Molecular mechanisms of disease heritability in atherosclerosis



Jiahao Jiang

Hertford College
Nuffield Department of Medicine
University of Oxford

A thesis submitted for the degree of
DPhil in Clinical Medicine

Trinity Term 2024

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise, by reference or acknowledgement, the work presented is entirely my own.

Abstract

Atherosclerotic coronary artery disease (CAD), the leading cause of mortality worldwide, is influenced by both genetic and environmental factors. Despite the identification of over 250 genetic risk loci through genome-wide association studies (GWAS), the specific causal variants, their regulatory mechanisms and their interactions with environmental risk factors remain largely unexplored.

Here, I focused on two crucial pathogenic events in early atherogenesis – the formation of lipid-laden foam cells and the onset of endothelial dysfunction, both triggered by atherogenic oxidised low-density lipoprotein cholesterol (ox-LDL). Utilising high-throughput multi-omics sequencing techniques, I identified the transcriptional regulatory elements and regulatory networks that underpin the cellular response to ox-LDL, prioritised causal CAD risk variants, and examined the molecular mechanisms of their regulatory effects.

Specifically, I developed a computational pipeline to systemically evaluate the relative genetic contribution to CAD across heterogeneous human macrophage subpopulations using a single-cell multi-omics approach. This highlighted a novel CD52^{hi} lipid-handling macrophage subpopulation with increased disease heritability. Functionally, lipid-handling macrophages displayed activated lipid metabolism pathways, accumulated fewer lipids upon ox-LDL exposure, and were reproducibly present in human atherosclerotic plaques. Mechanistically, I investigated the cis-regulatory effect of a specific CAD risk variant, rs10488763, that was activated in these lipid-handling macrophages and implicated its target *FDX1* as a disease gene for the first time.

In human aortic endothelial cells, our results implicated a largely shared transcriptional regulatory network in response to ox-LDL involving AP-1 and C/EBP-beta. The ox-LDL-induced changes of transcription factor bindings were enriched for CAD heritability and activated endothelial cell migration. An integrative analysis using genetic, transcriptomic and epigenomic data further revealed candidate regulatory mechanisms for rs62172376 at the *TFPI* locus.

Altogether, this study offers a comprehensive characterisation of the transcriptional regulation of ox-LDL response in macrophages and endothelial cells, providing new insights into the genotype-environment interactions underlying atherosclerotic disease. Our work presents an unbiased framework for interpreting GWAS results using functional genomic data in disease-relevant primary cell models.

Acknowledgements

As Oxford basks in its cherished sunny days, I find myself reflecting on the remarkable journey that has brought me to this point. Genetics, once an insipid high-school biology term printed alongside Mendel's pea flowers, has blossomed into the cornerstone of my research career, igniting a passion I hope to nurture for years to come. Looking back, my heart swells with gratitude for the countless individuals who have contributed their ideas, labour, and emotional support to this ambitious endeavour.

First and foremost, I extend my deepest thanks to my supervisor, Prof. Chris O'Callaghan, whose wisdom, genuine care for people, and great sense of humour have been a beacon throughout my journey. Chris has been an amazing mentor, teaching me not only the art of science but also the decency in life. I would also like to thank my second supervisor, Prof. Lucy Davison, for her unwavering support and infectious passion for scientific discovery.

To the members of the O'Callaghan group - Tom, Yash, Thomas, Nathali, Sara, Marsha, and Toby - thank you for being excellent colleagues and friends and for providing a lovely environment. A special thanks to Dr Tom Hiron for his guidance and supervision during my training years.

I am also grateful to the funding bodies and people who reviewed this work and believed in me. This project would not have been possible without the generous support of the University of Oxford, the China Scholarship Council, and the Nuffield Department of Medicine. In addition, I would like to thank my examiners, Drs Julian Knight, Mark

Crabtree, Katherine Bull, and Nicky Whiffin, for their invaluable feedback at various stages of this project.

I am deeply thankful for all my friends in Oxford and beyond, with whom I have shared the highs and lows of this challenging journey. Though I don’t have the space to list all your names, know that I treasure every moment we spent together, from road trips around Oxford and the UK to adventures around the world. I also wish to thank my dear neighbour Louis, my Goldador companion on countless muddy trails, for reminding me to always stay curious and hungry.

Finally, I wish to express my deepest gratitude to my family, to whom I owe everything: My parents, Mr Xiangyang Jiang 江向阳 and Mrs Lanhui Su 苏兰辉, chemical engineers who regained the momentum for learning English to understand their sons' work; and my twin brother Jiayu Su 苏嘉昱 - friend, collaborator, picky reviewer, and part-time nemesis. Your love and unconditional support have been everything to me. With immense pride and gratitude, I humbly accept the honour of being the first doctor in our family. I dedicate all this work to my family.

Table of Contents

Abstract ..... **i**
Acknowledgements ..... **iii**
Table of Contents ..... **v**
List of Figures ..... **ix**
List of Tables ..... **xii**
Abbreviations ..... **xiii**
Ch 1. Background **1**
1.1. Understanding the cellular context of atherosclerosis 1
1.1.1. Endothelial cell 2
1.1.2. Monocyte and macrophage 3
1.1.3. Other cell types involved in atherosclerosis 4
1.2. The atherogenicity of LDL cholesterol and its derivatives 6
1.3. The genetic component of atherosclerosis 7
1.3.1. Early genetic studies on human atherosclerosis 7
1.3.2. Genome-wide association studies 8
1.4. Genetics-led investigation of causal cell types 10
1.5. Translating GWAS to function 11
1.6. Aims and objectives 13
Ch 2. Materials and Methods **15**
2.1. Cell culture 15
2.1.1. Human monocyte-derived macrophages 15
2.1.2. Donor genotyping 16
2.1.3. Human aortic endothelial cells (HAECs) 16
2.1.4. Immortalised cell lines 16
2.2. Cell treatments 17
2.2.1. Preparation of LDL and DiI-LDL 17
2.2.2. Preparation of ox-LDL and DiI-ox-LDL 18
2.2.3. Plasmid Nucleofection 19
2.2.4. RNA Interference (RNAi) 20
2.3. Sequencing sample preparation 20
2.3.1. HAEC RNA sequencing 20
2.3.2. HAEC ATAC sequencing 21
2.3.3. Macrophage single-cell RNA sequencing 22
2.3.4. Macrophage single-cell multiome sequencing 22
2.4. Bulk sequencing data analysis 23
2.4.1. RNA-seq 23
2.4.2. ATAC-seq 24
2.4.3. ChIP-seq 26

2.5. Single-cell sequencing analysis 27
2.5.1. Meta-analysis of human atherosclerotic plaque data 27
2.5.2. Single-cell RNA-seq pre-processing 28
2.5.3. Integrative analysis between plaque and *ex vivo* macrophages 28
2.5.4. Single-cell multiome pre-processing 29
2.5.5. Transcription factor (TF) motif deviation analysis 30
2.5.6. Cis-regulatory analysis 30
2.5.7. Quantitative Trait Loci (QTL) SNPs 31
2.5.8. GWAS-related SNPs 31
2.5.9. Enrichment analysis for functional annotations 32
2.5.10. Heritability analysis using s-LDSC 32
2.6. Immunofluorescence imaging 33
2.6.1. Fixed cell imaging 33
2.6.2. Live cell lipid uptake assay 35
2.6.3. Endothelial migration assay 35
2.7. Flow cytometry 36
2.8. RT-qPCR 37
2.9. Luciferase assay 38
2.10. Seahorse assay 39
Ch 3. Single-cell multi-omics sequencing identifies cis-regulatory elements and recovers transcriptional regulatory networks associated with CAD variants 40
3.1. Introduction 40
3.2. Clustering and integration of single-cell multiome data 43
3.2.1. Re-analysing the single-cell RNA-seq data 43
3.2.2. Multiome data QC 46
3.2.3. snRNA-seq is highly concordant with scRNA-seq data 48
3.2.4. Joint clustering analysis using both modalities improves the resolution of the allocation of macrophage subpopulations 50
3.3. Heterogeneity in human macrophages and their response to ox-LDL 52
3.3.1. Heterogeneity in transcriptome 52
3.3.2. Heterogeneity in transcription factor activation 54
3.4. Mapping cis-regulatory elements 57
3.5. Macrophage CREs are enriched for functional annotations 58
3.5.1. Enrichment for public annotation 59
3.5.2. Enrichment for H3K27Ac signals 61
3.6. Constructing transcriptional regulatory networks 63
3.6.1. Domain of regulatory chromatins 64
3.6.2. TF-DORC regulatory networks 66
3.7. Discussion 68
Ch 4. A novel macrophage subpopulation conveying increased genetic risk of coronary artery disease 71

4.1. Introduction 71
4.2. Macrophage CREs prioritise causal variants and link them to target genes 74
4.2.1. Macrophage CREs are enriched for CAD risk 75
4.2.2. CRE-based variant prioritisation outperforms other methods 77
4.2.3. CAD-associated transcriptional regulatory networks 79
4.3. Development of a computational framework of partitioning disease heritability .. 80
4.4. Lipid-handling macrophages show enrichment in both subpopulation and subpopulation-oxLDL programmes 85
4.5. LHM display transcriptomic and epigenomic signatures related to lipid metabolism 88
4.6. LHM show much less lipoprotein accumulation after ox-LDL exposure 91
4.7. A CD52-high lipid-handling phenotype is robustly present in human atherosclerotic plaques 94
4.8. Discussion 100
Ch 5. Molecular explanations of the enriched heritability in lipid-handling macrophages 104
5.1. Introduction 104
5.2. The polygenetic signals at the *MITF* locus with rs6772383 and rs11452399 ... 106
5.3. The polygenetic signals at three *LPL* loci 110
5.4. The polygenetic signals at the *FDX1* locus with rs10488763 and rs1443120 ... 112
5.5. The allelic effect of rs10488763 on local chromatin accessibility and enhancer activity 114
5.6. The role of CEBPB in the cis-regulatory effect of rs10488763 116
5.7. *FDX1* knock-down reduces ox-LDL uptake in human macrophages 119
5.8. Discussion 120
Ch 6. Ox-LDL exposure in endothelial cells modulates gene regulatory programmes and genetic risk of CAD 123
6.1. Introduction 123
6.2. Ox-LDL accumulation triggers metabolic shift in HAECs 125
6.3. Transcriptomic analysis reveals disrupted fatty-acid metabolism 128
6.4. Integrative analysis highlights a shared ox-LDL-response network between endothelial cells and macrophages 133
6.4.1. In-silico deletion analysis 134
6.4.2. TF motif footprinting analysis 138
6.5. Differential transcription factor bindings induced by ox-LDL activate endothelial cell migration 140
6.6. Prioritisation of CAD variants using the DBSs in endothelial cells 143
6.6.1. rs56170783 at the *PLPP3* locus 146
6.6.2. Prioritised variants at the *CALCRL*/*TFPI* locus 147
6.7. Discussion 150

Ch 7. General discussion 154
7.1. An adaptable and scalable framework for exploring the interactions between genetic predisposition and environmental risk factors 155
7.2. Lipid-handling macrophage, lipid-associated macrophage, and human atherosclerosis 156
7.3. Reconciling the ox-LDL effect in endothelial cells 158
7.4. Limitations and future directions 159
7.5. Conclusions 160
Appendix 161
Supplemental materials 161
Published work 161
Bibliography 182

List of Figures

Figure 1.1 The cellular basis of atherogenesis featuring endothelial and myeloid lineages. 1
Figure 3.1 Representative images from immunofluorescence staining of primary human macrophages. 42
Figure 3.2 Uniform Manifold Approximation and Projection (UMAP) visualisation of the single-cell RNA-seq data. 44
Figure 3.3 Expression of selected marker genes for different macrophage subpopulations. 45
Figure 3.4 Ox-LDL triggers major transcriptomic changes in human macrophages. 46
Figure 3.5 Proportion of spliced and unspliced transcripts. 49
Figure 3.6 Ox-LDL-induced gene expression changes observed in macrophages using different techniques. 49
Figure 3.7 Uniform Manifold Approximation and Projection visualisation of the single-cell multiome data 50
Figure 3.8 Visualisation of macrophage subpopulations in each sample. 52
Figure 3.9 Dot plot of the expression of selected marker genes for the 9 macrophage subpopulations. 53
Figure 3.10 Motif deviation scores for the top 15 most deviated transcription factors. 55
Figure 3.11 Negative correlation between the gene expression and motif activity for FOS. 56
Figure 3.12 NF-kB family factors are activated in M1-like macrophages 57
Figure 3.13 Location of cis-regulatory elements (CREs). 58
Figure 3.14 Macrophage cis-regulatory elements (CREs) are enriched for functional annotations 61
Figure 3.15 Heatmap showing the aggregated H3K27ac signals in macrophage open chromatin regions. 62
Figure 3.16 Aggregated H3K27ac signals at CRE-peak and non-CRE-peak regions. 63
Figure 3.17 Domain of regulatory chromatin (DORC) analysis identified new transcriptional regulations compared to public bulk sequencing data. 64
Figure 3.18 Pearson correlation between ox-LDL–induced changes in cis-regulatory element chromatin accessibility and in gene expression. 65
Figure 3.19 Heatmap of regulation scores for the top 0.1% of TF-DORC enrichments. 67
Figure 4.1 Prioritisation of causal genetic variants using macrophage cis-regulatory elements 74
Figure 4.2 Permutation-based enrichment test for CAD risk variants. 75
Figure 4.3 Enrichment test for CAD heritability. 76
Figure 4.4 Permutation-based enrichment test for CAD risk variants using annotations from Activity-By-Contact (ABC) enhancers. 79
Figure 4.5 Network visualization of significant TF-DORC pairs for CAD-DORCs. 80
Figure 4.6 Number of independent CAD loci mapped onto CREs targeting subpopulation-specific genes. 82
Figure 4.7 GWAS heritability enrichment framework. 83
Figure 4.8 The enrichment for CAD heritability in subpopulation gene programmes. 85
Figure 4.9 Enrichments for CAD heritability in gene programmes. 86

Figure 4.10 Heatmap showing the correlation (Pearson correlation coefficient) between the SNP annotation matrices. 87
Figure 4.11 UMAP visualisation of the expression of selected marker genes for lipid-handling macrophages. 89
Figure 4.12 Pathways enriched in the marker genes of lipid-handling macrophages. 89
Figure 4.13 The co-expression gene module 51 identified in cross-tissue sequencing of a large CAD cohort[212]. 90
Figure 4.14 Motif analysis for transcription factors (TF) driving the expression of marker genes of lipid-handling macrophages. 91
Figure 4.15 Experimental validation of CD52, CHI3L1 and CD36 as markers for lipid-handling macrophages. 92
Figure 4.16 $CD52^{+}$ lipid-handling macrophages displayed reduced for lipid accumulation after ox-LDL exposure. 93
Figure 4.17 The expression of CD36 in lipid-handling macrophages as well as in other macrophages. 94
Figure 4.18 Meta-analysis of scRNA-seq profiles in 38 human atherosclerotic samples. 95
Figure 4.19 Plaque macrophage atlas constructed from the meta-analysis. 96
Figure 4.20 Monocyte-derived macrophages closely mirror the transcriptomic signatures of the $TREM2^{hi}$ lipid-associated macrophage (LAM) in plaque. 97
Figure 4.21 The $CD52^{hi}$ lipid-handling phenotype is present in human atherosclerotic plaques. 99
Figure 5.1 The regulatory landscape at the prioritised *MITF* locus. 107
Figure 5.2 Functional validation of MITF in ox-LDL-exposed macrophages. 109
Figure 5.3 The regulatory landscape at the *LPL* gene with three prioritised loci. 111
Figure 5.4 The regulatory landscape at the prioritised *FDX1*/*RDX* locus. 112
Figure 5.5 Polygenetic signals at the FDX1/RDX locus. 113
Figure 5.6 The alternative risk T allele of the prioritised SNP rs10488763 alters the regulatory property of its local chromatin. 116
Figure 5.7 The cis-regulatory effect of rs10488763 on FDX1 expression is mediated by CEBPB. 118
Figure 5.8 FDX1 knock-down reduces lipid accumulation in primary macrophages. 119
Figure 5.9 Model showing the role of AP-1 factors and CEBPB in modulating the regulatory effect of CAD risk SNP rs10488763. 121
Figure 6.1 Ox-LDL exposure triggers intracellular accumulation of lipid droplets in primary aortic endothelial cells (HAECs). 125
Figure 6.2 Ox-LDL induces metabolic reprogramming in HAECs. 126
Figure 6.3 Ox-LDL alone does not trigger pro-inflammatory phenotypes in HAECs. 127
Figure 6.4 Volcano plot showing differentially expressed genes induced by ox-LDL in human aortic endothelial cells. 128
Figure 6.5 Pathways disrupted by ox-LDL exposure. 130
Figure 6.6 Exposure to ox-LDL but not ox-PAPC recapitulates the transcriptomic signatures associated with endothelial cells in human atherosclerosis. 132
Figure 6.7 Limited effect of ox-LDL on the global open chromatin landscape in endothelial cells. 133

Figure 6.8 In-silico deletion (ISD) analysis based on public sequencing data identifies potential regulators underlying the transcriptomic response to ox-LDL. 135
Figure 6.9 CEBPB is as a potential regulator of endothelial cell response to ox-LDL. 137
Figure 6.10 Ox-LDL triggers redistribution of transcription factor binding within open chromatin sites. 140
Figure 6.11 Changes in transcription factor binding activity indicate the activation of endothelial migration pathway in response to ox-LDL. 142
Figure 6.12 Ox-LDL-induced differential transcription factor binding sites are enriched for CAD heritability. 144
Figure 6.13 The regulatory landscape at the prioritised *PLPP3* locus. 146
Figure 6.14 The regulatory landscape at the prioritised *CALCRL*/*TFPI* locus. 147
Figure 6.15 TFPI is the candidate causal gene at the prioritised locus. 148
Figure 6.16 Alternative splicing of CALCR-AS1 correlates with TFPI expression in endothelial cells. 150

List of Tables

Table 2.1 Densities of KBr solution for LDL preparation..... 17
Table 2.2 Plasmids used in this thesis 19
Table 2.3 Antibody dilutions for flow cytometry 36
Table 2.4 PCR primers used in this thesis 38
Table 3.1 Description and summary statistics for the snRNA-seq library of the multiome data 47
Table 3.2 Description and summary statistics for the snATAC-seq library of the multiome data..... 48
Table 6.1 Top 10 most upregulated protein-coding genes by ox-LDL in HAECs 129
Table 6.2 Top 10 most downregulated protein-coding genes by oxLDL in HAECs 129
Table 6.3 Prioritised CAD-risk variants in ox-LDL-induced DBSs 145

Online Table S1 OxLDL-induced differentially expressed genes in macrophages identified by single-cell transcriptomics. 161
Online Table S2 Marker genes of macrophage subpopulations identified by single-cell multi-omics. 161
Online Table S3 Predicted macrophage cis-regulatory elements..... 161
Online Table S4 TF-DORC pairs with strong cis-regulations. 161
Online Table S5 Prioritized CAD risk variants. 161

Abbreviations

AP-1	Activating protein 1
ATAC	Assay for transposase accessible chromatin
CAD	Coronary artery disease
caQTL	Chromatin accessibility QTL
CD	Cluster of differentiation
ChIP	Chromatin immuno-precipitation
CRE	Cis-regulatory element
DBS	Differential transcription factor binding site
DEG	Differentially expressed gene
DORC	Domain of regulatory chromatin
dsDNA	Double-stranded DNA
EC	Endothelial cells
eQTL	Expression quantitative trait locus
FCCP	Carbonyl cyanide-p-trifluoromethoxyphenylhydrazone
FDR	False discovery rate
GO	Gene ontology
GWAS	Genome-wide association study
H3K27ac	Histone H3 Lysine 27 acetylation
HAEC	Human aortic endothelial cell
HUVEC	Human umbilical vein endothelial cell
IF	Immunofluorescence
ISD	In-silico deletion
KEGG	Kyoto Encyclopedia of Genes and Genomes
LAM	Lipid-associated macrophage
LAP	Liver-enriched transcriptional activator protein
LD	Linkage disequilibrium
LDL	Low density lipoprotein
LDSC	Linkage-disequilibrium-score regression
LHM	Lipid-handling macrophage

LPS	Lipopolysaccharide
M-CSF	Macrophage colony stimulating factor
MΦ	Macrophage
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
ox-LDL	Oxidised low-density lipoprotein
ox-PAPC	Oxidised 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphocholine
PBS	Phosphate buffered saline
PC	Principal component
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
scATAC-seq	Single-cell ATAC-seq
scRNA-seq	Single-cell RNA-seq
siRNA	Small interference RNA
SMC	Smooth muscle cell
SNP	Single nucleotide polymorphism
snRNA-seq	Single-nucleus RNA-seq
TF	Transcription factor
TFBS	Transcription factor binding site
TSS	Transcription start site
UMAP	Uniform manifold approximation and projection
UTR	Untranslated region

Ch 1. Background

1.1. Understanding the cellular context of atherosclerosis

Atherosclerosis is a chronic inflammatory disease characterised by the formation of lipid-rich substances called plaques within the arterial wall, leading to the narrowing and hardening of the arteries. It is the primary underlying cause of coronary artery disease (CAD), which eventually may lead to myocardial infarction and stroke, two of the leading causes of global mortality [1]. The pathogenesis of atherosclerosis is complex and multifactorial, involving various cell types and numerous molecular pathways (Figure 1.1). Understanding the cellular and molecular context of atherogenesis is crucial for the development of effective strategies for therapeutic intervention and disease prevention.

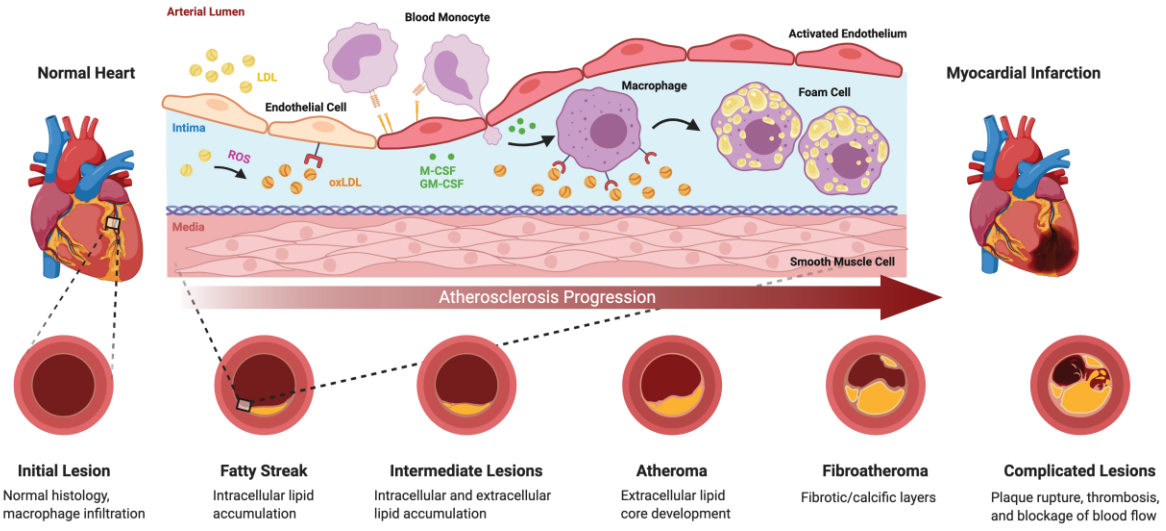


Figure 1.1 The cellular basis of atherogenesis featuring endothelial and myeloid lineages.
Circulating monocytes adhere to activated endothelium and infiltrate the arterial wall. Once in the arterial intima, monocytes differentiate into macrophages, which engulf lipids and become lipid-laden foam cells. Atherosclerotic lesions continue to evolve under regional inflammation and other environmental stimuli, form fibrotic layers and may eventually rupture provoking acute thrombosis. Illustration adapted from [2] and [3].

1.1.1. Endothelial cell

Endothelial cells form the innermost layer of the arterial wall that border the luminal blood flow. Although the exact causes of the initiation of atherosclerosis are still poorly understood, the activation and subsequent dysfunction of endothelial cells are likely amongst the earliest pathological events that drive atherosclerosis [4]. Endothelial cell dysfunction is characterised by a shift in the vascular homeostasis towards reduced vasodilation, proinflammatory phenotypes, and prothrombic properties. During this process, the size of cell-cell junctions in the endothelial barrier increases, allowing circulating cholesterol and plasma proteins to move freely into the subendothelial space [5,6]. The elevated expression of cell adhesion molecules like ICAM-1 and VCAM-1 in dysfunctional endothelial cells facilitates the recruitment and transmigration of immune cells from circulation, further exacerbating the inflammatory and necrotic microenvironment within the arterial wall [7]. Additionally, endothelial cells secrete chemokines and growth factors that promote the abnormal migration and proliferation of smooth muscle cells [8,9].

Endothelial cell dysfunction is often triggered by risk factors such as hypercholesterolemia, hypertension, smoking and diabetes [10,11], all of which linked to the increased incidence of CAD [12]. However, even under these systemic manifestations, the occurrence of atherosclerotic lesions follows regional patterns and is concentrated at arterial branches and curvatures. It is later recognised that shear stress, the frictional force exerted by blood flow on the endothelial surface, is also a significant determinant of endothelial function and atherosclerosis [13]. Laminar shear stress, typically found in straight parts of the artery, promotes a quiescent, anti-inflammatory endothelial phenotype [14]. In contrast, disturbed

oscillating shear stress, which typically occurs at artery bifurcations, induces endothelial activation and promotes atherogenesis [15,16].

1.1.2. Monocyte and macrophage

Regardless of the initial trigger or triggers of atherosclerosis, there is a growing consensus from both murine and human studies that the recruitment of blood leukocytes, mediated by activated endothelium, marks the early onset of lesion formation [17-19]. In particular, circulating monocytes adhere to the inner lining of the arterial wall, infiltrate the intima, and differentiate into macrophages and dendritic cells. Studies in murine models of atherosclerosis have demonstrated that impaired function in any of the adhesion, transmigration and differentiation processes reduces the overall atherosclerosis burden [20,21], suggesting a central role of myeloid cells in the initiation and progression of atherosclerosis.

The formation of lipid-laden foam cells is a key step in the atherogenesis, marking dysregulated lipid homeostasis triggered by excessive atherogenic cholesterols. The precise origin of foam cells in the atherosclerotic lesions is still under debate, particularly because of the macrophage-like phenotype discovered in trans-differentiated vascular smooth muscle cells [22]. However, macrophage undoubtably play a crucial role in lipid accumulation by engulfing the retained cholesterol-rich lipoproteins in the subendothelial space, as well as influencing the plaque microenvironment through the secretion of various cytokines, chemokines and matrix metalloproteinases [19].

Historically, macrophages in atherosclerotic lesions were categorised based on the established M1/M2 diagram [23], with the pro-inflammatory M1 phenotype linked to the

progression and anti-inflammatory M2 phenotype linked to the resolution of atherosclerosis. This dichotomy, derived from *in vitro* studies of macrophage polarisation, oversimplifies the remarkable plasticity of myeloid cells *in vivo*, and has been increasingly challenged by high-resolution characterisations of atherosclerotic plaques powered by single-cell analysis [24-27]. Currently, one of the most important macrophage subpopulations trending in the macrophage research in a broad range of human diseases is the lipid-associated macrophage (LAM). LAM can be distinguished by the expression of *TREM2*, and is widely found in liver, adipose tissue, and atherosclerotic plaques [28,29]. Despite its well-recognised function in maintaining the metabolic homeostasis, the role of LAM in atherogenesis is still to be determined. Some report that the inflammatory transition of naïve LAM increases the risk of cerebrovascular events [30], implying a protective role of LAM. Conversely, a recent CRISPR screen has shown that deficiency in *TREM2* leads to a reduced ability of lipid uptake in foamy macrophages, which subsequently attenuates plaque progression [31].

1.1.3. Other cell types involved in atherosclerosis

Arterial smooth muscle cells (SMCs) are the primary cellular component of the vascular wall and directly control blood vessel contraction. Normally, SMCs reside quiescent in the media of the artery, characterised by the expression of specific markers such as alpha-smooth muscle actin, smooth muscle myosin, and calponin [9]. In response to pro-inflammatory stimuli and growth factors, SMCs migrate to the intima and undergo clonal expansion [32], exhibiting a remarkable spectrum of cellular functions. Single-cell lineage tracing and fate mapping in mice reveal SMC cell states that resemble fibroblasts, macrophages and osteoblasts more than the conventional contractile phenotype of SMC

[33-35]. Notably, the trans-differentiated SMCs can interact with various atherogenic stimuli including retained LDL particles, and can exhibit a lipid-loaded foamy phenotype. These SMC-derived foam cells, however, differ significantly in lysosome function and lipid-metabolism compared foam cells with myeloid origin [22]. SMCs are also the major cellular source of the fibrous cap in advanced atherosclerotic lesions [36]. Deciphering the functional consequences of the diverse SMC states involved in atherogenesis could provide promising therapeutic interventions for the disease's advanced stages, yet still require considerable work in both mouse and human models of atherosclerosis.

T cells are typically believed to be recruited by a series of apoptotic and necrotic signals released from foam cells, significantly driving the progression of late-stage anthogenesis [37]. The abundance of various subsets of T cells have been observed in mouse and human atherosclerotic plaques for a long time [38-40]. With the rapid development of single-cell techniques, T cells in atherosclerotic plaques have been extensively characterised and are broadly assigned to functional subsets based on membrane expression of CD4 and CD8 [41,42]. Like macrophages, different T cell subsets are linked to different roles in atherosclerosis. For instance, cytotoxic $CD8^+$ T cells are enriched in human atherosclerotic plaque compared with blood, and display transcriptomic signatures related to exhaustion in patients underwent cerebrovascular complications [42]. Conversely, tissue-resident $CD4^+$ T cells are more activated in asymptomatic patients, accompanied with a relative depletion of effector memory $CD4^+$ T cells. The recent discovery of a lipoprotein-reactive $CD4^+$ subset and its transformation into pathogenic Th1/Th17 cell highlights the autoimmune T cell response as an important pathway that influences different stages of atherosclerosis [43].

1.2. The atherogenicity of LDL cholesterol and its derivatives

Low-density lipoprotein (LDL) cholesterol, often referred to as 'bad cholesterol', has been a subject of extensive research due to its pivotal role in the pathogenesis of atherosclerosis and subsequent cardiovascular complications. The observation of excessive cholesterol in human atherosclerotic lesions dates back to more than a century ago [44,45]. At that time, pre-mature death of children with hypercholesterolemia from serious heart problems had been reported [46], but the nature of circulating lipids and their atherogenicity was not well recognised, partly due to the failure in early experimental work feeding pure cholesterol to dogs and rats [47]. The development of ultracentrifugation techniques greatly facilitated the separation of fractions in human plasma, leading to the first discovery of LDL [48]. It was later found that approximately 70% of total cholesterol in circulation was carried in the form of LDL particles [49,50]. Since then, LDL cholesterol has been closely monitored as an important marker in clinical lipidology [48].

Epidemiological evidence has consistently demonstrated a positive correlation between elevated LDL in circulation and the incidence of CAD. The Framingham Heart Study, one of the most influential studies in cardiovascular epidemiology, established that high LDL cholesterol is a major risk factor for CAD [51]. Similarly, the Multiple Risk Factor Intervention Trial (MRFIT) confirmed the association between LDL cholesterol and CAD mortality [52]. Today, LDL is included in the majority of clinical indexes for prediction of CAD risks, as well as in the guidelines for disease prevention [53].

On a molecular level, LDL cholesterol contributes to atherosclerosis through a series of complex processes. The search for the disease gene in familial hypercholesterolemia led

to the discovery of a membrane receptor for LDL, known as *LDLR*, which mediates the endocytosis of LDL particles [54]. However, the uptake of native LDL in most cells including macrophages was so slow that even at very high concentrations, no lipid-laden phenotype could develop. This paradox was eventually resolved with the identification of *in vitro* modified LDL derivatives and their corresponding scavenger receptors, which increase the cellular cholesterol content at a much faster rate [55-57]. One of the most important LDL modifications *in vivo* is oxidised LDL (ox-LDL). Originally called "endothelial cell-modified LDL", ox-LDL is derived from the infiltrated LDL particles within the artery wall and is recognised by scavenger receptors on macrophages with high affinity [58-60]. Failure in the clearance of ox-LDL leads to the formation of lipid-laden foam cells, a hallmark of atherosclerotic plaques. Other than its evident role in lipid metabolism, ox-LDL has been shown to induce endothelial dysfunction, inflammation, and thrombosis, all of which contribute to the progression of atherosclerosis [57,61].

1.3. The genetic component of atherosclerosis

1.3.1. Early genetic studies on human atherosclerosis

Atherosclerosis is a complex disorder with multifactorial aetiology. Like most complex traits, atherosclerosis is also largely influenced by genetic factors inherited in the human genome. Driven by a growth of eugenics, the study of disease heredity, termed as "genetics", surged in early 1900s shortly after the rediscovery of Gregor Mendel's work in pea plants. In late 1930s, the onset of atherosclerosis was associated to an inherited hypercholesterolemia disorder [62], which was further categorised into a mild heterozygous and a more severe homozygous phenotype [63]. However, perhaps overshadowed by the

success in the lipoprotein research, the genetic component of atherosclerosis was not widely recognised until the establishment of twin registries like Danish Twin Registry [64] and Swedish Twin Registry [65].

Even before the central dogma of biology was introduced, researchers have been using population studies in pairs of identical and fraternal twins to estimate the heritability of a trait [66]. The basic hypothesis underlying classical twin studies is that environmental influences can be well-controlled between twin pairs, so the genetic contribution to a specific trait will make monozygotic twins more similar to each other than dizygotic twins. In general, heritability of CAD was estimated to be 40% to 60% [67], with certain disease complications like calcified plaque volume having a heritability of up to 80% [68]. The correlation in twins is also heterogenous across sex and age, where male and preadolescent twins sharing higher correlation [68,69].

1.3.2. Genome-wide association studies

For the past twenty years, the genome-wide association study (GWAS) has emerged as a revolutionary tool in the field, providing unprecedented insights into the genetic basis of complex diseases. Recent advances in genotyping and high-throughput sequencing technology drastically lower the cost for mapping genetic variations in increasingly large population cohorts, leading to the discovery of more than 250 genetic loci carrying increased CAD risk [70-74]. Within a single risk locus, multiple statistically independent signals can exist, leading to an even greater number of underlying causal risk variants. While the use of whole exon sequencing has uncovered the disease modulating effect of numerous rare genetic variants in the coding region [72,75], most of the GWAS-identified

risk loci are in the non-protein-coding region of human genome with unknown functional consequences [76,77].

Unravelling the genetic basis of CAD risk loci, and more specifically, how a genetic variant mechanistically affects atherogenesis, has evolved slowly during the past decade. One major challenge is that GWAS-identified variants are in linkage disequilibrium with many other variants in haplotype, often creating significant GWAS signals that can span hundreds of thousands base pairs. While various statistical methods have been developed to delineate the correlated GWAS signals [78,79], the biology that segregates neighbouring genetic variants during meiosis creates chunks of genomic regions in tight LD that are virtually indistinguishable without additional functional annotations. In other words, genetic variants showing the highest signal in GWAS are not necessarily causal, and each variant in the LD block, even not genotyped in the GWAS cohort, could have a disease modulating effect.

Additionally, as discussed previously, the pathogenesis of atherosclerosis involves interactions between cells from multiple lineages, all carrying the same genetic information within individual patients. The observation that disease-associated loci are enriched in cell-type-specific regulatory regions suggests that genetic risk variants modulate disease risk under specific cellular contexts [80]. However, current knowledge of the exact cells and tissues in which risk variants operate is still limited for most complex human diseases including CAD.

1.4. Genetics-led investigation of causal cell types

GWAS-identified risk loci provide leverage points for exploring gene regulatory mechanism underlying complex diseases in a hypothesis-free way. However, translating the genetic discoveries into therapeutic opportunities necessitates robust models and experimental systems where the function of candidate genes can be manipulated with disease-relevant readouts. For example, LDL receptor pathway in hepatic cells has been a fundamental model for studies in LDL lowering therapies [81], and ultimately led to the development of PSCK9 inhibitors that reduce the risk of cardiovascular events [82]. These model systems are invaluable for high-throughput drug target screening, but require extensive knowledge of the causal cellular mechanisms, which remains unidentified for most complex human diseases.

Methodological and technological advances now allow the systemic mapping of transcription, chromatin structure, and histone modification across a wide range of tissues or cell types. Large consortium projects like ENCODE [83] and FANTOM5 [84] have identified hundreds of thousands regulatory DNA elements with cell-type-specific activity. Many statistical methods have been developed that combine this type of information with GWAS data, with expectation that enrichment test for GWAS variants in cell-type-specific regulatory regions would identify the causal cellular context for disease. For example, GREGOR [85] and GoShifter [86] test for overrepresentation of GWAS significant variants in functional annotations. RolyPoly [87], fGWAS [88] and GARFIELD [89] test for larger effect sizes of GWAS variants in regulatory regions.

Enrichment analyses using full GWAS summary statistics instead of the subset of statistically significant variants has been increasingly popular. It is clear that for most polygenetic traits, the genetic predisposition is shared across thousands of common variants with individually modest effect [90]. Under this model, a significant proportion of common variants will have some spurious association signals purely due to LD with one or a few causal variants, making the GWAS signals correlated with LD-dependent genetic architectures within tested populations. Conversely, the extent to which the GWAS statistics are inflated due to LD infers the general heritability of the disease, and can be estimated using tools like LDAK [91] or LDSC [92]. By dividing the genome into different functional categories information, these heritability-based analysis can be used to infer the cell-type-specific contribution for complex diseases. For instance, variants in enhancers in cells from the central nervous system have been shown to explain more heritability of educational attainment and smoking behaviour compared to variants in enhancers active in other cell types [93].

1.5. Translating GWAS to function

Identifying causal genes from GWAS results is not straightforward, as the majority of risk loci fall into non-coding regions of the genome [94] which do not directly influence gene products. Apart from simply assigning risk loci to their nearest genes, the most common approach to interpret the functional consequences of GWAS hits is through integration with quantitative trait loci (QTL) for molecular markers, especially gene expression [95,96]. However, it is estimated that almost half of common variants are associated with gene expression in at least one tissue [97], suggesting that many GWAS-QTL overlaps happen simply by chance. A number of sophisticated methods have been designed to pinpoint

the systematic colocalization of GWAS and QTL signals in gene region of interest [98-100]. Other approaches based on Mendelian-Randomisation provide further insights into the causal relationship between gene expression and complex diseases [101,102]. Although it has successfully identified the causal genes and tissues for multiple CAD risk loci [97,103], the statistical power of colocalization analysis is typically low. In particular, it has been reported that eQTL SNPs only explain about 11% of the disease heritability, partly due to systemic differences between the regulatory effects of eQTL and GWAS-associated variants [104-106]. This observation highlights the need for complementary approaches to identify causal genes [94].

Another promising approach in the search for causal mechanisms lies in the direct profiling of cis-regulatory relationships in DNA elements, in the hope that causal variants residing in these regions share the regulatory effect on the same target gene. This is supported by the observation that disease-associated variants predominantly located in accessible chromatin, and are enriched in the regulatory regions such as promoters and enhancers [85,93]. Thanks to the recent advances in genome editing techniques, the gold-standard functional proof of non-coding DNA elements can be obtained through endogenous knock-down experiments using CRISPR interference [107,108]. A recent paper reported that the influence of distal enhancer on gene expression can be predicted by its enhancer activity and its contact frequency with the promoter of target gene, and established the Activity-by-Contact model to map the regulatory function using bulk epigenomic sequencing data from consortium projects [109]. This genome-wide enhancer map across hundreds of different tissues and cell types were further integrated with GWAS results, and linked 5,036 risk variants across 72 traits to their target genes [110]. For

CAD, these causal genes often do not act in isolation: a large transcriptomic in study in CAD patients reported that networks of co-expressed genes, instead of individual risk genes identified by GWAS, explain most of the CAD heritability [111]. In addition, Perturb-seq in human endothelial cells recently showed the convergence of 43 CAD risk genes onto 5 co-expressed gene programmes [112].

In summary, connecting genetic variants to function requires high-resolution molecular datasets spanning disease-relevant cell types and states, coupled with statistical methods and experimental systems for rigorous testing. While the specific genes, molecular pathways, and cellular functions underlying the genetic risks of CAD remain to be fully elucidated, advances in polygenic modelling, multi-omics profiling and genome editing promise to help bridge the gap from GWAS to mechanism.

1.6. Aims and objectives

The atherogenicity of ox-LDL is well-documented, yet the transcriptional regulation underlying the cellular response to ox-LDL and its interaction with the genetic component of atherosclerosis remain largely unexplored. The present study aims to address these two areas: first, by employing multi-omics sequencing approaches on single-cell and bulk samples to examine the regulatory elements and networks that drive the response to ox-LDL; and secondly, by integrating the resulting data with GWAS signals to uncover novel cellular and molecular mechanisms of disease heritability in atherosclerosis.

Chapter 2 provides a detailed description of all the computational and experimental methods applied in this study. In chapter 3, single-cell RNA-seq and single-cell multiome sequencing are utilised to recover cis-regulatory elements (CREs) and transcriptional

regulatory networks in primary human macrophages. Chapter 4 features a lipid-handling macrophage subpopulation with an increased genetic risk for CAD, identified through a novel computational pipeline that integrates macrophage CREs and GWAS data. Chapter 5 delves into the molecular mechanisms for prioritised variants, achieved through the integration of functional annotations and experimental validation. In chapter 6, the effect of ox-LDL on primary human aortic endothelial cells is systemically profiled in bulk, which aids in the prioritisation of endothelial-specific causal variants for CAD. Finally, chapter 7 summarises the findings presented in this thesis and discusses the direction for future research.

Ch 2. Materials and Methods

2.1. Cell culture

2.1.1. Human monocyte-derived macrophages

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.

Human adult volunteers recruited in this study were mixed gender healthy individuals residing in the UK. Donors with ongoing medical problems such as diabetes or acute illness were excluded from the study and no other healthcare data was collected from these healthy volunteers.

Human peripheral blood mononuclear cells (PBMCs) were isolated by overlaying peripheral blood over Ficoll-Paque Plus solution (Cytiva, Marlborough, US) following the manufacturer's protocol. $CD14^+$ cells were then isolated using magnetic microbeads (Milltenyi Biotec, Bergisch Gladbach, Germany) and differentiated to macrophages in R10 medium (RPMI 1640 medium supplemented with 10% foetal bovine serum, 2mM L-glutamine, 100 units/mL penicillin and 100 ug/mL streptavidin) containing 50ng/mL human macrophage colony stimulating factor (M-CSF) (eBioscience, San Diego, US) for 7-9 days. The medium was replaced at day 3 post-isolation, and every other day afterwards. When necessary, cells were lifted with gentle scraping after 30 minutes of Accutase treatment (STEMCELL, Vancouver, Canada) and sub-cultured at desired densities. Unless otherwise stated, all cells were maintained in a humidified tissue culture incubator at 37°C and 5% CO2.

2.1.2. Donor genotyping

Genomic DNA was extracted from PBMCs using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. Genotyping assays were performed on a QuantStudio 6 machine (ThermoFisher Scientific, Waltham, US) with corresponding rhAmp Genotyping Master Mix and Reporter mixes (Integrated DNA technologies, Coralville, US).

2.1.3. Human aortic endothelial cells (HAECs)

HAECs were purchased (Promocell, Heidelberg, Germany) and expanded at passage 3 before usage. HAECs were maintained in Endothelial Cell Growth Medium MV2 (Promocell, Heidelberg, Germany) supplemented with 100 units/mL penicillin and 100 ug/mL streptomycin. At ~80% confluency, cells were detached with either Typsin/EDTA or Accutase solution and sub-cultured between passage 4 to passage 7. The catalogue numbers for the HAEC lines in this thesis are: #4082102.16 (donor-1), #422Z037 (donor-2), #431Z002.3 (donor-3), #452Z031.1 (donor-4).

2.1.4. Immortalised cell lines

THP-1 cells were obtained from ATCC (TIB-202, ATCC, Manassas, US) and cultured in R10 medium (as described above) with 50 uM of β-mercaptoethanol. HEK293T cells were obtained from ATCC (CRL-3216, ATCC) and cultured in DMEM medium supplemented with 10% foetal bovine serum, 2mM L-glutamine, 100 units/mL penicillin and 100 ug/mL streptavidin.

2.2. Cell treatments

2.2.1. Preparation of LDL and DiI-LDL

Undiluted human plasma was collected during the isolation of PBMCs, centrifuged again at 2,500xg for 30 minutes to remove platelets and any remaining cells, and density-adjusted by adding 0.325 g KBr per mL of plasma.

Low-density lipoprotein (LDL) cholesterols were isolated by isopycnic ultracentrifugation [113,114]: First, a sterile stock KBr solution with the density of 1.221 g/mL was made by dissolving 17.5 g KBr in 50 mL DPBS (Thermo Scientific); then, two KBr solutions with different densities were made according to Table 2.1:

Table 2.1 Densities of KBr solution for LDL preparation

50mL KBr solution	Density	KBr stock / mL	PBS / mL
Solution I	1.019	2.3	51.4
Solution II	1.063	13	38.5

Afterwards, the discontinuous density gradient was gently established by transferring 7.5 mL of DPBS, Solution I, Solution II and density-adjusted plasma, each underneath the previous layer, using a stainless blunt needle in an Optiseal ultracentrifuge tube (Beckman Coulter, Brea, CA). Immediately after balancing, the ultracentrifuge tubes were centrifuged in a Beckman L5-50 ultracentrifuge with a SW28 rotor at 28,000 rpm for 24 hours at 4ºC, with the brake disabled in the deceleration step.

After ultracentrifugation, the LDL layer of density 1.019 to 1.063 g/mL (between Solution I and II) was carefully extracted by piercing the tube with a needle, and sterile filtered

through a 0.2 um filter. At this point, LDLs from different biological donors were pooled to minimize the batch effect due to inter-individual variation, and were dialysed against 200X DPBS to remove residual KBr in a sterile 10KMWCO Slide-A-Lyzer Dialysis Cassette (ThermoFisher Scientific). The concentration of LDL was determined using the BCA protein assay (ThermoFisher Scientific) according to the manufacturer's protocol.

A fluorescent dye 3,3'-dioctadecylindocarbocyanine (DiI) was used to label LDL for confocal imaging and flow cytometry. Dialysed LDL was incubated with 300 ug DiI (Invitrogen, Carlsbad, US) per mg LDL at 37ºC overnight and purified by isopycnic ultracentrifugation. To prevent aggregation and remove excessive dye, DiI-LDL was filtered again through a 0.2 um filterer before dialysis against 200x DPBS.

2.2.2. Preparation of ox-LDL and DiI-ox-LDL

Following the isolation, $CuCl_2$ was added to LDL or DiI-labelled LDL to a final concentration of 25 uM to allow an overnight oxidation at 37ºC. A change in colour (from yellow to colourless for ox-LDL, and from dark pink to pale pink for DiI-ox-LDL) should be observed within 1~2 hours upon successful oxidation. The oxidation was terminated by addition of EDTA to a concentration of 0.5 mM. EDTA-chelated copper was removed through dialysis against 200x DPBS. DPBS solution from the last round of dialysis was sterile filtered and used as the control buffer.

The concentration of ox-LDL / DiI-ox-LDL was determined using the BCA protein assay. Unless otherwise specified, a final concentration of 50 ug/mL for ox-LDL or an equal volume of control buffer was used to treat the cells for up to 48 hours.

2.2.3. Plasmid Nucleofection

Transfection of primary human macrophages was performed using a Nucleofector 2b device (Lonza, Basel, Switzerland). Macrophages were detached and resuspended to a final density of 7x106 cells/mL in 100 uL Nucleofector Solution per transfection (Human Macrophage Nucleofector Kit, Lonza). For dual luciferase assay, 4 ug of pGL4.27 and 0.08 ug of pRL-SV40 plasmids were used per transfection; For CEBPB overexpression, 1.75 ug of pcDNA3 plasmids were used per transfection. A transfection programme of Y-010 was selected and the transfection was performed according to the manufacturer's protocol. Macrophages were resuspended in 300uL of pre-warm R10 w/o antibiotics medium and were split into two wells per transfection in a 24-well plate. When necessary, culture media were replaced 24 hours post transfection after gentle PBS washes. The catalogue numbers for plasmids used in this thesis are provided in Table 2.2.

Table 2.2 Plasmids used in this thesis

ID	Backbone	Insert	Notes
pOC1533	pGL4.27	-	Firefly luciferase
pOC1527	pGL4.27	COC9703	Rs10488763-150bp-ALT
pOC1528	pGL4.27	COC9704	Rs10488763-150bp-REF
pOC1529	pGL4.27	COC9803	Rs10488763-800bp-ALT
pOC1530	pGL4.27	COC9804	Rs10488763-800bp-REF
pOC1534	pRL-SV40	-	Renilla luciferase
pOC167	pcDNA3.1	-	
pOC1252	pcDNA3.1	LAP	Cloned by Dr. Michael Reschen
pOC1531	pGL4.27	Rs62172376-150bp-ALT	Cloned by GenScript
pOC1532	pGL4.27	Rs62172376-150bp-REF	Cloned by GenScript

2.2.4. RNA Interference (RNAi)

All RNAi reagents were purchased as part of the anti-FDX1 TriFECTa RNAi Kit (Integrated DNA Technologies), including three different Dicer-substrate small interfering RNAs (Dsi-RNA) targeting FDX1 and one scrambled negative control Dsi-RNA. Differentiated macrophages were seeded at a density of 60,000 per well into an 8-well ibiTreat u-slide (Ibidi, Gräfelfing, Germany) one day prior to the transfection. A final concentration of 25nM in opti-MEM was used for all Dsi-RNA (TriFECTa DsiRNA kit, Integrated DNA technologies), and was mixed with an equal volume of 1:80 diluted DharmaFECT 3 reagent (Horizon Discovery, Lafayette, US) at room temperature for 20 minutes. Cells were incubated with a final volume of 50uL transfection mix in 200 uL R10 w/o antibiotics for 48 hours. Fresh medium containing DiI-labelled ox-LDL or control buffer was added to each well on top of the transfection mix for another 24 hours before immunofluorescence imaging.

2.3. Sequencing sample preparation

2.3.1. HAEC RNA sequencing

Total RNA from HAECs was purified using RNeasy Micro Kit (Qiagen) after TRIzol and chloroform extraction according to the manufacturer's protocol. The quality of RNA samples was analysed using a TapeStation 2200 with High Sensitivity RNA ScreenTape (Agilent Technologies, Santa Clara, US). Samples with high integrity (RIN score>9) were submitted directly to the Oxford Genomics Centre for subsequent PolyA-enrichment and TruSeq library preparation. For each sample, 1 ug of total RNA was submitted and the

final library was sequenced on a NovaSeq 6000 device (Illumina, San Diego, US) with a target depth of 37.5 million pairs of 150 bp reads.

2.3.2. HAEC ATAC sequencing

ATAC-seq libraries were prepared according to the Omni-ATAC protocol [115]. In brief, HAECs were detached at ~80% confluency using Trypsin/EDTA or Accutase and resuspended in ice cold PBS. Nuclei were isolated by incubation with ATAC-Resuspension Buffer (ATAC-RSB) containing 0.1% NP40 (Roche, Basel, Switzerland), 0.1% Tween-20 (Roche, Basel, Switzerland) and 0.01% Digitonin (Promega, Madison, WI) on ice for 3 minutes, and were washed with 1 mL of ATAC-RSB containing 0.1% Tween-20. For each Tn5 tagmentation reaction, 50 µL of transposition mix was prepared from 25 µL 2X Tagment DNA Buffer and 2.5 µL Tagment DNA Enzyme 1 (Tn5 transposase) from the Nextera kit (Illumina), as well as 0.1% Tween-20 and 0.01% Digitonin. Nuclei were resuspended in transposition mix and mixed thoroughly by gently pipetting up and down 6 times, followed by incubation at 37°C for 30 minutes.

After tagmentation, transposed DNA fragments were purified using a MinElute PCR Purification Kit (Qiagen) and PCR amplified using NEBNext High-Fidelity PCR master mix (New England Biolabs, Ipswich, US) and index primers from Nextera Kit (Illumina). 10% of the pre-amplification mix was used for a pilot qPCR experiment to determine the optimal number of PCR cycles where one-third of the maximum fluorescent intensity was met. The quality of DNA samples was analysed using a TapeStation 2200 with High Sensitivity D1000 ScreenTape (Agilent), with successful transposition indicated by fragments piling up with a nucleosome-free peak at around 25 bp and larger peaks

between 160~200 bp apart. Indexed libraries were quantified by qPCR using a KAPA Library Quantification kit (Roche) and pooled in equimolar amounts for sequencing with the Oxford Genomics Centre. Libraries were sequenced on a NovaSeq 6000 device (Illumina with the target depth of 112.5 M pairs of 150 bp reads per sample.

2.3.3. Macrophage single-cell RNA sequencing

Differentiated human macrophages were lifted by trypsinisation and washed gently in fresh PBS. Samples with ~90% viability were submitted directly to the Oxford Genomics Centre for preparation of single-cell RNA-seq libraries. Barcoded sequencing libraries of macrophages treated with ox-LDL or control buffer (n=4 biological donors) were generated from single cells using 10X Genomics Chromium system with 3' chemistry (v2; 10x Genomics, Pleasanton, US) according to the manufacturer's instructions and sequenced on the Illumina HiSeq4000 platform with a target depth of at least 50M pairs per sample (approximately 0.1M pairs per single cell).

2.3.4. Macrophage single-cell multiome sequencing

Sequencing libraries from macrophages treated with ox-LDL or buffer (n=4 donors) were generated from single nuclei using 10X Genomics Multiome system according to the manufacturer's instructions at the Oxford Genomics Centre. Each sample was loaded onto a separate channel of a 10X chip with the target capture of 500 cells. Barcoded libraries for both gene expression and ATAC were sequenced on the Illumina NovaSeq6000 platform.

2.4. Bulk sequencing data analysis

Unless otherwise specified, all sequencing analyses were performed on the high-performance computing clusters of the Biomedical Research Computing (BMRC) facility. For all bulk sequencing libraries, read qualities were checked using FastQC [116]. When sequencing adaptors were detected (e.g., fragments were shorter than 150bp so sequencing reads extended to the adaptor region), the 3' end of affected reads was trimmed using NGMerge [117].

2.4.1. RNA-seq

For bulk RNA-seq analysis, adaptor-trimmed reads were first aligned to the reference human genome using the splice-aware STAR aligner [118]. Human genome assembly GRCh38.p13 and gene annotation Release 43 were downloaded from the GENCODE site and were indexed using STAR. For standard RNA-seq differential gene expression analysis, one-pass alignment with options of –outFilterMultimapNmax 10, --alignSJoverhangMin 5, --alignSJDBoverhangMin 3, --outFilterMismatchNmax 10, --outFilterMismatchNoverLmax 0.04, --alignIntronMin 21, --alignIntronMax 0, --alignMatesGapMax 0, --outFilterIntronMotifs RemoveNoncanonical was performed. For RNA-based genotyping, a second-pass alignment was performed with default parameters after the first pass.

Post sequencing alignment, only uniquely mapping pair-end reads were retained using SAMtools [119], and gene-wise count matrices were generated using featureCounts from the command line or the R interface of the Subread package [120]. The count matrix were subsequently imported as a DGEList object in R and analysed using edgeR [121]. Raw read

counts were normalized by sequencing depth and log1p transformed. Genes with low expression (min.count<50 in more than half of the samples) were filtered out, and the remaining count matrices were further normalized by Trimmed Mean of M-values (TMM) method implemented in the edgeR package to compensate for the differing library compositions. Differential gene expression between treatment groups was tested using the quasi-likelihood test, with a generalized linear model controlling for experimental batches and biological donors. The test statistics were adjusted using the Benjamini-Hochberg (B-H) method and genes with adjusted p-values <0.05 were considered as differentially expressed. Pathway enrichment analyses were performed using the XGR package [122], with functional annotations collected from Gene Ontology, KEGG, and Reactome.

2.4.2. ATAC-seq

Adaptor-trimmed reads were aligned to the hg38 reference genome using Bowtie2 with --very-sensitive settings [123]. PCR duplicates were marked using Picard, and were further removed together with mitochondrial reads using SAMtools. Post-alignment QC including PCR Bottleneck Coefficients (PBC), fragment size distribution, and Transcription Start Site Enrichment (TSSE) was performed as instructed by the ENCODE ATAC-seq processing pipeline. More specifically, a Non-Redundant Fraction (NRF) >0.9 with PBC1>0.9, PBC2>3 indicates an ideal library complexity, and a TSSE score >7 with hg38 annotation indicates high ATAC-seq quality.

After alignment and post-alignment QC, sequencing reads were further trimmed using BEDTools [124] to retain only the 9-bp Tn5 cutting sites at both ends of the DNA fragments.

This step is required for high-resolution open chromatin mapping as only the 5' and 3' ends of any transposed DNA fragments are cut by the Tn5 transposase and thus truly accessible. The 9-bp reads from different biological replicates were merged with sub-sampling in SAMtools to generate a consensus bam file per treatment group. Open chromatin peaks were then called on these consensus bam files using MACS3 [125] with options --qvalue 1e-5, --nomodel, --shift -46, --extsize 100, --broad, which extends 50bp to both ends and re-centres DNA fragments to the cutting sites. Peaks overlapping the ENCODE blacklisted region were removed before read counting in Subread. Consensus peaks from buffer-treated and ox-LDL-treated samples that were less than 50bp apart were merged using BEDtools. Raw count matrices were imported as DGEList objects in R and analysed using edgeR, as described above in the RNA-seq differential expression analysis. A window-based differential accessibility test was also performed by counting reads within 100bp sliding windows across the genome in csaw [126].

In-silico deletion (ISD) tests were performed using the "FromCoverage" function of the Lisa package [127] with 500 background genes and the default 'enhanced_10k' regulatory potential model. The consensus bam file of ox-LDL-exposed samples were used for all ISD analysis. For the ISD of ox-LDL-specific CEBPB bindings, fragments overlapping ox-LDL-induced CEBPB binding sites were filtered out using BEDTools, and the gene-wise regulatory potentials were calculated using the "genescore" function implemented through the MAESTRO package[128]; empirical null distribution was estimated from n=1000 sets of expression-matched background genes (i.e., genes randomly drawn from n=20 quantiles of all expressed genes, with the number of genes per quantile matched for that in the input gene list).

TOBIAS [129] was used to analyse the motif foot printing signatures in ATAC-seq data. In brief, the merged 9-bp consensus bam file was processed using ATACorrect function in TOBIAS to correct for the intrinsic bias in Tn5-cut sites; Footprint scores were calculated at base-pair resolution within open chromatin peaks using ScoreBigwig; differential foot printing was then conducted using BINDetect, with transcription factor (TF) motifs downloaded from JASPAR core 2018 [130]. For ontology enrichment and heritability analysis, differential TF binding sites was extended 500bp at each end; the enrichment for functional annotations was tested using GREAT v4 [131].

2.4.3. ChIP-seq

The macrophage ChIP-seq data on H3K27ac and CEBPB were generated by Dr Michael Reschen [113], and the HAEC ChIP-seq data on H3K27ac was generated by Dr Anil Chalisey [132]. As described above, adaptor-trimmed reads were aligned to the hg38 reference genome using Bowtie2 followed by PCR duplicates removal using Picard. For H3K27ac ChIP-seq in macrophages and HAECs, aligned bam files were merged with subsampling per treatment group. Aggregated ChIP-seq signals were computed over 100bp bins and normalized by sequencing depth using the bamCoverage from DeepTools, with 150bp extension of each fragment. For function annotation of macrophage cis-regulatory elements (CREs), ChIP-seq reads were counted per 100bp bin for each 4kb window centred at each CRE peak, normalized by sequencing depth, and smoothed using a generalized additive model implemented through ggplot2. The heatmap showing the normalized ChIP-seq signals over CREs and background peaks were plotted using the plotHeatmap function from Deeptools. For CEBPB ChIP-seq, a consensus peak set was called using MACS3 with all samples, and peaks overlapping

the ENCODE blacklisted region were removed. Differential binding tests were performed in edgeR, with both donor and treatment included as covariates in the quasi-likelihood test. Peaks with p-values lower than 0.05 were retained (n=8,718) for ISD analysis.

2.5. Single-cell sequencing analysis

The single-cell analyses presented in this thesis are part of our recently published work [133]. The computational codes for reproducing published results are available at https://github.com/jhjiang2020/multiome_paper.

2.5.1. Meta-analysis of human atherosclerotic plaque data

Sequencing data of 38 scRNA-seq datasets on human coronary plaques, carotid plaques, adjacent control artery tissues and PBMCs reported in five studies were retrieved from Gene Expression Omnibus (GEO) using accession numbers GSE131778 [25] GSE155512 [33], GSE159677 [26], and GSE196943 [27], GSE210152 [134]. For samples where the fastq files were provided, I re-aligned the sequencing reads using Cell Ranger (v5.0; 10x Genomics, Pleasanton, CA) against the GRCh38 reference genome. Count matrices were then aggregated in Seurat V4 [135], providing n=142,751 cells.

Low-quality cells were first removed with a threshold of (detected genes <200 | detected reads < 500 | mt% > 20%). Potential doublets were scored using DoubletFinder [136] assuming doublet rates of 5%, leaving n= 126,685 singlets for further analysis. I next performed three rounds of sample integration and clustering analysis, each round with an increased resolution (see Supplementary Notes). For the first two rounds of clustering, the number of retained principal components (PCs) were determined heuristically based on inspection of the variance ratio plots. For subclustering of lipid-associated

macrophages, the number of PCs, clustering resolution, and the k value for the KNN graph were optimized using the scclusteval package.

2.5.2. Single-cell RNA-seq pre-processing

Barcoded sequencing libraries of primary human monocyte-derived macrophages treated with ox-LDL or control buffer (n=4 biological donors) were generated and sequenced as described above. Sequencing reads were aligned using Cell Ranger (v5.0; 10x Genomics) against the GRCh38 reference genome. Count matrices were then aggregated in Seurat V4 [135], providing n=5,140 cells passing the above QC threshold. For clustering analysis, raw count matrices were first normalized using SCTransform, and then integrated using the CCA algorithm. Clustering parameters were optimized by stability analysis (n_PCs = 25, resolution=0.4 and k = 40) using the package scclusteval [137]. Clusters with high levels of nuclear transcripts were discarded, leaving n=4,881 cells.

2.5.3. Integrative analysis between plaque and *ex vivo* macrophages

Anchor-based reference mapping of *ex vivo* macrophages was performed in Seurat. Raw counts of *ex vivo* macrophages were normalized using the SCTransform algorithm and integrated with plaque macrophage atlas to find anchor genes. Using these anchors, *ex vivo* cells were projected onto the reference UMAP, and were labelled with corresponding plaque annotations.

For deep-learning-based macrophage classifier, a semi-supervised variational autoencoder model was trained using scANVI [138] based on n=5000 highly variable genes from the plaque macrophage atlas (n_hidden=128, n_latent=10, dropout_rate=0.1,

max_epochs=100, train_size=0.8). The trained model was then used to classify *ex vivo* macrophages and predict the associated probabilities.

For transcriptome-wide correlation analysis, single-cell count matrices across different macrophage populations were aggregated into pseudo-bulk matrices, and were normalized by sequencing depth and log1p transformed. Pearson correlation coefficient was used to evaluate the transcriptomic similarities between each pair of two macrophage populations.

2.5.4. Single-cell multiome pre-processing

Sequencing libraries from primary human monocyte-derived macrophages treated with ox-LDL or control buffer (n=4 donors) were generated from single nuclei using 10X Genomics Multiome system according to the manufacturer's instructions and sequenced on the Illumina NovaSeq6000 platform. Reads were aligned and counted using Cell Ranger ARC (v2; 10x Genomics) with the GRCh38 reference genome. For snATAC-seq, peaks were trimmed and re-called with MACS3 [125]. Count matrices were processed with Seurat (v4.0.4) and Signac (v1.4.0). Low-quality cells and potential doublets were filtered out with the threshold of (TSSE <1 | detected gene/peak < 1000 | detected gene/peak > 50,000).

For RNA integration, counts were normalized with SCTransform and integrated using Seurat's CCA integration workflow. For ATAC integration, count matrices were normalized in Signac and batch corrected using Harmony. The first 26 RNA reduced dimensions and first 10 ATAC reduced dimensions were selected by inspecting the variance ratio plots for integrated clustering using Seruat's WNN workflow [139]. For all

differential gene expression analyses and visualization, sequencing-depth-normalized gene expression was used, quantified as count-per-10k-reads. Cluster marker genes were identified by a logistic regression test with mitochondrial percentage as a latent variable, and p-values were adjusted using the Benjamini-Hochberg procedure. Pathway enrichment was performed using XGR [122]. Enriched GO terms were reduced and visualized using REVIGO [140].

2.5.5. Transcription factor (TF) motif deviation analysis

After generating background peak sets matching GC content and average accessibility, motif deviation scores were computed using ChromVAR [141]. Human TF motifs were downloaded and manually curated from cisBP, as described in [142].

2.5.6. Cis-regulatory analysis

Peak to gene correlation was calculated as described [139,142]. The peak-gene correlation coefficient was calculated using the mean-centred ATAC-seq signals for a given peak and a corresponding gene expression. Statistical significance of a given peak-gene pair was estimated from a null distribution computed using a set of background peaks matched for GC content and average accessibility. CREs were annotated using the ChIPseeker R package [143].

For Domain-Of-Regulatory-Chromatin (DORC) analysis, genes with at least 9 identified CREs were selected as DORCs. By summarizing the number of reads mapped to the CREs of each DORC gene, the snATAC-seq peak matrix was transformed into a DORC score matrix. The DORC-TF regulatory network was constructed as described [142]. In brief, DORC genes were pooled with their k-nearest neighbours based on the DORC scores

(k=30). The enrichment of each TF motif was computed using the CREs of each DORC gene pool over the matched background peaks. The smoothed RNA expression of each TF tested was correlated with DORC-scores in all cells, and the significance tested using standardized Spearman correlation levels. A regulatory score per DORC-TF pairs was calculated by combining two significance levels (motif enrichment and correlation). Network plots for CAD-DORCs (DORC genes with CREs overlapping CAD-related SNPs) were drawn for the strongest associations (score > 1.5) using the ggnet2 R package.

2.5.7. Quantitative Trait Loci (QTL) SNPs

Expression QTL SNPs were obtained from GTEx v8 (whole blood) and eQTLGen (Blood) database. SNPs at hg19 coordinates were converted to hg38 using annotations from NCBI dbSNP Build 155. Chromatin accessibility QTL SNPs (human iPSC-derived macrophage) were downloaded from supplementary materials at the publisher's website [144].

2.5.8. GWAS-related SNPs

GWAS statistics were curated by NHGRI-EBI Catalog and downloaded from the UCSC hg38 database (April 2022). We identified SNPs with significant genome-wide association ($p<5x10^{-8}$), pruned the list ($r^2 > 0.1$ identified by 1000 Genome Project phase 3 in EUR samples, hg38 build) to remove redundant loci, and then pooled all SNPs in high LD within each locus ($r^2 > 0.8$ in EUR samples) to establish a set of trait-associated variants. Note that due to the presence of multiple statistically independent signals within one risk locus, the total number of pruned loci was expected to be higher than the number reported in

the original GWAS study. Only dbSNP common variants (MAF>1%, dbSNP155) were considered in this study. LD information was extracted using PLINK v1.90b6.26 [145].

2.5.9. Enrichment analysis for functional annotations

We used chromVAR to create background peaks matched for GC content and chromatin accessibility levels across cells for each open chromatin region. For each annotation considered, SNPs/annotated regions were intersected with peaks using the findOverlaps R function, and the null distributions were constructed with n=1000 background peak sets. The significance of each trait-region overlap was determined using Z-statistics computed from the background null distributions.

2.5.10. Heritability analysis using s-LDSC

We stratified cis-regulatory elements for each macrophage subpopulation by applying a logistic regression test for differential expression between subpopulations (one vs the rest), between treatment (ox-LDL vs buffer) of all cells, or between treatment within a subpopulation. The p-values for genes in each test were then transformed to the relative weights (gene scores) as described previously [146]. In brief, gene scores were computed by min-max normalizing X scores, which follow a X2 distribution and were calculated as X = -2 log (p). Each CRE was then weighted by the sum of the scores of their target genes.

LDSC v1.01 was used to partition CAD heritability and test the enrichment for each subpopulation as described [93,147]. In brief, all 1000G SNPs were annotated and weighted by their overlaps with subpopulation CREs or subpopulation-ox-LDL CREs; we then inferred the LD score for each SNP using the 1000G EUR reference panel, and computed the heritability associated with each subpopulation using a multivariate linear regression

model, conditioned on background annotations; the standard errors of enrichment statistics and regression coefficients are computed using a block jack-knife (n=1,000) [93]. The full summary statistics for CAD used in this study were downloaded from the GWAS Catalog, study ID GCST005194 [70].

We evaluated the extent and the significance of heritability enrichment using two metrics: the relative enrichment score E_c over background annotations and the standardized effect size τ^*, as implemented by the LDSC software [147]. E_c follows the normal distribution and is defined as: $E_c = \frac{h^2(C)}{|C|} - \frac{h^2 - h^2(C)}{M - |C|}$, where $h^2(C)$ and $h^2$are the heritability of SNPs in category C and all SNPs respectively, M is the total number of SNPs included, and $|C|$ is the number of SNPs in C. The relative enrichment over background was calculated by comparing E_c to the background enrichment score $E_{background}$ constructed using either all macrophage open chromatin regions, or the pseudo-bulk ox-LDL-response program.

For *ex vivo* macrophage subpopulation programmes and plaque macrophage programmes, we computed τ^* conditioned on all other subpopulations but the subpopulation of interest using a multivariate linear regression model. For *ex vivo* macrophage subpopulation-ox-LDL programmes, we computed the τ^* conditioned on the pseudo-bulk ox-LDL program.

2.6. Immunofluorescence imaging

2.6.1. Fixed cell imaging

For fluorescence microscopy, $CD14^+$ monocytes were differentiated at 100,000 cells/well in 8-well ibiTreat µ-slides (Ibidi) for 7 days, as described above, before treatment with DiI-

labelled ox-LDL or control buffer. HAECs were seeded at 15,000 cells per well in 8-well ibiTreat µ-slides and imaged at ~80% confluency.

For detection of CHI3L1, cells were treated with protein transport inhibitor cocktail (eBioscience) overnight according to the manufacturer's instructions. Treated macrophages and endothelial cells were fixed at room temperature for 15 mins with 4% PFA in PBS, followed by permeabilization with 0.1% Triton X-100 (Thermo Scientific) and 0.2% Tween-20 (Roche) in PBS. Fixed cells were incubated in blocking buffer (1% BSA, 2% FCS, 0.3M glycine and 0.1% Tween-20 in PBS) at room temperature for 1 hour. Primary antibody staining was carried out at 4°C overnight with the following antibody dilutions in blocking buffer: CD52-APC (clone HI186, BioLegend) 1:200, CHI3L1 (MM0628-9X24, Abcam) 1:500, MITF (HPA003259, Atlas Antibodies) 1:100, CD36 (clone NL07, eBioscience) 1:1000, tubulin-AF488 (clone DM1A, eBioscience) 1:500 and CEBPB (sc-7962, Santa Cruz Biotechnology) 1:100. For detection of the CHI3L1, MITF, and CD36/CEBPB primary antibodies, cells were incubated with FITC anti-rat IgG (Sigma-Aldrich), AF633 anti-rabbit IgG (Invitrogen), and AF647 anti-mouse IgM (Invitrogen) respectively, at 1:1000 in blocking buffer in the dark at room temperature for 1 hour. Cells were washed three times with PBS-0.1% Tween-20 for 5 mins between stains. Nuclei were counterstained with Hoechst 33342 (Santa Cruz Biotechnology) diluted 1:10,000 in PBS. Stained cells were imaged on a LSM 900 with Airyscan2 in confocal mode (Zeiss). Quantitative analysis was performed using arivis Vision4D version 4.1 (Zeiss).

2.6.2. Live cell lipid uptake assay

Differentiated macrophages were seeded at 100,000 cells/well in 24-well ibiTreat µ-Plate (Ibidi) 24 hours prior to the assay and were stained with 1 ug/mL Fc-muted CD52 antibody (LT205, Leinco Technologies, Fenton, MO) labelled with Zip Alexa Fluor 488 Antibody Labelling Kit (Z11233, Thermo Scientific) and 50 nM MemGlow-640 (Cytoskeleton, Denver, CO) at room temperature for 1 hour prior to the assay. Live cells were imaged using the Olympus SpinSR SoRa and ScanR High Content system (Olympus LS, Tokyo, Japan) at 5-minute intervals for 6 hours in assay media containing 50nM MemGlow-640 (Cytoskeleton) and 50 ug/mL DiI-labelled ox-LDL.

2.6.3. Endothelial migration assay

Cell culture plates with 2-well silicone inserts (80241, Ibidi) were used to create a 500um cell-free gap for uniform imaging of endothelial migration. Prior to the assay, HAECs were seeded at the density of 10,000 cells per culture insert well and were cultured until reaching ~80% confluency. Cells were then starved in supplement-depleted EC culture medium containing 50ug/mL ox-LDL or control buffer overnight, and were pre-stained with 300 nM SiR-DNA for at least 2 hours. Right before live-cell confocal imaging on the Olympus SpinSR SoRa and ScanR High Content system (Olympus), culture inserts were carefully lifted using a sterile tweezer, and cells were gently washed with sterile PBS twice. Live cells were then imaged every 18 minutes for 24 hours in assay media containing 300 nM SiR-DNA and 50 ug/mL ox-LDL or equivalent volume of control buffer. The location of nuclei was recovered using arivis Vision4D version 4.1 (Zeiss), and the migration rate was defined as the change of the percentage of nuclei overlapping the initial gap region.

2.7. Flow cytometry

Cells were detached using either Typsin/EDTA or Accutase solution and re-suspended in ice-cold FACS buffer (PBS, 1% BSA and optional 1% azide). For macrophages, 10M cells were re-suspended in 80uL buffer and were incubated with 20uL of FcR Blocking Reagent (Miltenyi Biotec, Bergisch-Gladbach, Germany) at 4ºC for 10min. Cells were adjusted to a density of 100,000 to 500,000 cells per well in a 96-well U-bottom plate, and were pelleted at 1,500 rpm at 4ºC for 5min. For primary staining, cell pellets were resuspended in 100uL of primary antibody solution and were stained in the dark at 4ºC for 30min. Cells were then washed with ice-cold FACS buffer for three times as described previously. When secondary staining was required, cells were incubated with secondary antibody solution in the dark at 4ºC for 30min, washed and resuspended in FACS buffer for flow cytometry analysis. Data was exported to FlowJo 10 (FlowJo LLC, Ashland, USA) for quantitative analysis. The antibody dilution used is described in Table 2.3:

Table 2.3 Antibody dilutions for flow cytometry

Antibody/Dye	Supplier	Catalogue Number	Dilution
Anti-ICAM-1(CD54)	BioLegend	353102	1:200
Anti-VCAM-1 (CD106)	Invitrogen	14-1069-82	1:500
Anti-Mouse-IgG-647	Invitrogen	A-21240	1:1000
Zombie Violet	BioLegend	423113	1:200

2.8. RT-qPCR

Total RNA was extracted from TRIzol-lysed samples using RNeasy Micro Kit (Qiagen) and was kept on ice before reverse-transcription (RT) to cDNA. For each RT reaction, 0.2 to 1ug of RNA in 10uL RNase-free water, 1uL of 10mM dNTP mix, 0.5uL of 0.4mg/mL random hexamer and 0.5uL of RNase-free water were added to the PCR plate/tube. Then, the sealed plate/tube was incubated at 70ºC for 5min followed by 2min on ice. After the initial denaturation, 8uL of RT master mix containing 4uL 5x RT buffer, 2uL H2O, 1uL RNase-inhibitor (40U/uL) and 1uL Tetro Reverse transcriptase (20U/uL; Meridian Bioscience, Cincinnati, US) was added to the reaction. The mixture was incubated at 25ºC for 10min, 45ºC for 1h and 85ºC for 5min. The tube was pulse spined between incubations to ensure through mixing and to prevent change of reaction volume caused by evaporation. The reaction mix was chilled on ice for 10min before qPCR. Alternatively, samples can be stored at -20ºC upon successful RT reaction.

For qPCR, 4uL of cDNA from previous steps (up to 200ng of total RNA input) was mixed with 1uL of 2uM primer set (1:50 diluted forward and reverse primers from the 100uM stocks) and 5uL of 2x SYBR master mix (Applied Biosystems, Foster City, US) in a 384-well qPCR plate. After sealing and pulse spin at 4ºC, samples were analysed by a QuantStudio 6 machine (ThermoFisher) following the fast qPCR-384 protocol. GAPDH was used as the internal reference to control for variations in template amount and RT efficiency. The primers are stored in the COC oligonucleotide archive with accession number and sequence as the following:

Table 2.4 PCR primers used in this thesis

Target	Direction	Oligo_ID	Sequence
CEBPB	forward	CO5472	GCCAAGAAGACCGTGGACAA
CEBPB	reverse	CO5473	TGCGCACGGCGATGT
FDX1	forward	CO9678	GTGATTCTCTGCTAGATGTTGTGG
FDX1	reverse	CO9679	CCATATGCCAGATCGAGCATGTC
GAPDH	forward	CO8063	AATGAAGGGGTCATTGATGG
GAPDH	reverse	CO8064	AAGGTGAAGGTCGGAGTCAA

2.9. Luciferase assay

Two dsDNA constructs were synthesized (IDT gBlocks™, Integrated DNA technologies) for each allele of the rs10488763, one covering -75bp to +75bp, and one covering -100bp to +700bp of the targeted variant. DsDNA inserts were cloned into empty pGL4.27 Firefly luciferase reporter (Promega, Fitchburg, WI), and successful cloning was verified by Sanger sequencing prior to transfection. For the CALCRL variants rs77376996 and rs62172376, 150bp dsDNA constructs of each haplotype were synthesized and cloned into pGL4.27 by GenScript (GenScript Biotech, Piscataway, US).

For dual luciferase assay, pGL4.27 Firefly luciferase reporter plasmid and pRL-SV40 Renilla luciferase plasmid were transfected in a 50:1 ratio and cells were lysed 24 or 48 hours post transfection using the Dual Luciferase Reporter Assay System (Promega). Luciferase activities were measured using a CLARIOstar Microplate Reader (BMG Labtech, Ortenberg, Germany), and Firefly luciferase activity was normalized to Renilla luciferase activity as relative light units.

For transfection of THP-1 and HEK293T cells, 500ng of total plasmid DNA per 100k cells was used with 3uL of ViaFect Transfection Reagent (Promega) following the manufacturer's protocol. pcDNA3.1 plasmid containing the activating isoform of *CEBPB* (LAP) was co-transfected in a 1:4 ratio with the luciferase plasmids. For primary human macrophages, 700,000 cells were transfected with 4ug of pGL4.27 and 80ng of pRL-SV40 per transfection using the Human Macrophage Nucleofector kit (Amaxa, Lonza, Portsmouth, NH). Transfected cells were split into two wells in R10 w/o antibiotics and were allowed to recover for 24 hours cells before treatment with 50 ug/ml ox-LDL or buffer for a further 24 hours.

2.10. Seahorse assay

Metabolic measurements of HAEC were performed with Seahorse XF Mito Stress Test and XF Glycolytic Rate Assay kits using a Seahorse XFe96 Analyzer (Agilent Technologies, Santa Clara, US). 24 hours prior to the assay, 50,000 endothelial cells were seeded into the Seahorse 96-well plates. For mito stress test, 2uM of Oligomycin, 1uM of FCCP and 0.5uM of Rotenone/Antimycin A were used based on a pilot titration experiment; For glycolytic rate, 0.5uM of Rotenone/Antimycin A and 50mM of 2-DG were used. Results were analysed using Seahorse Wave Desktop (Agilent) and were manually inspected to exclude spurious readings due to air bubbles or other technical issues.

Ch 3. Single-cell multi-omics sequencing identifies cis-regulatory elements and recovers transcriptional regulatory networks associated with CAD variants

Collaboration Notes: The work described this chapter was part of a research project in collaboration with Dr. Thomas K. Hiron who was a postdoctoral researcher in the O'Callaghan lab at the time. Dr. Hiron acquired the original single-cell RNA-seq and single-cell multiome data, and was responsible for most of the phlebotomy and tissue culture during macrophage isolation. All computational analyses presented were performed solely by me. This work has been published in:

Jiang J, Hiron TK, Agbaedeng T, Malhotra Y, Drydale E, Bancroft J, Ng E, Reschen ME, Davison LJ, O'Callaghan CA. A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease. *Circ Res.* 2024 May 15. doi: 10.1161/CIRCRESAHA.123.324172. PMID: 38747151.

3.1. Introduction

Monocytes and macrophages of the innate immune system represent the first line of host defence and are involved in maintaining tissue homeostasis even in the absence of intruding pathogens. Cells of this lineage often display remarkably plastic responses to their microenvironment [148,149], with two extensively studied polarised states M1 and M2 representing pro- and anti-inflammatory phenotypes [23]. While early studies reported macrophage functions based on this established M1/M2 paradigm, modern views are shifted towards a model where macrophage subpopulations coexist *in vivo* with distinct functions and origins [150-152].

As reviewed in 1.1.2, monocyte and macrophages play a crucial role in the initiation and progression of atherosclerosis via engulfment of modified cholesterols retained in the arterial intima [61]. Within atherosclerotic plaques, excessive oxidised lipoprotein (ox-LDL) result in the imbalanced cholesterol influx, esterification and efflux in macrophages, which

subsequently drives them into a lipid-laden phenotype commonly recognised as "foam cells" [17-19] [153]. While recent advances in single-cell sequencing offer powerful tools to investigate the disease-relevant macrophage populations in atherosclerotic plaques[24-27], the lesions profiled are typically in the late stages of atherogenesis. This means that the investigation of individual pro-atherogenic stimulus is confounded by the complicated microenvironment where contradicting signals from modified cholesterols, cellular necrosis and immune activation act concomitantly to influence the cells[18].

Sharing the origin with most macrophages in the atherosclerotic lesions [154], human macrophages derived from circulating monocytes are extensively used to study the molecular basis of atherosclerosis, including the response to ox-LDL [113,155-157]. For *in vitro* studies, macrophages are typically differentiated in the presence of macrophage colony-stimulating factor (M-CSF) [158], which some considered to be in a non-activating resting state with anti-inflammatory characteristics [159,160].

The macrophages' response to ox-LDL is fundamental to atherosclerosis, our group has well established the experimental system to study this cellular response. Using M-CSF-derived macrophages, we have previously examined the dose and time-dependent response to ox-LDL (unpublished data generated by Dr Anil Chalisey [132]), and have confirmed the lipid-laden phenotype in macrophages exposed to 50ug/mL of ox-LDL for 48 hours [113].

In culture, monocyte-derived macrophages exhibit remarkable heterogeneity in cell morphology (Figure 3.1A), and they display various lipid-accumulation and distribution patterns after ox-LDL treatment (Figure 3.1B). Therefore, we hypothesised that the

response to ox-LDL may differ across heterogeneous subpopulations. Gaining a deeper understanding of these heterogenous responses could help to identify new potential therapeutic targets.

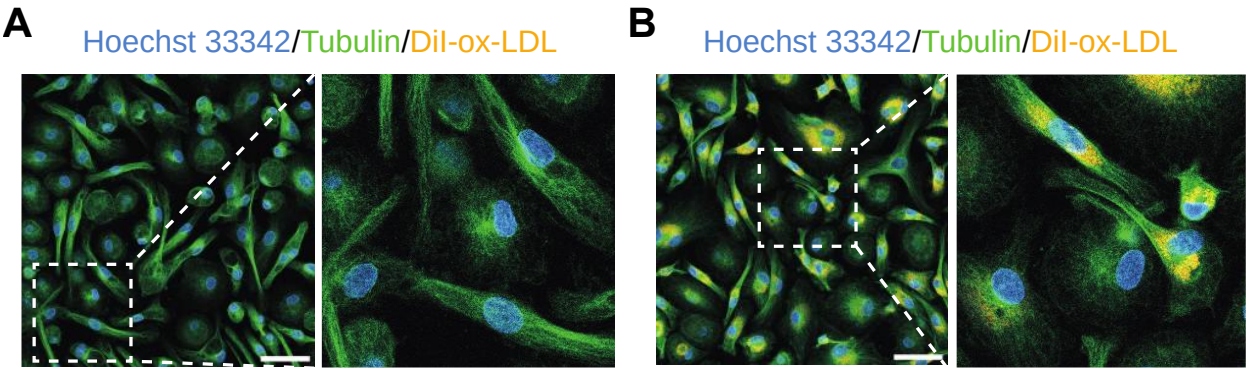


Figure 3.1 Representative images from immunofluorescence staining of primary human macrophages.
Cells were treated with control buffer (**A**) or 50ug/mL of DiI-labelled ox-LDL (**B**) for 24 hours before imaging, showing the intracellular accumulation of lipid droplets. Images were obtained by Dr. Thomas Hiron. Scale bar: 50um.

In this chapter, I aim to explore the gene regulation dynamics associated with ox-LDL exposure in macrophages at the single-cell resolution. Through single-cell RNA and single-cell multiome sequencing, I first describe the heterogenous macrophage subpopulations and their transcriptomic response to ox-LDL. The regulatory DNA elements underlying both the naïve un-treated state and the ox-LDL-exposed state are then identified from integration of the gene expression and chromatin accessibility of single cells. In addition, I recovered the transcriptional regulatory networks underlying the genetic risk genes for coronary artery disease (CAD), and demonstrate the central role of AP-1 family factors.

3.2. Clustering and integration of single-cell multiome data

We have previously characterised the macrophage response to ox-LDL in bulk samples, and have identified major alterations in open chromatin and enhancer activity within lipid-laden foam cells [113]. Given the remarkable plasticity of macrophages and their rapid adaptation to external stimulus, we hypothesized that monocyte-derived macrophages may comprise distinct subpopulations with heterogenous responses to ox-LDL. In a pilot experiment, we collected single-cell gene expression profiles using single-cell RNA-seq (scRNA-seq), and defined macrophage subpopulations in both ox-LDL-exposed and unexposed macrophages (unpublished data generated by Dr Thomas Hiron [161]).

Due to technical limitations, the expression profiles from the single-cell multiome data were generated from nuclear rather than whole-cell transcripts. As pre-RNA requires splicing before it can be transported to the cytoplasm, single-nucleus RNA-seq (snRNA-seq) data inevitably contains a higher proportion of nascent and immature RNA compared to single-cell RNA-seq data. To ensure our single-cell multiome data robustly capture the transcriptomic diversity of monocyte-derived macrophages, I re-analysed the scRNA-seq data and compared the results with the snRNA-seq modality in the multiome data.

3.2.1. Re-analysing the single-cell RNA-seq data

To avoid spurious clusters due to over-clustering, I employed an automated pipeline with cluster stability tests (See Ch.2 for details). After removing low-quality cells (detected genes <200 | detected reads < 500 | mt% >20%), 5,140 cells were retained. After batch-effect correction, a shared nearest-neighbour graph was constructed by calculating the neighbourhood overlap (shared nearest-neighbourhood, SNN) between every cell and its

k=40 nearest neighbours using the first 25 principal components. Further clustering at a resolution of 0.4 revealed 9 transcriptomically distinct clusters, including a spurious cluster with high nuclear transcript (possibly damaged cells with intact nuclei) and a small T-cell cluster (possibly from inefficient $CD14^{+}$ selection during blood prep). After post-clustering QC, 4,881 cells from the macrophage clusters were retained, representing two cycling populations, a $FUCA1^{hi}$ population, a $PCSK6^{hi}$ population, an MHC^{hi} population, an M1-like population and a lipid-handling population (Figure 3.2). These clusters were named by their marker genes (Figure 3.3), and a detailed description of each subpopulation will be provided later in the chapter with additional epigenetic information from the multiome data.

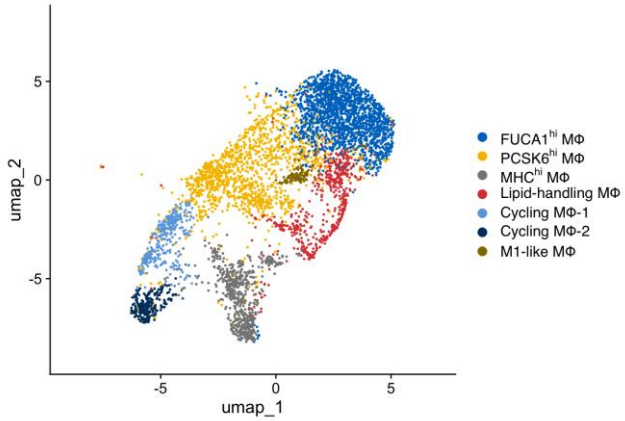


Figure 3.2 Uniform Manifold Approximation and Projection (UMAP) visualisation of the single-cell RNA-seq data.
Primary human macrophages were treated with or without 50 ug/mL of ox-LDL and were collected for single-cell RNA-sequencing. Cells (n=4,881) are coloured by subpopulation inferred from un-supervised clustering based on the transcriptomic similarity, and projected on the UMAP embeddings.

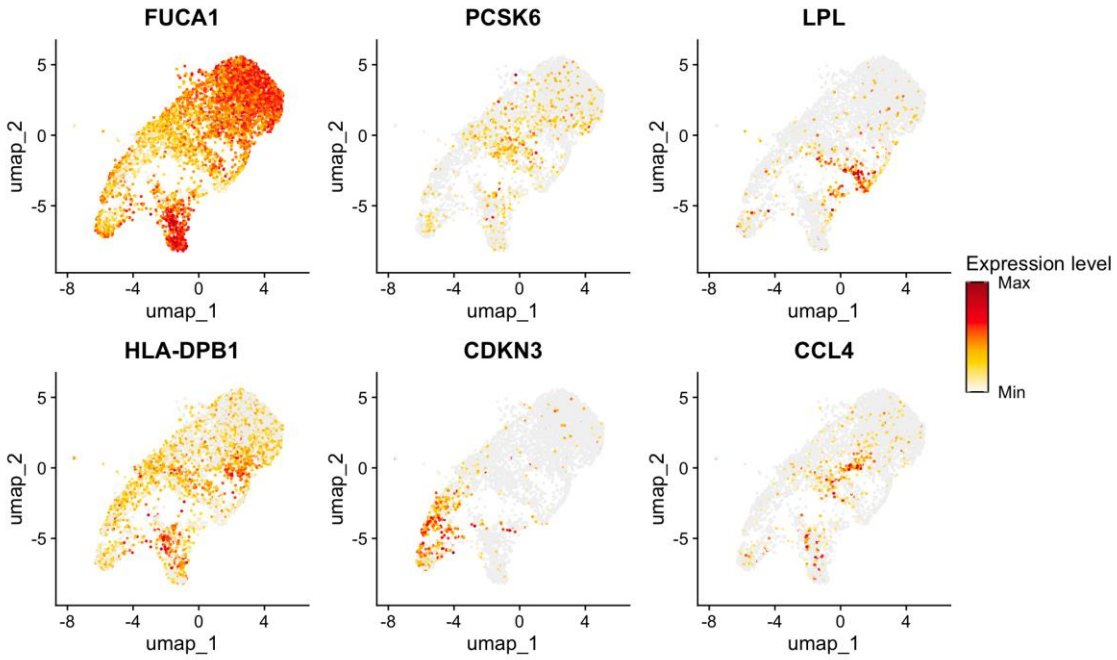


Figure 3.3 Expression of selected marker genes for different macrophage subpopulations.
Macrophage subpopulations were identified using the scRNA-seq data as shown in Figure 3.2 Read counts were normalized by sequencing depth, and were minimum-maximum normalized for visualization purposes. Genes were displayed on the same UMAP visualisation as that in Figure 3.2.

I next examined whether our single-cell transcriptomic data successfully captured the cellular response to ox-LDL by a differential expression (DE) test using 'pseudo-bulk' populations. More specifically, reads from all cells within each biological replicate were aggregated into a 'pseudo-bulk' sample, representing an average global expression profile similar to bulk sequencing, and were processed using the bulk DE pipeline in edgeR. This pseudo-bulk method is considered as the gold standard test to control for false-positive discoveries in single-cell transcriptomic data [162]. Consistent with our previous findings in bulk macrophages [113,156], exposure to ox-LDL significantly upregulated the expression of genes involved in lipid uptake (*CD36*, *FABP5*), cytoplasmic lipid deposit (*PLIN2*), and lipid transport (*ABCA1* and *ABCG1*) (Figure 3.4 A, Online Table

S1)[163,164]. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis demonstrated that ox-LDL triggered lysosomal activation and inhibited cholesterol biosynthesis (Figure 3.4 B).

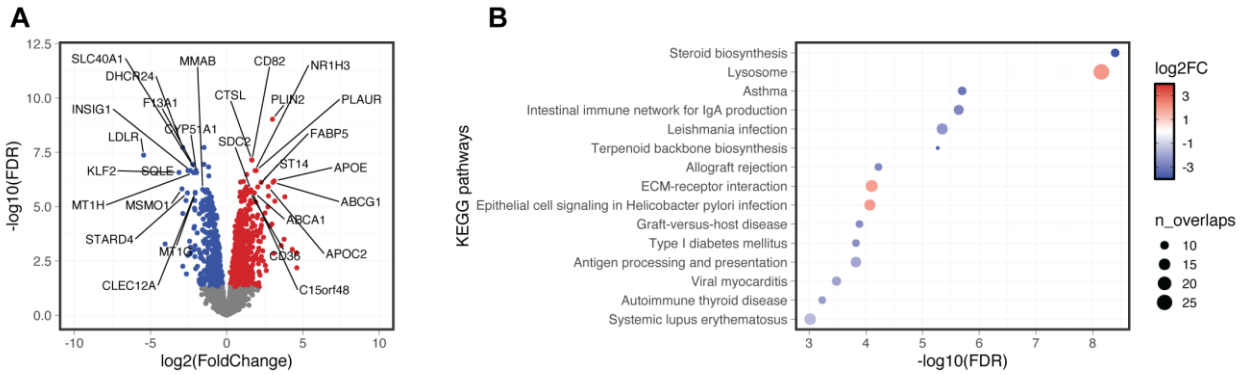


Figure 3.4 Ox-LDL triggers major transcriptomic changes in human macrophages. **A**, Volcano plot of genes up-regulated (red) and down-regulated (blue) by ox-LDL, with fold changes plotted on the x axis and false-discovery rates (FDR) plotted on the y axis. **B**, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in the differentially-expressed genes induced by ox-LDL. Colour represents log2-fold change and size represents the number of overlapped genes in each pathway.

3.2.2. Multiome data QC

To better understand the epigenetic regulation of macrophage gene expression, we undertook single-cell multi-omics sequencing using the same experimental model. Macrophages were derived from four healthy adults of mixed gender, all within the working age range, who self-reported European ancestry. The transcriptomic and open chromatin profile of each single nucleus were characterized simultaneously, with or without the exposure to ox-LDL.

Due to an unexpected issue in sample loading at the Oxford Genomic Centre, one ox-LDL sample (oxLDL_2) failed the in-house pre-alignment QC and was excluded from downstream analysis. In total, 5,478 cells have passed the QC threshold of:

- TSS enrichment >1
- 1000 < detected gene < 50,000
- 1000 < detected peak < 50,000.

For QC of the single-nucleus RNA-seq (snRNA-seq) assay, unlike in the scRNA-seq analysis, I did not apply any mitochondrial read percentage filter because of the presence of excessive mitochondrial transcripts in all buffer samples. QC metrics from the snATAC assay suggests that this potential mitochondrial contamination came at a late stage of the library prep, as only the mtRNA but not mtDNA was found in libraries recovered from the same single nucleus. To mitigate the effect of mitochondrial contamination on downstream analyses, a latent variable of mitochondrial RNA% was included in both the sample integration and the differential expression step. The detailed QC metrics for all samples in the RNA and the ATAC assay were provided in Table 3.1 and Table 3.2, respectively.

Table 3.1 Description and summary statistics for the snRNA-seq library of the multiome data

Sample ID	Donor	Sex	n_cell	Median genes per cell	Median RNA counts per cell	Median mito%	Median ribo%
buffer_1	D001	M	892	4,773.5	18,490.5	19.12	1.54
buffer_2	D002	F	855	3,903	12,168	16.54	1.08
buffer_3	D003	F	714	4,252	15,388	18.79	0.88
buffer_4	D004	M	682	3,843.5	12,368	23.67	1.00
oxldl_1	D001	M	803	3,395	9,269	3.87	0.55
oxldl_3	D003	F	734	3,683	10,776.5	2.48	0.47
oxldl_4	D004	M	798	3,293.5	7,984	0.12	0.39

Table 3.2 Description and summary statistics for the snATAC-seq library of the multiome data

Sample ID	Donor	n_cell	Median ATAC fragments per cell	Median ATAC counts per cell	Median TSS enrichment	Mean mtDNA%
buffer_1	D001	892	8,422.5	19,017.5	5.12	1.02
buffer_2	D002	855	10,647	24,889	5.20	1.15
buffer_3	D003	714	8,549	19,847.5	5.28	1.17
buffer_4	D004	682	13,252	31,172	4.73	0.72
oxldl_1	D001	803	8,795	20,344	5.40	1.12
oxldl_3	D003	734	10,106	23,999.5	5.20	0.75
oxldl_4	D004	798	11,070.5	26,346	5.20	0.74

3.2.3. snRNA-seq is highly concordant with scRNA-seq data

To test the robustness of our model system, I first compared the gene expression profiles generated from the snRNA-seq and our previous scRNA-seq. As expected, transcripts collected from single cells were mostly spliced mRNA (Figure 3.5A), while transcripts collected from nuclei contained more pre-mRNA in the unspliced form (Figure 3.5B).

While the absolute counts per gene is highly dependent on the sequencing depth and the nature of samples (e.g., more nuclear-retained transcripts in the snRNA-seq compared to the scRNA-seq data), the relative expression relationship (i.e., the rank of expression among all expressed genes) were highly concordant between the two datasets (spearman's correlation coefficient $\rho=0.83$, $p<2.2\times10^{-16}$). In addition, the changes in gene expression induced by ox-LDL were consistent between the scRNA-seq and snRNA-seq data (Figure 3.6), although the magnitude these changes seems to be smaller in snRNA-seq data.

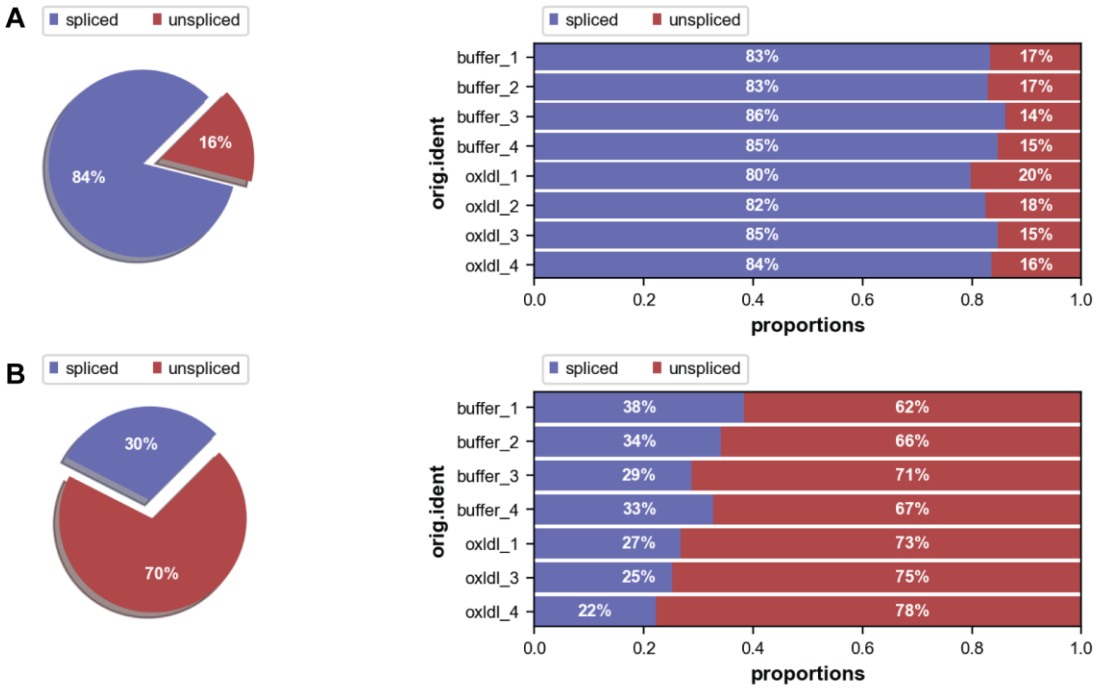


Figure 3.5 Proportion of spliced and unspliced transcripts.
Spliced and unspliced transcripts were quantified by running the velocyto pipeline [165]. The proportion of unspliced transcripts was plotted for the single-cell (**A**) and the single-nucleus RNA-seq library (**B**), showing high abundance of unspliced transcripts in the single-nucleus data. The per-sample splicing ratios are shown to the right.

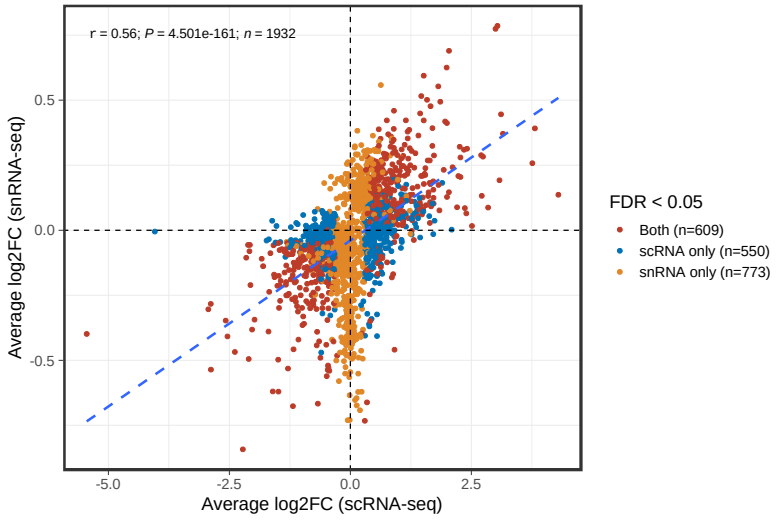


Figure 3.6 Ox-LDL-induced gene expression changes observed in macrophages using different techniques.
Data are plotted as log2 fold changes (log2FC) in single-cell RNA-seq (scRNA-seq, x axis) and in single-nucleus RNA-seq (snRNA-seq, y axis). The 95% confidence interval of the fitted regression line is shaded in grey. Significant differentially expressed genes identified in scRNA-seq, snRNA-seq or both datasets are coloured in blue, yellow or red, respectively. Spearman's correlation coefficient and p-value are shown alongside the plot.

3.2.4. Joint clustering analysis using both modalities improves the resolution of the allocation of macrophage subpopulations

Next, I integrated the snATAC assay with the snRNA profile from the same cell to improve the resolution of macrophage subtype allocation. Both assays were subject to stringent batch-effect removal and were used to construct two independent SNN graphs (Figure 3.7A and B). These two graphs were further integrated using the 'weighted-nearest neighbour' (WNN) method by unsupervised learning of the cell-specific modality weights. As demonstrated in Figure 3.7, while the RNA assay performed well in separating transcriptomically distinct macrophage subpopulations, integrating additional open chromatin information increased the clustering resolution. In particular, the FUCA1^{hi} cluster was now separated into two clusters, and a small AK5^{+} population (*ANKH*, *AK5* and *MYO1B*) was newly identified with some similarity in gene expression to osteoclasts [166].

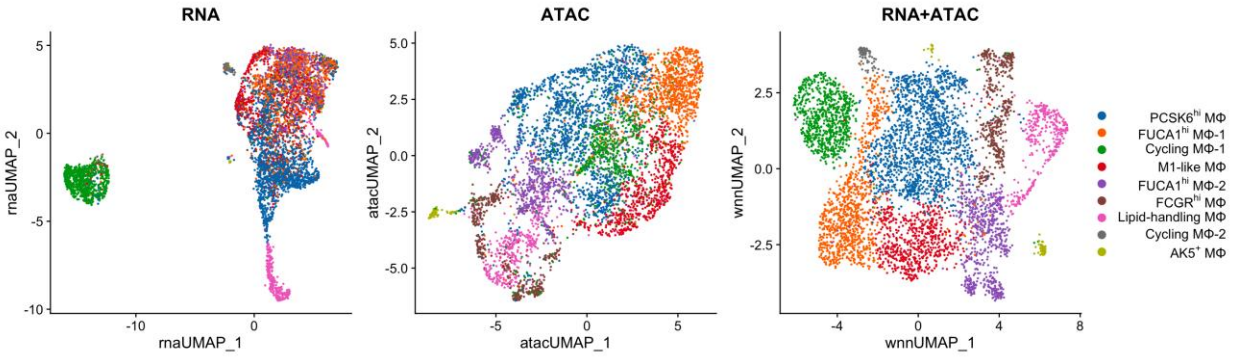


Figure 3.7 Uniform Manifold Approximation and Projection visualisation of the single-cell multiome data.
Reduced dimensions were computed based on the RNA assay (**left**), ATAC assay (**middle**) or both assays (**right**). Cells (n=5,478) were clustered by their similarity in both assays and are coloured by subpopulations assigned based on jointly clustering.

In total, WNN-based clustering identified 9 macrophage subpopulations (Figure 3.7). During the initial sample integration step, each sample (e.g., buffer_2) was treated as an individual batch. This is under the assumption that macrophages from different donors should have similar functional subpopulations, and ox-LDL treatment would not lead to the formation of new subpopulations. It is important to note that these batch-corrected transcriptomic or epigenomic profiles were only used in the integration and clustering step, and depth-normalized raw counts were used in differential expression analyses. Therefore, violation of these assumptions will only affect the interpretation of clustering results and does not affect the downstream analysis of gene regulation dynamics.

As shown in Figure 3.8, the subpopulations identified in the multiome data were unbiasedly present in all samples, indicating successful batch-effect removal during integration. We found one cluster showing high expression of ox-LDL-induced genes (*APOE*, *APOC1*, *PLPP3*, etc.) even in the naïve state without ox-LDL exposure; since most of its marker genes are involved in lipoprotein and fatty-acid metabolism [167], I termed this cluster lipid-handling macrophages. A detailed investigation of this subpopulation will be presented in Chapter 4.

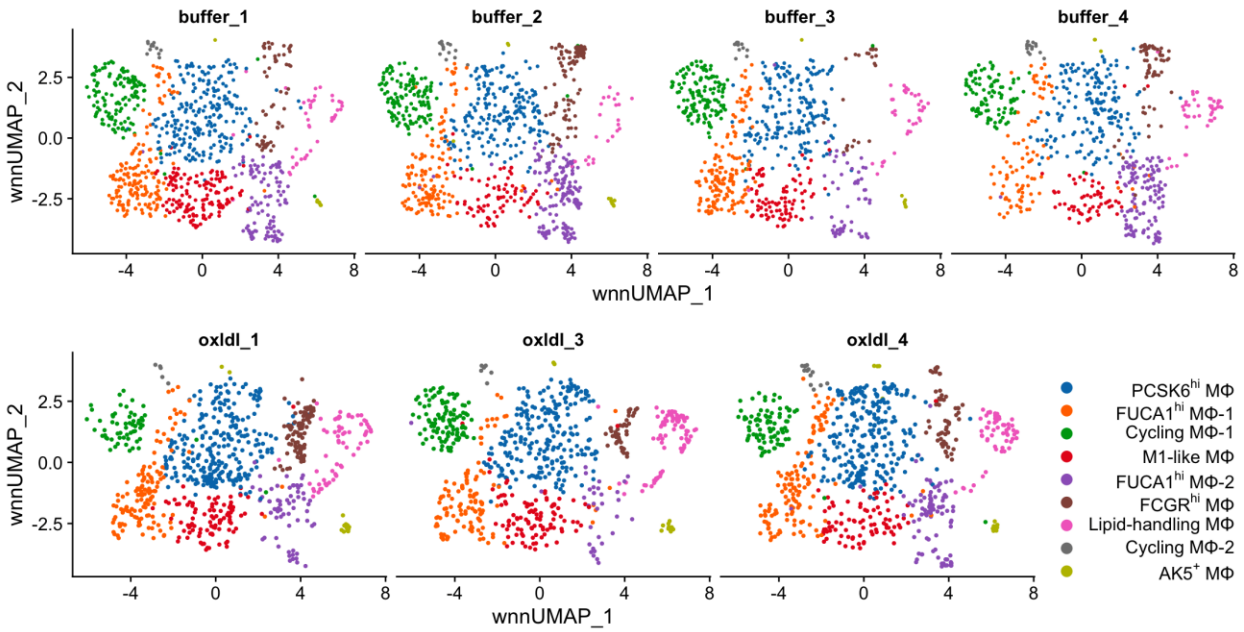


Figure 3.8 Visualisation of macrophage subpopulations in each sample.
Cells were clustered by their similarity in both assays as in Figure 3.7 and plotted separately by sample. The presence of all clusters in each sample indicates successful correction for batch effect.

3.3. Heterogeneity in human macrophages and their response to ox-LDL

3.3.1. Heterogeneity in transcriptome

I next applied a logistic regression test using mtRNA% as a latent variable to investigate the subpopulation-specific gene expression patterns (Figure 3.9). The majority of macrophages (57.0%) demonstrated classic resting M0 phenotypes, including a PCSK6^{hi} population, (*PCSK6*, *ITGB5* and *SHC3*) and two FUCA1^{hi} populations (*FUCA1*, *WWP1* and *SELENOP*). Interestingly, even under the culture condition absent of traditional M1 stimulus like LPS or IFNγ, a macrophage cluster showing relatively-high expression of pro-inflammatory cytokines (*CCL2*, *ICAM1* and *CCL3*) was present in all four donors,

indicating the complex nature of macrophage functions that deviates from the classic stimulus-dependent M1/M2 model.

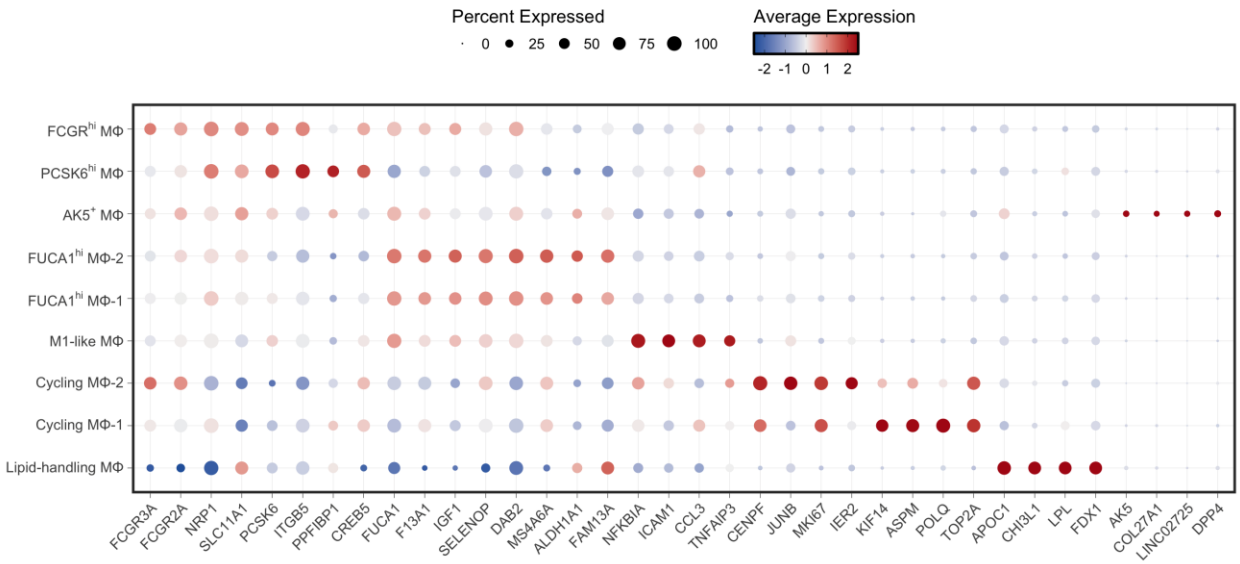


Figure 3.9 Dot plot of the expression of selected marker genes for the 9 macrophage subpopulations.
The colour represents the average gene expression in each subpopulation, normalized by sequencing depth and scaled across all cells; the size represents the percentage of cells expressing the gene.

In addition, a small cluster constitute only 0.1% of the total population was found to have gene expression signatures (*ANKH*, *AK5* and *MYO1B*) similar to those in human osteoclasts [166], possibly representing osteoclast pre-cursors. Cross-referencing with an unpublished scRNA-seq dataset in murine osteoclasts that I worked on in collaboration with the Simon group at the Kennedy Institute of Rheumatology also confirmed this match expression signature. It has been long established that osteoclasts develop from cell fusion of macrophages under M-CSF and RANK ligand (RANKL) stimulation [168,169]. However, the *TNFSF11* gene that encodes RANKL was not detected in macrophages profiled in this study, and it is unclear why and how these AK5$^+$ cells arose.

Other subpopulations included two proliferating populations (*TOP2A*, *CENPA* and *CENPE*), and a transitional monocyte/macrophage FCGR^{hi} population (*FCGR3A*, *FCGR2A* and *PCSK6*). The full list of marker genes for each macrophage subpopulation is detailed in Online Table S2.

3.3.2. Heterogeneity in transcription factor activation

To explore the transcription factors (TFs) activated in the different macrophage subpopulations, we assessed TF binding activity in individual cells by computing motif deviation scores using chromVAR [141]. In brief, a set of GC%- and average-accessibility-matched background peaks was created for each macrophage open chromatin region; cell-specific deviation scores were then computed by calculating the relative enrichment for the TF binding motifs in cell-specific open chromatin regions against the background peak set. In other words, this motif deviation represents motif-specific gain or loss of chromatin accessibility relative to the average cell profile. Similar to gene expression, TFs that were differentiated regulated would have higher variation across different cells, and the top 15 TFs with highest variations are plotted in Figure 3.10.

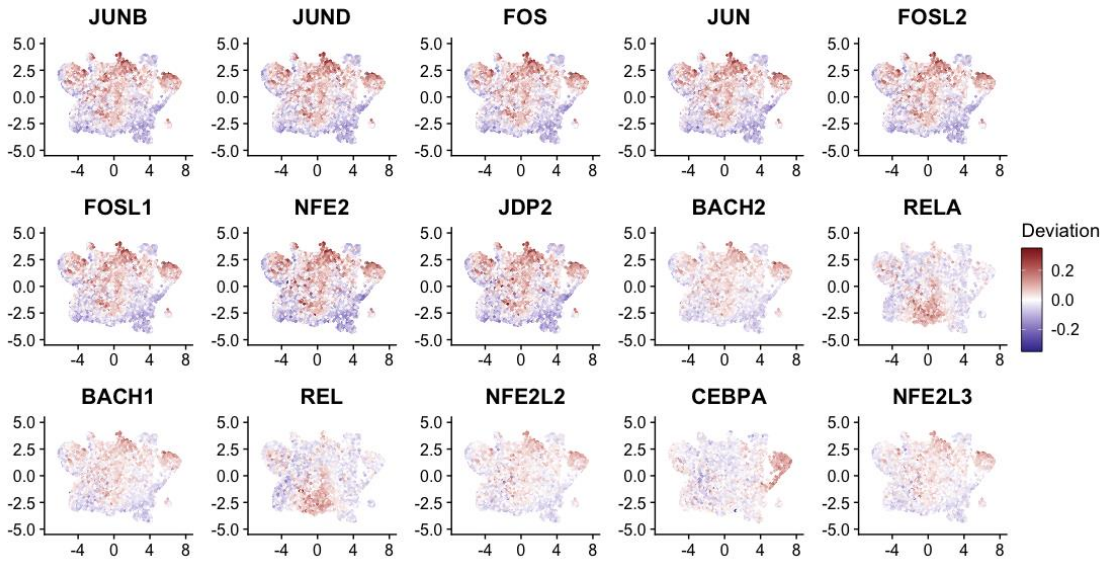


Figure 3.10 Motif deviation scores for the top 15 most deviated transcription factors. The relative activity of transcription factor binding was quantified as the motif-specific gain or loss of chromatin accessibility relative to the average cell profile (motif deviation score). Cells are coloured by the associated motif deviation score, with positive deviation meaning increased TF activity and vice versa. Single cells are projected onto the same UMAP as shown in Figure 3.8.

Most of these differential regulated TFs belong to the AP-1 family (JUN, JUNB, FOS, FOSL1) and the NFE-2 family (NFE2, BACH2, NFE2L2). Using bulk approaches, we have previously identified enrichment of the binding motifs of AP-1 and NFE-2 factors in dynamic enhancer sites following exposure of human macrophages to ox-LDL [113]. Both AP-1 factors and NFE-2 factors respond to oxidative stress [170,171]. However, it is challenging to pinpoint the specific TFs involved in the ox-LDL-response in macrophages, as they all share similar binding motifs.

To address this challenge and to delineate the differential role of TFs within the same family, I further examined the correlation between the motif deviation scores and the expression of each individual TF in paired single cells. As shown in Figure 3.11A, While most AP-1/NFE-2 factors have high motif deviation scores in ox-LDL-exposed macrophages, motif-accessibilities for some TFs do not correlate well with their

corresponding gene expression, meaning that the observed accessibility changes are likely due to other members in the family. With additional information from the correlation analysis, we can further recover the directional influence of individual AP-1 and NFE-2 TFs. For example, expression of two AP-1 family members *FOSL1* and *FOSL2* positively correlated with the chromatin accessibility of AP-1 motifs (Figure 3.11A), suggesting that their expression activates the bound chromatin; conversely, higher expression of another AP-1 factor *FOS* is accompanied with lower chromatin accessibility of its binding sites (Figure 3.11A-C), which indicates a repressive role in transcriptomic regulation. Expression of the NFE-2 family member *BACH1* was positively correlated with its motif binding (Figure 3.11A), but no significant correlation was observed for other NFE-2 factors.

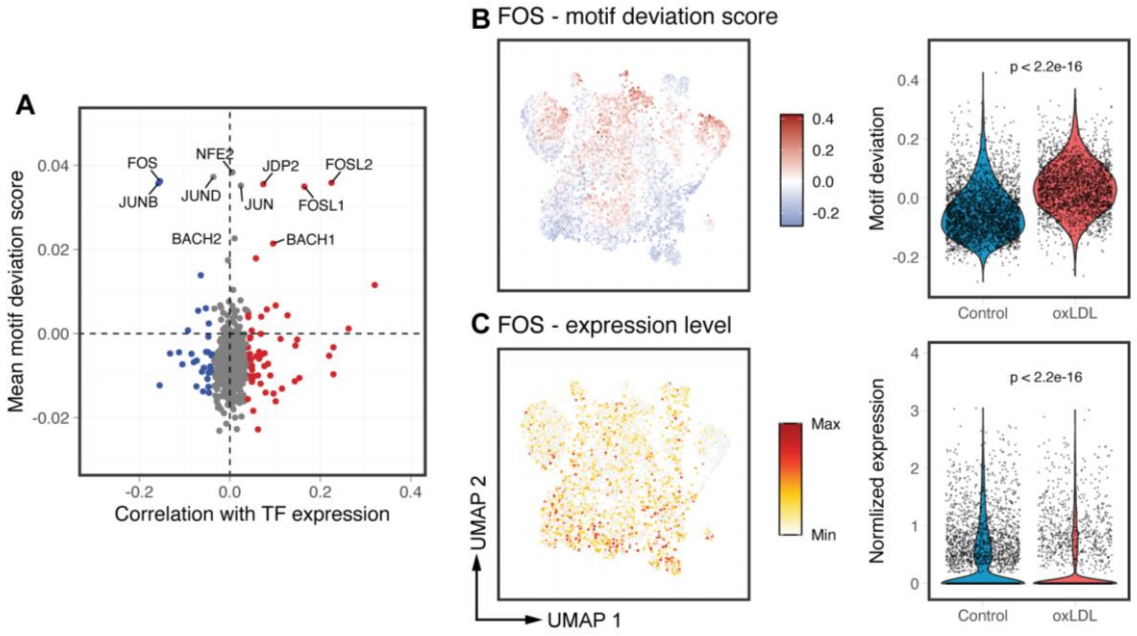


Figure 3.11 Negative correlation between the gene expression and motif activity for FOS.
A, Mean TF motif deviation scores in ox-LDL–treated macrophages (y axis, relative enrichment compared with background) plotted against their correlation with TF RNA expression (x axis, Pearson coefficient). TFs are coloured by significance (FDR <0.05) for negative (blue) or positive (red) correlation. **B**, Motif deviation scores for *FOS*, shown on the UMAP (left) as well as summarised in the violin plots between control and ox-LDL-exposed macrophages (right). **C**, Gene expression of *FOS*, shown on the UMAP (left) as well as summarised in violin plots (right).

This motif-expression correlation analysis was also applied to analyse subpopulation-specific TF activation. As expected, M1-like macrophages display strong and specific activation of NF-kB family TFs, as demonstrated by their high motif scores and positive correlation with gene expression (Figure 3.12).

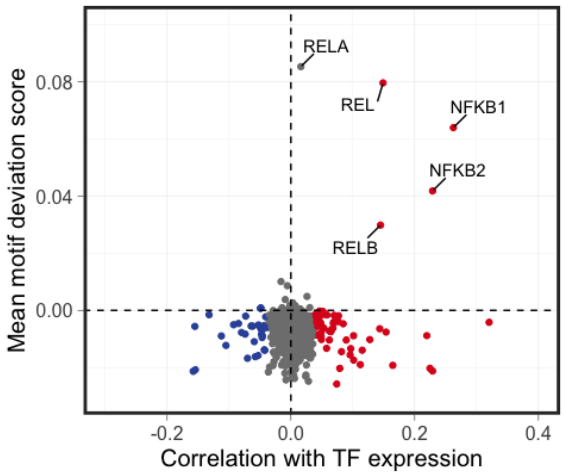


Figure 3.12 NF-kB family factors are activated in M1-like macrophages
Mean TF motif deviation scores in M1-like macrophages (y axis, relative enrichment compared with background) plotted against their correlation with TF RNA expression (x axis, Pearson coefficient). TFs are coloured by significance (FDR <0.05) for negative (blue) or positive (red) correlation.

3.4. Mapping cis-regulatory elements

Single cell multi-omics data allowed us to identify transcriptomic regulatory interactions by correlating chromatin accessibility at non-coding peaks with expression of their potential target genes [172]. To mitigate the multiple-test burden and to exclude complicated trans-interaction which biased toward transcription factors, I limited the correlation test to the peak-gene pairs where the TSS and the peak are <500kb apart. Similar to the motif deviation analysis, background peaks were tested alongside with the peak of interest to control for false positive discoveries.

In total, 18,782 unique cis-regulatory elements (CREs) were identified, spanning 6,381 different genes (Online Table S3). Of these CREs, 2,642 were correlated with expression of more than one gene, resulting in a total of 22,555 peak-gene pairs. For 94.9% (21,416 out of 22,555) of the CRE-gene pairs, the accessibility of CRE correlates positively with the target gene expression, suggesting an overall activating role for identified CREs.

The major locations of CREs were in introns (42.3%), promoters (24.4%, defined as up to 3kb upstream of TSSs), and intergenic regions (24.3%) (Figure3.13A). A significant proportion (42.6%) of CREs were located >10kb away from any recognized transcription start site (Figure 3.13B).

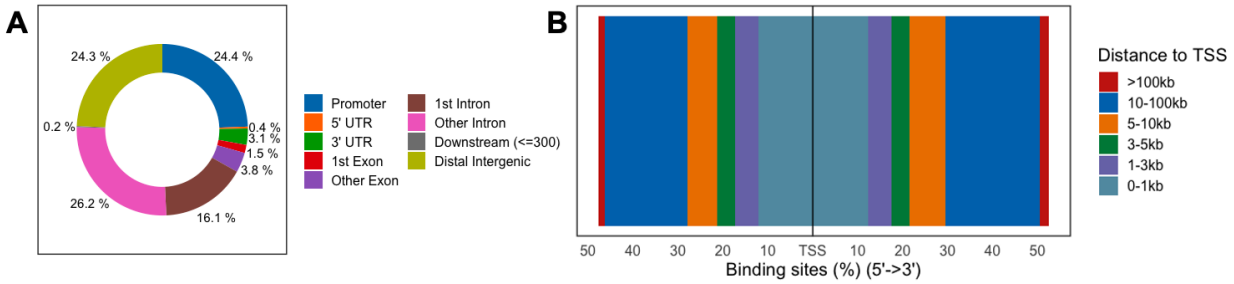


Figure 3.13 Location of cis-regulatory elements (CREs).
A, genomic annotations for identified n=18,782 CREs; **B**, distance to the nearest transcription start site (TSS) for identified CREs. For genes with multiple TSSs, only the canonical TSS was included.

3.5. Macrophage CREs are enriched for functional annotations

Traditionally, regulatory elements were identified by a combination of epigenomic signatures (e.g., histone modifications, CBP/p300 bindings, chromatin looping) using bulk sequencing techniques [173,174]. To validate whether our computationally predicted macrophage CREs also carry these epigenomic markers, I performed an enrichment test

against a wide range of functional annotations including blood eQTLs, Activity-By-Contact (ABC) enhancers, promoter-interacting elements and H3K27ac signals.

3.5.1. Enrichment for public annotation

The eQTL datasets collected in whole human blood were downloaded from GTEx [175] and eQTLGen [176]; gene ids were converted to the HUGO Gene Nomenclature Committee (HGNC) approved gene symbols; variants were lifted over to hg38 reference genome. Permutation-based enrichment test showed that blood eQTL SNPs (eSNPs) were significantly overrepresented in macrophage CREs ($p=6.2\times10^{-10}$, Figure 3.14A), and 41.2% of CREs contain at least one eSNP.

As eQTL signals from whole blood samples are likely a mixture from multiple cell types, I next sought to test the enrichment for myeloid-specific annotations. A recent paper proposed a simplified model of quantifying enhancer strength by multiplying enhancer histone signals with the contact frequency of their target genes [109]. Later the authors expanded their model to samples from the ENCODE Project to generate genome-wide enhancer-gene maps in 131 human samples, covering 38 samples from the mononuclear phagocyte lineage [110]. These mononuclear samples included primary cells like human $CD14^{+}$ monocytes and monocyte-differentiated dendritic cells, as well as immortalized cell lines like THP-1. ABC enhancers from all 38 samples were concatenated and used as the reference set for enrichment test and the benchmarking analysis in Chapter 4. As expected, macrophage CREs were also enriched for mononuclear ABC super enhancers (Figure 3.14B).

Another important feature of distal enhancers is that they often physically interact with the promoters of their target genes via long-range DNA looping. In a similar experimental system where macrophages were differentiated from circulating monocytes with M-CSF treatment, over 15,000 regions were shown to interact with one or more promoters [177]. Overall, our macrophage CREs displayed significant enrichment for these promoter-interacting regions (Figure 3.14C). I also performed separate tests splitting promoter-interacting elements collected from M0 (M-CSF), M1 (M-CSF+LPS), or M2 (M-CSF + IL-13 + Rosiglitazone) macrophages, and the differences were negligible (data not shown). This suggest that although M1 and M2 macrophages present polarized phenotypes, the regulatory elements controlling their gene expression are largely shared and can be profiled at the resting M0 stage.

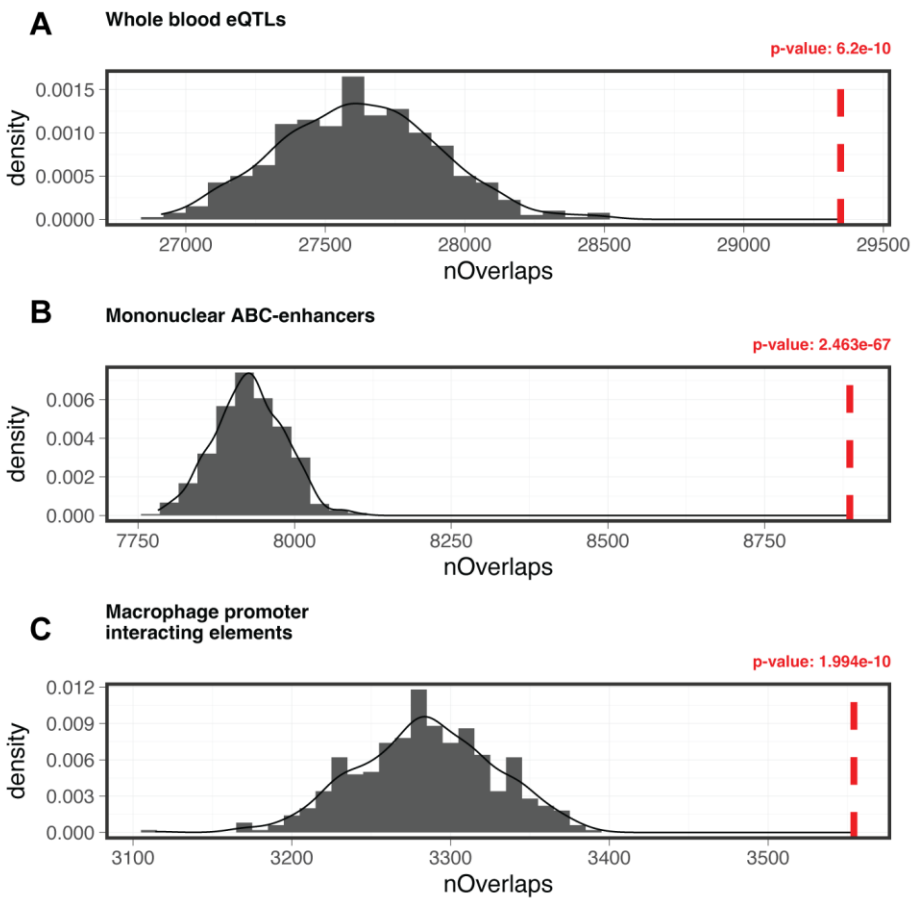


Figure 3.14 Macrophage cis-regulatory elements (CREs) are enriched for functional annotations.
A through **C**, Histogram showing the null distribution of overlaps between whole blood eQTLs (**A**), mononuclear ABC-enhancer (**B**) and macrophage promoter-interacting elements (**C**) and n=1000 sets of background peaks matched for GC% bias and average accessibility for all macrophage CREs. The observed overlap is marked by the red dashed line. P-values calculated using Z-test.

3.5.2. Enrichment for H3K27Ac signals

One of the most common epigenetic markers used to identify active enhancers is histone H3K27ac [178,179]. In our group we have previously generated a genome-wide enhancer annotation for macrophages using bulk H3K27ac sequencing, and have shown that ox-LDL treatment triggers change in enhancer signals [113]. Here, I integrated the previous bulk H3K27ac ChIP-seq data and assessed the average active enhancer signals in macrophage CREs (Figure 3.15).

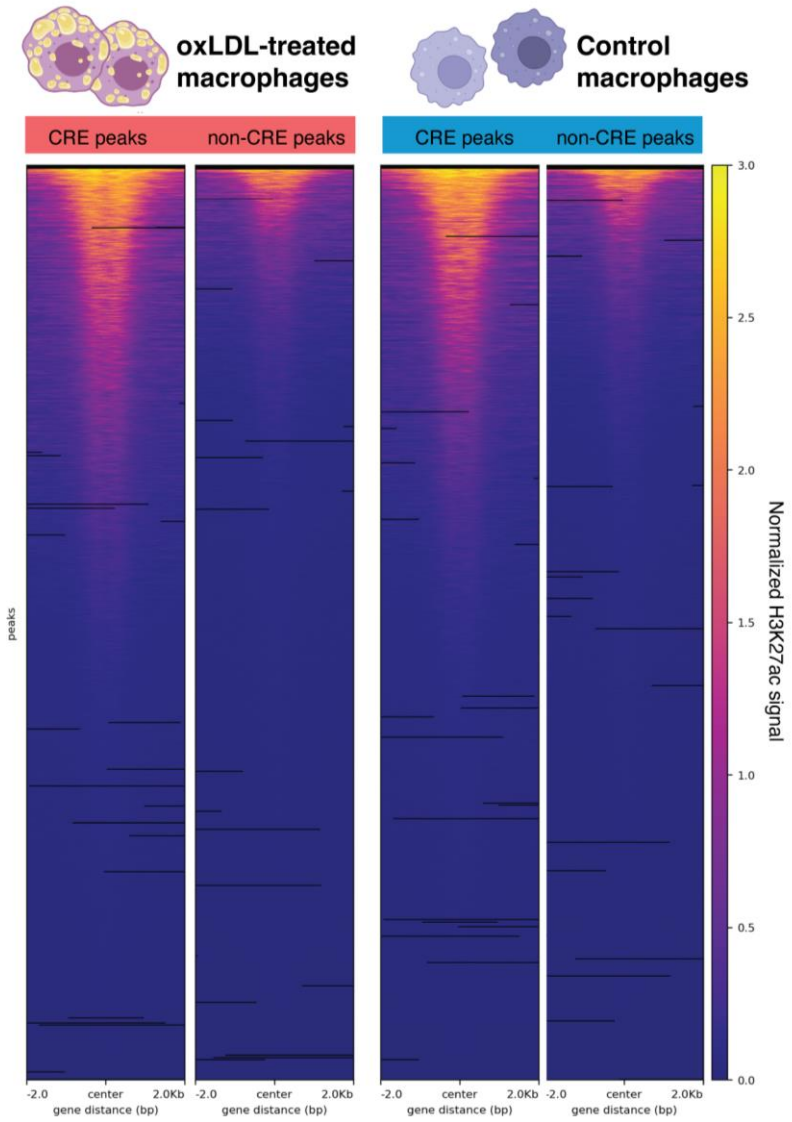


Figure 3.15 Heatmap showing the aggregated H3K27ac signals in macrophage open chromatin regions.
H3K27ac signals from the bulk ChIP-seq were aggregated by CRE peaks and non-CRE peaks, showing higher signals for enhancer histone marks in CREs. **left**: ChIP-seq signals in ox-LDL-exposed macrophages. **right**: ChIP-seq signals in unexposed control macrophages.

As expected, CREs displayed significantly higher H3K27ac signal than non-CRE open chromatin regions ($p < 2.2\times10^{-16}$; Figure 3.15 and Figure 3.16). Ox-LDL exposure was found to further increase the H3K27ac signals in CREs ($p = 6.458\times10^{-6}$), but not in non-CRE peaks (Figure 3.16). While it requires extensive experimental validations to prove

the predicted cis-regulations, the enrichment for traditional enhancer signatures collectively suggest that single-cell multi-omics data can robustly map the enhancer-gene under disease-relevant context.

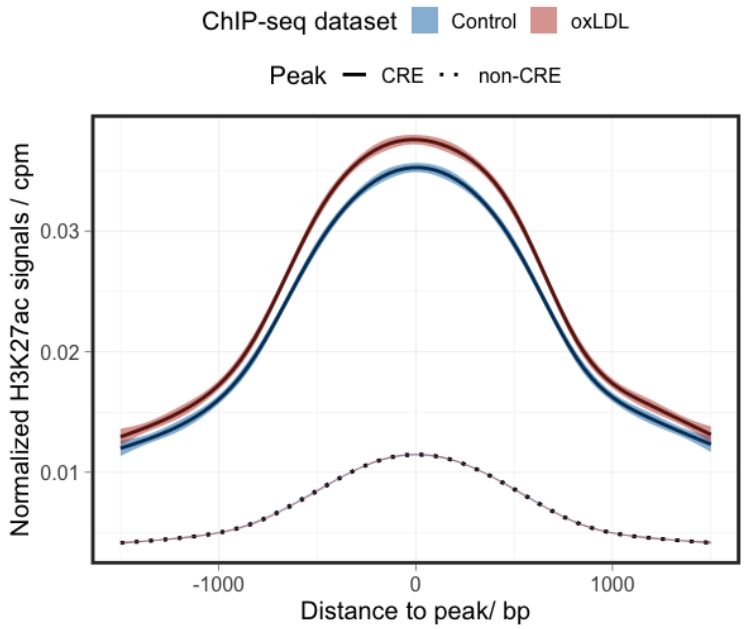


Figure 3.16 Aggregated H3K27ac signals at CRE-peak and non-CRE-peak regions. The enhancer histone signals were significantly higher in CRE peaks compared to non-CRE open chromatin regions. Signals from unexposed (blue) and ox-LDL–exposed (red) human macrophages are plotted separately. Shaded regions indicate 95% CIs.

3.6. Constructing transcriptional regulatory networks

Recent advances in single-cell genomics included the recovery of transcription-regulatory networks [142,180-183]. Largely following the Functional interference of gene regulation (FigR) framework [142], we carried out a series of correlation tests to infer the gene regulation dynamics in macrophages. Briefly, we first filtered out genes with low densities of CRE to reduce random noises and to mitigate the multiple test burden (section 3.6.1). The remaining chromatin accessibility profiles were then used to infer gene-regulatory network through correlation with the expression of candidate TF regulators (section 3.6.2). Only

TFs whose binding motif was enriched in the regulatory regions were retained for down-stream analysis.

3.6.1. Domain of regulatory chromatins

In total, 636 genes linked to more than 8 CREs were retained (Figure 3.17A). These genes and their associated CREs constitute domains of regulatory chromatin (DORCs), and overlapped largely with Activity-By-Contact (ABC) super-enhancer-to-gene pairs [110] (529 out of 636 DORC genes; super enhancer identified as ABC score >0.1; Figure 3.17B).

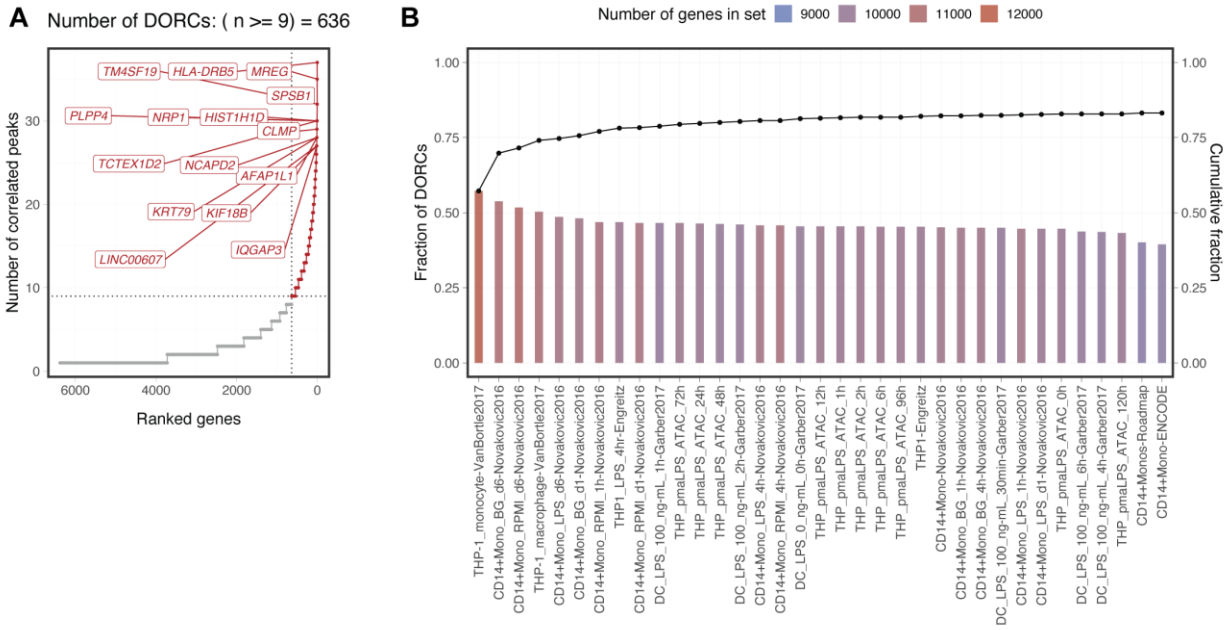


Figure 3.17 Domain of regulatory chromatin (DORC) analysis identified new transcriptional regulations compared to public bulk sequencing data.
A, Genes plotted according to the number of significant CREs mapped across all macrophages. Top scoring genes are labelled. Genes with at least 9 CREs are marked as DORCs in red. **B**, Individual (bars) and cumulative (dots on line) fraction of the DORCs also being identified as strong enhancers (score>0.1) using the activity-by-contact (ABC) approach across the different mononuclear cell samples shown.

As previously demonstrated, the vast majority of the identified CREs correlated positively with their target gene expressions. Therefore, for each DORC gene, the open chromatin signals overlapping all its CREs were summarised to a DORC score, which represented the epigenomic regulatory activity of this gene. Upon ox-LDL exposure, changes in DORC scores closely correlated with changes in the expression of DORC genes (R=0.62, $p=2.1\times10^{-68}$; Figure 3.18).

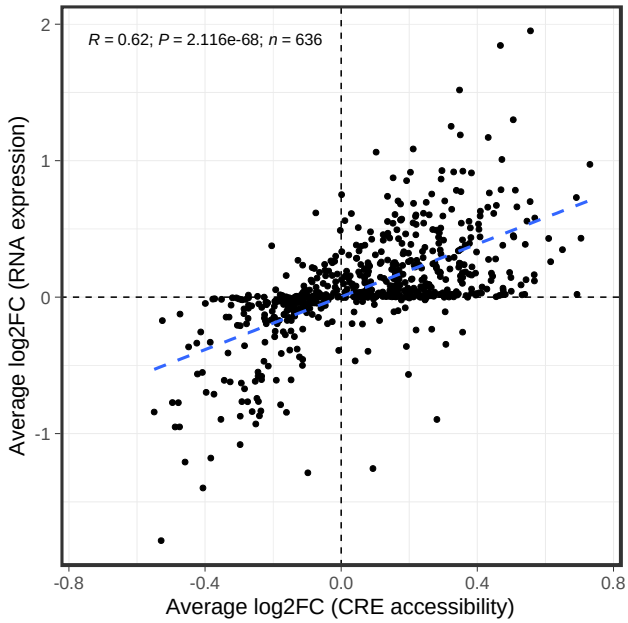


Figure 3.18 Pearson correlation between ox-LDL–induced changes in cis-regulatory element chromatin accessibility and in gene expression.
For each DORC, ATAC-seq signals in all CREs were aggregated and were tested for differential accessibility between ox-LDL-exposed and unexposed cells. In total n=636 DORC gene as identified in Figure 3.17A were analysed.

3.6.2. TF-DORC regulatory networks

To infer the potential TF regulators underlying the gene-CRE pairs for each DORC, I tested the expression correlation and motif enrichment for each TF in each DORC as implemented through the FigR pipeline. To overcome the data sparsity, k=20 nearest neighbours from the WNN graph were extracted and used to smooth the DORC scores and TF expression. Then, for each DORC, a second KNN graph (k=30) was constructed based on the smoothed single-cell DORC scores (a graph for DORCs instead of cells as in the previous step). This step is essential for the motif enrichment because the total number of CREs for one single DORC can be as low as 9, which is insufficient to test motif enrichment for all TFs. By pooling the CREs of the tested DORC and its 30 neighbours, with the assumption that co-variable DORCs are regulated shared TFs, motif enrichment tests were performed for all expressed TFs. At the same time, Spearman correlation tests were performed to evaluate the relationship between TF gene expression and DORC accessibility.

Combining these two metrics, I analysed the interactions between 925 expressed TFs and 636 DORCs. Out of the 588,300 total TF-DORC pairs, 1612 pairs involving 95 TFs and 526 DORCs exhibited strong interactions (|regulatory score| > 1.5) (Online Table S4). The top 0.1% of TF regulators and their target DORCs with the maximum regulatory scores higher than 2 are visualised on a heatmap in Figure 3.19. Notably, several TFs, primarily in the AP-1 family, display opposite regulatory effects across different DORC genes. For example, JDP2 has an activating effect on *CD109*, but negatively regulates *PLPP3* (marked with arrows in Figure 3.19).

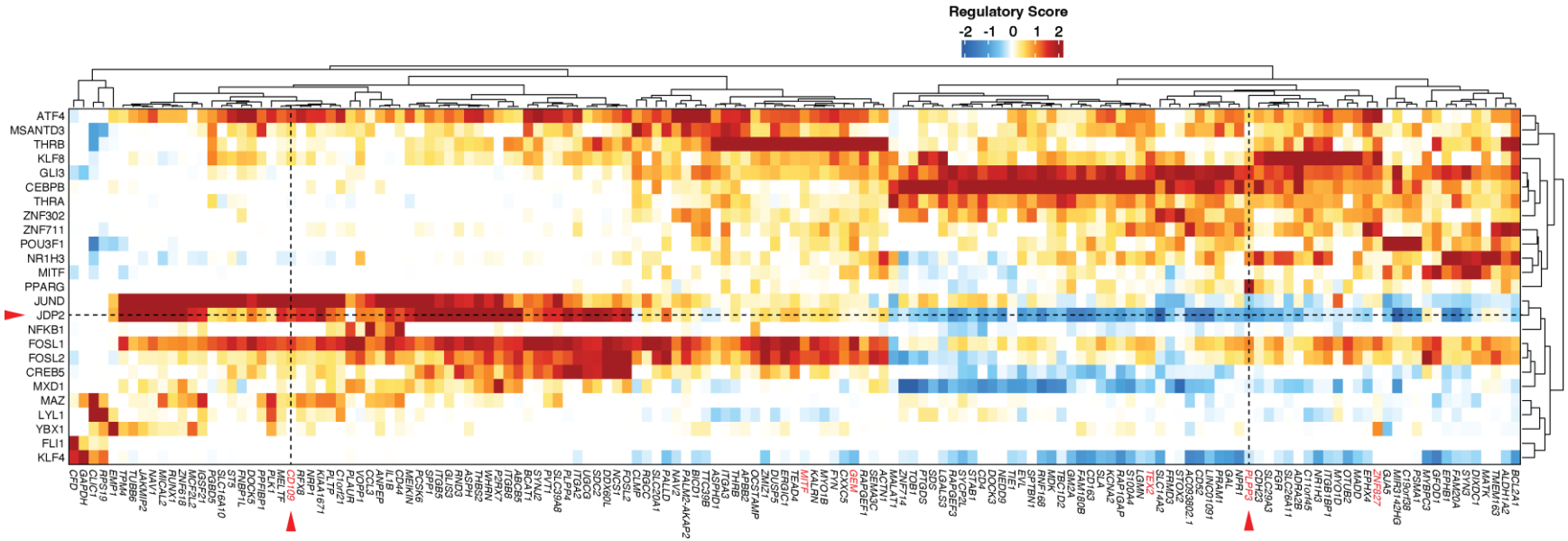


Figure 3.19 Heatmap of regulation scores for the top 0.1% of TF-DORC enrichments.
DORC genes are in columns and TFs are in rows (n=25 TFs and n=147 DORCs). DORCs with CREs overlapping CAD-associated variants are marked with red. Positive and negative regulatory effects are shown in red and blue, respectively. Red arrows highlight the opposite regulatory effect of JDP2 on CD109 and PLPP3.

3.7. Discussion

The innate immune system, centred at the monocyte-macrophage lineage, have long been associated with the progression and resolution of atherosclerosis [184]. Decades of epidemiological research have identified LDL cholesterol, together with its modified form ox-LDL, as the major risk factor of atherosclerosis, and pathological studies have revealed their *in vivo* accumulation in lipid-laden macrophages [61]. Despite extensive study in preclinical models and in human plaque samples, the molecular mechanisms underlying the response to ox-LDL in macrophages remain unclear, mainly due to the heterogeneity of macrophages and the complexity of microenvironment within plaques.

Advances in single-cell technologies offer unprecedented opportunity to directly profile the cellular diversity in atherosclerotic plaque samples[24,185,186]. However, almost all these cells are sampled from advanced lesions, where the functional pathogenic activity of ox-LDL and other environmental stimuli may be historic. This is particular problematic for investigations focused on macrophages, as these cells are highly plastic and gene regulatory programmes are strongly influenced by prevailing stimuli and neighbouring cellular activity [144,187]. To overcome this, we established a simplified experimental system using primary monocyte-derived macrophages, and employed single-cell RNA and multi-omics sequencing to characterize the transcriptional regulatory dynamics before and after exposure to the atherogenic lipid, ox-LDL.

During batch effect correction and sample integration, cells from the same biological donor but receiving different treatment were treated as independent batches. This means that most ox-LDL related transcriptomic and epigenomic signatures have been corrected

before allocation of macrophage subpopulations, under the assumption that ox-LDL treatment alone does not change the biology of macrophage heterogeneity. While this step is critical to ensure the representativeness of identified macrophage subpopulations, the assumption itself might be violated. For example, a recent single-cell study reported that the TREM2^{+} macrophage subpopulation in plaque transition into a new PLIN2^{hi}/TREM1^{hi} subpopulation partly mediated by TLR2 signalling [30]. Although trajectory analysis in macrophages profiled in this chapter does not support an ox-LDL-induced transition in cellular identities, it is certainly possible that the identified subpopulations represent dynamic functional states instead of static terminal subsets.

Another major focus of this chapter is the identification of cis-regulatory elements (CREs). Although regulatory regions like promoters and enhancers can be easily tracked by their unique epigenomic marks, a major challenge is to map these regulatory elements to their target genes. Traditionally, the gene-enhancers relationships were inferred either based on proximity (e.g., to assign the enhancer to all genes within an arbitrary window) [188], or by Topologically-Associated Domains (TAD) assayed through Hi-C [109]. These approaches were limited in resolution as there are often many genes in a specific TAD and the target genes are not necessarily the nearest ones. Experimentally, such enhancer-gene interactions can be validated through CRISPR interference [189], which requires esoteric experimental expertise and is often limited to a defined region of interests.

Alternatively, gene-enhancer maps can be derived from correlation analysis between the activity of an enhancer (accessibility, H3K27ac signal, etc.) and the expression of its potential targets [84]. However, such correlation tests typically require very large sample

sizes, which in practice are often collected from immortalized cell lines with a wide range of tissue origins. This means that the inferred cis-regulation would only provide generic information for those tightly-regulated genes but fail to capture the gene regulation dynamics under a disease context. Here, by joint profiling of the expression and the open chromatin profile of individual cells, the statistical power of correlation tests can be greatly improved, without losing the cellular context of interests. In total, I have mapped the cis-regulatory effect of 18,782 unique CREs and linked them to 6,381 target genes. In the DORC analysis, I showed that at least 20% of the identified strong cis-regulatory relationships were previously uncharacterised in bulk samples using the state-of-the-art ABC enhancer model. This improved resolution then provides the foundation for the identification of transcription regulatory network coronary artery disease.

Ch 4. A novel macrophage subpopulation conveying increased genetic risk of coronary artery disease

Collaboration Notes: The work described this chapter was part of a research project in collaboration with Dr. Thomas K. Hiron who was a postdoctoral researcher in the O'Callaghan lab at the time. Dr. Hiron acquired the original single-cell RNA-seq and single-cell multiome data, and was responsible for most of the phlebotomy and tissue culture during macrophage isolation. I performed most of the fixed and live-cell confocal experiments, with additional input from Dr. Hiron clearly indicated. All computational analyses presented were performed solely by me. This work has been published in:

Jiang J, Hiron TK, Agbaedeng T, Malhotra Y, Drydale E, Bancroft J, Ng E, Reschen ME, Davison LJ, O'Callaghan CA. A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease. *Circ Res.* 2024 May 15. doi: 10.1161/CIRCRESAHA.123.324172. PMID: 38747151.

4.1. Introduction

Coronary artery disease (CAD) and the underlying pathological process atherosclerosis are influenced by both the environmental and the genetic factors [190]. Twin studies have estimated the disease heritability to be up to 60% [67-69], suggesting a major contribution from inherited genetic variants. Understanding the genetic component not only helps in identifying patients with high disease risks, but also provides insights into the biology of atherosclerosis, paving the way to future therapeutic intervention.

To date, more than 250 genomic risk loci have been identified by large genome-wide association studies (GWAS) in hundreds of thousands of participants [70,71,74]. The fact that multiple statistically independent signals can exist within a single risk locus leads to an even greater number of underlying causal risk variants for CAD. However, the majority of these variants are found in non-coding regions of the genome, which makes their biological consequence difficult to decipher. In addition, GWAS-identified variants are

often in linkage disequilibrium (LD) with multiple other variants [191,192], and in most cases only one of these LD variants would be causal. Since disease variants have been widely reported to be enriched in regulatory regions in the human genome [110,193], a common practice to prioritise causal variants is to integrate GWAS signals with functional annotations obtained from multi-omics sequencing. However, the cellular context from which functional annotations are retrieved is often arbitrarily chosen and might not reflect the true pathological process. Despite considerable efforts [194], the cellular populations and the molecular mechanisms through which genetic risk loci transmit disease risk are largely unknown.

The rapid development in single-cell sequencing techniques provides new opportunities for investigating the function of non-coding variants without prior knowledge of causal cell types. By recovering cis-regulatory interactions across different cell types [141,172,195,196], both the causal variant and the associated causal cell type can be prioritised at the same time [185,197-199]. Although the pathological characteristic of macrophages is undoubtable, and several CAD risk variants may act specifically through macrophages [200], there is surprisingly little genetic evidence supporting the linkage between macrophage and the disease. Recent single-cell studies in human plaque and cardiac samples highlighted how CAD risk loci affect specific cell types such as endothelial and smooth muscle cells, yet none of them captured such genetic contribution in macrophages [24,185,186,201].

The challenge in identifying such contributions is mainly due to the limitations of using single-cell RNA-seq (scRNA-seq) or single-cell ATAC-seq (scATAC-seq) in isolation. Almost all single-cell transcriptome-based approaches need to convert the expression signatures to genomic regions for enrichment testing, and the accuracy of this conversion

crucially affects the power of such approaches. Since the high-quality gene-enhancer maps are typically constructed from bulk sequencing of immortalized cell lines or whole tissues, they do not fully capture gene regulation dynamics in disease contexts. Conversely, scATAC-seq seems to circumvent the need for established gene-enhancer map by focusing on cell-type open chromatin regions. However, only a limited proportion of accessible chromatin sites have regulatory activity [202], leading to excessive false positive signals in the prioritised variants. These limitations call for a better regulatory elements map under a cellular system that are more relevant to the disease.

To overcome the limitations of previous studies that only utilized one of the two modalities, I integrate GWAS results with cis-regulatory interactions identified by direct multi-omics profiling of individual primary macrophages (Figure 4.1). The information is then incorporated with heterogenous transcriptomic profiles to systemically quantify the disease relevancy of different macrophage subpopulations. In this chapter, I develop an unbiased framework for functional fine-mapping of CAD GWAS results through partitioning of heritability among macrophage subpopulations. I identify a novel CD52^{hi} lipid-handling macrophage subpopulation that conveys more heritable risks of CAD, and experimentally validate their presence and enhanced lipid-processing capacity *in vitro*. The work presented in this chapter delivers the first genetic evidence for the diverse influences of different human macrophage subpopulations on atherosclerosis.

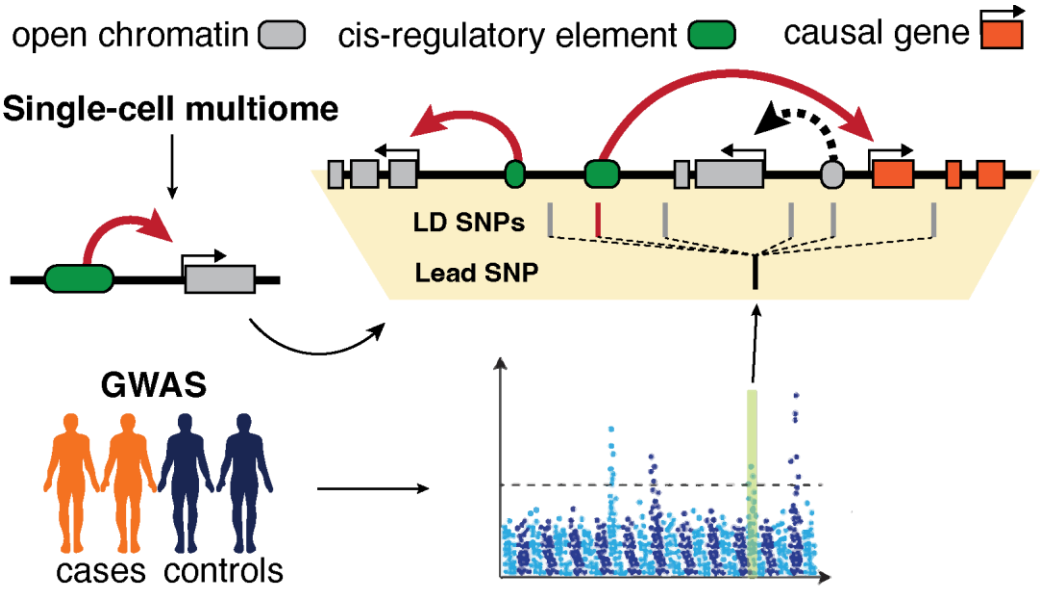


Figure 4.1 Prioritisation of causal genetic variants using macrophage cis-regulatory elements.
The linkage disequilibrium (LD) structure of each GWAS risk locus was estimated using the 1000G European panel, and all SNPs in high LD (r^2>0.8) were retrieved to form a candidate pool for prioritisation. Putative causal variants as well as its target genes were inferred by their overlaps with CREs.

4.2. Macrophage CREs prioritise causal variants and link them to target genes

Many non-coding genetic variants influence disease risks by altering the transcription-regulatory properties of the local DNA [203]. This observation has fuelled numerous efforts to prioritise GWAS variants using function annotations [204-207]. Despite the differences in methodological implementation, all these methods upweight variants overlapping functional annotations. In other word, variants residing in CREs are more likely to be causal compared to their LD buddies sharing the same GWAS signals but located elsewhere in the genome.

4.2.1. Macrophage CREs are enriched for CAD risk

To test if the CAD-associated variants are indeed enriched in macrophage CREs, I performed a permutation test using background peaks matched for GC% and accessibility. Reassuringly, CREs overlapped significantly more CAD-associated variants than comparable background peaks did (p=0.00548; Figure 4.2A). In addition, the subset of CREs associated with ox-LDL-regulated genes (ox-LDL-response CREs) were also enriched for CAD-associated variants (p=0.004455; Figure 4.2B).

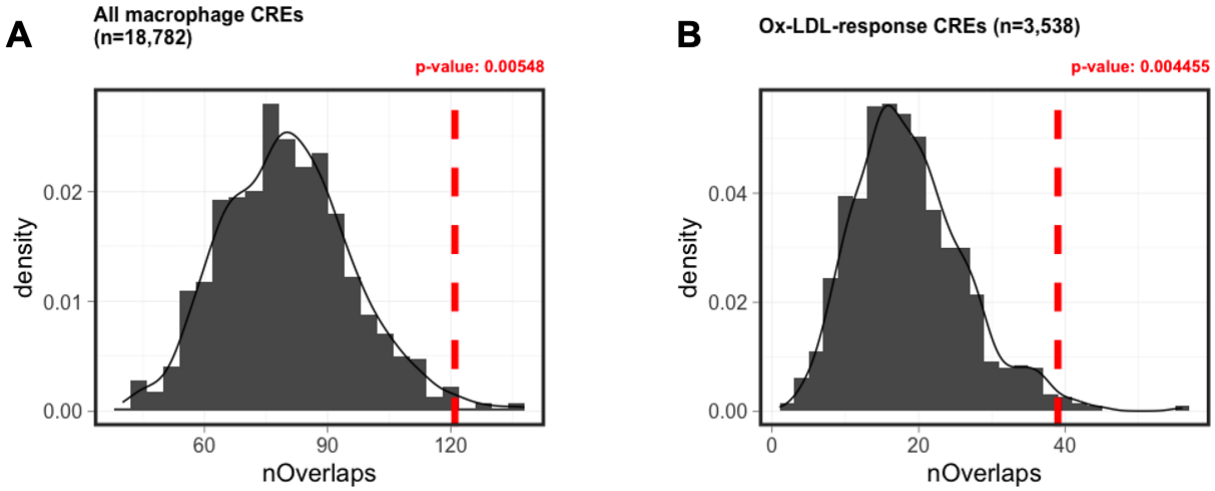


Figure 4.2 Permutation-based enrichment test for CAD risk variants.
A, Histogram showing the null distribution of overlaps between CAD-associated SNPs and n=1000 sets of background peaks matched for GC% bias and average accessibility for all macrophage CREs. The observed overlap is marked by the red dashed line and the p-value is calculated assuming Z statistics. **B**, similar to **A** but for a subset of CREs targeting ox-LDL-induced genes (ox-LDL-response CREs).

As discussed in Chapter 1.4, the extent to which GWAS test statistics are inflated across assayed DNA variants can be used to estimate the heritability of the profiled trait, also known as SNP heritability. Since risk variants with genome-wide significance ($p<5\times10^{-8}$) only explain a small portion of the genetic variance of CAD [208], I further extended this test

to evaluate the enrichment for the overall SNP heritability, rather than individual risk variants. This was achieved by applying the stratified linkage disequilibrium score regression (s-LDSC) [93] to the full GWAS summary statistics [70].

As expected, CREs and ox-LDL CREs were all enriched for CAD heritability, and this enrichment was specific to CAD compared to non-immune traits (Figure 4.3). For example, the enrichment signals for body mass index, a trait previously linked to macrophages [209], were considerably smaller than that for CAD. Additionally, no significant enrichment was observed for smoking behaviour, a neurological trait that has been associated with brain and neural cell types [210]. These results strongly support the hypothesis that macrophages and their response to ox-LDL are causally involved in the heritability of CAD. In turn, they also support the use of macrophage CREs to prioritise causal variants that increase CAD risks.

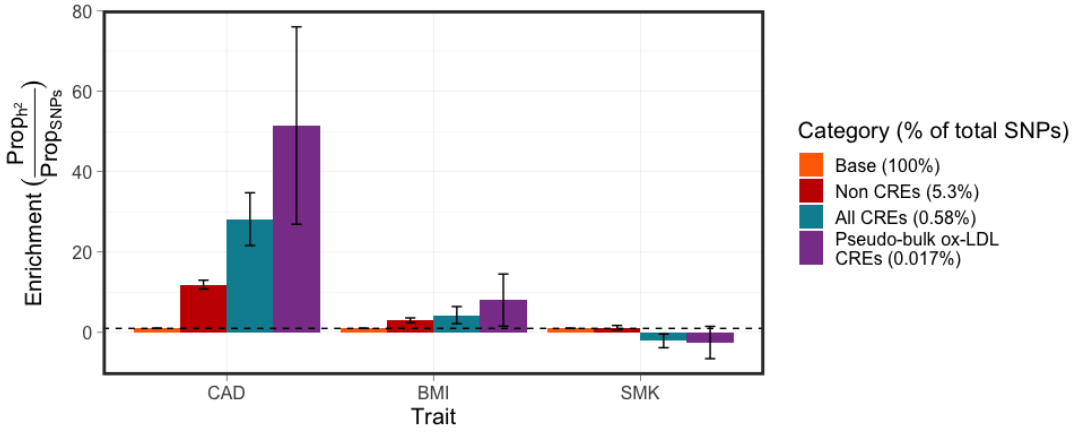


Figure 4.3 Enrichment test for CAD heritability.
CAD heritability was partitioned into different genomic regions to evaluate the enrichment for genetic risks, defined as the proportion of SNP heritability explained divided by the proportion of SNPs included. Base: whole genome; non-CREs: open chromatins without cis-regulatory interactions; CREs: cis-regulatory elements; Pseudo-bulk ox-LDL CREs: CREs targeting ox-LDL-induced genes as identified in Chapter 3. Standard errors for enrichment statistics were estimated using s-LDSC. GWAS summary statistics were obtained from GWAS Catalog using study ID GCST005194 [70], GCST006900 [211], and GCST007327 [72] for coronary artery disease (CAD), body-mass index (BMI) and smoking behaviour (SMK), respectively.

The integration of a CRE-gene map into the analysis of CAD GWAS results also links the potential causal variants to their target genes, should their functional consequences act through the local regulatory elements. Of the 9516 variants in high linkage disequilibrium (r^2 >0.8 in the 1000G EUR panel) with the lead CAD variants in the NHGRI-EBI Catalog, 121 SNPs (44 independent risk loci) were located within macrophage CREs. These SNPs were further linked to 56 cis-regulated target genes, with a total of 154 SNP-gene pairs (one CRE can regulate the expression of multiple genes; the full list of prioritised variants can be found in Online Table S5).

4.2.2. CRE-based variant prioritisation outperforms other methods

In most of the SNP-gene pairs (101 out of 156, 71.4%), the SNP and cis-regulated gene were >10 kb apart, with a median distance of 42.9 kb. In addition, the cis-regulated gene was the nearest gene to the SNP only in 48 SNP-gene pairs (34.3%). These results demonstrate that traditional proximity-based approaches (i.e., linking variants to their nearest genes) lack accuracy in identifying disease genes and may suffer significantly from false positive results.

Furthermore, although 110 of the SNPs (90.9%) within macrophage CREs were eSNPs in blood samples [175,176], only 34.3% (53 out of the 154 SNP-gene pairs) had the same target gene for both the CRE and the eQTL effect; 19 genes in the identified SNP-gene pairs were missed by blood eQTL overlaps, all of which suggesting that eQTL analysis alone will not identify the regulatory effects of many risk variants.

I then benchmarked our macrophage CRE map against state-of-the-art ABC-enhancer mapping developed from 35 public mononuclear cell datasets (based on enhancer activity

and promoter-enhancer contact information, as described in the previous chapter) [110]. While macrophages are known for their involvement in atherosclerosis [19], no enrichment for CAD risk variants was found after annotating macrophage open chromatin regions with the ABC-enhancer map (61,950 peaks, targeting 19,969 genes; Figure 4.4A). That is, prioritising variants based on overlaps with the ABC enhancer map performs no better than randomly selecting variants in macrophage open chromatin regions. Unlike what I have shown in Figure 4.2 and 4.3 with macrophage CREs, limiting ABC-enhancers to the subset whose target genes response to ox-LDL exposure did not improve the enrichment signal (9,153 peaks, Figure 4.4B). These results suggest that ABC-mapped mononuclear enhancers are limited in their power to prioritize causal CAD variants from background noise in primary macrophages.

Together, these findings demonstrate the value of using single cell multi-omics to generate a CRE map for the prioritization and interpretation of CAD risk variants. This focused approach provides greatly enhanced power and precision for investigating macrophage-specific CAD variants compared to traditional approaches.

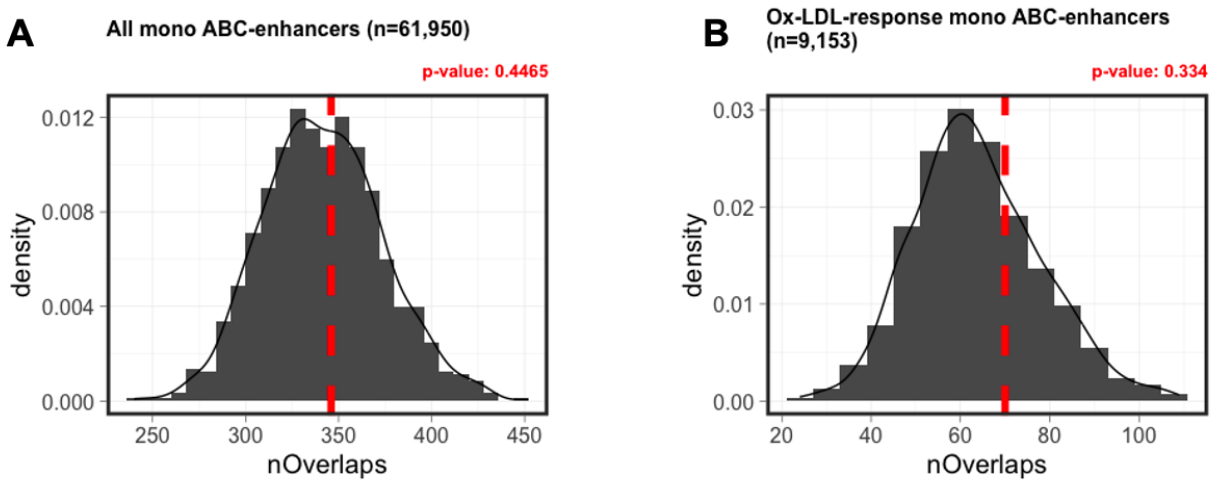


Figure 4.4 Permutation-based enrichment test for CAD risk variants using annotations from Activity-By-Contact (ABC) enhancers.
Similar to that in Figure 4.2, the empirical null distribution was estimated by overlapping CAD-associated variants with n=1000 matched background peak sets. The observed overlap in all mononuclear ABC enhancers (**A**), or in ABC-enhancers targeting ox-LDL-induced genes (**B**) is marked by the red dashed line. P-values are calculated assuming Z statistics.

4.2.3. CAD-associated transcriptional regulatory networks

While these genetic risk loci are statistically independent, their target genes do not function in isolation. To determine if and which prioritised risk genes are co-regulated by the same TF, I closely inspected a subset of macrophage transcriptional regulatory networks where the DORC CREs overlapped CAD-associated variants. All strong regulatory interactions (|regulatory score| > 1.5) for these CAD DORCs were extracted and visualised on a node graph in Figure 4.5. This network centred around members of the AP-1 family, with FOS and JUNB acting as repressors and FOSL1, ATF4, and JUND as activators. FOS in particular, was shown to repress the expression of multiple CAD DORCs. Other TF families, including CEBP and KLF, also targeted multiple CAD DORCs. The TF-DORC relationships can be complex; for example, MITF, one of the CAD DORC

regulated by AP-1 factors, is itself a TF regulator and activates a wide range of downstream DORCs (Figure 4.5).

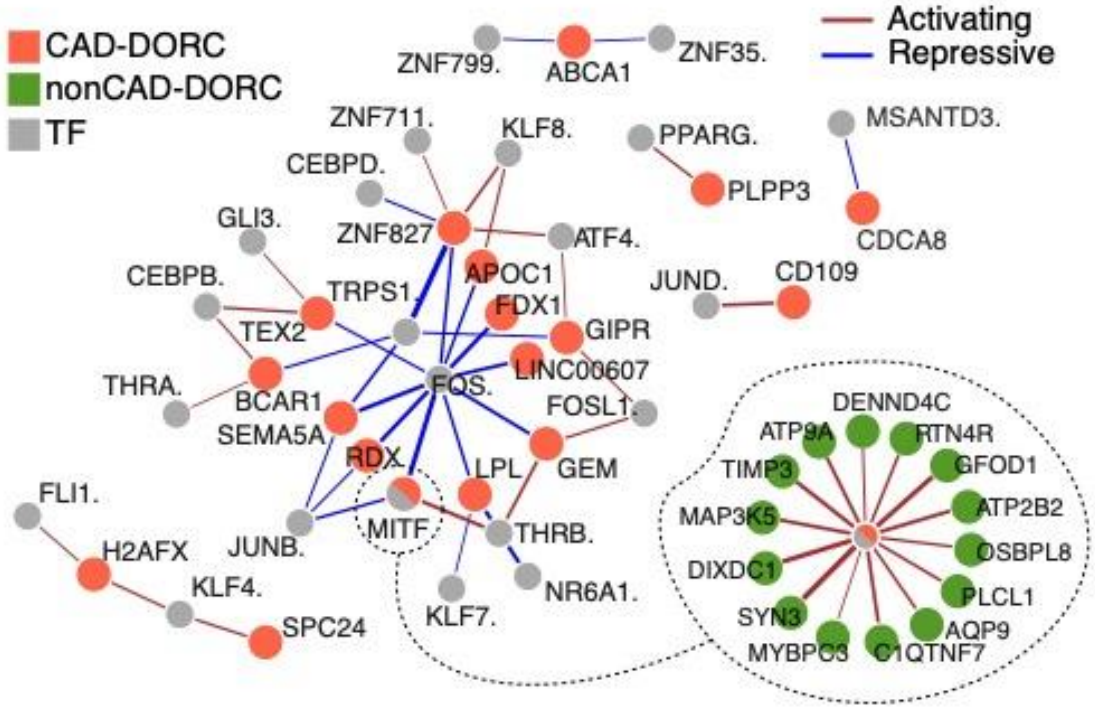


Figure 4.5 Network visualization of significant TF-DORC pairs for CAD-DORCs. Nodes represent CAD-DORCs (orange) and TF regulators (grey) in TF-DORC pairs with an absolute regulation score >1.5. Non-CADDORCs regulated by MITF are plotted separately (green discs). Red lines indicate activating effects, and blue lines indicate repressive effects. Line widths are weighted by the regulation score.

4.3. Development of a computational framework of partitioning disease heritability

As demonstrated in the previous chapter, macrophage subpopulations differ in their transcriptomic and open chromatin profiles, and can have distinct TF activation patterns. Therefore, I reasoned that this would make them differ in the extent to which they are

functionally influenced by CAD-risk variants. In other words, despite all carrying the same variants, the contribution of different macrophage subpopulations to CAD heritability might be different.

In accordance with this hypothesis, more than half of the prioritised variants (29 out of the 44 independent loci) were linked to subpopulation-specific genes (Figure 4.6). For example, 6 risk loci overlapped CREs whose target genes were specially expressed in lipid-handling macrophages, indicating that their disease risk-modifying influence is likely to be specific to this subpopulation.

However, due to the differing statistical power in differential expression tests, applying the same significant cut-off (e.g., adjusted p-value <0.05) inevitably leads to marker gene sets with various sizes (Figure 4.6, left panel). This means that attempts to prioritise disease-critical subpopulations based on significant DE results will be biased towards larger subpopulation with more distinct transcriptomic signatures simply because of higher statistical power, regardless of their actual contribution to the disease heritability.

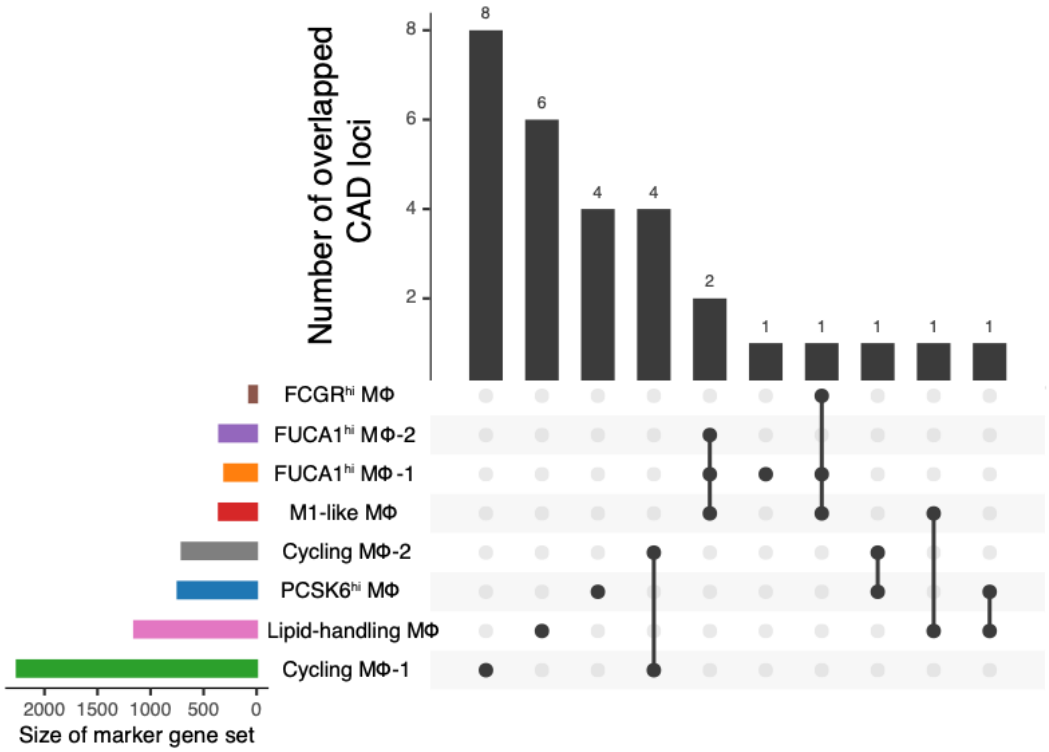


Figure 4.6 Number of independent CAD loci mapped onto CREs targeting subpopulation-specific genes.
Each set of subpopulation marker genes was binarily selected by a differential gene expression test under the FDR threshold of 0.05. Rows represent marker gene sets and intersections between different sets are plotted as connected dots. The size of each intersection is plotted on the top of the matrix, and the size of each marker gene set is plotted to the left.

To overcome this issue and to test the statistical significance of the subpopulation-specific patterns in Figure 4.6, I developed a computational framework (Figure 4.7) that integrates single-cell multiome data and CAD GWAS results based on the published s-LDSC method[147]. This framework involves two major steps: 1. Generate genetic annotations from single-cell sequencing data; 2. Partition CAD hereditability based on different annotations and test for significance.

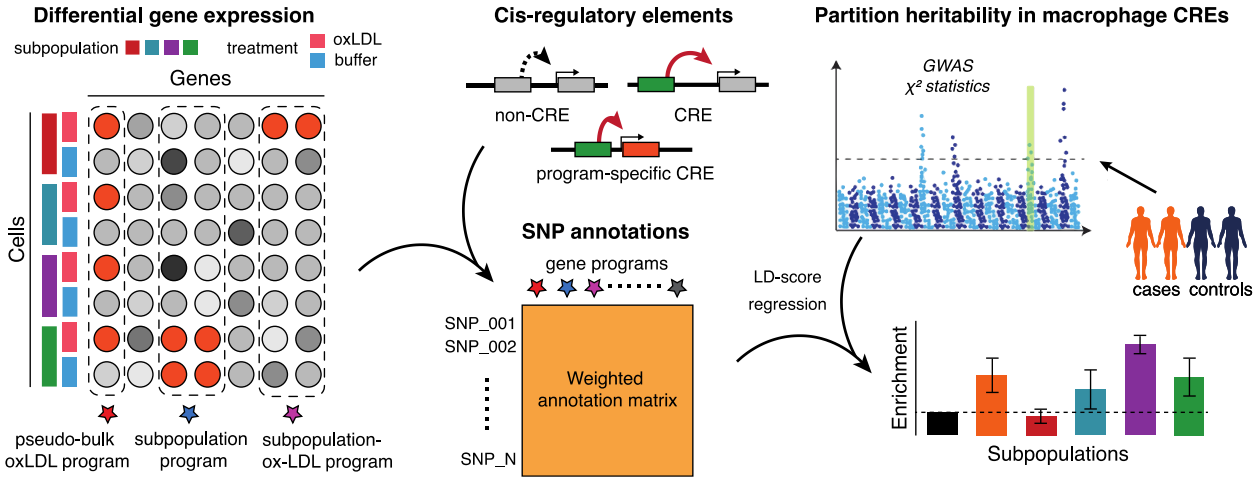


Figure 4.7 GWAS heritability enrichment framework.
Left: Deriving macrophage subpopulation gene programmes, pseudo-bulk ox-LDL gene programme, and subpopulation-ox-LDL gene programmes from single-cell transcriptomic data. **Middle:** Converting gene programmes into weighted genomic regions using the macrophage CRE map (top), then constructing SNP annotation matrices by assigning the corresponding weights to 1000G SNPs overlapping each region (bottom). **Right:** the relative enrichment of GWAS heritability for stratified subpopulation-specific / subpopulation-specific-ox-LDL-specific annotations are tested using the LD score regression model[93].

The idea of partitioning disease heritability into different functional annotations is largely inspired by published work in the field that integrates cell-type-specific information with the standard LDSC analysis [80,197]. As the CAD heritability is already shown to be enriched in macrophage CREs (Figure 4.3), I reasoned that it could be further stratified to evaluate the genetic contribution of each macrophage subpopulation to the disease.

In principle, the genetic contribution of macrophages can be evaluated based on their gene expression patterns. In our dataset, the gene expression patterns reflect both the baseline and the ox-LDL-response signatures in the transcriptome. To quantify these gene expression patterns, I applied the Pi strategy previously published by our colleagues at Oxford [146] by scoring differential gene expression statistics (see Methods). Unlike

GWAS-identified variants which always have one risk allele and one protective allele, the directional effects of GWAS-linked genes are largely unknown. Therefore, I treated the up-and down-regulated genes equally when scoring differential expression genes. The derived continuously-weighted gene sets constitute gene programmes, which characterise either the baseline expression of a subpopulation (termed a subpopulation gene programme), or the response of a subpopulation to ox-LDL (termed a subpopulation-ox-LDL gene programme). As a comparator, an ox-LDL-response programme was constructed from the DE test of the pseudo-bulk populations; this programme captured the generic response of the average macrophage population to ox-LDL, while the heritability enrichment in this programme was plotted previously in Figure 4.3.

Subsequently, the gene programmes were converted into weighted genomic regions by assigning each CRE with the sum of programme scores for all its target genes. Scores were passed to individual SNPs in the 1000G reference panel that will be used in the subsequent LD score regression analysis, in the form of annotation matrices required by the LDSC software. I next performed a conditional s-LDSC analysis using a multi-variate regression model to disentangle the conditional independent signal from a specific programme from the aggregated enrichment signals. As demonstrated in the original s-LDSC paper [147], the average heritability per programme can be recovered from the coefficient estimates of this multiple regression model. More specifically, heritability enrichment was defined as the proportion of SNP heritability explained divided by the proportion of SNPs included in each annotation matrix. As subpopulation-specific annotation matrices were weighted on a continuous scale, I also computed the

standardized effect sizes τ*, defined as the proportional change in per-SNP heritability with a 1 standard deviation increase in the SNP annotation weight, to assess the strength and statistical significance of heritability enrichment (see Chapter 2.5.10 for detailed methods) [147,197].

4.4. Lipid-handling macrophages show enrichment in both subpopulation and subpopulation-oxLDL programmes

Compared to the background annotation constructed using all macrophage open chromatin regions, subpopulation gene programmes show higher enrichment for CAD heritability (Figure 4.8; this relative enrichment over background was also visualised as the dot sizes in Figure 4.9A). This is expected as regulatory elements have been widely reported to be enriched for disease and complex trait heritability [93].

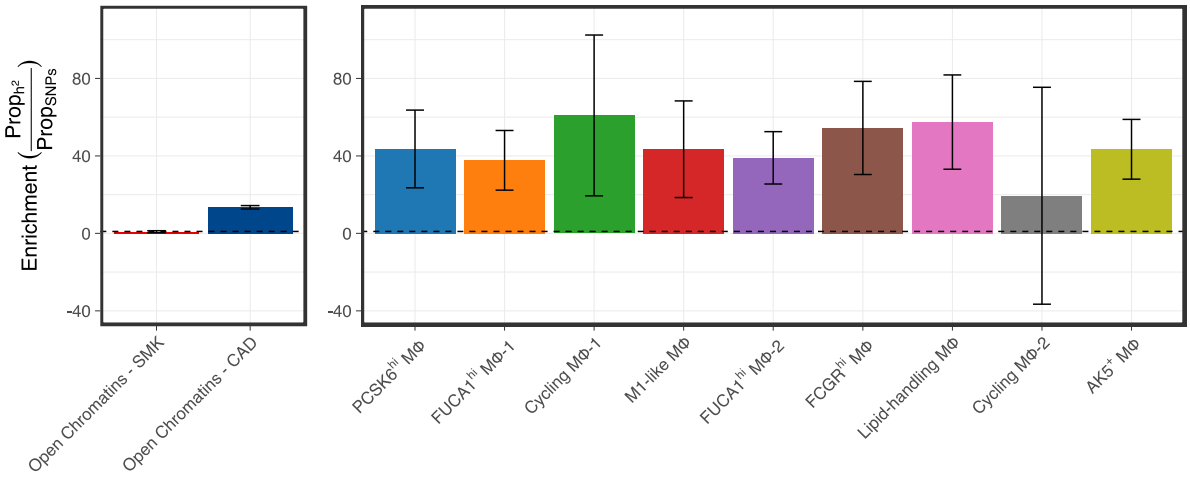


Figure 4.8 The enrichment for CAD heritability in subpopulation gene programmes. The left panel shows the heritability enrichment in macrophage open chromatin regions for smoking behaviour (a non-immune polygenic trait reflecting the expected null distribution) and CAD as shown in Figure 4.3.

Using a multivariate linear regression model controlling subpopulation programmes on each other, I further analysed the conditional independent signals of each programmes to the overall enrichment for CAD heritability. The results showed that CAD heritability was relatively depleted (τ*=-2.24) in FUCA1^{hi} macrophages, and remained significantly enriched (τ*=1.28) in lipid-handling macrophages (Figure 4.9A). This suggests that the baseline expression patterns in lipid-handling macrophages convey higher genetic risks for CAD with respect to other macrophage subpopulations.

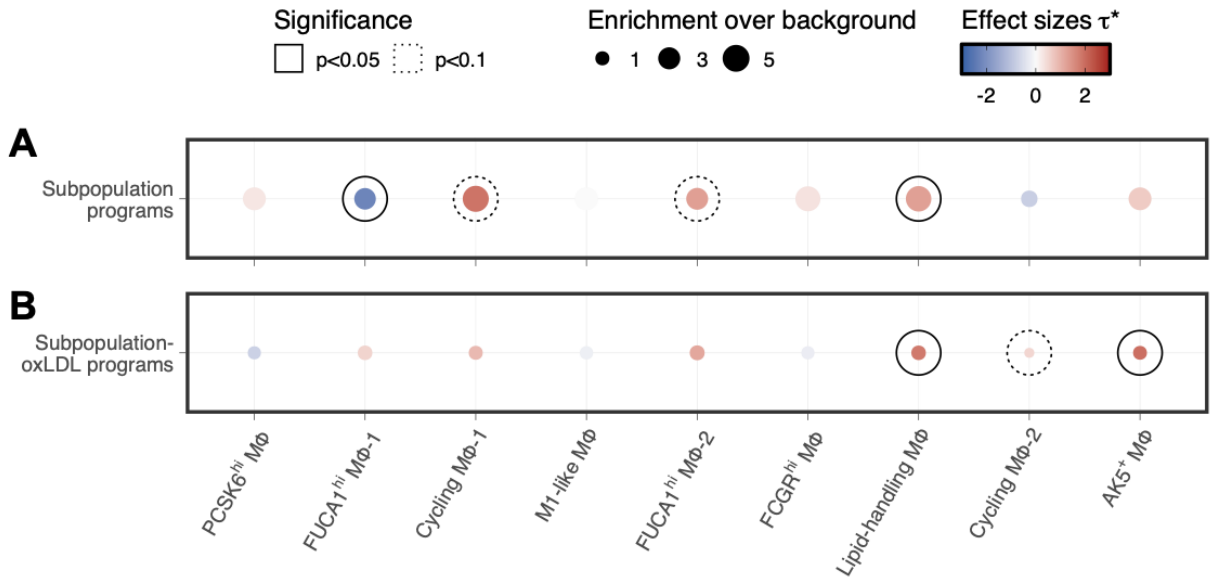


Figure 4.9 Enrichments for CAD heritability in gene programmes.
A, heritability enrichment in subpopulation programmes as compared to that in macrophage open chromatins. Dot size denotes the relative enrichment over background annotations. Colour denotes the standardised effect sizes (see text) after conditioning on each other. **B**, heritability enrichment in subpopulation-ox-LDL programmes as compared to that in the pseudo-bulk ox-LDL-response program. Colour denotes the standardised effect sizes after conditioning on the pseudo-bulk program. Nonzero τ* estimates reported at two significance levels (solid: p<0.05, dotted: p<0.1).

Unlike the subpopulation programmes that were diverse and distinguishable from one another (Figure 4.10A), the subpopulation-ox-LDL gene programmes were highly correlated (Figure 4.10B). This means that macrophage subpopulations shared an overall ox-LDL-response, and it's statistically underpowered to delineate their independent contribution. Therefore, instead of conditioning on each other, I used the pseudo-bulk ox-LDL gene programme as the reference to compute the standardized effect size τ*, measuring the performance of adding subpopulation-specific information on top of the shared ox-LDL response. For most subpopulations, τ* no longer deviated from zero (Figure 4.9B), with a relative enrichment close to 1 (as visualised by the dot sizes in Figure 4.9B), meaning that the ox-LDL effect in these subpopulations is largely not subpopulation-specific. However, there was still significant enrichment of heritability in the ox-LDL response in lipid-handling and AK5^{+} macrophages, with τ* of 1.81 and 1.97, respectively (Figure 4.9B).

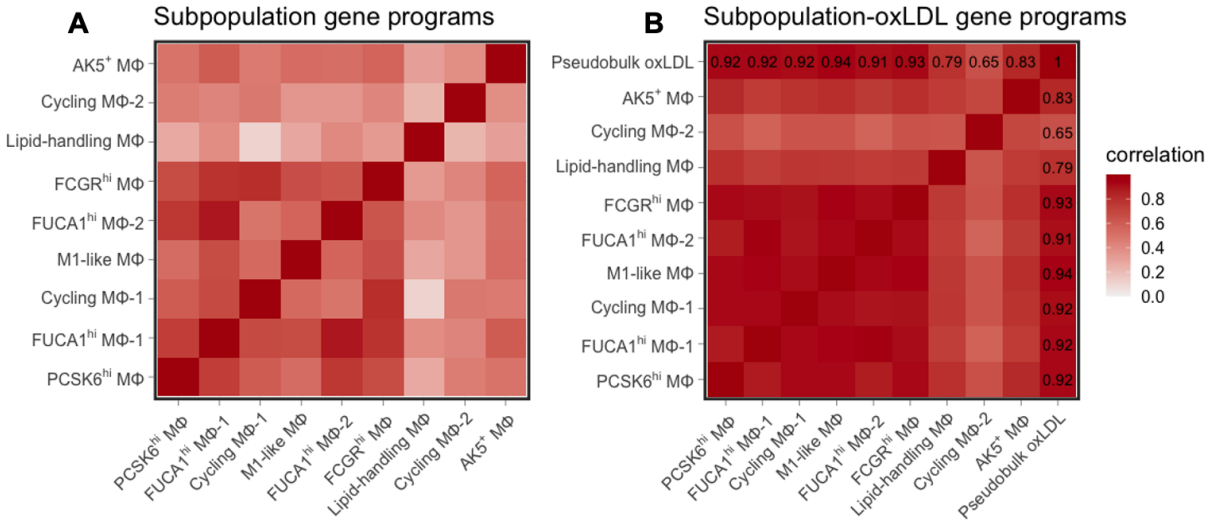


Figure 4.10 Heatmap showing the correlation (Pearson correlation coefficient) between the SNP annotation matrices.
Annotation matrices constructed using subpopulation-programmes (**A**) or subpopulation-ox-LDL programmes (**B**). High correlation means two gene programmes are genetically similar.

Overall, our heritability analysis demonstrated that lipid-handling macrophages contribute more to the heritable component of CAD than other macrophage subpopulations, both in unexposed state and upon exposure to ox-LDL.

4.5. LHM display transcriptomic and epigenomic signatures related to lipid metabolism

So far, I have shown that CAD heritability is not evenly distributed across heterogenous macrophage subpopulations. In addition, the identified increased heritability in gene programmes of the lipid-handling macrophages leads to a more mechanistic question: what is different in these cells compared to other macrophage subpopulations?

As described in the previous Chapter, lipid-handling macrophages were named by their transcriptomic signatures associated with lipid-metabolism genes. For example, these cells displayed a significantly increased expression of apolipoprotein *APOC1* and lipoprotein lipase *LPL* in both snRNA-seq (Figure 4.11A) and scRNA-seq data (Figure 4.11B). A closer inspection of the KEGG and GO-BP pathways enriched in their marker genes identified enhanced phagocytosis (Figure 4.12A) and fatty acid transport (Figure 4.12B) within these cells. Of note, the membrane glycoprotein *CD52* and non-enzymatic chitinase-3 like-protein-1 *CHI3L1* were also uniquely expressed in lipid-handling macrophages (Figure 4.11A and B), while their roles in macrophage lipid metabolism, if any, are uncharacterized.

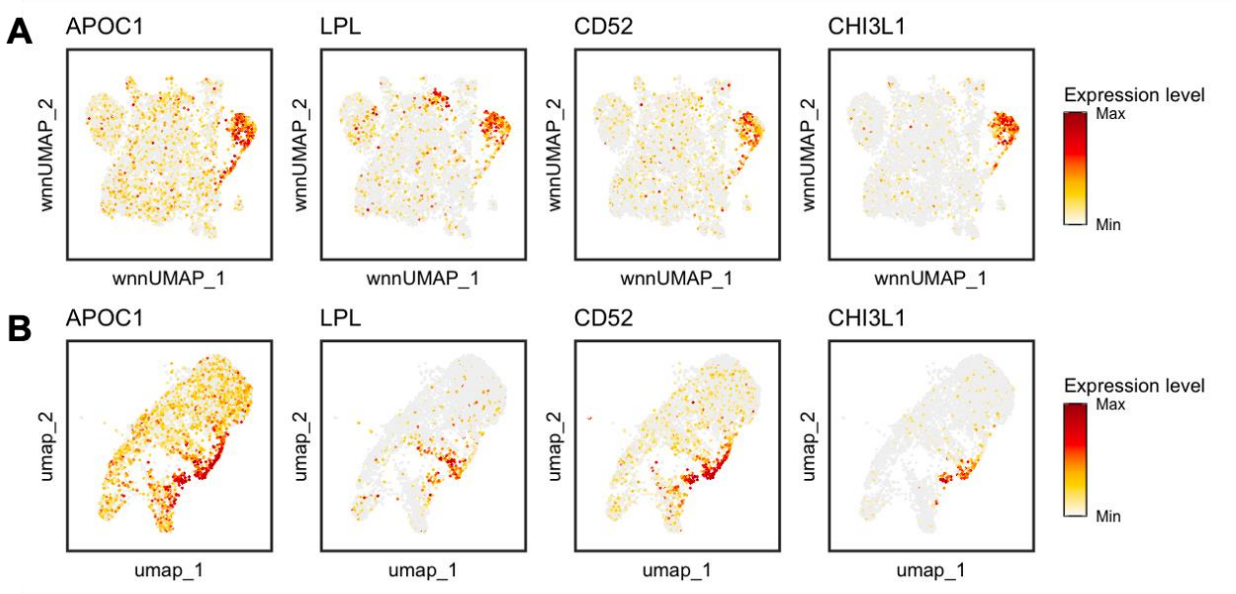


Figure 4.11 UMAP visualisation of the expression of selected marker genes for lipid-handling macrophages.
Single-nucleus (**A**) and single-cell (**B**) expression was min-max normalised and plotted on the corresponding UMAP, demonstrating the co-expression of *APOC1*, *LPL*, *CD52* and *CHI3L1* in only a subset of cells.

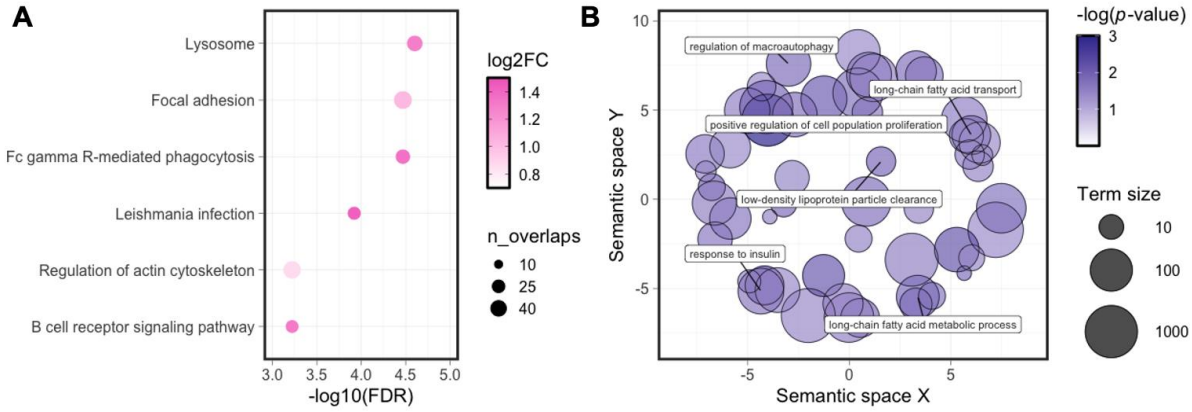


Figure 4.12 Pathways enriched in the marker genes of lipid-handling macrophages.
A, dot plot showing enriched KEGG pathways, with colour representing log2 fold change and size representing the number of overlapped genes in each pathway. **B**, enriched Gene Ontology – Biological Process (GO-BP) terms projected onto a 2D semantic space, showing clusters of GO-BP terms associated with fatty-acid metabolisms. The similarity between different terms is summarised using REVIGO[140] and can be represented by their distance in the semantic space.

A recent study has recovered more than 200 co-expression modules from bulk RNA-seq of a large CAD cohort [212], and have linked the expression of these gene modules to a wide range of CAD-related phenotypes in patients. Interesting, one of such gene modules (GRN51, ref:) significantly overlaps with the marker genes of lipid-handling macrophages (FDR=1.42×10^{-20}; 133 marker genes with FDR < 0.05 and log2FC > 1 tested). This gene module was mostly expressed in subcutaneous fat and liver (Figure 4.13A), with expression strongly correlates with BMI and blood triglyceride levels in CAD patients (Figure 4.13B).

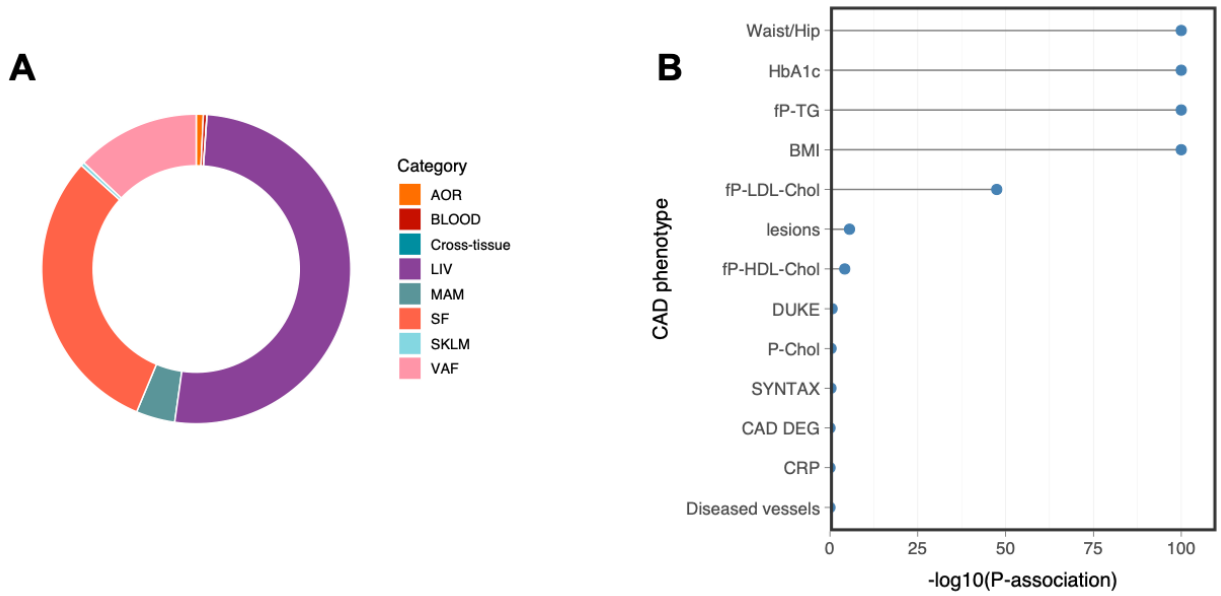


Figure 4.13 The co-expression gene module 51 identified in cross-tissue sequencing of a large CAD cohort[212].
A, The tissue of origin (shown as the percentage of tissue-specific genes) for gene nodes in this co-expression module. LIV: liver; MAM: free internal mammary artery; AOR: atherosclerotic aortic root; SF: subcutaneous fat; VAF: visceral abdominal fat. SKLM: skeletal muscle. **B**, Associations between the gene module expression and CAD phenotypes. The module expression was calculated using all samples collected from expressing tissues as shown in (**A**). fP-TG, fasting plasma triglyceride levels; fP-LDL-Chol, fasting plasma LDL cholesterol levels; fP-HDL-Chol, fasting plasma HDL cholesterol levels; P-Chol, plasma cholesterol levels; CRP, C-reactive protein.

The transcriptional regulators specific to lipid-handling macrophages was also identified using the motif deviation analysis described in the previous chapter. CEBP family TFs showed the largest motif deviation score for these macrophages (Figure 4.14A), with CEBPB expression correlating positively with CEBP motif binding activity (Figure 4.14A and B). A strong correlation was also observed between expression and motif usage for NR1H3 (Figure 4.14A and C), a key transcriptional regulator of genes involved in fatty acid metabolism [213].

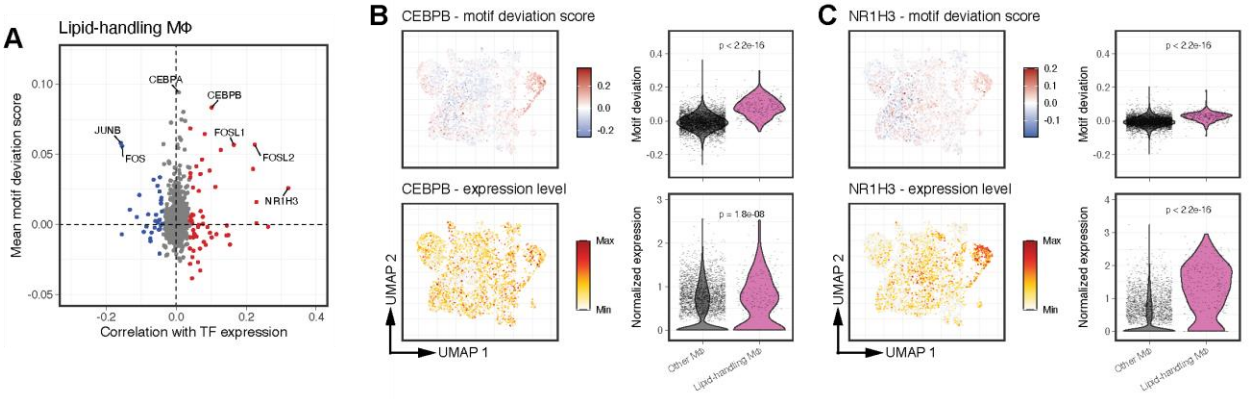


Figure 4.14 Motif analysis for transcription factors (TF) driving the expression of marker genes of lipid-handling macrophages.
A, Mean TF motif deviation scores (y axis, the relative enrichment compared to background) compared with their correlation with TF RNA expression (x axis, Pearson coefficient) in lipid-handling macrophages. TFs are coloured by significance (FDR< 0.05) for negative (blue) or positive (red) correlation. **B**, Motif deviation scores (**top**) and RNA expression (**bottom**) for *CEBPB*. **C**, Motif deviation scores (**top**) and RNA expression (**bottom**) for *NR1H3*. P-values in **B** and **C** are computed using Wilcoxon tests.

4.6. LHM show much less lipoprotein accumulation after ox-LDL exposure

To functionally validate these findings, I next tested whether lipid-handling macrophages can be separated by their marker genes *in vitro* using immunofluorescence staining. Indeed, a strong co-expression of CD52 and CHI3L1 was observed exclusively in a

subset of macrophages (Figure 4.15A and B). In addition, a specific expression of the well-documented ox-LDL uptake receptor CD36 [214] was also confirmed at the protein level in un-exposed CD52^{+} cells (Figure 4.15C and D).

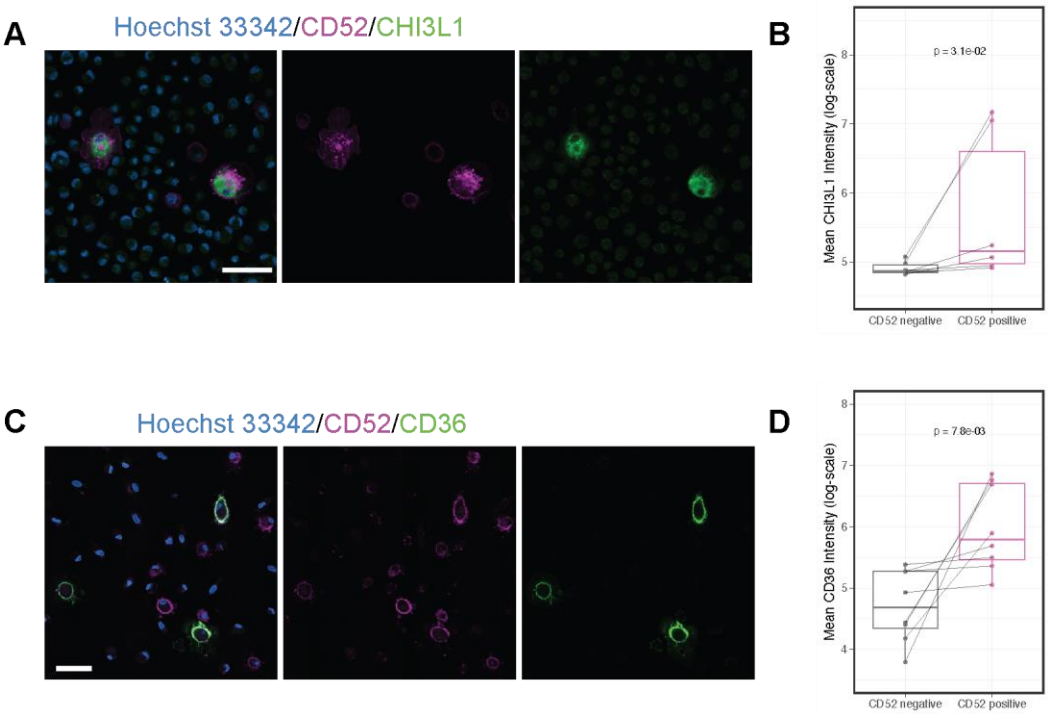


Figure 4.15 Experimental validation of CD52, CHI3L1 and CD36 as markers for lipid-handling macrophages.
A, Representative immunostaining of primary macrophages for lipid-handling macrophage markers CD52 and CHI3L1. **B**, quantification of the CHI3L1 signals in **A**, showing significant up-regulation of CHI3L1 in CD52^{+} cells (n=6 biological replicates, including 2 obtained by Dr. Thomas Hiron). **C**, Representative immunostaining for lipid-handling macrophage markers CD52 and CD36. **D**, quantification of the CD36 signals in **C**, showing significant up-regulation of CD36 in CD52^{+} cells (n=8 biological replicates). Note that cells in **A** were treated with protein transport inhibitor to stain the secretive CHI3L1 and thus have different morphology compared to cells in **C**.

Next, a lipid-uptake assay was performed using DiI-labelled ox-LDL, and CD52^{+} macrophages demonstrated significantly reduced accumulation of intracellular ox-LDL in compared to CD52- macrophages (Figure 4.16A and Figure 4.16B). I repeated the assay

with live-cell imaging techniques and found this reduced lipid accumulation was visible as early as 2 hours post ox-LDL exposure (Figure 4.16C, Online Video).

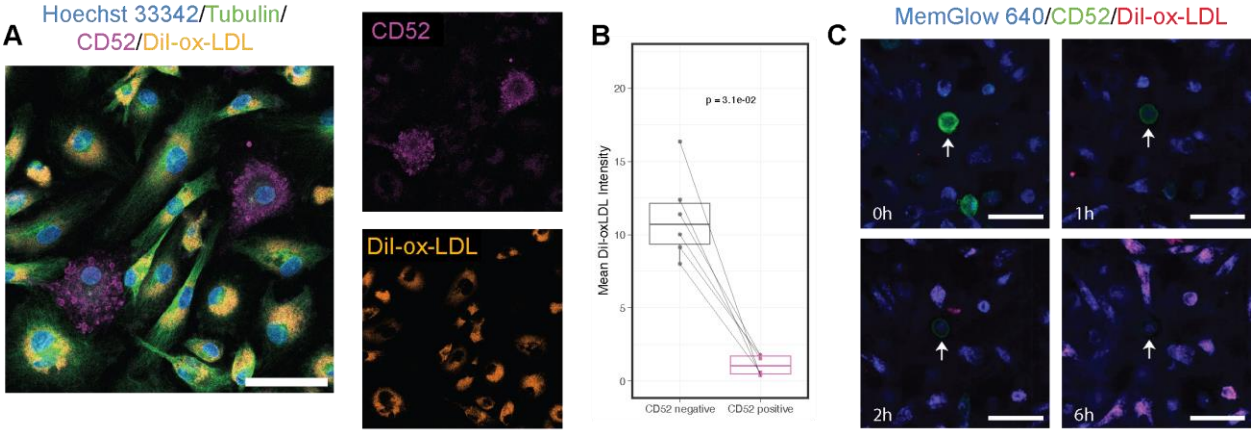


Figure 4.16 CD52⁺ lipid-handling macrophages displayed reduced for lipid accumulation after ox-LDL exposure.
A, Representative immunostaining of primary macrophages after 48 hr treatment with DiI-ox-LDL. **B**, Quantification of the intracellular level of DiI-ox-LDL in (**A**), shown as average fluorescence intensities (arbitrary units) per biological replicate (n=6). The initial IF staining with 3 biological replicates were obtained by Dr. Thomas Hiron. **C**, Representative live-cell imaging of primary macrophages for lipid accumulation at different time points post DiI-ox-LDL exposure, with one CD52⁺ cell per view marked by the white arrow. Cells were briefly incubated with Fc-muted CD52 antibody for 1 hour prior to the assay to minimize the effect of antibody binding. Scale bar: 50μm. P-value calculated using a Wilcoxon test in **B.**

After ox-LDL treatment, lipid-handling macrophages exhibited increased expression of *CD36* (Figure 4.17), a trend also seen in other subpopulations. The up-regulation of *CD36,* even in the unexposed state, would be expected to increase ox-LDL uptake, so the reduced ox-LDL accumulation observed is consistent with CD52⁺ lipid-handling macrophages having more efficient lipid/cholesterol processing and clearance capacities.

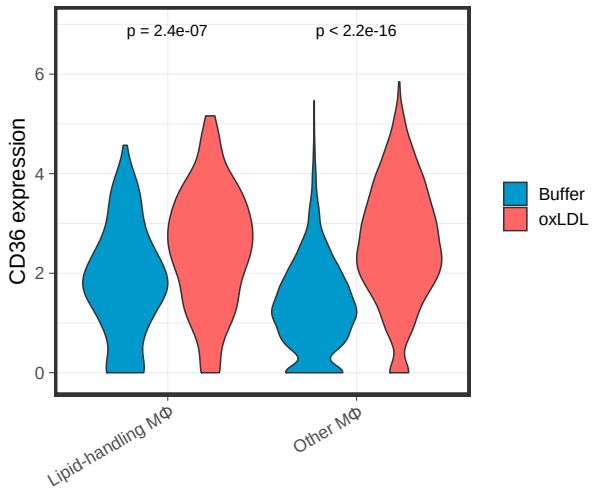


Figure 4.17 The expression of CD36 in lipid-handling macrophages as well as in other macrophages.
Gene expression was normalized by sequencing depth (count per 10k reads) and log1p transformed. P-values were calculated with Wilcoxon test.

4.7. A CD52-high lipid-handling phenotype is robustly present in human atherosclerotic plaques

Most macrophages in atherosclerotic lesions are derived from circulating monocytes [17,154,215] as were the macrophages profiled in this work. However, unlike our experimental system where macrophages are differentiated with minimum disturbance, macrophages in plaques are exposed to a mixture of atherogenic and pro-inflammatory influences, so might have very different transcriptomic profiles. Therefore, it is useful to evaluate the transcriptomic similarities between plaque macrophages and monocyte-derived macrophages, and to determine whether the identified CD52^{hi} lipid-handling phenotype is a reproducible feature of macrophages in human atherosclerosis.

I started with a meta-analysis of scRNA-seq data from 38 human coronary and carotid artery plaque samples from 5 independent studies [25-27,33,134]. After read alignment, sample QC and doublet removal, I retained a total of 126,685 cells, which clustered into 8 broad

cell types after batch effect correction (Figure 4.18A and B). I then focused on myeloid cells (n=20,458) based on expression of *CD14*, *CD68* and *AIF1*.

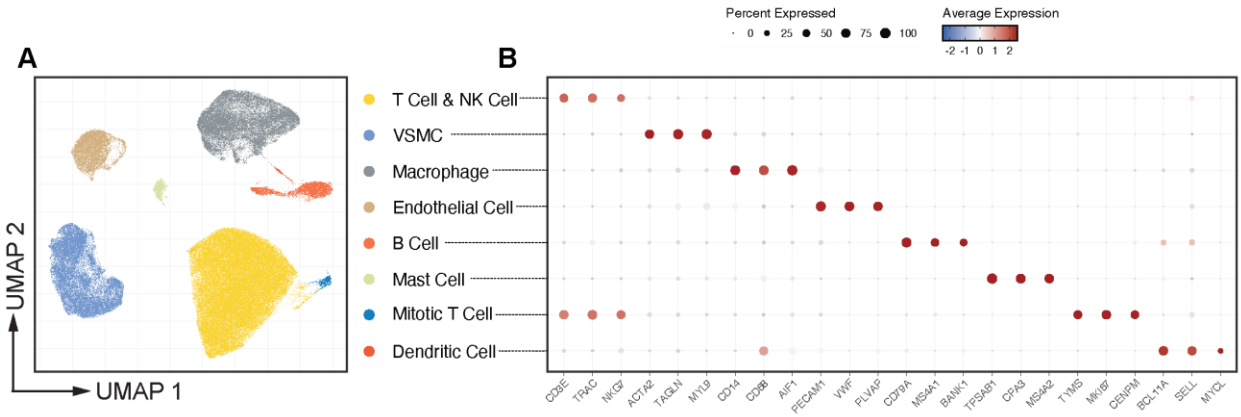


Figure 4.18 Meta-analysis of scRNA-seq profiles in 38 human atherosclerotic samples.
A, UMAP visualisation of the human plaque cell atlas. **B**, The expression of reported markers for different cell types in (**A**) is summarized in the dot plot, where the colour represents the average gene expression, normalized by sequencing depth and scaled across all cells; the size represents the percentage of cells expressing the gene.

Next, I re-integrated the scRNA-seq data to replicate the macrophage subpopulations reported in previous studies [42,134,216]. In total, 6 heterogeneous macrophage clusters were identified (Figure 4.19 A and B) including two S100A8^{hi} clusters S100A8/S100A12^{+} and S100A8/S100A12- macrophages, an IFNG-activated cluster, an IGSF21^{+} cluster and two TREM2^{+} clusters expressing marker genes associated with lipid metabolism (*SPP1*, *CD9*, *FABP5*). These two clusters were termed as lipid-associated macrophages (LAM) and PLIN2^{hi}/TREM1^{hi} macrophages for consistency with previous reports [28,134]. I also identified 3 previously reported clusters for dendritic cells, termed cDC1, cDC2 and mature cDC2 [134].

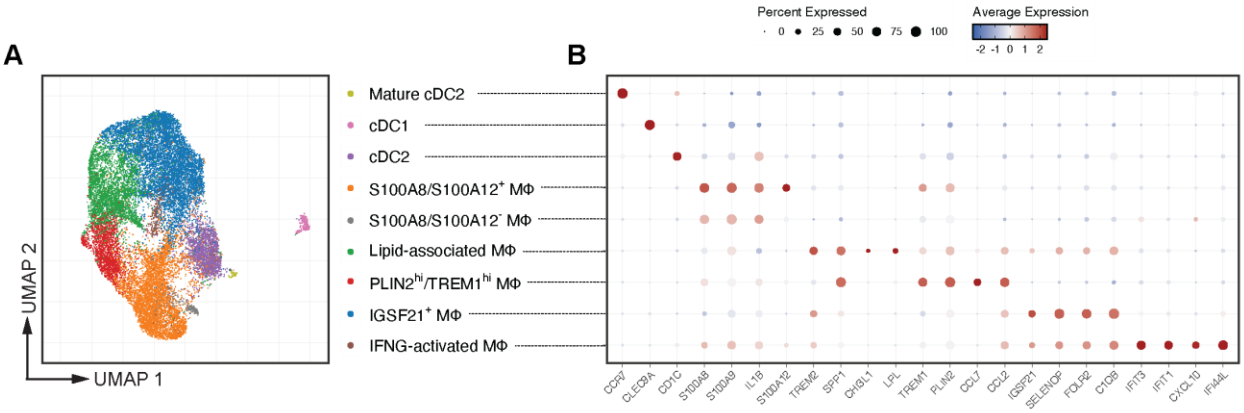


Figure 4.19 Plaque macrophage atlas constructed from the meta-analysis.
A, UMAP visualisation of different macrophage subpopulations in human atherosclerotic plaques. Clustering resolution was chosen to reproduce subpopulations identified previously[25-27,33,134]. **B**, The expression of selected marker genes for subpopulations in (**A**) is summarized in the dot plot, where the colour represents the average gene expression, normalized by sequencing depth and scaled across all cells; the dot size represents the percentage of cells expressing the gene in the subpopulation.

To investigate how monocyte-derived macrophages map to *in vivo* plaque macrophages, I conducted an integrative analysis using our primary macrophage scRNA-seq profiles and the plaque macrophage atlas from the meta-analysis. Surprisingly, most of the monocyte-derived macrophages, both unexposed and exposed to ox-LDL, mapped closely to the TREM2^{+} LAM population in plaques (Figure 4.20A). To further validate this inferred linkage, I trained a machine-learning macrophage classifier with a deep generative model [138] on the plaque macrophage atlas, and found almost all *ex-vivo*-derived macrophages were predicted to be LAM with high confidence (Figure 4.20B). Analysis of the pseudo-bulk gene expression profiles also supported this transcriptomic similarity, with the correlation coefficient between LAM and *ex vivo* macrophages higher than all other macrophage populations in the plaque (Figure 4.20C). Importantly, when I extended the heritability analysis to plaque macrophage populations, LAM, like monocyte-derived macrophages, showed a strong enrichment for CAD heritability (Figure 4.20D).

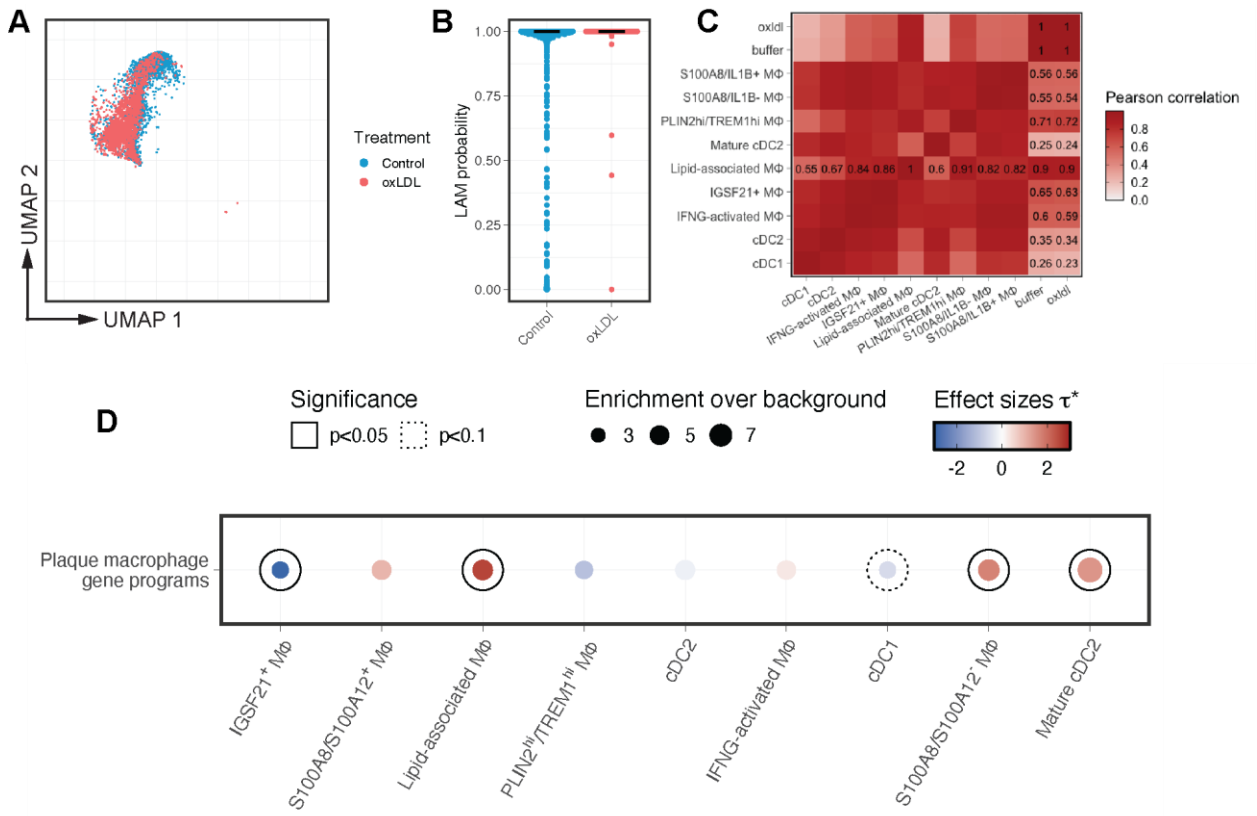


Figure 4.20 Monocyte-derived macrophages closely mirror the transcriptomic signatures of the TREM2^{hi} lipid-associated macrophage (LAM) in plaque.
A, Single-cell transcriptomic profiles of monocyte-derived macrophages projected onto the plaque macrophage atlas, coloured by treatment. **B**, Predicted probabilities of *ex vivo* macrophages being classified as plaque LAM, shown as individual data points along with the median, coloured by treatment. **C**, Heatmap showing the correlation between the pseudo-bulk transcriptomic profiles of plaque macrophage subpopulations, unexposed and ox-LDL-exposed *ex vivo* macrophages. Colours represent the Pearson correlation coefficient and values for comparisons involving LAM and *ex vivo* macrophages are shown as text. **D**, Enrichments for CAD heritability in subpopulation gene programmes of plaque macrophages. Dot size denotes the relative enrichment over all macrophage open chromatin regions. Colour denotes the standardized effect sizes after conditioning on each other. Nonzero τ* estimates reported at two significance levels (solid: p<0.05, dotted: p<0.1).

As have been demonstrated in the previous chapter, macrophages derived from circulating monocytes possess striking heterogeneity in both their transcriptomic and epigenomic profiles. Given that they robustly mapped to the LAM population in plaques,

I next asked if the plaque LAMs would also be heterogeneous with functionally distinct subpopulations. With the greatly increased sample size in our meta-analysis, I anticipated that macrophage heterogeneity could be investigated at a much higher resolution than previously achieved, leading to the discovery of rare populations not characterized in previous studies.

To ensure the robustness and reproducibility of newly identified macrophage subpopulations, I performed a semi-automated cluster stability analysis to determine the optimal clustering condition. The assumption is that if a subpopulation is robust, random subsampling and re-clustering will maintain the cell identities within this subpopulation, thus the cluster should be stable.

After testing 48 different sets of n_PCs, resolution and the k value for KNN graphs in n=50 subsampled datasets, I identified 4 stable subpopulations within plaque LAMs (n=3,645, Figure 4.21A and 4.21B) in both coronary and carotid artery samples. In general, these LAM subpopulations broadly shared the gene expression patterns that had been identified in various *ex vivo* subpopulations, including a MHC^{hi} phenotype, a proliferating phenotype, and a resting phenotype (Figure 4.21C). Notably, a subset of $CD52^{hi}$ LAM displayed strong expression for genes involved in lipid processing and metabolism (*LPL*, *CD36*, *PPARG*) (Figure 4.21C and D). Although *CD52* was expressed in more than one plaque macrophage subpopulation, the co-expression of *CD52*, *CHI3L1* and *LPL* was only observed in $TREM2^{+}$ LAMs (Figure 4.19B). This is consistent with the lipid-handling phenotype we identified in *ex vivo* macrophages being a unique feature specific to LAMs.

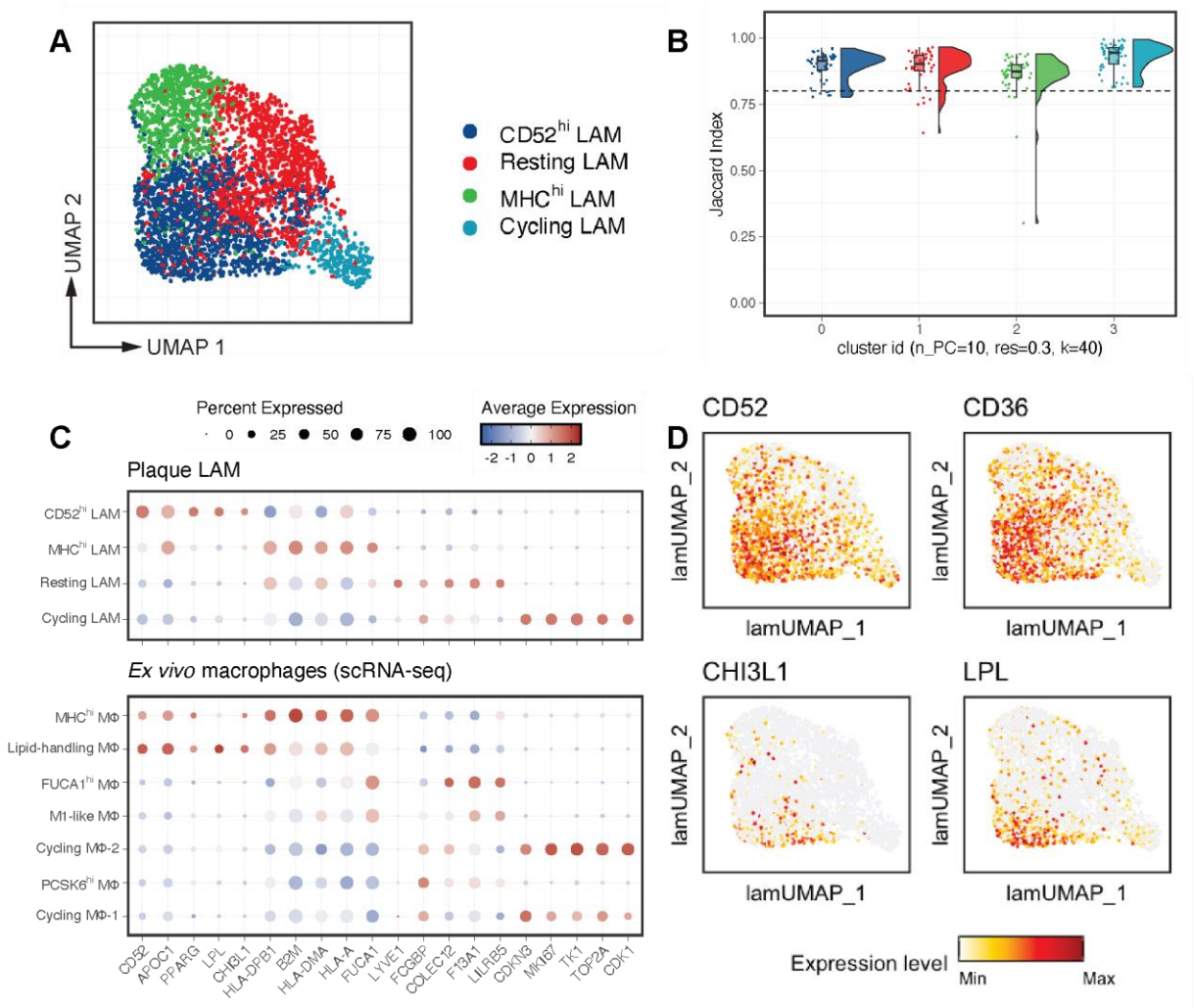


Figure 4.21 The CD52^{hi} lipid-handling phenotype is present in human atherosclerotic plaques.

A, UMAP visualisation of plaque LAMs re-clustered at a higher resolution. **B**, Jaccard similarity indices for n=50 sub-sampled datasets, showing 4 stable clusters within LAMs at the specified clustering resolution. Clusters with Jaccard indices higher than 0.8 (dashed line) in at least 75% of the sub-sampled datasets were considered as stable. **C**, Gene expression patterns of selected marker genes for different plaque LAM subpopulations (top) and different *ex vivo* subpopulations (bottom). Colour denotes average expression scaled across all cells and size denotes the percentage of the population expressing the gene. **D**, The expression of selected lipid-handling macrophage markers displayed on the UMAP as in **A**. Expression levels were min-max normalised for visualization purposes.

4.8. Discussion

For the vast majority of GWAS-identified loci with increased CAD risks, the identity of the causal gene(s) and associated cellular context(s) remains unknown, limiting the biological insight gained into the mechanisms of atherosclerosis. Single-cell sequencing approaches circumvent the need for prior identification and segregation of disease-relevant cell types for functional genomic studies [217].

To date, most single-cell-based studies prioritised genetic variants by their physical presence in either arbitrarily-defined gene-flanking regions or accessible chromatins, yet such regions are not necessarily regulatory [185,201]. In our macrophage data, only 6.9% of open chromatin peaks have cis-regulatory effects, so colocalization analysis using all open chromatin sites would likely include many false positives for causal variants. In a rare but certainly possible case, the specificity of such approaches might be no better than randomly selecting variants, making the downstream validation extremely difficult. In addition, the presence of many non-regulatory peaks in accessible chromatin regions drowns out true enrichment signals, leading to false negative results in prioritisation of causal cell types or cell states. This is particularly an issue in macrophages: we have previously profiled open chromatin and enhancer regions in bulk samples, and found no global enrichment for CAD risk loci [113]. Similar false negatives were also observed in single-cell ATAC-seq studies on human plaque and heart macrophages [185,201].

The analysis of CREs as presented in the previous chapter offers a powerful route to further understand the macrophage contribution to CAD heritability. In this chapter, I show that the identified macrophage CREs account for 11.5% of the total SNP heritability, and

contain 121 risk variants with genome-wide significance. I have also developed a computational workflow to partition CAD heritability among macrophage subpopulations. This is largely inspired by a recent work [197] done at the Broad Institute, with two important modifications for a fine-grained focus on macrophages: first, the multiome-derived CRE map, instead of the ABC-enhancer map, is adapted to annotate genomic regions with transcriptomic information, with great improvement in specificity demonstrated in our benchmark analysis. Secondly, a multivariate regression model is employed to retrieve the conditional significance for heritability enrichment in each gene program. Typically, enrichment tests using s-LDSC were conditioned on a "baseline model" containing function annotations of open chromatins, histone marks, CpGs, sequence-conservation and other LD-related annotations of population genetics forces [147]. This drastically reduces the power for heritability analysis and often results in close-to-zero τ* estimates. Here, I restrict the "base annotations" to the subpopulation gene programmes and a pseudo-bulk ox-LDL gene program to evaluate the informativeness of including each subpopulation-specific annotation on top of the average macrophage profile.

One common pitfall for multiple regression is multicollinearity, where two or more predictor variables in a multiple regression model are highly correlated with each other. As the annotation matrices were derived by scoring the differential expression statistics without discriminating up- and down-regulation, they will inevitably correlate with each other when overall heterogeneity level is low. When multicollinearity is severe, even small changes in the data might cause large changes in coefficient estimates, leading to the reduced power due to wider confidence intervals. An alternative approach may be to

define gene programmes based on the relative expression to a narrower set of cells (instead of all other cells in the current analysis).

Results from the conditional s-LDSC analysis demonstrate that CAD heritability is not uniformly distributed across macrophage subpopulations. In particular, I show that CAD heritability is enriched in the marker gene program and the ox-LDL-response gene program specific to a subset of macrophages, namely the lipid-handling macrophages. This macrophage subpopulation is distinguished by its activation of lipid metabolism genes and transcriptional networks driven by CEBP and NR1H3 family factors. I further track this subpopulation *in vitro* through immunofluorescence staining of a membrane glycoprotein CD52, and demonstrate that these $CD52^{hi}$ cells accumulate much less ox-LDL than other macrophages. Interestingly, the frequency of $CD52^{hi}$ lipid-handling macrophages varies significantly, ranging from 1% to 10% across different biological donors. While in the current study only healthy donors were recruited, this can be clinically relevant considering the significant capacities of lipid clearance these cells possess.

The reproducible presence of a $CD52^{hi}$ lipid-handling macrophage subset is further supported by the comprehensive meta-analysis of 38 human atherosclerotic samples from 5 independent studies. I have constructed by far the largest single-cell atlas for plaque macrophages, with a total of 20,458 cells representing 9 previously reported subpopulations. Counterintuitively, almost all macrophage subpopulations profiled in our study display tight transcriptomic similarity to one specific subpopulation in plaque, the lipid-associated macrophage (LAM), a population first discovered in liver and adipose tissue [28,216]. A recent spatial transcriptomic study in human liver samples reported that LAMs are concentrated around portal veins, unlike the general monocyte/macrophage

compartment which is distributed across different zones [218]. It is therefore reasonable to believe that plaque LAMs originate from circulating monocytes *in vivo*, and future lineage tracing work in atherosclerotic plaques is required to validate this hypothesis.

The greatly increased resolution of our plaque macrophage atlas uncovers the transcriptomic heterogeneity within plaque LAMs, and identifies, for the first time, a $CD52^{hi}$ subset expressing high levels of lipid processing genes. Although it is impractical to quantify the lipid-processing capacity directly in human plaque macrophages, this property can be inferred from our *in vitro* work given the remarkable similarity in transcriptomic profiles between the lipid-handling macrophages and the $CD52^{hi}$ LAM subset. These findings may provide mechanistic insights into the rare adverse cardiovascular events including myocardial infarction that have been reported with anti-CD52 therapy (alemtuzumab) [219]. In addition to the lymphocyte depletion triggered by alemtuzumab, acute damage to the $CD52^{hi}$ LAMs within atherosclerotic lesions may drive the progression of atherosclerosis, giving the potentially significant role for these cells in *in vivo* lipid clearance.

Ch 5. Molecular explanations of the enriched heritability in lipid-handling macrophages

Collaboration Notes: The work described this chapter was part of a research project in collaboration with Dr. Thomas K. Hiron who was a postdoctoral researcher in the O'Callaghan lab at the time. Dr. Hiron was responsible for most of the phlebotomy and performed the Western blot analysis of MITF. All other computational and experimental validations were performed by myself. This work has been published in:

Jiang J, Hiron TK, Agbaedeng T, Malhotra Y, Drydale E, Bancroft J, Ng E, Reschen ME, Davison LJ, O'Callaghan CA. A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease. *Circ Res.* 2024 May 15. doi: 10.1161/CIRCRESAHA.123.324172. PMID: 38747151.

5.1. Introduction

A common molecular mechanism through which genetic variants influence disease risks is the alteration of the transcriptional regulation of disease genes[203]. Since both the disease gene and its regulatory machinery are often heterogeneously expressed in macrophages, different subpopulations could be differentially affected by the same set of risk variants. In other words, the impact of a particular CAD-associated variant may be subpopulation-specific and might be missed in the analysis of bulk samples. So far, I have systemically demonstrated the enrichment for CAD heritability in a subset of human monocyte-derived macrophages, namely the $CD52^{+}$ lipid-handling macrophage (LHM) subpopulation. The potential clinical importance of this finding was further strengthened by the reproducible presence of these cells in human atherosclerosis. At a molecular level, the increased genetic risks means that at least one risk variant exerts larger regulatory effect through the LHM, possibly by utilising a transcriptional regulatory network specifically activated in these cells.

A closer inspection of the prioritised list of CAD-associated variants reveal 19 variants at 6 independent risk loci whose predicted target genes are significantly up-regulated in lipid-handling macrophages. These variants were present at CREs targeting *MITF*, *LPL*, *FDX1*/*RDX*, *BCAR1*, *TEX2*, and *PLPP3*. The regulatory mechanism of one such variant in the *PLPP3* gene has been demonstrated in previous work from our group [113], while others remain to be studied. In this chapter, I focus on the CAD risk variants affecting the $CD52^{+}$ LHMs at the *MITF*, *LPL* and *FDX1* loci, and investigated the molecular regulatory mechanism of one such variant rs10488763 in details.

Microphthalmia-associated transcription factor, as the name suggests, has been long associated with the development of melanin-producing cells ever since its first discovery in laboratory animals [220]. In human, this helix–loop–helix leucine zipper transcription factor is encoded by the gene *MITF*, and binds DNA in heterodimers with other members in the MiT family [221]. Unlike its widely-recognised role in melanin-related diseases like melanoma [222,223], MITF has only recently been linked to macrophage biology [224] and lysosome function [225,226]. Although other factors of the MiT family have demonstrated an athero-protective effect in preclinical models [227], the function of MITF is still largely unknown.

LPL encodes lipoprotein lipase, a highly conserved protein that has been studied for decades [228,229]. LPL protein hydrolyses triglycerides in circulating lipoproteins, and can serve as a bridging factor for receptor-mediated lipoprotein uptake [228]. Deficiency in *LPL* results in hypertriglyceridemia, and poses significant risks of atherosclerosis [230]. Mendelian randomisation analysis suggested a protective role of triglyceride-lowering *LPL* variants, most of which are linked to increased *LPL* expression in blood [231]. Due to

its soluble and secretory nature, it is difficult to track the cellular origin of LPL protein. Particularly, it is unclear what role macrophages are playing in LPL homeostasis within vascular walls. Contradictory observations have been reported in experimental atherosclerosis: LPL was identified as a key regulator for the retention of low-density lipoprotein in atherosclerotic lesions, yet its enzymatic function did not affect the extent of atherosclerosis [232]; myeloid-specific knock-out of *LPL* attenuates atherosclerosis [233], but has no effect on the lipid accumulation in macrophages nor the resolution of atherosclerotic plaques [234]. Clearly, there is still much to be learned about *LPL* beyond its function in regulating triglycerides in circulation.

Ferredoxin 1 has come under the spotlight as one of the key players underlying the recently discovered copper-dependent death called cuproptosis [235]. Encoded by the *FDX1* gene in the nuclear DNA, this iron-sulphur protein reductase serves as an electron donor for many reducing reactions in mitochondria [236]. A genome-wide CRISPR screen demonstrated that *FDX1* is essential for oxidative phosphorylation [237], and it regulates key respiratory complexes through lipoylation [238]. While mitochondrial dysfunction is widely believed to trigger the initiation of atherosclerosis [239], the role of ferredoxin has not characterized under the context of coronary artery disease.

5.2. The polygenetic signals at the *MITF* locus with rs6772383 and rs11452399

In a trans-ancestry GWAS analysis, an intronic variant rs12714757 at the *MITF* locus has been associated with CAD risk in both European and East Asian populations (β=0.033 and 0.048, respectively) [70,73]. This lead SNP was in high linkage disequilibrium with many

other variants, including rs6772383 and rs11452399 which resided in an intronic CRE of *MITF* (Figure 5.1). At rs6772383, the risk T allele disrupts a binding site of TEAD1, and is associated with decreased *MITF* expression in GTEx whole blood eQTL data (Figure5.2A), indicating that *MITF* expression is protective against atherosclerosis.

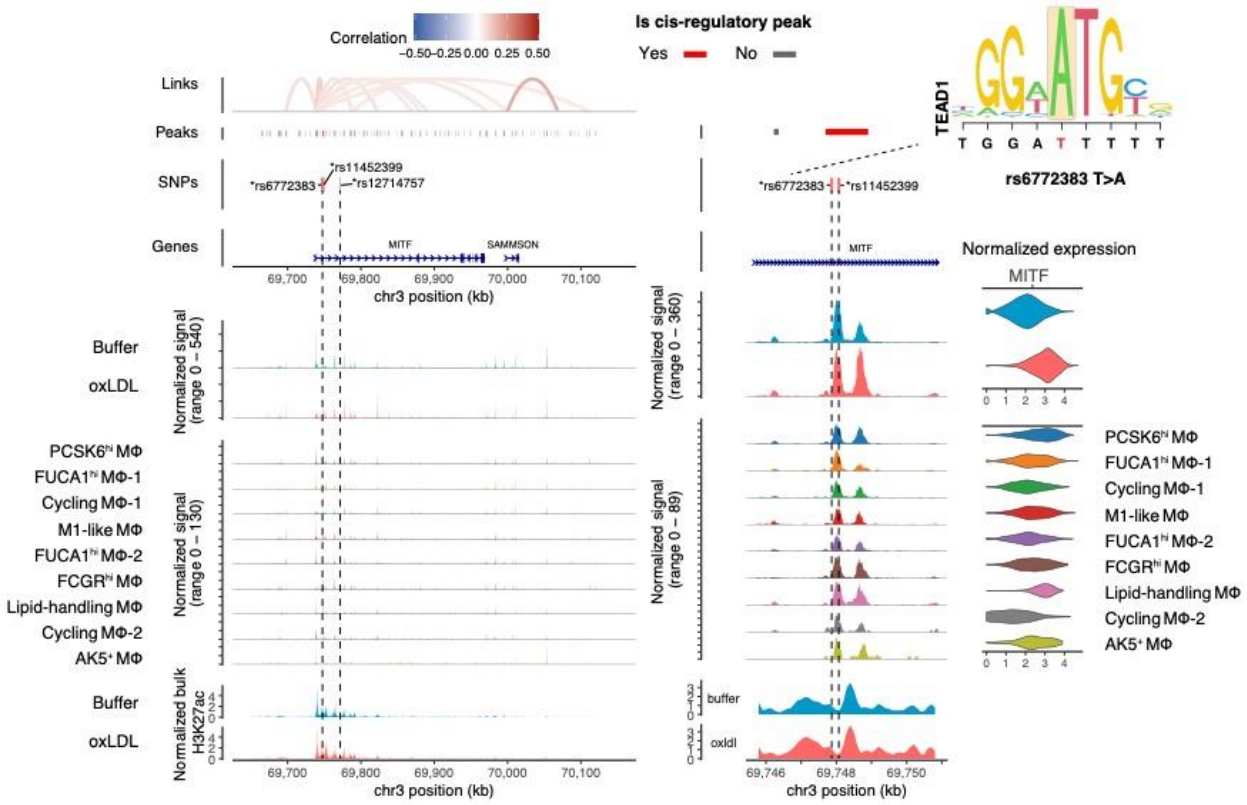


Figure 5.1 The regulatory landscape at the prioritised *MITF* locus.
The tagged variant rs12714757 is marked in grey and two candidate causal variants rs6772383 and rs11452399 are marked in red. Tracks (from top to bottom) showing: cis-regulatory interactions (links), ATAC-seq peaks (peaks), CAD-associated variants (SNPs), genomic annotations (Genes), ATAC-seq signals aggregated by treatment, ATAC-seq signals aggregated by subpopulations, enhancer signals from H3K27ac ChIP-seq. Higher resolution coverage plots centred on the cis-regulatory peak are plotted in the middle panel. Violin plots of aggregated gene expression are shown to the right. A red bar indicates significant CREs and eQTL SNPs are marked with an asterisk.

In primary macrophages, exposure to ox-LDL resulted in only a moderate increase in enhancer histone signals flanking this CRE (Figure 5.1). However, there is a significant increase in the *MITF* expression ($p<2.2\times10^{-16}$; Figure 5.1), accompanied by the increased chromatin accessibility of this intronic CRE ($p=3.3\times10^{-8}$). This ox-LDL-induced up-regulation of *MITF* was confirmed at the protein level using immunofluorescence staining (Figure 5.2B), and a similar trend was observed using Western blot analysis (Figure 5.2C and D). In plaque macrophages, *MITF* was predominantly expressed in the LAM subpopulation (Figure 5.3E), and its expression was up-regulated in LAMs collected from disease tissue as compared to adjacent control tissue (Figure 5.3F). Collectively, these results suggest a protective role of *MITF* in macrophages response to lipids and potentially other atherogenic stimuli.

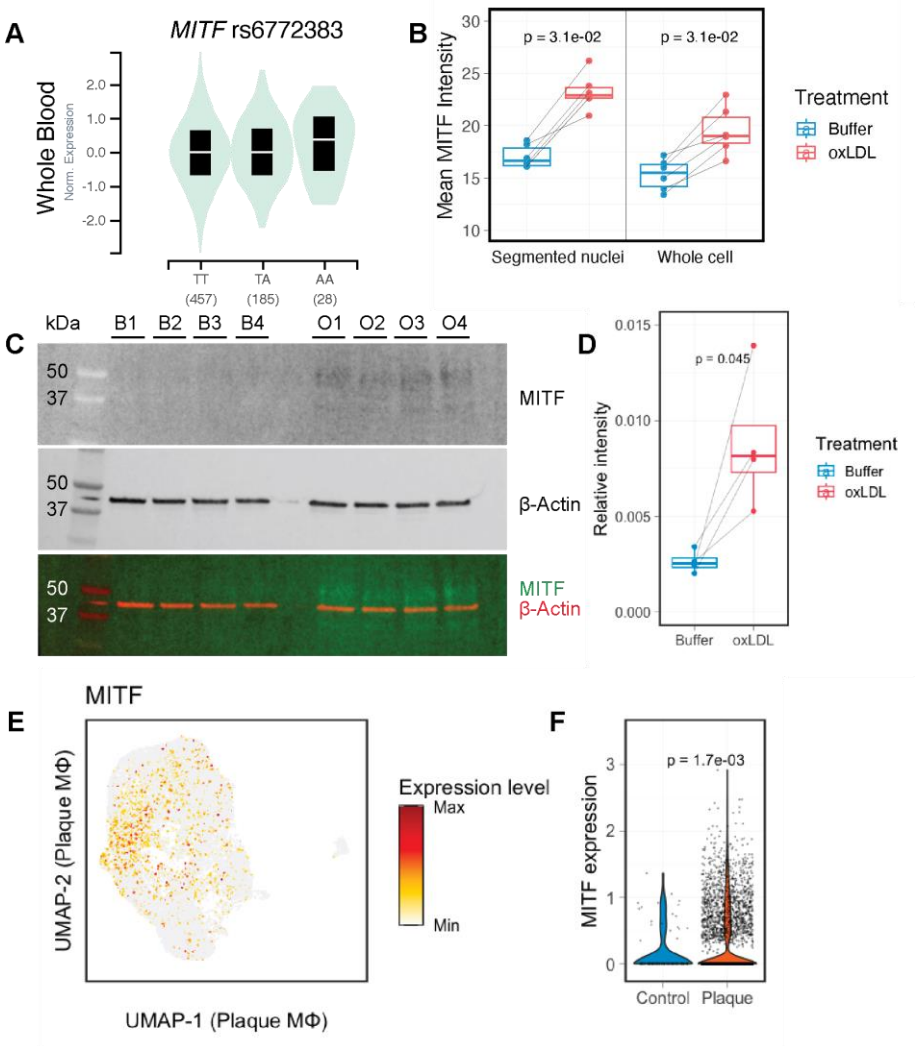


Figure 5.2 Functional validation of MITF in ox-LDL-exposed macrophages.
A, Violin plot showing the allele-specific effect of rs6772383 on MITF expression in human blood based on GTEx data. **B**, Quantification of MITF expression using immunofluorescence staining in macrophages with (red) or without (blue) ox-LDL treatment, shown as average nuclear and intracellular fluorescence intensities (arbitrary units) per biological replicate (n=6). **C**, Detection of MITF by Western blotting. Whole-cell extracts from unexposed or ox-LDL-exposed primary human macrophages (n=4 biological replicates) were analysed by Western blotting using anti-MITF antibody. Molecular weight standards are indicated to the left. β-Actin was probed on the same blot and used as the internal reference for quantification normalization. The fluorescent signals (MITF, 800nm, green; β-Actin, 700nm, red) on the developed blot were simultaneously detected and the overlay is shown at the bottom. **D**, Quantitative analysis of the MITF band at ~52kDa in **C**. Western Blot analysis was performed by Dr. Thomas Hiron. P-value calculated using a paired Wilcoxon test. **E**, *MITF* expression shown in the UMAP of plaque macrophage atlas, showing increased the LAM subset. **F**, *MITF*

expression in LAMs grouped by the tissue of origin, showing higher expression in plaques LAMs vs LAMs from control unaffected arteries. P-value calculated using a Wilcoxon test.

5.3. The polygenetic signals at three *LPL* loci

The *LPL* gene encodes the secretive lipoprotein lipase which plays a role in lipoprotein uptake [240]. A handful genetic variants located in the *LPL* gene were linked to lower triglyceride level and lower risk of CAD [231]. In the present study, I have identified *LPL* as one of the strongest marker genes for lipid-handling macrophage (Figure 4.11A and Figure 5.3), a macrophage subpopulation carrying increased level of CAD heritability. Although its expression was below the detection threshold in many cells, *LPL* was significantly upregulated by the exposure to ox-LDL ($p=3\times10^{-4}$).

In the variant prioritisation analysis, I identified 6 CAD risk variants in 3 different CREs at the *LPL* locus (Figure 5.3): rs1800590 in a 5' UTR CRE (β=0.091 in European population); rs6999612 and rs59811201 in an intronic CRE (β=0.126); rs11204085, rs11204086 and rs11204087 clustered in a distal intergenic CRE (β=0.028). The latter locus was located 182kb downstream of the TSS of *LPL* gene, with *LPL* being its third nearest gene. This locus would have been missed by traditional "nearest-gene" approaches that prioritise variants based on proximity only.

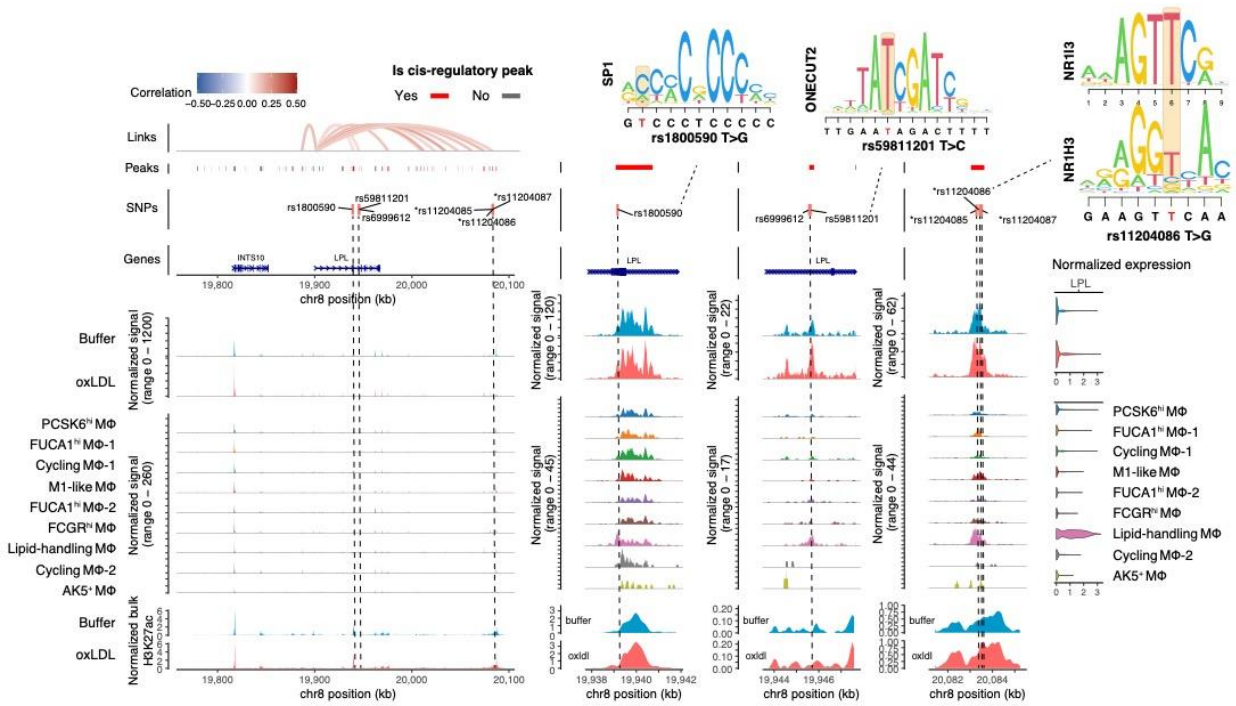


Figure 5.3 The regulatory landscape at the *LPL* gene with three prioritised loci. Similar to that in Figure 5.1, zoom-in views centred on the associated CREs are plotted in the middle panel.

No significant change in chromatin accessibility was found in any of the three CREs with ox-LDL exposure, likely because they were already highly accessible. Nevertheless, enhancer histone signals increased 70.9% in the distal CRE containing rs11204086, and 43.0% in the 5'UTR CRE containing rs1800590. This suggests that the effect of ox-LDL on *LPL* expression might be mediated through changes in enhancer activities rather than chromatin remodelling.

The protective G allele of rs11204086 disrupts a binding motif for NR1H3 (Figure 5.3) and is associated with higher blood expression of *LPL* in eQTL data. As illustrated in the Chapter 4, NR1H3 itself is activated notably in lipid-handling macrophages (Figure 4.14C). Therefore, it is reasonable to conclude that the cis-regulatory effect of rs11204086 may

be specific to lipid-handling macrophages, mediated by a transcriptional regulatory network that is only activated in these cells.

5.4. The polygenetic signals at the *FDX1* locus with rs10488763 and rs1443120

From the CRE-based variant prioritization analysis, two CAD-associated risk SNPs (rs10488763 and rs1443120) in tight linkage disequilibrium (r^2=1 in EUR panel) overlapped a distal CRE for *FDX1* and *RDX*. This CRE resided in a long non-coding RNA and is 56 kb and 77 kb upstream of the transcriptional start sites of *FDX1* and *RDX* respectively (Figure 5.4).

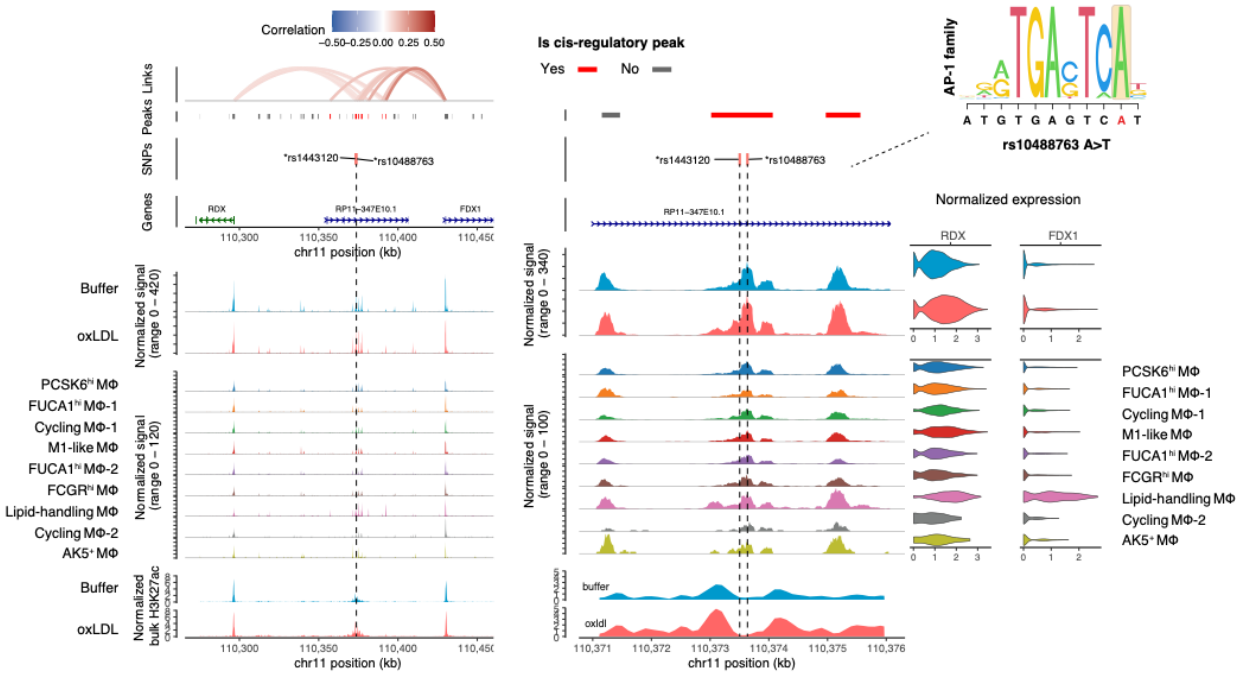


Figure 5.4 The regulatory landscape at the prioritised *FDX1*/*RDX* locus.
The CRE overlapping rs10488763 and rs1443120 was linked to both the *FDX1* and the *RDX* gene. Similar to that in Figure 5.1, genomic tracks with zoom-in views are plotted with corresponding plots for gene expression to the right.

The alternative alleles of these two SNPs were associated with higher CAD risk (β=0.036) [70,73] and lower expression of both *FDX1* and *RDX* in whole blood (Figure 5.5A). However, CAD summary statistics and eQTL mapping suggest that the GWAS signal aligns most clearly with the eQTL signal of *FDX1* rather than of *RDX* (Figure 5.5B and C). In addition, the expression of *FDX1* was elevated in plaque LAMs ($p=7.7\times10^{-5}$; Figure 5.5D) compared to non-plaque tissue LAMs, while *RDX* expression remained unchanged (Figure 5.5E). To rule out the any spurious differential expression due to sample imbalance, I also repeated the DE test in down-sampled datasets (Figure 5.5D and E). Reassuringly, *FDX1* but not *RDX* was significantly up-regulated in plaque compared to control tissues. These results indicate that *FDX1* is the likely causal gene influenced by one or both CAD variants rs10488763 and rs1443120 at this locus.

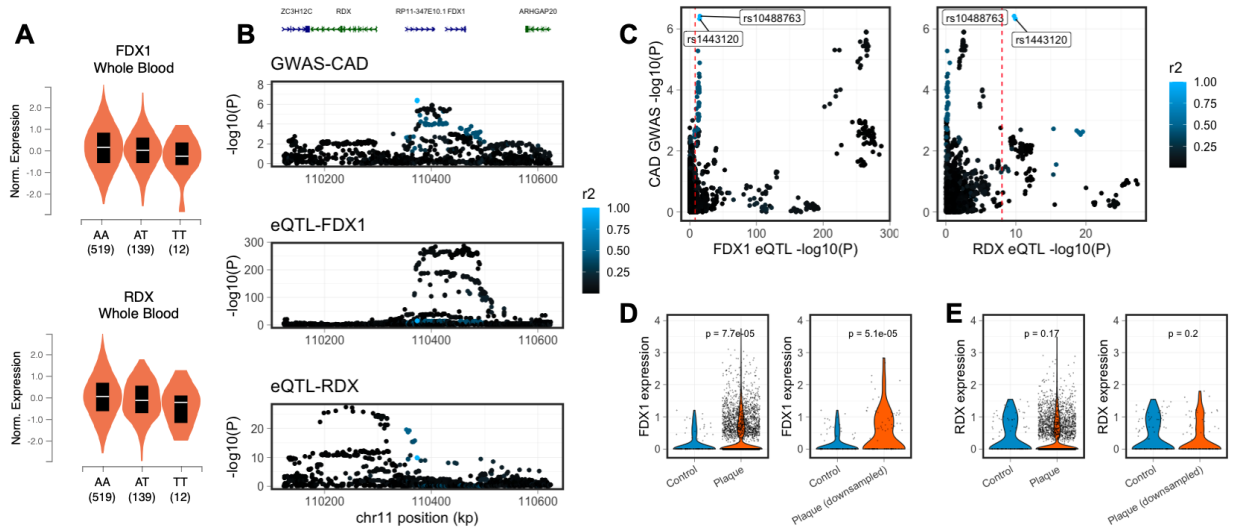


Figure 5.5 Polygenetic signals at the FDX1/RDX locus.
A, Violin plot showing the allele-specific effect of rs10488763 on *FDX1* (top) and RDX (bottom) expression in human blood based on GTEx data. **B**, Association of rs10488763 variant with CAD (GWAS-CAD), *FDX1* expression (eQTL-FDX1), *RDX* expression (eQTL-RDX, note smaller scale of y axis), and chromatin accessibility (caQTL) at the *FDX1/RDX* locus. SNPs are coloured by their LD with rs10488763. **C**, LocusCompare plots overlaying polygenetic signals of SNPs within 1Mb window of rs10488763. x axis:

the blood eQTL signals for FDX1 (left), RDX (right). Dashed line showing the genome-wide significance cut-off for each QTLs. y axis: GWAS signals for CAD. SNPs are coloured by their LD level (r^2) to rs10488763. Note that due to the availability of the full GWAS summary statistics, only the CAD signals from the UK Biobank cohort are plotted in **B** and **C**, which are underpowered to produce genome-wide significant signals ($p<5x10^{-8}$) at this locus. **D** and **E**, The expression of *FDX1* (**D**) and *RDX* (**E**) in LAMs collected from control tissue and plaques. Full and down-sampled datasets are plotted in the left and right panel, respectively.

5.5. The allelic effect of rs10488763 on local chromatin accessibility and enhancer activity

The risk T allele of rs10488763 perturbs the binding motif for AP-1 factors (Figure 5.4). As AP-1 family members have been reported to interact with chromatin remodelling complex [241], I hypothesized that this SNP exerts a cis-regulatory effect by modulating local chromatin structure.

In a multi-modal QTL study of iPSC-derived macrophages [144], the rs10488763 SNP has been linked to changes in local chromatin accessibility (chromatin accessibility QTL, caQTL; as shown in Figure 5.6A), although no significant eQTL signal was detected possibly due the limited sample size. Consistent with this caQTL signal, ox-LDL-treated macrophages carrying the risk T allele showed lower chromatin accessibility compared to those that are homozygous for the protective A allele ($p=1.8\times10^{-4}$; Figure 5.6B). In heterozygous samples, I noticed a striking allelic imbalance where the ATAC-seq coverage for the A allele is nearly 10-fold of that of the T allele, indicating lower accessibility in the chromosome carrying the risk T allele (Figure 5.6C). This allele-specific difference in chromatin accessibility was also observed in snATAC-seq samples of plaque cells from donors heterozygous for rs10488763 [242] ($p=0.041$; Figure 5.6C).

On top of the chromatin modulating effect of variant itself, chromatin accessibility at the locus of this variant is also increased by ox-LDL (p=0.015; Figure 5.4). In addition, ox-LDL exposure induced a 45.3% increase in the local H3K27ac enhancer signal (Figure 5.4). To test if the two rs10488763 alleles differentially affect this ox-LDL-induced enhancer activation, I performed a dual-luciferase reporter assay in primary macrophages. In brief, a 150-bp region containing each allele of rs10488763 (but not rs1443120; see Figure 5.7A) was cloned into a luciferase reporter assay and transfected into primary macrophages. Cells were then exposed to ox-LDL or control buffer before lysed for luminescence assay. Consistent with the observed H3K27ac signals, ox-LDL treatment triggered the enhancer activity with the protective A allele (average 1.58-fold, p=0.031; Figure 5.6D). This ox-LDL-induced activation was disrupted with the risk T allele (Figure 5.6D).

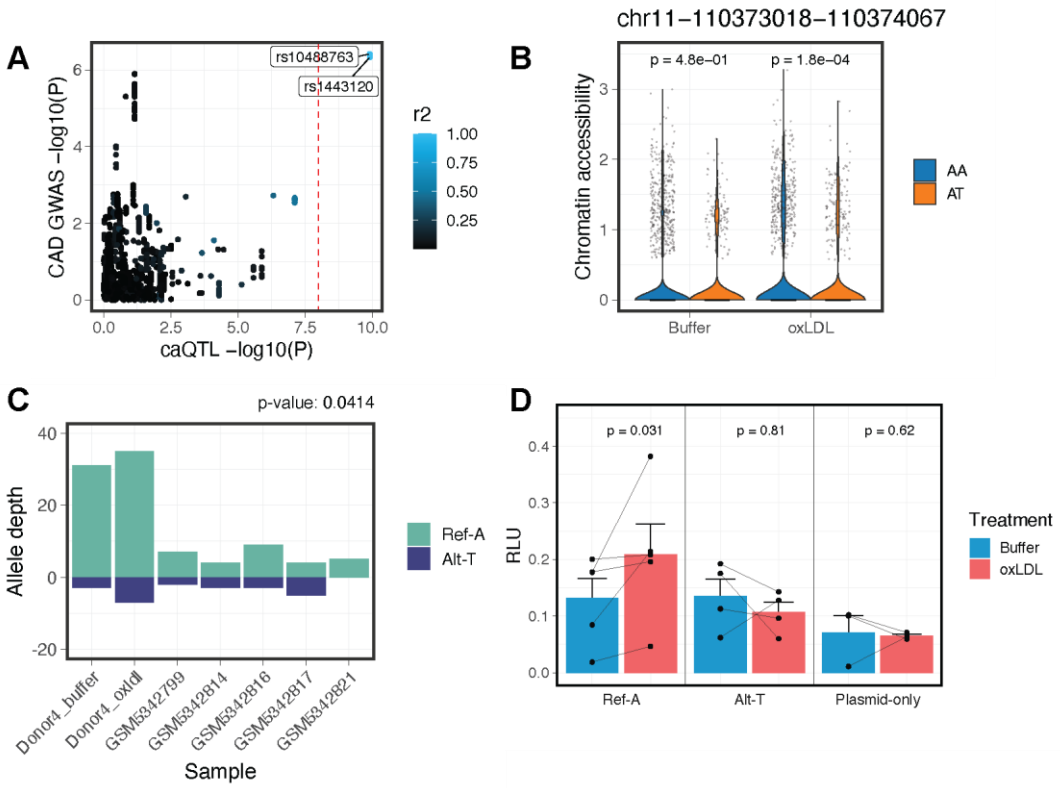


Figure 5.6 The alternative risk T allele of the prioritised SNP rs10488763 alters the regulatory property of its local chromatin.
A, Similar to that in Figure 5.5C, LocusCompare plots overlaying the GWAS signals (y axis) and chromatin-accessibility-QTL (caQTL, x axis) of SNPs within 1Mb window of rs10488763. Dashed line showing the genome-wide significance cut-off for each QTLs. SNPs are coloured by their LD level (r^2) to rs10488763. **B**, Chromatin accessibility of the CRE overlapping rs10488763, grouped by genotype from the single-cell multiome data. (n=3 for A/A, n=1 for A/T). **C**, Allelic sequencing coverage of rs10488763 in 7 heterozygous samples, showing higher accessibility in the chromosome containing the protective A allele compared to the T allele. Reads are aggregated from snATAC-seq datasets of primary human macrophages and human plaques [242]. Sequencing coverage on the alternative allele shown in negative values for visualization purposes. **D**, Dual luciferase reporter assays in primary human macrophages with and without ox-LDL treatment testing the allelic enhancer activity of rs10488763. (n=5 biological replicates). Signals normalised to Renilla luciferase activities.

5.6. The role of CEBPB in the cis-regulatory effect of rs10488763

As shown in Figure 5.4, the CRE overlapping rs10488763 is constituted by three local-accessibility summits, flanked by histones with H3K27ac signals. These summits were

separated by around 200 bp of DNA, close to the average nucleosome spacing of 190 bp [243]. This suggests that the three accessibility summits are likely connected by tightly positioned nucleosomes and thus are physically proximal to each other. Specifically, I reasoned that the additional summits might relate to the binding of other transcription factors, which could also contribute to the regulatory effect of rs10488763. Through motif scanning and inspection of ChIP-seq data generated previously in the O'Callaghan group in bulk macrophages[113], I identified a CEBPB binding site 300bp downstream of the rs10488763 SNP within the same CRE (Figure 5.7A). A longer DNA construct containing both the CEBPB binding site and the rs10488763 (but not rs1443120) was then cloned into the luciferase plasmid and assayed for enhancer activity. After overexpression of the activator isoform of CEBPB (liver-enriched activator protein, LAP), the enhancer activity increased over 7-fold (Figure 5.7B). Similar to the shorter 150bp constructs, no significant difference in enhancer activity was observed between the risk T and protective A alleles (Figure 5.6D and 5.7B). This is probably because the plasmid DNA, unlike histone-bound genomic DNA, is already accessible to TFs to initiate transcription of downstream luciferase reporter.

To test whether the cis-regulatory effect of rs10488763 is mediated by the recruitment of CEBPB, I overexpressed LAP in primary macrophages and quantified the expression of *FDX1* using RT-qPCR. Macrophages were obtained from genotyped donors so that the effect of rs10488763 could be investigated in its native form. Consistent with the eQTL signals in whole blood samples (Figure 5.5A), *FDX1* expression was significantly lower in macrophages carrying the risk T allele ($p=0.0039$, Figure 5.7C). Overexpression of CEBPB compensated for this reduction of *FDX1* expression in macrophages carrying the

T allele (p=0.016, Figure 5.7C), while having no additional effect in macrophages homozygous for the A allele (p=0.63, Figure 5.7C).

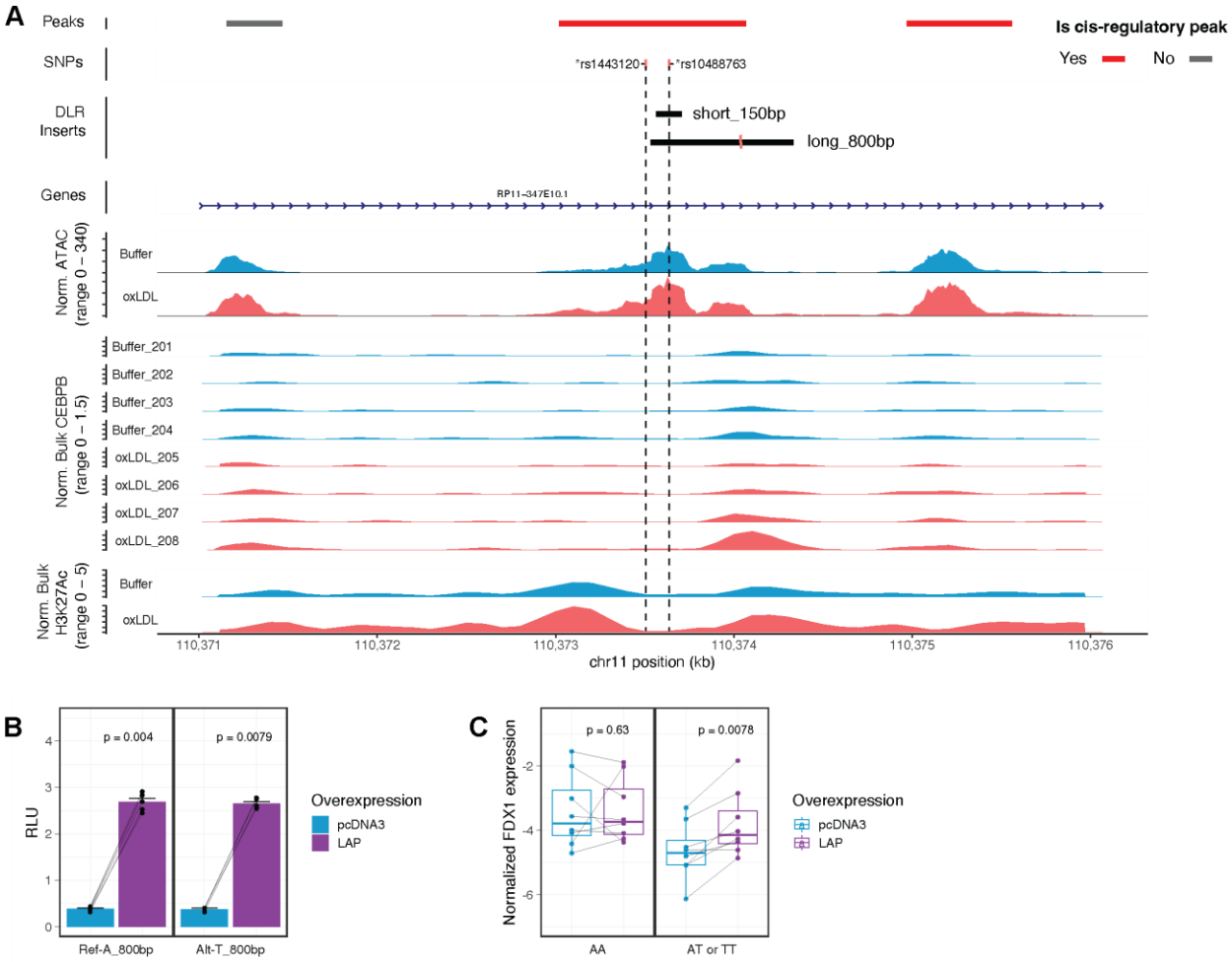


Figure 5.7 The cis-regulatory effect of rs10488763 on FDX1 expression is mediated by CEBPB.

A, Zoom-in view of CAD-associated GWAS variant rs10488763 at the FDX1/RDX locus. CRE peaks are marked in red (top panel). Compared to the 150bp dsDNA construct used in Figure 5.7D, a longer 800bp construct containing the motif-predicted CEBPB binding site (marked in red) was designed and tested for enhancer activities using dual luciferase assay. Genomic tracks show pseudo-bulk chromatin accessibility profiles aggregated by treatment (**top**), bulk CEBPB ChIP-seq signals aggregated by sample (**middle**), and bulk H3K27ac ChIP-seq signals (**bottom**). All sequencing signals are normalised by total read counts. **B**, Luciferase reporter assays of the long dsDNA construct as shown in **A** in THP-1 cells with or without liver activator protein (LAP; the activator isoform of CEBPB protein) overexpression. (n=5 independent transfections). **C**, RT-qPCR measurement of *CEBPB* expression in primary macrophages (n=8, 6, 2 for AA, AT, and TT, respectively) with or

without LAP overexpression. GAPDH was used as the reference gene. P-values were calculated with paired Wilcoxon tests.

5.7. *FDX1* knock-down reduces ox-LDL uptake in human macrophages

Given that *FDX1* is highly expressed in lipid-handling macrophages (Figure 5.4) and also up-regulated by ox-LDL (p=0.00472), I hypothesised that it is involved in the lipid metabolism in human macrophages. To test this hypothesis, I knocked down the expression of *FDX1* using small interference RNA (siRNA). Although the knock-down efficiency is relatively low (~10% decrease in fluorescence intensity, as shown in the Figure 5.8A and B), the intracellular DiI-ox-LDL accumulation reduced for up to 40% (Figure 5.8A and C). This result is somewhat contradictory to the observation in the previous chapter (Figure 4.16) that lipid-handling macrophages, with elevated *FDX1* expression (Figure 5.4), accumulate significantly less DiI-ox-LDL. Given that lipid-handling macrophages also showed elevated *CD36* expression, one plausible explanation is that *FDX1*, similar to *CD36*, plays a role in lipid uptake in macrophages.

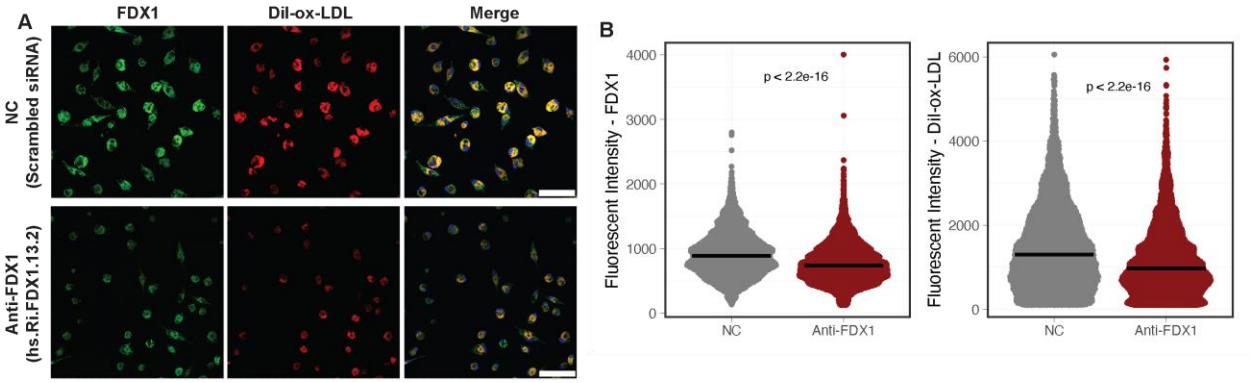


Figure 5.8 FDX1 knock-down reduces lipid accumulation in primary macrophages. **A**, Representative immunostaining of primary macrophages for FDX1 and DiI-ox-LDL. Cells were transfected with anti-FDX1 or scrambled siRNA for 48 hours and exposed to DiI-ox-LDL for another 24 hours. Scale bar: 100um. Successful knock-down was also

verified in macrophages without DiI-ox-LDL treatment. **B**, quantification of the FDX1 level (left) and DiI signals (right) in immunofluorescence images. Cells were pooled (n=3,099 for scrambled siRNA and 4,600 for siRNA_2) from n=4 biological donors, and p-values were calculated using Wilcoxon tests.

5.8. Discussion

For non-coding genetic risk variants to have functional effects, the presence of both the regulatory DNA and the associated molecular regulator is required. Through comprehensive integration of multi-omics sequencing data, polygenetic signals, and the TF binding motifs, I have highlighted the potential regulatory mechanisms at CAD risk loci targeting *MITF*, *LPL* and *FDX1*. These loci were selected because the predicted causal genes were up-regulated in lipid-handling macrophages, a novel subpopulation that I have linked to increased genetic risk in the previous chapter. In addition, all three genes are involved in the transcriptomic and epigenomic response to ox-LDL, showing changes in gene expression, chromatin accessibility and/or enhancer activity.

One of the key advantages of our GWAS-informed functional genomic analyses is that the identification of disease gene is largely hypothesis-free. Traditionally, investigations of disease mechanisms are driven by candidate-gene approaches, often a euphemism for guessing. Although the choice of candidate genes can be guided by pre-existing knowledge of the disease, the presence of such knowledge itself inevitably biases the "guess and test" process against poorly-studied genes, or even well-recognised genes outside of their supposed function scopes. The results presented in this chapter highlight the previously underappreciated role of *MITF*, and provide the first evidence of the *FDX1* regulation in the context of coronary artery disease.

In previous chapters I have shown the central role for AP-1 family factors in regulating the expression of ox-LDL-induced genes (Figure 3.11) and CAD-risk genes (Figure 4.5). A recent single-cell ATAC-seq study also reported enriched AP-1 bindings in macrophage-specific accessible sites [185]. Here, I have followed up the SNP rs10488763 experimentally and linked its regulatory effect on the expression of *FDX1* to the binding of AP-1 factors and CEBPB. Our computational and experimental results show that the CAD risk T allele lowers *FDX1* expression in macrophages through allele-specific disruption of enhancer activation. This effect is likely caused by a complex hierarchical mechanism where both the binding and subsequent replacement of AP-1 factor is required (Figure 5.9).

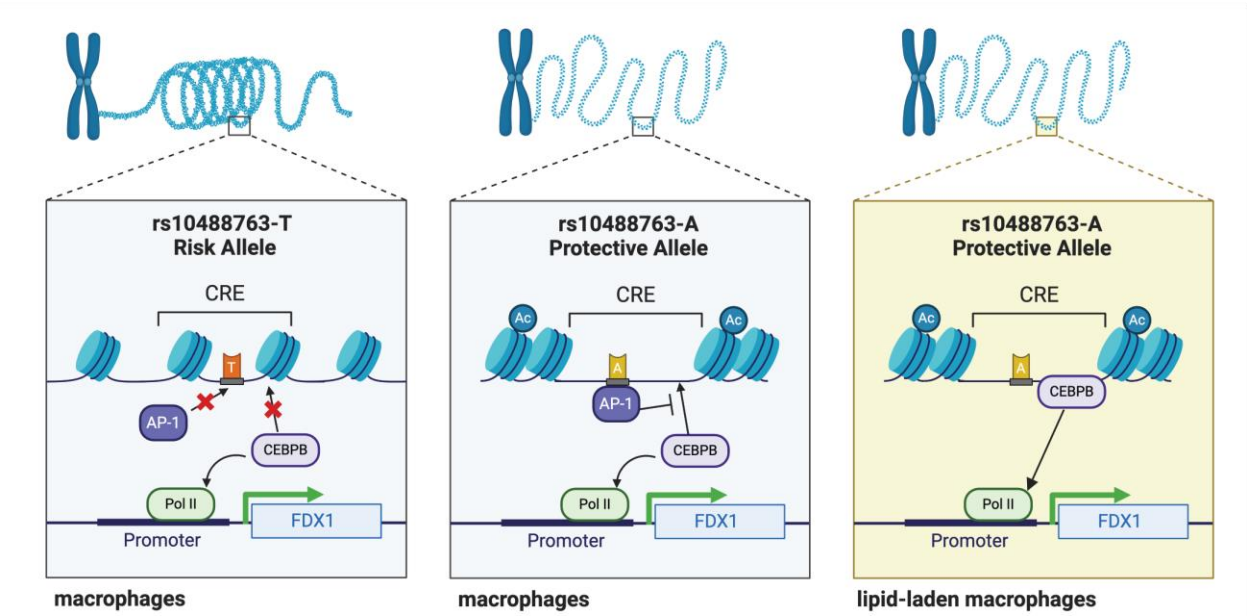


Figure 5.9 Model showing the role of AP-1 factors and CEBPB in modulating the regulatory effect of CAD risk SNP rs10488763.
Left: The risk T allele disrupts the binding motif for AP-1 family factors, making the local chromatin inaccessible and thereby inactivating the CRE. **Middle:** AP-1 binding remodels the local chromatin and exposes the CEBPB motif within the CRE. However, the physical presence of AP-1 factors hinders the binding of CEBPB. **Right:** Upon ox-LDL treatment, the expression of AP-1 factors decreases, while the expression of CEBPB increases. As CEBPB replaces AP-1 factors at this locus, the transcription is initiated and the expression of *FDX1* increases.

As shown in Figure 3.19, AP-1 factor FOS is a negative regulator of *FDX1* expression, yet disrupting its binding motif at the SNP rs10488763 decreases *FDX1* expression. It is possible that FOS is replaced by an activating subunit in the heterodimer AP-1 complex when the enhancer is activated by ox-LDL, although our computational prediction does not show the presence of activating subunit in this particular CRE. In addition, AP-1 factors have been reported to bind collaboratively with cell-type-specific TFs [244], leading to another plausible explanation whereby AP-1 factors select and establish accessible chromatin for other TFs without an expression modulating effect *per se*. This is supported by the observation that CEBPB overexpression bypasses the allelic effect of rs10488763 and restores *FDX1* expression in macrophages with the disrupted AP-1 binding site.

At the time of writing, one study has reported that Fdx1 deficiency disrupted lipid homeostasis in mouse embryo fibroblasts (homozygous Fdx1 knock-out leads to preweaning death in mice) [245], and a preprint linked *FDX1* to cholesterol-induced bile acid synthesis in THP-1 macrophages[246]. In our model, knock-down of FDX1 in macrophages decreases the overall accumulation of lipids upon ox-LDL treatment, which seems to contradict the protective role suggested by genetic studies. However, the observed lipid accumulation is a combinatorial effect involving both the uptake, metabolism and efflux of ox-LDL cholesterol, and more work needs to be done to delineate the role of *FDX1* in each of these processes.

Ch 6. Ox-LDL exposure in endothelial cells modulates gene regulatory programmes and genetic risk of CAD

6.1. Introduction

Cholesterol-carrying low-density lipoprotein (LDL) particles are reported to be retained and chemically modified (e.g., by oxidation) in the vessel wall at lesion-prone sites during the initiation of atherosclerosis [247,248]. Endothelial cells (ECs) are one of the most important of the arterial cells and constitute the first barrier of the arterial vasculature against circulating cholesterol. In health, ECs tightly control vascular tone, prevent blood clotting and maintain redox balance; in disease or inflammatory conditions, damage to ECs triggers the recruitment of immune cell to the sub-endothelial space and provides intercellular autocrine and paracrine signalling loops (e.g., M-CSF and GM-CSF for macrophage maturation) within the vessel wall [4,249,250].

After almost thirty years since the discovery of an ox-LDL receptor in bovine endothelial cells [251], it is somewhat surprising to find that the response to ox-LDL in ECs is still poorly characterised. Although the concept of ox-LDL-mediated endothelial cell dysfunction is widely accepted [252], the existence of such an effect is not reproducibly reported in the literature: using ICAM-1 and VCAM-1 as markers, some reports show that ox-LDL triggers pro-inflammatory response in human endothelial cells from coronary artery [253], aorta [254] and umbilical vein [255,256]; whereas other reports find no evidence of this response using the same models [257-261]. Since most of the early studies on ox-LDL and ECs were conducted with a narrow focus on specific pathways or functions, the experimental readout in these studies is typically low-throughput and inconclusive. A comprehensive,

unbiased assessment of the ox-LDL response in human ECs is crucial for a more complete understanding of the pathophysiological processes underlying atherosclerosis and related cardiovascular conditions.

The development of high-throughput multi-omics sequencing approaches enables genome-wide, hypothesis-free investigation of the endothelial response to atherogenic lipids. One of the bioactive components of ox-LDL, oxidised 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphocholine (ox-PAPC) has been extensively characterised [262-264]. In primary human aortic endothelial cells (HAECs), ox-PAPC has been shown to alter the transcriptome [265], epigenome [266], and metabolome [267]. While serving as a convenient synthetic alternative to the plasma-derived ox-LDL, ox-PAPC represents only a fraction of the lipid peroxidation products found in ox-LDL, and its pathological similarity with ox-LDL has not been systemically evaluated. Given that oxidised lipids do not act in isolation during atherosclerosis, measuring the overall effect of ox-LDL instead of its bioactive components might better reflect its multifaceted atherogenic properties *in vivo*.

To address this, I employed high-throughput sequencing methods to analyse the influence of ox-LDL exposure on gene expression and chromatin accessibility in primary endothelial cells. Integrating the multi-omics sequencing data highlights an ox-LDL-response regulatory network conserved between macrophages and ECs. Using a combination of computational and experimental approaches, I find that ox-LDL triggers endothelial cell migration through epigenomic rewiring of transcription factor binding. In addition, I show that the endothelial response to ox-LDL is associated with the genetic risk of coronary artery disease (CAD), and prioritise 8 CAD risk variants that overlap differential TF binding sites.

6.2. Ox-LDL accumulation triggers metabolic shift in HAECs

It is well-recognized that endothelial cells play an important role in the transcytosis of lipids from the circulation to the sub-arterial space [268]. To test whether exposure to ox-LDL results in an intracellular accumulation of lipids, primary human aortic endothelial cells (HAECs) were incubated with DiI-labelled ox-LDL and profiled using immunofluorescence microscopy. After 48 hours of DiI-ox-LDL treatment, build-up of lipid droplets was clearly visible in the cytoplasm (Figure 6.1A and B), although the magnitude of lipid accumulation is significantly lower compared to that in macrophages (Figure 4.16). The dose-dependent manner of intra-cellular lipid accumulation was further confirmed by flow cytometry (Figure 6.1C).

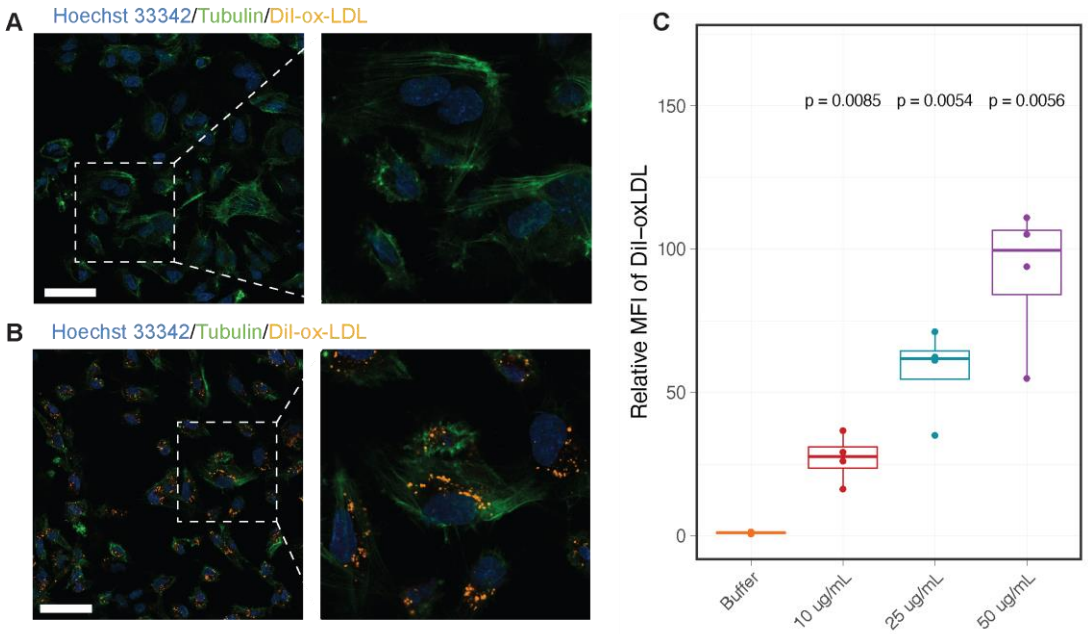


Figure 6.1 Ox-LDL exposure triggers intracellular accumulation of lipid droplets in primary aortic endothelial cells (HAECs).
A and **B**, Representative images from immunofluorescence staining of unexposed (**A**) and ox-LDL-exposed (**B**) HAECs. Cells were treated with 50ug/mL of DiI-labelled ox-LDL or control buffer for 48 hours before imaging. Scale bar: 50um. **C**, Quantification of intracellular DiI signals using flow cytometry in HAECs treated with DiI-ox-LDL for 24

hours at indicated concentrations (n=4 biological replicates). Signals were normalised to the buffer group and shown as the relative mean fluorescence intensity.

To investigate whether there is any metabolic consequence of ox-LDL exposure, I next measured the cellular respiratory rates and glycolytic rates using Seahorse assays (Figure 6.2A to D). While the basal respiration rate remains constant, ox-LDL treatment drastically reduced the spare respiration capacity the cells can utilized when the energy demand is high (Figure 6.2 A and B). In addition, cells exposed to ox-LDL displayed a significant increase in basal glycolytic rate (Figure 6.2 C and D); the addition of rotenone and antimycin A, which inhibits electron transport complexes and thereby shuts down mitochondrial respiration, did not further increase the glycolytic rate (Figure 6.2C). These results suggest that ox-LDL induces metabolic stress and pushes the energy balance toward glycolysis.

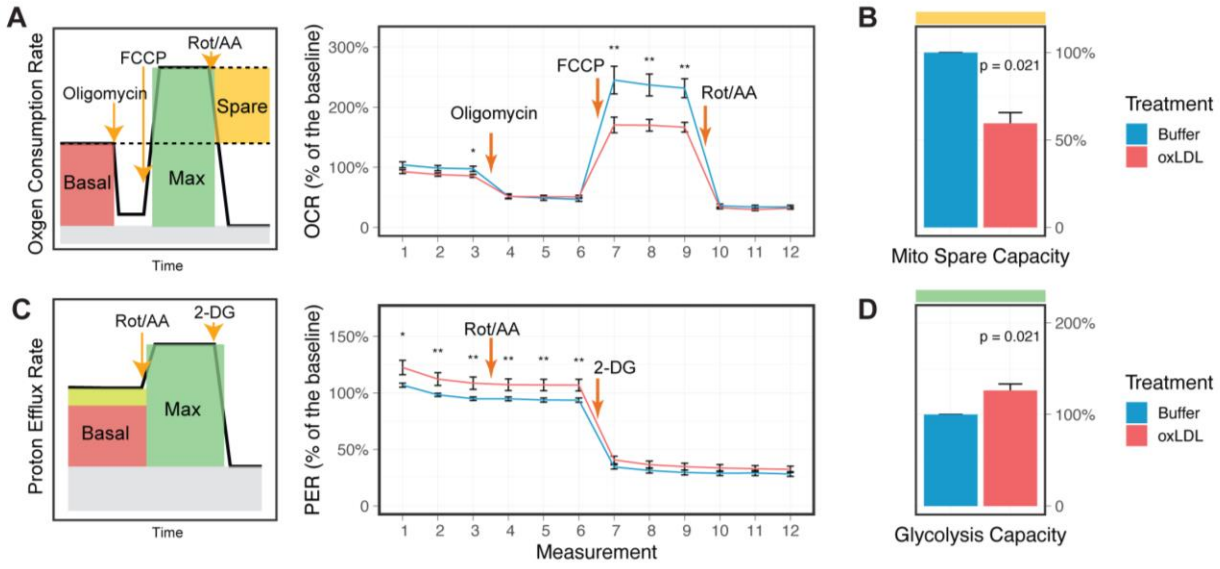


Figure 6.2 Ox-LDL induces metabolic reprogramming in HAECs.
A, Oxygen consumption rate (OCR) of unexposed (blue) and ox-LDL-exposed (red) HAECs, shown as the average value along with the standard error across n=4 biological

replicates. Overview of the mitochondrial stress test assay is plotted to the left. **B**, The spare respiration capacity as marked in yellow in **A**, normalised to the buffer group. **C**, Proton efflux rate (OCR) of unexposed (blue) and ox-LDL-exposed (red) HAECs, shown as the average value along with the standard error across n=4 biological replicates. Overview of the glycolytic rate assay is plotted to the left. **D**, The glycolysis capacity as marked in green in **C**, normalised to the buffer group. Seahorse assays were conducted with a minimum of three technological replicates. P-values were calculated using Wilcoxon tests. * p-value< 0.05, ** p-value <0.01.

Previous studies have shown that ox-LDL induced endothelium dysfunction, an important atherogenic process that is typically characterised by elevated expression of cellular adhesion molecules [253]. However, despite its intracellular accumulation and the metabolic consequences, ox-LDL-exposure alone was not found to affect the expression of ICAM-1 and VCAM-1 in HAECs (Figure 6.3). This unchanged expression of adhesion molecules was further validated in the following RNA-seq analysis from two independent experiments.

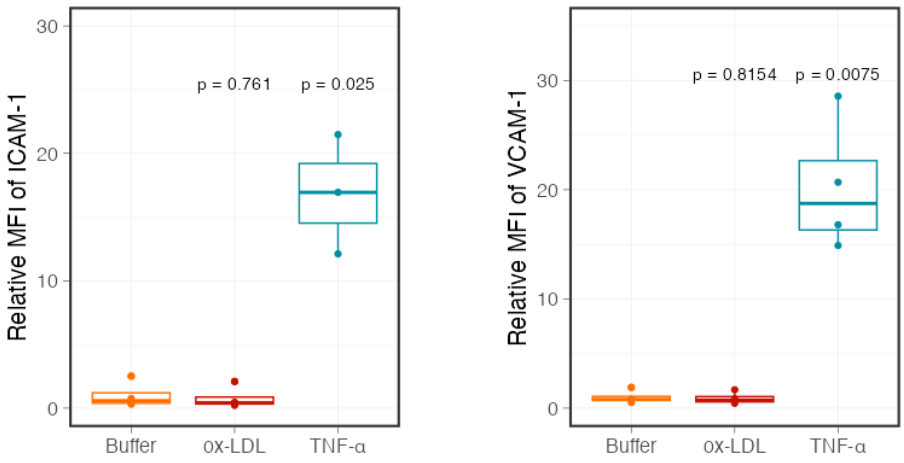


Figure 6.3 Ox-LDL alone does not trigger pro-inflammatory phenotypes in HAECs. HAECs were treated with 50ug/mL ox-LDL or 10ng/mL TNF- α for 24 hours (n=4 biological replicates). The membrane level of ICAM1 (left) and VCAM1 (right) was quantified using flow cytometry. Signals were normalised to the buffer group and shown as the relative mean fluorescence intensity. P-values were calculated using t-tests.

6.3. Transcriptomic analysis reveals disrupted fatty-acid metabolism

Next, I performed high-throughput bulk RNA-seq to systematically characterise the effect of ox-LDL on HAECs. Differential expression analysis was performed using edgeR [269], with a generalised linear model controlling for donor and batch effect. In total, 304 genes were identified as differentially expressed under the FDR threshold of 0.05, with similar numbers in each direction (144 genes being up-regulated by ox-LDL and 160 genes down-regulated; Figure 6.4). The most upregulated and most downregulated genes are listed in (Table 6.1 and Table 6.2). Among these genes were several with previous known role in cholesterol efflux (*ABCA1*, *ABCC3*) [270] and oxidative stress response (*NQO1*, *HMOX1*) [271].

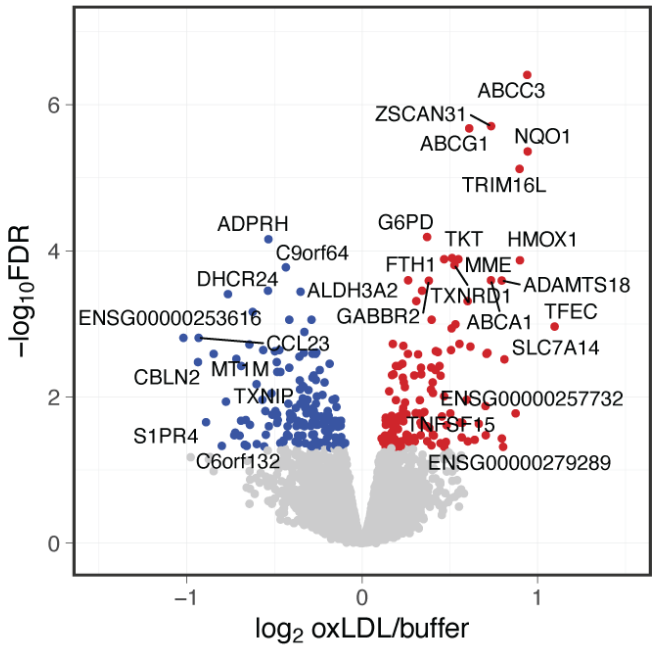


Figure 6.4 Volcano plot showing differentially expressed genes induced by ox-LDL in human aortic endothelial cells.
Significant results are coloured in red and blue for up-regulated and down-regulated genes, respectively. Log2 fold changes in gene expression are plotted on the x axis and false-discovery rates (FDR) are plotted on the y axis.

Table 6.1 Top 10 most upregulated protein-coding genes by ox-LDL in HAECs

Gene	$\log2\left(\frac{\text{oxLDL}}{\text{Buffer}}\right)$	FDR	Description
TFEC	1.097	1.09E-03	Transcription factor EC
NQO1	0.943	4.36E-06	NAD(P)H quinone dehydrogenase 1
ABCC3	0.940	3.91E-07	ATP binding cassette subfamily C member 3
HMOX1	0.899	1.35E-04	Heme oxygenase 1
SLC7A14	0.810	3.05E-03	Solute carrier family 7 member 14
TNFSF15	0.796	3.70E-02	TNF superfamily member 15
ADAMTS18	0.795	2.55E-04	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 18
ZSCAN31	0.735	1.97E-06	Zinc finger and SCAN domain containing 31
ABCA1	0.734	2.53E-04	ATP binding cassette subfamily A member 1
SLC7A11	0.708	2.57E-03	Solute carrier family 7 member 11

Table 6.2 Top 10 most downregulated protein-coding genes by oxLDL in HAECs

Gene	$\log2\left(\frac{\text{oxLDL}}{\text{Buffer}}\right)$	FDR	Description
CBLN2	-0.936	3.33E-03	Cerebellin 2 Precursor
CCL23	-0.932	1.55E-03	Chemokine ligand 23
S1PR4	-0.890	2.21E-02	Sphingosine-1-phosphate receptor 4
MT1M	-0.847	2.57E-03	Metallothionein 1E
C6orf132	-0.800	4.67E-02	Chromosome 6 Open Reading Frame 132
TXNIP	-0.776	1.16E-02	Thioredoxin-interacting protein
DHCR24	-0.765	3.91E-04	24-dehydrocholesterol reductase
BDKRB2	-0.729	3.34E-02	Bradykinin Receptor B2
CD14	-0.724	3.08E-02	CD14 molecule
COL21A1	-0.717	3.01E-03	Collagen Type XXI Alpha 1 Chain

The most upregulated protein-coding gene is *TFEC*, a member of the MiT/TFE transcription factor family. The MiT/TFE family has been linked to lysosome biogenesis in cancer cells [272], and in the previous chapter I have demonstrated ox-LDL-induced expression of *MITF* in macrophages (Figure 5.2). However, the cellular function of *TFEC* is poorly characterised, and no other MiT/TFE members were differentially expressed in ox-LDL-exposed endothelial cells.

To investigate the biological pathways affected by ox-LDL, I performed an enrichment test for KEGG and MSigDB pathways (Figure 6.5). Consistent with our observation in macrophages, ox-LDL treatment inhibited intrinsic cholesterol biosynthesis and triggered the anti-oxidative-stress NFE2 signalling pathways. Interestingly, ferroptosis, an iron-dependent form of necrosis was identified as the most significant pathway up-regulated by ox-LDL exposure (Figure 6.5). In murine endothelial cells, it has been reported that accumulation of ox-LDL induced ferroptosis through disrupted lipid peroxidation and disrupted mitochondria, which further promoted atherosclerosis [273].

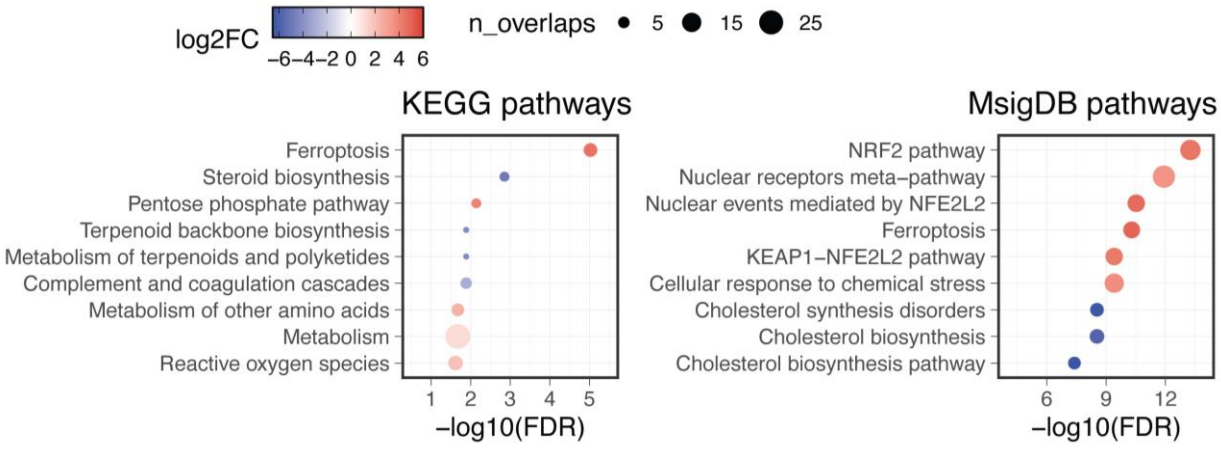


Figure 6.5 Pathways disrupted by ox-LDL exposure.
KEGG and MsigDB pathways enriched in the differentially-expressed genes induced by ox-LDL. Colour represents log2 fold change and size represents the number of overlapped genes in each pathway.

In Chapter 4 I conducted a meta-analysis of single-cell RNA-seq data from 38 atherosclerotic samples (Figure 4.18). As an external reference, I subsetted the plaque endothelial cells and performed a differential expression test between disease tissue and adjacent control tissue. Although plaque endothelial cells were exposed to different microenvironments from our *in vitro* ox-LDL model, they shared an ox-LDL-related transcriptomic signature: the overall expression of ox-LDL-repressed genes was significantly lower in plaque ECs, while ox-LDL-induced genes were moderately upregulated in plaques (Figure 6.6A); a set of sliding hypergeometric tests identified significant overlaps at both ends of the differential expression gene list (Figure 6.6B), showing the same set of genes being up- or down-regulated in both plaque ECs and ox-LDL-exposed ECs. As a comparator, similar analysis was performed on a public RNA-seq dataset of HAEC treated with ox-PAPC [266], one of the active components of ox-LDL [274]. However, unlike our ox-LDL data, the transcriptomic signatures related to the ox-PAPC response did not align well with that related to plaque endothelial cells (Figure 6.6C). Together, these results highlighted the disease relevancy of our *in vitro* ox-LDL-exposure model, which partly captures the *in vivo* transcriptomic dynamics in plaques and outperforms existing studies with a narrower focus on ox-PAPC.

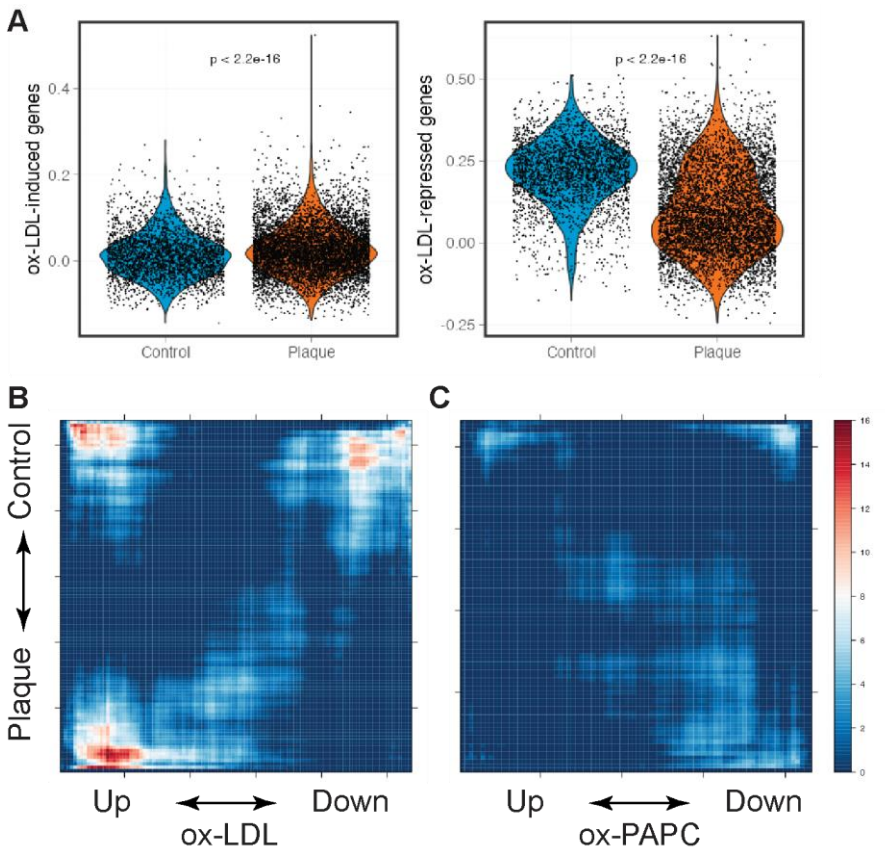


Figure 6.6 Exposure to ox-LDL but not ox-PAPC recapitulates the transcriptomic signatures associated with endothelial cells in human atherosclerosis.

A, Average expression levels of ox-LDL-induced (left) or ox-LDL-repressed genes (right) in single-cell RNA-seq profiles of endothelial cells from human atherosclerotic plaques (orange) or adjacent control tissue (blue). **B** and **C**, Rank-rank hypergeometric overlap tests showing similarity between transcriptomic signatures related to plaque ECs (y axis), ox-LDL-treated HAECs (x axis in **B**) and ox-PAPC-treated HAECs (x axis in **C**). Note that genes were divided into bins by their differential expression statistics and only the end of x or y axis represent significant differentially-expressed genes. The colour denotes the significance of each hypergeometric test ($-\log_{10}$FDR), measuring the extent of overlap between the two gene set on a sliding basis.

6.4. Integrative analysis highlights a shared ox-LDL-response network between endothelial cells and macrophages

To gain insights into the epigenetic changes induced by ox-LDL, I conducted bulk ATAC-seq on HAECs from 4 different donors. The ATAC-seq libraries were prepared in parallel with the RNA-seq samples to minimise any potential random effects of experimental conditions. In ATAC-seq experiments, open chromatin regions are cut by Tn5 transposases and flanked by sequencing adapters. To enable a high-resolution mapping of chromatin accessibility, I resized ATAC-seq read-pairs to 9-bp fragments at each end, and performed downstream analysis on these fragments (See Methods). Unlike differential expression, differential accessibility analysis can be severely affected by the choice of data normalisation method [275], and a wide range of parameters have been tested to rule out the possibility of misspecification in the differential testing model. Surprisingly, although the promoter accessibility positively correlates with the gene expression (Figure 6.7A), no region was confidently identified as differential accessible site (Figure 6.7B).

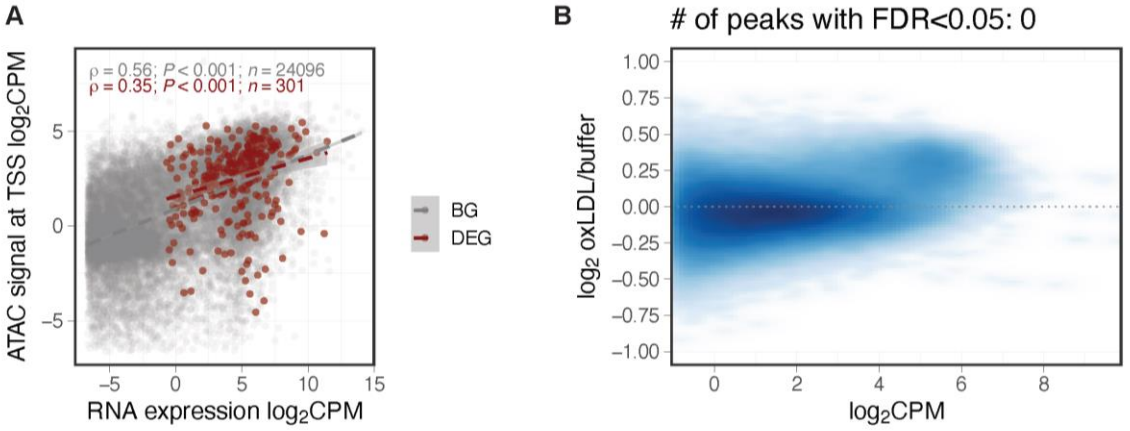


Figure 6.7 Limited effect of ox-LDL on the global open chromatin landscape in endothelial cells.

A, Correlation between gene expression and chromatin accessibility at the 2-kb window centred at the transcription start site (TSS) in differentially expressed genes by ox-LDL (red) and all other genes (grey). **B**, Fold change in chromatin accessibility plotted against the average accessibility across all open chromatin regions in HAECs. No open chromatin peak is identified as differentially accessible using the FDR threshold of 0.05.

6.4.1. In-silico deletion analysis

I reasoned that gene expression changes in the absence of chromatin accessibility change could result from the re-organisation of transcription regulator binding in pre-existing open chromatin. To investigate the underlying regulators driving the ox-LDL-induced gene expression, I performed in-silico-deletion (ISD) analysis using Lisa [127], integrating both the RNA-seq and ATAC-seq data. In brief, Lisa first transforms the open chromatin information into a gene-wise score called regulatory potentials (RPs); TF bindings, inferred from public ChIP-seq datasets, were then deleted *in silico* from the ATAC-seq signals. For each ISD, the statistical significance was evaluated by comparing the RP changes in differentially-expressed genes versus random background genes. This was done under the assumption that the ISD of the causal TF regulator would result in a larger RP change in differentially expressed genes than others. Candidate transcriptional regulators with significant ISD effects on ox-LDL-induced and ox-LDL-repressed genes are shown in Figure 6.8A and B.

Previous work from our group has demonstrated that ox-LDL exposure activates AP-1 and NFE-2 family factors in human macrophages [113]. Using single-cell multi-omics techniques, I have identified the directional effect of different factors within the AP-1 family, including an activating role of FOSL2 and a repressive role of FOS (Figure 3.11). Here in endothelial cells, AP-1 and NFE-2 members were also found to regulate ox-LDL-response

genes (Figure 6.8A and B; AP-1 and NFE-2 family factors are marked in red). Specifically, FOSL2 emerged as a candidate regulator of ox-LDL-induced genes, while FOS and JUN were linked to genes that were down-regulated by ox-LDL exposure.

The transcription factor CEBPB, which was linked to the newly discovered lipid-handling macrophage subpopulation in previous chapters, displayed the strongest ISD effect on ox-LDL-induced genes in HAECs (Figure 6.8A). It was estimated that CEBPB regulates about 20% of these genes, a prediction that remained consistent across various ChIP-seq datasets (Figure 6.8C).

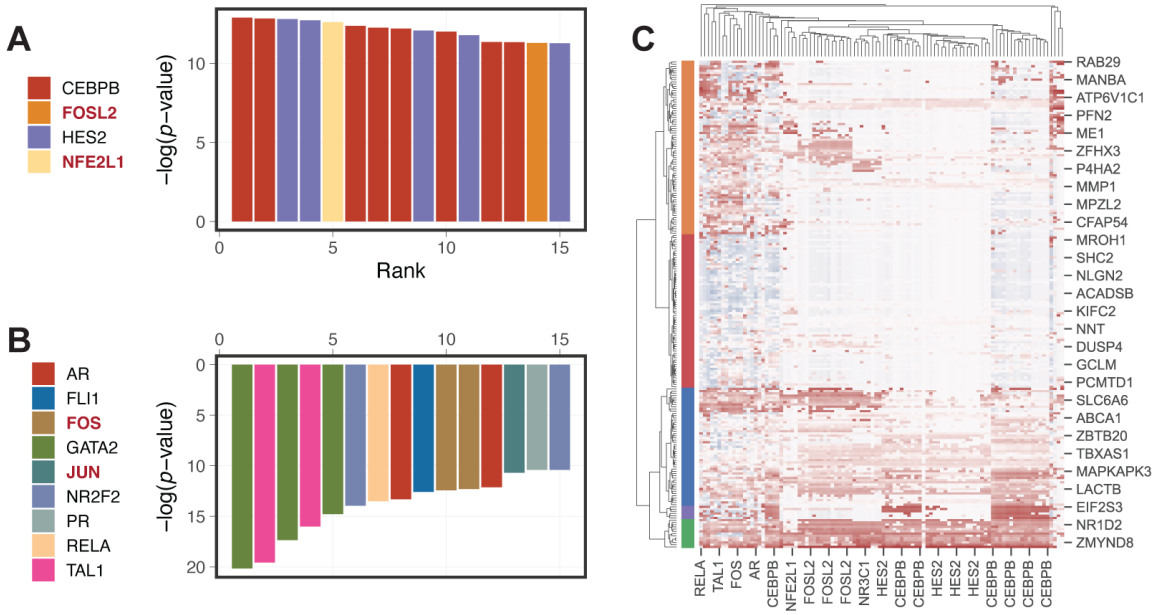


Figure 6.8 In-silico deletion (ISD) analysis based on public sequencing data identifies potential regulators underlying the transcriptomic response to ox-LDL. **A** and **B**, Candidate transcription factors regulating ox-LDL-induced genes (**A**) and ox-LDL-repressed genes (**B**) identified using the ISD analysis in LISA, ranked by significance. Note that each unique rank represents an independent ChIP-seq experiment and the same TF may appear more than once in the plot. **C**, Heatmap showing the predicted regulatory effect of each TF ISD (column) on the expression of individual genes (row) induced by ox-LDL treatment. The colour in heatmap represents predicted regulatory score (in an arbitrary unit), with red denoting activating effects. The ox-LDL-induced genes are clustered based on similarities in TF-gene regulatory effect and are plotted in the row.

A major limitation for the ISD analysis is the potential misalignment between the cellular context in which the TF bindings were inferred and the context where differential gene expression was measured. Specifically, it is unclear whether the inferred CEBPB bindings from cancer cell lines are involved in the cellular response to ox-LDL. To address this, I performed an ox-LDL-specific ISD analysis using a macrophage CEBPB ChIP-seq dataset previously collected by our group using our standardised ox-LDL model system [113].

Adopting the same approach, I computationally deleted all endothelial ATAC-seq signals from the open chromatin regions where we previously observed altered CEBPB binding in ox-LDL-exposed macrophages (n=8,718 peaks out of 100,602 total CEBPB binding sites); the remaining open chromatin signals were then transformed to RPs for each ox-LDL-induced genes (n=135 genes analysed). For comparison, the empirical null distribution of RP changes was estimated based on n=1000 set of expression-matched background genes. Consistent with previous ISD results from Lisa, there were significantly more genes being disrupted in ox-LDL-induced genes compared to the random background, with in total 11 (8.1%) genes showing >10% change in their RPs (p=0.002; Figure 6.9A). This indicates that the observed transcriptomic changes in ox-LDL-exposed endothelial cells are partly driven by the differential binding of CEBPB, highlighting a shared ox-LDL-response network between endothelial cells and macrophages.

However, unlike macrophages, the effect of ox-LDL on the expression of *CEBPB* is weak and statistically insignificant. Specifically, RNA-seq data support the up-regulation of *CEBPB* by ox-LDL in HAECs from two out of the four donors (Figure 6.9B), which were confirmed by immune-fluorescence staining (Figure 6.9C). However, this change was inconsistent and was not seen across other donors (Figure 6.9B).

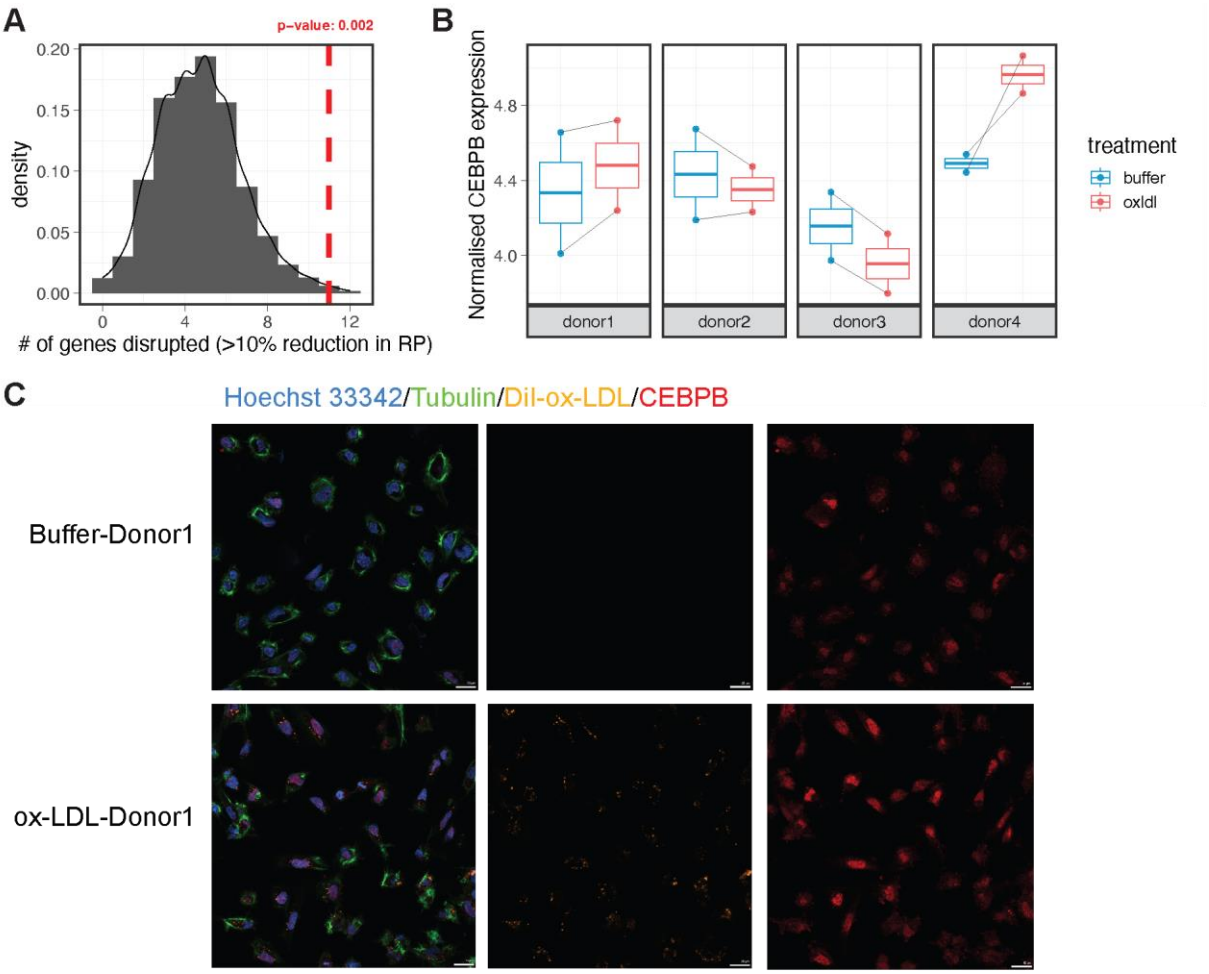


Figure 6.9 CEBPB is as a potential regulator of endothelial cell response to ox-LDL. **A**, In-silico deletion of ox-LDL-altered CEPBP-binding sites significantly disrupts the regulatory potential of ox-LDL-induced genes as compared to expression-matched background genes. Genes with >10% reduction in regulatory potential after the in-silico deletion were considered as being disrupted. The empirical p-value was estimated from the null distribution using n=1000 expression-matched background genes. **B**, expression of *CEBPB* across 4 biological replicates in 2 independent batches. **C**, Representative

images from immunofluorescence staining of CEBPB in unexposed and ox-LDL-exposed HAECs from donor 1.

6.4.2. TF motif footprinting analysis

Another popular approach to detect potential TF regulators is motif footprinting analysis. Because the presence of individual TFs at the open chromatin region sterically blocks Tn5 transposition, each ATAC-seq peak often contains many small regions with low accessibility signals (called footprints). In the previous section, I have demonstrated that the ox-LDL-induced differential gene expression can be explained in part by the differential binding of certain transcription factors. To investigate whether these differential binding patterns can be recovered from the ATAC-seq profiles, I next performed a genome-wide scan for changes in footprint signals using TOBIAS [129].

In brief, potential footprints were first identified by drops in local ATAC-seq signals in all accessible regions; an activity score was then assigned to each footprint per condition based on magnitude of accessibility drop, reflecting the strength of TF binding. Similar to differential accessibility analysis, relative TF bindings (i.e., changes in activity score) were tested between ox-LDL and control samples. Finally, footprints were grouped by matches for known TF binding motifs, and the statistical significance was evaluated by comparing relative binding scores within a TF group over an empirical null distribution. This analysis recovered two sets of information for each TF: the differential binding sites (DBSs), and the shift in global binding patterns induced by ox-LDL exposure.

Consistent with the ISD analysis, AP-1 / NFE-2 family factors were globally activated in ox-LDL-treated endothelial cells (Figure 6.10A), although the overall activation is weak

(8.0%, Figure 6.10B). CEBPB was not among the significant hits from footprinting analysis, possibly due to a known discrepancy between its *in vivo* binding pattern and the reported motif [276]. For the potential TFs repressed by ox-LDL, limited overlaps were observed between the footprint and the ISD analysis. This could be explained by the fact that ISD analysis relies on the TF binding information from public ChIP-seq datasets, thereby having limited power when testing poorly characterised TFs.

After merging the nearby footprints, I identified 2,453 TF binding sites that displayed at least two-fold change in footprint scores (n=1,172 and 1,281 sites for ox-LDL-activated and repressed sites, respectively). These DBSs located across the genome (Figure 6.10C), and were enriched for H3K27ac signals compared to average open chromatins (Figure 6.10D).

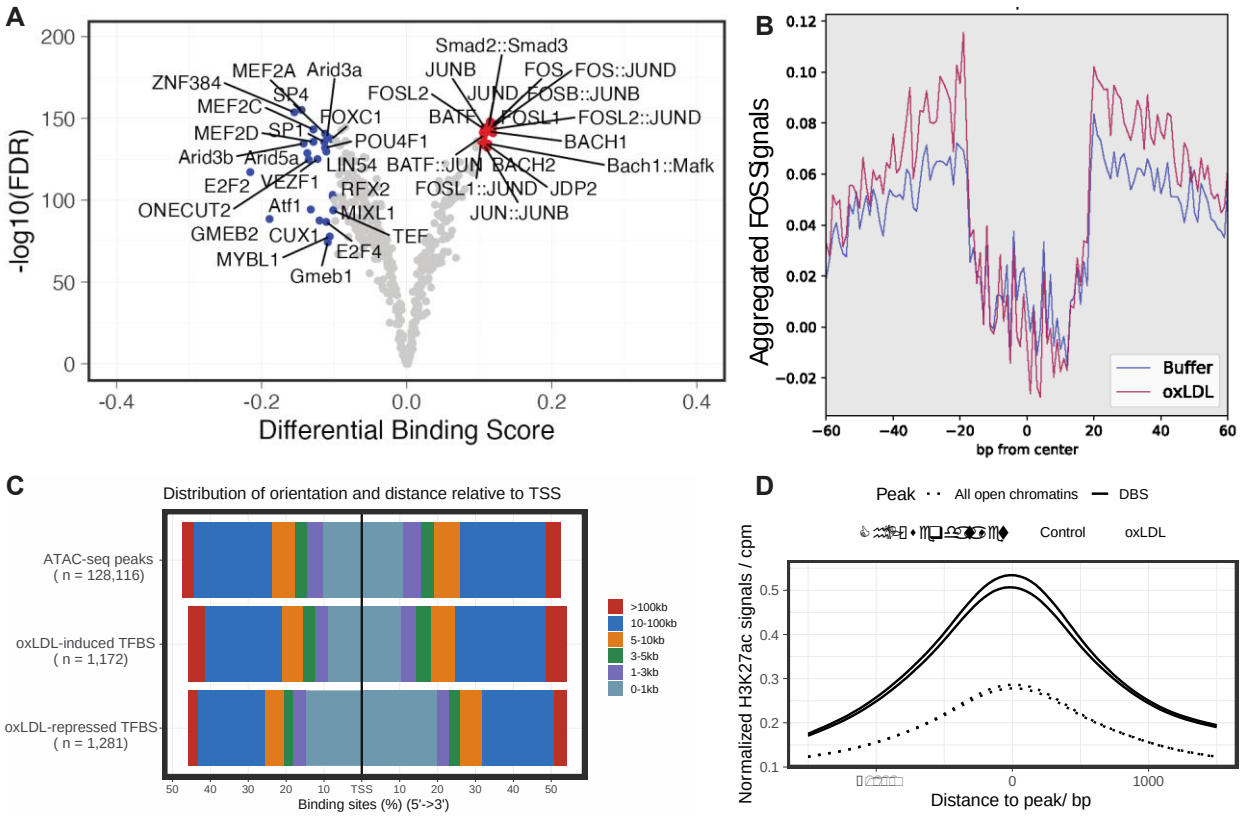


Figure 6.10 Ox-LDL triggers redistribution of transcription factor binding within open chromatin sites.

A, Volcano plot of transcription factors with genome-wide binding patterns changed after ox-LDL exposure. Significant effect induced or repressed by ox-LDL are coloured in red and blue, respectively. Only TF expressed in HAECs were included in this analysis. **B**, Aggregated foot print signals for FOS at n=1,223 potential binding sites with ox-LDL-induced activity. **C**, Genomic annotation of the identified TF binding sites induced or repressed by ox-LDL. **D**, Aggregated H3K27ac signals at peaks containing differential TF binding sites (DBS) or background open chromatin regions. Signals from unexposed (blue) and ox-LDL-exposed (red) HAECs are plotted separately. Shaded regions indicate 95% confidence intervals.

6.5. Differential transcription factor bindings induced by ox-LDL activate endothelial cell migration

To better understand the biological consequence of the observed change in TF binding, I conducted a pathway enrichment test in DBSs. Functional annotations related to endothelial migration were highly enriched in genes near both the ox-LDL-activated and

ox-LDL-repressed DBSs (marked in red in Figure 6.11A). Although some of these proximal genes with migration-related annotations showed elevated expression in ox-LDL samples, none passed the FDR correction. This is likely the reason why migration pathways were not highlighted in the differential gene expression analysis in Figure 6.5.

To experimentally validate this finding, I carried out a migration assay in HAECs stimulated with or without ox-LDL. Because ox-LDL was reported to promote proliferation in HAECs [277], the cells were starved in serum-depleted medium before and during the assay. This is to ensure that any observed migration differences would not be secondary to the change in cellular proliferation. Consistent with the pathway analysis of the DBSs, ox-LDL exposure in HAECs significantly increased EC migration (Figure 6.11B and C). To test the effect of ox-LDL priming, this experiment was also repeated in cells pre-treated with ox-LDL for 24 hours. I found a slight increase in migration at the start of the assay in ox-LDL-primed cells, possibly triggered by residual ox-LDL. After 20 hours, no significant difference in migration was observed in ox-LDL-primed cells (Figure 6.11D), indicating that the activated migration is a rapid response directly triggered by ox-LDL-exposure.

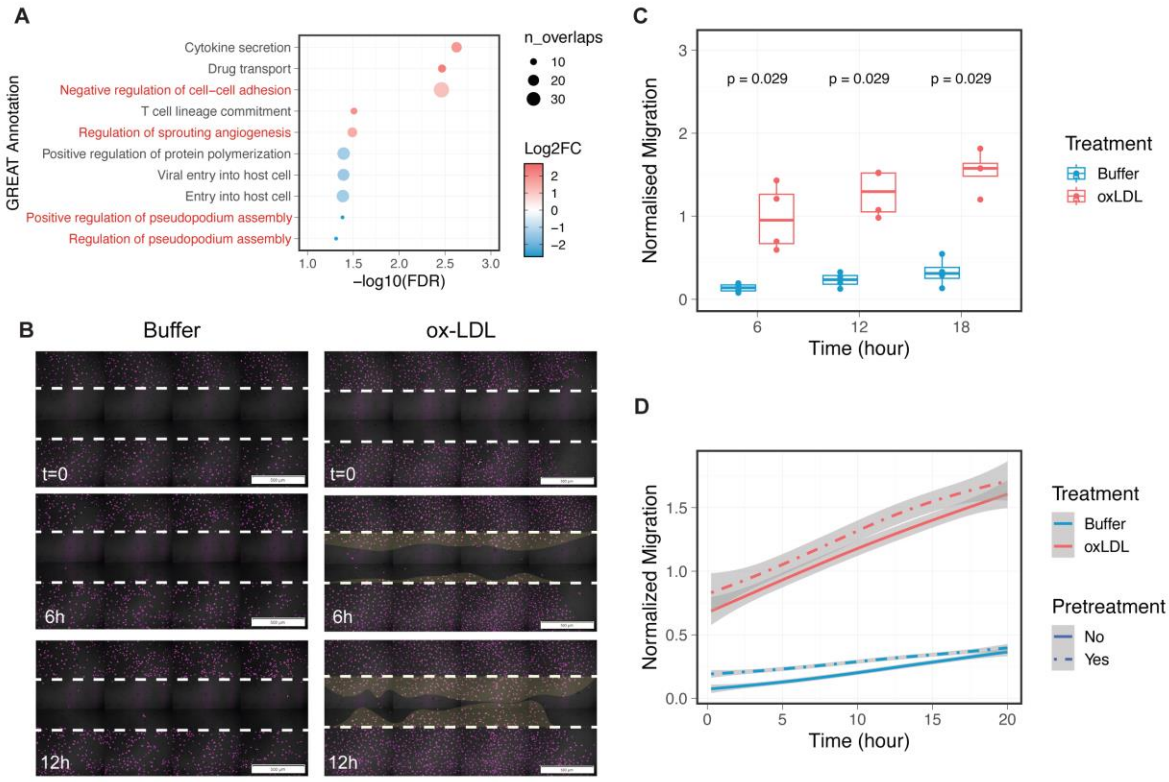


Figure 6.11 Changes in transcription factor binding activity indicate the activation of endothelial migration pathway in response to ox-LDL.

A, Predicted functions of the differential binding sites (DBS) using the Genomic Regions Enrichment of Annotations Tool (GREAT). Colour represents log2-fold change and size represents the number of overlapped genes in each annotation. Pathways associated with endothelial migration are highlighted in red. **B**, Representative images of the migration assay at specified time points, with cells migrated into the gap region highlighted. HAECs were incubated in serum-depleted medium with or without ox-LDL for up to 20 hours. Nuclei were labelled using 300nM of SiR-DNA. **C**, Endothelial migration quantified as the percentage of cells in the gap region normalised to that in buffer-treated cells at t=20h to compensate for donor variation (n=4 biological replicates). **D**, Normalised migration of HAECs with or without 24 hours of ox-LDL pre-treatment. Shaded regions indicate 95% confidence intervals estimated from n=4 biological replicates.

6.6. Prioritisation of CAD variants using the DBSs in endothelial cells

So far I have demonstrated that endothelial cells undergo major ox-LDL-induced changes in metabolism, transcriptome and epigenome. It is widely recognized that cells with a known biological role in a disease often show an enrichment of GWAS signals related to that disease [80,197,217]. Therefore, I hypothesised that the cellular response to ox-LDL in endothelial cells is also genetically linked to the development of atherosclerosis.

To test this hypothesis, I next profiled the relationship between the ox-LDL-induced changes and the SNP heritability of coronary artery disease using the s-LDSC software. Similar to the patterns observed in macrophages, the TF binding sites in endothelial open chromatin regions are enriched for CAD heritability. In addition, the ox-LDL-induced differential binding sites carry even higher per-SNP heritability than that in the background binding sites (Figure 6.12). Conversely, no enrichment was observed in background nor ox-LDL-induced binding sites for irrelevant traits like smoking behaviour. These results suggest that the genetic risk of CAD is concentrated in the TF binding sites altered by ox-LDL exposure, which in turn supports the prioritisation of causal variants by their overlaps with ox-LDL-induced DBSs.

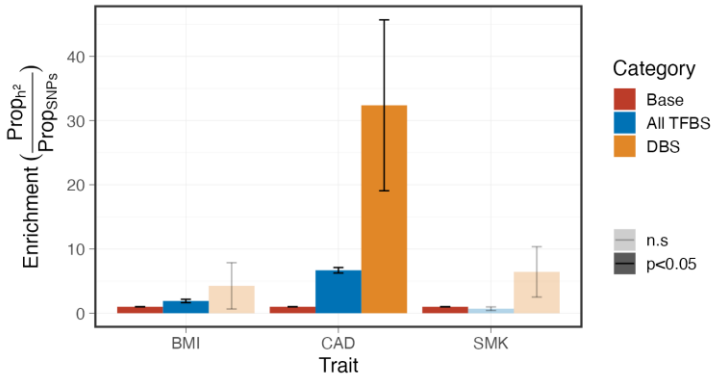


Figure 6.12 Ox-LDL-induced differential transcription factor binding sites are enriched for CAD heritability.
Similar to that in Figure 4.3, CAD heritability was partitioned into different genomic regions to evaluate the enrichment for genetic risks. Base: whole genome; All TFBS: open chromatins showing foot prints from TF binding; DBS: differential TF binding sites. Standard errors for enrichment statistics were estimated using s-LDSC. Non-significant enrichment results are plotted in transparent. GWAS summary statistics were obtained from GWAS Catalog using study ID GCST005194 [70], GCST006900 [211], and GCST007327 [72] for coronary artery disease (CAD), body-mass index (BMI) and smoking behaviour (SMK), respectively.

In total, 8 variants at 5 independent risk loci reaching genome-wide significance threshold ($p<5x10^{-8}$) located within a DBS (Table 6.3). The risk locus at chromosome 9 (9p21.3) has been extensively characterised using knockout mice models [278], although no QTL signals were found at the locus in human endothelial cells and vascular smooth muscle cells. The *PLPP3* locus overlaps an EC-specific enhancer reported in a previous variant prioritisation study [266], and has been linked to endothelial barrier function in various experimental models [279,280].

Table 6.3 Prioritised CAD-risk variants in ox-LDL-induced DBSs

Rsid	Chr	Pos	Ref	Alt	DBS	Annotation
rs56170783	chr1	56550459	A	C	chr1-56549815-56550830	PLPP3 intron
rs17048367	chr1	218660548	A	G,T	chr1-218660506-218661540	Intergenic region
rs7068966	chr10	12235993	C	G,T	chr10-12234966-12236001	CDC123 intron
rs77376996	chr2	187424445	T	TCTC,TCTCA	chr2-187423908-187424925	CALCRL intron
rs62172376	chr2	187424447	C	A	chr2-187423908-187424925	CALCRL intron
rs1537372	chr9	22103184	G	A,T	chr9-22103132-22104146	CDKN2B intron
rs1537373	chr9	22103342	T	A,G	chr9-22103132-22104146	CDKN2B intron
rs1333042	chr9	22103814	A	C,G	chr9-22103132-22104146	CDKN2B intron

6.6.1. rs56170783 at the *PLPP3* locus

The risk variant rs56170783 (β=0.109 in European population) overlapped with an ox-LDL-induced TF binding site at the *PLPP3* locus, and its risk allele A disrupt the binding motif for nuclear factor Y family (NFY) factors (Figure 6.13A). Interestingly, an allele-specific increase in ATAC-seq signals was observed in the risk allele of the heterozygous donor (Figure 6.13B), indicating a repressive role of NFY factor binding. This is consistent with a previous report that NFY protein complex represses the activation of an endothelial-specific promoter of the *VWF* gene [281].

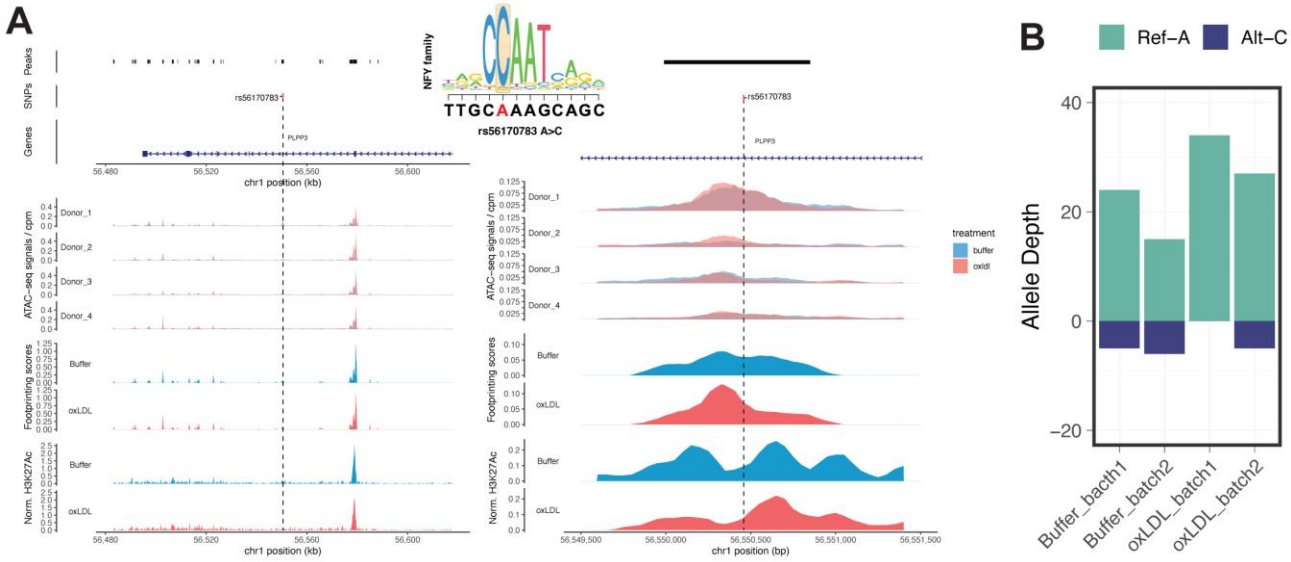


Figure 6.13 The regulatory landscape at the prioritised *PLPP3* locus.
A, The candidate causal variant rs56170783 is marked in red. Tracks (from top to bottom) showing: ATAC-seq peaks (peaks), CAD-associated variants (SNPs), genomic annotations (Genes), ATAC-seq signals aggregated by biological donors with signals from paired samples overlaid together, Motif footprinting score aggregated by treatment, and enhancer signals from H3K27ac ChIP-seq. Higher resolution coverage plots centred on the risk variant are plotted in the right panel. **B**, Allelic sequencing coverage at rs56170783 in the ATAC-seq libraries of four heterozygous samples from donor 4.

6.6.2. Prioritised variants at the *CALCRL*/*TFPI* locus

At the *CALCRL*/*TFPI* locus, a genetic variant rs840616 has been associated with CAD risks (β=0.0402) in European population [70]. In linkage disequilibrium with this lead variant (r^2=0.81), two closely-linked variants rs77376996 and rs62172376 overlapped with an intronic DBS in the *CALCRL* gene (Figure 6.14). Ox-LDL-induced increase in chromatin accessibility was observed in endothelial cells from three out of the four donors (Figure 6.14), and motif footprinting analysis suggested increased TF binding at this region.

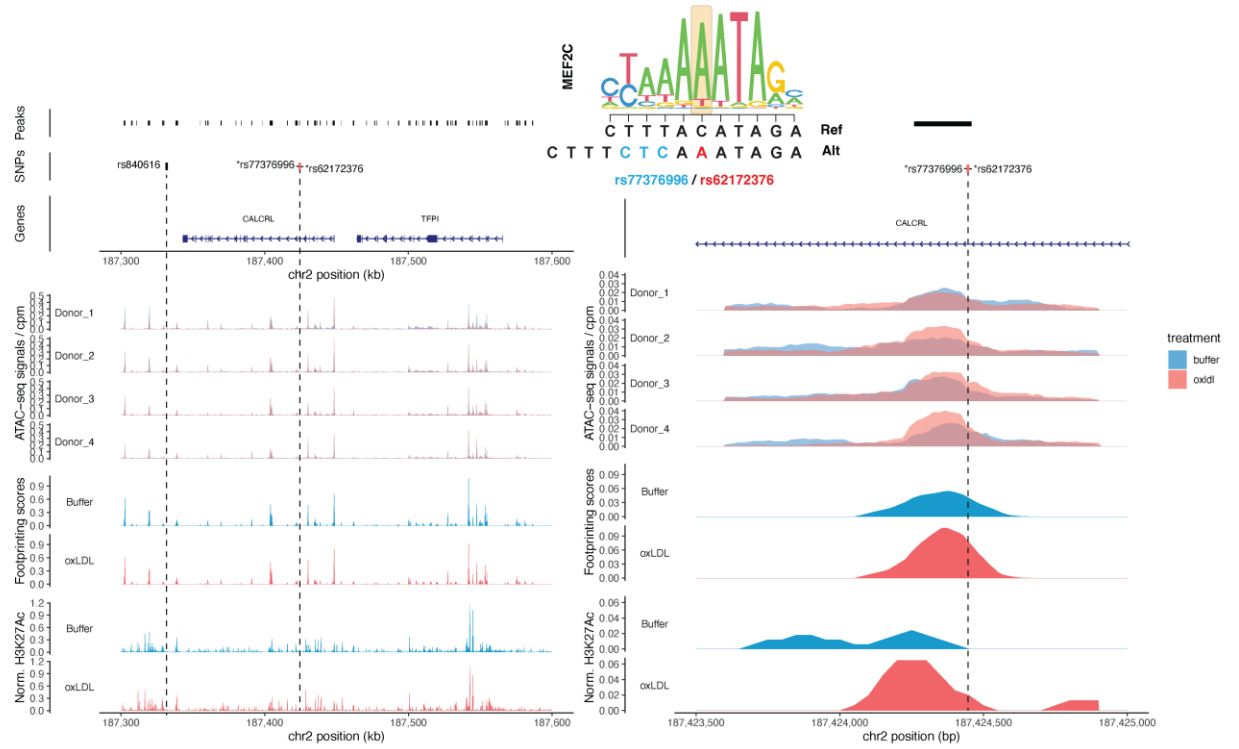


Figure 6.14 The regulatory landscape at the prioritised *CALCRL*/*TFPI* locus.
The tagged variant rs840616 is marked in grey and two candidate causal variants rs77376996 and rs62172376 are marked in red. Similar to that in Figure 6.13A, zoom-in views centred on the risk variant are plotted in the right panel.

There are two protein-coding genes *CALCRL* and *TFPI* expressed at this locus, together with a long non-coding RNA *CALCRL-AS1* antisense to these two genes. The protective

A allele of rs62172376 is linked to higher expression of both genes *CALCRL* and *TFPI* in human artery (Figure 6.15A). While the expression of *CALCRL* and *CALCRL-AS1* did not change, the nearby gene *TFPI* is weakly down-regulated in ox-LDL-treated endothelial cells (log2FC=-0.12, p=0.016; Figure 6.15B). A structural analogue from the TFPI family, *TFPI2*, is also amongst the differentially-expressed genes, although its expression increases after ox-LDL exposure (log2FC=0.34, p=7.8x10^{-3}; Figure 6.15C).

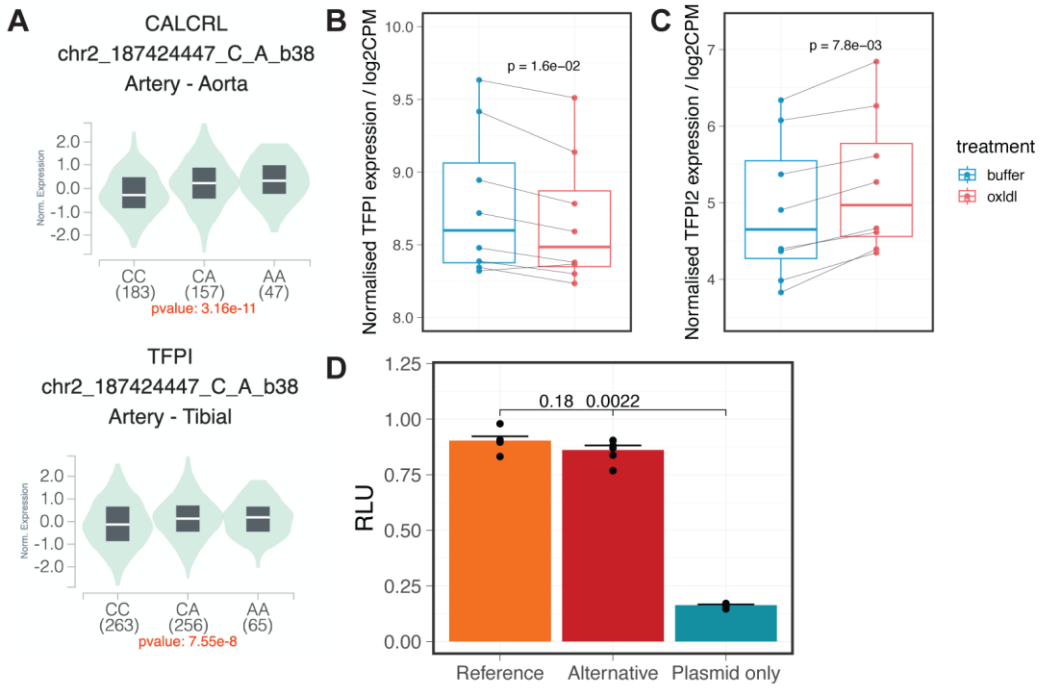


Figure 6.15 TFPI is the candidate causal gene at the prioritised locus.
A, Violin plot showing the allele-specific effect of the haplotype on the expression of *CALCRL* and *TFPI* in human artery based on the GTEx data. **B** and **C**, Normalised gene expression (log2 count-per-million-reads) of *TFPI* (**B**) and *TFPI2* (**C**) in unexposed and ox-LDL-exposed HAECs (n=8 samples per treatment group). P-values calculated with paired Wilcoxon tests. **D**, Luciferase reporter assays of the 150bp dsDNA construct containing each haplotype as shown in Figure 6.14. HEK293T cells were transfected with the indicated plasmids and lysed for luciferase activity measurement after 24 hours (n=6 independent transfections). Signals normalised to Renilla luciferase activities.

Given that the protective haplotype creates a binding site for MEF2 family factors (Figure 6.14), I next sought to determine whether it also alters the enhancer activity. Both the reference and alterative haplotypes containing these two SNPs were cloned into the reporter plasmid, and were then transfected into HEC293T cells. While the 150bp dsDNA insert exhibited significant enhancer activity, I did not observe a significant allelic-difference between the two haplotypes (Figure 6.15D), possibly due to low intrinsic expression of MEF2 factors in HEC293T cells (only 36.5% of that in human endothelial cells).

A closer inspection of the other molecular QTL signals provided an alternative explanation for the allelic effect of rs62172376 on TFPI expression: rs62172376 has been linked to a splicing QTL of the anti-sense gene *CALCRL-AS1*, where the alternative haplotype is associated with the retention of the intron overlapping the promoter region of *TFPI* in human coronary artery (Figure 6.16A and B). In HAECs, strand-specific usage of this intron positively correlated with the expression of *TFPI* (Figure 6.16C), but not *CALCRL* (Figure 6.16D).

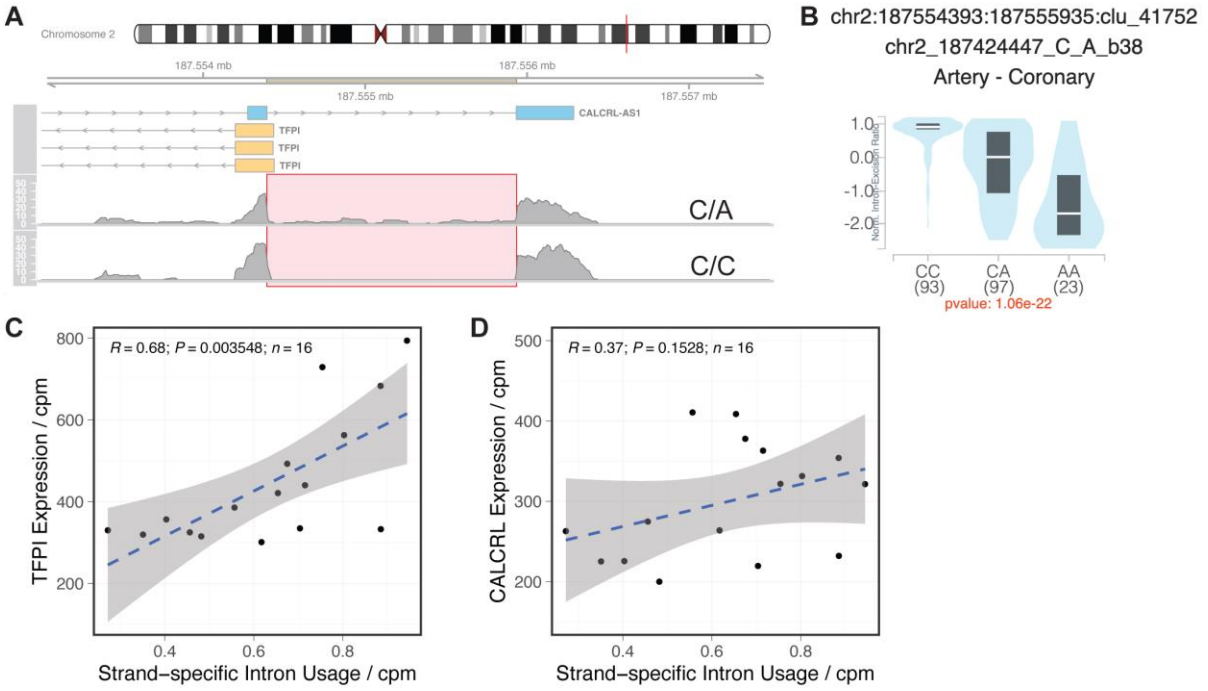


Figure 6.16 Alternative splicing of CALCR-AS1 correlates with TFPI expression in endothelial cells.

A, A zoom-in view of the annotated *CALCRL-AS1* transcripts at the promoter region of *TFPI*. The genomic region with an intron retention event is highlighted in red. The lower panel displays representative tracks that illustrate the strand-specific coverage of the above *CALCRL-AS1* transcript from both heterozygous and homozygous reference samples. **B**, Normalised excision ratio for the intron chr2-187554393-187555935 of *CALCRL-AS1* as shown in **A** in human artery based on the GTEx data. **C** and **D**, Pearson correlation between the strand-specific usage of the intron chr2-187554393-187555935 and gene expression of *TFPI* (**C**) and *CALCRL* (**D**). Gene and intron expression normalised to sequencing depth (count-per-million-reads). Shaded regions indicate 95% confidence intervals estimated from n=16 HAEC samples sequenced in this study.

6.7. Discussion

Exposure to ox-LDL in endothelial cells marks one of the earliest pathological events in atherosclerosis. In this chapter, I provide the first detailed transcriptomic and epigenomic characterisation of the response to ox-LDL in primary human aortic endothelial cells. Through integrative analysis of multi-omics high-throughput sequencing, I identify the

transcriptional regulators underlying the endothelial response to ox-LDL. These regulators feature a conservative ox-LDL-response network largely shared with human macrophages, as well as endothelial-specific pathways including migration. Importantly, I show that ox-LDL-induced change in transcription factor binding is enriched for CAD heritability, and can be used as a leverage point to prioritise genetic variants with potential regulatory effect on the endothelial responses to atherogenic lipids.

In HAECs, exposure to ox-LDL reduces mitochondrial respiration and shifts the energy balance towards glycolysis. This is contradictory to the effect of ox-PAPC, a widely studied bioactive component in ox-LDL, which has been reported to increase the respiration and metabolic activity in HAECs [267]. In addition, our differential expression analyses indicate a strong alignment between the *in vivo* findings in human atherosclerotic lesions and the effect of ox-LDL, while no alignment was found for ox-PAPC-induced transcriptomic signatures. Clearly, the proatherogenic effect of ox-LDL is a combination of complex, sometimes opposing influences from individual components. Studying one specific element within the ox-LDL mixture may miss the whole picture and so could lead to misleading conclusions.

While ferroptosis is one of the most up-regulated pathway upon ox-LDL exposure and has been linked to accelerate endothelial dysfunction [282], ox-LDL alone is not sufficient to trigger the pro-inflammatory phenotype in HAECs as described in previous studies [253,254,261]. I noticed that in one of these papers [261], activation of inflammation was only observed when the lectin-type oxidized LDL receptor 1 (OLR1) was overexpressed in HAECs. Despite being first identified as an endothelial receptor in bovine cells [251], OLR1 is barely expressed in human endothelial cells from either disease-free vasculatures or

atherosclerotic plaques. This means that previous research focused on the ox-LDL-OLR1 pathway in endothelial cells may be severely confounded by the cellular response to the artificially activated OLR1 pathway, and possibly have missed the intrinsic property of ox-LDL on endothelial cells under physiological conditions.

The effect of ox-LDL on endothelial migration pathway is controversial, as opposite phenotypes were observed in different endothelial lines [277,283,284]. Most early research on vasculature uses endothelial cells from human umbilical vein (HUVECs). However, HUVECs may not be representative of the actual *in vivo* conditions as they are derived from immune-privileged foetal tissue and may differ significantly from adult vascular endothelium. More specifically, ox-LDL was reported to inhibit migration via dephosphorylation of Akt in HUVECs[283], but increased Akt phosphorylation and promoted proliferation in HAECs[277]. In our study, an activating effect of ox-LDL on endothelial migration was identified from unbiased genome-wide TF footprinting analysis and was functionally validated using live-cell microscopy. The enriched disease heritability in differential TF binding sites further supported the observed *in vitro* migration phenotype being causally involved in the *in vivo* progression of human atherosclerosis.

Since many non-coding genetic variants influence disease risks by altering the function of the regulatory DNA [203], profiling the colocalization of regulatory regions with genetic risk variants in LD offers a powerful approach to decipher the causal variants and their target genes [94]. One of the prioritised risk loci around the *CALCRL* and *TFPI* gene is particularly interesting, as only one of the two coding genes is differentially expressed by ox-LDL exposure. Unlike the *FDX1* variant characterised in the previous chapter, the prioritised variants at this locus display no allelic effect on enhancer activity. Although the

luciferase assay was performed in an epithelial-like kidney cell line HEK293T due to the extremely low transfection efficiency in HAECs, the molecular mechanisms underlying the regulatory variants are likely to be the same across different human cell lines. Therefore, it is likely that the prioritised variants influence gene expression in a more complex and controlled manner. One possible mechanism operates through the enhancer-associated noncoding transcript: enhancers have been found to produce long noncoding RNA transcripts to augment their regulatory activity [285], and antisense noncoding RNA upregulates gene expression through inhibition of non-productive alternative splicing [286]. In human artery samples, rs62172376 was linked to an intron-retention event that generates a noncoding transcript antisense to the promoter of *TFPI* gene. Our transcriptomic data in HAECs show that this retained antisense intron positively correlates with the expression of *TFPI* but not *CALCRL*, highlighting the role of alternative splicing in mediating the genetic risk at this locus.

Ch 7. General discussion

The central question of this thesis revolves around two key stages of atherogenesis induced by exposure to oxidised low-density lipoprotein (ox-LDL), the formation of foam cells and endothelial cell dysfunction. The gene expression patterns and transcriptomic regulation underlying these processes, along with their interaction with the genetic component of atherosclerosis, hold promise for discovering new targets for clinical diagnosis and precise therapeutic intervention. However, these areas remain largely unexplored.

To address this, I utilised high-throughput multi-omics sequencing in primary human macrophage and endothelial cell models. This enabled us to systemically profile the molecular signatures associated with ox-LDL exposure and identify key regulatory DNA elements that underpin this pathogenic response. In conjunction, the baseline and ox-LDL-induced transcriptional regulation networks were integrated with GWAS data to interpret the functions of genetic risk variants, leading to the prioritisation of 129 causal variants at 49 independent loci in macrophage and endothelial cells.

Using cutting-edge single-cell techniques, our recently published work [133] described in chapter 3, 4 and 5, delivered the first genetic evidence for the diverse influences of heterogenous human macrophage subpopulations on CAD heritability. It also established the molecular mechanism by which a CAD risk variant rs10488763 is specifically activated in the risk-enriched lipid-handling macrophage subpopulation, and, for the first time, implicated its target *FDX1* as a disease gene.

7.1. An adaptable and scalable framework for exploring the interactions between genetic predisposition and environmental risk factors

Twin studies have provided compelling evidence that nearly all human traits are heritable [67]. Genetic studies, despite their limitations [287], provide a robust platform to infer cause-and-effect relationships [288]. Except for those in close linkage-disequilibrium, genetic variants are naturally shuffled during meiosis, meaning that any regulatory effect of these variants on a phenotype should be causal. This rationale largely attenuates the burden of proof for causation in functional genomic analysis. Unlike traditional molecular investigations that focused on pre-specified candidate genes, GWAS-informed high-throughput omics studies offer unbiased insights into the causal mechanisms underlying complex diseases, including CAD.

In the present study, I demonstrate a computational and experimental framework that identifies causal genetic variants and connects these variants to their molecular and cellular functions. While my primary focus is on the cellular response to ox-LDL, a well-known atherogenic stimulus, the strategy applied is highly generalisable to other environmental stimuli and diseases. Indeed, many previous studies, including our own, have illustrated context-specific effect of genetic variants through high-throughput sequencing [113,289,290].

There is also a growing interest in integrating genetic data with single-cell transcriptomic [186,291] and single-cell epigenomic profiling [292] in primary disease tissues. The high-resolution of genomic maps and the ability of capturing understudied, yet critical cell types

make single-cell sequencing a promising strategy for functional genomic analysis. In this thesis, I have developed a computational pipeline that systematically tests the enrichment for disease heritability in transcriptomic gene programmes, which substantially outperforms existing methods, including the “nearest gene” approach and the ABC-enhancer model. Our computational pipeline is generalisable to future single-cell multiome study not only for standard cell-type-enrichment analysis, but also for fine-grained tasks such as delineating intertwined enrichment signals through conditional analysis.

7.2. Lipid-handling macrophage, lipid-associated macrophage, and human atherosclerosis

The primary macrophages that we have profiled in this study are differentiated *ex vivo* from circulating monocytes with minimal disturbance. Although macrophages in atherosclerotic lesions are largely derived from circulating monocytes [154], they are exposed to complex microenvironments and might have very different transcriptomic profiles. Clearly, it is an important objective to investigate whether the heterogeneous macrophage subpopulations we observed, especially the lipid-handling phenotype, are reproducibly present in human atherosclerosis.

It is somewhat counterintuitive that integrative single-cell transcriptomic analysis links almost all macrophages profiled in our study to one specific subpopulation in plaque, the lipid-associated macrophage (LAM). LAMs are first discovered in liver and adipose tissue [28,216], and predominantly occur in lesions with metabolic complications compared to healthy tissues [218,293,294]. The ontogeny of LAMs has not been fully resolved, with both

recruitment of circulating monocytes [295,296] and repolarisation from tissue-resident macrophages [297] being proposed as the cellular origin of LAMs in murine adipose tissues. Our analysis provides evidence supporting a monocyte-derived origin for LAMs in human atherosclerosis.

In mouse model of atherosclerosis, LAM corresponds to a lipid-rich foamy macrophage phenotype with lipid-metabolism and anti-inflammatory signatures [298-300]. On the contrary, the lipid-handling macrophages identified in this study, while also display activated lipid-metabolism pathways, do not accumulate intracellular lipids upon ox-LDL exposure. Given that monocyte-derived macrophages exhibit tight transcriptomic similarity to plaque LAMs, this suggests the presence of distinct functional subsets or cellular states within the LAM subpopulation. Taking advantage of the greatly increased sample size from our single-cell meta-analysis of human plaque samples, I identify a novel CD52^{hi} subset of plaque LAM, which shows high and specific expression of *ex vivo* lipid-handling macrophage markers including *CD36*, *CHI3L1* and *LPL*. Targeting LAM as a whole through myeloid-specific know-down of *TREM2* has shown contradictory effects on the overall burden of atherosclerosis in murine models [31,301], partly due to the complex role of different LAM subsets at various stages of atherogenesis. Our findings uncover the transcriptomic and functional heterogeneity within LAM, and highlight the CD52^{hi} lipid-handling phenotype as a more precise target for therapeutic exploitation of LAMs.

7.3. Reconciling the ox-LDL effect in endothelial cells

Endothelial cells, due to their direct exposure to circulating blood, have long been believed to be the first cell type to experience the pathogenic effects of LDL and its derivatives. The classical model of endothelial activation at the onset of atherosclerosis suggests that the retention and oxidisation of LDL particles in the vascular walls damage the endothelial barrier, leading to the activation of adhesion molecules on the endothelial surface and the recruitment of circulating leukocytes [252].

Contrary to much previous research on ox-LDL and endothelial function, this study finds a relatively small effect of ox-LDL on human aortic endothelial cells (HAECs). While approximately 300 genes were identified as differentially expressed, the expression changes for most of these genes were moderate, with only 55 genes showing changes higher than 50%. These transcriptomic alterations do not coincide with global changes in chromatin accessibility, but are rather associated with redistributions of transcription factor bindings within open chromatin regions. Specifically, no activation of inflammatory pathways, including the previously reported cell adhesion molecule pathway [254], was observed in current study. As a comparator, an acute inflammatory response triggered by TNF-α stimulation in HAECs significantly increased the surface levels of adhesion molecules, and was shown to modulate the chromatin accessibility at more than 8,000 regions [302].

In vivo, ox-LDL does not operate in isolation. Analysis of supernatant from *ex vivo* cultured human plaque cells revealed the presence of TNF-α, IL1-β and various other proinflammatory cytokines [303]. While our data suggest that ox-LDL alone is insufficient to

trigger endothelial activation, it has been reported to exacerbate the inflammatory phenotype in already damaged endothelium exposed to TNF-α [258], infected with Chlamydia pneumoniae [304], or overexpressing ox-LDL receptor 1 [261].

Additionally, the apical surface of endothelial cells is constantly exposed to the shear stress from circulating blood under physiological conditions. Our group has previously developed an *in vitro* shear stress model of HAECs and has uncovered a drastic difference in cells grown under laminar shear stress compared to standard static cell culture conditions. It is certainly possible that the pathological endothelial phenotype triggered by ox-LDL is obscured by the atherogenic effect of growing endothelial cells under static conditions [305].

7.4. Limitations and future directions

The experimental design of this study is centred around the exposure of cells to ox-LDL. While the response to ox-LDL is undoubtably crucial in human atherosclerosis, our findings may not account for other atherogenic signals, potentially leading to false negatives in heritability analysis. For example, certain subpopulations may only display disease-related transcriptomic changes under specific cellular contexts, which may have been missed in our experimental system. Future research could benefit from single-cell multiome data on macrophages directly profiled from atherosclerotic lesions. This, combined with multi-omics data on macrophages treated with other atherogenic stimuli, could potentially increase the power of heritability analysis.

In addition, we only collected sequencing data after macrophage differentiation, which could potentially overlook transitional subpopulations and the associated information on

cell fate decisions. The distinct subpopulations identified in this thesis may represent different *in vivo* functional states, with possible interconversion between each other. Future research is required to better understand the origin and plasticity of lipid-handling macrophages, possibly through time-course single-cell sequencing or flow cytometry analysis.

This thesis research operates on the assumption that causal genetic variants modulate disease risk by altering the transcriptional regulation of disease genes. While RNA abundance is undeniably important, it only represents a portion of the functional spectrum of effects of non-coding DNA variations. Other mechanisms, such as the splicing and translation of transcripts, as well as the post-translational modification of proteins, may also play significant roles in the molecular mechanisms of disease heritability and could be characterised in future omics studies.

7.5. Conclusions

The findings of this thesis provide a comprehensive understanding of the transcriptional regulation of ox-LDL response in macrophages and endothelial cells, as well as its interaction with the genetic component of atherosclerosis. Our research illustrates the substantial value of employing a multi-omics approach to improve the genomic resolution of the genetic architecture of human diseases. The computational and experimental framework outlined in this thesis is generalisable to other complex human diseases, and will greatly expedite the translation of genetic discoveries in the emerging era of genomic medicine.

Appendix

Supplemental materials

Supplemental materials including Online Table S1-S5 and Online Video can be found at:

https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.123.324172#supplementary-materials

Online Table S1 OxLDL-induced differentially expressed genes in macrophages identified by single-cell transcriptomics.

Online Table S2 Marker genes of macrophage subpopulations identified by single-cell multi-omics.

Online Table S3 Predicted macrophage cis-regulatory elements.

Online Table S4 TF-DORC pairs with strong cis-regulations.

Online Table S5 Prioritized CAD risk variants.

Published work

The work described in Chapter 3, 4, and 5 has been published in a peer-reviewed research article, which is attached in the following pages:

Jiang J, Hiron TK, Agbaedeng T, Malhotra Y, Drydale E, Bancroft J, Ng E, Reschen ME, Davison LJ, O'Callaghan CA. A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease. *Circ Res.* 2024 May 15. doi: 10.1161/CIRCRESAHA.123.324172. PMID: 38747151

Check for updates

Circulation Research

ORIGINAL RESEARCH

A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease

Jiahao Jiang,* Thomas K. Hiron,* Thomas A. Agbaedeng, Yashaswat Malhotra, Edward Drydale, James Bancroft, Esther Ng, Michael E. Reschen, Lucy J. Davison, Chris A. O'Callaghan

BACKGROUND: Coronary artery disease (CAD), the leading cause of death worldwide, is influenced by both environmental and genetic factors. Although over 250 genetic risk loci have been identified through genome-wide association studies, the specific causal variants and their regulatory mechanisms are still largely unknown, particularly in disease-relevant cell types such as macrophages.

METHODS: We utilized single-cell RNA-seq and single-cell multiomics approaches in primary human monocyte–derived macrophages to explore the transcriptional regulatory network involved in a critical pathogenic event of coronary atherosclerosis—the formation of lipid-laden foam cells. The relative genetic contribution to CAD was assessed by partitioning disease heritability across different macrophage subpopulations. Meta-analysis of single-cell RNA-seq data sets from 38 human atherosclerotic samples was conducted to provide high-resolution cross-referencing to macrophage subpopulations in vivo.

RESULTS: We identified 18 782 cis-regulatory elements by jointly profiling the gene expression and chromatin accessibility of >5000 macrophages. Integration with CAD genome-wide association study data prioritized 121 CAD-related genetic variants and 56 candidate causal genes. We showed that CAD heritability was not uniformly distributed and was particularly enriched in the gene programs of a novel CD52-hi lipid-handling macrophage subpopulation. These CD52-hi macrophages displayed significantly less lipoprotein accumulation and were also found in human atherosclerotic plaques. We investigated the cis-regulatory effect of a risk variant rs10488763 on *FDX1*, implicating the recruitment of AP-1 and C/EBP-β in the causal mechanisms at this locus.

CONCLUSIONS: Our results provide genetic evidence of the divergent roles of macrophage subsets in atherogenesis and highlight lipid-handling macrophages as a key subpopulation through which genetic variants operate to influence disease. These findings provide an unbiased framework for functional fine-mapping of genome-wide association study results using single-cell multiomics and offer new insights into the genotype-environment interactions underlying atherosclerotic disease.

GRAPHIC ABSTRACT: A graphic abstract is available for this article.

Key Words: atherosclerosis ■ coronary artery disease ■ cholesterol, LDL ■ genome-wide association study ■ macrophages ■ multiomics ■ single cell analysis

In This Issue, see p 2 | Meet the First Author, see p 3

Coronary artery disease (CAD) is a leading global cause of morbidity and mortality[1] and results from atherosclerosis, a chronic inflammatory process within the artery wall.[2] Monocytes and macrophages play a central role in atherosclerosis and can engulf modified cholesterol-carrying lipoproteins to form

Correspondence to: Chris A. O'Callaghan, DM, DPhil, FRCP, Nuffield Department of Medicine, Wellcome Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom. Email chris.ocallaghan@ndm.ox.ac.uk
*J. Jiang and T.K. Hiron contributed equally.
Supplemental Material, the podcast, and transcript are available with this article at https://www.ahajournals.org/doi/suppl/10.1161/CIRCRESAHA.123.324172.
For Sources of Funding and Disclosures, see page 22.
© 2024 The Authors. *Circulation Research* is published on behalf of the American Heart Association, Inc., by Wolters Kluwer Health, Inc. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial-NoDerivs License, which permits use, distribution, and reproduction in any medium, provided that the original work is properly cited, the use is noncommercial, and no modifications or adaptations are made.
Circulation Research is available at www.ahajournals.org/journal/res

Downloaded from http://ahajournals.org by on June 25, 2024

Novelty and Significance

What Is Known?

- Macrophages play a central role in atherosclerosis and in the uptake of atherogenic lipids such as oxidized low-density lipoprotein cholesterol.
- Human atherosclerotic lesions contain multiple macrophage subpopulations, but the distinct contribution of the different subpopulations to the disease process is unclear.
- Genome-wide association studies have identified over 250 genetic risk loci for coronary artery disease (CAD), many of which might act specifically through macrophages.

What New Information Does This Article Contribute?

- High-resolution genome-wide mapping of CRE (cis-regulatory elements) generated from single-cell multiomics links 121 CAD risk variants to 56 candidate genes and uncovers CAD-relevant transcriptional networks in the macrophage response to oxidized low-density lipoprotein cholesterol.
- The genetic heritability of CAD varies substantially across macrophage subpopulations and is significantly enriched in a novel CD52-hi lipid-handling macrophage subpopulation.
- *FDX1* is a novel disease gene regulated by risk variant rs10488763 specifically in lipid-handling macrophages.

Macrophages have long been recognized as central players in atherosclerosis, but their involvement in the genetic component of CAD remains poorly understood. In this study, we constructed a high-resolution map of macrophage CREs, which allows linkage of 121 candidate risk variants at 44 genetic loci to their 56 target genes. Based on this CRE map, we systemically quantified the disease relevance across heterogenous macrophage subpopulations, and discovered that CAD heritability was particularly enriched in a novel CD52-hi lipid-handling macrophage subpopulation. Functionally, these macrophages displayed activated lipid metabolism pathways, accumulated less lipid upon oxidized low-density lipoprotein cholesterol exposure, and were present in human atherosclerotic plaques. Mechanistically, we identified a specific CAD risk variant, rs10488763, which was activated in these lipid-handling macrophages and, for the first time, implicated its target *FDX1* as a disease gene. Altogether, our work provides novel insights into the genetic underpinnings of CAD and macrophage heterogeneity, paving the way for future therapeutic intervention targeting specific macrophage subpopulations in atherosclerosis.

Nonstandard Abbreviations and Acronyms

ABC	activity by contact
CAD	coronary artery disease
CRE	cis-regulatory element
DORC	domain of regulatory chromatin
GWAS	genome-wide association study
LAM	lipid-associated macrophage
LAP	liver-enriched activator protein
ox-LDL	oxidized low-density lipoprotein
scRNA-seq	single-cell RNA sequencing
TF	transcription factor

lipid-laden foam cells within the lesions.[3–5] Cells of this lineage can display remarkably plastic responses to their microenvironment,[6,7] and human atherosclerotic lesions contain macrophages with proinflammatory M1-like to proresolution M2-like profiles.[8–10] Although single-cell techniques have demonstrated heterogeneity in human plaque macrophage populations,[11–14] the lesions profiled are typically in the late stages of atherogenesis, when phenotypes may be polarized by long-established factors including modified cholesterols, cellular necrosis, sustained immune cell activation, and concomitant cytokine release.[4] The response of macrophages to cholesterol, its vectors, and modified derivatives is fundamental to atherosclerosis but has not been well defined at the single-cell level and may differ across macrophage subpopulations.[15,16] Deeper insight into the responses of different macrophage subpopulations to atherogenic lipid could help to identify new potential therapeutic targets.

CAD has a large heritable component,[17] and >250 genomic risk loci have been identified by large genome-wide association studies (GWAS).[18,19] However, the majority of these variants are found in noncoding regions of the genome, often in linkage disequilibrium with multiple other variants.[20,21] Despite considerable efforts,[22] the cellular populations and the molecular mechanisms through which these genetic loci transmit disease risk are largely unknown. Recent advances in single-cell techniques provide new opportunities for investigating the function of noncoding variants by identifying cis-regulatory interactions in cellular subpopulations.[23–26] This information can then be used to prioritize candidate causal genetic variants.[27–30] Although it has been

Downloaded from http://ahajournals.org by on June 25, 2024

established that several CAD risk variants act specifically through macrophages,[31] current single-cell studies have failed to address the genetic contribution of macrophages to the disease.[11,32,33]

The challenge in identifying such contributions is likely due to the limitations of using single-cell RNA-seq (scRNA-seq) or single-cell assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq) in isolation. The power of single-cell–transcriptomic-based approaches is heavily dependent on the accuracy of gene-enhancer mapping. Because these maps are typically constructed from bulk sequencing of immortalized cell lines or whole tissues, they do not fully capture gene regulation dynamics in disease contexts. Conversely, variants prioritized based on single-cell studies of chromatin accessibility alone have excess false positives, as only a subset of accessible chromatin sites have regulatory activity.[34] We reason that joint single-cell profiling of both the transcriptome and chromatin accessibility directly in primary cells will bridge the gap between previous studies that only utilized 1 of the 2 modalities, providing a more comprehensive understanding of both the cellular and genetic components of human diseases.

In the current study, we undertake comprehensive single-cell RNA and multiomics profiling of human macrophages to characterize their transcriptomic and epigenomic response to atherogenic oxidized low-density lipoprotein cholesterol (ox-LDL). We construct a high-resolution map of macrophage CRE (cis-regulatory elements), enabling us to prioritize candidate causal CAD risk variants and partition CAD heritability across macrophage subpopulations. We identify a novel CD52-hi lipid-handling macrophage subpopulation that conveys more heritable risks of CAD and demonstrate the reduced lipoprotein accumulation in these cells. In addition, we illustrate how this approach provides mechanistic insight into the cis-regulatory effect of the CAD-risk single nucleotide polymorphism (SNP) rs10488763 at the *FDX1/RDX* locus. Our results provide an unbiased framework for functional fine-mapping of CAD GWAS results through partitioning of heritability among macrophage subpopulations and deliver the first genetic evidence for the diverse and divergent influences of different human macrophage subpopulations on atherosclerosis and its heritability.

METHODS

Data Availability

All sequencing data supporting the findings in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive under accession number PRJNA1102756. Processed data have been made available at the Zenodo repository with record ID 11001642. Computational analysis codes for reproducing results and figures presented in this study are available at https://github.com/jhjiang2020/multiome_paper.

See the Supplemental Material and the Major Resources Table for a detailed description of materials and methods.

RESULTS

Single-Cell Sequencing Identifies Cellular Heterogeneity in Human Monocyte-Derived Macrophages

Monocytes adhere to the vessel wall at sites of atherosclerotic lesions and are the predominant source of plaque macrophages.[3,4,35] Primary human macrophages derived ex vivo from circulating monocytes are extensively used to study the molecular basis of atherosclerosis,[36–39] and transform into a lipid-laden foam cell phenotype when exposed to ox-LDL. We observed significant heterogeneity in cell morphology among these macrophages (Figure 1A), and they exhibited various lipid-accumulation and distribution patterns after ox-LDL treatment (Figure 1B). To characterize the transcriptomic heterogeneity in macrophages and its impact on ox-LDL response, we first performed single-cell RNA sequencing in both unexposed and ox-LDL–exposed human macrophages (Figure S1A). After removing low-quality cells, we retained 4881 cells comprising 7 transcriptomically distinct clusters (Figure 1C). Consistent with our previous findings,[37,38] exposure to ox-LDL significantly upregulated the expression of genes involved in lipid transport, such as *CD36*, *ABCA1*, and *ABCG1* (Figure S1B; Table S1).[40,41] Kyoto Encyclopedia of Genes and Genomes pathway analysis demonstrated that ox-LDL triggered lysosomal activation and inhibited cholesterol biosynthesis (Figure S1C).

To better understand the epigenetic regulation of macrophage gene expression, we undertook single-cell multiomics, simultaneously characterizing the transcriptome and open chromatin profile of each cell, using single-nucleus RNA-seq and ATAC-seq, respectively (Figure S1A). The single-nucleus RNA-seq profiles were highly concordant with the scRNA-seq profiles (Figure S1D), and integration with the ATAC-seq open chromatin profile for each cell improved the resolution of the allocation of macrophage subpopulations. With this improved analysis, we identified 9 macrophage subpopulations (Figure 1D; Figure S1E and S1F; Table S2), including a cluster with high expression of lipid–metabolism genes (*APOC1* and *LPL*) even without ox-LDL exposure; therefore, we termed these lipid-handling macrophages (Figure 1D and 1E; Table S2).

Other subpopulations included an M1-like population showing high expression of proinflammatory cytokines (*CCL2*, *ICAM1*, and *CCL3*), 2 proliferating populations (*TOP2A*, *CENPA*, and *CENPE*), and a small AK5$^+$ population (*ANKH*, *AK5*, and *MYO1B*) with some

Downloaded from http://ahajournals.org by on June 25, 2024

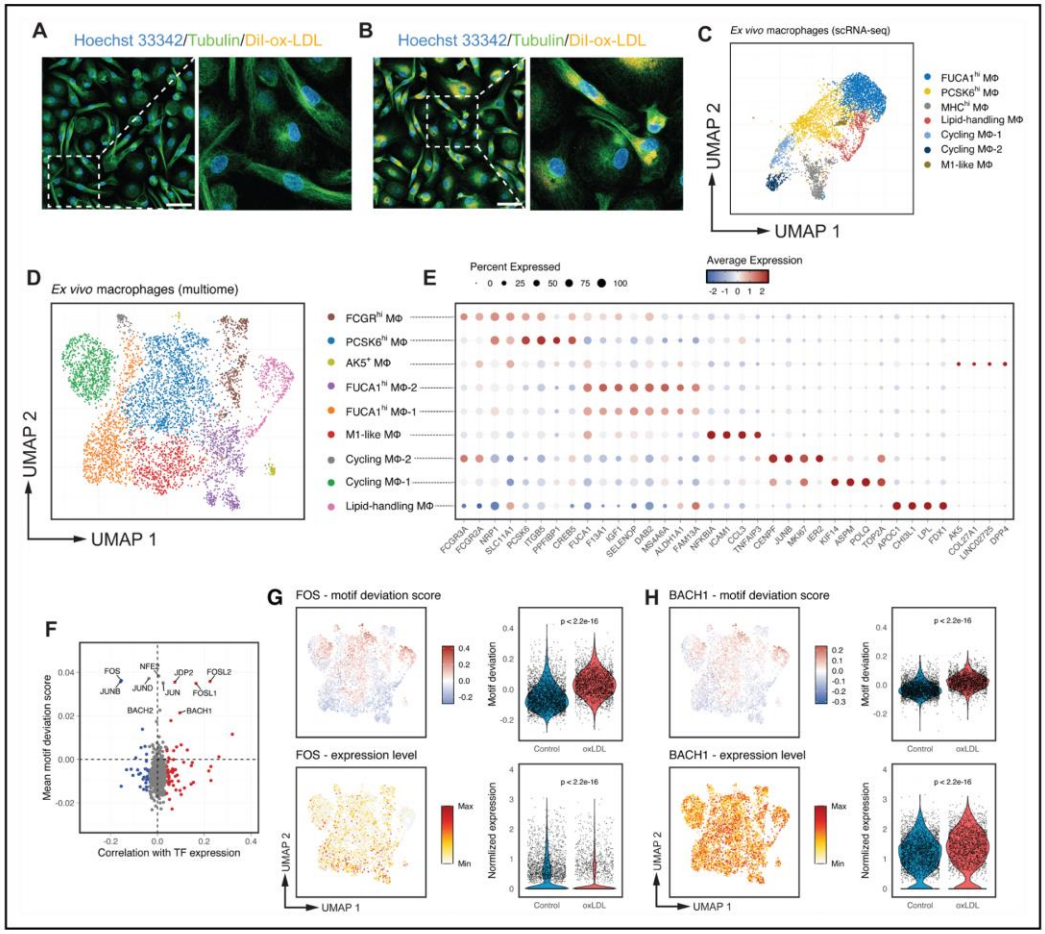


Figure 1. Multiomics single-cell sequencing identifies cellular heterogeneity in human monocyte-derived macrophages and captures transcriptional regulators driving their response to oxidized low-density lipoprotein (ox-LDL) cholesterol.
A and **B**, Representative immunostaining of unexposed (**A**) and lipid-laden ex vivo macrophages (**B**), showing variations in cellular morphology and lipid accumulation. Scale bar: 50 µm. **C**, Uniform manifold approximation and projection (UMAP) visualization showing transcriptomically distinct macrophage subpopulations identified by single-cell RNA-sequence (scRNA-seq; n=4881 cells). **D**, Refined ex vivo macrophage subpopulations based on integrated single-nucleus RNA-seq and assay for transposase-accessible chromatin with high throughput sequencing (ATAC-seq) profiles and displayed as UMAP. **E**, Dot plot of the expression of selected marker genes for the 9 subpopulations in **D**; for each subpopulation, the color represents the average gene expression, normalized by sequencing depth and scaled across all cells; the size represents the percentage of cells expressing the gene. **F**, Mean TF (transcription factor) motif deviation scores (*y* axis, relative enrichment compared with background) plotted against their correlation with TF RNA expression (*x* axis, Pearson coefficient) in ox-LDL–treated cells. TFs are colored by significance (FDR <0.05) for negative (blue) or positive (red) correlation. **G**, Motif deviation scores (**top**) and normalized gene expression (**bottom**) for the AP-1 family TF *FOS*, shown on the UMAP and summarized by ox-LDL exposure status in the violin plot. Gene expression levels were min-max normalized for visualization purposes. **H**, Motif deviation scores (**top**) and normalized gene expression (**bottom**) for NFE2 family TF *BACH1*. *P* values were calculated using a Wilcoxon test. SNP indicates single nucleotide polymorphism.

similarity in gene expression to osteoclasts.[42] Subpopulations analogous to classic resting M0 macrophages included a PCSK6^{hi} population (*PCSK6*, *ITGB5*, and *SHC3*), 2 FUCA1^{hi} populations (*FUCA1*, *WWP1*, and *SELENOP*) and a transitional monocyte/macrophage FCGR^{hi} population (*FCGR3A*, *FCGR2A*, and *PCSK6*; Figure 1D and 1E).

Single-Cell Multiomics Captures the Transcriptional Regulators Driving the Response to ox-LDL in Macrophages

We next assessed TF (transcription factor)-binding activity in individual cells by computing motif deviation scores, which represent motif-specific gain or loss of chromatin

Downloaded from http://ahajournals.org by on June 25, 2024

accessibility relative to the average cell profile (see Methods; Figure S1G). As TFs belonging to the same family usually share binding motifs, we examined the correlation between the motif deviation scores and the expression of each individual TF to understand the differential role of TFs within the same family. Using bulk approaches, we have previously identified enrichment of the binding motifs of AP-1 and NFE-2 factors in dynamic enhancer sites following exposure of human macrophages to ox-LDL.[38] Both AP-1 and NFE-2 factors respond to oxidative stress.[43,44] However, it is challenging to pinpoint the specific TFs involved in the ox-LDL response in macrophages, as they all share similar binding motifs. Here, with single-cell multiomics, we can now dissect the directional influence of individual AP-1 and NFE-2 family member TFs (Figure 1F). For example, the expression of 2 AP-1 family members *FOSL1* and *FOSL2* positively correlated with the chromatin accessibility of AP-1 motifs (Figure 1F), whereas another AP-1 factor *FOS* displayed a reverse correlation (Figure 1F and 1G). Expression of the NFE-2 family member *BACH1* was positively correlated with its motif binding (Figure 1F and 1H), but no significant correlation was observed for other NFE-2 family members.

Analysis of Cis-Regulatory Elements Identifies Complex CAD-Associated Transcriptional Regulatory Networks

Single-cell multiomics data allowed us to identify cis-regulatory interactions by correlating chromatin accessibility at noncoding peaks with expression of nearby genes (<500 kb).[26] We identified 18 782 unique CREs (cis-regulatory elements) spanning 6381 genes (Table S3). Of these CREs, 2642 were correlated with the expression of >1 gene, resulting in a total of 22 555 peak-gene pairs. The major locations of CREs were in introns (42.3%), promoters (24.4%), and intergenic regions (24.3%), with 42.6% of all CREs located >10 kb from any recognized transcription start site (Figure 2A).

To validate the predicted peak-gene interactions, we tested our macrophage CRE sets for enrichment against a wide range of functional annotations generated from public data sets. By incorporating blood expression quantitative trait loci (eQTL) data from GTEx[45] and eQTLGen,[46] we showed that eQTL SNPs were significantly overrepresented in the CREs (P=6.2×10^{-10}; Figure S2A), with at least 1 eQTL SNP in 41.2% of CREs. The CREs were also enriched for activity-by-contact (ABC) super enhancers identified from enhancer-promoter contacts in mononuclear phagocytes[47] (Figure S2B) and for genomic regions shown to interact with macrophage promoters by promoter capture Hi-C[48] (Figure S2C).

In addition, we quantified the active enhancer signals over all open chromatin peaks using an H3K27ac ChIP-seq data set we generated in bulk macrophages.[38] The average H3K27ac signal in CRE peaks was significantly higher than in non-CRE peaks (P<2.2×10^{-16}) and exposure to ox-LDL increased the H3K27ac signals in CRE peaks (P=6.458×10^{-6}) but not in non-CRE peaks (Figure 2B; Figure S2D).

We next applied the FigR framework[49] to the multiomics data to understand the gene regulation dynamics in macrophages. To further reduce the background noise, we limited the downstream correlation analysis to a subset of 636 genes with a high density of CREs (≥9 CREs per gene; Figure S2E). These genes and their associated CREs constitute domains of regulatory chromatin (DORCs). Of these DORC genes, 529 (83.2%) were regulated by strong mononuclear cell ABC enhancers (ABC score >0.1; Figure S2F).[47] Upon ox-LDL exposure, changes in chromatin accessibility at DORC CREs closely correlated with changes in the expression of DORC genes (R=0.62; P=2.1×10^{-68}; Figure 2C).

For each DORC gene, we assessed their interactions with all detected TFs by 2 metrics: TF motif enrichment in CRE peaks, and the correlation between TF expression and CRE chromatin accessibility (Figure 2D). In total, we analyzed 588 300 TF-DORC pairs representing 925 TFs and 636 DORCs and identified 1612 strong interactions (|regulatory score| >1.5) involving 95 TFs and 526 DORCs (Table S4). The top 0.1% of TF regulators and their target DORCs with the maximum regulatory scores >2 are plotted in Figure 2E. Notably, several TFs, primarily in the AP-1 family, display opposite regulatory effects across different DORC genes (Figure 2E). For example, JDP2 has an activating effect on *CD109* but negatively regulates *PLPP3* (marked with arrows in Figure 2E).

Different DORCs that share similar TF associations constitute integrated gene regulatory networks. We collated all strong TF-DORC interactions (|regulatory score| >1.5) with CAD DORCs (DORC with CREs overlapping CAD-associated variants) and constructed a TF regulatory network for CAD in macrophages (Figure 2F). This network centered around members of the AP-1 family, with *FOS* and *JUNB* acting as repressors and *FOSL1*, *ATF4*, and *JUND* as activators. Other TF families, including CEBP and KLF, also targeted multiple CAD DORCs. The TF-DORC relationships can be complex; for example, one of the CAD DORC genes regulated by AP-1 factors includes *MITF*, which is itself a TF regulator and activates a wide range of downstream DORCs (Figure 2F).

Macrophage CRE Map Prioritizes 121 CAD-Associated Variants Targeting 56 Genes

Noncoding disease-associated genetic variants often influence disease risk by altering the gene regulatory properties of CRE.[50] The integration of CREs into the analysis of CAD GWAS results allows candidate causal variants to be linked to their target genes. Permutation-based testing demonstrated that macrophage CREs

Downloaded from http://ahajournals.org by on June 25, 2024

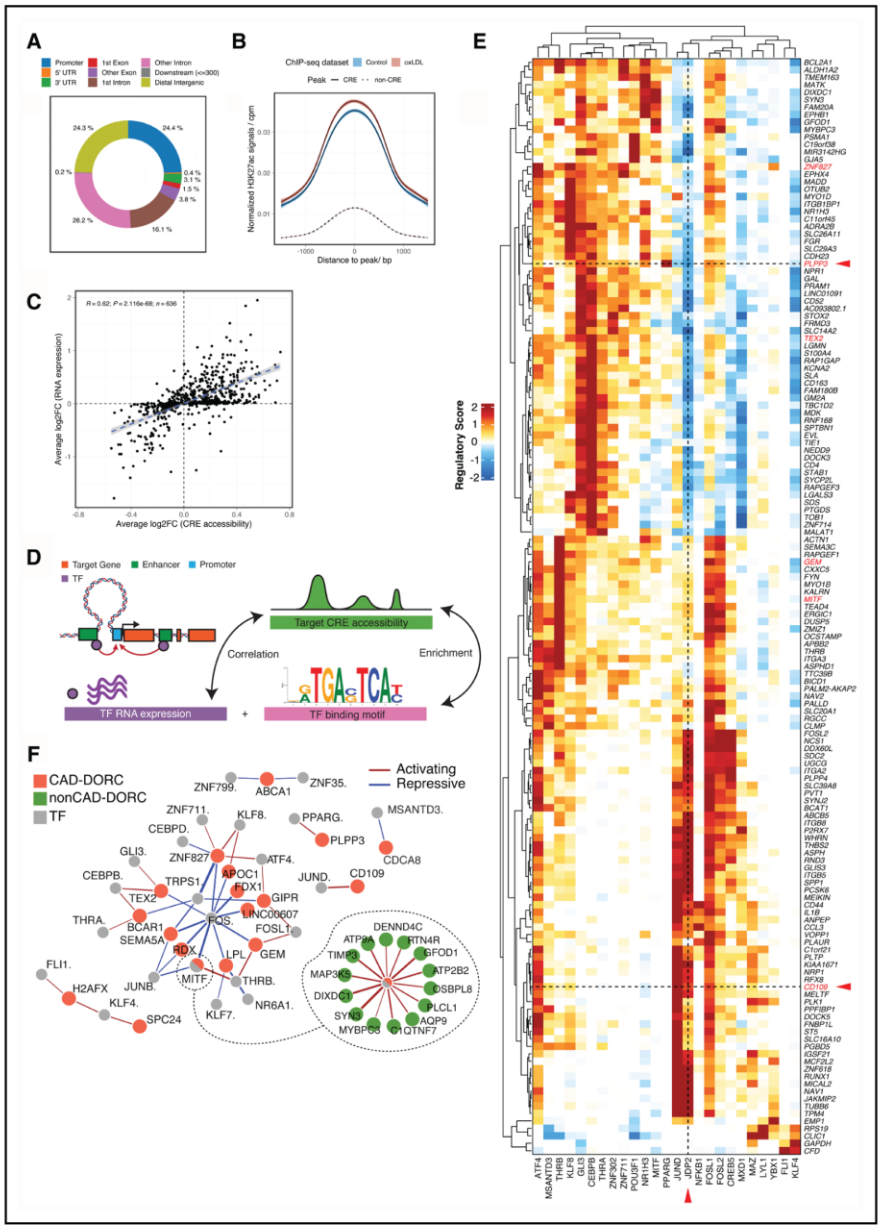


Figure 2. Multiomics analysis identifies cis-regulatory elements and key transcriptional regulatory networks for coronary artery disease (**CAD).**

A, Location of CREs (cis-regulatory elements; n=22 555 CREs). **B**, Aggregated H3K27ac signals at CRE-peak and non-CRE-peak regions. Signals from unexposed (blue) and oxidized low-density lipoprotein cholesterol (ox-LDL)–exposed (red) human macrophages are plotted separately. Shaded regions indicate 95% CIs. **C**, Pearson correlation between ox-LDL–induced changes in CRE chromatin accessibility and in gene expression. n=636 DORC genes analyzed. **D**, Schematic diagram of the workflow for identification of transcription regulatory networks.[49] The significance of each TF-gene interaction is tested by assessing whether the TF motif is enriched in the target CRE and whether the expression of the TF correlates with the accessibility of the target CRE (see Methods). **E**, Heatmap of regulation scores for the top 0.1% of TF (transcription factor)-DORC enrichments (n=25 TFs and n=147 DORCs). DORC genes are in rows and TFs are in columns. DORCs (*Continued*)

Downloaded from http://ahajournals.org by on June 25, 2024

were enriched for CAD-associated variants—these CREs overlapped significantly more CAD-associated variants than comparable background control peaks did (P=0.00548; Figure S3A). The subset of CREs associated with ox-LDL–regulated genes (ox-LDL–response CREs) was also enriched for CAD-associated variants (P=0.004455; Figure 3A). We extended GWAS significant SNPs (adjusted $P<5\times10^{-8}$) to full summary statistics[18] using stratified linkage disequilibrium score regression[51] and showed that this enrichment was specific for CAD compared with nonrelevant traits (Figure S3B). These results strongly support the hypothesis that macrophages and their response to ox-LDL are causally involved in the heritability of CAD.

Of the 9516 variants in high linkage disequilibrium ($r^2>0.8$ in the 1000G EUR panel) with the lead CAD variants in the NHGRI-EBI catalog, 121 SNPs were located within macrophage CREs and we linked these to 56 cis-regulated target genes (154 SNP-gene pairs; Table S5). In 101 pairs (71.4%), the SNPs and cis-regulated genes were >10 kb apart (median distance, 42.9 kb), and the cis-regulated gene was the nearest gene to the SNP only in 48 SNP-gene pairs (34.3%). These results demonstrate the limitations of traditional approaches, which prioritize variants by proximity. Furthermore, although 110 of the SNPs (90.9%) within macrophage CREs were eQTL SNPs,[45,46] only 34.3% (53 of the 154 SNP-gene pairs) had the same target gene for both the CRE and the eQTL effect, suggesting that eQTL analysis alone will not identify the regulatory effects of many risk variants.

We benchmarked our macrophage CRE map against state-of-the-art ABC-enhancer mapping based on promoter-enhancer contact information from 35 public mononuclear cell data sets.[47] Although macrophages are known for their involvement in atherosclerosis,[5] we found no significant enrichment for CAD risk variants after annotating macrophage open chromatin regions with the ABC-enhancer map (61 950 peaks, targeting 19 969 genes; Figure S3C). We also found no significant enrichment for CAD risk variants in the subset of 9153 ABC-enhancer peaks targeting ox-LDL–response genes (Figure S3D). These results suggest that ABC-mapped mononuclear enhancers are limited in their power to prioritize causal CAD variants from background noise in primary macrophages.

Together, these findings demonstrate the value of using single-cell multiomics to generate a CRE map for the prioritization and interpretation of CAD risk variants. This focused approach provides greatly enhanced power and precision for investigating macrophage-specific CAD variants compared with traditional approaches.

CAD Heritability Varies in Macrophage Subpopulations With Enrichment in Lipid-Handling Macrophages

As demonstrated above (Figure 1C through 1E), macrophage subpopulations differ in their transcriptomic and epigenetic profiles. We, therefore, reasoned that this would make them differ in the extent to which they are functionally influenced by CAD–risk variants and so the nature of their contribution to CAD heritability. More specifically, we hypothesized that the CAD heritability that is enriched in macrophage CREs (Figure S3A and S3B) could be further stratified to evaluate the genetic contribution of each macrophage subpopulation to the disease. In accordance with this hypothesis, of the 44 loci at which CAD SNPs overlap CREs, 29 of these loci overlap CREs whose target genes are only expressed in one or a few macrophage subpopulations (Figure 3B), indicating that their disease risk-modifying influence is likely to be specific to certain subpopulations. Because assayed DNA variants with genome-wide significance only explain a small portion of the genetic variance of CAD,[52] we again extended the significant variants to all SNPs included in the GWAS summary statistics. For each macrophage subpopulation, we applied stratified linkage disequilibrium score regression to systematically assess the enrichment for genome-wide CAD heritability, rather than the enrichment for individual risk variants (Figure 3C).

In principle, the genetic contribution of macrophages can be divided into subpopulations based on their gene expression patterns, and the gene expression patterns reflect both the baseline and the ox-LDL-response signatures in the transcriptome. Using our previously published Pi strategy,[53] we determined a series of weighted gene sets (gene programs) by scoring differential gene expression statistics for each subpopulation. A gene program characterizes either the baseline expression of a subpopulation (termed a subpopulation gene program) or the response of a subpopulation to ox-LDL (termed a subpopulation–ox-LDL gene program). As a comparator, we pooled all subpopulations and constructed a pseudo-bulk ox-LDL program characterizing the response of the whole bulk macrophage population to ox-LDL.

The gene programs were then converted into weighted genomic regions using our macrophage CRE map; all overlapping 1000G SNPs were assigned the

Figure 2 Continued. with CREs overlapping CAD-associated variants are marked with red. Positive and negative regulatory effects are shown in red and blue, respectively. Red arrows highlight the opposite regulatory effect of JDP2 on *CD109* and *PLPP3*. **F**, Network visualization of significant TF-DORC pairs for CAD-DORCs (absolute regulation score >1.5). Nodes represent CAD-DORCs (orange) and TF regulators (grey). Non-CAD-DORCs regulated by MITF are plotted separately (green discs). Red lines indicate activating effects, and blue lines indicate repressive effects. Line widths are weighted by the regulation score.

Downloaded from http://ahajournals.org by on June 25, 2024

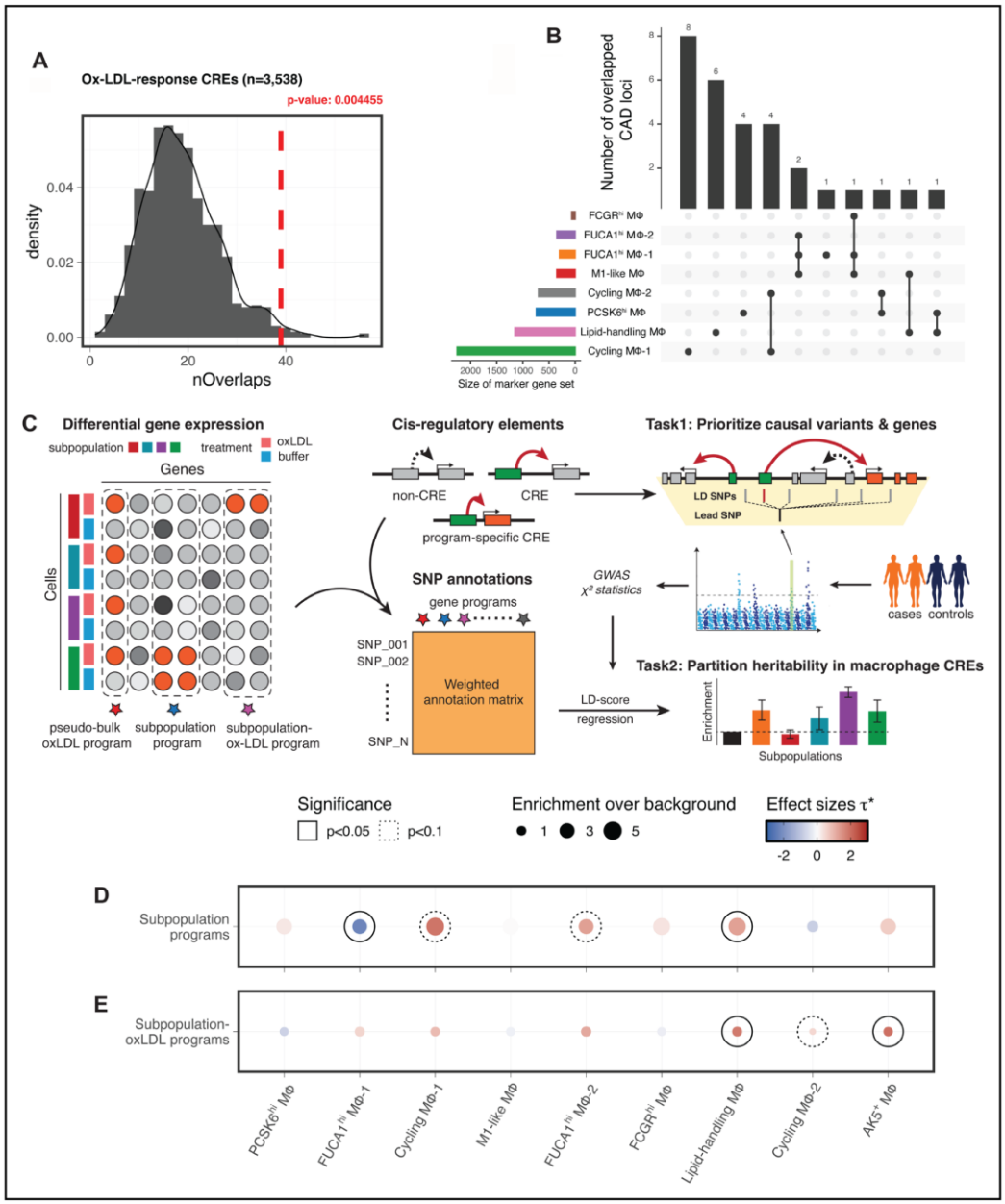


Figure 3. Partitioning the genetic risk of coronary artery disease (CAD) prioritizes disease-critical macrophage subpopulations. **A**, Histogram showing the null distribution of overlaps between CAD-associated SNPs and n=1000 sets of background peaks matched for GC% bias and average accessibility for all oxidized low-density lipoprotein cholesterol (ox-LDL)–response CREs (cis-regulatory elements). The observed overlap is marked by the red dashed line. **B**, Number of independent CAD loci mapped onto CREs targeting subpopulation-specific genes. Each set of subpopulation marker genes was binarily selected by a differential gene expression test using an FDR threshold of 0.05. Rows represent marker gene sets, and intersections between different sets are shown as connected dots. The size of each marker gene set is plotted to the left of the matrix, and the size of each intersection is plotted on the top. **C**, Genome-wide association studies (GWAS) heritability enrichment framework. **Left**, Deriving macrophage subpopulation gene programs, pseudo-bulk ox-LDL gene program, and (*Continued*)

Downloaded from http://ahajournals.org by on June 25, 2024

corresponding weight for each region to construct the SNP annotation matrices required for heritability stratification. Heritability enrichment was defined as the proportion of SNP heritability explained divided by the proportion of SNPs included in each annotation matrix. As subpopulation-specific annotation matrices were weighted on a continuous scale, we also computed the standardized effect sizes τ^*, defined as the proportional change in per-SNP heritability with a 1-SD increase in the SNP annotation weight, to assess the strength and statistical significance of heritability enrichment (see Methods).[27,54]

Compared with the background annotation constructed using all macrophage open chromatin regions, subpopulation gene programs show higher enrichment for CAD heritability (Figure 3D; Figure S3E). This is expected as regulatory elements have been widely reported to be enriched for disease and complex trait heritability.[51] To test the independent contribution of each subpopulation gene program to the overall enrichment for CAD heritability, we further controlled subpopulation gene programs on each other using a multivariate linear regression model. The results showed that CAD heritability was relatively depleted (τ^*=−2.24) in FUCA1^{hi} macrophages and remained significantly enriched (τ^*=1.28) in lipid-handling macrophages (Figure 3D). This suggests that the baseline expression patterns in lipid-handling macrophages convey higher genetic risks for CAD with respect to other macrophage subpopulations.

Although different macrophage subpopulations had distinct subpopulation gene programs (Figure S3F), their subpopulation–ox-LDL gene programs were highly correlated, indicating a shared overall ox-LDL–response across macrophage subpopulations (Figure S3G). Therefore, instead of conditioning on each other, we used the pseudo-bulk ox-LDL gene program as the reference to compute the standardized effect size τ^* for each subpopulation–ox-LDL program. For most subpopulations, τ^* no longer deviated from zero (Figure 3E), meaning that the ox-LDL effect in these subpopulations is largely not subpopulation specific. However, there was still significant enrichment of heritability in the ox-LDL response in lipid-handling and AK5$^+$ macrophages, with τ^* of 1.81 and 1.97, respectively (Figure 3E).

Overall, our heritability analysis demonstrated that lipid-handling macrophages contribute more to the heritable component of CAD compared with other macrophage subpopulations, both in unexposed state and upon exposure to ox-LDL.

Lipid-Handling Macrophages Display Activation of Genes and Transcription Regulators Involved in Lipid Processing and Clearance

Having identified a dominant role for lipid-handling macrophages in the heritability of CAD, we sought to further characterize this macrophage subpopulation. Both scRNA-seq and single-nucleus RNA-seq data revealed a significantly increased expression of apolipoprotein *APOC1* and lipoprotein lipase *LPL* in these cells (Figure 4A; Figure S4A), both of which are involved in lipoprotein metabolism.[55] Kyoto Encyclopedia of Genes and Genomes and Gene Ontology Biological Process enrichment analyses suggested an enhancement in phagocytosis (Figure 4B) and fatty acid metabolism (Figure S4B) within these cells. The membrane glycoprotein *CD52* and nonenzymatic chitinase-3 like-protein-1 *CHI3L1* were also uniquely expressed in lipid-handling macrophages (Figure 4A; Figure S4A), while their roles in macrophage lipid metabolism, if any, are uncharacterized.

Notably, the marker genes of lipid-handling macrophages (133 genes with FDR <0.05 and log2FC >1; Table S1) were significantly overrepresented in a gene regulatory network associated with subcutaneous fat and liver (GRN51, FDR=1.42×10^{-20}; Figure S4C).[56] The expression of this network strongly correlates with body mass index and blood triglyceride levels in a CAD patient cohort (Figure S4D).

We then identified the transcriptional regulators specific to lipid-handling macrophages. CEBP family TFs showed the largest motif deviation score for these macrophages (Figure 4C), with *CEBPB* expression correlating positively with CEBP–motif binding activity (Figure 4C and 4D). A strong correlation was also observed between expression and motif usage for *NR1H3*, a key transcriptional regulator of genes involved in fatty acid metabolism (Figure 4C and 4E).[57]

To functionally validate these findings, we first used immunofluorescence to confirm key marker gene expression and observed coexpression of *CD52*, *CHI3L1*, and the well-documented ox-LDL uptake receptor *CD36*[58] in unexposed macrophages (Figure 4F and 4G; Figure S4E and S4F). Next, we performed a lipid uptake assay using DiI-labeled ox-LDL, which demonstrated reduced

Figure 3 Continued. subpopulation-ox-LDL gene programs from single-cell transcriptomic data. **Middle**, Converting gene programs into weighted genomic regions using the macrophage CRE map (**top**), then constructing single nucleotide polymorphism (SNP) annotation matrices by assigning the corresponding weights to 1000G SNPs overlapping each region (**bottom**). **Right**, After accounting for the linkage disequilibrium (LD) structure within GWAS risk loci, the putative causal variants and their target genes are prioritized by their overlaps with macrophage CREs (task 1); the relative enrichment of GWAS heritability for stratified subpopulation-specific/subpopulation-specific–ox-LDL–specific annotations are tested using the LD score regression model[51] (task 2). **D** and **E**, Enrichments for CAD heritability in subpopulation gene programs (**D**) and subpopulation-ox-LDL gene programs (**E**). Dot size denotes the relative enrichment over background annotations. Color denotes the standardized effect sizes (see text) after conditioning on each other (**D**) or on a pseudo-bulk ox-LDL–response program (**E**) Nonzero τ^* estimates reported at 2 significance levels (solid: $P<0.05$, dotted: $P<0.1$).

Downloaded from http://ahajournals.org by on June 25, 2024

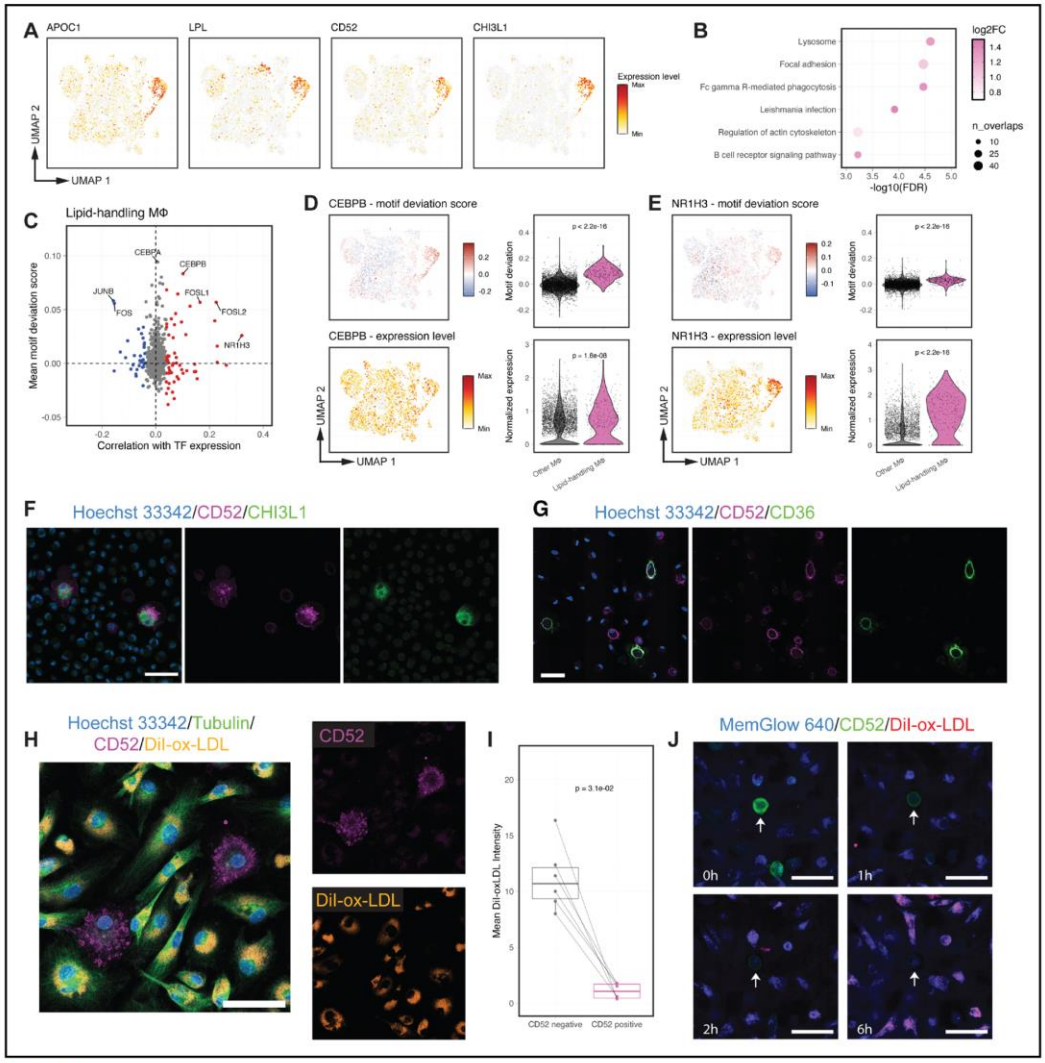


Figure 4. Multiomics characterization of the lipid-handling macrophage subpopulation.
A, The expression of selected marker genes displayed on the uniform manifold approximation and projection (UMAP). Expression levels were min-max normalized for visualization purposes. **B**, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in the marker genes of lipid-handling macrophages. Color represents log2-fold change and size represents the number of overlapped genes in each pathway. **C**, Mean TF (transcription factor) motif deviation scores (*y* axis, the relative enrichment compared with background) compared with their correlation with TF RNA expression (*x* axis, Pearson coefficient) in lipid-handling macrophages. TFs are colored by significance (FDR<0.05) for negative (blue) or positive (red) correlation. **D**, Motif deviation scores (**top**) and RNA expression (bottom) for *CEBPB*. **E**, Motif deviation scores (**top**) and RNA expression (**bottom**) for *NR1H3*. **F** and **G**, Representative immunostaining of ex vivo macrophages for lipid-handling macrophage markers CD52, CHI3L1 (**F**), and CD36 (**G**). **H**, Representative immunostaining of ex vivo macrophages for lipid accumulation after 48 hours treatment with DiI-oxidized low-density lipoprotein cholesterol (ox-LDL). **I**, Quantification of the intracellular level of DiI-ox-LDL in **H**, shown as average fluorescence intensities (arbitrary units) per biological replicate (n=6). **J**, Representative live-cell imaging of ex vivo macrophages for lipid accumulation at different time points post-DiI-ox-LDL exposure, with a CD52⁺ cell marked by the white arrow. Cells were briefly incubated with Fc-muted CD52 antibody for 1 hour before the assay to minimize the effect of antibody binding. Scale bar: 50 μm. *P* values were calculated using a Wilcoxon test in **D** and **E**, and a paired Wilcoxon test in **I**.

Downloaded from http://ahajournals.org by on June 25, 2024

accumulation of intracellular ox-LDL in CD52$^+$ macrophages compared with CD52$^-$ macrophages (Figure 4H and 4I). This reduced lipid accumulation was further supported by live-cell imaging, as early as 2 hours post ox-LDL exposure (Figure 4J; Video S1). After ox-LDL treatment, lipid-handling macrophages exhibited increased expression of *CD36* (Figure S4G), a trend also seen in other subpopulations. The upregulation of *CD36* even in the unexposed state would be expected to increase ox-LDL uptake, so the reduced ox-LDL accumulation observed is consistent with CD52$^+$ lipid-handling macrophages having more efficient lipid/cholesterol processing and clearance capacities.

Meta-Analysis of 38 Human Atherosclerotic Samples Enables High-Resolution Mapping of Macrophages in Plaques

Most macrophages in atherosclerotic lesions are derived from circulating monocytes[3,35,59] as were the macrophages we profiled; as plaque macrophages are exposed to multiple atherogenic influences, we sought to determine whether the CD52-hi lipid-handling macrophage subpopulation is a reproducible subpopulation in human atherosclerotic plaque.

We started with a meta-analysis of scRNA-seq data from 38 human coronary and carotid artery plaque samples from 5 independent studies[12–14,60,61] (Figure S5; Supplemental Notes). After read alignment, sample QC, and doublet removal, we retained a total of 126 685 cells across 8 broad cell types (Figure S6A and S6B). We then focused on myeloid cells (n=20 458) based on expression of *CD14*, *CD68*, and *AIF1*.

Next, we reintegrated the scRNA-seq data to replicate the macrophage subpopulations reported in previous studies.[61–63] In total, we identified 6 heterogeneous macrophage clusters (Figure 5A and 5B; Figure S6C) including 2 S100A8^{hi} clusters S100A8/S100A12$^+$ and S100A8/S100A12$^-$ macrophages, an IFNG-activated cluster, an IGSF21$^+$ cluster and 2 TREM2$^+$ clusters expressing marker genes associated with lipid metabolism (*SPP1*, *CD9*, and *FABP5*). We termed these 2 clusters lipid-associated macrophages (LAMs) and PLIN2^{hi}/TREM1^{hi} macrophages for consistency with previous reports.[61,64] We also identified 3 previously reported clusters for dendritic cells, termed cDC1, cDC2, and mature cDC2.[61]

To investigate how monocyte-derived macrophages map to in vivo plaque macrophages, we conducted an integrative analysis using scRNA-seq profiles from our ex vivo macrophages and our plaque macrophage atlas. To our surprise, most ex vivo macrophages, both unexposed and exposed to ox-LDL, mapped closely to the TREM2$^+$ LAM population in plaques (Figure 5C). This transcriptomic similarity was further supported by correlation analysis of the pseudo-bulk profiles (Figure S6D). In addition, we trained a machine-learning macrophage classifier with a deep generative model[65] on the plaque macrophage atlas, and almost all ex vivo macrophages were predicted to be LAM with high confidence (Figure 5D). Importantly, we also expanded our heritability analysis to plaque macrophage populations and found that LAM, like ex vivo macrophages, showed a strong enrichment for CAD heritability (Figure 5E).

CD52-hi Subpopulation Expressing High Levels of Lipid-Processing Genes Is Also Found Among Plaque Lipid-Associated Macrophages

We have demonstrated the transcriptomic, epigenomic, and functional heterogeneity among ex vivo macrophages. Because they robustly mapped to the LAM population in plaques, we hypothesized that plaque LAMs would also be heterogeneous with functionally distinct subpopulations. We anticipated that the increased sample size of our plaque macrophage atlas would allow us to discern macrophage heterogeneity at a much higher resolution than previously achieved and so discover rare populations not characterized in previous studies.

To ensure the robustness and reproducibility of newly identified macrophage subpopulations, we performed cluster stability analysis, testing a wide range of clustering parameters after correcting for batch effects (Figure S5; Methods and Supplemental Notes). As a result, we identified 4 transcriptomically distinct subpopulations within plaque LAMs (n=3645; Figure 5F; Figure S6E) in both coronary and carotid artery samples (Figure S6F). These LAM subpopulations broadly shared the gene expression patterns we had identified in various ex vivo subpopulations (Figure S6G and S6H). Notably, a subset of CD52-hi LAM displayed strong expression for genes involved in lipid processing and metabolism (*LPL*, *CD36*, and *PPARG*; Figure 5G; Figure S6G). Although CD52 was expressed in >1 plaque macrophage subpopulation, the coexpression of *CD52*, *CHI3L1*, and *LPL* was only observed in TREM2$^+$ LAMs (Figure S6I). This is consistent with the lipid-handling phenotype we identified in ex vivo macrophages being a unique feature specific to LAMs.

Integrative Analysis Reveals the Regulatory Mechanisms of CAD Risk Loci Targeting *MITF* and *LPL*

Given the finding of plaque macrophages with the lipid-handling phenotype and the strong enrichment for CAD heritability in lipid-handling macrophage gene programs, we next sought to investigate the CAD risk variants affecting this subpopulation and their regulatory mechanisms. Among the variants conveying CAD heritability, 19 variants (at 6 independent risk loci) with genome-wide significance were mapped to the marker genes of lipid-handling macrophages (Figure 3B). These risk variants

Downloaded from http://ahajournals.org by on June 25, 2024

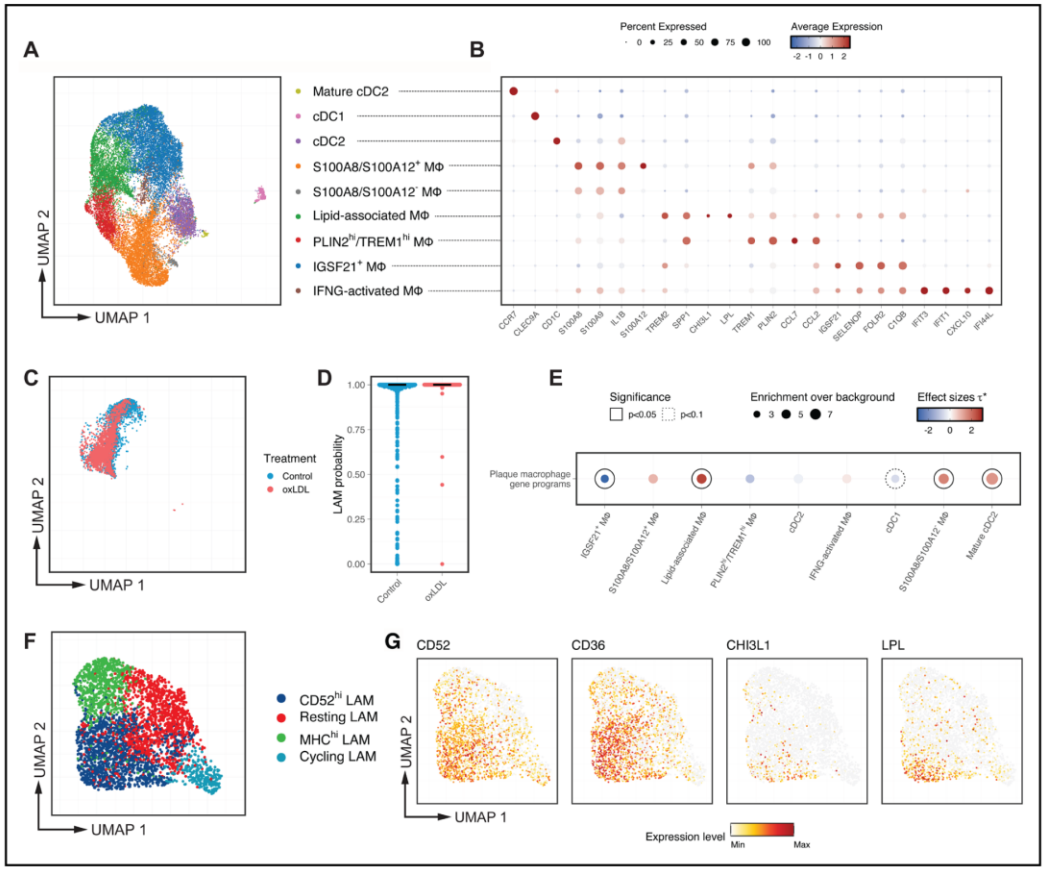


Figure 5. Meta-analysis of 38 human plaque samples enables high-resolution mapping of plaque macrophages in vivo. **A**, Uniform manifold approximation and projection (UMAP) visualization of the plaque macrophage atlas: multiple subpopulations of plaque macrophages demonstrated by meta-analysis of 38 samples from 5 independent studies.[12–14,60,61] **B**, The expression of selected marker genes for subpopulations in **A** is summarized in the dot plot, where the color represents the average gene expression, normalized by sequencing depth and scaled across all cells; the dot size represents the percentage of cells expressing the gene in the subpopulation. **C**, Single-cell transcriptomic profiles of ex vivo macrophages projected onto the plaque macrophage atlas (**A**) colored by treatment, demonstrating close mapping to the lipid-associated macrophage (LAM) subpopulation in plaques. **D**, Predicted probabilities of ex vivo macrophages being classified as plaque LAM, shown as individual data points along with the median, colored by treatment. **E**, Enrichments for coronary artery disease (CAD) heritability in subpopulation gene programs of plaque macrophages. Dot size denotes the relative enrichment over all macrophage open chromatin regions. Color denotes the standardized effect sizes after conditioning on each other. Nonzero τ^* estimates reported at 2 significance levels (solid: $P<0.05$, dotted: $P<0.1$). **F**, UMAP visualization of plaque LAMs in **A** but clustered at a higher resolution. **G**, The expression of selected lipid-handling macrophage markers displayed on the UMAP in **F**. Expression levels were minimum-maximum normalized for visualization purposes.

were present at CREs targeting *MITF*, *LPL*, *FDX1/RDX*, *BCAR1*, *TEX2*, and *PLPP3*. We have previously characterized the regulatory mechanism of one such variant in the *PLPP3* gene based on studies of bulk macrophage populations,[38] providing important orthogonal independent validation of the current approach.

MITF is a key TF influencing lysosome function and autophagy.[66,67] An intronic variant rs12714757 has been associated with CAD risk in both European and East Asian populations (β=0.033 and 0.048, respectively).[18,68] In our analysis, 2 closely linked SNPs rs6772383 and rs11452399 overlapped an intronic CRE of *MITF* (Figure 6A). At rs6772383, the risk T allele disrupts a TEAD1-binding site and is associated with decreased *MITF* expression in GTEx whole blood eQTL data (Figure 6A and 6B; Figure S7A), indicating that *MITF* expression is protective against atherosclerosis. Exposure of cells to ox-LDL resulted in a moderate increase in enhancer histone signals flanking this CRE (Figure 6A) and highly significant increases in both the chromatin accessibility

Downloaded from http://ahajournals.org by on June 25, 2024

ORIGINAL RESEARCH

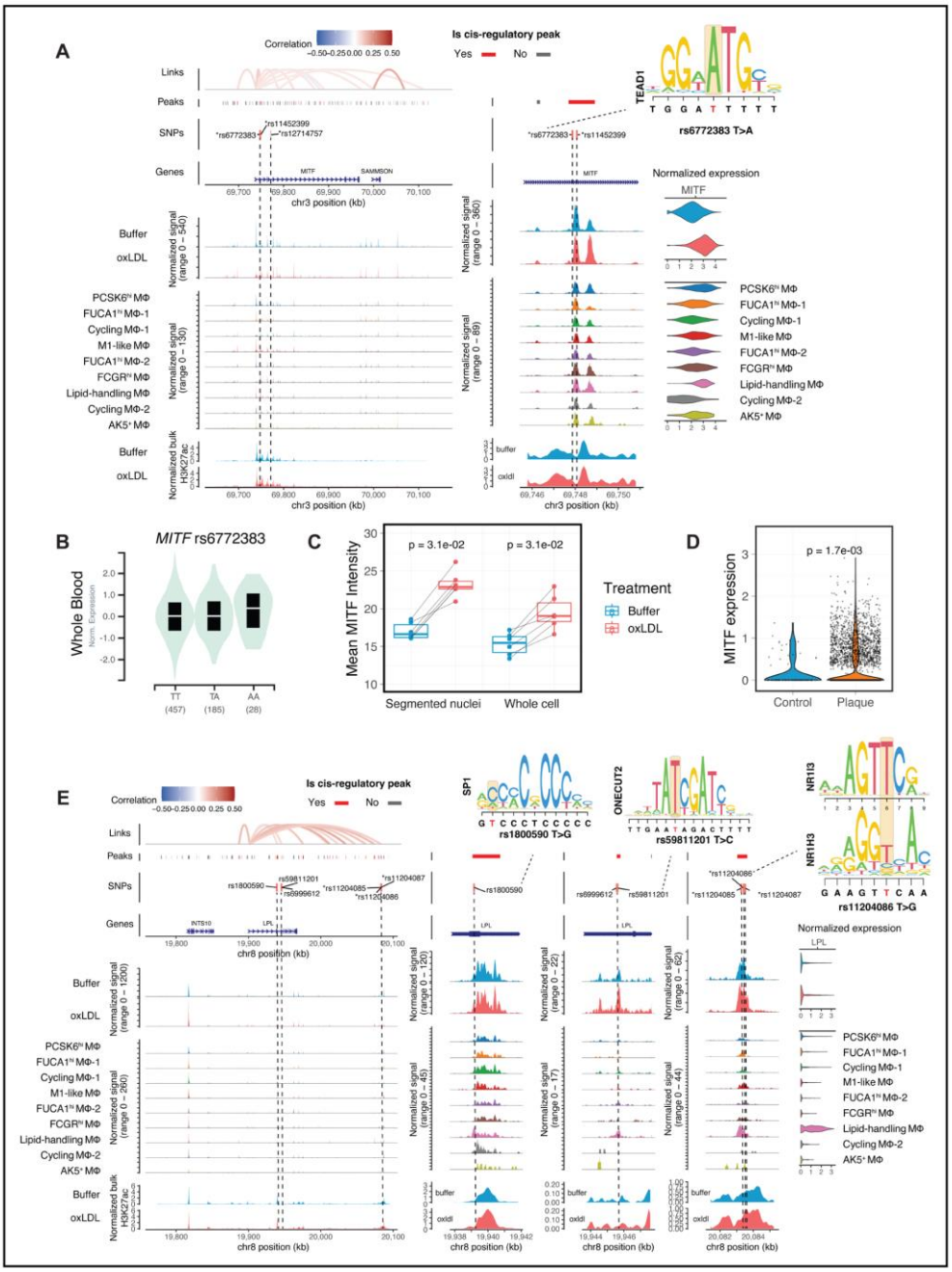


Figure 6. Chromatin regulatory landscapes of coronary artery disease (CAD) risk variants in lipid-handling macrophages.
A, Genomic plots for the region encoding CAD risk variants rs6772383 and rs11452399, which localize to a CRE (cis-regulatory element) at the *MITF* locus. Tracks (from **top** to **bottom**) showing cis-regulatory interactions (links), ATAC-seq peaks (peaks), CAD-associated variants (SNPs), genomic annotations (genes), ATAC-seq signals aggregated by treatment, ATAC-seq signals aggregated by subpopulations, enhancer signals from H3K27ac ChIP-seq. Higher resolution coverage plots centered on the cis-regulatory peak are plotted in the **middle**. (*Continued*)

Downloaded from http://ahajournals.org by on June 25, 2024

of this CRE (P=3.3×10^{-8}) and *MITF* expression (average log2FC=0.39; P<2.2×10^{-16}). Ox-LDL–induced expression of *MITF* was confirmed at the protein level using immunofluorescence staining (Figure 6C; Figure S7B), and a similar trend was observed using Western blot analysis (Figure S7C and S7D). In plaque macrophages, *MITF* was predominantly expressed in the LAM subpopulation (Figure S7E), and its expression was upregulated in LAMs from disease tissue as compared with adjacent control tissue (Figure 6D).

The *LPL* gene encodes lipoprotein lipase, which plays a role in lipoprotein uptake and is one of the strongest marker genes of lipid-handling macrophages (Figure 4A); its expression is increased by ox-LDL exposure (average log2FC=0.32; P=3×10^{-4}). We identified 6 CAD risk variants in 3 different CREs at the *LPL* locus (Figure 6E): rs1800590 in a 5′ untranslated region (UTR) CRE, rs6999612 and rs59811201 in an intronic CRE, and rs11204085/6/7 clustered in a distal intergenic CRE (β=0.091, 0.126, and 0.028 in European population for these 3 loci, respectively). Although no significant accessibility change was found in any of the 3 CREs with ox-LDL exposure (likely because they were already highly accessible), enhancer histone signals increased 70.9% in the CRE containing rs11204086 and 43.0% in CRE containing rs1800590. The protective G allele of rs11204086 disrupts a binding motif for NR1H3 (Figure 6E) and is associated with higher blood expression of *LPL* in eQTL data (Figure S7F). NR1H3 itself is activated notably in lipid-handling macrophages (Figure 4E) indicating that the cis-regulatory effect of rs11204086 may be mediated by a subpopulation-specific transcriptional network.

CAD-Associated Variant rs10488763 Alters Chromatin Structure at an AP-1 Site and Regulation of *FDX1* Expression by CEBPB

From the CRE-based variant prioritization analysis, we identified 2 eQTL SNPs (rs10488763 and rs1443120) in tight linkage disequilibrium (r^2=1 in EUR panel) that overlap a distal CRE for *FDX1* and *RDX*. This CRE is 56 and 77 kb upstream of the transcriptional start sites of *FDX1* and *RDX*, respectively (Figure 7A). The alternative alleles of these 2 SNPs were associated with higher CAD risk (β=0.036)[18,68] and lower expression of both *FDX1* and *RDX* in whole blood (Figure S8A). However, CAD summary statistics and eQTL mapping suggest that the GWAS signal aligns most clearly with the eQTL signal of *FDX1* rather than of *RDX* (Figure 7B; Figure S8B). We also found elevated *FDX1* in plaque LAMs (P=7.7×10^{-5}; Figure S8C) compared with non-plaque tissue LAMs, while *RDX* expression remained unchanged (Figure S8D). These results indicate that *FDX1* is the likely causal gene influenced by one or both CAD variants rs10488763 and rs1443120 at this locus.

As the risk T allele of rs10488763 perturbs the binding motif for AP-1 factors (Figure 7A), we hypothesized that this SNP exerts a cis-regulatory effect by modulating local chromatin structure. The rs10488763 SNP has been linked to a chromatin accessibility QTL locus (caQTL) in iPSC-derived macrophages in an infection model[69] (Figure 7B; Figure S8B). Consistent with this caQTL signal, we found higher chromatin accessibility in ox-LDL–treated macrophages that are homozygous for the protective A allele, compared with those that are heterozygous for this allele (P=1.8×10^{-4}; Figure 7C). We further compared the single-nucleus ATAC-seq (snATAC-seq) signal of rs10488763 in heterozygous donors and found that on the chromosome containing the protective A allele, this site was more accessible in both ex vivo macrophages and in human plaque cells[70] (P=0.041; Figure 7D).

In addition to the chromatin modulating effect of rs10488763, chromatin accessibility at the locus of this variant is also increased by ox-LDL (P=0.015; Figure 7A), and this is associated with a 45.3% increase in the local H3K27ac enhancer signal (Figure 7A, bottom). To test whether rs10488763 alleles differentially affect this ox-LDL–induced enhancer activation, we transfected primary macrophages with luciferase reporter plasmids containing each allele. We observed a significant allelic difference, with ox-LDL triggering increased enhancer activity with the protective A allele (average 1.58-fold, P=0.031) but not at all with the risk T allele (Figure 7E).

Inspection of our CEBPB ChIP-seq data from bulk macrophages[38] revealed a CEBPB-binding site 300 bp downstream of the rs10488763 SNP within the same CRE (Figure S8E). Overexpression of the activator isoform of CEBPB (LAP [liver-enriched activator protein]) induced a 7-fold increase in the enhancer activity of this CRE (Figure S8F). To test whether the cis-regulatory effect of rs10488763 is mediated by the recruitment of CEBPB, we overexpressed LAP in primary macrophages and quantified the expression of *FDX1* using RT-qPCR (Figure 7F; Figure S8G). Consistent with the eQTL signals in whole blood samples (Figure 7B), *FDX1*

Figure 6 Continued. Violin plots of aggregated gene expression are shown to the **right**. A red bar indicates significant CREs, and eQTL SNPs are marked with an asterisk. **B**, Violin plot showing the allele-specific effect of rs6772383 on *MITF* expression in human blood based on GTEx data. **C**, Quantification of the immunofluorescence staining of ex vivo macrophages for MITF, shown as average nuclear and intracellular fluorescence intensities (arbitrary units) per biological replicate (n=6). **D**, *MITF* expression in lipid-associated macrophages (LAMs) from plaques vs LAMs from control unaffected arteries. **E**, Genomic plot similar to that in **A**, but showing CAD risk variants rs1800590, rs59811201, rs11204085, rs11204086, and rs11204087 overlapping CREs at the *LPL* locus. *P* values were calculated using a paired Wilcoxon test in **C**, and a Wilcoxon test in **D**.

Downloaded from http://ahajournals.org by on June 25, 2024

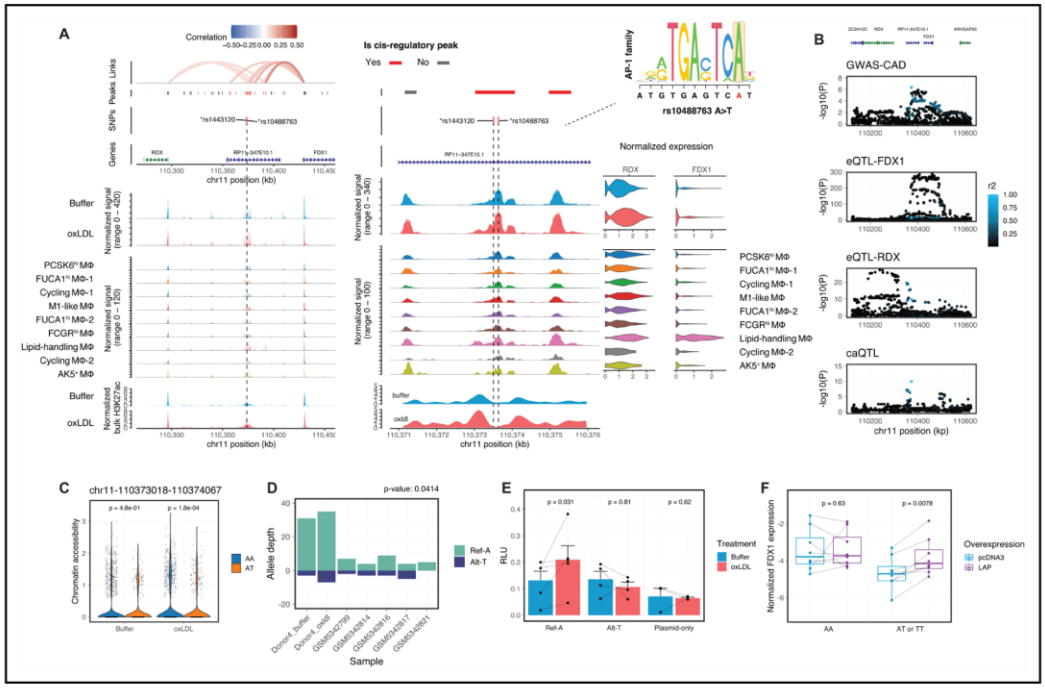


Figure 7. Coronary artery disease (CAD) risk variants at the cis-regulatory region of *FDX1* and *RDX*.
A, Genomic plots of CAD-associated genome-wide association studies (GWAS) variant rs10488763 at the FDX1/RDX locus, similar to the plots shown in Figure 6A. **B**, Association of rs10488763 variant with CAD (GWAS-CAD), *FDX1* expression (eQTL-FDX1), *RDX* expression (eQTL-RDX, note smaller scale of *y* axis), and chromatin accessibility (caQTL) at the *FDX1/RDX* locus. SNPs are colored by their LD with rs10488763. **C**, Chromatin accessibility of the cis-regulatory peak in **A** grouped by genotype from multiomics data (n=3 for A/A, n=1 for A/T). **D**, Allelic sequencing coverage of rs10488763 in 7 heterozygous samples, showing higher accessibility in the chromosome containing the protective A allele compared with the T allele. Reads are aggregated from snATAC-seq data sets of primary human macrophages and human plaques.[70] Sequencing coverage on the alternative allele shown in negative values for visualization purposes. **E**, Dual luciferase reporter assays in primary human macrophages with and without ox-LDL treatment testing the allelic enhancer activity of rs10488763 (n=5 biological replicates). Signals normalized to Renilla luciferase activities. **F**, RT-qPCR measurement of *FDX1* expression. Primary macrophages (n=8, 6, 2 for AA, AT, and TT samples, respectively) were transfected with LAP (liver-enriched activator protein) overexpression plasmid or control pcDNA3 plasmid. *GAPDH* was used as the reference gene. *P* values were calculated using a Wilcoxon test in **C** and a paired Wilcoxon test in **D**, **E**, and **F**.

expression was significantly lower in ex vivo macrophages carrying the risk T allele (P=0.0039; Figure 7F). Overexpression of CEBPB compensated for this reduction of *FDX1* expression in macrophages carrying the T allele (P=0.016; Figure 7F), while having no additional effect in macrophages homozygous for the A allele (P=0.63; Figure 7F).

Together, our analyses link the GWAS signal at the rs10488763 locus to *FDX1* and demonstrate an allelic effect on the ox-LDL–induced enhancer activation and change in chromatin accessibility. Our data are consistent with a model in which the interaction of AP-1 TFs is altered by allelic variation at this SNP, which influences CEBPB recruitment to the CRE and so upregulation of *FDX1* (Figure S9). Given that *FDX1* and *CEBPB* were specifically expressed in lipid-handling macrophages, our results highlight the crucial role of lipid-handling macrophages as the causal subpopulation with an active transcriptional regulatory network required for the CAD risk variant rs10488763 to take effect.

DISCUSSION

Beyond Current Single-Cell Studies on Plaque Cells

Large-scale genome-wide association studies have identified many genomic loci associated with CAD, but the molecular mechanisms accounting for these associations are largely unknown. Recent studies using single-cell sequencing approaches on atherosclerotic lesions have provided insights into how these risk loci affect specific cell types such as endothelial and smooth muscle cells, yet failed to capture the genetic contribution of macrophages.[11,30,32,33] The number of cells available from lesions is typically limited, and most are sampled from

Downloaded from http://ahajournals.org by on June 25, 2024

advanced lesions. In these lesions, the functional pathogenic activity of environmentally derived stimuli, such as atherogenic lipids, may be historic, making genotype-environment interactions difficult to evaluate. For macrophages, these limitations are especially problematic as macrophages are highly plastic and gene regulatory programs are strongly influenced by prevailing stimuli and neighboring cellular activity.[69,71] To overcome this, we have studied primary human macrophages derived ex vivo from circulating monocytes, a major source of plaque macrophages in vivo.[15,35] These ex vivo macrophages allow us to characterize the transcriptional regulatory dynamics before and after exposure to the atherogenic lipid, ox-LDL, and to comprehensively evaluate the contributions of different macrophage subpopulations to CAD heritability.

CAD Variant Prioritization in Macrophages

The analysis of CREs offers a powerful route to further understand CAD heritability. The macrophage CREs that we identified account for 11.5% of the total SNP heritability for CAD and contain 121 risk variants with genome-wide significance. The advantages of the CRE approach over eQTL analysis are evident from the small fraction of these 121 variants that have a cis-regulatory link to the gene with which they are associated by eQTL analysis. Indeed, multiple studies have now reported limited overlaps between eQTLs and GWAS trait-associated variants,[72–74] which highlights the need for other approaches to identify causal genes.[75]

Single-cell sequencing approaches circumvent the need for prior identification and segregation of disease-relevant cell types for functional genomic studies.[76] Most single-cell variant prioritization studies integrating single-cell epigenome data only use open chromatin regions, but such regions are not necessarily regulatory.[30,33] Only 6.9% of open chromatin peaks have cis-regulatory effects in our macrophage data, so colocalization analysis using all open chromatin sites would likely include many false positives for causal variants. A further issue is that the presence of many nonregulatory peaks in accessible chromatin regions drowns out enrichment signals from true regulatory elements, leading to false-negative results in the prioritization of causal cell types or cell states. This is evident in macrophages for which we have previously profiled open chromatin regions in bulk samples and found no global enrichment for CAD risk loci.[38] Similar false negatives were also observed in single-cell ATAC-seq studies on human plaque and heart macrophages.[30,33]

In this study, we have developed a computational workflow to partition SNP heritability among macrophage subpopulations. Our results demonstrate that CAD heritability is not uniformly distributed across these subpopulations. CAD-associated variants likely act by altering gene regulation, so different macrophage subpopulations with distinct transcriptomes and epigenetic profiles are likely to be differently affected by these causal variants. Therefore, the impact of a particular CAD-associated variant may differ between subpopulations. In particular, we show that CAD heritability is enriched in the marker gene program and the ox-LDL response gene program specific to a subset of macrophages, namely the lipid-handling macrophages. We have provided further orthogonal validation of the relevance of such subpopulation-specific CAD variants at the *MITF*, *LPL*, and *FDX1* loci.

Functional Characterization of Lipid-Handling Macrophages

The lipid-handling macrophage subpopulation is distinguished by its activation of lipid-processing genes and transcriptional networks driven by CEBP and NR1H3 family factors. This subpopulation can be tracked in vitro through immunofluorescence staining of a membrane glycoprotein CD52. These CD52-hi macrophages exhibit a higher expression of ox-LDL receptor CD36, yet surprisingly, accumulate much less ox-LDL than other subpopulations. In addition, we noted a variation in the proportion of CD52-hi cells, ranging from 1% to 10% across different biological donors. Although in the current study we only recruited healthy donors, this can be clinically relevant considering the significant lipid-clearing capacities these cells possess.

Through a comprehensive meta-analysis of 38 human atherosclerotic samples from 5 independent studies, we constructed by far the largest single-cell atlas for plaque macrophages with a total of 20 458 cells representing 9 previously reported subpopulations. We demonstrate that monocyte-derived macrophages closely mirror the transcriptomic signatures of the lipid-associated macrophage (LAM) in plaque, a population first discovered in liver and adipose tissue.[63,64] A recent spatial transcriptomic study in the liver reported that LAMs are concentrated around portal veins, while the general monocyte/macrophage compartment is distributed across different zones.[77] Although currently there is no spatial mapping of LAMs in human atherosclerotic plaques, it is reasonable to believe that plaque LAMs in vivo also originate from circulating monocytes.

With the high-resolution plaque macrophage atlas, we were able to further illustrate the transcriptomic heterogeneity within plaque LAMs and identify, for the first time, a CD52-hi subset expressing high levels of lipoprotein processing genes. Although direct quantitative measurement of lipid uptake in human plaque macrophages presents a challenge, the remarkable similarity in transcriptomic profiles between the lipid-handling macrophages and the CD52-hi LAM subset suggests a potentially significant role for these cells in in vivo lipid clearance. These findings may cast light on the

Downloaded from http://ahajournals.org by on June 25, 2024

rare adverse cardiovascular events including myocardial infarction that have been reported with anti-CD52 therapy (alemtuzumab).[78] Binding of alemtuzumab to plaque LAMs within arterial walls may trigger intramural inflammation from acute damage to these cells and depletion of CD52-hi LAM from the vessel walls, thereby increasing CAD risk.

FDX1/RDX Locus

FDX1 encodes an important mitochondrial iron-sulfur protein reductase, which is an electron donor for many reducing reactions[79] and is itself reduced by the NADPH-dependent ferredoxin reductase. Key mitochondrial complexes involved in respiration are regulated by FDX1-mediated lipoylation, and FDX1 deficiency results in loss of respiration.[80] A genome-wide CRISPR screen demonstrated that it was essential for oxidative phosphorylation.[81] While mitochondrial dysfunction is recognized as a contributory mechanism in the initiation of atherosclerosis,[82] the role of ferredoxin has not been characterized in the context of coronary artery disease.

Our computational and experimental results show that the CAD risk T allele at rs10488763 in a distal enhancer region of the *FDX1* gene increases CAD risk by lowering *FDX1* expression in macrophages. The mechanism of this effect is likely complex. Specifically, we demonstrate the role of both AP-1 family factors and CEBPB in modulating the cis-regulatory effect of this CAD SNP. While the AP-1-binding motif is required for ox-LDL–induced enhancer activation, the expression of the AP-1 family member *FOS* negatively correlates with the expression of *FDX1* (Figure 2F). Overexpression of CEBPB bypasses the allelic effect of rs10488763 on enhancer activity and restores *FDX1* expression in macrophages with the disrupted AP-1-binding site. These results indicate a model in which AP-1 binding at the locus does not directly affect *FDX1* expression but rather remodels the local chromatin to create space for other TFs like CEBPB (Figure S9). This model is consistent with the observation that AP-1 factors bind collaboratively with cell-type–specific TFs to select enhancers and establish accessible chromatin.[83]

Overview

Overall, we have developed a comprehensive framework that enables the systematic quantification of the disease relevance of heterogeneous macrophage subpopulations. This framework greatly facilitates the investigation of molecular mechanisms that account for the heterogeneity in the genetic risk conveyed by these subpopulations. Our work illustrates the substantial value of using single-cell information to improve the genomic resolution of the genetic architecture of human diseases.

Downloaded from http://ahajournals.org by on June 25, 2024

ARTICLE INFORMATION

Received December 21, 2023; revision received April 26, 2024; accepted May 1, 2024.

Affiliations

Nuffield Department of Medicine, Wellcome Centre for Human Genetics (J.J., T.K.H., T.A.A., Y.M., E.D., J.B., L.J.D., C.A.O.) and Nuffield Department of Orthopaedics, Kennedy Institute of Rheumatology, Rheumatology and Musculoskeletal Sciences (E.N.), University of Oxford, United Kingdom. Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, United Kingdom (M.E.R.). Department of Clinical Science and Services, Royal Veterinary College, Hatfield, United Kingdom (L.J.D.).

Acknowledgments

The authors thank all healthy volunteers who agreed to participate in this study. The authors also thank the Oxford Genomic Center for their support. Illustrations in Figure S1A, Figure S2D, and Figure S9 were created with BioRender.com.

Author Contributions

C.A. O'Callaghan, T.K. Hiron, and J. Jiang contributed to conceptualization and design; T.K. Hiron, J. Jiang, T.A. Agbaedeng, Y. Malhotra, E. Drydale, J. Bancroft, and M.E. Reschen contributed to data collection; J. Jiang, T.K. Hiron, and C.A. O'Callaghan contributed to analysis and interpretation of results; C.A. O'Callaghan contributed to resources and funding acquisition and supervision. J. Jiang and C.A. O'Callaghan contributed to the manuscript drafting. All authors evaluated the results and commented on and approved the article.

Sources of Funding

The research was supported by the Novo Nordisk Foundation (NNF15CC0018346 and NNF0064142), the British Hearth Foundation (RG/F/22/110085), the Wellcome Trust Core Award grant number 203141/Z/16/Z with funding from the National Institute for Health and Care Research (NIHR) Oxford Biomedical Research Centre. J. Jiang received funding from the China Scholarship Council (202006320024). T.A. Agbaedeng is supported by a Novo Nordisk Postdoctoral Fellowship run in partnership with the University of Oxford. The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR or the Department of Health.

Disclosures

None.

Supplemental Material

Expanded Materials and Methods
Supplemental Notes
Video S1
Tables S1–S6
Figures S1–S9
Major Resources Table
References 12–14,18,23,38,49,51,53,54,56,60,61,65,69,84–98

REFERENCES

1. Kassebaum NJ, Bertozzi-Villa A, Coggeshall MS, Shackelford KA, Steiner C, Heuton KR, Gonzalez-Medina D, Barber R, Huynh C, Dicker D, et al. Global, regional, and national levels and causes of maternal mortality during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384:980–1004. doi: 10.1016/S0140-6736(14)60696-6
2. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol*. 2019;234:16812–16823. doi: 10.1002/jcp.28350
3. Woollard KJ, Geissmann F. Monocytes in atherosclerosis: subsets and functions. *Nat Rev Cardiol*. 2010;7:77–86. doi: 10.1038/nrcardio.2009.228
4. Chinetti-Gbaguidi G, Colin S, Staels B. Macrophage subsets in atherosclerosis. *Nat Rev Cardiol*. 2015;12:10–17. doi: 10.1038/nrcardio.2014.173
5. Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis. *Cell*. 2011;145:341–355. doi: 10.1016/j.cell.2011.04.005
6. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol*. 2005;5:953–964. doi: 10.1038/nri1733
7. Guilliams M, Mildner A, Yona S. Developmental and functional heterogeneity of monocytes. *Immunity*. 2018;49:595–613. doi: 10.1016/j.immuni.2018.10.005
8. Bouhlel MA, Derudas B, Rigamonti E, Dievart R, Brozek J, Haulon S, Zawadzki C, Jude B, Torpier G, Marx N, et al. PPARgamma

activation primes human monocytes into alternative M2 macrophages with anti-inflammatory properties. *Cell Metab.* 2007;6:137–143. doi: 10.1016/j.cmet.2007.06.010
9. Chinetti-Gbaguidi G, Baron M, Bouhlel MA, Vanhoutte J, Copin C, Sebti Y, Derudas B, Mayi T, Bories G, Tailleux A, et al. Human atherosclerotic plaque alternative macrophages display low cholesterol handling but high phagocytosis because of distinct activities of the PPARgamma and LXRalpha pathways. *Circ Res.* 2011;108:985–995. doi: 10.1161/CIRCRESAHA.110.233775
10. Barrett TJ. Macrophages in atherosclerosis regression. *Arterioscler Thromb Vasc Biol.* 2020;40:20–33. doi: 10.1161/ATVBAHA.119.312802
11. Depuydt MAC, Prange KHM, Slenders L, Ord T, Elbersen D, Boltjes A, de Jager SCA, Asselbergs FW, de Borst GJ, Aavik E, et al. Microanatomy of the human atherosclerotic plaque by single-cell transcriptomics. *Circ Res.* 2020;127:1437–1455. doi: 10.1161/CIRCRESAHA.120.316770
12. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J, Koyano TK, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med.* 2019;25:1280–1289. doi: 10.1038/s41591-019-0512-5
13. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. *Commun Biol.* 2022;5:1084. doi: 10.1038/s42003-022-04056-7
14. Chowdhury RR, D'Addabbo J, Huang X, Veizades S, Sasagawa K, Louis DM, Cheng P, Sokol J, Jensen A, Tso A, et al. Human coronary plaque T cells are clonal and cross-react to virus and self. *Circ Res.* 2022;130:1510–1530. doi: 10.1161/CIRCRESAHA.121.320090
15. Tabas I, Lichtman AH. Monocyte-macrophages and T cells in atherosclerosis. *Immunity.* 2017;47:621–634. doi: 10.1016/j.immuni.2017.09.008
16. Libby P. Inflammation in atherosclerosis. *Nature.* 2002;420:868–874. doi: 10.1038/nature01323
17. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet.* 2017;18:331–344. doi: 10.1038/nrg.2016.160
18. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res.* 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086
19. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang M, Hindy G, Zhou W, et al; Biobank Japan. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet.* 2022;54:1803–1815. doi: 10.1038/s41588-022-01233-6
20. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science.* 2012;337:1190–1195. doi: 10.1126/science.1222794
21. Zeggini E, Gloyn AL, Barton AC, Wain LV. Translational genomics and precision medicine: moving from the lab to the clinic. *Science.* 2019;365:1409–1413. doi: 10.1126/science.aax4588
22. Hekselman I, Yeger-Lotem E. Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nat Rev Genet.* 2020;21:137–150. doi: 10.1038/s41576-019-0200-9
23. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods.* 2017;14:975–978. doi: 10.1038/nmeth.4401
24. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Mol Cell.* 2018;71:858–871.e8. doi: 10.1016/j.molcel.2018.06.044
25. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, Kannan S. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using scribe. *Cell Syst.* 2020;10:265–274.e11. doi: 10.1016/j.cels.2020.02.003
26. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell.* 2020;183:1103–1116.e20. doi: 10.1016/j.cell.2020.09.056
27. Jagadeesh KA, Dey KK, Montoro DT, Mohan R, Gazal S, Engreitz JM, Xavier RJ, Price AL, Regev A. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet.* 2022;54:1479–1492. doi: 10.1038/s41588-022-01187-9
28. Watanabe K, Umicevic Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun.* 2019;10:3222. doi: 10.1038/s41467-019-11181-1
29. Kim SS, Jagadeesh K, Dey KK, Shen AZ, Raychaudhuri S, Kellis M, Price AL. Leveraging single-cell ATAC-seq to identify disease-critical fetal and adult brain cell types. *bioRxiv.* Preprint posted online May 21, 2021. doi: 10.1101/2021.05.20.445067
30. Ord T, Ounap K, Stolze LK, Aherrahrou R, Nurminen V, Toropainen A, Selvarajan I, Lonnberg T, Aavik E, Yla-Herttuala S, et al. Single-cell epigenomics and functional fine-mapping of atherosclerosis GWAS loci. *Circ Res.* 2021;129:240–258. doi: 10.1161/CIRCRESAHA.121.318971
31. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, Kohlmaier A, Herbst A, Northoff BH, Nicolaou A, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun.* 2016;7:12429. doi: 10.1038/ncomms12429
32. Slenders L, Landsmeer LPL, Cui K, Depuydt MAC, Verwer M, Mekke J, Timmerman N, van den Dungen NAM, Kuiper J, de Winther MPJ, et al. Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. *Eur Heart J Open.* 2022;2:oeab043. doi: 10.1093/ehjopen/oeab043
33. Hocker JD, Poirion OB, Zhu F, Buchanan J, Zhang K, Chiou J, Wang TM, Zhang Q, Hou X, Li YE, et al. Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci Adv.* 2021;7:eabf1444. doi: 10.1126/sciadv.abf1444
34. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet.* 2019;20:207–220. doi: 10.1038/s41576-018-0089-8
35. Hilgendorf I, Swirski FK, Robbins CS. Monocyte fate in atherosclerosis. *Arterioscler Thromb Vasc Biol.* 2015;35:272–279. doi: 10.1161/ATVBAHA.114.303565
36. Das R, Ganapathy S, Mahabeleshwar GH, Drumm C, Febbraio M, Jain MK, Plow EF. Macrophage gene expression and foam cell formation are regulated by plasminogen. *Circulation.* 2013;127:1209–18, e1. doi: 10.1161/CIRCULATIONAHA.112.001214
37. Reschen ME, Lin D, Chalisey A, Soilleux EJ, O'Callaghan CA. Genetic and environmental risk factors for atherosclerosis regulate transcription of phosphatase and actin regulating gene PHACTR1. *Atherosclerosis.* 2016;250:95–105. doi: 10.1016/j.atherosclerosis.2016.04.025
38. Reschen ME, Gaulton KJ, Lin D, Soilleux EJ, Morris AJ, Smyth SS, O'Callaghan CA. 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:e1005061. doi: 10.1371/journal.pgen.1005061
39. Domschke G, Linden F, Pawig L, Hafner A, Akhavanpoor M, Reymann J, Doesch AO, Erbel C, Weber C, Katus HA, et al. Systematic RNA-interference in primary human monocyte-derived macrophages: a high-throughput platform to study foam cell formation. *Sci Rep.* 2018;8:10516. doi: 10.1038/s41598-018-28790-3
40. Park YM. CD36, a scavenger receptor implicated in atherosclerosis. *Exp Mol Med.* 2014;46:e99. doi: 10.1038/emm.2014.38
41. Chistiakov DA, Bobryshev YV, Orekhov AN. Changes in transcriptome of macrophages in atherosclerosis. *J Cell Mol Med.* 2015;19:1163–1173. doi: 10.1111/jcmm.12591
42. Georgess D, Mazzorana M, Terrado J, Delprat C, Chamot C, Guasch RM, Perez-Roger I, Jurdic P, Machuca-Gayet I. Comparative transcriptomics reveals RhoE as a novel regulator of actin dynamics in bone-resorbing osteoclasts. *Mol Biol Cell.* 2014;25:380–396. doi: 10.1091/mbc.E13-07-0363
43. Huang HC, Nguyen T, Pickett CB. Regulation of the antioxidant response element by protein kinase C-mediated phosphorylation of NF-E2-related factor 2. *Proc Natl Acad Sci U S A.* 2000;97:12475–12480. doi: 10.1073/pnas.220418997
44. Zhou LZ, Johnson AP, Rando TA. NF kappa B and AP-1 mediate transcriptional responses to oxidative stress in skeletal muscle cells. *Free Radic Biol Med.* 2001;31:1405–1416. doi: 10.1016/s0891-5849(01)00719-5
45. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, et al; GTEx Consortium. A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv Biobank.* 2015;13:311–319. doi: 10.1089/bio.2015.0032
46. Vosa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, et al; BIOS Consortium. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet.* 2021;53:1300–1310. doi: 10.1038/s41588-021-00913-z
47. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature.* 2021;593:238–243. doi: 10.1038/s41586-021-03446-x

Downloaded from http://ahajournals.org by on June 25, 2024

48. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, et al; BLUEPRINT Consortium. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell*. 2016;167:1369–1384.e19. doi: 10.1016/j.cell.2016.09.037
49. Kartha VK, Duarte FM, Hu Y, Ma S, Chew JG, Lareau CA, Earl A, Burkett ZD, Kohlway AS, Lebofsky R, et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom*. 2022;2:100166. doi: 10.1016/j.xgen.2022.100166
50. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24:R102–R110. doi: 10.1093/hmg/ddv259
51. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al; ReproGen Consortium. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228–1235. doi: 10.1038/ng.3404
52. Bjorkegren JLM, Kovacic JC, Dudley JT, Schadt EE. 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;65:830–845. doi: 10.1016/j.jacc.2014.12.033
53. Fang H, Consortium U-D, De Wolf H, Knezevic B, Burnham KL, Osgood J, Sanniti A, Lledo Lara A, Kasela S, De Cesco S, et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet*. 2019;51:1082–1091. doi: 10.1038/s41588-019-0456-1
54. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, Schoech A, Bulik-Sullivan B, Neale BM, Gusev A, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet*. 2017;49:1421–1427. doi: 10.1038/ng.3954
55. Rouland A, Masson D, Lagrost L, Verges B, Gautier T, Bouillet B. Role of apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: a systematic review. *Cardiovasc Diabetol*. 2022;21:272. doi: 10.1186/s12933-022-01703-5
56. Koplev S, Seldin M, Sukhavasi K, Ermel R, Pang S, Zeng L, Bankier S, Di Narzo A, Cheng H, Meda V, et al. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat Cardiovasc Res*. 2022;1:85–100. doi: 10.1038/s44161-021-00009-1
57. Wang B, Tontonoz P. Liver X receptors in lipid signalling and membrane homeostasis. *Nat Rev Endocrinol*. 2018;14:452–463. doi: 10.1038/s41574-018-0037-x
58. Nicholson AC, Frieda S, Pearce A, Silverstein RL. Oxidized LDL binds to CD36 on human monocyte-derived macrophages and transfected cell lines. Evidence implicating the lipid moiety of the lipoprotein as the binding site. *Arterioscler Thromb Vasc Biol*. 1995;15:269–275. doi: 10.1161/01.atv.15.2.269
59. Williams JW, Zaitsev K, Kim KW, Ivanov S, Saunders BT, Schrank PR, Kim K, Elvington A, Kim SH, Tucker CG, et al. Limited proliferation capacity of aortic intima resident macrophages requires monocyte recruitment for atherosclerotic plaque progression. *Nat Immunol*. 2020;21:1194–1204. doi: 10.1038/s41590-020-0768-4
60. Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, Yang DY, Trignano SB, Liu W, Shi J, et al. Single-cell genomics reveals a novel cell state during smooth muscle cell phenotypic switching and potential therapeutic targets for atherosclerosis in mouse and human. *Circulation*. 2020;142:2060–2075. doi: 10.1161/CIRCULATIONAHA.120.048378
61. Dib L, Koneva LA, Edsfeldt A, Zurke Y-X, Sun J, Nitulescu M, Attar M, Lutgens E, Schmidt S, Lindholm MW, et al. Lipid-associated macrophages transition to an inflammatory state in human atherosclerosis, increasing the risk of cerebrovascular complications. *Nat Cardiovasc Res*. 2023;2:656–672. doi: 10.1038/s44161-023-00295-x
62. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, Khan NS, Wong CK, Shamailova R, Hill CA, et al. Single-cell immune landscape of human atherosclerotic plaques. *Nat Med*. 2019;25:1576–1588. doi: 10.1038/s41591-019-0590-4
63. Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, Wang J, Van Wittenberghe N, Rouhana JM, Waldman J, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*. 2022;376:eabl4290. doi: 10.1126/science.abl4290
64. Jaitin DA, Adlung L, Thaiss CA, Weiner A, Li B, Descamps H, Lundgren P, Bleriot C, Liu Z, Deczkowska A, et al. Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell*. 2019;178:686–698.e14. doi: 10.1016/j.cell.2019.05.054
65. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol*. 2021;17:e9620. doi: 10.15252/msb.20209620
66. Perera RM, Stoykova S, Nicolay BN, Ross KN, Fitamant J, Boukhali M, Lengrand J, Deshpande V, Selig MK, Ferrone CR, et al. Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. *Nature*. 2015;524:361–365. doi: 10.1038/nature14587
67. Ploper D, Taelman VF, Robert L, Perez BS, Titz B, Chen HW, Graeber TG, von Euw E, Ribas A, De Robertis EM. MITF drives endolysosomal biogenesis and potentiates Wnt signaling in melanoma cells. *Proc Natl Acad Sci U S A*. 2015;112:E420–E429. doi: 10.1073/pnas.1424576112
68. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, Matsunaga H, Ieki H, Ozaki K, Onouchi Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet*. 2020;52:1169–1177. doi: 10.1038/s41588-020-0705-3
69. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, Consortium H, Hale C, Dougan G, Gaffney DJ. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet*. 2018;50:424–431. doi: 10.1038/s41588-018-0046-7
70. Turner AW, Hu SS, Mosquera JV, Ma WF, Hodonsky CJ, Wong D, Auguste G, Song Y, Sol-Church K, Farber E, et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet*. 2022;54:804–816. doi: 10.1038/s41588-022-01069-0
71. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell*. 2016;167:657–669.e21. doi: 10.1016/j.cell.2016.09.025
72. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet*. 2020;52:626–633. doi: 10.1038/s41588-020-0625-2
73. Umans BD, Battle A, Gilad Y. Where are the disease-associated eQTLs? *Trends Genet*. 2021;37:109–124. doi: 10.1016/j.tig.2020.08.009
74. Mostafavi H, Spence JP, Naqvi S, Pritchard JK. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet*. 2023;55:1866–1875. doi: 10.1038/s41588-023-01529-1
75. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424. doi: 10.3389/fgene.2020.00424
76. Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, Satpathy AT, Kartha VK, Salem RM, Hirschhorn JN, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet*. 2019;51:683–693. doi: 10.1038/s41588-019-0362-6
77. Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, Bujko A, Martens L, Thone T, Browaeys R, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell*. 2022;185:379–396.e38. doi: 10.1016/j.cell.2021.12.018
78. Basquiera AL, Berretta AR, Garcia JJ, Palazzo ED. Coronary ischemia related to alemtuzumab therapy. *Ann Oncol*. 2004;15:539–540. doi: 10.1093/annonc/mdh114
79. Sheftel AD, Stehling O, Pierik AJ, Elsasser HP, Muhlenhoff U, Webert H, Hobler A, Hannemann F, Bernhardt R, Lill R. Humans possess two mitochondrial ferredoxins, Fdx1 and Fdx2, with distinct roles in steroidogenesis, heme, and Fe/S cluster biosynthesis. *Proc Natl Acad Sci U S A*. 2010;107:11775–11780. doi: 10.1073/pnas.1004250107
80. Dreishpoon MB, Bick NR, Petrova B, Warui DM, Cameron A, Booker SJ, Kanarek N, Golub TR, Tsvetkov P. FDX1 regulates cellular protein lipoylation through direct binding to LIAS. *J Biol Chem*. 2023;299:105046. doi: 10.1016/j.jbc.2023.105046
81. Arroyo JD, Jourdain AA, Calvo SE, Ballarano CA, Doench JG, Root DE, Mootha VK. A Genome-wide CRISPR death screen identifies genes essential for oxidative phosphorylation. *Cell Metab*. 2016;24:875–885. doi: 10.1016/j.cmet.2016.08.017
82. Madamanchi NR, Runge MS. Mitochondrial dysfunction in atherosclerosis. *Circ Res*. 2007;100:460–473. doi: 10.1161/01.RES.0000258450.44413.96
83. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME. AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol Cell*. 2017;68:1067–1082.e12. doi: 10.1016/j.molcel.2017.11.026
84. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902.e21. doi: 10.1016/j.cell.2019.05.031
85. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst*. 2019;8:329–337.e4. doi: 10.1016/j.cels.2019.03.003
86. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16:1289–1296. doi: 10.1038/s41592-019-0619-0

Downloaded from http://ahajournals.org by on June 25, 2024

87. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, Chak S, Naikawadi RP, Wolters PJ, Abate AR, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol*. 2019;20:163–172. doi: 10.1038/s41590-018-0276-y
88. Tang M, Kaymaz Y, Logeman BL, Eichhorn S, Liang ZS, Dulac C, Sackton TB. Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics*. 2021;37:2212–2214. doi: 10.1093/bioinformatics/btaa956
89. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM; GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27:3641–3649. doi: 10.1093/hmg/ddy271
90. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, Benner C, Liu D, Locke AE, Balasubramanian S, et al; Regeneron Genetics Center. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599:628–634. doi: 10.1038/s41586-021-04103-z
91. Fang H, Knezevic B, Burnham KL, Knight JC. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*. 2016;8:129. doi: 10.1186/s13073-016-0384-y
92. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e29. doi: 10.1016/j.cell.2021.04.048
93. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575. doi: 10.1086/519795
94. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800. doi: 10.1371/journal.pone.0021800
95. van der Auwera G, O'Connor BD. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*. O'Reilly Media; 2020.
96. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–2383. doi: 10.1093/bioinformatics/btv145
97. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol*. 2008;9:R137. doi: 10.1186/gb-2008-9-9-r137
98. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–359. doi: 10.1038/nmeth.1923

Downloaded from http://ahajournals.org by on June 25, 2024

Bibliography

1. Collaborators GBDCoD. Global burden of 288 causes of death and life expectancy decomposition in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet*. 2024;403:2100-2132. doi: 10.1016/S0140-6736(24)00367-2

2. Stary HC, Chandler AB, Dinsmore RE, Fuster V, Glagov S, Insull W, Jr., Rosenfeld ME, Schwartz CJ, Wagner WD, Wissler RW. 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. *Circulation*. 1995;92:1355-1374. doi: 10.1161/01.cir.92.5.1355

3. Gistera A, Hansson GK. The immunology of atherosclerosis. *Nat Rev Nephrol*. 2017;13:368-380. doi: 10.1038/nrneph.2017.51

4. Gimbrone MA, Jr., Garcia-Cardena G. Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. *Circ Res*. 2016;118:620-636. doi: 10.1161/CIRCRESAHA.115.306301

5. Sukriti S, Tauseef M, Yazbeck P, Mehta D. Mechanisms regulating endothelial permeability. *Pulm Circ*. 2014;4:535-551. doi: 10.1086/677356

6. Weinbaum S, Tzeghai G, Ganatos P, Pfeffer R, Chien S. Effect of cell turnover and leaky junctions on arterial macromolecular transport. *Am J Physiol*. 1985;248:H945-960. doi: 10.1152/ajpheart.1985.248.6.H945

7. Davignon J, Ganz P. Role of endothelial dysfunction in atherosclerosis. *Circulation*. 2004;109:III27-32. doi: 10.1161/01.CIR.0000131515.03336.f8

8. Scott-Burden T, Vanhoutte PM. Regulation of smooth muscle cell growth by endothelium-derived factors. *Tex Heart Inst J*. 1994;21:91-97.

9. Bennett MR, Sinha S, Owens GK. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ Res*. 2016;118:692-702. doi: 10.1161/CIRCRESAHA.115.306361

10. Hadi HA, Suwaidi JA. Endothelial dysfunction in diabetes mellitus. *Vasc Health Risk Manag*. 2007;3:853-876.

11. Rajendran P, Rengarajan T, Thangavel J, Nishigaki Y, Sakthisekaran D, Sethi G, Nishigaki I. The vascular endothelium and human diseases. *Int J Biol Sci*. 2013;9:1057-1069. doi: 10.7150/ijbs.7502

12. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol*. 2019;234:16812-16823. doi: 10.1002/jcp.28350

13. Cheng H, Zhong W, Wang L, Zhang Q, Ma X, Wang Y, Wang S, He C, Wei Q, Fu C. Effects of shear stress on vascular endothelial functions in atherosclerosis and potential therapeutic approaches. *Biomed Pharmacother*. 2023;158:114198. doi: 10.1016/j.biopha.2022.114198

14. Parmar KM, Larman HB, Dai G, Zhang Y, Wang ET, Moorthy SN, Kratz JR, Lin Z, Jain MK, Gimbrone MA, Jr., et al. Integration of flow-dependent endothelial phenotypes by Kruppel-like factor 2. *J Clin Invest*. 2006;116:49-58. doi: 10.1172/JCI24787

15. Chappell DC, Varner SE, Nerem RM, Medford RM, Alexander RW. Oscillatory shear stress stimulates adhesion molecule expression in cultured human endothelium. *Circ Res*. 1998;82:532-539. doi: 10.1161/01.res.82.5.532

16. Matharu NM, McGettrick HM, Salmon M, Kissane S, Vohra RK, Rainger GE, Nash GB. Inflammatory responses of endothelial cells experiencing reduction in flow after conditioning by shear stress. *J Cell Physiol*. 2008;216:732-741. doi: 10.1002/jcp.21457

17. Woollard KJ, Geissmann F. Monocytes in atherosclerosis: subsets and functions. *Nat Rev Cardiol*. 2010;7:77-86. doi: 10.1038/nrcardio.2009.228

18. Chinetti-Gbaguidi G, Colin S, Staels B. Macrophage subsets in atherosclerosis. *Nat Rev Cardiol*. 2015;12:10-17. doi: 10.1038/nrcardio.2014.173

19. Moore KJ, Tabas I. Macrophages in the pathogenesis of atherosclerosis. *Cell*. 2011;145:341-355. doi: 10.1016/j.cell.2011.04.005

20. Kitagawa K, Matsumoto M, Sasaki T, Hashimoto H, Kuwabara K, Ohtsuki T, Hori M. Involvement of ICAM-1 in the progression of atherosclerosis in APOE-knockout mice. *Atherosclerosis*. 2002;160:305-310. doi: 10.1016/s0021-9150(01)00587-1

21. Qiao JH, Tripathi J, Mishra NK, Cai Y, Tripathi S, Wang XP, Imes S, Fishbein MC, Clinton SK, Libby P, et al. Role of macrophage colony-stimulating factor in atherosclerosis: studies of osteopetrotic mice. *Am J Pathol*. 1997;150:1687-1699.

22. Dubland JA, Allahverdian S, Besler KJ, Ortega C, Wang Y, Pryma CS, Boukais K, Chan T, Seidman MA, Francis GA. Low LAL (Lysosomal Acid Lipase) Expression by Smooth Muscle Cells Relative to Macrophages as a Mechanism for Arterial Foam Cell Formation. *Arterioscler Thromb Vasc Biol*. 2021;41:e354-e368. doi: 10.1161/ATVBAHA.120.316063

23. Sica A, Mantovani A. Macrophage plasticity and polarization: in vivo veritas. *J Clin Invest*. 2012;122:787-795. doi: 10.1172/JCI59643

24. Depuydt MAC, Prange KHM, Slenders L, Ord T, Elbersen D, Boltjes A, de Jager SCA, Asselbergs FW, de Borst GJ, Aavik E, et al. Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. *Circ Res*. 2020;127:1437-1455. doi: 10.1161/CIRCRESAHA.120.316770

25. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, Kundu R, Nagao M, Coller J, Koyano TK, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med*. 2019;25:1280-1289. doi: 10.1038/s41591-019-0512-5

26. Alsaigh T, Evans D, Frankel D, Torkamani A. Decoding the transcriptome of calcified atherosclerotic plaque at single-cell resolution. *Commun Biol*. 2022;5:1084. doi: 10.1038/s42003-022-04056-7

27. Chowdhury RR, D'Addabbo J, Huang X, Veizades S, Sasagawa K, Louis DM, Cheng P, Sokol J, Jensen A, Tso A, et al. Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. *Circ Res*. 2022;130:1510-1530. doi: 10.1161/CIRCRESAHA.121.320090

28. Jaitin DA, Adlung L, Thaiss CA, Weiner A, Li B, Descamps H, Lundgren P, Bleriot C, Liu Z, Deczkowska A, et al. Lipid-Associated Macrophages Control Metabolic Homeostasis in a Trem2-Dependent Manner. *Cell*. 2019;178:686-698 e614. doi: 10.1016/j.cell.2019.05.054

29. Xu R, Vujic N, Bianco V, Reinisch I, Kratky D, Krstic J, Prokesch A. Lipid-associated macrophages between aggravation and alleviation of metabolic diseases. *Trends Endocrinol Metab*. 2024. doi: 10.1016/j.tem.2024.04.009

30. Dib L, Koneva LA, Edsfeldt A, Zurke YX, Sun J, Nitulescu M, Attar M, Lutgens E, Schmidt S, Lindholm MW, et al. Lipid-associated macrophages transition to an inflammatory state in human atherosclerosis increasing the risk of cerebrovascular complications. *Nat Cardiovasc Res*. 2023;2:656-672. doi: 10.1038/s44161-023-00295-x

31. Patterson MT, Firulyova MM, Xu Y, Hillman H, Bishop C, Zhu A, Hickok GH, Schrank PR, Ronayne CE, Caillot Z, et al. Trem2 promotes foamy macrophage lipid uptake and survival in atherosclerosis. *Nat Cardiovasc Res*. 2023;2:1015-1031. doi: 10.1038/s44161-023-00354-3

32. Chappell J, Harman JL, Narasimhan VM, Yu H, Foote K, Simons BD, Bennett MR, Jorgensen HF. Extensive Proliferation of a Subset of Differentiated, yet Plastic, Medial Vascular Smooth Muscle Cells Contributes to Neointimal Formation in Mouse Injury and Atherosclerosis Models. *Circ Res*. 2016;119:1313-1323. doi: 10.1161/CIRCRESAHA.116.309799

33. Pan H, Xue C, Auerbach BJ, Fan J, Bashore AC, Cui J, Yang DY, Trignano SB, Liu W, Shi J, et al. Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. *Circulation*. 2020;142:2060-2075. doi: 10.1161/CIRCULATIONAHA.120.048378

34. Zhao G, Lu H, Chang Z, Zhao Y, Zhu T, Chang L, Guo Y, Garcia-Barrio MT, Chen YE, Zhang J. Single-cell RNA sequencing reveals the cellular heterogeneity of aneurysmal infrarenal abdominal aorta. *Cardiovasc Res*. 2021;117:1402-1416. doi: 10.1093/cvr/cvaa214

35. Kalucka J, de Rooij L, Goveia J, Rohlenova K, Dumas SJ, Meta E, Conchinha NV, Taverna F, Teuwen LA, Veys K, et al. Single-Cell Transcriptome Atlas of Murine Endothelial Cells. *Cell*. 2020;180:764-779 e720. doi: 10.1016/j.cell.2020.01.015

36. Alonso-Herranz L, Albarran-Juarez J, Bentzon JF. Mechanisms of fibrous cap formation in atherosclerosis. *Front Cardiovasc Med*. 2023;10:1254114. doi: 10.3389/fcvm.2023.1254114

37. Saigusa R, Winkels H, Ley K. T cell subsets and functions in atherosclerosis. *Nat Rev Cardiol*. 2020;17:387-401. doi: 10.1038/s41569-020-0352-5

38. Gewaltig J, Kummer M, Koella C, Cathomas G, Biedermann BC. Requirements for CD8 T-cell migration into the human arterial wall. *Hum Pathol*. 2008;39:1756-1762. doi: 10.1016/j.humpath.2008.04.018

39. van der Wal AC, Das PK, Bentz van de Berg D, van der Loos CM, Becker AE. Atherosclerotic lesions in humans. In situ immunophenotypic analysis suggesting an immune mediated response. *Lab Invest*. 1989;61:166-170.

40. Ali AJ, Makings J, Ley K. Regulatory T Cell Stability and Plasticity in Atherosclerosis. *Cells*. 2020;9. doi: 10.3390/cells9122665

41. Winkels H, Ehinger E, Vassallo M, Buscher K, Dinh HQ, Kobiyama K, Hamers AAJ, Cochain C, Vafadarnejad E, Saliba AE, et al. Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. *Circ Res*. 2018;122:1675-1688. doi: 10.1161/CIRCRESAHA.117.312513

42. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, Khan NS, Wong CK, Shamailova R, Hill CA, et al. Single-cell immune landscape of human atherosclerotic plaques. *Nat Med*. 2019;25:1576-1588. doi: 10.1038/s41591-019-0590-4

43. Wolf D, Gerhardt T, Winkels H, Michel NA, Pramod AB, Ghosheh Y, Brunel S, Buscher K, Miller J, McArdle S, et al. Pathogenic Autoimmunity in Atherosclerosis Evolves From Initially Protective Apolipoprotein B(100)-Reactive

CD4(+) T-Regulatory Cells. *Circulation*. 2020;142:1279-1293. doi: 10.1161/CIRCULATIONAHA.119.042863

44. Konstantinov IE, Mejevoi N, Anichkov NM. Nikolai N. Anichkov and his theory of atherosclerosis. *Tex Heart Inst J*. 2006;33:417-423.

45. Kapur NK, Ashen D, Blumenthal RS. High density lipoprotein cholesterol: an evolving target of therapy in the management of cardiovascular disease. *Vasc Health Risk Manag*. 2008;4:39-57. doi: 10.2147/vhrm.2008.04.01.39

46. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy: part II: the early evidence linking hypercholesterolemia to coronary disease in humans. *J Lipid Res*. 2005;46:179-190. doi: 10.1194/jlr.R400012-JLR200

47. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis. An interpretive history of the cholesterol controversy: part I. *J Lipid Res*. 2004;45:1583-1593. doi: 10.1194/jlr.R400003-JLR200

48. Siri-Tarino PW, Krauss RM. The early years of lipoprotein research: from discovery to clinical application. *J Lipid Res*. 2016;57:1771-1777. doi: 10.1194/jlr.R069575

49. Sampietro T, Tuoni M, Ferdeghini M, Ciardi A, Marraccini P, Prontera C, Sassi G, Taddei M, Bionda A. Plasma cholesterol regulates soluble cell adhesion molecule expression in familial hypercholesterolemia. *Circulation*. 1997;96:1381-1385. doi: 10.1161/01.cir.96.5.1381

50. O'Keefe JH, Jr., Cordain L, Harris WH, Moe RM, Vogel R. Optimal low-density lipoprotein is 50 to 70 mg/dl: lower is better and physiologically normal. *J Am Coll Cardiol*. 2004;43:2142-2146. doi: 10.1016/j.jacc.2004.03.046

51. Gordon T, Castelli WP, Hjortland MC, Kannel WB, Dawber TR. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med*. 1977;62:707-714. doi: 10.1016/0002-9343(77)90874-9

52. Stamler J, Wentworth D, Neaton JD. 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;256:2823-2828.

53. Mattiuzzi C, Sanchis-Gomar F, Lippi G. Worldwide burden of LDL cholesterol: Implications in cardiovascular disease. *Nutr Metab Cardiovasc Dis*. 2020;30:241-244. doi: 10.1016/j.numecd.2019.09.008

54. Goldstein JL, Brown MS. The low-density lipoprotein pathway and its relation to atherosclerosis. *Annu Rev Biochem*. 1977;46:897-930. doi: 10.1146/annurev.bi.46.070177.004341

55. Goldstein JL, Ho YK, Basu SK, Brown MS. Binding site on macrophages that mediates uptake and degradation of acetylated low density lipoprotein, producing massive cholesterol deposition. *Proc Natl Acad Sci U S A*. 1979;76:333-337. doi: 10.1073/pnas.76.1.333

56. Alquraini A, El Khoury J. Scavenger receptors. *Curr Biol*. 2020;30:R790-R795. doi: 10.1016/j.cub.2020.05.051

57. Steinberg D. Thematic review series: the pathogenesis of atherosclerosis: an interpretive history of the cholesterol controversy, part III: mechanistically defining the role of hyperlipidemia. *J Lipid Res*. 2005;46:2037-2051. doi: 10.1194/jlr.R500010-JLR200

58. Hessler JR, Morel DW, Lewis LJ, Chisolm GM. Lipoprotein oxidation and lipoprotein-induced cytotoxicity. *Arteriosclerosis*. 1983;3:215-222. doi: 10.1161/01.atv.3.3.215

59. Steinberg D, Parthasarathy S, Carew TE, Khoo JC, Witztum JL. Beyond cholesterol. Modifications of low-density lipoprotein that increase its atherogenicity. *N Engl J Med*. 1989;320:915-924. doi: 10.1056/NEJM198904063201407

60. Steinberg D. The LDL modification hypothesis of atherogenesis: an update. *J Lipid Res*. 2009;50 Suppl:S376-381. doi: 10.1194/jlr.R800087-JLR200

61. Poznyak AV, Nikiforov NG, Markin AM, Kashirskikh DA, Myasoedova VA, Gerasimova EV, Orekhov AN. Overview of OxLDL and Its Impact on Cardiovascular Health: Focus on Atherosclerosis. *Front Pharmacol*. 2020;11:613780. doi: 10.3389/fphar.2020.613780

62. Nutrition classics. Archives of Internal Medicine, Volume 64, October 1939: Angina pectoris in hereditary xanthomatosis. By Carl Muller. *Nutr Rev*. 1987;45:113-115. doi: 10.1111/j.1753-4887.1987.tb02723.x

63. Khachadurian AK. The Inheritance of Essential Familial Hypercholesterolemia. *Am J Med*. 1964;37:402-407. doi: 10.1016/0002-9343(64)90196-2

64. Skytthe A, Kyvik K, Bathum L, Holm N, Vaupel JW, Christensen K. The Danish Twin Registry in the new millennium. *Twin Res Hum Genet*. 2006;9:763-771. doi: 10.1375/183242706779462732

65. Magnusson PK, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, Halldner L, Lundstrom S, Ullen F, Langstrom N, et al. The Swedish Twin Registry:

establishment of a biobank and other recent developments. *Twin Res Hum Genet*. 2013;16:317-329. doi: 10.1017/thg.2012.104

66. Sahu M, Prasuna JG. Twin Studies: A Unique Epidemiological Tool. *Indian J Community Med*. 2016;41:177-182. doi: 10.4103/0970-0218.183593

67. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. 2015;47:702-709. doi: 10.1038/ng.3285

68. Drobni ZD, Kolossvary M, Karady J, Jermendy AL, Tarnoki AD, Tarnoki DL, Simon J, Szilveszter B, Littvay L, Voros S, et al. Heritability of Coronary Artery Disease: Insights From a Classical Twin Study. *Circ Cardiovasc Imaging*. 2022;15:e013348. doi: 10.1161/CIRCIMAGING.121.013348

69. Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252:247-254. doi: 10.1046/j.1365-2796.2002.01029.x

70. van der Harst P, Verweij N. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ Res*. 2018;122:433-443. doi: 10.1161/CIRCRESAHA.117.312086

71. Aragam KG, Jiang T, Goel A, Kanoni S, Wolford BN, Atri DS, Weeks EM, Wang M, Hindy G, Zhou W, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54:1803-1815. doi: 10.1038/s41588-022-01233-6

72. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, Benner C, Liu D, Locke AE, Balasubramanian S, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature*. 2021;599:628-634. doi: 10.1038/s41586-021-04103-z

73. Koyama S, Ito K, Terao C, Akiyama M, Horikoshi M, Momozawa Y, Matsunaga H, Ieki H, Ozaki K, Onouchi Y, et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet*. 2020;52:1169-1177. doi: 10.1038/s41588-020-0705-3

74. Tcheandjieu C, Zhu X, Hilliard AT, Clarke SL, Napolioni V, Ma S, Lee KM, Fang H, Chen F, Lu Y, et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat Med*. 2022;28:1679-1692. doi: 10.1038/s41591-022-01891-3

75. Tg, Hdl Working Group of the Exome Sequencing Project NHL, Blood I, Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22-31. doi: 10.1056/NEJMoa1307095

76. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, Raychaudhuri S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet*. 2013;45:124-130. doi: 10.1038/ng.2504

77. Pickrell JK. Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. *Am J Hum Genet*. 2014;94:559-573. doi: 10.1016/j.ajhg.2014.03.004

78. de Los Campos G, Grueneberg A, Funkhouser S, Perez-Rodriguez P, Samaddar A. Fine mapping and accurate prediction of complex traits using Bayesian Variable Selection models applied to biobank-size data. *Eur J Hum Genet*. 2023;31:313-320. doi: 10.1038/s41431-022-01135-5

79. Wang QS, Huang H. Methods for statistical fine-mapping and their applications to auto-immune diseases. *Semin Immunopathol*. 2022;44:101-113. doi: 10.1007/s00281-021-00902-8

80. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shoresh N, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet*. 2018;50:621-629. doi: 10.1038/s41588-018-0081-4

81. Chan JC, Piper DE, Cao Q, Liu D, King C, Wang W, Tang J, Liu Q, Higbee J, Xia Z, et al. A proprotein convertase subtilisin/kexin type 9 neutralizing antibody reduces serum cholesterol in mice and nonhuman primates. *Proc Natl Acad Sci U S A*. 2009;106:9820-9825. doi: 10.1073/pnas.0903849106

82. Sabatine MS, Giugliano RP, Keech AC, Honarpour N, Wiviott SD, Murphy SA, Kuder JF, Wang H, Liu T, Wasserman SM, et al. Evolocumab and Clinical Outcomes in Patients with Cardiovascular Disease. *N Engl J Med*. 2017;376:1713-1722. doi: 10.1056/NEJMoa1615664

83. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74. doi: 10.1038/nature11247

84. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, et al. An atlas of active enhancers across human cell types and tissues. *Nature*. 2014;507:455-461. doi: 10.1038/nature12787

85. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, Willer CJ. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics*. 2015;31:2601-2606. doi: 10.1093/bioinformatics/btv201

86. Trynka G, Westra HJ, Slowikowski K, Hu X, Xu H, Stranger BE, Klein RJ, Han B, Raychaudhuri S. Disentangling the Effects of Colocalizing Genomic Annotations

to Functionally Prioritize Non-coding Variants within Complex-Trait Loci. *Am J Hum Genet*. 2015;97:139-152. doi: 10.1016/j.ajhg.2015.05.016

87. Calderon D, Bhaskar A, Knowles DA, Golan D, Raj T, Fu AQ, Pritchard JK. Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *Am J Hum Genet*. 2017;101:686-699. doi: 10.1016/j.ajhg.2017.09.009

88. Huang C, Thompson P, Wang Y, Yu Y, Zhang J, Kong D, Colen RR, Knickmeyer RC, Zhu H, Alzheimer's Disease Neuroimaging I. FGWAS: Functional genome wide association analysis. *Neuroimage*. 2017;159:107-121. doi: 10.1016/j.neuroimage.2017.07.030

89. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, Timpson NJ, Consortium UK, Dunham I, Birney E, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet*. 2019;51:343-353. doi: 10.1038/s41588-018-0322-6

90. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467:832-838. doi: 10.1038/nature09410

91. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet*. 2012;91:1011-1021. doi: 10.1016/j.ajhg.2012.10.010

92. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, Patterson N, Daly MJ, Price AL, Neale BM. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291-295. doi: 10.1038/ng.3211

93. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47:1228-1235. doi: 10.1038/ng.3404

94. Cano-Gamez E, Trynka G. From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front Genet*. 2020;11:424. doi: 10.3389/fgene.2020.00424

95. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501:506-511. doi: 10.1038/nature12531

96. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45:1238-1243. doi: 10.1038/ng.2756

97. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet*. 2019;51:768-769. doi: 10.1038/s41588-019-0404-0

98. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10:e1004383. doi: 10.1371/journal.pgen.1004383

99. Giambartolomei C, Zhenli Liu J, Zhang W, Hauberg M, Shi H, Boocock J, Pickrell J, Jaffe AE, CommonMind C, Pasaniuc B, et al. A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics*. 2018;34:2538-2545. doi: 10.1093/bioinformatics/bty147

100. Foley CN, Staley JR, Breen PG, Sun BB, Kirk PDW, Burgess S, Howson JMM. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun*. 2021;12:764. doi: 10.1038/s41467-020-20885-8

101. Porcu E, Rueger S, Lepik K, e QC, Consortium B, Santoni FA, Reymond A, Kutalik Z. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat Commun*. 2019;10:3300. doi: 10.1038/s41467-019-10936-0

102. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48:481-487. doi: 10.1038/ng.3538

103. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, Lovering RC, Tajuddin SM, Winkler TW, Graff M, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun*. 2018;9:5141. doi: 10.1038/s41467-018-07340-5

104. Yao DW, O'Connor LJ, Price AL, Gusev A. Quantifying genetic effects on disease mediated by assayed gene expression levels. *Nat Genet*. 2020;52:626-633. doi: 10.1038/s41588-020-0625-2

105. Umans BD, Battle A, Gilad Y. Where Are the Disease-Associated eQTLs? *Trends Genet*. 2021;37:109-124. doi: 10.1016/j.tig.2020.08.009

106. Mostafavi H, Spence JP, Naqvi S, Pritchard JK. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet*. 2023;55:1866-1875. doi: 10.1038/s41588-023-01529-1

107. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013;8:2180-2196. doi: 10.1038/nprot.2013.132

108. Li K, Liu Y, Cao H, Zhang Y, Gu Z, Liu X, Yu A, Kaphle P, Dickerson KE, Ni M, et al. Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing. *Nat Commun*. 2020;11:485. doi: 10.1038/s41467-020-14362-5

109. Fulco CP, Nasser J, Jones TR, Munson G, Bergman DT, Subramanian V, Grossman SR, Anyoha R, Doughty BR, Patwardhan TA, et al. Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. 2019;51:1664-1669. doi: 10.1038/s41588-019-0538-0

110. Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, Jones TR, Nguyen TH, Ulirsch JC, Lekschas F, et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature*. 2021;593:238-243. doi: 10.1038/s41586-021-03446-x

111. Koplev S, Seldin M, Sukhavasi K, Ermel R, Pang S, Zeng L, Bankier S, Di Narzo A, Cheng H, Meda V, et al. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat Cardiovasc Res*. 2022;1:85-100. doi: 10.1038/s44161-021-00009-1

112. Schnitzler GR, Kang H, Fang S, Angom RS, Lee-Kim VS, Ma XR, Zhou R, Zeng T, Guo K, Taylor MS, et al. Convergence of coronary artery disease genes onto endothelial cell programs. *Nature*. 2024;626:799-807. doi: 10.1038/s41586-024-07022-x

113. Reschen ME, Gaulton KJ, Lin D, Soilleux EJ, Morris AJ, Smyth SS, O'Callaghan CA. 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:e1005061. doi: 10.1371/journal.pgen.1005061

114. Chapman MJ, Goldstein S, Lagrange D, Laplaud PM. A density gradient ultracentrifugal procedure for the isolation of the major lipoprotein classes from human serum. *J Lipid Res*. 1981;22:339-358.

115. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods*. 2017;14:959-962. doi: 10.1038/nmeth.4396

116. FastQC. In; 2015.

117. Gaspar JM. NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors. *BMC Bioinformatics*. 2018;19:536. doi: 10.1186/s12859-018-2579-2

118. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21. doi: 10.1093/bioinformatics/bts635

119. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079. doi: 10.1093/bioinformatics/btp352

120. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41:e108. doi: 10.1093/nar/gkt214

121. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139-140. doi: 10.1093/bioinformatics/btp616

122. Fang H, Knezevic B, Burnham KL, Knight JC. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*. 2016;8:129. doi: 10.1186/s13073-016-0384-y

123. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357-359. doi: 10.1038/nmeth.1923

124. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841-842. doi: 10.1093/bioinformatics/btq033

125. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137. doi: 10.1186/gb-2008-9-9-r137

126. Lun AT, Smyth GK. csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res*. 2016;44:e45. doi: 10.1093/nar/gkv1191

127. Qin Q, Fan J, Zheng R, Wan C, Mei S, Wu Q, Sun H, Brown M, Zhang J, Meyer CA, et al. Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. *Genome Biol*. 2020;21:32. doi: 10.1186/s13059-020-1934-6

128. Wang C, Sun D, Huang X, Wan C, Li Z, Han Y, Qin Q, Fan J, Qiu X, Xie Y, et al. Integrative analyses of single-cell transcriptome and regulome using MAESTRO. *Genome Biol*. 2020;21:198. doi: 10.1186/s13059-020-02116-x

129. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nat Commun*. 2020;11:4267. doi: 10.1038/s41467-020-18035-1

130. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Cheneby J, Kulkarni SR, Tan G, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res*. 2018;46:D1284. doi: 10.1093/nar/gkx1188

131. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010;28:495-501. doi: 10.1038/nbt.1630

132. Chalisey A. The cellular response to atherogenic lipid. In: University of Oxford; 2019.

133. Jiang J, Hiron TK, Agbaedeng T, Malhotra Y, Drydale E, Bancroft J, Ng E, Reschen ME, Davison LJ, O'Callaghan CA. A Novel Macrophage Subpopulation Conveys Increased Genetic Risk of Coronary Artery Disease. *Circ Res*. 2024. doi: 10.1161/CIRCRESAHA.123.324172

134. Dib L, Koneva LA, Edsfeldt A, Zurke Y-X, Sun J, Nitulescu M, Attar M, Lutgens E, Schmidt S, Lindholm MW, et al. Lipid-associated macrophages transition to an inflammatory state in human atherosclerosis, increasing the risk of cerebrovascular complications. *Nature Cardiovascular Research*. 2023. doi: 10.1038/s44161-023-00295-x

135. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177:1888-1902 e1821. doi: 10.1016/j.cell.2019.05.031

136. McGinnis CS, Murrow LM, Gartner ZJ. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst*. 2019;8:329-337 e324. doi: 10.1016/j.cels.2019.03.003

137. Tang M, Kaymaz Y, Logeman BL, Eichhorn S, Liang ZS, Dulac C, Sackton TB. Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics*. 2021;37:2212-2214. doi: 10.1093/bioinformatics/btaa956

138. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol*. 2021;17:e9620. doi: 10.15252/msb.20209620

139. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573-3587 e3529. doi: 10.1016/j.cell.2021.04.048

140. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6:e21800. doi: 10.1371/journal.pone.0021800

141. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*. 2017;14:975-978. doi: 10.1038/nmeth.4401

142. Kartha VK, Duarte FM, Hu Y, Ma S, Chew JG, Lareau CA, Earl A, Burkett ZD, Kohlway AS, Lebofsky R, et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom*. 2022;2. doi: 10.1016/j.xgen.2022.100166

143. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382-2383. doi: 10.1093/bioinformatics/btv145

144. Alasoo K, Rodrigues J, Mukhopadhyay S, Knights AJ, Mann AL, Kundu K, Consortium H, Hale C, Dougan G, Gaffney DJ. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet*. 2018;50:424-431. doi: 10.1038/s41588-018-0046-7

145. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575. doi: 10.1086/519795

146. Fang H, Consortium U-D, De Wolf H, Knezevic B, Burnham KL, Osgood J, Sanniti A, Lledo Lara A, Kasela S, De Cesco S, et al. A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat Genet*. 2019;51:1082-1091. doi: 10.1038/s41588-019-0456-1

147. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, Schoech A, Bulik-Sullivan B, Neale BM, Gusev A, et al. Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nat Genet*. 2017;49:1421-1427. doi: 10.1038/ng.3954

148. Gordon S, Taylor PR. Monocyte and macrophage heterogeneity. *Nat Rev Immunol*. 2005;5:953-964. doi: 10.1038/nri1733

149. Guilliams M, Mildner A, Yona S. Developmental and Functional Heterogeneity of Monocytes. *Immunity*. 2018;49:595-613. doi: 10.1016/j.immuni.2018.10.005

150. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep*. 2014;6:13. doi: 10.12703/P6-13

151. Hume DA, Millard SM, Pettit AR. Macrophage heterogeneity in the single-cell era: facts and artifacts. *Blood*. 2023;142:1339-1347. doi: 10.1182/blood.2023020597

152. Mass E, Nimmerjahn F, Kierdorf K, Schlitzer A. Tissue-specific macrophages: how they develop and choreograph tissue biology. *Nat Rev Immunol*. 2023;23:563-579. doi: 10.1038/s41577-023-00848-y

153. Yu XH, Fu YC, Zhang DW, Yin K, Tang CK. Foam cells in atherosclerosis. *Clin Chim Acta*. 2013;424:245-252. doi: 10.1016/j.cca.2013.06.006

154. Williams JW, Zaitsev K, Kim KW, Ivanov S, Saunders BT, Schrank PR, Kim K, Elvington A, Kim SH, Tucker CG, et al. Limited proliferation capacity of aortic intima resident macrophages requires monocyte recruitment for atherosclerotic plaque progression. *Nat Immunol*. 2020;21:1194-1204. doi: 10.1038/s41590-020-0768-4

155. Das R, Ganapathy S, Mahabeleshwar GH, Drumm C, Febbraio M, Jain MK, Plow EF. Macrophage gene expression and foam cell formation are regulated by plasminogen. *Circulation*. 2013;127:1209-1218, e1201-1216. doi: 10.1161/CIRCULATIONAHA.112.001214

156. Reschen ME, Lin D, Chalisey A, Soilleux EJ, O'Callaghan CA. Genetic and environmental risk factors for atherosclerosis regulate transcription of phosphatase and actin regulating gene PHACTR1. *Atherosclerosis*. 2016;250:95-105. doi: 10.1016/j.atherosclerosis.2016.04.025

157. Domschke G, Linden F, Pawig L, Hafner A, Akhavanpoor M, Reymann J, Doesch AO, Erbel C, Weber C, Katus HA, et al. Systematic RNA-interference in primary human monocyte-derived macrophages: A high-throughput platform to study foam cell formation. *Sci Rep*. 2018;8:10516. doi: 10.1038/s41598-018-28790-3

158. Jin X, Kruth HS. Culture of Macrophage Colony-stimulating Factor Differentiated Human Monocyte-derived Macrophages. *J Vis Exp*. 2016. doi: 10.3791/54244

159. Martinez FO, Gordon S, Locati M, Mantovani A. Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. *J Immunol*. 2006;177:7303-7311. doi: 10.4049/jimmunol.177.10.7303

160. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, et al. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*. 2014;40:274-288. doi: 10.1016/j.immuni.2014.01.006

161. Hiron TK. Environmental and genetic influences on atherosclerosis. In: University of Oxford; 2019.

162. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, Hudelle R, Qaiser T, Matson KJE, Barraud Q, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun*. 2021;12:5692. doi: 10.1038/s41467-021-25960-2

163. Park YM. CD36, a scavenger receptor implicated in atherosclerosis. *Exp Mol Med*. 2014;46:e99. doi: 10.1038/emm.2014.38

164. Chistiakov DA, Bobryshev YV, Orekhov AN. Changes in transcriptome of macrophages in atherosclerosis. *J Cell Mol Med*. 2015;19:1163-1173. doi: 10.1111/jcmm.12591

165. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lonnerberg P, Furlan A, et al. RNA velocity of single cells. *Nature*. 2018;560:494-498. doi: 10.1038/s41586-018-0414-6

166. Georgess D, Mazzorana M, Terrado J, Delprat C, Chamot C, Guasch RM, Perez-Roger I, Jurdic P, Machuca-Gayet I. Comparative transcriptomics reveals RhoE as a novel regulator of actin dynamics in bone-resorbing osteoclasts. *Mol Biol Cell*. 2014;25:380-396. doi: 10.1091/mbc.E13-07-0363

167. Rouland A, Masson D, Lagrost L, Verges B, Gautier T, Bouillet B. Role of apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: a systematic review. *Cardiovasc Diabetol*. 2022;21:272. doi: 10.1186/s12933-022-01703-5

168. Nijweide PJ, Burger EH, Feyen JH. Cells of bone: proliferation, differentiation, and hormonal regulation. *Physiol Rev*. 1986;66:855-886. doi: 10.1152/physrev.1986.66.4.855

169. Glantschnig H, Fisher JE, Wesolowski G, Rodan GA, Reszka AA. M-CSF, TNFalpha and RANK ligand promote osteoclast survival by signaling through mTOR/S6 kinase. *Cell Death Differ*. 2003;10:1165-1177. doi: 10.1038/sj.cdd.4401285

170. Huang HC, Nguyen T, Pickett CB. Regulation of the antioxidant response element by protein kinase C-mediated phosphorylation of NF-E2-related factor 2. *Proc Natl Acad Sci U S A*. 2000;97:12475-12480. doi: 10.1073/pnas.220418997

171. Zhou LZ, Johnson AP, Rando TA. NF kappa B and AP-1 mediate transcriptional responses to oxidative stress in skeletal muscle cells. *Free Radic Biol Med*. 2001;31:1405-1416. doi: 10.1016/s0891-5849(01)00719-5

172. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, Ding J, Brack A, Kartha VK, Tay T, et al. Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*. 2020;183:1103-1116 e1120. doi: 10.1016/j.cell.2020.09.056

173. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, et al. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity*. 2010;32:317-328. doi: 10.1016/j.immuni.2010.02.008

174. Yao L, Liang J, Ozer A, Leung AK, Lis JT, Yu H. A comparison of experimental assays and analytical methods for genome-wide identification of active enhancers. *Nat Biotechnol*. 2022;40:1056-1065. doi: 10.1038/s41587-022-01211-7

175. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank*. 2015;13:311-319. doi: 10.1089/bio.2015.0032

176. Vosa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53:1300-1310. doi: 10.1038/s41588-021-00913-z

177. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016;167:1369-1384 e1319. doi: 10.1016/j.cell.2016.09.037

178. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010;107:21931-21936. doi: 10.1073/pnas.1016071107

179. Zhu Y, Sun L, Chen Z, Whitaker JW, Wang T, Wang W. Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res*. 2013;41:10032-10043. doi: 10.1093/nar/gkt826

180. Li Z, Nagai JS, Kuppe C, Kramann R, Costa IG. scMEGA: single-cell multi-omic enhancer-based gene regulatory network inference. *Bioinform Adv*. 2023;3:vbad003. doi: 10.1093/bioadv/vbad003

181. Bravo Gonzalez-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, Poovathingal S, Wouters J, Aibar S, Aerts S. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods*. 2023;20:1355-1367. doi: 10.1038/s41592-023-01938-4

182. Kim D, Tran A, Kim HJ, Lin Y, Yang JYH, Yang P. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. *NPJ Syst Biol Appl*. 2023;9:51. doi: 10.1038/s41540-023-00312-6

183. Mitra S, Malik R, Wong W, Rahman A, Hartemink AJ, Pritykin Y, Dey KK, Leslie CS. Single-cell multi-ome regression models identify functional and disease-associated enhancers and enable chromatin potential analysis. *Nat Genet*. 2024;56:627-636. doi: 10.1038/s41588-024-01689-8

184. Witztum JL, Lichtman AH. The influence of innate and adaptive immune responses on atherosclerosis. *Annu Rev Pathol*. 2014;9:73-102. doi: 10.1146/annurev-pathol-020712-163936

185. Ord T, Ounap K, Stolze LK, Aherrahrou R, Nurminen V, Toropainen A, Selvarajan I, Lonnberg T, Aavik E, Yla-Herttuala S, et al. Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. *Circ Res*. 2021;129:240-258. doi: 10.1161/CIRCRESAHA.121.318971

186. Slenders L, Landsmeer LPL, Cui K, Depuydt MAC, Verwer M, Mekke J, Timmerman N, van den Dungen NAM, Kuiper J, de Winther MPJ, et al. Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. *Eur Heart J Open*. 2022;2:oeab043. doi: 10.1093/ehjopen/oeab043

187. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S, et al. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*. 2016;167:657-669 e621. doi: 10.1016/j.cell.2016.09.025

188. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934-947. doi: 10.1016/j.cell.2013.09.053

189. Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, Kane M, Cleary B, Lander ES, Engreitz JM. Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*. 2016;354:769-773. doi: 10.1126/science.aag2445

190. Khera AV, Kathiresan S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18:331-344. doi: 10.1038/nrg.2016.160

191. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190-1195. doi: 10.1126/science.1222794

192. Zeggini E, Gloyn AL, Barton AC, Wain LV. Translational genomics and precision medicine: Moving from the lab to the clinic. *Science*. 2019;365:1409-1413. doi: 10.1126/science.aax4588

193. Ying P, Chen C, Lu Z, Chen S, Zhang M, Cai Y, Zhang F, Huang J, Fan L, Ning C, et al. Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk. *Nat Commun*. 2023;14:5958. doi: 10.1038/s41467-023-41690-z

194. Hekselman I, Yeger-Lotem E. Mechanisms of tissue and cell-type specificity in heritable traits and diseases. *Nat Rev Genet*. 2020;21:137-150. doi: 10.1038/s41576-019-0200-9

195. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol Cell*. 2018;71:858-871 e858. doi: 10.1016/j.molcel.2018.06.044

196. Qiu X, Rahimzamani A, Wang L, Ren B, Mao Q, Durham T, McFaline-Figueroa JL, Saunders L, Trapnell C, Kannan S. Inferring Causal Gene Regulatory Networks from Coupled Single-Cell Expression Dynamics Using Scribe. *Cell Syst*. 2020;10:265-274 e211. doi: 10.1016/j.cels.2020.02.003

197. Jagadeesh KA, Dey KK, Montoro DT, Mohan R, Gazal S, Engreitz JM, Xavier RJ, Price AL, Regev A. Identifying disease-critical cell types and cellular processes by integrating single-cell RNA-sequencing and human genetics. *Nat Genet*. 2022;54:1479-1492. doi: 10.1038/s41588-022-01187-9

198. Watanabe K, Umicevic Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun*. 2019;10:3222. doi: 10.1038/s41467-019-11181-1

199. Kim SS, Jagadeesh K, Dey KK, Shen AZ, Raychaudhuri S, Kellis M, Price AL. Leveraging single-cell ATAC-seq to identify disease-critical fetal and adult brain cell types. *bioRxiv*. 2021:2021.2005.2020.445067. doi: 10.1101/2021.05.20.445067

200. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, Kohlmaier A, Herbst A, Northoff BH, Nicolaou A, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun*. 2016;7:12429. doi: 10.1038/ncomms12429

201. Hocker JD, Poirion OB, Zhu F, Buchanan J, Zhang K, Chiou J, Wang TM, Zhang Q, Hou X, Li YE, et al. Cardiac cell type-specific gene regulatory programs and disease risk association. *Sci Adv*. 2021;7. doi: 10.1126/sciadv.abf1444

202. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 2019;20:207-220. doi: 10.1038/s41576-018-0089-8

203. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24:R102-110. doi: 10.1093/hmg/ddv259

204. Mountjoy E, Schmidt EM, Carmona M, Schwartzentruber J, Peat G, Miranda A, Fumis L, Hayhurst J, Buniello A, Karim MA, et al. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat Genet*. 2021;53:1527-1533. doi: 10.1038/s41588-021-00945-5

205. Dong S, Boyle AP. Prioritization of regulatory variants with tissue-specific function in the non-coding regions of human genome. *Nucleic Acids Res*. 2022;50:e6. doi: 10.1093/nar/gkab924

206. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826. doi: 10.1038/s41467-017-01261-5

207. Weissbrod O, Hormozdiari F, Benner C, Cui R, Ulirsch J, Gazal S, Schoech AP, van de Geijn B, Reshef Y, Marquez-Luna C, et al. Functionally informed fine-mapping and polygenic localization of complex trait heritability. *Nat Genet*. 2020;52:1355-1363. doi: 10.1038/s41588-020-00735-5

208. Bjorkegren JLM, Kovacic JC, Dudley JT, Schadt EE. 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;65:830-845. doi: 10.1016/j.jacc.2014.12.033

209. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. Genetics of gene expression and its effect on disease. *Nature*. 2008;452:423-428. doi: 10.1038/nature06758

210. Xu K, Li B, McGinnis KA, Vickers-Smith R, Dao C, Sun N, Kember RL, Zhou H, Becker WC, Gelernter J, et al. Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nat Commun*. 2020;11:5302. doi: 10.1038/s41467-020-18489-3

211. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, et al. Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27:3641-3649. doi: 10.1093/hmg/ddy271

212. Koplev S, Seldin M, Sukhavasi K, Ermel R, Pang S, Zeng L, Bankier S, Di Narzo A, Cheng H, Meda V, et al. A mechanistic framework for cardiometabolic and coronary artery diseases. *Nature Cardiovascular Research*. 2022;1:85-100. doi: 10.1038/s44161-021-00009-1

213. Wang B, Tontonoz P. Liver X receptors in lipid signalling and membrane homeostasis. *Nat Rev Endocrinol*. 2018;14:452-463. doi: 10.1038/s41574-018-0037-x

214. Nicholson AC, Frieda S, Pearce A, Silverstein RL. Oxidized LDL binds to CD36 on human monocyte-derived macrophages and transfected cell lines. Evidence implicating the lipid moiety of the lipoprotein as the binding site. *Arterioscler Thromb Vasc Biol*. 1995;15:269-275. doi: 10.1161/01.atv.15.2.269

215. Hilgendorf I, Swirski FK, Robbins CS. Monocyte fate in atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015;35:272-279. doi: 10.1161/ATVBAHA.114.303565

216. Eraslan G, Drokhlyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, Wang J, Van Wittenberghe N, Rouhana JM, Waldman J, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science*. 2022;376:eabl4290. doi: 10.1126/science.abl4290

217. Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, Satpathy AT, Kartha VK, Salem RM, Hirschhorn JN, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet*. 2019;51:683-693. doi: 10.1038/s41588-019-0362-6

218. Guilliams M, Bonnardel J, Haest B, Vanderborght B, Wagner C, Remmerie A, Bujko A, Martens L, Thone T, Browaeys R, et al. Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell*. 2022;185:379-396 e338. doi: 10.1016/j.cell.2021.12.018

219. Basquiera AL, Berretta AR, Garcia JJ, Palazzo ED. Coronary ischemia related to alemtuzumab therapy. *Ann Oncol*. 2004;15:539-540. doi: 10.1093/annonc/mdh114

220. Arnheiter H. The discovery of the microphthalmia locus and its gene, Mitf. *Pigment Cell Melanoma Res*. 2010;23:729-735. doi: 10.1111/j.1755-148X.2010.00759.x

221. Goding CR, Arnheiter H. MITF-the first 25 years. *Genes Dev*. 2019;33:983-1007. doi: 10.1101/gad.324657.119

222. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature*. 2005;436:117-122. doi: 10.1038/nature03664

223. Guhan SM, Artomov M, McCormick S, Njauw C, Stratigos AJ, Shannon K, Ellisen LW, Tsao H. Cancer risks associated with the germline MITF(E318K) variant. *Sci Rep*. 2020;10:17051. doi: 10.1038/s41598-020-74237-z

224. Lee A, Lim J, Lim JS. Emerging roles of MITF as a crucial regulator of immunity. *Exp Mol Med*. 2024;56:311-318. doi: 10.1038/s12276-024-01175-5

225. Perera RM, Stoykova S, Nicolay BN, Ross KN, Fitamant J, Boukhali M, Lengrand J, Deshpande V, Selig MK, Ferrone CR, et al. Transcriptional control of autophagy-lysosome function drives pancreatic cancer metabolism. *Nature*. 2015;524:361-365. doi: 10.1038/nature14587

226. Ploper D, Taelman VF, Robert L, Perez BS, Titz B, Chen HW, Graeber TG, von Euw E, Ribas A, De Robertis EM. MITF drives endolysosomal biogenesis and potentiates Wnt signaling in melanoma cells. *Proc Natl Acad Sci U S A*. 2015;112:E420-429. doi: 10.1073/pnas.1424576112

227. Lu H, Fan Y, Qiao C, Liang W, Hu W, Zhu T, Zhang J, Chen YE. TFEB inhibits endothelial cell inflammation and reduces atherosclerosis. *Sci Signal*. 2017;10. doi: 10.1126/scisignal.aah4214

228. Wu SA, Kersten S, Qi L. Lipoprotein Lipase and Its Regulators: An Unfolding Story. *Trends Endocrinol Metab*. 2021;32:48-61. doi: 10.1016/j.tem.2020.11.005

229. Kumari A, Kristensen KK, Ploug M, Winther AL. The Importance of Lipoprotein Lipase Regulation in Atherosclerosis. *Biomedicines*. 2021;9. doi: 10.3390/biomedicines9070782

230. Kobayashi J, Hashimoto H, Fukamachi I, Tashiro J, Shirai K, Saito Y, Yoshida S. Lipoprotein lipase mass and activity in severe hypertriglyceridemia. *Clin Chim Acta*. 1993;216:113-123. doi: 10.1016/0009-8981(93)90144-s

231. Ference BA, Kastelein JJP, Ray KK, Ginsberg HN, Chapman MJ, Packard CJ, Laufs U, Oliver-Williams C, Wood AM, Butterworth AS, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C-Lowering LDLR Variants With Risk of Coronary Heart Disease. *JAMA*. 2019;321:364-373. doi: 10.1001/jama.2018.20045

232. Gustafsson M, Levin M, Skalen K, Perman J, Friden V, Jirholt P, Olofsson SO, Fazio S, Linton MF, Semenkovich CF, et al. Retention of low-density lipoprotein in atherosclerotic lesions of the mouse: evidence for a role of lipoprotein lipase. *Circ Res*. 2007;101:777-783. doi: 10.1161/CIRCRESAHA.107.149666

233. Takahashi M, Yagyu H, Tazoe F, Nagashima S, Ohshiro T, Okada K, Osuga J, Goldberg IJ, Ishibashi S. Macrophage lipoprotein lipase modulates the development of atherosclerosis but not adiposity. *J Lipid Res*. 2013;54:1124-1134. doi: 10.1194/jlr.M035568

234. Chang HR, Josefs T, Scerbo D, Gumaste N, Hu Y, Huggins LA, Barrett TJ, Chiang SS, Grossman J, Bagdasarov S, et al. Role of LpL (Lipoprotein Lipase) in Macrophage Polarization In Vitro and In Vivo. *Arterioscler Thromb Vasc Biol*. 2019;39:1967-1985. doi: 10.1161/ATVBAHA.119.312389

235. Tsvetkov P, Coy S, Petrova B, Dreishpoon M, Verma A, Abdusamad M, Rossen J, Joesch-Cohen L, Humeidi R, Spangler RD, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins. *Science*. 2022;375:1254-1261. doi: 10.1126/science.abf0529

236. Sheftel AD, Stehling O, Pierik AJ, Elsasser HP, Muhlenhoff U, Webert H, Hobler A, Hannemann F, Bernhardt R, Lill R. Humans possess two mitochondrial

ferredoxins, Fdx1 and Fdx2, with distinct roles in steroidogenesis, heme, and Fe/S cluster biosynthesis. *Proc Natl Acad Sci U S A*. 2010;107:11775-11780. doi: 10.1073/pnas.1004250107

237. Arroyo JD, Jourdain AA, Calvo SE, Ballarano CA, Doench JG, Root DE, Mootha VK. A Genome-wide CRISPR Death Screen Identifies Genes Essential for Oxidative Phosphorylation. *Cell Metab*. 2016;24:875-885. doi: 10.1016/j.cmet.2016.08.017

238. Dreishpoon MB, Bick NR, Petrova B, Warui DM, Cameron A, Booker SJ, Kanarek N, Golub TR, Tsvetkov P. FDX1 regulates cellular protein lipoylation through direct binding to LIAS. *J Biol Chem*. 2023;299:105046. doi: 10.1016/j.jbc.2023.105046

239. Madamanchi NR, Runge MS. Mitochondrial dysfunction in atherosclerosis. *Circ Res*. 2007;100:460-473. doi: 10.1161/01.RES.0000258450.44413.96

240. Wang H, Eckel RH. Lipoprotein lipase: from gene to obesity. *Am J Physiol Endocrinol Metab*. 2009;297:E271-288. doi: 10.1152/ajpendo.90920.2008

241. Wolf BK, Zhao Y, McCray A, Hawk WH, Deary LT, Sugiarto NW, LaCroix IS, Gerber SA, Cheng C, Wang X. Cooperation of chromatin remodeling SWI/SNF complex and pioneer factor AP-1 shapes 3D enhancer landscapes. *Nat Struct Mol Biol*. 2023;30:10-21. doi: 10.1038/s41594-022-00880-x

242. Turner AW, Hu SS, Mosquera JV, Ma WF, Hodonsky CJ, Wong D, Auguste G, Song Y, Sol-Church K, Farber E, et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat Genet*. 2022;54:804-816. doi: 10.1038/s41588-022-01069-0

243. Clark SC, Chereji RV, Lee PR, Fields RD, Clark DJ. Differential nucleosome spacing in neurons and glia. *Neurosci Lett*. 2020;714:134559. doi: 10.1016/j.neulet.2019.134559

244. Vierbuchen T, Ling E, Cowley CJ, Couch CH, Wang X, Harmin DA, Roberts CWM, Greenberg ME. AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. *Mol Cell*. 2017;68:1067-1082 e1012. doi: 10.1016/j.molcel.2017.11.026

245. Mohibi S, Zhang Y, Perng V, Chen M, Zhang J, Chen X. Ferredoxin 1 is essential for embryonic development and lipid homeostasis. *Elife*. 2024;13. doi: 10.7554/eLife.91656

246. Sadhu N, Dalan R, Jain PR, Lee CJM, Pakkiri LS, Tay KY, Mina TH, Low D, Min Y, Ackers-Johnson M, et al. Metabolome-wide association of carotid intima media thickness identifies FDX1 as a determinant of cholesterol metabolism and cardiovascular risk in Asian populations. *medRxiv*. 2024:2024.2005.2014.24307316. doi: 10.1101/2024.05.14.24307316

247. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*. 2005;352:1685-1695. doi: 10.1056/NEJMra043430

248. Virmani R, Kolodgie FD, Burke AP, Farb A, Schwartz SM. Lessons from sudden coronary death: a comprehensive morphological classification scheme for atherosclerotic lesions. *Arterioscler Thromb Vasc Biol*. 2000;20:1262-1275. doi: 10.1161/01.atv.20.5.1262

249. Chevre R, Gonzalez-Granado JM, Megens RT, Sreeramkumar V, Silvestre-Roig C, Molina-Sanchez P, Weber C, Soehnlein O, Hidalgo A, Andres V. High-resolution imaging of intravascular atherogenic inflammation in live mice. *Circ Res*. 2014;114:770-779. doi: 10.1161/CIRCRESAHA.114.302590

250. Pober JS, Sessa WC. Evolving functions of endothelial cells in inflammation. *Nat Rev Immunol*. 2007;7:803-815. doi: 10.1038/nri2171

251. Sawamura T, Kume N, Aoyama T, Moriwaki H, Hoshikawa H, Aiba Y, Tanaka T, Miwa S, Katsura Y, Kita T, et al. An endothelial receptor for oxidized low-density lipoprotein. *Nature*. 1997;386:73-77. doi: 10.1038/386073a0

252. Jiang H, Zhou Y, Nabavi SM, Sahebkar A, Little PJ, Xu S, Weng J, Ge J. Mechanisms of Oxidized LDL-Mediated Endothelial Dysfunction and Its Consequences for the Development of Atherosclerosis. *Front Cardiovasc Med*. 2022;9:925923. doi: 10.3389/fcvm.2022.925923

253. Li D, Mehta JL. Antisense to LOX-1 inhibits oxidized LDL-mediated upregulation of monocyte chemoattractant protein-1 and monocyte adhesion to human coronary artery endothelial cells. *Circulation*. 2000;101:2889-2895. doi: 10.1161/01.cir.101.25.2889

254. Sohrabi Y, Lagache SMM, Voges VC, Semo D, Sonntag G, Hanemann I, Kahles F, Waltenberger J, Findeisen HM. OxLDL-mediated immunologic memory in endothelial cells. *J Mol Cell Cardiol*. 2020;146:121-132. doi: 10.1016/j.yjmcc.2020.07.006

255. Zhu H, Xia M, Hou M, Tang Z, Li Y, Ma J, Ling W. Ox-LDL plays dual effect in modulating expression of inflammatory molecules through LOX-1 pathway in human umbilical vein endothelial cells. *Front Biosci*. 2005;10:2585-2594. doi: 10.2741/1722

256. Takei A, Huang Y, Lopes-Virella MF. Expression of adhesion molecules by human endothelial cells exposed to oxidized low density lipoprotein. Influences of degree of oxidation and location of oxidized LDL. *Atherosclerosis*. 2001;154:79-86. doi: 10.1016/s0021-9150(00)00465-2

257. Dwivedi A, Anggard EE, Carrier MJ. Oxidized LDL-mediated monocyte adhesion to endothelial cells does not involve NFkappaB. *Biochem Biophys Res Commun*. 2001;284:239-244. doi: 10.1006/bbrc.2001.4955

258. Khan BV, Parthasarathy SS, Alexander RW, Medford RM. Modified low density lipoprotein and its constituents augment cytokine-activated vascular cell adhesion molecule-1 gene expression in human vascular endothelial cells. *J Clin Invest*. 1995;95:1262-1270. doi: 10.1172/JCI117776

259. Chen C, Khismatullin DB. Oxidized low-density lipoprotein contributes to atherogenesis via co-activation of macrophages and mast cells. *PLoS One*. 2015;10:e0123088. doi: 10.1371/journal.pone.0123088

260. Amberger A, Maczek C, Jurgens G, Michaelis D, Schett G, Trieb K, Eberl T, Jindal S, Xu Q, Wick G. Co-expression of ICAM-1, VCAM-1, ELAM-1 and Hsp60 in human arterial and venous endothelial cells in response to cytokines and oxidized low-density lipoproteins. *Cell Stress Chaperones*. 1997;2:94-103. doi: 10.1379/1466-1268(1997)002<0094:ceoive>2.3.co;2

261. Mattaliano MD, Huard C, Cao W, Hill AA, Zhong W, Martinez RV, Harnish DC, Paulsen JE, Shih HH. LOX-1-dependent transcriptional regulation in response to oxidized LDL treatment of human aortic endothelial cells. *Am J Physiol Cell Physiol*. 2009;296:C1329-1337. doi: 10.1152/ajpcell.00513.2008

262. Watson AD, Leitinger N, Navab M, Faull KF, Horkko S, Witztum JL, Palinski W, Schwenke D, Salomon RG, Sha W, et al. Structural identification by mass spectrometry of oxidized phospholipids in minimally oxidized low density lipoprotein that induce monocyte/endothelial interactions and evidence for their presence in vivo. *J Biol Chem*. 1997;272:13597-13607. doi: 10.1074/jbc.272.21.13597

263. Chu LH, Indramohan M, Ratsimandresey RA, Gangopadhyay A, Morris EP, Monack DM, Dorfleutner A, Stehlik C. The oxidized phospholipid oxPAPC protects from septic shock by targeting the non-canonical inflammasome in macrophages. *Nat Commun*. 2018;9:996. doi: 10.1038/s41467-018-03409-3

264. Appleton BD, Palmer SA, Smith HP, Stephens LE, Major AS. Oxidized Phospholipid oxPAPC Alters Regulatory T-Cell Differentiation and Decreases Their Protective Function in Atherosclerosis in Mice. *Arterioscler Thromb Vasc Biol*. 2023;43:2119-2132. doi: 10.1161/ATVBAHA.123.319674

265. Romanoski CE, Che N, Yin F, Mai N, Pouldar D, Civelek M, Pan C, Lee S, Vakili L, Yang WP, et al. Network for activation of human endothelial cells by oxidized phospholipids: a critical role of heme oxygenase 1. *Circ Res*. 2011;109:e27-41. doi: 10.1161/CIRCRESAHA.111.241869

266. Hogan NT, Whalen MB, Stolze LK, Hadeli NK, Lam MT, Springstead JR, Glass CK, Romanoski CE. Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. *Elife*. 2017;6. doi: 10.7554/eLife.22536

267. Hitzel J, Lee E, Zhang Y, Bibli SI, Li X, Zukunft S, Pfluger B, Hu J, Schurmann C, Vasconez AE, et al. Oxidized phospholipids regulate amino acid metabolism through MTHFD2 to facilitate nucleotide release in endothelial cells. *Nat Commun*. 2018;9:2292. doi: 10.1038/s41467-018-04602-0

268. Abumrad NA, Cabodevilla AG, Samovski D, Pietka T, Basu D, Goldberg IJ. Endothelial Cell Receptors in Tissue Lipid Uptake and Metabolism. *Circ Res*. 2021;128:433-450. doi: 10.1161/CIRCRESAHA.120.318003

269. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res*. 2012;40:4288-4297. doi: 10.1093/nar/gks042

270. Moore JM, Bell EL, Hughes RO, Garfield AS. ABC transporters: human disease and pharmacotherapeutic potential. *Trends Mol Med*. 2023;29:152-172. doi: 10.1016/j.molmed.2022.11.001

271. Ade N, Leon F, Pallardy M, Peiffer JL, Kerdine-Romer S, Tissier MH, Bonnet PA, Fabre I, Ourlin JC. HMOX1 and NQO1 genes are upregulated in response to contact sensitizers in dendritic cells and THP-1 cell line: role of the Keap1/Nrf2 pathway. *Toxicol Sci*. 2009;107:451-460. doi: 10.1093/toxsci/kfn243

272. Perera RM, Di Malta C, Ballabio A. MiT/TFE Family of Transcription Factors, Lysosomes, and Cancer. *Annu Rev Cancer Biol*. 2019;3:203-222. doi: 10.1146/annurev-cancerbio-030518-055835

273. Bai T, Li M, Liu Y, Qiao Z, Wang Z. Inhibition of ferroptosis alleviates atherosclerosis through attenuating lipid peroxidation and endothelial dysfunction in mouse aortic endothelial cell. *Free Radic Biol Med*. 2020;160:92-102. doi: 10.1016/j.freeradbiomed.2020.07.026

274. Bochkov VN, Oskolkova OV, Birukov KG, Levonen AL, Binder CJ, Stockl J. Generation and biological activities of oxidized phospholipids. *Antioxid Redox Signal*. 2010;12:1009-1059. doi: 10.1089/ars.2009.2597

275. Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin*. 2020;13:22. doi: 10.1186/s13072-020-00342-y

276. Cohen DM, Lim HW, Won KJ, Steger DJ. Shared nucleotide flanks confer transcriptional competency to bZip core motifs. *Nucleic Acids Res*. 2018;46:8371-8384. doi: 10.1093/nar/gky681

277. Zhang C, Adamos C, Oh MJ, Baruah J, Ayee MAA, Mehta D, Wary KK, Levitan I. oxLDL induces endothelial cell proliferation via Rho/ROCK/Akt/p27(kip1) signaling: opposite effects of oxLDL and cholesterol loading. *Am J Physiol Cell Physiol*. 2017;313:C340-C351. doi: 10.1152/ajpcell.00249.2016

278. Kim JB, Deluna A, Mungrue IN, Vu C, Pouldar D, Civelek M, Orozco L, Wu J, Wang X, Charugundla S, et al. Effect of 9p21.3 coronary artery disease locus neighboring genes on atherosclerosis in mice. *Circulation*. 2012;126:1896-1906. doi: 10.1161/CIRCULATIONAHA.111.064881

279. Mueller PA, Yang L, Ubele M, Mao G, Brandon J, Vandra J, Nichols TC, Escalante-Alcalde D, Morris AJ, Smyth SS. Coronary Artery Disease Risk-Associated Plpp3 Gene and Its Product Lipid Phosphate Phosphatase 3 Regulate Experimental Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2019;39:2261-2272. doi: 10.1161/ATVBAHA.119.313056

280. Chatterjee I, Baruah J, Lurie EE, Wary KK. Endothelial lipid phosphate phosphatase-3 deficiency that disrupts the endothelial barrier function is a modifier of cardiovascular development. *Cardiovasc Res*. 2016;111:105-118. doi: 10.1093/cvr/cvw090

281. Peng Y, Jahroudi N. The NFY transcription factor functions as a repressor and activator of the von Willebrand factor promoter. *Blood*. 2002;99:2408-2417. doi: 10.1182/blood.v99.7.2408

282. Ouyang S, You J, Zhi C, Li P, Lin X, Tan X, Ma W, Li L, Xie W. Ferroptosis: the potential value target in atherosclerosis. *Cell Death Dis*. 2021;12:782. doi: 10.1038/s41419-021-04054-3

283. Chavakis E, Dernbach E, Hermann C, Mondorf UF, Zeiher AM, Dimmeler S. Oxidized LDL inhibits vascular endothelial growth factor-induced endothelial cell migration by an inhibitory effect on the Akt/endothelial nitric oxide synthase pathway. *Circulation*. 2001;103:2102-2107. doi: 10.1161/01.cir.103.16.2102

284. Murugesan G, Chisolm GM, Fox PL. Oxidized low density lipoprotein inhibits the migration of aortic endothelial cells in vitro. *J Cell Biol*. 1993;120:1011-1019. doi: 10.1083/jcb.120.4.1011

285. Gil N, Ulitsky I. Production of Spliced Long Noncoding RNAs Specifies Regions with Increased Enhancer Activity. *Cell Syst*. 2018;7:537-547 e533. doi: 10.1016/j.cels.2018.10.009

286. Lim KH, Han Z, Jeon HY, Kach J, Jing E, Weyn-Vanhentenryck S, Downs M, Corrionero A, Oh R, Scharner J, et al. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. *Nat Commun*. 2020;11:3501. doi: 10.1038/s41467-020-17093-9

287. Marian AJ. Causality in genetics: the gradient of genetic effects and back to Koch's postulates of causality. *Circ Res*. 2014;114:e18-21. doi: 10.1161/CIRCRESAHA.114.302904

288. Pingault JB, Richmond R, Davey Smith G. Causal Inference with Genetic Data: Past, Present, and Future. *Cold Spring Harb Perspect Med*. 2022;12. doi: 10.1101/cshperspect.a041271

289. Aylward A, Okino ML, Benaglio P, Chiou J, Beebe E, Padilla JA, Diep S, Gaulton KJ. Glucocorticoid signaling in pancreatic islets modulates gene regulatory programs and genetic risk of type 2 diabetes. *PLoS Genet*. 2021;17:e1009531. doi: 10.1371/journal.pgen.1009531

290. Zhang P, Amarasinghe HE, Whalley JP, Tay C, Fang H, Migliorini G, Brown AC, Allcock A, Scozzafava G, Rath P, et al. Epigenomic analysis reveals a dynamic and context-specific macrophage enhancer landscape associated with innate immune activation and tolerance. *Genome Biol*. 2022;23:136. doi: 10.1186/s13059-022-02702-1

291. Mosquera JV, Auguste G, Wong D, Turner AW, Hodonsky CJ, Alvarez-Yela AC, Song Y, Cheng Q, Lino Cardenas CL, Theofilatos K, et al. Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. *Cell Rep*. 2023;42:113380. doi: 10.1016/j.celrep.2023.113380

292. Chiou J, Geusz RJ, Okino ML, Han JY, Miller M, Melton R, Beebe E, Benaglio P, Huang S, Korgaonkar K, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature*. 2021;594:398-402. doi: 10.1038/s41586-021-03552-w

293. Mulder K, Patel AA, Kong WT, Piot C, Halitzki E, Dunsmore G, Khalilnezhad S, Irac SE, Dubuisson A, Chevrier M, et al. Cross-tissue single-cell landscape of human monocytes and macrophages in health and disease. *Immunity*. 2021;54:1883-1900 e1885. doi: 10.1016/j.immuni.2021.07.007

294. Norreen-Thorsen M, Struck EC, Oling S, Zwahlen M, Von Feilitzen K, Odeberg J, Lindskog C, Ponten F, Uhlen M, Dusart PJ, et al. A human adipose tissue cell-type transcriptome atlas. *Cell Rep*. 2022;40:111046. doi: 10.1016/j.celrep.2022.111046

295. Cottam MA, Caslin HL, Winn NC, Hasty AH. Multiomics reveals persistence of obesity-associated immune cell phenotypes in adipose tissue during weight loss and weight regain in mice. *Nat Commun*. 2022;13:2950. doi: 10.1038/s41467-022-30646-4

296. Stansbury CM, Dotson GA, Pugh H, Rehemtulla A, Rajapakse I, Muir LA. A lipid-associated macrophage lineage rewires the spatial landscape of adipose tissue in early obesity. *JCI Insight*. 2023;8. doi: 10.1172/jci.insight.171701

297. Sarvari AK, Van Hauwaert EL, Markussen LK, Gammelmark E, Marcher AB, Ebbesen MF, Nielsen R, Brewer JR, Madsen JGS, Mandrup S. Plasticity of Epididymal Adipose Tissue in Response to Diet-Induced Obesity at Single-

Nucleus Resolution. *Cell Metab*. 2021;33:437-453 e435. doi: 10.1016/j.cmet.2020.12.004

298. Kim K, Shim D, Lee JS, Zaitsev K, Williams JW, Kim KW, Jang MY, Seok Jang H, Yun TJ, Lee SH, et al. Transcriptome Analysis Reveals Nonfoamy Rather Than Foamy Plaque Macrophages Are Proinflammatory in Atherosclerotic Murine Models. *Circ Res*. 2018;123:1127-1142. doi: 10.1161/CIRCRESAHA.118.312804

299. Zernecke A, Winkels H, Cochain C, Williams JW, Wolf D, Soehnlein O, Robbins CS, Monaco C, Park I, McNamara CA, et al. Meta-Analysis of Leukocyte Diversity in Atherosclerotic Mouse Aortas. *Circ Res*. 2020;127:402-426. doi: 10.1161/CIRCRESAHA.120.316903

300. Cochain C, Saliba AE, Zernecke A. Letter by Cochain et al Regarding Article, "Transcriptome Analysis Reveals Nonfoamy Rather Than Foamy Plaque Macrophages Are Proinflammatory in Atherosclerotic Murine Models". *Circ Res*. 2018;123:e48-e49. doi: 10.1161/CIRCRESAHA.118.314120

301. Piollet M, Porsch F, Rizzo G, Kapser F, Schulz DJJ, Kiss MG, Schlepckow K, Morenas-Rodriguez E, Sen MO, Gropper J, et al. TREM2 protects from atherosclerosis by limiting necrotic core formation. *Nature Cardiovascular Research*. 2024;3:269-282. doi: 10.1038/s44161-024-00429-9

302. Alizada A, Khyzha N, Wang L, Antounians L, Chen X, Khor M, Liang M, Rathnakumar K, Weirauch MT, Medina-Rivera A, et al. Conserved regulatory logic at accessible and inaccessible chromatin during the acute inflammatory response in mammals. *Nat Commun*. 2021;12:567. doi: 10.1038/s41467-020-20765-1

303. Shalhoub J, Viiri LE, Cross AJ, Gregan SM, Allin DM, Astola N, Franklin IJ, Davies AH, Monaco C. Multi-analyte profiling in human carotid atherosclerosis uncovers pro-inflammatory macrophage programming in plaques. *Thromb Haemost*. 2016;115:1064-1072. doi: 10.1160/TH15-08-0650

304. Vielma SA, Mironova M, Ku JR, Lopes-Virella MF. Oxidized LDL further enhances expression of adhesion molecules in Chlamydophila pneumoniae-infected endothelial cells. *J Lipid Res*. 2004;45:873-880. doi: 10.1194/jlr.M300456-JLR200

305. Chatzizisis YS, Jonas M, Coskun AU, Beigel R, Stone BV, Maynard C, Gerrity RG, Daley W, Rogers C, Edelman ER, et al. Prediction of the localization of high-risk coronary atherosclerotic plaques on the basis of low endothelial shear stress: an intravascular ultrasound and histopathology natural history study. *Circulation*. 2008;117:993-1002. doi: 10.1161/CIRCULATIONAHA.107.695254