biophysical Research Communications, 2011. **411**(3): p. 543-548.
208. Jiang, G., et al., *RNA interference for HIF-1alpha inhibits foam cells formation in vitro.* Eur J Pharmacol, 2007. **562**(3): p. 183-90.
209. Wiesner, P., et al., *Low Doses of Lipopolysaccharide and Minimally Oxidized Low-Density Lipoprotein Cooperatively Activate Macrophages via Nuclear Factor κB and Activator Protein-1.* Circulation Research, 2010. **107**(1): p. 56-65.
210. Andersson, T., et al., *Novel candidate genes for atherosclerosis are identified by representational difference analysis-based transcript profiling of cholesterol-loaded macrophages.* Pathobiology, 2001. **69**(6): p. 304-14.
211. Scholz, H., et al., *Enhanced plasma levels of LIGHT in unstable angina: possible pathogenic role in foam cell formation and thrombosis.* Circulation, 2005. **112**(14): p. 2121-9.
212. Hung, Y.C., M.Y. Hong, and G.S. Huang, *Cholesterol loading augments oxidative stress in macrophages.* FEBS letters, 2006. **580**(3): p. 849-61.
213. Levula, M., et al., *Effects of oxidized low‐ and high‐density lipoproteins on gene expression of human macrophages.* Scandinavian Journal of Clinical & Laboratory Investigation, 2006. **66**(6): p. 497-508.
214. Leitinger, N. and I.G. Schulman, *Phenotypic Polarization of Macrophages in Atherosclerosis.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2013. **33**(6): p. 1120-1126.
215. Feig, J.E., et al., *Regression of atherosclerosis is characterized by broad changes in the plaque macrophage transcriptome.* PLoS One, 2012. **7**(6): p. e39790.
216. Consortium, E.P., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

217. Song, L., et al., *Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity.* Genome research, 2011. **21**(10): p. 1757-67.
218. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-268.
219. Biddie, S.C., et al., *Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding.* Mol Cell, 2011. **43**(1): p. 145-55.
220. Gaulton, K.J., et al., *A map of open chromatin in human pancreatic islets.* Nat Genet, 2010. **42**(3): p. 255-9.
221. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
222. Pasquali, L., et al., *Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants.* Nat Genet, 2014.
223. Ward, L.D. and M. Kellis, *Interpreting noncoding genetic variation in complex traits and human disease.* Nat Biotechnol, 2012. **30**(11): p. 1095-106.
224. Trynka, G., et al., *Chromatin marks identify critical cell types for fine mapping complex trait variants.* Nat Genet, 2012.
225. Wierda, R.J., et al., *Epigenetics in atherosclerosis and inflammation.* Journal of Cellular and Molecular Medicine, 2010. **14**(6a): p. 1225-1240.
226. Lenhard, B., A. Sandelin, and P. Carninci, *Metazoan promoters: emerging characteristics and insights into transcriptional regulation.* Nat Rev Genet, 2012. **13**(4): p. 233-45.
227. Sakabe, N.J., D. Savic, and M.A. Nobrega, *Transcriptional enhancers in development and disease.* Genome Biol, 2012. **13**(1): p. 238.
228. Visel, A., E.M. Rubin, and L.A. Pennacchio, *Genomic views of distant-acting enhancers.* Nature, 2009. **461**(7261): p. 199-205.
229. Dailey, L., *High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics.* Genomics, 2015. **106**(3): p. 151-8.
230. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.
231. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.* Nat Genet, 2011. **43**(3): p. 264-8.
232. Magnani, L., et al., *PBX1 genomic pioneer function drives ERalpha signaling underlying progression in breast cancer.* PLoS Genet, 2011. **7**(11): p. e1002368.
233. Sharma, P., et al., *Detection of altered global DNA methylation in coronary artery disease patients.* DNA Cell Biol, 2008. **27**(7): p. 357-65.
234. Stenvinkel, P., et al., *Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease?* J Intern Med, 2007. **261**(5): p. 488-99.
235. Choi, J.H., et al., *Trichostatin A exacerbates atherosclerosis in low density lipoprotein receptor-deficient mice.* Arterioscler Thromb Vasc Biol, 2005. **25**(11): p. 2404-9.
236. Dunn, J., S. Thabet, and H. Jo, *Flow-Dependent Epigenetic DNA Methylation in Endothelial Gene Expression and Atherosclerosis.* Arterioscler Thromb Vasc Biol, 2015. **35**(7): p. 1562-9.

237. Stranger, B.E., E.A. Stahl, and T. Raj, *Progress and promise of genome-wide association studies for human complex trait genetics.* Genetics, 2011. **187**(2): p. 367-83.
238. Horikoshi, M., et al., *Discovery and Fine-Mapping of Glycaemic and Obesity-Related Trait Loci Using High-Density Imputation.* PLoS Genet, 2015. **11**(7): p. e1005230.
239. Trynka, G., et al., *Chromatin marks identify critical cell types for fine mapping complex trait variants.* Nat Genet, 2013. **45**(2): p. 124-30.
240. Farh, K.K., et al., *Genetic and epigenetic fine mapping of causal autoimmune disease variants.* Nature, 2015. **518**(7539): p. 337-43.
241. Giresi, P.G. and J.D. Lieb, *Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements).* Methods, 2009. **48**(3): p. 233-9.
242. Burd, C.J., et al., *Analysis of chromatin dynamics during glucocorticoid receptor activation.* Mol Cell Biol, 2012. **32**(10): p. 1805-17.
243. Song, L. and G.E. Crawford, *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells.* Cold Spring Harb Protoc, 2010. **2010**(2): p. pdb prot5384.
244. Nord, A.S., et al., *Rapid and pervasive changes in genome-wide enhancer usage during mammalian development.* Cell, 2013. **155**(7): p. 1521-31.
245. Zhang, B., et al., *A dynamic H3K27ac signature identifies VEGFA-stimulated endothelial enhancers and requires EP300 activity.* Genome Res, 2013. **23**(6): p. 917-27.
246. Reschen, M.E., et al., *Lipid-Induced Epigenomic Changes in Human Macrophages Identify a Coronary Artery Disease-Associated Variant that Regulates PPAP2B Expression through Altered C/EBP-Beta Binding.* PLoS Genet, 2015. **11**(4): p. e1005061.
247. Hnisz, D., et al., *Super-Enhancers in the Control of Cell Identity and Disease.* Cell, 2013. **155**(0): p. 934-47.
248. Nerlov, C. and T. Graf, *PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors.* Genes Dev, 1998. **12**(15): p. 2403-12.
249. Huber, R., et al., *Regulation of C/EBPbeta and resulting functions in cells of the monocytic lineage.* Cell Signal, 2012. **24**(6): p. 1287-96.
250. Collaboration, I.R.G.C.E.R.F., et al., *Interleukin-6 receptor pathways in coronary heart disease: a collaborative meta-analysis of 82 studies.* Lancet, 2012. **379**(9822): p. 1205-13.
251. Ferreira, R.C., et al., *Functional IL6R 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases.* PLoS Genet, 2013. **9**(4): p. e1003444.
252. Stamler, J., D. Wentworth, and J.D. Neaton, *Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT).* JAMA, 1986. **256**(20): p. 2823-8.
253. Getz, G.S. and C.A. Reardon, *Animal models of atherosclerosis.* Arterioscler Thromb Vasc Biol, 2012. **32**(5): p. 1104-15.
254. Moore, K.J. and I. Tabas, *Macrophages in the pathogenesis of atherosclerosis.* Cell, 2011. **145**(3): p. 341-55.

255. Nammo, T., S.A. Rodriguez-Segui, and J. Ferrer, *Mapping open chromatin with formaldehyde-assisted isolation of regulatory elements.* Methods in molecular biology, 2011. **791**: p. 287-96.
256. Paul, D.S., et al., *Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci.* Genome Res, 2013. **23**(7): p. 1130-41.
257. Waki, H., et al., *Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation.* PLoS Genet, 2011. **7**(10): p. e1002311.
258. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.
259. Siersbaek, R., et al., *Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis.* EMBO J, 2011. **30**(8): p. 1459-72.
260. Rhie, S.K., et al., *Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells.* BMC Genomics, 2014. **15**(1): p. 331.
261. Nagy, L., et al., *Nuclear hormone receptors enable macrophages and dendritic cells to sense their lipid environment and shape their immune response.* Physiol Rev, 2012. **92**(2): p. 739-89.
262. Gupte, A.A., C.J. Lyon, and W.A. Hsueh, *Nuclear factor (erythroid-derived 2)-like-2 factor (Nrf2), a key regulator of the antioxidant response to protect against atherosclerosis and nonalcoholic steatohepatitis.* Curr Diab Rep, 2013. **13**(3): p. 362-71.
263. Rahman, S.M., et al., *CCAAT/enhancer-binding protein beta (C/EBPbeta) expression regulates dietary-induced inflammation in macrophages and adipose tissue in mice.* J Biol Chem, 2012. **287**(41): p. 34349-60.
264. Osada, S., et al., *DNA Binding Specificity of the CCAAT/Enhancer-binding Protein Transcription Factor Family.* Journal of Biological Chemistry, 1996. **271**(7): p. 3891-3896.
265. Ramji, D.P. and P. Foka, *CCAAT/enhancer-binding proteins: structure, function and regulation.* Biochem J, 2002. **365**(Pt 3): p. 561-75.
266. Jakobsen, J.S., et al., *Temporal mapping of CEBPA and CEBPB binding during liver regeneration reveals dynamic occupancy and specific regulatory codes for homeostatic and cell cycle gene batteries.* Genome Res, 2013. **23**(4): p. 592-603.
267. Park, B.H., L. Qiang, and S.R. Farmer, *Phosphorylation of C/EBPbeta at a consensus extracellular signal-regulated kinase/glycogen synthase kinase 3 site is required for the induction of adiponectin gene expression during the differentiation of mouse fibroblasts into adipocytes.* Mol Cell Biol, 2004. **24**(19): p. 8671-80.
268. Piwien-Pilipuk, G., O. MacDougald, and J. Schwartz, *Dual regulation of phosphorylation and dephosphorylation of C/EBPbeta modulate its transcriptional activation and DNA binding in response to growth hormone.* J Biol Chem, 2002. **277**(46): p. 44557-65.
269. Chien, M.W., et al., *OxLDL induces mitogen-activated protein kinase activation mediated via PI3-kinase/Akt in vascular smooth muscle cells.* J Lipid Res, 2003. **44**(9): p. 1667-75.

270. Coronary Artery Disease Genetics, C., *A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease.* Nat Genet, 2011. **43**(4): p. 339-44.
271. Kai, M., et al., *Cloning and characterization of two human isozymes of Mg2+-independent phosphatidic acid phosphatase.* The Journal of biological chemistry, 1997. **272**(39): p. 24572-8.
272. Wary, K.K. and J.O. Humtsoe, *Anti-lipid phosphate phosphohydrolase-3 (LPP3) antibody inhibits bFGF- and VEGF-induced capillary morphogenesis of endothelial cells.* Cell communication and signaling : CCS, 2005. **3**: p. 9.
273. Bianchini, L., et al., *Identification of PPAP2B as a novel recurrent translocation partner gene of HMGA2 in lipomas.* Genes Chromosomes Cancer, 2013. **52**(6): p. 580-90.
274. Hata, M., et al., *Expression of Th2-skewed pathology mediators in monocyte-derived type 2 of dendritic cells (DC2).* Immunol Lett, 2009. **126**(1-2): p. 29-36.
275. Burnett, C. and K. Howard, *Fly and mammalian lipid phosphate phosphatase isoforms differ in activity both in vitro and in vivo.* EMBO Rep, 2003. **4**(8): p. 793-9.
276. Pyne, S., K.C. Kong, and P.I. Darroch, *Lysophosphatidic acid and sphingosine 1-phosphate biology: the role of lipid phosphate phosphatases.* Seminars in cell & developmental biology, 2004. **15**(5): p. 491-501.
277. Mechtcheriakova, D., et al., *FTY720-phosphate is dephosphorylated by lipid phosphate phosphatase 3.* FEBS letters, 2007. **581**(16): p. 3063-8.
278. Simon, M.F., et al., *Expression of ectolipid phosphate phosphohydrolases in 3T3F442A preadipocytes and adipocytes. Involvement in the control of lysophosphatidic acid production.* The Journal of biological chemistry, 2002. **277**(26): p. 23131-6.
279. Jia, Y.J., et al., *Differential localization of lipid phosphate phosphatases 1 and 3 to cell surface subdomains in polarized MDCK cells.* FEBS letters, 2003. **552**(2-3): p. 240-6.
280. Humtsoe, J.O., et al., *Murine lipid phosphate phosphohydrolase-3 acts as a cell-associated integrin ligand.* Biochemical and Biophysical Research Communications, 2005. **335**(3): p. 906-19.
281. Long, J., et al., *Regulation of cell survival by lipid phosphate phosphatases involves the modulation of intracellular phosphatidic acid and sphingosine 1-phosphate pools.* The Biochemical journal, 2005. **391**(Pt 1): p. 25-32.
282. Kai, M., et al., *Lipid phosphate phosphatases 1 and 3 are localized in distinct lipid rafts.* Journal of biochemistry, 2006. **140**(5): p. 677-86.
283. Long, J.S., N.J. Pyne, and S. Pyne, *Lipid phosphate phosphatases form homo- and hetero-oligomers: catalytic competency, subcellular distribution and function.* The Biochemical journal, 2008. **411**(2): p. 371-7.
284. Nelson, W.J. and R. Nusse, *Convergence of Wnt, beta-catenin, and cadherin pathways.* Science, 2004. **303**(5663): p. 1483-7.
285. Tanyi, J.L., et al., *The human lipid phosphate phosphatase-3 decreases the growth, survival, and tumorigenesis of ovarian cancer cells: validation of the lysophosphatidic acid signaling cascade as a target for therapy in ovarian cancer.* Cancer research, 2003. **63**(5): p. 1073-82.
286. Chatterjee, I., et al., *Lipid phosphate phosphatase-3 regulates tumor growth via beta-catenin and CYCLIN-D1 signaling.* Mol Cancer, 2011. **10**: p. 51.

287. Panchatcharam, M., et al., *Mice with targeted inactivation of ppap2b in endothelial and hematopoietic cells display enhanced vascular inflammation and permeability.* Arterioscler Thromb Vasc Biol, 2014. **34**(4): p. 837-45.
288. Erbilgin, A., et al., *Identification of CAD candidate genes in GWAS loci and their expression in vascular cells.* J Lipid Res, 2013. **54**(7): p. 1894-905.
289. Knowlden, S. and S.N. Georas, *The autotaxin-LPA axis emerges as a novel regulator of lymphocyte homing and inflammation.* J Immunol, 2014. **192**(3): p. 851-7.
290. Yung, Y.C., N.C. Stoddard, and J. Chun, *LPA Receptor Signaling: Pharmacology, Physiology, and Pathophysiology.* J Lipid Res, 2014.
291. Bot, M., et al., *Atherosclerotic lesion progression changes lysophosphatidic acid homeostasis to favor its accumulation.* Am J Pathol, 2010. **176**(6): p. 3073-84.
292. Siess, W., et al., *Lysophosphatidic acid mediates the rapid activation of platelets and endothelial cells by mildly oxidized low density lipoprotein and accumulates in human atherosclerotic lesions.* Proc Natl Acad Sci U S A, 1999. **96**(12): p. 6931-6.
293. Smyth, S.S., et al., *Arguing the case for the autotaxin-lysophosphatidic acid-lipid phosphate phosphatase 3-signaling nexus in the development and complications of atherosclerosis.* Arterioscler Thromb Vasc Biol, 2014. **34**(3): p. 479-86.
294. Stegemann, C., et al., *Comparative lipidomics profiling of human atherosclerotic plaques.* Circ Cardiovasc Genet, 2011. **4**(3): p. 232-42.
295. Schober, A. and W. Siess, *Lysophosphatidic acid in atherosclerotic diseases.* Br J Pharmacol, 2012. **167**(3): p. 465-82.
296. Ren, H., et al., *Lipid phosphate phosphatase (LPP3) and vascular development.* Biochim Biophys Acta, 2013. **1831**(1): p. 126-32.
297. Mendelson, K., T. Evans, and T. Hla, *Sphingosine 1-phosphate signalling.* Development, 2014. **141**(1): p. 5-9.
298. Breart, B., et al., *Lipid phosphate phosphatase 3 enables efficient thymic egress.* J Exp Med, 2011. **208**(6): p. 1267-78.
299. Skoura, A., et al., *Sphingosine-1-phosphate receptor-2 function in myeloid cells regulates vascular inflammation and atherosclerosis.* Arterioscler Thromb Vasc Biol, 2011. **31**(1): p. 81-5.
300. Zeller, T., et al., *Genetics and beyond--the transcriptome of human monocytes and disease susceptibility.* PLoS One, 2010. **5**(5): p. e10693.
301. Wu, C., et al., *Mechanosensitive PPAP2B Regulates Endothelial Responses to Atherorelevant Hemodynamic Forces.* Circ Res, 2015. **117**(4): p. e41-e53.
302. Halazonetis, T.D., et al., *c-Jun dimerizes with itself and with c-Fos, forming complexes of different DNA binding affinities.* Cell, 1988. **55**(5): p. 917-24.
303. Escalante-Alcalde, D., R. Sanchez-Sanchez, and C.L. Stewart, *Generation of a conditional Ppap2b/Lpp3 null allele.* Genesis, 2007. **45**(7): p. 465-9.
304. Andersson, R., et al., *An atlas of active enhancers across human cell types and tissues.* Nature, 2014. **507**(7493): p. 455-61.
305. Lan, Q., K.O. Mercurius, and P.F. Davies, *Stimulation of transcription factors NF kappa B and AP1 in endothelial cells subjected to shear stress.* Biochem Biophys Res Commun, 1994. **201**(2): p. 950-6.

306. Lu, X., et al., *Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease.* Nat Genet, 2012. **44**(8): p. 890-894.
307. Hager, J., et al., *Genome-wide association study in a Lebanese cohort confirms PHACTR1 as a major determinant of coronary artery stenosis.* PLoS One, 2012. **7**(6): p. e38663.
308. Beaudoin, M., et al., *Myocardial Infarction-Associated SNP at 6p24 Interferes With MEF2 Binding and Associates With PHACTR1 Expression Levels in Human Coronary Arteries.* Arterioscler Thromb Vasc Biol, 2015. **35**(6): p. 1472-9.
309. Potthoff, M.J. and E.N. Olson, *MEF2: a central regulator of diverse developmental programs.* Development, 2007. **134**(23): p. 4131-40.
310. Ota, T., et al., *Complete sequencing and characterization of 21,243 full-length human cDNAs.* Nat Genet, 2004. **36**(1): p. 40-5.
311. Miralles, F., et al., *Actin dynamics control SRF activity by regulation of its coactivator MAL.* Cell, 2003. **113**(3): p. 329-42.
312. Wiezlak, M., et al., *G-actin regulates shuttling and PP1 binding by the RPEL protein Phactr1 to control actomyosin assembly.* J Cell Sci, 2012.
313. Sorokin, A.V., E.R. Kim, and L.P. Ovchinnikov, *Nucleocytoplasmic transport of proteins.* Biochemistry (Mosc), 2007. **72**(13): p. 1439-57.
314. Mouilleron, S., et al., *Structures of the Phactr1 RPEL Domain and RPEL Motif Complexes with G-Actin Reveal the Molecular Basis for Actin Binding Cooperativity.* Structure, 2012. **20**(11): p. 1960-70.
315. Moorhead, G.B., L. Trinkle-Mulcahy, and A. Ulke-Lemee, *Emerging roles of nuclear protein phosphatases.* Nat Rev Mol Cell Biol, 2007. **8**(3): p. 234-44.
316. Heroes, E., et al., *The PP1 binding code: a molecular-lego strategy that governs specificity.* FEBS J, 2013. **280**(2): p. 584-95.
317. Cougot, D., et al., *Inhibition of PP1 phosphatase activity by HBx: a mechanism for the activation of hepatitis B virus transcription.* Sci Signal, 2012. **5**(205): p. ra1.
318. Fardilha, M., et al., *The physiological relevance of protein phosphatase 1 and its interacting proteins to health and disease.* Curr Med Chem, 2010. **17**(33): p. 3996-4017.
319. Favot, L., et al., *Overexpression of a family of RPEL proteins modifies cell shape.* FEBS Lett, 2005. **579**(1): p. 100-4.
320. Fils-Aime, N., et al., *MicroRNA-584 and the protein phosphatase and actin regulator 1 (PHACTR1), a new signaling route through which transforming growth factor-beta Mediates the migration and actin dynamics of breast cancer cells.* J Biol Chem, 2013. **288**(17): p. 11807-23.
321. Zhang, X., et al., *Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes.* Hum Mol Genet, 2014. **23**(3): p. 782-95.
322. Allain, B., et al., *Neuropilin-1 regulates a new VEGF-induced gene, Phactr-1, which controls tubulogenesis and modulates lamellipodial dynamics in human endothelial cells.* Cell Signal, 2012. **24**(1): p. 214-23.
323. Flicek, P., et al., *Ensembl 2013.* Nucleic Acids Research, 2013. **41**(D1): p. D48-D55.
324. Consortium, F., et al., *A promoter-level mammalian expression atlas.* Nature, 2014. **507**(7493): p. 462-70.

325. Dinkel, H., et al., *The eukaryotic linear motif resource ELM: 10 years and counting.* Nucleic Acids Res, 2014. **42**(Database issue): p. D259-66.
326. Sotiropoulos, A., et al., *Signal-regulated activation of serum response factor is mediated by changes in actin dynamics.* Cell, 1999. **98**(2): p. 159-69.
327. Wiezlak, M., et al., *G-actin regulates the shuttling and PP1 binding of the RPEL protein Phactr1 to control actomyosin assembly.* J Cell Sci, 2012. **125**(Pt 23): p. 5860-72.
328. Lange, A., et al., *Classical nuclear localization signals: definition, function, and interaction with importin alpha.* J Biol Chem, 2007. **282**(8): p. 5101-5.
329. Medina, M.W., et al., *Coordinately regulated alternative splicing of genes involved in cholesterol biosynthesis and uptake.* PLoS One, 2011. **6**(4): p. e19420.
330. Drogemuller, C., et al., *Assignment of the PHACTR1 gene to bovine chromosome 23q24 by fluorescence in situ hybridization and radiation hybrid mapping.* Cytogenet Genome Res, 2005. **109**(4): p. 533.
331. Zhang, X., et al., *Genetic Associations with Expression for Genes Implicated in GWAS Studies for Atherosclerotic Cardiovascular Disease and Blood Phenotypes.* Hum Mol Genet, 2013.
332. Zelensky, A.N. and J.E. Gready, *The C-type lectin-like domain superfamily.* FEBS J, 2005. **272**(24): p. 6179-217.
333. Dambuza, I.M. and G.D. Brown, *C-type lectins in immunity: recent developments.* Curr Opin Immunol, 2015. **32**: p. 21-7.
334. Ogura, S., et al., *Lox-1: the multifunctional receptor underlying cardiovascular dysfunction.* Circ J, 2009. **73**(11): p. 1993-9.
335. Mehta, J.L., et al., *Deletion of LOX-1 reduces atherogenesis in LDLR knockout mice fed high cholesterol diet.* Circ Res, 2007. **100**(11): p. 1634-42.
336. Chen, Q., et al., *Genetic variation in lectin-like oxidized low-density lipoprotein receptor 1 (LOX1) gene and the risk of coronary artery disease.* Circulation, 2003. **107**(25): p. 3146-51.
337. Ishikawa, E., et al., *Direct recognition of the mycobacterial glycolipid, trehalose dimycolate, by C-type lectin Mincle.* J Exp Med, 2009. **206**(13): p. 2879-88.
338. Marakalala, M.J., L.M. Graham, and G.D. Brown, *The role of Syk/CARD9-coupled C-type lectin receptors in immunity to Mycobacterium tuberculosis infections.* Clin Dev Immunol, 2010. **2010**: p. 567571.
339. Tanaka, M., et al., *Macrophage-inducible C-type lectin underlies obesity-induced adipose tissue fibrosis.* Nat Commun, 2014. **5**: p. 4982.
340. Balch, S.G., et al., *Organization of the mouse macrophage C-type lectin (Mcl) gene and identification of a subgroup of related lectin molecules.* Eur J Immunogenet, 2002. **29**(1): p. 61-4.
341. Matsumoto, M., et al., *A novel LPS-inducible C-type lectin is a transcriptional target of NF-IL6 in macrophages.* J Immunol, 1999. **163**(9): p. 5039-48.
342. Yamasaki, S., et al., *Mincle is an ITAM-coupled activating receptor that senses damaged cells.* Nat Immunol, 2008. **9**(10): p. 1179-88.
343. Furukawa, A., et al., *Structural analysis for glycolipid recognition by the C-type lectins Mincle and MCL.* Proc Natl Acad Sci U S A, 2013. **110**(43): p. 17438-43.

344. Feinberg, H., et al., *Mechanism for recognition of an unusual mycobacterial glycolipid by the macrophage receptor mincle.* J Biol Chem, 2013. **288**(40): p. 28457-65.
345. Hattori, Y., et al., *Glycerol monomycolate is a novel ligand for the human, but not mouse macrophage inducible C-type lectin, Mincle.* J Biol Chem, 2014. **289**(22): p. 15405-12.
346. Stocker, B.L., et al., *On one leg: trehalose monoesters activate macrophages in a Mincle-dependant manner.* Chembiochem, 2014. **15**(3): p. 382-8.
347. Wells, C.A., et al., *The macrophage-inducible C-type lectin, mincle, is an essential component of the innate immune response to Candida albicans.* J Immunol, 2008. **180**(11): p. 7404-13.
348. Vijayan, D., et al., *Mincle polarizes human monocyte and neutrophil responses to Candida albicans.* Immunol Cell Biol, 2012. **90**(9): p. 889-95.
349. Yamasaki, S., et al., *C-type lectin Mincle is an activating receptor for pathogenic fungus, Malassezia.* Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1897-902.
350. Sousa, M.D., et al., *Restoration of Pattern Recognition Receptor Costimulation to Treat Chromoblastomycosis, a Chronic Fungal Infection of the Skin.* Cell Host & Microbe, 2011. **9**(5): p. 436-443.
351. Ishikawa, T., et al., *Identification of distinct ligands for the C-type lectin receptors Mincle and Dectin-2 in the pathogenic fungus Malassezia.* Cell Host Microbe, 2013. **13**(4): p. 477-88.
352. Ribbing, C., et al., *Mast cells generated from patients with atopic eczema have enhanced levels of granule mediators and an impaired Dectin-1 expression.* Allergy, 2011. **66**(1): p. 110-9.
353. Kawata, K., et al., *Mincle and human B cell function.* J Autoimmun, 2012. **39**(4): p. 315-22.
354. Ichioka, M., et al., *Increased expression of macrophage-inducible C-type lectin in adipose tissue of obese mice and humans.* Diabetes, 2011. **60**(3): p. 819-26.
355. Lee, W.B., et al., *Neutrophils Promote Mycobacterial Trehalose Dimycolate-Induced Lung Inflammation via the Mincle Pathway.* PLoS Pathog, 2012. **8**(4): p. e1002614.
356. Lang, R., *Recognition of the mycobacterial cord factor by Mincle: relevance for granuloma formation and resistance to tuberculosis.* Front Immunol, 2013. **4**: p. 5.
357. Schoenen, H., et al., *Cutting edge: Mincle is essential for recognition and adjuvanticity of the mycobacterial cord factor and its synthetic analog trehalose-dibehenate.* J Immunol, 2010. **184**(6): p. 2756-60.
358. Schweneker, K., et al., *The mycobacterial cord factor adjuvant analogue trehalose-6,6'-dibehenate (TDB) activates the Nlrp3 inflammasome.* Immunobiology, 2012.
359. Desel, C., et al., *The Mincle-activating adjuvant TDB induces MyD88-dependent, TLR-independent Th1 and Th17 responses through IL-1R signalling.* Immunology, 2012. **137**: p. 310-311.
360. Zollbrecht, C., et al., *Expression pattern in human macrophages dependent on 9p21.3 coronary artery disease risk locus.* Atherosclerosis, 2013. **227**(2): p. 244-9.

361. Schoenen, H., et al., *Differential Control of Mincle-Dependent Cord Factor Recognition and Macrophage Responses by the Transcription Factors C/EBPbeta and HIF1alpha.* J Immunol, 2014. **193**(7): p. 3664-75.
362. Rota, S. and S. Rota, *Mycobacterium tuberculosis complex in atherosclerosis.* Acta Med Okayama, 2005. **59**(6): p. 247-51.
363. Moody, D.B., et al., *Cd1b-Mediated T Cell Recognition of a Glycolipid Antigen Generated from Mycobacterial Lipid and Host Carbohydrate during Infection.* The Journal of Experimental Medicine, 2000. **192**(7): p. 965-976.
364. Andersen, C.S., et al., *A simple mycobacterial monomycolated glycerol lipid has potent immunostimulatory activity.* J Immunol, 2009. **182**(1): p. 424-32.
365. Miyake, Y., et al., *C-type lectin MCL is an FcRgamma-coupled receptor that mediates the adjuvanticity of mycobacterial cord factor.* Immunity, 2013. **38**(5): p. 1050-62.
366. Matsunaga, I. and D.B. Moody, *Mincle is a long sought receptor for mycobacterial cord factor.* J Exp Med, 2009. **206**(13): p. 2865-8.
367. Desel, C., et al., *The Mincle-activating adjuvant TDB induces MyD88-dependent Th1 and Th17 responses through IL-1R signaling.* PLoS One, 2013. **8**(1): p. e53531.
368. Kiyotake, R., et al., *Human Mincle Binds to Cholesterol Crystals and Triggers Innate Immune Responses.* J Biol Chem, 2015. **290**(42): p. 25322-32.
369. Boyle, A.P., et al., *High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells.* Genome Res, 2011. **21**(3): p. 456-64.
370. Mikkelsen, T.S., et al., *Comparative epigenomic analysis of murine and human adipogenesis.* Cell, 2010. **143**(1): p. 156-69.
371. Garber, M., et al., *A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals.* Mol Cell, 2012. **47**(5): p. 810-22.
372. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.
373. Hunter, C.A. and S.A. Jones, *IL-6 as a keystone cytokine in health and disease.* Nat Immunol, 2015. **16**(5): p. 448-57.
374. Kiyotake, R., et al., *Human Mincle Binds to Cholesterol Crystals and Triggers Innate Immune Responses.* J Biol Chem, 2015.