Environmental and genetic influences on atherosclerosis

Thomas Keith Hiron

The Queen's College

Nuffield Department of Medicine

University of Oxford

A thesis submitted for the degree of

DPhil in Clinical Medicine

Trinity Term 2019

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.

Environmental and genetic influences on atherosclerosis

Thomas Keith Hiron
The Queen's College

A thesis submitted for the degree of
DPhil in Clinical Medicine

Trinity Term 2019

Abstract

Background: Atherosclerosis is the underlying cause of coronary artery disease (CAD), which remains a major cause of death worldwide. It is a complex disease influenced by multiple environmental factors acting on different cell types and at different stages of pathogenesis. Genetic variants at more than 160 loci have been associated with CAD, but for the majority the causal molecular mechanisms remain unknown.

Aims: I aim to improve our understanding of the cellular response to environmental influences on atherosclerosis, and to identify molecular mechanisms of heritable disease risk. Specifically, I aim to characterise the responses of human macrophages to high glucose, of human aortic endothelial cells to laminar shear stress, and of single human macrophages to oxidised low-density lipoprotein.

Methods: I developed primary human cell models of atherosclerotic disease processes and applied high-throughput sequencing (HTS) methods to study changes in transcriptional regulation. RNA-seq and ATAC-seq were used to characterise responses to known environmental atherosclerosis risk factors at the molecular level. Resulting genome-wide maps were integrated with publicly available GWAS data to identify SNPs

located in dynamic regulatory elements. In addition, I used single-cell RNA-seq to study heterogeneity in human macrophages exposed to atherogenic lipid.

Key results: Human macrophages express CD28 at the cell surface, and expression is downregulated by high glucose. Laminar shear stress induces widespread changes in endothelial cell chromatin accessibility. Newly accessible regions are enriched for AP-1 and KLF transcription factor binding motifs, as well as for GWAS SNPs associated with CAD. Colocalisation of these features reveals candidate genetic mechanisms for disease risk. Human macrophages are heterogeneous in vitro and consist of subtypes with gene expression reflecting the full range of macrophage functions.

Conclusions: HTS methods are valuable tools for studying the intersection of environmental and genetic influences on atherosclerosis. Discovery of novel disease mechanisms could lead to the development of therapeutic strategies for cardiovascular disease.

Acknowledgements

I would like to express my gratitude to Dr Bill Frankland for funding my studentship. Without his generosity, belief, and inspiration, I would not be where I am today.

I would like to thank my supervisor, Professor Chris O'Callaghan, for his wisdom, his encouragement and, most of all, his patience over the last four years. I would also like to thank the other members of the O'Callaghan lab, Anil Chalisey, Da Lin, Lucy Davison, Marsha Wallace, Stephen Hare and Katherine Robins, for their constructive (and disruptive) influence, and for being the most excellent colleagues. In the same vein, I thank the members of the Cornall lab for their assistance, friendship, and most vitally, cake.

I would also like to thank my family, to whom I owe everything. My parents, Keith and Caroline, my siblings, Josie and Jack, and not forgetting our Miniature Schnauzer, Reggie - you are immeasurably important to me. Last, but by no means least, I would like to thank Amy for bringing joy to my life every single day.

Contents

List of figures

List of tables

Abbreviations

Abbreviation	Term
ATAC	assay for transposase accessible chromatin
ATAC-RSB	ATAC-resuspension buffer
ATF	activating transcription factor
BAM	binary alignment map
BDMA	benzyldimethylamine
BP	biological process
BSA	bovine serum albumin
bZIP	basic leucine zipper
CAD	coronary artery disease
CAGE	cap analysis of gene expression
CAM	cell adhesion molecule
CC	cellular component
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
ChREBP	carbohydrate response element binding protein
CPM	counts per million mapped reads
CREB	cAMP response element binding protein
DHS	Dnase hypersensitive site
ECAR	extracellular acidification rate
ECM	extracellular matrix
EDTA	ethylenediaminetetraacetic acid
eNOS	endothelial nitric oxide synthase
EPD	eukaryotic promoter database
eQTL	expression quantitative trait locus
ERG	ETS-related gene
eRNA	enhancer RNA
ETS	E26 transformation-specific
FAIRE	formaldehyde-assisted isolation of regulatory elements
FANTOM	functional annotation of the mammalian genome
FCCP	carbonyl cyanide-4-(trifluoromethoxy)phenylhydrazone
FcR	Fc receptor
FPKM	fragments per kilobase per million mapped reads
FRiP	fraction of reads in peaks
FSC	forward scatter
GLM	generalised linear model
GO	gene ontology
GPCR	G-protein coupled receptor
GRO	global run-on
GSH	reduced glutathione
GSSG	oxidised glutathione

GTF	gene transfer format
GWAS	genome-wide association studies
HAEC	human aortic endothelial cell
HDAC	histone deacetylase
HDL	high-density lipoprotein
HG	high glucose
HIF-1α	hypoxia-inducible factor 1α
HSMM	human skeletal muscle myoblasts
HTS	high-throughput sequencing
HUVEC	human umbilical vein endothelial cell
IFN-γ	interferon-γ
KLF	Krüppel-like factor
LD	linkage disequilibrium
LDL	low-density lipoprotein
LPA	lysophosphatidic acid
LPS	lipopolysaccharide
M-CSF	macrophage colony-stimulating factor
MACS	magnetic cell sorting
MAPK	mitogen-activated protein kinase
MCP-1	monocyte chemoattractant protein 1
MF	molecular function
MHC	major histocompatibility complex
mRNA	messenger RNA
mTOR	mammalian target of rapamycin
NAC	N-acetylcysteine
NB	negative binomial
NF-κB	nuclear factor kappa-light-chain-enhancer of activated B cells
NFY	nuclear transcription factor Y
NG	normal glucose
NGF	nerve growth factor
NK	natural killer
NO	nitric oxide
OCR	oxygen consumption rate
oxLDL	oxidised low-density lipoprotein
PBS	phosphate buffered saline
PCA	principal component analysis
PDGF	platelet-derived growth factor
PER	proton efflux rate
PWM	position weight matrix
QL	quasi-likelihood
RAR	retinoic acid receptor
ROS	reactive oxygen species
RT-qPCR	reverse transcription quantitative polymerase chain reaction

S1P	sphingosine 1-phosphate
SAM	sequence alignment map
scATAC-seq	single-cell ATAC-seq
scRNA-seq	single-cell RNA-seq
SNP	single nucleotide polymorphism
SOX	sry-related high mobility group box
SSC	side scatter
t-SNE	t-distributed stochastic neighbour embedding
TCA	tricarboxylic acid
TCH	thiocarbohydrazide
TEM	transmission electron microscopy
TLR	toll-like receptor
TMM	trimmed mean of m-values
TNFα	tumour necrosis factor α
TSS	transcription start site
UMAP	uniform manifold approximation and projection
UMI	unique molecular identifier
UTR	untranslated region
VEGF	vascular endothelial growth factor
VSMC	vascular smooth muscle cell

Chapter 1

Background

1.1 Introduction

Atherosclerosis is the most common cause of coronary artery disease (CAD), which remains a leading cause of morbidity and mortality worldwide[1,2]. It is a chronic and progressive disease characterised by deposition and accumulation of circulating lipids in the arterial intima, beginning with the formation of 'fatty streaks' in the vessel wall as early as childhood, and ultimately progressing to advanced, fibro-atheromatous lesions in later life[3]. Late-stage atherosclerosis may involve plaque formation and rupture, which releases plaque contents into the arterial lumen and can trigger thrombosis and vessel occlusion, leading to major cardiovascular events such as myocardial infarction or stroke[4]. Medium and large-sized arteries are mainly affected, and lesions typically form at sites of arterial injury or regions subjected to low and oscillatory haemodynamic shear stress[5]. A number of environmental risk factors are known to influence the development of atherosclerosis, including modifiable extrinsic factors such as smoking, dietary fat, and uncontrolled diabetes mellitus (types I and II), as well as intrinsic factors such as sex, age and haemodynamic forces at the vessel wall, such as shear stress and blood pressure[6]. A major risk factor is high levels of circulating low density lipoprotein (LDL) which, when oxidised (oxLDL), is highly atherogenic[7]. Deposition of circulating lipoproteins in the vessel wall at atheroprone sites can lead to their chemical modification (e.g. oxidation, acetylation)[8]. Circulating monocytes enter the developing lesion, wherein they differentiate to macrophages and take up these modified lipoproteins leading to intracellular accumulation of lipid and formation of 'foam cells', a hallmark of early 'fatty streak' type lesions[9]. These cells express inflammatory cytokines and contribute to the recruitment of additional immune cells, as well as modifying the behaviour of other cells in the vessel wall (e.g. endothelial cells, smooth muscle cells)[10]. Over time, the accumulation of lipid and cells leads to intimal thickening and occlusion of the affected

vessel. Given the complexity of the disease, with the involvement of multiple cell types and environmental factors in lesion formation and development, the precise aetiology of atherosclerosis is still poorly understood. Recently, genome-wide association studies (GWAS) have identified a large number of genomic loci associated with CAD and related traits, highlighting a substantial genetic contribution to, and mechanism for heritability of, CAD risk[11–13]. Functional characterisation of these loci may yield a deeper understanding of the development of atherosclerosis at the molecular level, potentially leading to targeted therapies and mitigation of cardiovascular disease burden.

1.2 Mechanisms of atherogenesis

1.2.1 Endothelial cell activation and dysfunction

The endothelium spans the entire circulatory system maintaining a physiological barrier between blood and tissue, and controls the transport of essential metabolites to all cells outside of the vasculature, while excluding pathogens and other potentially damaging blood-borne substances[14]. It also regulates extravasation of circulating leukocytes through the expression of cell adhesion molecules (CAMs) on the luminal surface of endothelial cells, which are recognised by complementary receptors expressed on the surface of immune cells[15]. Additionally, normal endothelium produces nitric oxide (NO) through the metabolism of L-arginine by endothelial nitric oxide synthase (eNOS), the primary function of which is to promote vasodilation through guanylyl cyclase signalling in vascular smooth muscle cells (VSMCs)[16]. These processes combine to maintain normal endothelial integrity, which is essential for haemostasis and prevention of thrombosis[17].

The early events in atherosclerosis have been attributed to both endothelial activation and dysfunction. Endothelial activation is defined as a proliferative, proinflammatory and prothrombotic state, in which cell surface expression of CAMs, such as ICAM-1 and VCAM-1, is increased and promotes transendothelial migration of leukocytes[15]. Endothelial dysfunction, on the other hand, is defined as a reduced propensity for NO production, which regulates vascular tone and also inhibits adhesion of platelets and leukocytes to the endothelium[18]. Although discovered separately, endothelial activation and dysfunction are often conflated, as both are features of endothelial cells in atherosclerosis, and are linked by common regulatory pathways and disease risk factors[19–21].

Transendothelial movement of LDL and other macromolecules is restricted for normal endothelium due to the barrier established by inter-endothelial cell junctions, comprising tight, adherens and gap junctions[22]. Under these conditions, deposition of LDL in the intimal space is largely dependent on coordinated transcytosis via LDL receptors (e.g. SCARB1, ALK1), caveolae and clathrin-coated vesicles[23,24]. In dysfunctional endothelium, where apoptosis and cell turnover are increased, the intercellular junctions are impaired and allow LDL particles to move freely through the open space between cells[25,26]. The result is deposition of LDL in the subendothelial space, which contributes to the development of an atherosclerotic lesion at the affected site. Impairment of inter-endothelial cell junctions may also promote diapedesis of leukocytes at nascent and early developing lesions. Endothelial cells interact directly with circulating immune cells by expressing CAMs such as ICAM-1, VCAM-1 and selectins, and these proteins are upregulated on the surface of activated endothelial cells[15]. Increased recruitment of immune cells to atheroprone sites is also mediated by the release of inflammatory cytokines, such as MCP-1 and IL-8, from activated endothelial cells[27,28]. These secreted

proteins are chemotactic for monocytes and lymphocytes, and further accelerate the infiltration of immune cells at the developing lesion. In addition to immunomodulatory molecules, activated endothelial cells express prothrombotic mediators, such as Tissue Factor and von Willebrand Factor, which can activate the coagulation cascade via interaction with Factor VII/VIIa, and platelet activation, respectively[17].

Further to these aberrations of normal endothelial function, activated endothelium can express MICA and MICB, which are ligands for the NKG2D receptor expressed on NK cells and cytotoxic T cells[29–31]. These proteins act as a molecular signal of stress, and upon ligation of NKG2D trigger the targeted lysis of the stressed cells. MICA expression is increased by TNFα *in vitro* and has been observed in atherosclerotic plaque from patients[29]. It is possible that immune-mediated killing of endothelial cells expressing MICA contributes to atherosclerosis by disrupting endothelial integrity, thereby allowing entry of lipid and immune cells to the arterial intima, exacerbating atherogenesis. MICA-mediated destruction of endothelial cells could also result in plaque destabilisation and rupture in the later stages of the disease, releasing the proinflammatory and prothrombotic contents of the atheromatous plaque into the circulation.

A number of factors regulate endothelial dysfunction/activation in atherosclerosis (e.g. hyperlipidaemia, reactive oxygen species, cigarette smoke, hypertension), but a major influence on both of these processes is shear stress exerted on the apical surface of the endothelium by circulating blood[19,32]. Evidence from autopsy samples and animal models has shown that the distribution of atheroma formation in the vasculature is non-random, and shows a preference for branch points, the outer walls of bifurcations, and the inner wall of curvatures, all of which experience low, disturbed or oscillatory shear stress[33–35]. Conversely, vessels subjected to physiological laminar flow and high shear stress (> 15

dynes/cm^2) are protected from atherogenesis[36]. Recently, studies have mapped the transcriptomic effects of defined laminar shear stress *in vitro* and demonstrated that shear stress directly alters the expression of pro- and anti-inflammatory genes[37]. Shear stress is sensed by various mechanosensory receptors at the endothelial cell surface, including GPCRs (e.g. GPR68)[38], ion channels (e.g. PIEZO1/2)[39] and components of inter-endothelial junctions (e.g. PECAM-1)[40]. These signals alter shear stress-responsive gene expression through a number of pathways, resulting in activation or repression of multiple transcription factors. Expression of the Krüppel-like Factor zinc finger proteins KLF2 and KLF4 is induced by physiological laminar shear stress and has been shown to induce expression of many of the genes associated with the atheroprotective effects of shear stress[41]. Expression of KLF2 in particular has been demonstrated in regions of human arteries protected from atherosclerosis[42]. Under conditions of low or oscillatory shear stress, NF-κB has been implicated as a key driver of inflammatory gene expression[32]. These factors are therefore considered to be antagonistic in the context of endothelial cells under laminar shear stress, and their activity is central to atherogenic processes occurring at the endothelium[43]. Transcription of shear stress-responsive genes is also regulated by epigenetic modifications, with evidence of DNA hypermethylation by DNMT1 at the KLF4 promoter in endothelial cells exposed to disturbed flow[44]. NF-κB is also involved in endothelial cell activation by TNFα, which binds to the MICA promoter and stimulates increased transcription of the gene, and therefore higher levels of protein at the cell surface[29]. The transcriptional networks controlling endothelial gene expression in atherosclerosis are complex, and much work remains to elucidate the full extent of their regulation at the molecular level.

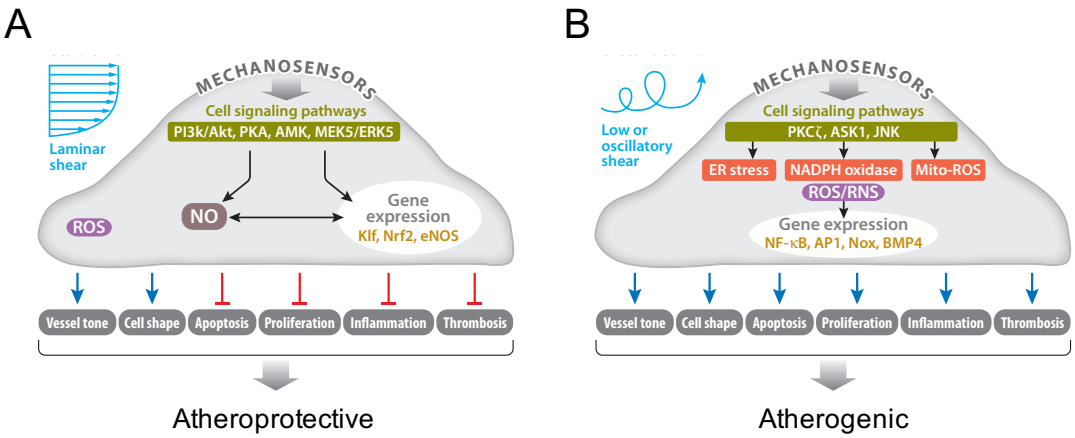

Figure 1.1: Effects of different shear stress patterns on endothelial cells. (A) Laminar shear stress induces an atheroprotective phenotype, whereas (B) low or oscillatory shear stress promotes an atherogenic phenotype characterised by increased endothelial cell turnover, increased expression of adhesion molecules and inflammatory cytokines, and reduced nitric oxide (NO) production. Signaling pathways and downstream transcription factors activated by shear stress are shown. Figure adapted from Tarbell et al, 2014[35].

1.2.2 Monocyte differentiation and macrophage lipid uptake

Monocyte infiltration is a major feature of atherogenesis, and after entering a lesion monocytes differentiate to macrophages[10]. Murine models of atherosclerosis have identified macrophage colony-stimulating factor (M-CSF) as important to atheroma formation, with almost complete ablation of aortic atherogenesis in homozygous knockout mice lacking M-CSF[45]. Further experiments on a *Ldlr*$^{-/-}$ background demonstrated that mice with a single functional copy of M-CSF had reduced atherosclerosis but a normal number of circulating monocytes, suggesting that M-CSF may contribute to fatty streak formation locally rather than systemically[46]. Expression of M-CSF by human endothelial and vascular smooth muscle cells has been demonstrated *in vitro* and has also been observed in human atheroma[47]. M-CSF stimulates adherence and differentiation of monocytes *in vitro,* and this is a commonly used method of obtaining macrophages experimentally[48,49]. Macrophages express a number of scavenger receptors known to interact with oxLDL, including CD36, SCARB1, MSR-1 and LOX-

1[50]. Some of these molecules have been shown to directly contribute to the progression of atherosclerosis in murine models. For example, *Cd36*-knockout mice (on both *ApoE*$^{-/-}$ and *Ldlr*$^{-/-}$ backgrounds) fed a high fat diet exhibited reduced formation of atherosclerotic lesions[51,52]. LDL particles are rich in triglycerides, phospholipids and cholesteryl esters, and endocytosis by macrophages leads to cytosolic accumulation of these lipid species. While some cholesterol efflux is facilitated by transporters in the macrophage plasma membrane (e.g. ABCA1, ABCG1), the overwhelming concentration of modified lipoprotein particles in a lesion results in a significant population of lipid-laden macrophages, which are known as 'foam cells' due to their 'foamy' appearance[9]. Transcriptomic analysis of foam cells has revealed profound changes in expression of genes involved in lipid and cholesterol metabolism, as well as inflammation and cytokine production[53]. *In vitro* studies of foam cells have demonstrated secretion of inflammatory cytokines, such as IL-1β[54], IL-8[55] and MCP-1[56], and the central role of these molecules has been demonstrated by knockout mouse models of atherosclerosis[57,58]. The cytokines released by macrophage-derived foam cells likely contribute to the progression of a fatty streak to a more advanced lesion by recruiting additional monocytes (and other immune cells) from the bloodstream, as well as modifying their intralesional behaviour.

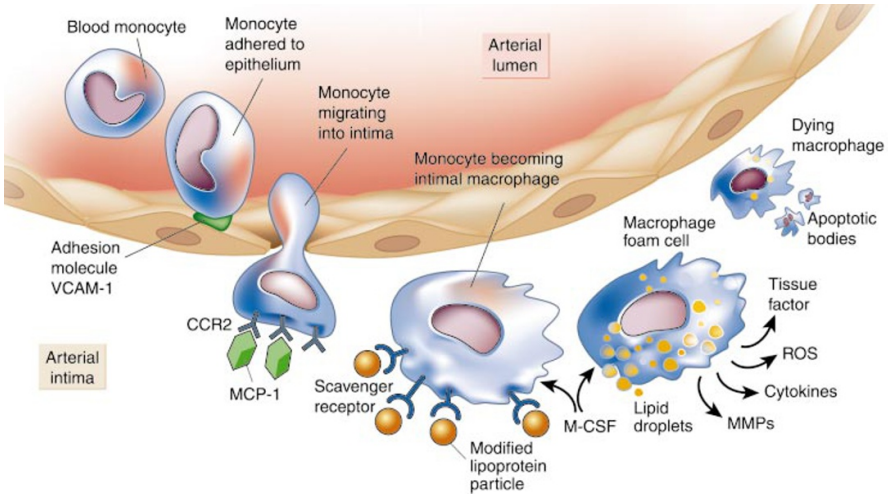

Figure 1.2: A model of monocyte infiltration in atherosclerosis. Circulating monocytes adhere to activated endothelial cells expressing adhesion molecules (e.g. VCAM-1) and traverse the endothelium to enter the arterial intima. Binding of monocyte chemoattractant protein 1 (MCP-1) released from lesional cells to its receptor C-C chemokine receptor 2 (CCR2) on monocytes promotes chemotaxis into the lesion. In the arterial intima, macrophage colony stimulating factor (M-CSF) promotes monocyte differentiation to macrophages, which take up modified lipoproteins (e.g. oxLDL) and become foam cells. These cells release a host of inflammatory mediators and reactive oxygen species (ROS), before eventually succumbing to apoptosis and contributing to the accumulation of lipid and cellular debris within the lesion. Figure from Libby, 2002[10].

Changes in foam cell gene expression are mediated through a number of transcription factors, as well as epigenetic modifications. Recognition and internalisation of oxLDL by CD36 on the surface of macrophages induces activation of NFκB, which in turn upregulates the transcription of inflammatory cytokines such as TNFα and IL-1β[59]. In addition, uptake of oxLDL via CD36 stimulates upregulation of CD36 itself, in a PPARγ-dependent manner[60,61]. These pathways integrate immune and metabolic regulation of gene expression to define the response of macrophages to atherogenic lipids. Human monocyte-derived macrophages treated with oxLDL *in vitro* show a profound alteration of chromatin accessibility at loci containing genes associated with atherosclerosis, as well as histone modifications associated with active gene expression (i.e. H3K27Ac)[53]. Further to these observations, genome-wide analysis of CEBP-β binding in response to oxLDL treatment demonstrated increased binding of this factor near genes related to

inflammatory processes, suggesting a role for CEBP-β in foam cell formation. Overlap of the regions identified in these epigenomic datasets with disease-associated genetic variants identified by GWAS led to the discovery that the single nucleotide polymorphism (SNP) rs72664324, at the *PLPP3* locus, is located within a CEBP-β binding site. This SNP demonstrated allelic differences in CEBP-β binding, enhancer activity, and consequently expression of *PLPP3*, the product of which is an enzyme catalysing the dephosphorylation and inactivation of pro-inflammatory mediators such as lysophosphatidic acid (LPA) and sphingosine 1-phosphate (S1P). This study demonstrates the utility of functional genomics in resolving biological function of disease-associated genetic variation, and thereby identifying molecular pathways central to disease progression. Similar studies targeting other key processes in atherosclerosis have the potential to further clarify the genetic component of the disease, and to discover novel molecular pathways influencing atherogenesis.

1.2.3 The role of vascular smooth muscle cells

The growth factors, inflammatory cytokines and matrix metalloproteinases released by endothelial and immune cells under atherogenic conditions stimulate quiescent VSMCs to take on a migratory and proliferative phenotype[62]. These activated cells, normally resident in the tunica media, traverse the internal elastic lamina and enter the tunica intima at sites of developing lesions, where they interact with a complex mixture of cells, secreted factors and metabolites (including oxLDL). This phenotypic switch also promotes the release of inflammatory cytokines, and the synthesis of extracellular matrix (ECM) components, including collagen and elastin, which contribute to the formation of the fibrous cap covering advanced lesions[63]. This synthetic phenotype can be induced in VSMCs by direct exposure to oxLDL, and foam cell-like lipid-laden VSMCs have been

observed in human atherosclerotic lesions[63,64]. Later in an affected individual's life, the thinning and eventual rupture of the fibrous cap can lead to release of inflammatory mediators, lipids, and cellular debris into the arterial lumen, triggering thrombosis and possibly fatal cardiovascular events through vessel occlusion and infarction. The repeated rupture and clotting of a late-stage atheroma can result in stenosis of a diseased vessel, restricting the blood supply to organs which, in the case of the coronary artery supplying the heart, can also be fatal. Although the precise molecular mechanisms governing the phenotypic switch from quiescent to activated VSMCs remain poorly understood, studies have shown that NO and growth factors produced by intact and functional endothelium reduces the migration of VSMCs to the intima[19]. This may partly explain the contribution of reduced NO production by dysfunctional endothelial cells to atherosclerosis.

1.2.4 Other immune cell types involved in atherosclerosis

Minor populations of several other immune cell subsets have been found in atherosclerotic lesions, but intense research has focused on the presence of CD4+ T lymphocytes in mature lesions[65]. Adaptive immunity is thought to drive advanced disease progression, with oxLDL autoantibodies[66] and oxLDL-specific T cells[67] enhancing localised inflammation by secreting chemokines and recruiting further immune cells to the lesion. T_H1 cells, in particular, are considered pro-atherogenic due to their secretion of IFNγ, which activates macrophages[68]. CD4+ T cells are skewed toward T_H1 by IL-12 and IL-18, which are secreted by activated macrophages. There is some evidence that oxLDL uptake upregulates the expression of these cytokines in macrophages, and that they play a direct role in the development of atherosclerosis[69,70]. Interactions between cell adhesion molecules on the surface of activated endothelium (e.g. ICAM-1) and

circulating lymphocytes (e.g. LFA-1, a ligand for ICAM-1) also certainly play a role in the development of atherosclerosis[71].

1.3 HTS enables genome-wide exploration of transcriptional regulation

The modern development of high throughput sequencing (HTS) technologies has enabled massive efforts to determine the function of non-coding regions of the genome, thought to contain regulatory elements such as enhancers, which can influence the expression of genes over large genomic distances[72]. It has also enabled exploration of the transcriptomic landscape in many organisms, not just limited to protein-coding mRNA transcripts, but identifying many non-coding RNAs, micro-RNAs, enhancer RNAs and more[73].

Contemporary efforts to characterise the human transcriptome have been driven by technologies such as RNA-seq, a successor of DNA microarrays. RNA-seq enables unbiased quantification of all transcripts expressed in a population of cells and can be used for many applications including, but not limited to, the study of differential gene expression, *de novo* transcript discovery, and investigation of allele-specific expression. The methodology of RNA-seq involves purification of RNA from the cells or tissue under investigation, followed by fragmentation and generation of complementary DNA (cDNA) by reverse transcription. The cDNA is then converted to a sequencing library by ligating platform-specific adapter sequences, and ultimately the libraries are loaded onto a flow cell for massively parallel sequencing. The preparation of sequencing libraries can be tailored, depending on the desired output, by additional steps such as selection of poly-A-tailed transcripts to enrich for mRNA, i.e. protein-coding transcripts. A multitude of bioinformatic tools have been developed for the analysis of the data generated by RNA-

seq, rendering this technology accessible to investigators worldwide. Most recently, single cell RNA-seq (scRNA-seq) platforms have been developed, which enable simultaneous quantification of the transcripts expressed in thousands of individual cells[74]. This allows the study of transcriptomic heterogeneity in tissues and can drive the discovery of novel cell sub-types. It also enables the determination of functional differences within ostensibly homogeneous cell populations.

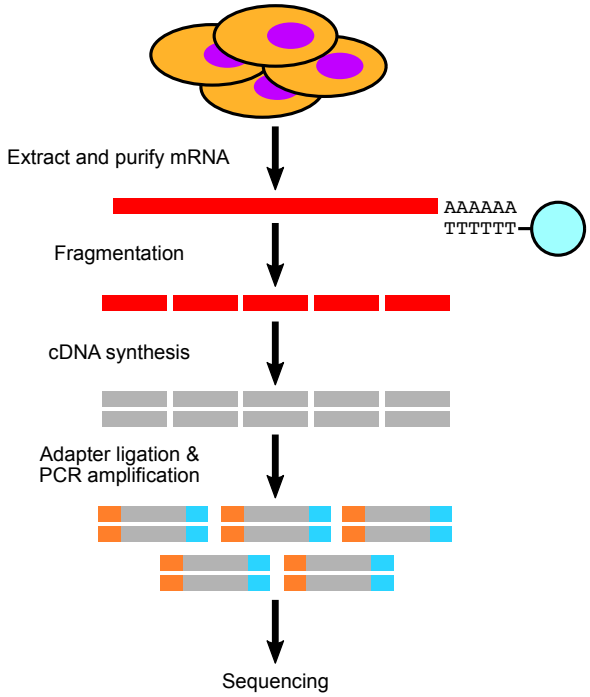

Figure 1.3: Overview of sequencing library preparation for RNA-seq. Total RNA is purified from cell or tissue lysate, and can be enriched for particular transcripts (e.g. polyadenylated) prior to fragmentation and cDNA synthesis. Platform-specific adapter sequences are ligated to the resulting cDNA, followed by PCR amplification and sequencing.

HTS has also been applied to the study of chromatin structure and function, with the aim of characterising the epigenetic factors responsible for establishing cell-specific transcriptional profiles. Chromatin is composed of nucleosomes in which ~147 bp of DNA is wrapped around an octamer containing two copies of each of the core histone proteins H2A, H2B, H3 and H4[75]. These nucleosomes are connected by naked 'linker

DNA' of variable length, and nucleosome structure is further stabilised by additional histone proteins, such as H1[76]. Regulation of this structure at the molecular level is integral to ensuring correct expression of genes in a cell type- and context-specific manner. Stretches of inactive chromatin are highly condensed and designated 'heterochromatin', while regions which are rich in genes, and hence actively transcribed, are referred to as 'euchromatin', and are typically less condensed. Euchromatin is accessible to proteins involved in active gene expression, including RNA polymerase, general transcription factors, as well as transcription factors that regulate gene expression by binding to sites distal to gene promoters, such as enhancers. The identification and characterisation of accessible sites in the genome using HTS has developed our understanding of epigenetic influences on gene expression[77]. One of the most widely used methods for probing chromatin accessibility genome-wide is the assay for transposase accessible chromatin (ATAC) with high throughput sequencing (ATAC-seq)[78]. This technology involves isolation of nuclei from a tissue or cell population under investigation, followed by treatment with a hyperactive mutant Tn5 transposase which has been pre-loaded with sequencing adapters. The transposase inserts the adapters in any accessible region of the genome, and simultaneously fragments and tags these regions preparing them for the generation of sequencing libraries[79]. ATAC-seq enables an unbiased quantitative analysis of chromatin accessibility across the entire genome and can be used to identify regulatory DNA elements which are differentially accessible in response to a particular stimulus. This technology is a successor to other technologies with a similar output, such as DNase-seq and FAIRE-seq. ATAC-seq has been widely adopted due to the speed and simplicity of the protocol, as well as the vastly reduced input requirements compared to DNase-seq and FAIRE-seq.

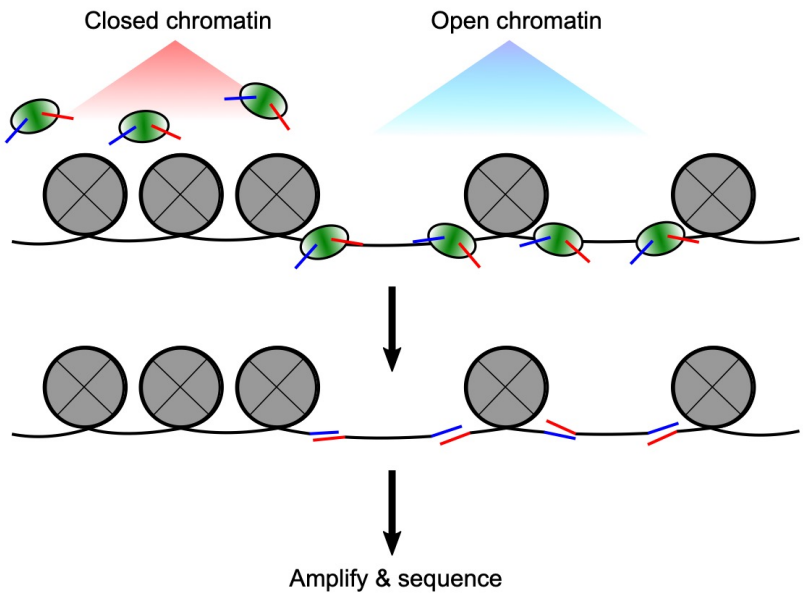

Figure 1.4: Transposition of accessible chromatin for ATAC-seq library preparation. Hyperactive Tn5 transposase (green) inserts pre-loaded sequencing adapters (blue and red) into regions of open chromatin, but not closed chromatin, within isolated nuclei. This results in the formation of DNA fragments amplifiable by PCR.

Another important aspect of the regulation of chromatin structure is the post-translational modification of histone proteins[80]. Lysine residues on histone proteins can be chemically modified in a variety of ways, which can influence inter- and intra-nucleosome interactions, and the best studied of these modifications are acetylation and methylation. Specific histone modifications are associated with particular chromatin states and transcriptional activity of nearby genes. Acetylation of lysine 27 on histone 3 (H3K27ac), for example, is associated with active gene expression and distinguishes active from poised enhancers[81]. Other markers of active gene expression include methylation of lysine 4 on histone 3 (H3K4me1, H3K4me2, H3K4me3), which is enriched at the promoters of actively transcribed genes[82]. Histone marks associated with repression of transcription have also been identified, such as H3K9me1, H3K9me3 and H3K27me3[80]. Chromatin immunoprecipitation (ChIP) with high throughput sequencing (ChIP-seq) has been used extensively to characterise the distribution of many of these marks in a range of tissues and cell types. ChIP-seq allows individual histone marks to be probed by

selection of an antibody specific to the mark of interest. Nucleosomes are crosslinked, and pull-down of histones bearing the mark recognised by the antibody results in co-precipitation of the DNA sequences bound to the histone. These DNA fragments can be used to prepare sequencing libraries, the analysis of which allows the mapping of genomic regions associated with each chromatin mark. ChIP-seq can also be used to explore other protein-DNA interactions genome-wide, such as the binding of specific transcription factors. Modern methods of exploring higher order chromatin structure, such as interactions in cis and in trans between different regions of the genome, include chromatin conformation capture (e.g. 3C), which fragments and ligates sequences physically near to one another in three-dimensional space within the nucleus. This technology has also been integrated with high throughput sequencing, resulting in the technique known as Hi-C, which identifies interacting regulatory elements genome-wide[83]. These methods have been used to successfully identify intergenic regulatory elements involved in disease processes[84].

Several large-scale projects, utilising a range of sequencing technologies, have made significant progress in mapping elements of the human epigenome. The ENCODE project, launched in 2003, set out to comprehensively identify functional elements across the human genome. To date, ENCODE has released data from more than 15,000 experiments, of which the most common assays are ChIP-seq of histone marks and transcription factors, DNase-seq to identify regions of accessible chromatin, and polyA RNA-seq. Datasets from human cell lines, primary cells and tissues have provided an extensive view of the sequence elements regulating transcription of the human genome[85]. The Roadmap Epigenomics Mapping Consortium was initiated in 2008, with the aim of constructing publicly available reference epigenomes in many different human tissues to facilitate research into the epigenomic factors influencing human disease. This program

expanded the set of cell lines originally explored by ENCODE to survey the chromatin features of primary human tissues in both foetal and adult samples[86]. An alternative and complementary approach was adopted by the FANTOM consortium, wherein transcriptional regulation was investigated at the level of RNA. The investigators employed cap analysis of gene expression with HTS (CAGE-seq)[87] to survey the 5'-most 27 bp of all transcripts contained in a broad range of mammalian cell types. This enabled the comprehensive identification of promoter elements genome wide[88]. In addition, following on from the discovery of transcribed enhancers (enhancer RNAs, eRNAs)[89], FANTOM researchers also catalogued transcribed enhancers across the entire genome[90], as well as the dynamics of enhancer transcription in cellular models of development and disease[91].

While the efforts to catalogue regulatory elements of the human genome have been hugely successful, the role of epigenetics in disease is still under investigation. A number of studies have identified epigenetic mechanisms involved in atherosclerosis. For example, direct methylation of cytosine residues within CpG dinucleotides, which represses transcription through the recruitment of chromatin remodelling factors and disruption of transcription factor binding sites, has been shown to regulate endothelial gene expression in response to disturbed blood flow[44]. Histone modifications have also been linked to the endothelial response to shear stress, with a recent study identifying regulatory elements with differential H3K27Ac between endothelial cells cultured in static and oscillatory shear stress conditions[92]. In support of the role of histone modifications, several lines of evidence have implicated histone deacetylases (HDACs) in the endothelial response to shear stress[93]. The effect of shear stress on chromatin accessibility has yet to be studied, and the epigenetics of atheroprotective laminar shear stress are underrepresented in current research, with the focus instead on atherogenic

patterns of shear stress. As described previously, regulation of chromatin structure has also been demonstrated for human macrophages treated with oxLDL[53], and chromatin accessibility is additionally altered in VSMCs in response to atherogenic stimuli[94]. These studies highlight how analysis of epigenetic regulation during key stages of pathogenesis can provide novel insights into the molecular mechanisms driving pathogenesis.

1.4 GWAS have identified a large number of CAD-associated loci

To date, more than 160 loci associated with CAD have been identified by genome-wide association studies (GWAS), as well as additional loci associated with related traits, such as circulating LDL-cholesterol measurement and blood pressure[11–13]. These studies demonstrate a substantial heritable component to cardiovascular disease but fail to identify the causal mechanisms underlying a variant-trait association. Although annotation of CAD-associated loci has highlighted molecular pathways in the vessel wall, the precise causal variants and molecular mechanisms remain largely unknown, in part due to linkage disequilibrium masking causal variants[12]. In addition, many of the CAD-associated variants are intergenic and distal to any coding sequence, that is they do not directly alter the transcription and ultimately translation of a gene product either by altering the primary amino acid sequence of the resulting protein or disrupting a splice site to modify the mRNA transcript. They are however enriched in DNase hypersensitive sites (DHS) present in the chromatin of a number of different cell types, suggesting an influence of genetic variation on the activity of distal regulatory elements in CAD[13].

1.5 Aims and objectives

The epigenetic regulation of disease processes provides an interface between environmental and genetic influences on atherosclerosis. Although many of the pathways

underlying the complexity of atherosclerosis are well understood, the genetic component of the disease remains largely unexplained. The aim of this study is to apply HTS methods to the study of transcriptional and epigenetic regulation at key stages of atherogenesis, and to integrate the resulting data with GWAS SNPs to identify novel mechanisms of susceptibility in atherosclerosis. In chapter 2, a detailed description of the methods applied is provided. In chapter 3, the application of RNA-seq and ATAC-seq to the study of the human macrophage response to hyperglycaemia, as a model of diabetic atherosclerosis, is described. Chapter 4 details a scRNA-seq study of primary human macrophage heterogeneity, and its influence on the response to oxLDL. In chapter 5 RNA-seq and ATAC-seq are used to study the endothelial cell response to atheroprotective laminar shear stress. Finally, in chapter 6 the findings presented in this thesis are summarised and discussed.

Chapter 2

Materials and methods

2.1 Ethical approval

Participants were recruited as part of the 'AtherOx1' study, with ethical approval from the NHS Research Ethics Committee (South Central - Hampshire B, reference 13/SC/0392, approval granted 09/08/13). All participants provided informed consent. The participant information leaflet and a blank copy of the consent form are included in Appendix 2.

2.2 Primary cell culture

The primary human cell types used were CD14+ monocytes/macrophages and aortic endothelial cells (HAECs). CD14+ cells were separated from peripheral blood of healthy adult volunteers and cultured immediately following isolation. HAECs (Promocell, Heidelberg, Germany) from 3 different donors were purchased at passage 3 and used at passage 4-5. CD14+ monocytes were differentiated to macrophages in R10 medium (RPMI 1640 supplemented with 10% heat inactivated (55°C for 30 minutes) foetal calf serum, 100 units/ml penicillin and 100 μg/ml streptomycin) containing 50 ng/ml recombinant human macrophage colony stimulating factor (M-CSF) (eBioscience, San Diego, CA) for 7 days. The medium was replaced 3 days post-isolation (to allow cells to become adherent), and again 5 days post-isolation. HAECs were maintained in Endothelial Cell Growth Medium MV2 (Promocell, Heidelberg, Germany) supplemented with 100 units/ml penicillin and 100 μg/ml streptomycin. Endothelial cells were passaged by trypsinisation at ~80% confluence. All cells were maintained in a humidified tissue culture incubator at 37°C and 5% CO_2. Cells were pelleted in an Allegra X-12R benchtop centrifuge (Beckman Coulter, Brea, CA) at 1500 rpm for 5 minutes at

room temperature, unless otherwise stated. Phase contrast images of cells in culture were acquired using an EVOS XL Core microscope (ThermoFisher Scientific, Waltham, MA).

2.3 Cell treatments

For investigation of the effects of glucose on human macrophages, CD14+ cells purified from human peripheral blood were differentiated in R10 formulated using glucose-free RPMI 1640 (Life Technologies, Carlsbad, CA) supplemented with appropriate quantities of D-glucose (Sigma-Aldrich, St. Louis, MO) for each condition (normal – 5 mM; high – 25 mM) and 50 ng/ml M-CSF. Normal and high glucose media were replaced 3, 5 and 7 days after isolation, and cells were harvested on day 8. For studies of the macrophage response to oxLDL, macrophages cultured in standard R10 were treated on day 7 post-isolation with 50 μg/ml oxidized low-density lipoprotein (oxLDL) or an equal volume of control buffer for 48 hours.

2.4 Isolation of CD14+ cells

50 ml of peripheral blood was drawn from healthy adult volunteers (REC approval reference 13/SC/0392) and centrifuged at 1200 rpm for 30 minutes with no brake over Ficoll-Paque PLUS (GE Healthcare Life Sciences, Pittsburgh, PA) in a blood:ficoll ratio of 2:1. After centrifugation, the buffy coat containing mononuclear cells was transferred to a clean tube and washed twice in 50 ml R10. The supernatant was discarded, and the cell pellet was resuspended in 12 ml ice cold MACS buffer (PBS with 0.5% BSA and 2mM EDTA) before centrifugation at 1500 rpm for 5 minutes. The supernatant was again discarded, and the cell pellet resuspended in 180 μl ice cold MACS buffer. 20 μl of anti-CD14 conjugated magnetic microbeads (Milltenyi Biotec, Bergisch Gladbach, Germany) was added to the cell suspension to make a final volume of 200 μl, and the cell/bead

mixture was incubated at 4°C on an orbital shaker for 15 minutes. Unbound beads were removed by washing the suspension with 12 ml ice cold MACS buffer, and pelleted cells were resuspended in 500 µl MACS buffer for magnetic sorting. For each cell suspension, a LS column (Milltenyi Biotec, Bergisch Gladbach, Germany) was placed in a magnetic field and equilibrated with 3 ml cold MACS buffer before addition of the cells. Unbound cells were allowed to pass through the column, and bound cells were washed with 3 ml cold MACS buffer three times. After the final wash, the LS column was removed from the magnetic field and CD14+ cells were eluted by adding 5 ml cold MACS buffer and flushing the column using the supplied plunger. Monocyte purity was ≥ 95%, as determined by flow cytometry. The cells were pelleted, resuspended in 10 ml R10, and cell number and viability were determined using a haemocytometer and staining with trypan blue, respectively. Cell concentration was adjusted to 5 x 10^5 cells/ml R10, and M-CSF was added to a final concentration of 50 ng/ml. 2 ml of the suspension was added to each well of a 6-well plate.

2.5 Parallel-plate flow chamber and flow circuit assembly

HAECs were subjected to laminar shear stress using a custom-made parallel plate flow chamber with a Masterflex L/S Digital Economy Drive and L/S Easy Load II Head (Cole-Parmer, Vernon Hills, IL) to pump Endothelial Cell Growth Medium MV2 around a continuous flow circuit[95]. For assembly of the flow circuit, tubing, a pulse dampener, stopcocks and luer connectors were all purchased from Cole-Parmer. The flow rate required to exert a defined shear stress on the endothelial cell monolayer was calculated using the following equation:

$$Q = \frac{\tau \cdot w \cdot h^2}{6 \cdot \mu}$$

where Q is the flow rate (ml/s), τ is the desired shear stress (dynes/cm^2), w and h are the width and height of the flow chamber, respectively, and μ is the viscosity of the flowed culture medium (0.009 g/cm/s). Fully developed laminar flow was achieved by maintaining a ratio of inertial forces to viscous forces (Reynolds number, Re) of < 2300. Values for Re greater than 4000 signify turbulent flow. Re was calculated according to the following equation:

$$Re = \frac{\rho \cdot Q \cdot D_h}{\mu \cdot w \cdot h}$$

where ρ is the fluid density and D_h is the hydraulic diameter defined as follows:

$$D_h = \frac{4 \cdot Cross - sectional\ Area}{Wetted\ Perimeter} = \frac{4 \cdot w \cdot h}{2 \cdot (w + h)}$$

The value of Re calculated for high laminar shear stress (20 dynes/cm^2) in the flow chamber was 41.4, which is well below the maximum threshold value for fully developed laminar flow, beyond which flow patterns become disturbed. For each run, two sterile 25 x 75 mm glass microscope slides were seeded with 1 x 10^6 cells each and cultured for 24-48 hours until ~50% confluent. For flow versus control experiments, one seeded slide was then cultured in the flow chamber under high laminar shear stress (20 dynes/cm^2), and the other in static medium (0 dynes/dm^2). Cells were harvested after 48 hours in either flowed or static medium.

2.6 Preparation of oxLDL

LDL was purified from human plasma by isopycnic ultracentrifugation. 50 ml of peripheral blood from healthy adult volunteers was centrifuged at 800 *g* for 30 minutes at room temperature to pellet cells, and the upper layer of plasma was transferred to a

clean tube. The plasma was again centrifuged at 2500 g for 30 minutes at room temperature to exclude platelets and any remaining cells. The supernatant was transferred to a clean tube, and the density was adjusted to 1.210 g/ml by addition of potassium bromide (KBr, 0.325 g KBr / ml plasma).

A sterile stock solution of 1.346 g/ml KBr was used to prepare solutions of 1.019 g/ml KBr (solution I) and 1.063 g/ml KBr (solution II) by dilution with sterile PBS. The volumes of stock solution and diluent were determined using the following equation:

$$V_D = \frac{V_S\,(d_f\ -\ 1.346)}{1.01 - \ d_f}$$

where V_D is the volume of diluent to be added (PBS), V_S is the volume of stock solution to be diluted, and d_f is the desired or final density of the solution. 1.346 and 1.01 are the densities of the stock solution and PBS, respectively. 7.5 ml of PBS was transferred to an Optiseal ultracentrifuge tube (Beckman Coulter, Brea, CA) and underlaid with 7.5 ml of solution I, followed by 7.5 ml of solution II, and lastly by 7.5 ml of density-adjusted plasma. Immediately after establishing the discontinuous density gradient, the tube was centrifuged in a Beckman L5-50 ultracentrifuge with a SW28 rotor at 28,000 rpm for 24 hours at 4°C, with the brake disabled. Following centrifugation, the discrete LDL layer found between solutions of density 1.019 – 1.063 g/ml was removed by piercing the ultracentrifuge tube with a 21G needle. LDL fractions from multiple donors' plasma were pooled at this stage to limit the impact of inter-individual variation. A sterile Slide-A-Lyzer G2 Dialysis Cassette, 10K MWCO, 15 ml (ThermoFisher Scientific, Waltham, MA) was hydrated by immersion in sterile PBS for 5 minutes, and the pooled LDL was transferred to the cassette. The sample was extensively dialysed against at least 200X the sample volume of PBS to remove KBr. The dialysed LDL was passed through a 0.2 µm

filter for sterilisation and removal of aggregates. For production of oxidised LDL (oxLDL), sterile 25mM $CuCl_2$ in distilled water was added to the LDL to a final concentration of 25 μM and the mixture was incubated at 37°C for 18 hours. The oxidation was terminated by addition of EDTA to a final concentration of 0.5 mM and the oxLDL was then dialysed against at least 200X the sample volume of PBS to remove EDTA and any remaining copper. The resulting dialysate was sterilised by passage through a 0.2 μm filter and used to treat control samples.

2.7 Flow cytometry

Adherent cells were detached from plates by removal of culture medium, two washes with sterile PBS, and treatment with enzyme-free Cell Dissociation Buffer (Invitrogen, Carlsbad, CA) for 5 minutes at 37°C. Cells were counted and a volume containing 5 x 10^5 cells was transferred to separate wells of a 96-well plate. Cells were pelleted at 1500 rpm for 5 minutes and stained with 0.2 μg primary antibody or isotype control, in 100 μl PBS with 1% BSA, for 30 minutes at 4°C. The cells were then pelleted at 1500 rpm for 5 minutes and washed three times with ice cold PBS/1% BSA. For staining with unconjugated primary antibodies, cells were resuspended in 100 μl ice cold PBS/1% BSA containing 0.2 μg of conjugated secondary antibody and stained in the dark for 30 minutes at 4°C. The cells were then washed three times as previously and resuspended in 150 μl PBS/1% BSA for analysis by flow cytometry. Samples were analysed on a BDFACSCanto2 using BD FACSDiva software. Unstained cells were used to set forward scatter (FSC), side scatter (SSC) and fluorescence channel voltages such that events fit within a detectable window. In all cases, initial live cell gates were determined using the height of the FSC trace and the area of the SSC trace. In order to restrict the need for compensation, propidium iodide staining was not carried out where FSC/SSC

characteristics were sufficient to distinguish live and dead cells. Baseline fluorescence voltages were set using signal from unstained cells as zero for single fluorochrome analysis. For multiple fluorochrome analysis, compensation controls were used. 30,000 events were recorded for all experiments, and biological triplicates were performed in each case, unless otherwise stated. Data was exported to FlowJo 10 for analysis, where initial gating was carried out as described previously. Anti-CD28 (clone CD28.2) and mouse IgG1 isotype control primary antibodies (ThermoFisher Scientific, Waltham, MA), and Alexa Fluor 647-conjugated goat anti-mouse IgG secondary antibody (ThermoFisher Scientific, Waltham, MA) were used for detection of CD28 on human macrophages. Prior to staining, macrophages were treated with FcR blocking reagent (Milltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's instructions.

2.8 RNA purification

For preparation of RNA, cells were lysed in 1 ml TRIzol reagent (Invitrogen, Carlsbad, CA) and stored at -80°C before purification. Lysates were thawed on ice and 200 µl chloroform was added to each tube, followed by vortexing for 15 seconds to ensure thorough mixing. Tubes were incubated on ice for 5 minutes before centrifugation at 13,200 rpm for 15 minutes at 4°C to separate the phases. 500 µl of the upper aqueous phase containing nucleic acids was transferred to a new tube, and 500 µl of 70% ethanol prepared using RNase-free water was added and the tube mixed by inverting several times. 700 µl of this mixture was transferred to a spin column from a PureLink RNA Mini kit (Invitrogen, Carlsbad, CA) and the column was centrifuged at 12,000 *g* for 15 seconds at room temperature. The flow-through was discarded, and the remaining nucleic acid/70% ethanol mixture was added to the column followed by centrifugation at 12,000

g for 15 seconds. The flow-through was again discarded, and the column was washed with 350 µl Wash Buffer I and centrifuged at 12,000 *g* for 15 seconds. On-column DNase digestion was performed using DNase solution prepared from a PureLink DNase Set (Invitrogen, Carlsbad, CA). The DNase mixture was added to the column and the column was incubated at room temperature for 15 minutes. Following DNase treatment, the column was washed once with Wash Buffer I, and twice with Wash Buffer II, each with centrifugation at 12,000 *g* for 15 seconds. The column was dried by centrifugation at 13,200 rpm for 1 minute in a clean 2 ml collection tube. The RNA was eluted into RNase-free tubes by addition of a minimum of 50 µl dH_2O, incubation for 1 minute at room temperature, and finally centrifugation at 13,200 rpm for 1 minute. RNA concentration was determined using a NanoDrop ND-1000 Spectrophotometer (ThermoFisher Scientific, Waltham, MA), and RNA was either used immediately for reverse transcription or stored at -80°C.

2.9 RT-qPCR

For each cDNA synthesis reaction, 500 ng of RNA was diluted in 10 µl dH_2O and mixed with 1 µl 10mM dNTP, 0.5 µl 0.4 mg/ml random hexamers (Qiagen, Hilden, Germany) and 0.5 µl dH_2O (dNTP, random hexamer and dH_2O were prepared as a master mix). The mixture was incubated at 70°C for 5 minutes and then chilled on ice for 2 minutes to denature the RNA, before a pulse spin at 4°C to collect the mixture at the bottom of the tube. The reverse transcription reaction mixture was made up to 20 µl by adding 0.5 µl Bioscript reverse transcriptase (200 U/ml) and 4 µl 5X reaction buffer (Bioline, London, UK), in addition to 0.5 µl 40 U/ml Rnasin (Promega, Madison, WI) and 3 µl dH_2O. After pipetting up and down to ensure thorough mixing, and a pulse spin at 4°C to collect the reaction mixture at the bottom of the tube, the mixture was incubated at 25°C for 10

minutes followed by 42°C for 1 hour. Following incubation, the mixture was collected at the bottom of the tube by a pulse spin at room temperature, and the reaction was terminated by incubating tubes at 85°C for 5 minutes and then chilling on ice for 10 minutes. The cDNA was diluted to 500 µl in dH_2O and the diluted cDNA was either used immediately for qPCR or stored at -20°C.

Samples were analysed by qPCR using a StepOnePlus instrument (Applied Biosystems, Foster City, CA). Each 20 µl reaction contained 10 µl Fast SYBR Green Master Mix (Applied Biosystems, Foster City, CA), 0.5 µl forward primer (10 µM), 0.5 µl reverse primer (10 µM), 4 µl previously diluted cDNA, and 5 µl dH_2O. For each assay, temperature was held at 95°C for 20 seconds to denature and mix reactants, before 40 cycles of 95°C for 3 seconds for denaturation followed by 60°C for 30 seconds for annealing/extension. Primer specificity was assessed using melt curve analysis. Data was analysed using the double-delta-C_t method to determine fold-change in expression between treatment groups, and C_t values were internally normalised to either GAPDH or ACTB expression. The primers used for validation of *CD28* differential expression in normal and high glucose-treated human macrophages were: forward 5'-GCTTGCTAGTAACAGTGGCC - 3' and reverse 5' - GTAGTCACTGTGCAGGAGCC - 3'.

2.10 RNA-seq sample preparation and data analysis

For RNA-seq, RNA samples were prepared using chloroform extraction and a RNeasy Micro Kit (Qiagen, Hilden, Germany) to purify total RNA from TRIzol lysates. RNA quality was assessed using either a Tapestation 2200 with RNA ScreenTapes (Agilent Technologies, Santa Clara, CA) or a Nanodrop ND-1000 Spectrophotometer

(ThermoFisher Scientific, Waltham, MA). High-quality RNA with RIN score > 9 (from Tapestation analysis) or A260/280 = ~2.0 and A260/230 = ~2.0 - 2.2 (from Nanodrop analysis) was submitted to Oxford Genomics Centre for library preparation and sequencing. For each sample, 1 μg of total RNA was used for TruSeq library preparation with polyA selection, and the libraries were sequenced as 75 bp paired-end reads on an Illumina HiSeq 4000 with a target depth of at least 3 x 10^7 reads per sample.

Sequencing reads in fastq format were quality-checked using the seqTools[96] R package distributed through Bioconductor[97], in combination with FastQC[98]. Reads were aligned to the GRCh38/hg38 reference genome (downloaded from the UCSC Genome Browser) using the splice-aware STAR aligner[99] with the options --outFilterMultimapNmax 10, --alignSJoverhangMin 5, --alignSJDBoverhangMin 3, --outFilterMismatchNmax 10, --outFilterMismatchNoverLmax 0.04, --alignIntronMin 21, --alignIntronMax 0, --alignMatesGapMax 0, --outFilterIntronMotifs RemoveNoncanonical. The resulting BAM files were filtered using SAMtools[100] to retain only unique alignments with properly paired reads. Gene-wise read counts in each sample were obtained using featureCounts[101], a program contained in the Rsubread package[102], with a GTF file generated from GENCODE v29 annotation. The resulting count matrix was loaded as a DGEList object and analysed using edgeR[103,104]. Genes with 0 counts in all samples were excluded from further analysis. Trimmed Mean of *M*-values (TMM) normalisation[105] was used to calculate scaling factors accounting for differences in composition between libraries. After normalisation the count matrix was further filtered to remove genes with very low counts, in order to retain only genes with biologically meaningful levels of expression and prevent large changes in low counts from affecting the statistical analysis of differential expression. For all experiments the filtering strategy was that all genes where it was not possible for the gene to be expressed above the expression threshold

(counts per million (CPM) > 1) in either condition (50% of the samples) were removed. Remaining genes were tested for differential expression using the quasi-likelihood framework[106] in edgeR, including coefficients for cell treatment and donor, and genes with an adjusted p-value (Benjamini-Hochberg) < 0.05 were considered significant. Gene ontology (GO), KEGG and Reactome pathway enrichment analyses were performed using the the goana function in edgeR.

2.11 ATAC-seq sample preparation

Libraries for ATAC-seq were prepared according to either the original ATAC-seq protocol[78,107], or the Omni-ATAC-seq protocol[108]. For each sample, cells were lifted from cultureware by trypsinisation at 37°C for 3 minutes with pre-warmed trypsin-EDTA solution (Sigma-Aldrich, St. Louis, MO). Trypsin was inactivated by addition of 7 ml cell culture medium, and cells were pelleted at 1500 rpm for 5 minutes. Cells were resuspended in ice cold PBS and counted using a haemocytometer. 50,000 viable cells were transferred to a sterile 1.7 ml microtube and pelleted at 500 *g* for 5 minutes at 4°C. The supernatant was completely aspirated using two pipetting steps – down to 100 µl using a P1000 pipette, followed by removal of the final 100 µl using a p200 pipette. For Omni-ATAC-seq, subsequent steps required ATAC Resuspension Buffer (ATAC-RSB) consisting of 10 mM Tris-HCl pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$ in dH_2O. Cell pellets were resuspended in 50 µl cold ATAC-RSB containing 0.1% NP40 (Roche, Basel, Switzerland), 0.1% Tween-20 (Roche, Basel, Switzerland) and 0.01% Digitonin (Promega, Madison, WI) and the suspension kept on ice for 3 minutes to release the nuclei. The lysate was washed with 1 ml ATAC-RSB containing 0.1% Tween-20 and nuclei were pelleted by centrifugation at 500 *g* for 10 minutes at 4°C. Supernatant was carefully removed by two pipetting steps, as previously, to avoid disturbing the pelleted

nuclei. For each 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, San Diego, CA), as well as 16.5 µl PBS, 0.5 µl 1% Digitonin (0.01% final concentration), 0.5 µl 10% Tween-20 (0.1% final concentration) and 5 µl dH_2O. Nuclei were resuspended in 50 µl transposition mix and mixed thoroughly by gently pipetting up and down 6 times, followed by incubation at 37°C for 30 minutes. Reactions were mixed by gently pipetting up and down 6 times halfway through this incubation step. Transposed DNA fragments were purified using a MinElute PCR Purification Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions and eluted in 20 µl of the provided EB buffer (10 mM Tris-HCl, pH 8). Purified fragments were either immediately amplified by PCR or stored at -20°C. The primary differences between the Omni-ATAC-seq and original ATAC-seq protocols are the inclusion of additional detergents during lysis (NP40 only in original ATAC-seq) and addition of PBS and detergent to the transposition mixture in Omni-ATAC-seq.

Transposed DNA fragments were amplified in a PCR reaction containing 20 µl transposed DNA, 25 µl NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, MA), 2.5 µl primer Ad1 (25 µM), and 2.5 µl primer Ad2.x (25 µM). The reverse primer sequence (Ad2.x) contained a barcode sequence which was used to index each library before pooling and sequencing. Primer sequences used for indexing ATAC-seq libraries are listed in Table 2.1. The reaction was incubated at 72°C for 5 minutes to allow extension of both primer ends and generation of amplifiable fragments after transposition, followed by 98°C for 30 seconds to denature DNA. After the initial extension, fragments were amplified for 5 cycles at 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute. The number of additional PCR cycles required for library

amplification was determined by a separate qPCR reaction to reduce GC and size bias introduced during PCR. Each 15 µl qPCR reaction was prepared by adding 4.41 µl dH_2O, 0.25 µl primer Ad1 (25 µM), 0.25 µl primer Ad2.x (25 µM), 0.09 µl 100X SYBR Green I prepared from a 10,000X stock in DMSO (Invitrogen, Carlsbad, CA), and 5 µl NEBNext High-Fidelity 2X PCR Master Mix to 5 µl of the previously amplified PCR mixture, while the remaining 45 µl was set on ice. The cycling conditions used for qPCR were a single step at 98°C for 30 seconds, followed by 20 cycles of 98°C for 10 seconds, 63°C for 30 seconds and 72°C for 1 minute. Linear Rn was plotted against cycle number and the cycle at which the fluorescence intensity was one third of the maximum was taken as the number of additional PCR cycles. The remaining 45 µl of the original PCR reaction was amplified under the previously defined conditions for the number of cycles determined by qPCR. Library quality was assessed using the Agilent DNA High Sensitivity Kit with the Agilent 2200 TapeStation system, and successful transposition was indicated by complex fragment size distributions with characteristic nucleosome peaks every ~200bp. Indexed libraries were quantified by qPCR using a KAPA Library Quantification kit (Roche, Basel, Switzerland) compatible with Illumina sequencing libraries and pooled in equimolar amounts for sequencing as 75 bp paired-end reads on an Illumina HiSeq4000 with a target depth of at least 4×10^7 reads per sample.

Table 2.1: Primers used for PCR amplification and indexing of ATAC-seq and Omni-ATAC-seq libraries. Ad1 was used as the forward primer for all samples, and Ad2.1-8 were used to index fragments from individual samples. Indexed libraries were pooled prior to sequencing.

Name	Barcode	Sequence
Ad1		AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
Ad2.1	TAAGGCGA	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
Ad2.2	CGTACTAG	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
Ad2.3	AGGCAGAA	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
Ad2.4	TCCTGAGC	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Ad2.5	GGACTCCT	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGGAGATGT
Ad2.6	TAGGCATG	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGGAGATGT
Ad2.7	CTCTCTAC	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGGAGATGT
Ad2.8	CAGAGAGG	CAAGCAGAAGACGGCATACGAGATCCTCTCTGGTCTCGTGGGCTCGGAGATGT

2.12 ATAC-seq data analysis

Raw sequence reads in fastq format were quality-checked as for RNA-seq, and contaminating adapter sequences were trimmed using PEAT[109]. Trimmed reads were aligned to the GRCh38/hg38 reference genome (with ALT contigs removed, from ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz) using Rsubread[102] with the options minFragLength = 0 and maxFragLength = 2000 to fully capture the library complexity of each sample. Duplicate sequences were marked using Picard[110] and BAM files were filtered to remove duplicates and reads mapping to the mitochondrial chromosome, and to retain only properly-paired, uniquely aligned reads. Quality of ATAC alignments was assessed by analysing the fragment size distribution using csaw[111], with high quality samples recapitulating the ~200 bp periodicity observed during electrophoresis of the libraries prior to sequencing. In order to represent the centre of each transposition event for peak calling and differential analysis, reads aligning to the + strand were shifted +4 bp and reads aligning to the -

strand were shifted -5 bp, as per the original ATAC-seq protocol[78]. For identifying regions of accessible chromatin, sample BAM files were downsampled to the minimum depth of all libraries, merged, and peaks were called for the merged reads from all samples using MACS2[112] with options --nomodel, --shift 50, --extsize 100, and a q-value threshold of 10^{-5}. Peaks overlapping regions included in the ENCODE Blacklist[113], located on chrY, or located on unplaced or unlocalised scaffolds were removed.

Transposition events in each sample, defined as the 5' ends of the shifted reads, overlapping the merged peaks were counted using featureCounts, producing a peak-wise count matrix. Scale factors for TMM normalisation were calculated using csaw and supplied to edgeR for differential analysis using the quasi-likelihood framework. Coefficients for cell treatment and donor were included in edgeR differential analysis, and peaks with an adjusted p-value (Benjamini-Hochberg) < 0.01 were considered differentially accessible between treatment conditions. Differential peaks were annotated using ChIPseeker[114], and ATAC-seq signal in differential peaks was inspected using TMM-normalised Tn5 insertion tracks generated with rtracklayer[115] and visualised in the UCSC[116] and IGV[117] genome browsers. GO, KEGG and Reactome pathway enrichment analyses for differential peaks were performed using the annotatePeaks.pl script included with the HOMER suite of tools[118]. Motif enrichment analysis was performed using HOMER's findMotifsGenome.pl script on peaks with increased and decreased accessibility separately, using peaks with an absolute fold change < 2 and an adjusted p-value > 0.1 (i.e. non-differential peaks) as background sequences. For analysis of chromatin accessibility at transcription factor binding motifs with chromVAR[119], position weight matrices (PWMs) for HOMER 'known' motifs were used. Aggregate profiles of Tn5 insertions were centred on motif instances identified using HOMER PWMs with chromVAR's matchMotifs function. Raw 1 bp insertions +/- 500 bp of the

motif centre were counted using deepTools[120] and counts were normalised to the average signal in the terminal 100 bp at both ends of the window. For SNP enrichment analysis, the complete NHGRI-EBI GWAS catalog of trait-associated variants was downloaded from https://www.ebi.ac.uk/gwas/api/search /downloads/full. The list was filtered down to SNPs associated with disease traits of interest, and then filtered for genome wide significance ($p < 5 \times 10^{-8}$). For each trait, independent SNPs (defined here as SNPs in low linkage disequilibrium (LD), i.e. $r^2 < 0.1$, and > 500 kb from one another) were selected using PLINK[121], and all SNPs from 1000G Phase 3 CEU data in high LD with the independent SNPs (i.e. $r^2 > 0.8$) were identified, also using PLINK. This represented the full set of potentially causal SNPs for the trait under study. The linked SNPs were overlapped with differentially accessible peaks identified by ATAC-seq, and significance was determined by comparison with the overlap of 10,000 sets of random 1000G Phase 3 SNPs matched for minor allele frequency, gene density, distance to nearest gene and LD buddies (i.e. other SNPs in LD, $r^2 > 0.8$) using a binomial test.

2.13 scRNA-seq sample preparation and data analysis

Monocytes from 4 healthy donors were differentiated to macrophages in R10 supplemented with 50 ng/ml M-CSF, as described previously, and single cell suspensions were prepared by trypsinasation for 15 minutes followed by gentle flushing with a P1000 pipette. Cells were washed in PBS, resuspended in fresh PBS, and submitted to the Oxford Genomics Centre for preparation of single-cell RNA-seq libraries. Viability of submitted cells was determined to be ~90% for all buffer- and oxLDL-treated samples by staining with Trypan Blue. Barcoded sequencing libraries were generated from single cells using the 10X Genomics Chromium system with 3' chemistry (v2), with a target capture of 500 cells per sample. The resulting libraries were sequenced over 2 lanes of a

flow cell on an Illumina HiSeq 4000 with the aim of sequencing a minimum of 1 x 10^5 reads per cell. Using Cell Ranger[74] v3.1, reads from each sample produced across multiple lanes were combined and aggregated into a single gene-barcode matrix for all samples. The matrix was filtered in Cell Ranger to remove barcoded sequences likely arising from empty droplets[125], and remaining barcodes were loaded into Seurat[126] v3.1 for further analysis. Cell barcodes were further filtered by percentage read content aligning to mitochondrial or ribosomal protein genes, as a measure of overall cell quality, and filtered for total number of genes detected (> 500 and < 4000). Unique molecular identifier (UMI) counts from each sample were normalised using SCTransform[127], with effects from mitochondrial and ribosomal protein read content regressed out. For regression of cell cycle effects, cell cycle scores for each cell were calculated in Seurat using built-in cell cycle marker genes, and the difference between S and G2M scores was regressed out during re-analysis. Cells were clustered and marker genes identified by following the standard integration workflow in Seurat (https://satijalab.org/seurat/vignettes.html). Differential gene expression analysis was performed on summed counts across cells from each sample with a common annotation (i.e. treatment group, cluster) in edgeR's quasi-likelihood framework, with inclusion of coefficients for treatment and donor. Immediately prior to differential analysis, lowly expressed genes (i.e. with an average $\log_2$ CPM < 2) were removed. For GO and KEGG pathway enrichment analyses, differentially expressed or marker genes were overlapped with gene sets for each database annotation and significance of overlap was tested using Fisher's exact test with the GeneOverlap[128] package in R.

2.14 Transmission electron microscopy

Sample preparation, sectioning and imaging for electron microscopy were performed in the Electron Microscopy Facility at the William Dunn School of Pathology, University of Oxford. For preparation of each sample for standard transmission electron microscopy (TEM), 2.5 x 10^5 monocytes were seeded onto a 13 mm glass coverslip in a 24-well plate in R10 supplemented with 50 ng/ml M-CSF and differentiated to macrophages as described previously. Differentiated macrophages on coverslips were treated with 50 μg/ml oxLDL or an equivalent volume of control buffer for 48 or 72 hours. Cells were fixed for 1 hour at room temperature in primary fixation buffer (2.5% glutaraldehyde and 4% paraformaldehyde in 0.1M cacodylate buffer) then stored at 4°C overnight. Samples were washed for 5-10 minutes 5 times each with 0.1M cacodylate buffer, followed by a 15-minute wash with 50mM glycine in 0.1M cacodylate buffer, and then a final 10-minute wash with 0.1M cacodylate. Secondary fixation was carried out by incubating samples in 1% osmium tetroxide and 1.5% potassium ferrocyanide in 0.1M cacodylate buffer at 4°C for 1 hour. Samples were then washed 5 times with milliQ water for 5-10 minutes. 10 mg/ml thiocarbohydrazide (TCH) was prepared in milliQ water, and samples were incubated in TCH solution for 20 minutes at room temperature. After 3 washes with milliQ water for 10 minutes, samples were incubated with 2% osmium tetroxide in milliQ water for 30 minutes at 4°C. Samples were washed a further 3 times with milliQ before overnight incubation in 0.5% uranyl acetate at 4°C in the dark. Samples were rinsed with milliQ water 3 times before dehydration by successive 10 minutes incubations in ice cold 30%, 50%, 70%, 80%, 90% and 95% ethanol, followed by 90 minutes in 100% anhydrous ethanol. Epoxy resin infiltration of the dehydrated samples was performed by incubation in 3:1 100% anhydrous ethanol:Agar100 resin for 1 hour, followed by 1:1 ethanol:resin

for 2 hours, and then 1:3 ethanol:resin for 1 hour all at room temperature. Catalysed 100% Agar100 resin was prepared by adding 300 µl of the accelerator benzyldimethylamine (BDMA) to 10 ml resin, and samples were incubated in the catalysed resin overnight at room temperature. Following resin infiltration, samples were embedded into fresh 100% Agar100 resin and polymerised for at least 24 hours at 60°C. Ultrathin sections (90 mm) were taken using a Diatome diamond knife on a Leica UC7 ultramicrotome and mounted onto 200-mesh Cu grids. Grids were placed section side down in a droplet of Reynold's lead citrate and incubated at room temperature for 5 minutes. Mounted sections were washed by passing grids over 5 droplets of degassed water and dried before imaging on a FEI Tecnai T12 Transmission Electron Microscope.

2.15 Seahorse assays

Metabolic assays were performed using a Seahorse XFe96 Analyzer with mito stress test and glycolytic rate assay kits (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. 24 hours prior to running each assay, 40,000 primary human macrophages were seeded in each well, and all assays were performed on cells from 4 donors with 5 replicates each. For the mito stress test assay, a final well FCCP concentration of 0.5 µM was used based on results of FCCP titration during pilot experiments. All other treatments were used at the manufacturer-recommended concentration. Immediately following each assay, cell viability was determined by dual staining with calcein AM and ethidium homodimer-1 (ThermoFisher Scientific, Waltham, MA). Stained cells were incubated at room temperature for 30 minutes, before measurement in a fluorescence microplate reader with fluorescein (ex/em: 485/530 nm) and rhodamine (ex/em: 530/645 nm) optical filters for calcein AM and ethidium homodimer-1, respectively. Readings for each dye were acquired separately to minimise

fluorescence spillover. Only results from assays with consistently high viability across all samples were used.

2.16 Redox assays

For measurement of intracellular reduced glutathione (GSH) concentration, 1 x 10^6 cells per sample were lysed in 100 µl ice cold PBS/0.5% NP40 and samples were deproteinised using a Deproteinising Sample Preparation Kit - TCA (Abcam, Cambridge, UK). GSH concentrations were measured using a fluorescence based GSH/GSSH Ratio Detection Assay Kit (Abcam, Cambridge, UK), and final measured GSH concentrations were adjusted for the addition of deproteinising reagents. Assays were performed using samples diluted 1:100 in assay buffer from 4 donors with 2 replicates each. Superoxide assays were performed using a ROS/Superoxide Detection Assy Kit (Abcam, Cambridge, UK). The day before each assay, 40,000 cells were seeded into each well of a black wall/clear bottom 96-well plate and incubated overnight at 37°C and 5% CO_2. On the day of the assay, cells were washed with wash buffer included in the assay kit, and subsequently stained with superoxide detection solution, as per the kit's instructions. One treatment group consisted of cells pre-treated with 5 mM N-acetylcysteine for 30 minutes at 37°C before simultaneous treatment with 200 µM pyocyanin and staining with superoxide detection reagent. Other cells were either left untreated or treated with 200 µM pyocyanin without N-acetylcysteine pre-treatment. After 30 minutes at 37°C/5% CO_2 in the pyocyanin/superoxide detection mix, fluorescence was measured at ex/em 550/620 nm in a fluorescence microplate reader. Background fluorescence was subtracted from all readings. Assays were performed using cells from 4 donors with 4 replicates each.

Chapter 3

The human macrophage response to high glucose

3.1 Introduction

Cardiovascular disease is a major complication in diabetes, and atherosclerotic plaque forms at an accelerated rate in these individuals. The exact mechanism by which diabetes expedites atherosclerosis remains uncertain, however there is some evidence that chronic hyperglycaemia can impact disease progression directly[129]. Clinical trials have shown that intensive glycaemic control significantly reduces the risk of microvascular complications (e.g. retinopathy, nephropathy), and to a lesser extent, cardiovascular disease risk in patients with diabetes mellitus types 1 and 2, compared with conventional insulin therapy[130]. The reduction in macrovascular risk only became apparent years after the studies were concluded, and other studies investigating the effects of restoring normoglycemia on cardiovascular disease in patients with type 2 diabetes have found no significant reduction in cardiovascular disease risk with intensive therapeutic blood glucose reduction[131,132]. Therefore, though there is a link between diabetes and atherosclerosis, there is conflicting evidence surrounding the notion that current glycaemic control protocols, comprising the primary therapeutic strategy for diabetes, effectively remedy accelerated atherogenesis in diabetic patients.

Further evidence for an association between hyperglycaemia and progression of atherosclerosis has come from animal models. A murine model of diabetic atherosclerosis, in which viral infection provoked targeted T-cell mediated destruction of pancreatic β-cells in *LDLR*$^{-/-}$ mice, has shown that the rate of appearance of early signs of atherosclerosis (i.e. fatty streak formation) is increased by chronic hyperglycaemia, independent of lipid levels[133]. The exact biological pathway by which hyperglycaemia enhances initiation of atherogenesis remains unclear, however several mechanisms have been proposed. *In vitro* studies have shown an increase in the expression of adhesion

molecules by endothelial cells, and augmented cytokine secretion by monocytes/macrophages cultured in hyperglycaemic conditions [15,134]. RNA-seq has been performed with the human monocyte-like THP1 cell line cultured in high glucose, which showed upregulation of genes involved in establishing an inflammatory phenotype[135]. Additionally, there are reports that macrophages cultured in media containing elevated glucose concentrations have increased cell-surface expression of the oxLDL scavenger receptors CD36[136] and LOX-1[137], and both of these molecules have been shown to mediate foam cell formation[138,139]. The regulation of CD36 expression by high glucose has been attributed to an increase in the efficiency of CD36 mRNA translation through a glucose-sensitive 5'-UTR[136], but the mechanism of glucose-dependent LOX-1 regulation remains to be characterised.

Increased flux through the glycolytic pathway is associated with immune activation in a range of cells, including human macrophages[140–142]. This state, known as 'Warburg metabolism', is characterised by increased glucose uptake and utilisation, increased lactate output, and a reduction in TCA cycle metabolism[143]. Exposure of cells to elevated glucose levels *in vitro* has been shown to induce this state and regulates the expression of proteins involved in immune processes, such as MICA, at the level of transcription[144]. Macrophages exposed to activating stimuli (e.g. LPS) in the presence of elevated glucose show increased output of inflammatory cytokines, and Warburg metabolism has been implicated in enhancing the function of activated immune cells in high glucose environments[145]. Transcription factors linking glucose metabolism directly to transcriptional regulation have been identified, including the carbohydrate response element binding protein (ChREBP)[146]. This protein is translocated to the nucleus in response to elevated intracellular glucose 6-phosphate, although the precise mechanism remains controversial[147–149]. In the liver, ChREBP binds to the promoters, and activates

transcription, of genes involved in glycolysis and lipogenesis[150]. ChREBP is expressed in lesional macrophages, protecting against cell death and reducing inflammatory responses upon TLR4 stimulation or oxLDL exposure[151]. Interestingly, transplanted *Chrebp*$^{-/-}$ bone marrow cells accelerated atherosclerosis in a *Ldlr*$^{-/-}$ murine model, supporting a protective role for this glucose-sensitive transcription factor in cardiovascular disease[151]. How exactly this factor influences macrophage gene expression in diabetic atherosclerosis remains to be determined.

The aim of this chapter is to explore transcriptomic and epigenetic changes in primary human monocyte-derived macrophages exposed to elevated levels of glucose. The experimental approach will be RNA-seq and ATAC-seq of monocyte-derived macrophages from three individuals cultured in media containing 5 mM (normal) or 25 mM (high) glucose. The results of these differential analyses will be integrated with publicly available data from GWAS. Characterisation of gene expression and chromatin accessibility changes between normal and high glucose conditions may yield insights into some of the immune effects of hyperglycaemia in diabetic atherosclerosis, and integration with GWAS data may lead to identification of molecular mechanisms influenced by disease-associated genetic variation.

3.2 Preparation and sequencing of RNA-seq libraries

CD14+ monocytes were isolated from the peripheral blood of 3 healthy adult volunteers and differentiated to macrophages over 7 days in media containing normal (5 mM) or high (25 mM) concentrations of glucose and supplemented with 50 ng/ml M-CSF. Differentiated macrophages were lysed and total RNA purified by phenol-chloroform extraction. RNA quality was determined by spectrophotometry on a Nanodrop ND-1000 instrument, and samples with with a 260/280 ratio of ~2.0 and a 260/230 ratio between

2.0-2.2 were submitted to Oxford Genomics Centre, Wellcome Centre for Human Genetics for library preparation and sequencing. Libraries were enriched for polyA-tailed transcripts and prepared for sequencing using Illumina TruSeq reagents, before sequencing as 75 bp paired-end reads split over two lanes of an Illumina HiSeq 4000 flow cell. Reads from each lane were quality checked separately in case of lane-specific effects on sequence quality. Across both sequencing lanes, the mean number of paired-end reads per sample was 97,206,943 (Figure 3.1A and Figure 3.2A), with slightly lower counts for samples from donor 3. Quality was assessed by calculating the % GC content of all reads in each sample and comparing the distribution to the theoretical distribution calculated from the human transcriptome (GRCh38/Hg38 GENCODE v29 annotation). All samples closely followed the theoretical distribution, suggesting they were virtually free from contaminating bacterial or other ambient sequences (Figure 3.1B and Figure 3.2B). Phred scores, which provide a measure of the confidence of base calls at each position of the read, were high on average across the entire length of all reads from each lane (Figure 3.1C and D, and Figure 3.2C and D).

3.3 RNA-seq read alignment and quality control

Reads were aligned to the GRCh38/Hg38 reference genome using STAR[99], and alignments containing reads from the same sample split over separate lanes were merged to give a single alignment per sample. Alignment rate was high, with > 95% of the BAM outputs comprising mapped reads for all samples. Aligned reads were filtered to remove multi-mapping reads, as well as reads not mapped within a proper pair, resulting in > 60 million high-quality properly paired reads in all samples (Table 3.1). After filtering, each pair of reads was counted as a single sequenced fragment in further analyses. Fragments

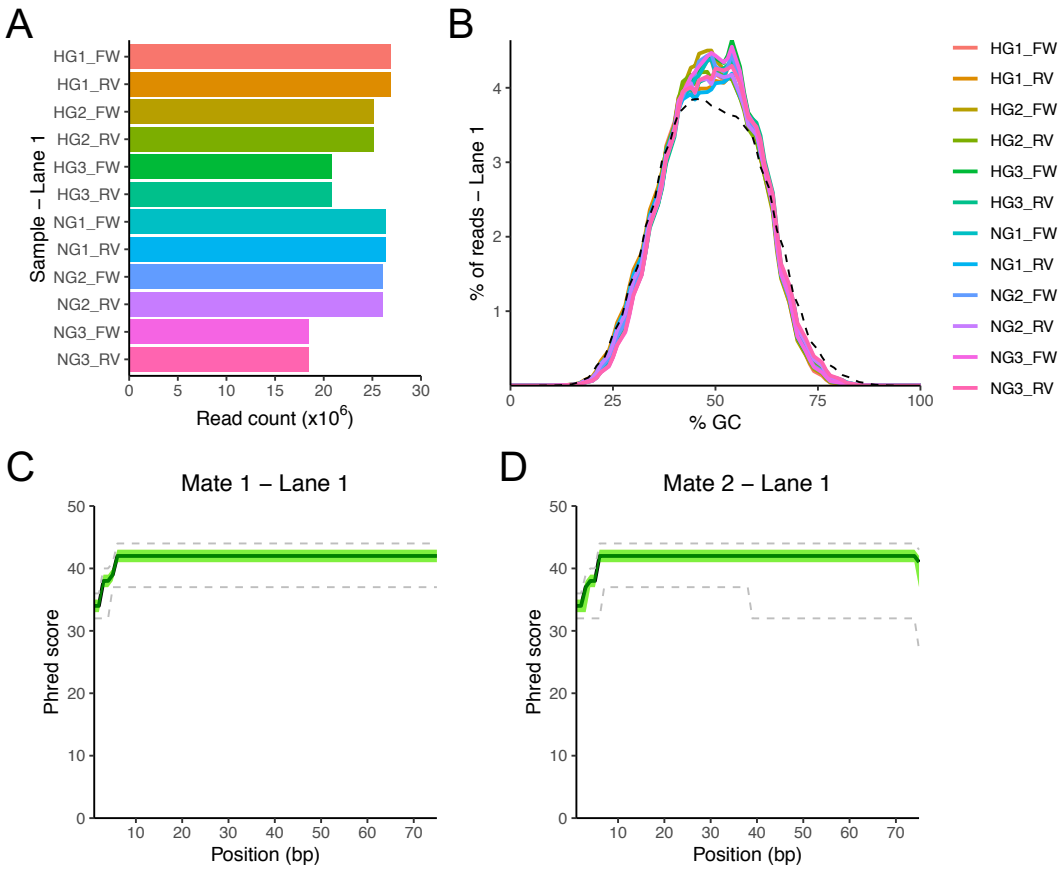

Figure 3.1: Quality control plots for RNA-seq fastq sequence files from lane 1. (A) Read counts for each sample. Mate 1 and mate 2 paired-end reads are denoted by '_FW' and '_RV' following the sample name, respectively. (B) Percentage of G or C bases called in all reads in each sample. The theoretical %GC distribution for the human transcriptome (GRCh38/hg38) is shown by the dashed black line. (C and D) Per-base Phred scores for all reads in all fastq files containing mate 1 (C) and mate 2 (D) reads. The dark green line shows the median Phred score at each position, and the light green shaded area marks the lower and upper quartiles. 10th and 90th percentiles are marked by the dashed grey lines.

aligning to genes (GENCODE v29 annotation) were counted using Rsubread's featureCounts function[101], and for all samples the majority of fragments were unambiguously assigned to annotated genes (Figure 3.3A). Normalisation of counts was performed with a trimmed mean of M-values (TMM) approach in edgeR[103,104]. This approach relies on an assumption of non-differential expression for most of the genes in the dataset, and computes scale factors for each sample to adjust the observed expression ratio (or M-value) between conditions to zero for the majority of genes. This mitigates the effect of composition bias, or the apparent decrease in counts from under-sampling

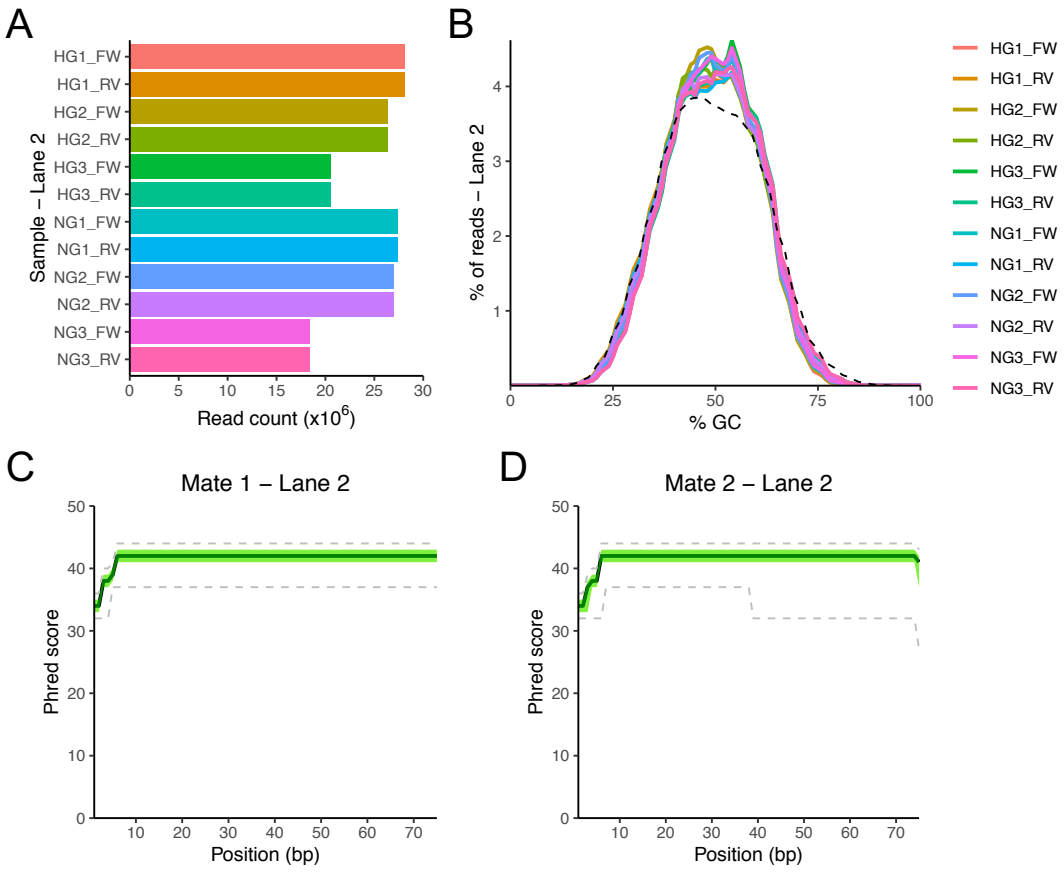

Figure 3.2: Quality control plots for RNA-seq fastq sequence files from lane 2. (A) Read counts for each sample. Mate 1 and mate 2 paired-end reads are denoted by '_FW' and '_RV' following the sample name, respectively. (B) Percentage of G or C bases called in all reads in each sample. The theoretical %GC distribution for the human transcriptome (GRCh38/hg38) is shown by the dashed black line. (C and D) Per-base Phred scores for all reads in all fastq files containing mate 1 (C) and mate 2 (D) reads. The dark green line shows the median Phred score at each position, and the light green shaded area marks the lower and upper quartiles. 10th and 90th percentiles are marked by the dashed grey lines.

caused by sequencing of highly expressed genes. The distribution of normalised counts for all genes, expressed as log-transformed counts per million (CPM), gave comparable distributions for all samples (Figure 3.3B). It also revealed that a substantial proportion of genes were expressed at a very low level (i.e. a $\log_2$ CPM value of -5 is equal to a ~1 count in a sample with a total of 30 million reads). Extremely low counts are unlikely to reflect true, biologically meaningful gene expression, and due to their discreteness can impact statistical modelling during differential analysis. As such, genes with $\log_2$ CPM < 1 in all samples were removed from the expression matrix, leaving 12,438 genes

expressed at a level above the threshold. Pearson correlation coefficients were calculated for each pair of samples, in order to explore the relationships between samples, which revealed strong correlation between samples from the same donor, rather than within treatment groups (Figure 3.4A). This observation was recapitulated by principal component analysis (PCA), which showed that the first principal component clearly contained the variance explained by expression differences between donors (Figure 3.4B). The second principal component did not separate the samples by treatment group, and neither did the third principal component (Figure 3.4C). There was clear separation of samples by treatment group in the fourth principal component, however this principal component explained only 1.5% of total variance, suggesting that high glucose exerted a minor effect on gene expression in this study (Figure 3.4D).

Table 3.1: Summary statistics for alignment of RNA-seq reads. Numbers represent records contained in BAM files from mapping paired-end reads, and the total includes primary and secondary alignments, as well as unmapped reads. Proportions are shown as percentage of total records contained in the BAM output from the STAR aligner.

Sample	Total	Mapped	% Mapped	Filtered reads	% Passed
HG1	117,794,994	112,233,828	95.28%	101,336,916	86.03%
HG2	110,993,580	107,472,166	96.83%	96,543,156	86.98%
HG3	89,457,548	85,595,790	95.68%	76,248,232	85.23%
NG1	118,411,466	116,266,160	98.19%	102,026,080	86.16%
NG2	114,078,966	111,335,706	97.60%	100,199,502	87.83%
NG3	79,484,536	75,757,626	95.31%	67,918,802	85.45%

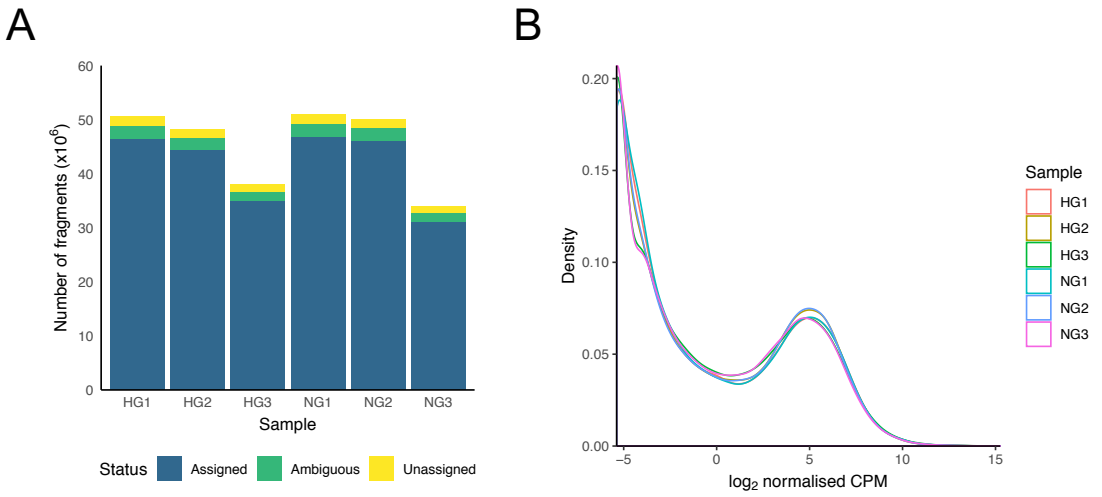

Figure 3.3: Quality control of RNA-seq counts. (A) Count of fragments assigned to genes, ambiguously assigned (i.e. to multiple genes), or unassigned (i.e. mapping outside of annotated genes). (B) Distribution of $\log_2$ normalised counts (expressed as counts per million mapped reads - CPM) for each sample prior to filtering.

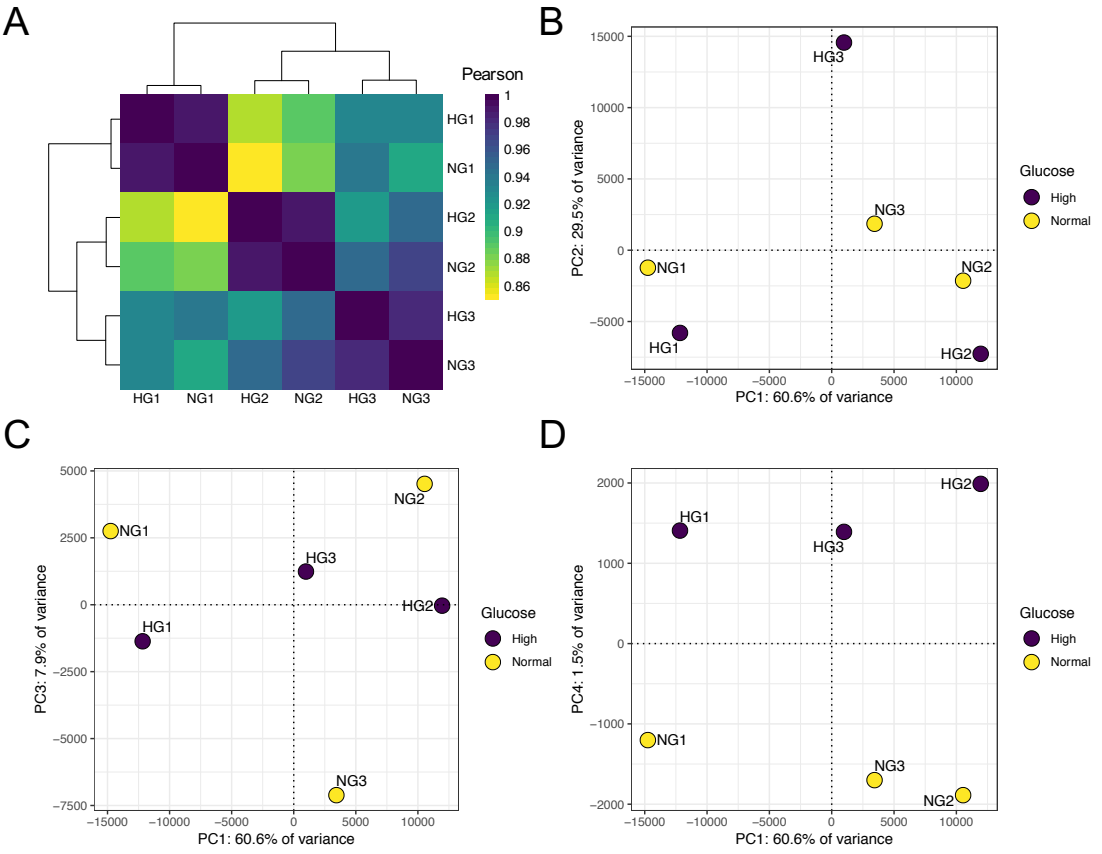

Figure 3.4: Inter-sample relationships of RNA-seq samples. (A) Heatmap showing Pearson correlation coefficients between the normalised counts for each pair of samples. Samples are clustered based on Euclidean distance between normalised counts. (B-D) Principal component

analysis of normalised counts, showing the distribution of samples in the second (B), third (C) and fourth (D) principal components, and in all cases plotted against the first principal component.

3.4 Differential gene expression analysis

To effectively model the variability from biological and technical sources in the RNA-seq count data, differential analysis was performed using the edgeR quasi-likelihood (QL) framework[106]. Negative binomial (NB) dispersion estimates were obtained for all genes and plotted against the average normalised expression ($\log_2$ CPM) estimate (Figure 3.5A). This showed a clear mean-variance trend, and variance was observed to increase for highly expressed genes. After fitting generalised linear models (GLMs) to the expression estimates, including coefficients for treatment group and the original donor to model inter-donor variability, the effect of empirical Bayesian moderation on the QL dispersion estimates was visualised and demonstrated shrinkage towards the dispersion trend, which again increased for genes with higher expression (Figure 3.5B). QL F-tests were performed to identify genes differentially expressed between normal and high glucose conditions and resulting p-values were adjusted for multiple testing (Benjamini-Hochberg procedure). There were no genes differentially expressed at a threshold of adjusted p-value < 0.1 using the QL framework. Although it was clear that glucose had at most a minor influence on gene expression, it was considered possible that the observed mean-dispersion trends at high expression for both NB and QL dispersions were restricting the power of the QL framework to detect differential expression. To explore this possibility, the standard edgeR GLM pipeline was used, which utilises NB dispersion estimates obtained for each gene directly, rather than moderating the estimates towards the mean-dispersion trend. This dispenses with the QL dispersion entirely, as there is no need to model the uncertainty in the moderation of NB dispersions. After fitting GLMs

based on gene-wise NB dispersion estimates, likelihood ratio tests were performed and resulting p-values adjusted for multiple testing (Benjamini-Hochberg). At a threshold adjusted p-value of 0.05, 65 genes were differentially expressed between normal and high glucose samples, of which 41 were more highly expressed in high glucose samples (2 > two-fold), and 24 were more highly expressed in normal glucose samples (3 > two-fold) (Figure 3.6). The 5 genes differentially expressed by more than two-fold were *GJB2*, *ASB2*, *PCDH12*, *CD28* and *CXCR2P1*. Among the upregulated genes were several with roles in immune functions (*ASB2, CCL22*, *TNFSF14*), signaling (*AK4, OCSTAMP, HCAR3*) and metabolism (*FBP1*, *HPGD*, *PLPP3*). The downregulated genes also included genes with immune functions (*CD28*, *FCER1A*, *CEBPD*) (Table 3.2 and Table 3.3). The full list of differentially expressed genes was tested for enrichment of annotations from the Gene Ontology: Biological Process and Molecular Function databases, but after correcting for multiple testing there were no terms with an adjusted p-value < 0.05.

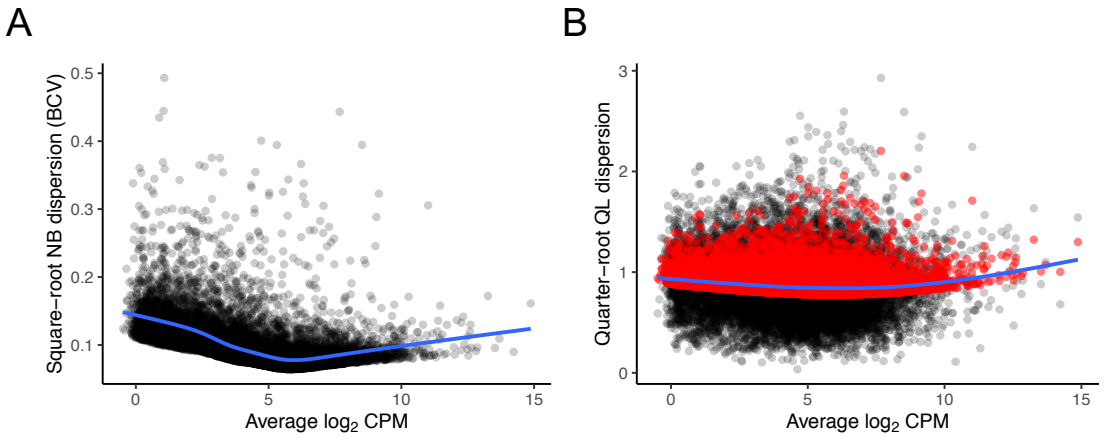

Figure 3.5: Estimation of dispersion for edgeR differential gene expression analysis. (A) Square root of the negative binomial (NB) dispersion estimates (Biological coefficient of variance, BCV) for each gene in the expression matrix against average expression. (B) Quarter-root of the quasi-likelihood (QL) dispersion estimates against average expression. Prior estimates are shown in black, and squeezed estimates produced by empirical Bayesian moderation shown in red. The mean-variance trend in each plot is shown by the blue line.

Table 3.2: Top 10 most upregulated genes in macrophages exposed to high glucose.

Symbol	log_2 high/normal	Adj. p-value	Name
GJB2	1.51	1.03E-18	Gap Junction Protein Beta 2
ASB2	1.14	3.56E-03	Ankyrin Repeat And SOCS Box Containing 2
CCL22	0.88	1.40E-03	C-C Motif Chemokine Ligand 22
AK4	0.85	4.66E-06	Adenylate Kinase 4
OCSTAMP	0.80	1.07E-03	Osteoclast Stimulatory Transmembrane Protein
HCAR3	0.80	2.77E-03	Hydroxycarboxylic Acid Receptor 3
SCN9A	0.77	1.16E-02	Sodium Voltage-Gated Channel Alpha Subunit 9
CDH23	0.77	1.02E-02	Cadherin Related 23
TDRD9	0.67	1.75E-06	Tudor Domain Containing 9
TNFSF14	0.66	1.77E-04	TNF Superfamily Member 14

Table 3.3: Top 10 most downregulated genes in macrophages exposed to high glucose.

Symbol	log_2 high/normal	Adj. p-value	Name
PCDH12	-1.16	2.46E-04	Protocadherin 12
CD28	-1.06	1.41E-03	CD28 Molecule
CXCR2P1	-1.02	8.03E-04	C-X-C Motif Chemokine Receptor 2 Pseudogene 1
GRIN3B	-0.92	1.98E-02	Glutamate Ionotropic Receptor NMDA Type Subunit 3B
SERPINB2	-0.89	2.62E-04	Serpin Family B Member 2
ADAT3	-0.82	3.10E-03	Adenosine Deaminase TRNA Specific 3
FCER1A	-0.79	2.15E-02	Fc Fragment Of IgE Receptor Ia
EDNRB	-0.67	2.11E-02	Endothelin Receptor Type B
IGFBP4	-0.67	2.85E-02	Insulin Like Growth Factor Binding Protein 4
GFRA2	-0.65	5.21E-03	GDNF Family Receptor Alpha 2

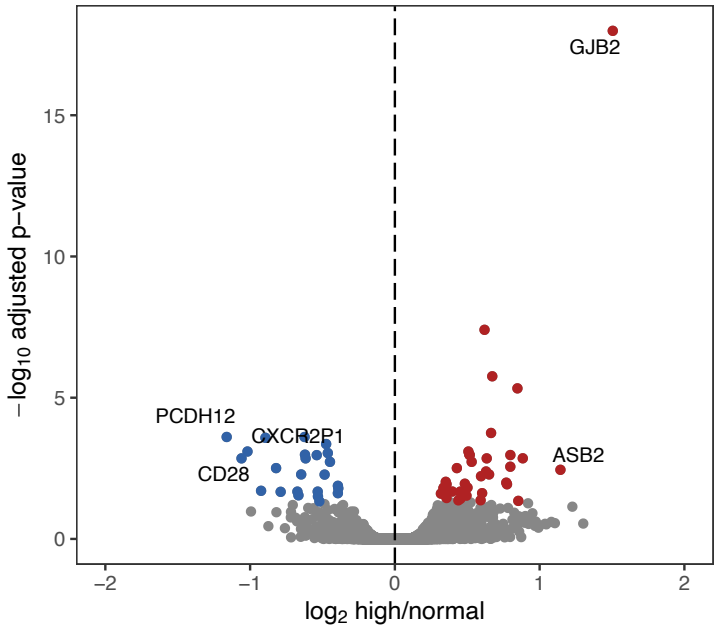

Figure 3.6: Volcano plot of differential expression analysis results. Points mark the $\log_2$ fold change in gene expression between high glucose and normal glucose against the adjusted p-value. Blue and red points represent downregulated and upregulated genes with an adjusted p-value < 0.05, respectively.

3.5 Human macrophage CD28 is downregulated by high glucose

Given the lack of evidence for differential expression using the QL framework in edgeR for RNA-seq analysis, differential expression was validated by RT-qPCR. From the list of differentially expressed genes identified using the standard GLM pipeline, *CD28* was selected for validation. The encoded receptor is well studied in its co-stimulatory role during activation of T lymphocytes[152], and expression has been reported in other immune cell lineages, such as eosinophils[153–155]. It has also been linked to metabolism of glucose, with CD28 ligation increasing glycolytic flux in activated T lymphocytes in a PI3K/Akt-dependent manner[156,157]. To validate the RNA-seq result, macrophages from 9 independent donors were cultured in normal or high glucose media, and cDNA libraries were prepared from extracted total RNA. RT-qPCR showed that *CD28* expression was decreased in all high glucose samples relative to the normal glucose sample from the

same donor ($p < 0.0001$, Student's *t*-test), and the expression in high glucose samples was 58% the expression in normal glucose samples, on average (Figure 3.7A). Expression of the CD28 receptor on the macrophage cell membrane was confirmed by flow cytometry, and the CD28-specific fluorescence was decreased on macrophages exposed to high glucose relative to normal glucose (Figure 3.7B). This supports a link between CD28 and glucose metabolism and is the first demonstration of CD28 expression on the surface of human macrophages.

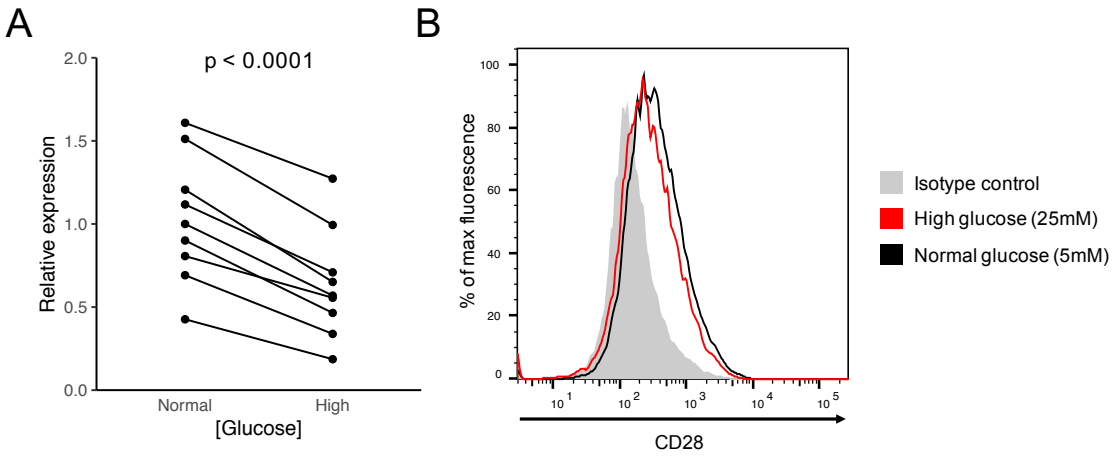

Figure 3.7: High glucose downregulates CD28 in human macrophages. (A) Expression of CD28, determined by RT-qPCR, in primary monocyte-derived macrophages from 9 healthy donors cultured in either normal or high glucose media. Data were normalised to GAPDH expression and analysed using the ΔΔCt method. Displayed values are relative to the median expression in macrophages in the normal glucose treatment group. P-value was determined using a paired Student's t-test. (B) Flow cytometry analysis of CD28 expression on the surface of human macrophages. Data are representative of cells from 3 donors.

3.6 Preparation and sequencing of ATAC-seq libraries

To determine the effect of high glucose on macrophage chromatin accessibility, and to identify changes associated with differentially expressed genes, ATAC-seq was performed using normal and high glucose macrophages from the same three healthy donors as for RNA-seq. Prepared libraries were submitted to Oxford Genomics Centre and sequenced as 75bp paired-end reads on a single lane of an Illumina HiSeq 4000 flow cell. Samples were sequenced to a minimum depth of 69,955,700 paired-end reads, and there was notable variability in the library size between samples (Figure 3.8A). This was likely due to the inaccuracy of spectrophotometric quantification for ATAC-seq libraries, leading to poorly balanced library molarity during pooling. The percentage of GC content for each read was calculated and the distribution of all reads was compared to a theoretical distribution of reads extracted from the human genome (GRCh38/hg38) *in silico* (Figure 3.8B). For ATAC-seq reads from all samples, the distribution was shifted towards a higher average % GC relative to the theoretical distribution, however the distribution was approximately Gaussian and consistent between samples, so the higher average % GC was not expected to be problematic for differential analysis. A probable explanation is the enrichment of sequences from promoter regions in the ATAC-seq data, which often overlap CpG islands. Phred quality scores for base calls were analysed and determined to be high on average across the entire length of all reads, with a slight decrease in quality for reverse or 'mate 2' reads (Figure 3.8C and D). Sequences matching Illumina adapters were detected in all samples and these sequences were trimmed in order to improve the mapping of shorter fragments expected from insertion of multiple transposases between nucleosomes.

3.7 ATAC-seq read alignment and quality control

Trimmed reads were aligned to GRCh38/hg38 using Rsubread[102] and were mapped at a rate of > 90% for all samples (Table 3.4). On average 13.5% of the alignments in each sample mapped to the mitochondrial chromosome (chrM). Alignments with identical 5' and 3' coordinates were considered far more likely to arise by duplication during PCR amplification of libraries, rather than by multiple pairs of transposase insertions producing duplicate fragments, and these alignments were marked for removal using Picard. After filtering to remove mitochondrial, duplicate, multi-mapping and improperly paired reads, samples contained a minimum of 41,500,392 reads. The significant loss of reads during filtering was largely due to mitochondrial contamination and PCR duplication, which are known issues with ATAC-seq. In addition, an average of 7.8% of reads were unmapped across all samples, and the cause of this was unknown as there was no obvious spike in high GC content, which is indicative of contamination with reads derived from bacterial DNA, and there was no significant improvement in alignment rate

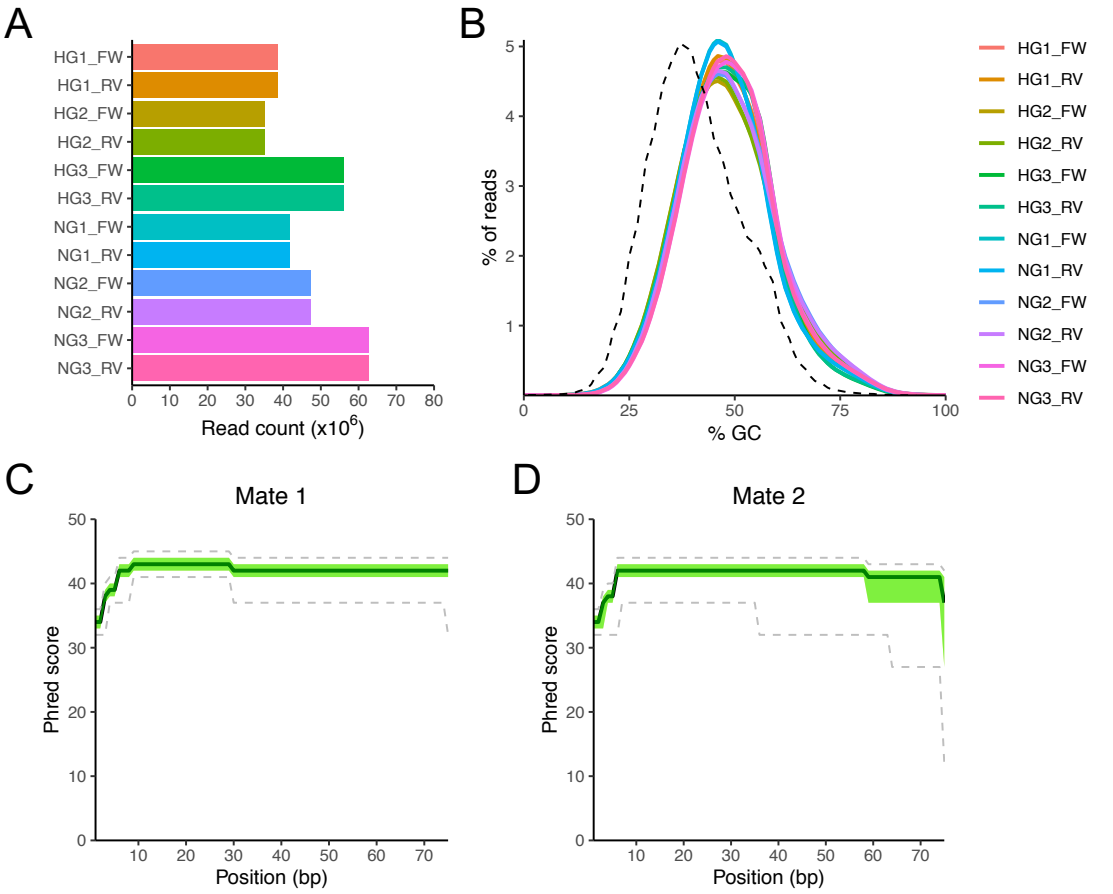

Figure 3.8: Quality control plots for adapter-trimmed ATAC-seq fastq sequence files. (A) Read counts for each sample. Mate 1 and mate 2 paired-end reads are denoted by '_FW' and '_RV' following the sample name, respectively. (B) Percentage GC content in all reads in each sample. The theoretical %GC distribution for the human genome (GRCh38/hg38) is shown by the dashed black line. (C and D) Per-base Phred scores for all reads in all fastq files containing mate 1 (C) and mate 2 (D) reads. The dark green line shows the median Phred score at each position, and the light green shaded area marks the lower and upper quartiles. 10th and 90th percentiles are marked by the dashed grey lines.

when using alternative aligners (e.g. Bowtie2[158]). Despite the substantial loss of reads, samples were still of reasonable depth to infer differences in chromatin accessibility between samples, based on the recommendations of the published protocol. Sample quality was further characterised by assessing the size distribution of the sequenced fragments. The expectation from successful transposition is a periodic distribution with peaks every ~150-200 bp, corresponding to fragments generated by insertion of Tn5 transposases either side of one or multiple nucleosomes, and this was observed for all

samples (Figure 3.9). For two of the samples, the density of ~200 bp and ~400 bp fragments was higher than for other samples, which suggested some differences in the efficiency of transposition between samples. However, these two samples originated from the same donor and due to the paired experimental design, this was therefore not expected to impact downstream analysis. Another quality metric for ATAC-seq samples is transcription start site (TSS) enrichment score, which is used by the ENCODE consortium as a measure of signal-to-noise ratio. The score for each sample was calculated from an aggregate distribution of reads in 2000bp windows centred on Refseq GRCh38/hg38 TSSs. Signal was normalised to the average coverage at each end of the window (i.e. in 100bp bins for 200bp of averaged coverage total), and the maximum fold change above background signal for each sample was taken as the TSS enrichment score. The minimum acceptable TSS enrichment score recommended by ENCODE is 5, with ideal scores > 7, and scores for all samples were > 7 (Figure 3.10).

Table 3.4: Summary statistics for alignment of ATAC-seq reads. Percentages shown are proportions of all alignments for each sample ('Total'), which includes unmapped and multi-mapping reads. 'Filtered reads' shows the final number of properly-paired reads taken forward for full analysis.

Sample	Total	Mapped	% Mapped	Marked duplicates	% Marked	Filtered reads	% Passed
NG1	83,371,726	75,882,272	91.02%	26,940,103	35.50%	43,049,892	51.64%
NG2	94,851,376	89,383,059	94.23%	20,844,368	23.32%	62,291,526	65.67%
NG3	125,315,774	114,798,332	91.61%	35,815,878	31.20%	71,132,454	56.76%
HG1	76,878,344	69,840,697	90.85%	22,727,062	32.54%	41,500,392	53.98%
HG2	69,955,700	65,789,324	94.04%	12,988,658	19.74%	47,913,928	68.49%
HG3	112,203,792	102,413,313	91.27%	28,914,133	28.23%	66,104,798	58.91%

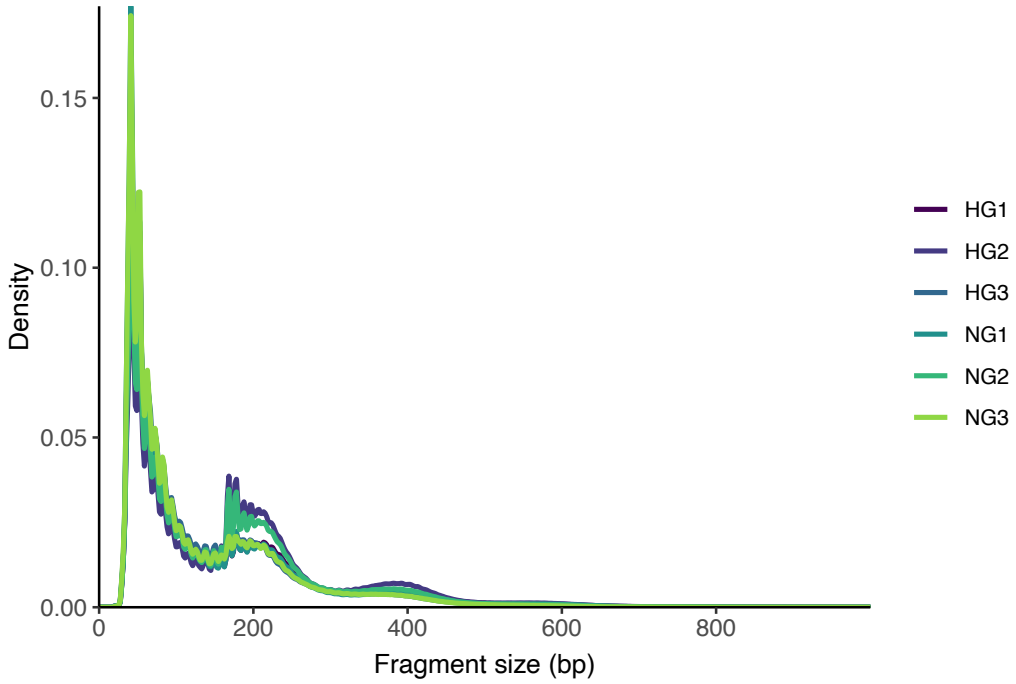

Figure 3.9: Distribution of sequenced fragment sizes for all ATAC-seq sample.

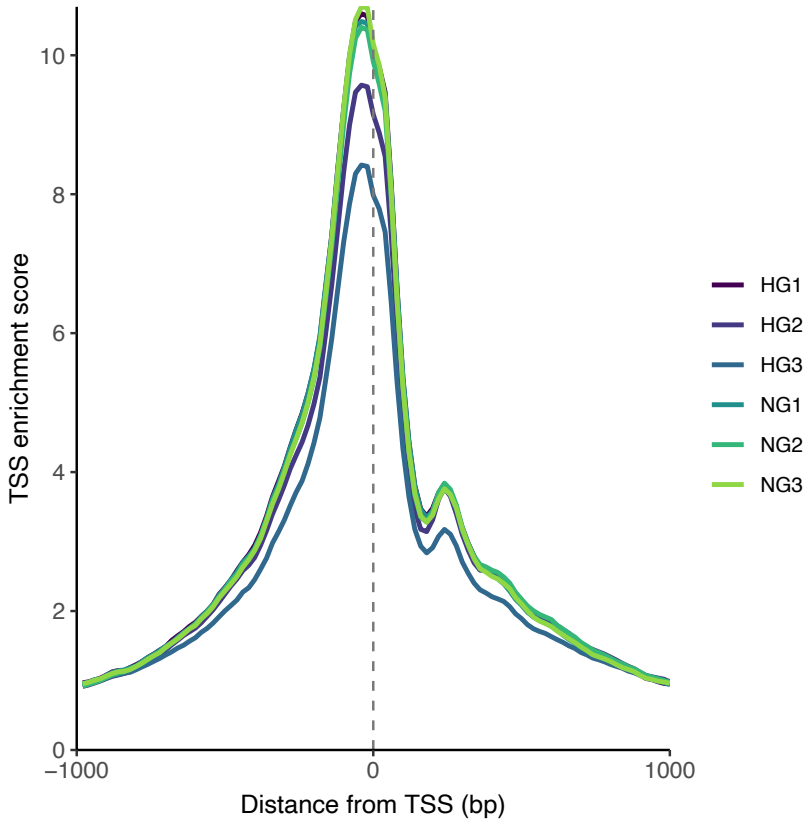

Figure 3.10: TSS enrichment plots for all ATAC-seq samples. The score at each position is calculated from the coverage in 2000bp windows centred on GRCh38/hg38 Refseq TSSs and normalised to background signal in the terminal 100bp bins.

3.8 Differential chromatin accessibility analysis

Accessible regions across the entire genome were identified by merging sample BAM files and calling peaks of signal with MACS2[112], thus producing a set of consensus accessible regions across both conditions for input to differential analysis. This approach was selected over windowed methods to allow the use of edgeR's QL framework during count-based differential analysis, and over other consensus peak-based methods to avoid loss of type I error control by pre-selecting 'differential' or variable peaks during preliminary filtering[159]. An initial set of 345,816 peaks were identified by MACS2, which were filtered to remove peaks on chrY, peaks overlapping ENCODE blacklisted regions, and peaks with an adjusted p-value > 10^{-5}. The remaining 246,889 accessible regions were representative of the normalised signal above background, based on visual inspection of bigwig signal tracks in the IGV genome browser (Figure 3.11A). Peaks were annotated with overlapping genomic features based on GENCODE v29 gene annotations, which revealed that peaks were primarily situated in introns (47.9%) or in distal intergenic regions (29.8%) (Figure 3.11B). A smaller proportion of peaks were located in gene promoters (14.2%, with promoters defined as the region -1000 bp to +200 bp from the transcription start site for each gene), and all other annotations made up the remaining ~8% of peaks. Differential analysis with edgeR revealed that the largest changes in chromatin accessibility occurred at regions with relatively low average signal (expressed as $\log_2$ counts per million mapped reads), and after correcting for multiple testing there were zero peaks with an adjusted p-value < 0.1 (Figure 3.11C). Regions with relatively large fold changes were attributed to variable peaks with low average signal, and therefore unlikely to be truly responsive to high glucose. The observed changes in gene expression could therefore not be associated with changes in chromatin accessibility, as there was no significant genome-wide effect of high glucose.

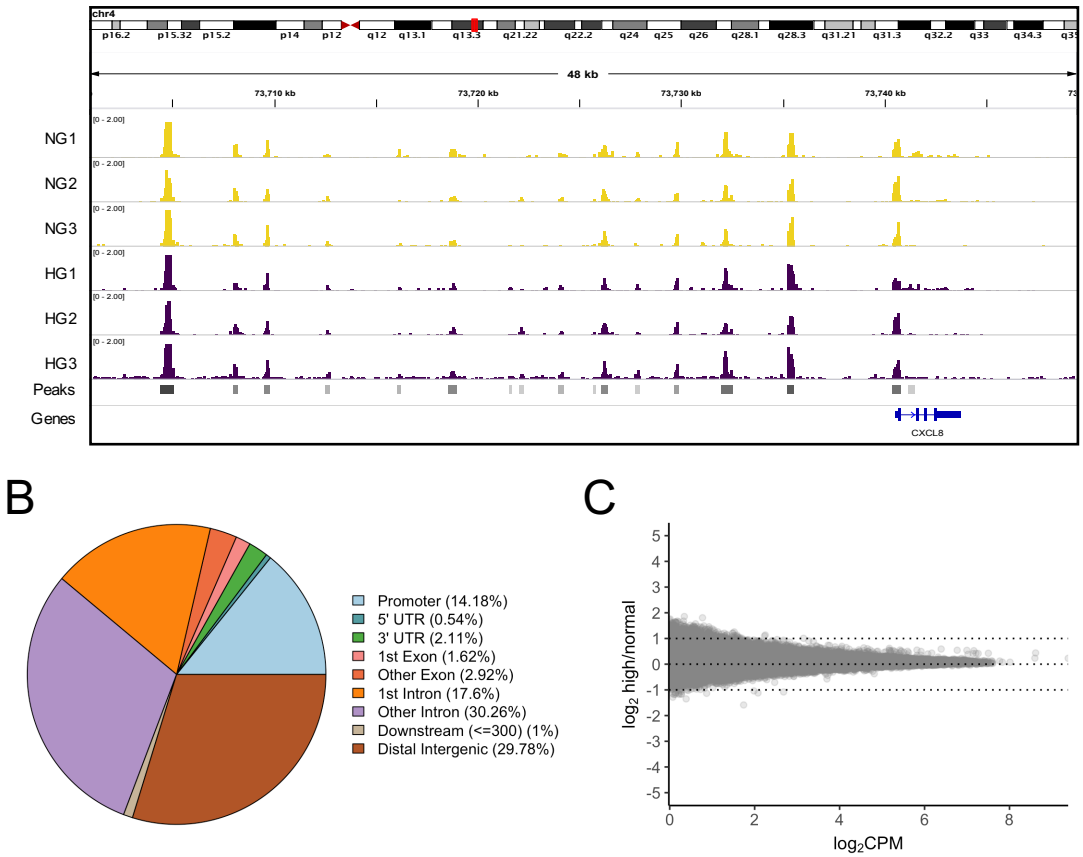

Figure 3.11: Chromatin accessibility in macrophages is not altered by high glucose. (A) Normalised Tn5 transposase insertion tracks showing open chromatin regions at an example locus upstream of CXCL8. Also shown are peaks called by MACS2, which clearly mark regions with significant ATAC-seq signal. (B) Genomic location of peaks relative to GENCODE v29 annotated genes. (C) Plot of fold change in accessibility between normal and high glucose macrophages against the average peak signal across all samples, expressed as $\log_2$ counts per million mapped reads (CPM).

3.9 Discussion

The mechanism of accelerated atherosclerosis in diabetes mellitus remains controversial. Although there is an association between high blood glucose and vascular disease, even in patients without diabetes[160], a number of metabolic effects beyond hyperglycaemia are also present in diabetics, and clinical manifestation of diabetes mellitus (particularly type II) often only occurs after many years of sustained metabolic stress. A common confounding factor in studies of diabetic atherosclerosis is dyslipidaemia, in which high

levels of triglycerides, low levels of high-density lipoprotein (HDL) and an abundance of small dense LDL particles could all contribute to disease progression[161]. As such, isolating the effects of hyperglycaemia on atherosclerotic lesions *in vivo* has proven challenging.

To address this, a number of studies have treated cell types found in atherosclerotic lesions, of which macrophages are the most abundant, with high glucose *in vitro,* and reported a number of effects consistent with enhanced atherogenesis. Many of these studies focus on the regulation of scavenger receptors for oxLDL (e.g. CD36, LOX-1)[137–139] or inflammatory cytokines (e.g. TNF-α, IL-1β)[162–164] by glucose, as these molecules have been clearly linked to atherogenesis in previous studies and could provide a straightforward link between hyperglycaemia and atherosclerosis. Unbiased, genome-wide studies of the effects of high glucose on human macrophages are limited, and existing reports of glucose-dependent proinflammatory stimulation have relied on the use of immortalised cell lines such as THP-1[135], which can exhibit substantial differences in activity from primary human cells[165]. Some of these studies also propose roles for epigenetic modifications in the macrophage response to high glucose, which may have implications for the establishment of trained immunity and innate immune 'memory'[135,166].

In this study, the effect of high glucose on primary human monocyte-derived macrophage gene expression was characterised by RNA-seq. There was no significant change in the expression of genes encoding scavenger receptors or inflammatory cytokines between normal and high glucose macrophages in this study, and significant expression changes were limited in number and magnitude. These changes did not appear to be mediated by changes in chromatin accessibility, as analysis of genome-wide chromatin accessibility

by ATAC-seq showed that there were no significant changes between normal and high glucose macrophages. Of unknown relevance to atherogenesis but of wider immunological interest, *CD28* was found to be significantly downregulated in human macrophages exposed to high glucose, and this finding was confirmed at the mRNA level by RT-qPCR, and at the protein level by flow cytometry. The CD28 receptor is well-studied for its costimulatory role during activation of T lymphocytes but reports of this molecule's expression by human macrophages are absent from the literature. There is a precedent for CD28-signaling in cells other than T lymphocytes, as ligation of CD28 on the surface of human eosinophils has been reported to induce secretion of a range of cytokines, including IL-2, IL-6, IFN-γ and IL-13[153–155]. CD28 has been linked to glucose metabolism in T lymphocytes, as costimulation via CD28 is required for increased glycolytic flux during activation. The significance of CD28 expression, signaling and relation to glucose metabolism in human macrophages remains to be determined, but this may represent a novel pathway in macrophage immunometabolism. Overall, treatment of human macrophages with high glucose did not significantly alter gene expression or chromatin accessibility and did not confer a proatherogenic phenotype *in vitro*. It is therefore unlikely that alteration of macrophage activity by glucose alone can explain the early onset of atherosclerosis in patients with diabetes.

Chapter 4

Heterogeneity in the human macrophage response to oxLDL

4.1 Introduction

Monocytes and macrophages are central to the development of atherosclerosis through their recruitment to and transmigration across activated endothelium, endocytosis of modified lipoproteins contained in the subendothelial space, and ultimately formation of lipid-laden 'foam cells'[10,167,168]. Cells of this lineage demonstrate remarkable plasticity, both *in vitro* and *in vivo*, and are highly adaptable to their microenvironment[169–171].

Different monocyte subsets have been identified in humans, characterised primarily by differential expression of the cell-surface markers CD14, a TLR4 co-receptor, and CD16, the type III Fcγ receptor, and even within these subsets heterogeneity has been observed[169,172]. $CD14^{++}CD16^{-}$ monocytes are the predominant subset in circulation, comprising ~80-95% of total monocytes[173]. These cells function primarily as phagocytes, and express high levels of the MCP-1 receptor CCR2 and the adhesion molecule L-selectin, facilitating their rapid recruitment to sites of inflammation in the vasculature[174]. The remainder of circulating monocytes are classified as either $CD14^{++}CD16^{+}$ or $CD14^{+}CD16^{++}$, of which the former exhibit moderate expression of CCR2 and secrete large quantities of the inflammatory cytokines IL-1β and TNF-α in response to LPS[175,176]. They also express CCR5, which is a receptor for the T cell chemotactic molecule CCL5 and implicates monocytes of this subset in activation of T cells and amplification of the inflammatory response[177]. The third identified subset, $CD14^{+}CD16^{++}$ monocytes, are regarded as 'patrolling' monocytes with low expression of CCR2, but high binding affinity for the endothelium[178,179]. The $CD14^{++}CD16^{-}$ subset is considered the major contributor to the development of atherosclerosis, due largely to their similarity to the murine $Ly6C^{hi}$ monocyte subset, including high expression of CCR2 and chemotaxis towards MCP-1, which is produced at sites of progressing lesions[180–182]. These

monocytes are presumed to comprise the bulk of infiltrating immune cells, differentiating to give rise to most of the intralesional macrophage population, but this has yet to be definitely proven and may differ depending on the stage of disease[183]. The presence of circulating intermediate $CD14^{++}CD16^{+}$ monocytes has been shown to independently predict cardiovascular events in patients[184], but the precise role of these cells in atherosclerotic lesions remains undefined.

Heterogeneity and plasticity of macrophages enables their diverse functions in pathogen destruction, inflammatory signalling and tissue repair[169]. Functional polarisation of cultured macrophages in response to defined molecular signals has been widely reported, and the two states most studied have been designated M1 and M2[171,185]. Macrophages adopt the M1 phenotype in response to stimulation with both LPS, which binds to TLR4 and results in NF-κB activation, and IFN-γ, which signals through JAK1 and JAK2 to activate the transcription factor STAT1[186,187]. These cells exhibit increased secretion of inflammatory cytokines, including IL-1β, IL-6 and TNF-α, as well as enhanced phagocytosis and intracellular killing through the production of reactive oxygen species[188,189]. These features render M1 macrophages effective responders to bacterial infection, and this state typifies the pro-inflammatory potential of macrophages. The M2 phenotype represents an alternative functional state of macrophages and is induced by treatment with a combination of IL-4 and IL-13[190,191]. Overall these cells are anti-inflammatory and engage in tissue repair pathways, secreting growth factors and chemokines to recruit fibroblasts and facilitate deposition of extracellular matrix and wound healing[192,193]. The establishment of the M1/M2 paradigm has resulted largely from early *in vitro* studies of murine macrophages treated with defined activating stimuli, but the modern view of macrophages *in vivo* and in humans has challenged the notion of macrophage 'polarisation'. Rather, it has been proposed that macrophages exist in a

spectrum of functional states depending on diverse signals in their microenvironment[194,195]. Furthermore, the M1 and M2 states do not represent terminal stages of differentiation, rather they are transient states in which repertoires of gene expression are induced to drive a response appropriate to the stimulus. After the stimulus is removed, macrophages will revert to a state designated 'M0', or even switch to a functionally opposite state in the presence of new stimuli[196,197]. For *in vitro* studies, macrophages are typically differentiated from circulating monocytes by treatment with macrophage colony-stimulating factor (M-CSF), which some have argued primes macrophages towards a M2-like state[198,199]. Production of M-CSF locally in the vessel wall by smooth muscle cells and endothelial cells has been demonstrated to contribute to atherogenesis, presumably by inducing the differentiation of infiltrating monocytes to macrophages[45]. Macrophages derived *in vitro* by treatment with M-CSF therefore provide a commonly used model of intra-lesional macrophages, albeit in a naïve state likely to be present only in early-stage lesions. These cells exhibit a range of different morphologies within cultured populations (Figure 4.1), but the impact of this heterogeneity on macrophage gene expression remains unexplored.

Macrophage subtypes unique to atherosclerotic lesions have been identified, such as the Mox phenotype identified in murine models of atherosclerosis, which exhibit Nrf2-mediated expression of genes involved in redox homeostasis in response to oxidised phospholipids[200]. The relevance of these cells in human disease however remains uncertain. Macrophages isolated from human atheromata display mixed expression of markers used in the M1/M2 classification, and there is some evidence that these subtypes are differentially distributed both spatially and temporally within developing lesions[201–203]. A number of studies have investigated the response of human macrophages to oxLDL *in vitro*, and these have shown oxLDL-induced expression of oxLDL-binding scavenger

receptors (CD36, MSR1)[204], cholesterol efflux transporters (ABCA1, ABCG1)[205] and inflammatory cytokines (MCP-1, IL-1β, TNF-α)[54,206]. It has also been shown that oxLDL induces widespread changes in macrophage chromatin accessibility, and that these changes occur close to genes with roles in both inflammation and cholesterol homeostasis[53]. To date, none of these studies have endeavoured to dissect the role of human macrophage heterogeneity in the response to oxLDL.

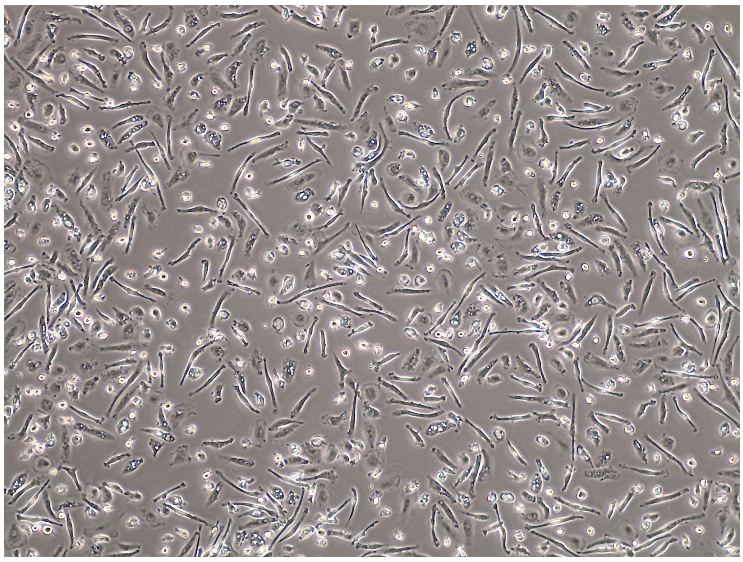

Figure 4.1: Morphological heterogeneity of human monocyte-derived macrophages in culture. CD14+ monocytes were purified from peripheral blood of healthy adult donors and differentiated to macrophages in the presence of 50 ng/ml M-CSF over 7 days.

The work described in this chapter aims to characterise the heterogeneity in primary human monocyte-derived macrophages using scRNA-seq of cells from four individuals treated with oxLDL or control buffer, and to identify differentially expressed genes between macrophage subtypes which could provide methods of distinguishing these subtypes *in vivo*. Additionally, it aims to determine whether different macrophage subtypes demonstrate a unique or conserved response to oxLDL, which may have implications for the development of atherosclerosis.

4.2 Purification and sequencing of RNA from single human macrophages

Primary human macrophages from 4 donors were treated with 50 μg/ml oxLDL or an equivalent volume of control buffer for 48 hours (8 samples in total), after which cells were lifted and resuspended in PBS before submission to Oxford Genomics Centre for library preparation and sequencing. Gene expression libraries for all 8 samples were prepared using 10X Genomics' Chromium Single Cell 3' v2 chemistry with a target capture rate of 500 cells per sample, and resultant libraries were sequenced over two lanes of a HiSeq 4000 (Illumina) flow cell for a target depth of ~100,000 reads per cell. Sequence files returned in fastq format were processed using the 10X Genomics Cell Ranger[74] software package to differentiate between reads originating from transcripts contained in intact cells and reads produced from ambient RNA contained in empty droplets. For further analyses, only barcodes with total unique molecular identifier (UMI) counts deviating significantly from the background distribution, as determined within Cell Ranger, were retained. Gene-wise counts for each cell in each sample were obtained using cellranger count with the GRCh38 reference provided by 10X Genomics (v3.0.0, built from Ensembl GRCh38 release 93). Preliminary statistics from cellranger count indicated that the target number of cells was achieved for all samples, and target depth was achieved in all but 2 samples. This resulted in detection of ~1300-2000 genes per cell on average across all 8 samples (Table 4.1).

Table 4.1: Gene expression metrics for all samples calculated in Cell Ranger.

Sample	Estimated no. of cells	Mean reads per cell	Median genes per cell
B1	685	138,028	1,630
B2	629	126,208	1,864
B3	748	91,707	1,358
B4	849	97,111	1,416
O1	601	140,021	1,899
O2	680	110,038	1,952
O3	572	151,758	2,016
O4	742	103,149	1,810

4.3 Quality control and selection of cells

Filtered feature-barcode matrices containing raw UMI counts for each gene (feature) in each cell (barcode) were loaded into Seurat v3[126] for quality control and further analysis. A common cell quality metric for scRNA-seq is the number of detected genes (i.e. genes with a count of at least 1 UMI), as low-quality cells may give rise to low numbers of detected genes, and droplets containing multiple cells could inflate the number genes detected for a particular barcode. Additionally, the percentages of total reads aligning to mitochondrial or ribosomal protein genes are frequently used as a measure of cell quality, as high relative counts for these genes can indicate cell death or damage. For all samples there was a linear relationship between the total UMI counts and the number of genes detected up to a threshold of 4000 detected genes, after which the increase in number of detected genes diminished with increasing total UMI count (Figure 4.2A). Therefore, a cut-off of 4000 detected genes was selected to remove barcodes from droplets suspected to contain multiple cells. A lower threshold of 500 detected genes was chosen to remove cells assumed to be low-quality. For mitochondrial and ribosomal read content, there was no significant difference in the average proportion between treatment groups (mitochondrial $p = 0.61$, ribosomal $p = 0.16$, paired t-test). A small number of cells did

contain large proportions of mitochondrial or ribosomal reads and fell outside of the observed distributions for each sample, and an upper threshold of 3 median absolute deviations calculated from all cells across all samples was chosen to exclude these cells (Figure 4.2B and C). After filtering cell barcodes using the metrics and thresholds described, all samples retained more than 500 cells, and as expected these high-quality cells contained a larger number of detected genes, on average (Table 4.2).

Table 4.2: Statistics for high-quality cells. Cells were selected based on number of detected genes (between 500 and 4000) and proportion of reads mapping to mitochondrial or ribosomal protein genes (less than 3 median absolute deviations for all cells across all samples).

Sample	Filtered cells	Median genes per cell
B1	605	1,703
B2	513	1,954
B3	629	1,429
B4	730	1,457
O1	504	2,005
O2	521	2,138
O3	509	2,030
O4	634	1,819

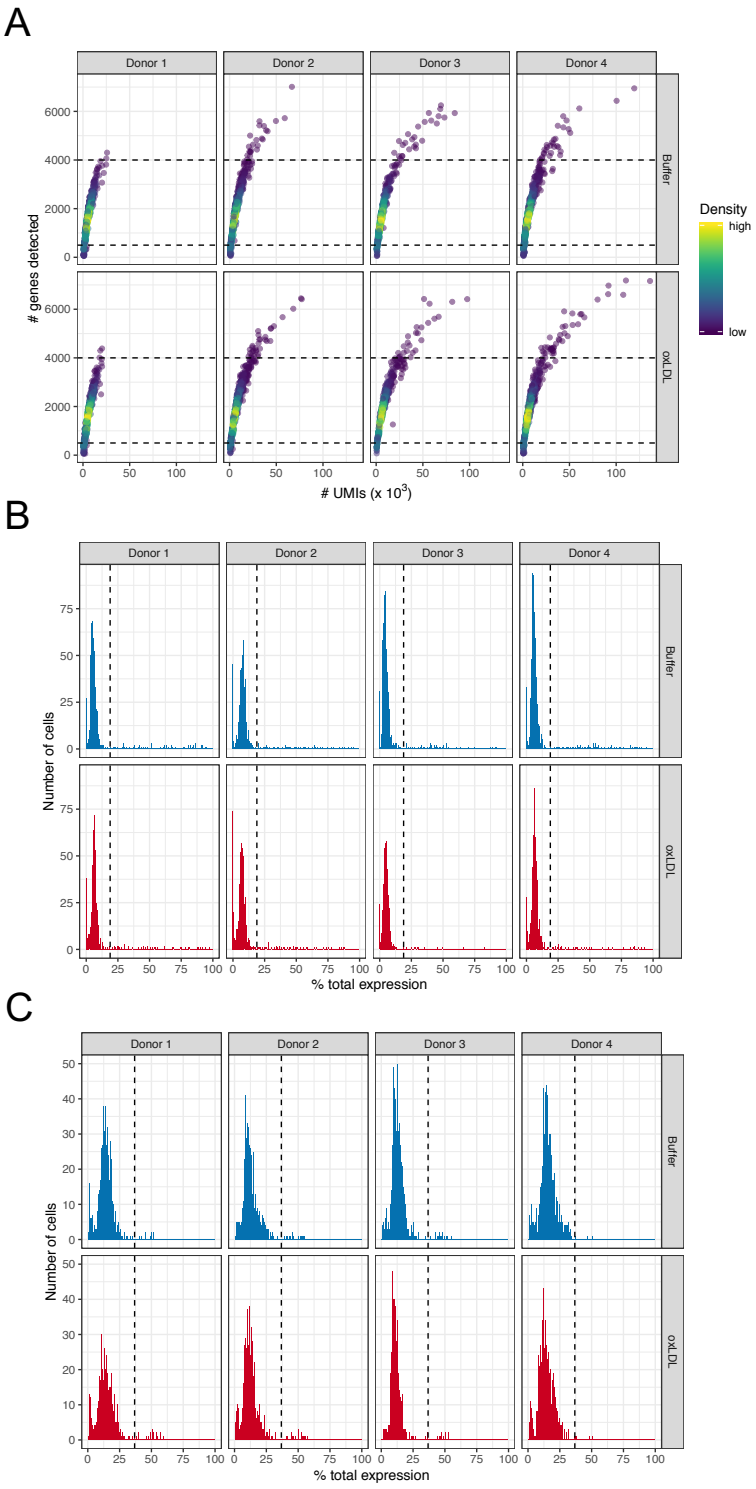

Figure 4.2: Quality metrics used for filtering in all scRNA-seq samples. (A) Correlation between total counts and number of genes detected for all samples. Dashed black lines indicate thresholds used in filtering to exclude low-quality cells and suspected multiplets. (B) Distribution of mitochondrial read content for all samples. Dashed black lines mark the upper threshold used in filtering, positioned at 3 median absolute deviations calculated from all cells across all samples. (C) Distribution of ribosomal protein read content for all samples. Dashed black lines again mark the upper filtering threshold, and this was calculated using the same approach as in (B).

4.4 Sample integration and initial clustering

In order to identify macrophage subtypes conserved between donors, and across treatment groups, the Seurat integration workflow was used to identify 'anchors' between the samples and align cells with related patterns of gene expression. Prior to integration, samples were normalised using the SCTransform[127] function in Seurat, which derives Pearson residuals from regularised negative binomial regression and uses these to remove the influence of technical factors, such as cellular sequencing depth, from downstream analyses. Other uninteresting potential sources of variability from technical factors are the percentages of reads from mitochondrial and ribosomal protein genes, and these were explicitly regressed out during normalisation. A set of 'anchor' genes for integration were identified by Seurat, and dimensionality reduction by principal component analysis was applied to the integrated library containing all cells. The cells were then embedded with UMAP[207] for visualisation, and initial cell clusters were identified using Seurat's FindClusters with a resolution of 0.5, which was lowered from the default resolution of 0.8 to create broader cell classifications. The majority of cells formed a single large group on the UMAP projection, which was divided into smaller clusters by FindClusters (Figure 4.3A). Smaller clusters were separated from the main body of cells, suggesting sub-populations of cells with distinct gene expression patterns. Cells were visualised by treatment (Figure 4.3B) and donor (Figure 4.3C) to check for effects of these factors on clustering, and there was no clear separation of cells by donor or treatment. The total UMI count (Figure 4.3D), percentage of reads in mitochondrial genes (Figure 4.3E) and percentage of reads in ribosomal protein genes (Figure 4.3F) were also visualised for all cells, revealing some differences between clusters. While these differences may be due to technical influences, it is also possible that different cell types or subtypes contain different proportions of mitochondrial or ribosomal protein reads for biological reasons,

e.g. some cells may have elevated levels of ribosomal protein expression due to increased protein synthesis.

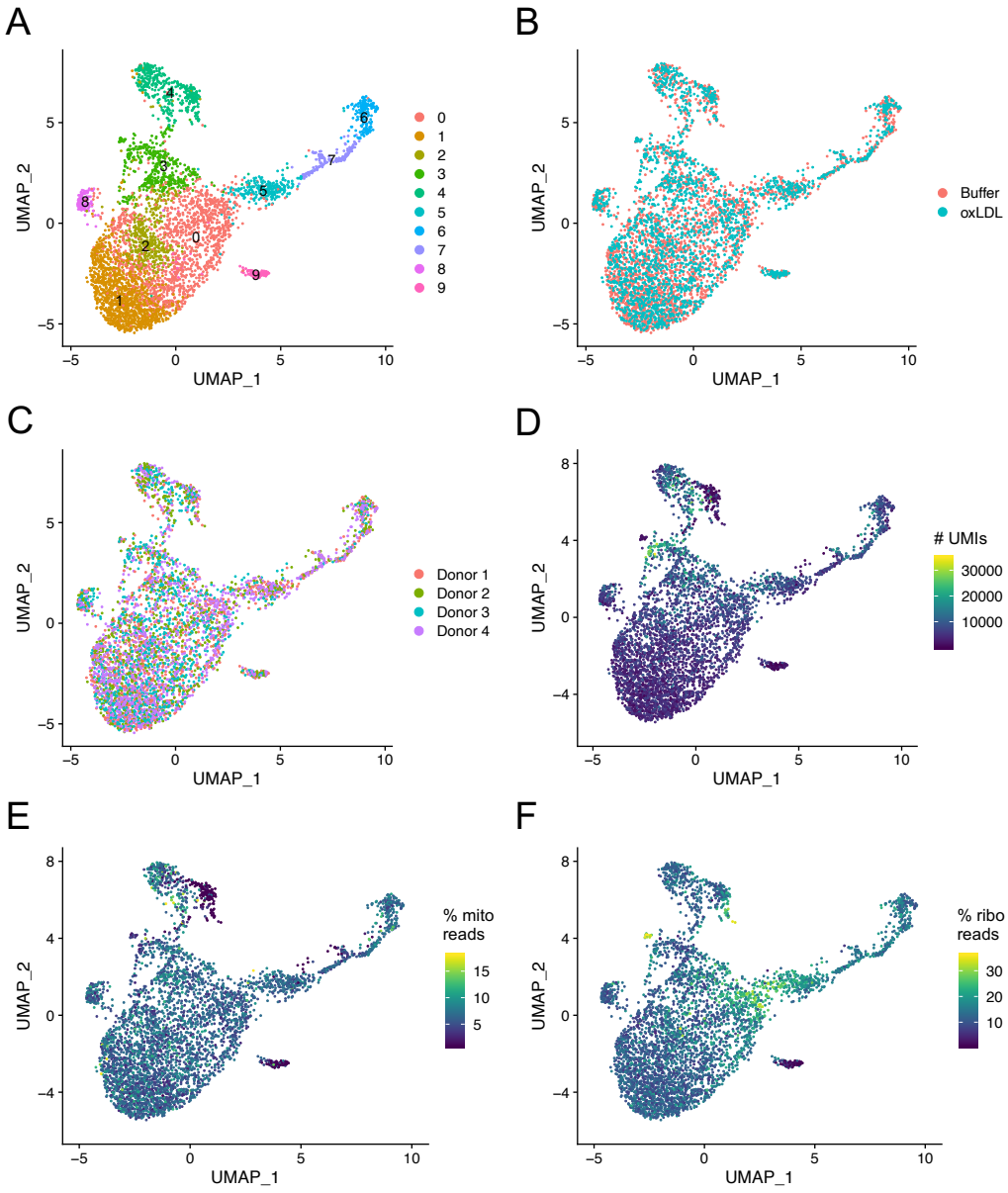

Figure 4.3: UMAP embeddings of all cells from all samples after integration in Seurat. Plots show cells coloured by (A) cluster, as determined by graph-based clustering with a resolution parameter of 0.5, (B) treatment, and (C) original donor. Cells in (D) were coloured by the number of UMIs detected, and in (E) and (F) coloured by percentage of reads aligning to mitochondrial and ribosomal protein genes, respectively.

4.5 Identification of cluster marker genes

Differential expression analysis was performed using Seurat to test for highly expressed genes within each cluster of cells, compared to all other cells. Such 'marker' genes were identified for all clusters and suggested the presence of functionally distinct subpopulations of macrophages (Figure 4.4A). Clusters 0 and 1 exhibited broadly similar patterns of gene expression but were separated by higher expression of genes encoding proteins with known roles in inflammation and atherosclerosis in cluster 0, namely *SPP1* and *CCL2* encoding Osteopontin and Monocyte Chemoattractant Protein 1 (MCP-1), respectively. In addition, cells in cluster 1 were marked by higher expression of *FUCA1* and *FTL*. Clusters 2 and 4 showed increased expression of genes encoding MHC class II molecules, as well as *CD74*, with cluster 4 additionally marked by *CD63* and *CYBA* expression, representing sub-populations of antigen-presenting and phagocytic macrophages. Markers for cluster 3 included *FABP5* and *CXCL16,* the latter of which has previously been identified as a scavenger receptor for oxLDL[208]. Clusters 5, 6 and 7, comprising the line of cells projecting from the main body of the UMAP visualisations, showed relatively high expression of a number of genes involved in the regulation of DNA replication and the cell cycle (Figure 4.4B). Cells in cluster 8 most closely resembled classically activated (M1) macrophages as described in the literature and expressed a number of genes encoding inflammatory cytokines at high levels relative to cells in other clusters; these cytokines included *CCL4*, *CXCL8*, *CXCL2*, as well as *TNF* and *IL1B*. The strongest marker genes for cluster 9 were *MALAT1* and *NEAT1*, both of which encode transcripts expressed in the nucleus (Figure 4.4C). Analysis of the full list of marker genes (adjusted p-value < 0.05, Bonferroni) for cluster 9 revealed enrichment of annotations from the Gene Ontology: Cellular Component (CC) database such as

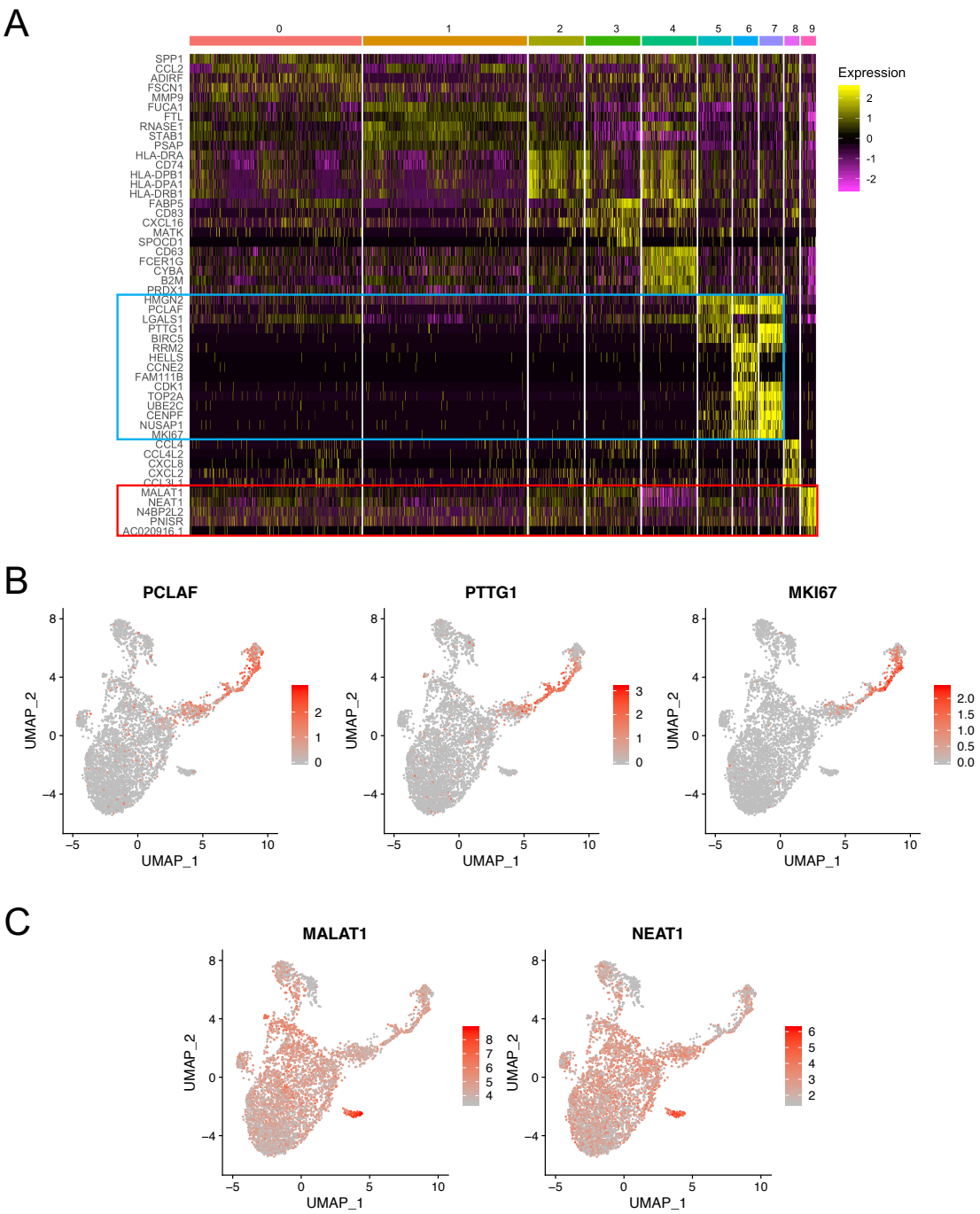

Figure 4.4: Marker genes for preliminary clusters. (A) Heatmap of scaled marker gene expression for each cluster, with rows corresponding to labelled genes and columns representing single cells. Cells are grouped according to their assigned cluster, and cluster labels are shown at the top of each group. The blue box marks genes associated with cell proliferation, DNA replication and the cell cycle. The red box marks genes highly expressed in the nucleus. (B-C) UMAPs coloured by log-normalised expression per cell for cell cycle-associated genes (B), and for nuclear transcripts (C).

'nuclear speck' ($p = 9.8 \times 10^{-4}$), 'nuclear lumen' ($p = 1 \times 10^{-3}$) and 'nuclear part' ($p = 2 \times 10^{-3}$). As such, cluster 9 was deemed to contain nuclei stripped of their cytoplasm rather than intact cells and was excluded from further analysis. In addition, cells were scored based on expression of marker genes for S or G2/M phases of the cell cycle using Seurat (Figure 4.5A), and cells with low scores for both S and G2M markers were classed as G1 phase, i.e. not actively cycling. For all clusters except 5, 6 and 7, the majority of cells were scored as G1 phase (Figure 4.5B). 60.3% of cells in cluster 6 were scored as S phase, with the remainder as G2/M, and 100% of the cells in cluster 7 were placed in G2M phase. Cells in these clusters therefore appeared to be actively proliferating. Although an interesting observation for differentiated macrophages, the increased expression of genes involved in cell proliferation was expected to impact the detected UMI counts for other genes. Therefore, rather than attempting to regress out all signal associated with cell cycle marker genes, the difference between S and G2M scores for all cells was regressed out of the model with the aim of removing signal associated with different stages of the cell cycle while retaining the signal separating cycling from non-cycling cells.

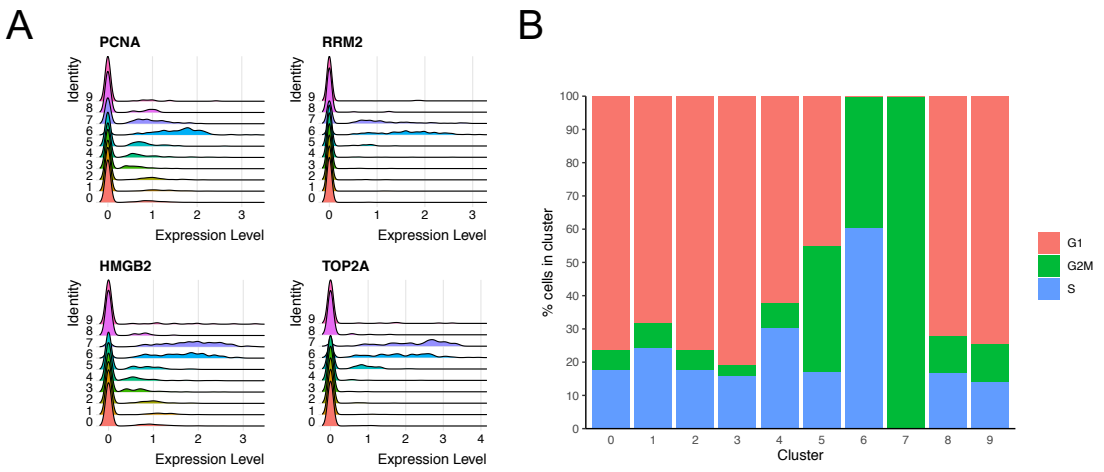

Figure 4.5: Macrophage expression of cell cycle genes. (A) Log-normalised expression of cell cycle marker genes within each cluster. (B) Cell cycle phase assignments for each cluster, determined with Seurat.

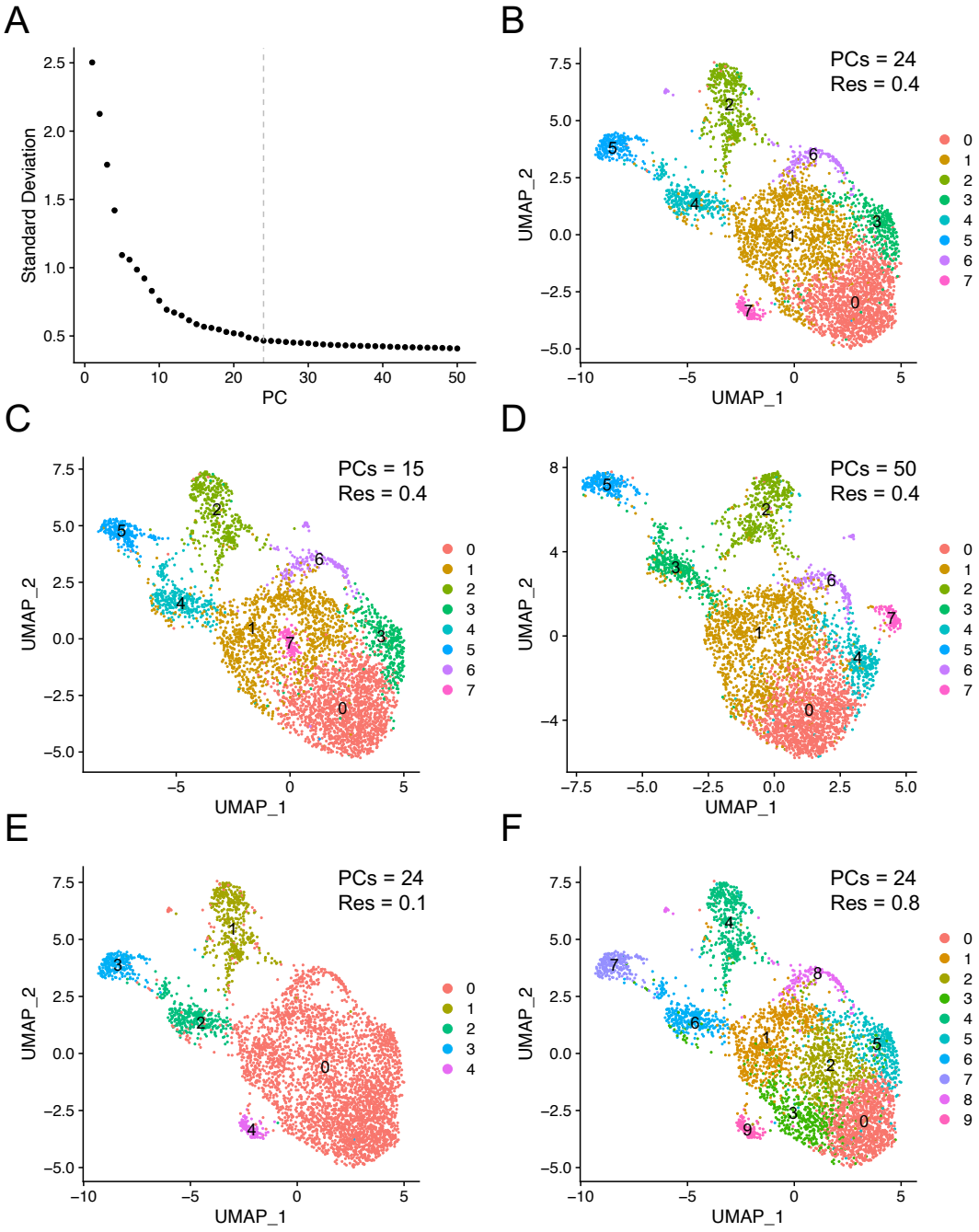

Figure 4.6: Optimisation of clustering parameters. (A) Scree plot showing the standard deviation of each principal component. The dashed grey line marks the 'elbow' at 24 principal components, i.e. the number of principal components selected for downstream analysis. (B) UMAP embedding of single cells, coloured by cluster assigned using 24 principal components and resolution = 0.4. (C-D) UMAP embeddings after adjusting the number of principal components, giving similar results. (E-F) Clustering with alternative resolution parameters.

4.6 Re-clustering and characterisation of macrophage subtypes

After removal of nuclei and calculation of cell cycle scores in Seurat, counts for the remaining cells were normalised using SCTransform, with the difference between S and G2M scores included as a variable to regress out of the model. Following normalisation cells underwent dimensionality reduction by PCA and UMAP embedding, as previously. For UMAP embedding and clustering, the number of principal components used was determined by plotting the standard deviation for each principal component and identifying the point at which additional principal components resulted in a negligible change to the standard deviation (Figure 4.6A). Using fewer principal components resulted in poorer separation of clusters when visualised by UMAP (Figure 4.6C), but similar results were obtained using a greater number of principal components (Figure 4.6D). Clustering with different values for the resolution parameter of Seurat's FindClusters either resulted in too few clusters with clear intra-cluster differences in gene expression (Figure 4.6E), or too many clusters with cells marked by very few unique differentially expressed genes (Figure 4.6F). As such, resolution = 0.4 was chosen and this achieved similar clustering of cells to that observed during preliminary analysis. All clusters contained cells from both buffer- and oxLDL-treated samples (Figure 4.7A), and the proportion of cells from each donor showed minor variation within clusters (Figure 4.7B). Interestingly, for all donors the proportion of cells in cluster 0 decreased between treatment groups, and the proportion of cells in cluster 6 increased, indicating cluster-specific responses to oxLDL.

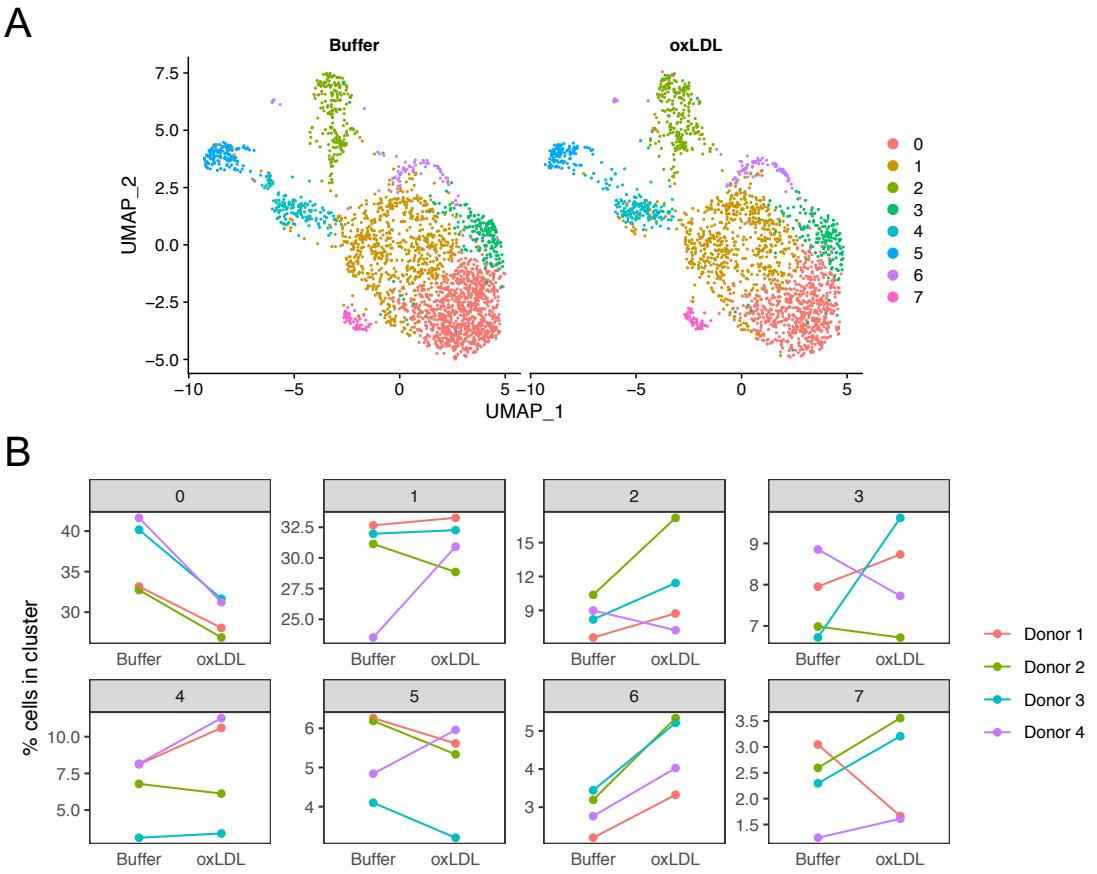

Figure 4.7: Subpopulations identified in buffer- and oxLDL-treated macrophages. (A) UMAPs of single macrophages treated with control buffer (left) and oxLDL (right), coloured by cluster. (B) Proportion of cells from each sample assigned to each cluster.

Marker genes were identified for all clusters using 'corrected' log-normalised UMI counts returned by normalisation with SCTransform, and results were filtered to remove genes with an adjusted p-value (Bonferroni) greater than 0.05, as well as genes detected in less than 50% of the cells in the cluster. Expression of canonical macrophage markers was detected in all clusters (Figure 4.8A), and for cluster 2 *CD14*, *CD68* and *FCGR3A* were more highly expressed relative to all other clusters (adjusted p-value < 0.05, Bonferroni). In addition, clusters were negative for expression of the dendritic cell markers *FCER1A* and *CLEC10A*. Markers identified as differentially expressed following re-clustering were consistent with those identified during initial clustering, with all major clusters retained and identifiable between analyses based on the top marker

genes. The primary difference following re-clustering was the re-distribution of the cells from the three initial clusters with high expression of proliferation markers into two clusters. Plotting the expression of the top marker for each cluster (determined by average log-fold change) revealed that some markers were not unique to specific clusters, but rather their selection was driven by relatively large differences in expression between larger clusters (Figure 4.8B). For example, differential expression analysis between clusters 0 and 1, which comprised the majority of cells in the dataset, identified *SPP1* as the most differentially expressed gene (Figure 4.8C). However, *SPP1* also showed relatively strong expression in several other clusters, indicating that *SPP1* expression could be considered a negative marker of clusters 0 and 3, rather than a true marker of cluster 1 (Figure 4.8D). For other clusters, high expression of the top marker was more useful in distinguishing clusters of cells and hinted at the biological functions of the represented subpopulations. Clusters 4 and 5 contained actively proliferating cells, with more than 50% of the cells in each cluster labelled as G2M phase based on the previously calculated cell cycle scores. The most differentially expressed genes between these clusters were related to DNA replication and the cell cycle (e.g. *HIST1H4C*, *RRM2*, *PCNA*, *TOP2A*), and these clusters were therefore considered to still be separated by cell cycle phase despite accounting for this during normalisation. Indeed, this was reflected in the proportions of cells labelled as in S or G1 phase in each cluster, with 31.3% in G1 and 11.2% in S for cluster 4, and 0% in G1 and 48.1% in S for cluster 5. Marker genes identified for cluster 3 included *CD74* and genes for MHC class II molecules (*HLA-DRA*, *HLA-DPB1*, *HLA-DPA1*, *HLA-DRB1*), indicating the presence of a subset of antigen-presenting macrophages colocalising with the major cluster of the UMAP embeddings. Cluster 7 marker genes included a number of inflammatory cytokines (*CCL4*, *CXCL2*,

CCL3, *CXCL3*, *CCL2*), as well as *ICAM1* and *NFKBIA*, identifying this cluster as containing activated, inflammatory macrophages.

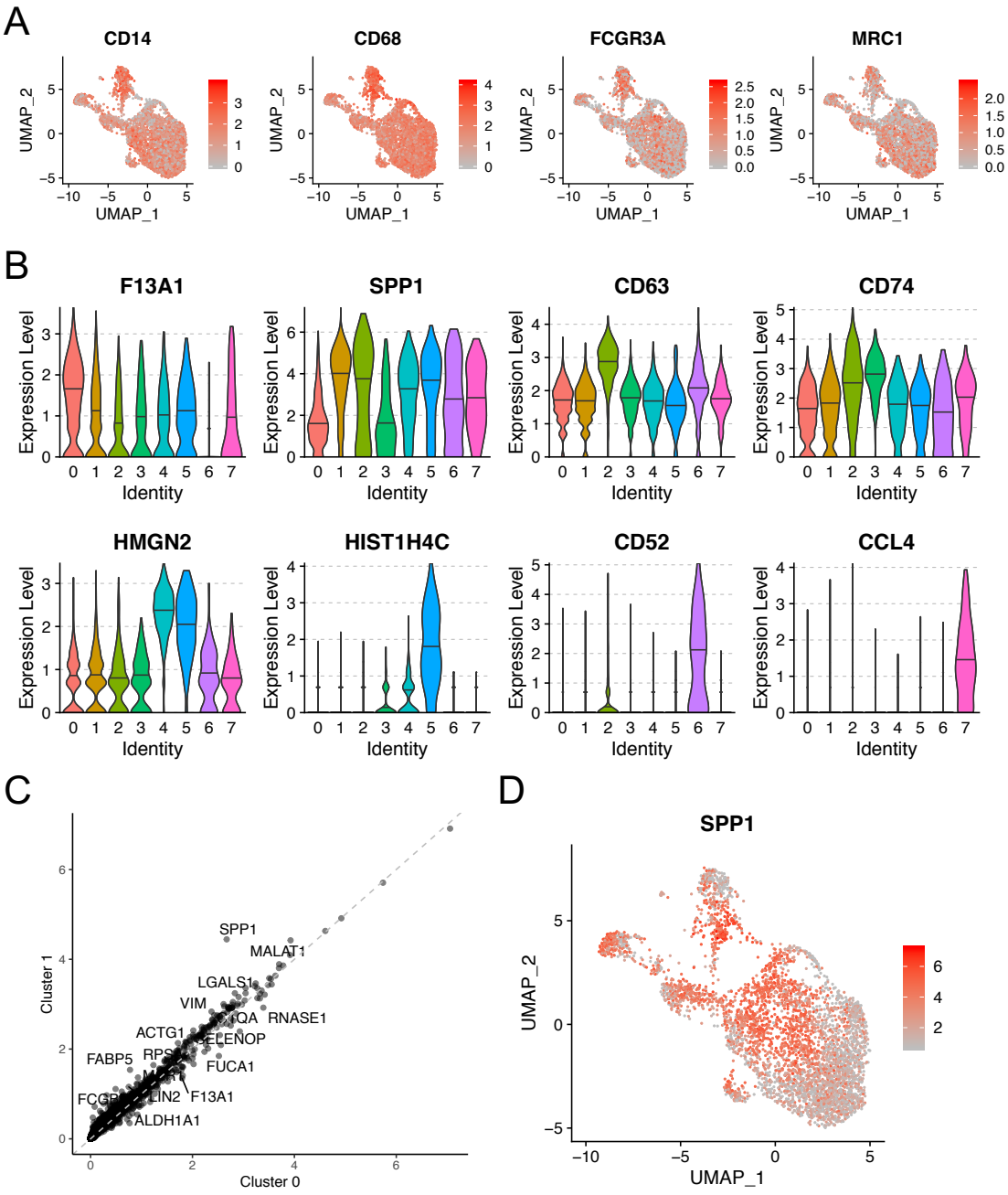

Figure 4.8: Macrophage subpopulations are marked by differentially expressed genes. (A) UMAPs coloured by log-normalised expression of canonical macrophage markers. (B) Violin plots showing log-normalised expression of 'top' markers (lowest adjusted p-value) for each cluster. Top row, left to right - cluster 0, cluster 1, cluster 2, cluster 3. Bottom row, left to right - cluster 4, cluster 5, cluster 6, cluster 7. (C) Scatter plot of log-normalised expression for genes expressed in clusters 0 and 1. Labelled genes are significant markers of either cluster. (D) UMAP embedding coloured by log-normalised expression of SPP1.

For cluster 2, 290 marker genes were identified, which was the largest number for any of the clusters. Among the cluster 2 markers with the greatest average fold change (i.e. greater than two-fold) were genes involved in antigen presentation (*HLA-DRB1*, *HLA-DPA1*, *HLA-DRA, CD74*, *HLA-C, B2M*), the complement cascade (*C1QC*, *CFD*), and phagocytosis (*FCER1G*, *CYBA*)(Figure 4.9A). Other genes with relatively high expression included *FABP5* and *PLTP*, which are involved in lipoprotein metabolism, and several genes involved in the maintenance of cellular redox homeostasis, i.e. *PRDX1*, *MGST3*, *GSTO1*, *TALDO1.* Interestingly, this cluster also showed relatively high expression of *TREM2* and *CD9* (Figure 4.9B), both of which have been implicated in atherosclerosis, and were recently identified as markers of a distinct population of macrophages found in aortic lesions of *Ldlr*$^{-/-}$ and *Apoe*$^{-/-}$ mice[209]. Marker genes for cluster 2 were tested for enrichment of gene ontology and pathway annotations, and the most strongly enriched terms related to antigen presentation, immune processes and the mitochondrial electron transport chain (Table 4.3). 75 genes were identified as markers of cluster 6, and the top markers included a number of genes involved in cholesterol metabolism, e.g. *CYP27A1*, *APOC1*, *APOE* and *NR1H3* (Figure 4.10A). Several of the strongest markers of cluster 6 were also expressed at relatively high levels in cluster 2, such as *TREM2*, *CD9*, *LGALS3* and *LIPA*, and cluster 6 markers showed enrichment of similar annotations from the GO:BP and KEGG databases, but showed no enrichment for terms related to metabolism (Table 4.4). The relatively high expression of genes involved in cholesterol metabolism, paired with the observation that the proportion of cells in cluster 6 was greater in oxLDL-treated samples for all donors, may mark this cluster as containing foam cells or an intermediate subset of lipid-laden macrophages.

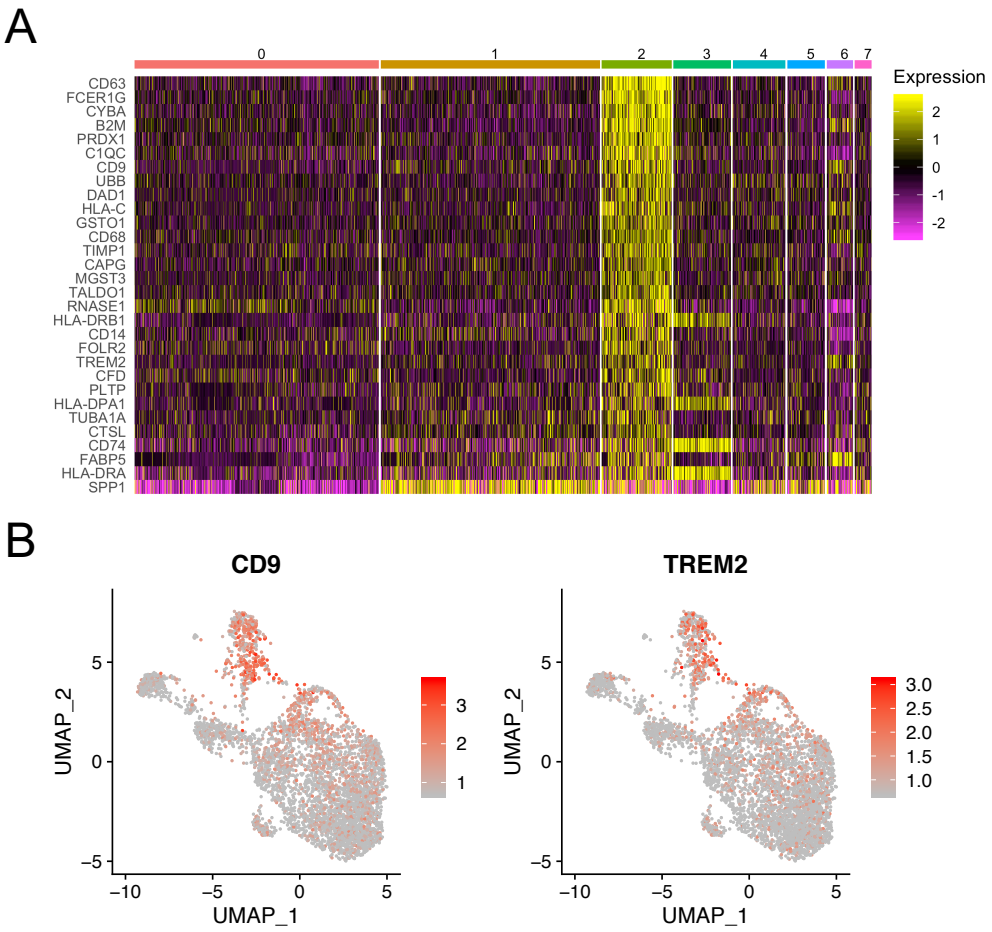

Figure 4.9: Cluster 2 marker genes. (A) Heatmap showing scaled expression of markers of cluster 2 (rows) in all cells (columns) grouped by cluster. (B) UMAPs coloured by log-normalised expression of cluster 2 markers CD9 and TREM2.

Table 4.3: Enriched terms from the GO: Biological Process and KEGG databases for cluster 2 marker genes.

GO:BP Term	Adjusted p-value	KEGG Pathway	Adjusted p-value
GO:0002478 antigen processing and presentation of exogenous peptide antigen	3.12E-32	hsa04612 Antigen processing and presentation	2.22E-19
GO:0002376 immune system process	1.99E-25	hsa05012 Parkinson's disease	2.34E-19
GO:0006091 generation of precursor metabolites and energy	6.61E-22	hsa00190 Oxidative phosphorylation	4.47E-18
GO:0022900 electron transport chain	1.32E-20	hsa04145 Phagosome	3.94E-17
GO:0006955 immune response	2.36E-20	hsa05016 Huntington's disease	1.94E-13
GO:0006950 response to stress	2.61E-20	hsa05010 Alzheimer's disease	2.01E-13
GO:0022904 respiratory electron transport chain	5.12E-20	hsa04142 Lysosome	8.72E-12
GO:0045333 cellular respiration	1.41E-18	hsa03050 Proteasome	5.47E-10
GO:0006952 defence response	5.16E-18	hsa05150 Staphylococcus aureus infection	6.48E-10
GO:0006915 apoptotic process	6.39E-17	hsa05323 Rheumatoid arthritis	4.61E-08

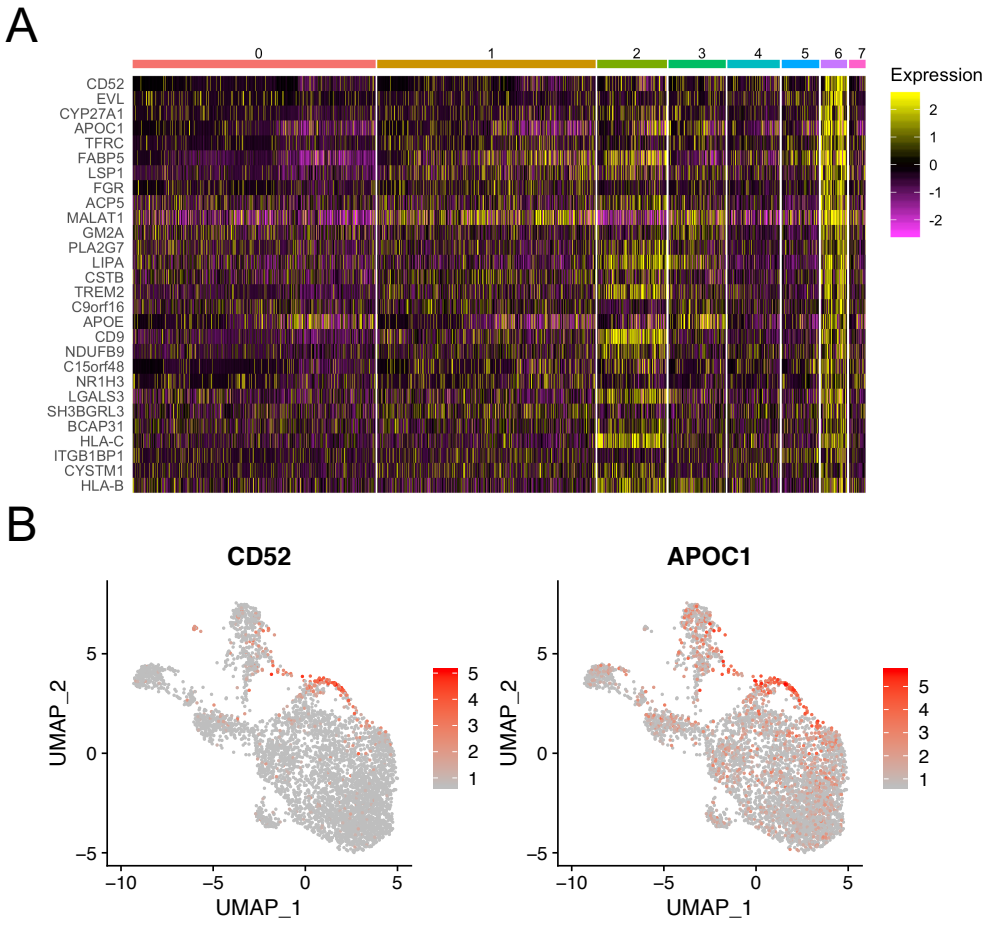

Figure 4.10: Cluster 6 marker genes. (A) Heatmap showing scaled expression of markers of cluster 6 (rows) in all cells (columns) grouped by cluster. (B) UMAPs coloured by log-normalised expression of cluster 6 markers CD52 and APOC1.

Table 4.4: Enriched terms from the GO: Biological Process and KEGG databases for cluster 6 marker genes.

GO:BP Term	**Adjusted p-value**	**KEGG Pathway**	**Adjusted p-value**
GO:0002376 immune system process	9.01E-09	hsa04142 Lysosome	7.58E-06
GO:0006950 response to stress	2.44E-07	hsa05416 Viral myocarditis	8.95E-04
GO:0006955 immune response	2.70E-07	hsa04612 Antigen processing and presentation	1.21E-03
GO:0009611 response to wounding	3.72E-07	hsa04145 Phagosome	2.40E-03
GO:0050896 response to stimulus	6.44E-07	hsa04650 Natural killer cell mediated cytotoxicity	2.63E-03
GO:0002474 antigen processing and presentation of peptide antigen via MHC class I	8.64E-07	hsa04514 Cell adhesion molecules (CAMs)	8.68E-03
GO:0050878 regulation of body fluid levels	9.52E-07	hsa03320 PPAR signaling pathway	1.69E-02
GO:0016192 vesicle-mediated transport	1.07E-06	hsa05330 Allograft rejection	2.27E-02
GO:0042060 wound healing	1.24E-06	hsa05332 Graft-versus-host disease	2.72E-02
GO:0007596 blood coagulation	1.26E-06	hsa04940 Type I diabetes mellitus	2.96E-02

4.7 Identification of genes regulated by oxLDL across all cells

To identify differentially expressed genes between buffer- and oxLDL-treated cells, a 'pseudo-bulk' approach was adopted wherein read counts across all cells from each donor, within treatment groups, were summed to give gene-wise counts for each sample. Summed counts were then analysed using the edgeR QL framework[106] with a paired-sample experimental design, analogous to a bulk RNA-seq analysis. This approach allows the explicit modelling of donor-specific effects on gene expression during differential analysis, which is not currently possible within Seurat. Results were filtered to retain genes with an adjusted p-value (Benjamini-Hochberg) less than 0.05, leaving 1,229 differentially expressed genes, of which 657 were upregulated (139 by > 2-fold) and 572 were downregulated (113 by > 2-fold) (Figure 4.11A). Expression changes were observed across all clusters for many of the most differentially expressed genes, but heterogeneity in gene expression between clusters was also present (Figure 4.11B). Among the upregulated genes with the largest fold changes were a number of genes with known roles in the macrophage response to oxLDL, such as *CD36* and *MSR1*, which encode scavenger receptors that recognise and trigger endocytosis of oxLDL, as well as *ABCA1* and *ABCG1*, which encode major transporters required for cholesterol efflux in foam cells. Also upregulated were *PLIN2, FABP4* and *FABP5*, which encode proteins involved in intracellular storage of lipids. Other genes significantly upregulated by oxLDL were *APOE* and *APOC1*, which encode apolipoproteins involved in the transport and exchange of cholesterol between tissues, and genetic variants at these loci have been associated with coronary artery disease. *F13A1*, encoding the fibrin stabilising Factor XIII of the blood coagulation cascade, was among the genes most downregulated by oxLDL. This gene is considered a marker of M2 polarisation in the literature, and in this study was the top marker of cluster 0. Other putative markers of M2 macrophages,

SELENOP and *MS4A6A*, were also downregulated by oxLDL on average across all cells. Upregulated genes were enriched for gene ontologies from the GO: Biological Process and GO: Molecular Function databases relating to cholesterol storage, lipid transport, leukocyte migration, response to hypoxia, carbohydrate metabolism, and oxidoreductase activity, among others (Figure 4.11C). For downregulated genes, several of the terms with the strongest enrichment related to the innate immune response, antigen presentation via MHC class II, cholesterol biosynthesis, the electron transport chain, and generation of superoxide (Figure 4.11D). The overlap between differentially expressed gene sets and annotated gene sets further identified genes regulated by oxLDL with the potential to alter macrophage function under atherogenic conditions (Figure 4.11E and F). For example, oxLDL upregulated *CCL2* which encodes Monocyte Chemotactic Protein 1 (MCP-1) and has been demonstrated to play a significant role in the development of atherosclerotic lesions. Also upregulated was *PKM*, encoding the M1 and M2 pyruvate kinase isozymes, which catalyse the conversion of phosphoenolpyruvate to pyruvate during the final step of glycolysis. Enrichment analysis also highlighted the presence of differentially expressed transcription factor encoding genes, such as *MAF*, which has recently been implicated in regulating the enhancer landscape of macrophages during inflammatory activation. Several genes involved in NADPH-dependent superoxide production were downregulated by oxLDL, including *CYBA, CYBB*, *NCF1* and *NCF4*, suggesting that oxLDL may disrupt this process in macrophages during atherogenesis.

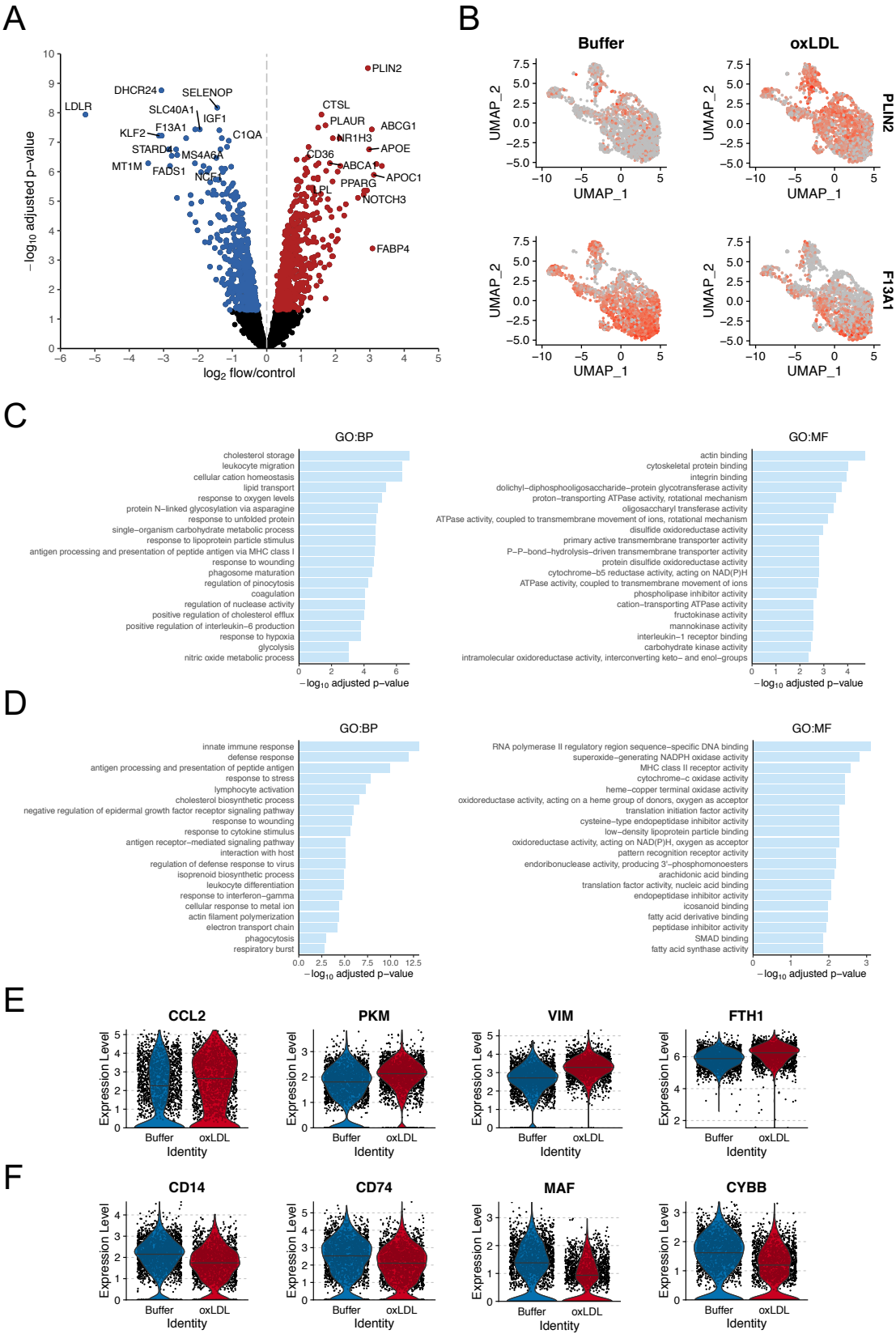

Figure 4.11: 'Pseudo-bulk' differential gene expression analysis. (A) Volcano plot of all genes detected by scRNA-seq. Red and blue points are genes upregulated and downregulated by oxLDL, respectively, with an adjusted p-value < 0.05. (B) UMAPs coloured by log-normalised

expression of genes identified as differentially expressed between all buffer- and all oxLDL-treated cells. (C-D) Enrichment of terms from the GO: Biological Process (left) and GO: Molecular Function (right) databases for (C) genes upregulated by oxLDL and (D) genes downregulated by oxLDL. (E-F) Examples of genes (E) upregulated and (F) downregulated by oxLDL with roles in biological functions relevant to macrophage immune function and atherosclerosis.

4.8 Cluster-specific differential gene expression analysis

To explore how the response to oxLDL differed between the identified macrophage subpopulations, cells within each cluster were tested for differential expression between buffer- and oxLDL-treated samples using summed counts in edgeR. The number of differentially expressed genes (adjusted p-value < 0.05, Benjamini-Hochberg) identified for each cluster was proportional to the number of cells within the cluster, due to the lower summed read counts from smaller clusters decreasing the confidence of expression estimates in edgeR (Figure 4.12A). All clusters showed a similar change in gene expression for the strongest markers of oxLDL treatment, as identified during the 'pseudo-bulk' analysis using all cells (Figure 4.12B). Differentially expressed genes in cluster 0 and cluster 1, the largest clusters containing 1,529 and 1,373 cells respectively, and together comprising 64.0% of all cells in the dataset, were compared to identify genes common to both clusters that were regulated by oxLDL, as well as uniquely regulated genes in each cluster (Figure 4.12C). The overlap between differentially expressed gene sets contained 478 genes, including many of the most differentially expressed genes between all buffer- and oxLDL-treated cells, such as *PLIN2, FABP4, FABP5, APOC1, APOE, ABCA1,* and *CD36*. These genes were also differentially expressed in all other clusters, representing a conserved response to oxLDL shared by all macrophage subpopulations. After ranking the differentially expressed genes unique to either cluster 0 or cluster 1 by adjusted p-value, the top gene for cluster 0 was *CCL2*. Visualisation of

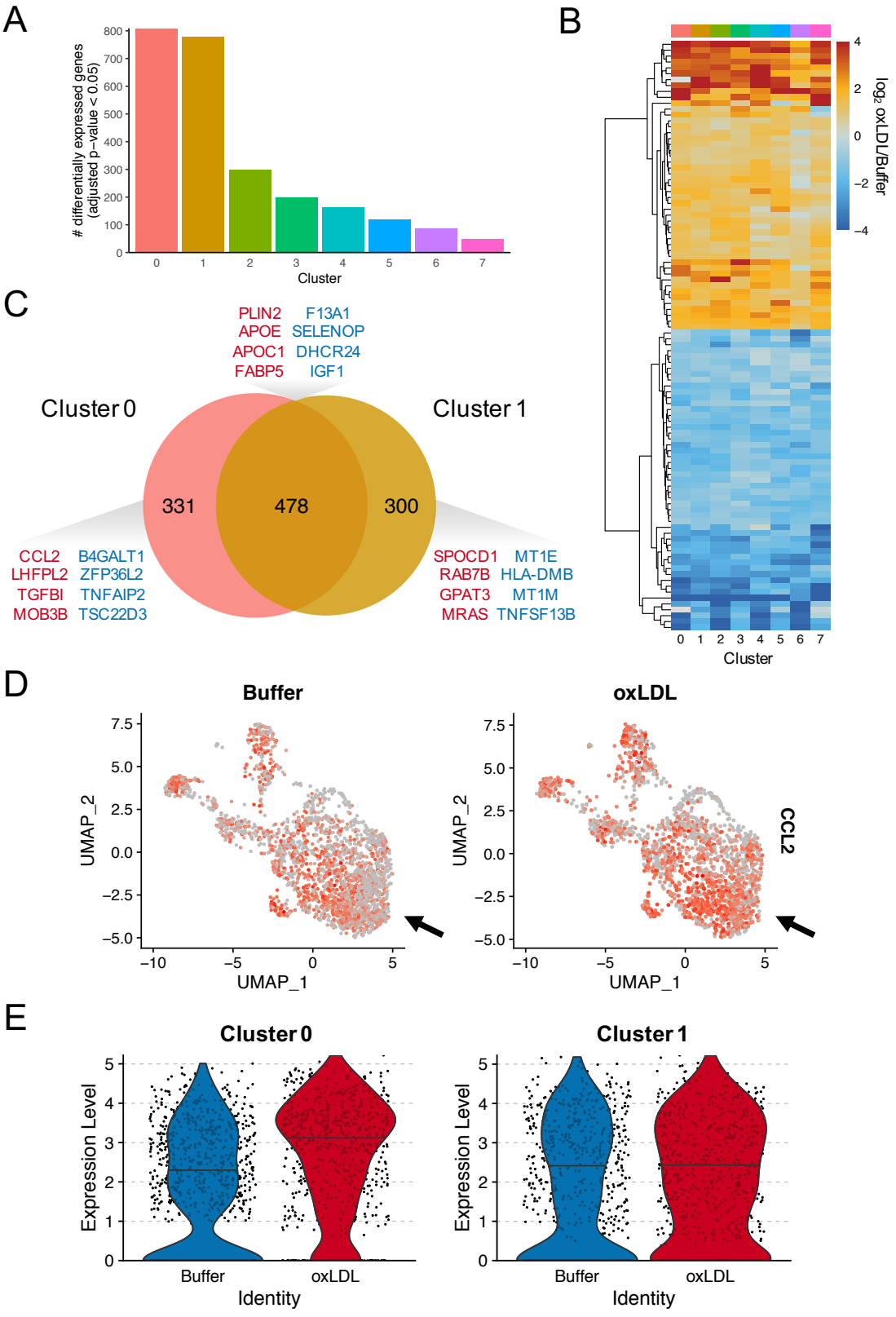

Figure 4.12: Identification of cluster-specific responses to oxLDL. (A) Bar plot showing the number of genes differentially expressed (adjusted p-value < 0.05, Benjamini-Hochberg) between buffer- and oxLDL-treated cells in each cluster. (B) Heatmap of per-cluster $\log_2$ fold change in expression for the top 100 (ranked by adjusted p-value) genes identified as

differentially expressed across all cells. Fold changes were calculated using edgeR. (C) Overlap of differentially expressed genes in clusters 0 and 1. (D-E) Visualisation of log-normalised CCL2 expression in (E) all cells, and (F) in cluster 0 (left) and cluster 1 (right) cells only.

the log-normalised expression per cell for *CCL2* showed a clear increase in expression for oxLDL-treated samples relative to cells in buffer-treated samples, which was not observed for cells in cluster 1 (Figure 4.12D and E). *CCL2* was also significantly upregulated by oxLDL in clusters 2, 3, 4 and 5, but the largest fold change in expression was observed in cluster 0. Given the proportion of cells contained in clusters 0 and 1 for each sample, this may represent a significant example of functional heterogeneity in the macrophage response to oxLDL, with implications for the development of atherosclerosis *in vivo*.

4.9 Electron microscopy reveals ultrastructural heterogeneity in oxLDL-treated macrophages

Based on the clear transcriptomic signature of lipid and cholesterol storage identified in oxLDL-treated macrophages, as well as previous descriptions of lipid-laden macrophages in atherosclerosis, transmission electron micrographs were obtained from buffer- and oxLDL-treated monocyte-derived macrophages from four donors to visualise ultrastructural changes occurring between the conditions. Cell monolayers were fixed and stained with osmium tetroxide, which binds preferentially to phospholipids and creates contrast during imaging. Samples were subsequently embedded in resin, sectioned to obtain ultra-thin 90 nm sections, and imaged using transmission electron microscopy. After fixation and sample preparation, mitochondria and membranes were well preserved, allowing for high resolution imaging of organelles and other cell contents. Cells in both treatment groups contained heterogeneous populations of lipid droplets with varying degrees of staining by osmium tetroxide. Buffer-treated macrophages contained

lipid droplets distributed throughout the cytoplasm (Figure 4.13A and C), with some cells containing very darkly stained droplets (Figure 4.13B) when compared with other buffer-treated macrophages (Figure 4.13D). Other compartments, exhibiting little to no internal staining with osmium tetroxide, were also present in the cytoplasm of both buffer- and oxLDL-treated macrophages. No clear differences in overall cell shape, mitochondrial shape and distribution, or lipid droplet composition were apparent from electron micrographs of oxLDL-treated cells, however distortion of the nuclear membrane was observed for many of the oxLDL-treated cells (Figure 4.14A-D) compared with the typically more rounded nuclei of buffer-treated cells. Approximately 50% of the oxLDL treated cells exhibited massive proliferation of multilamellar lipoid lysosomes (Figure 4.14E and F and Figure 4.15), identified as such based on similarity to structures observed in a previous study of murine peritoneal macrophages exposed to oxLDL[210], and these lipoid compartments were absent in buffer-treated macrophages. Given that the samples imaged during this study were treated for 48 hours, it remains unclear whether the observed heterogeneity is due to functional differences between macrophage subtypes or is a consequence of temporally separated cells representing different stages of the same process. In addition, the ultra-thin sections imaged provide limited information on the spatial organisation of lipids throughout entire cells. Therefore, further work is required to determine the significance of the observed lipoid structures and heterogeneity in the macrophage response to oxLDL.

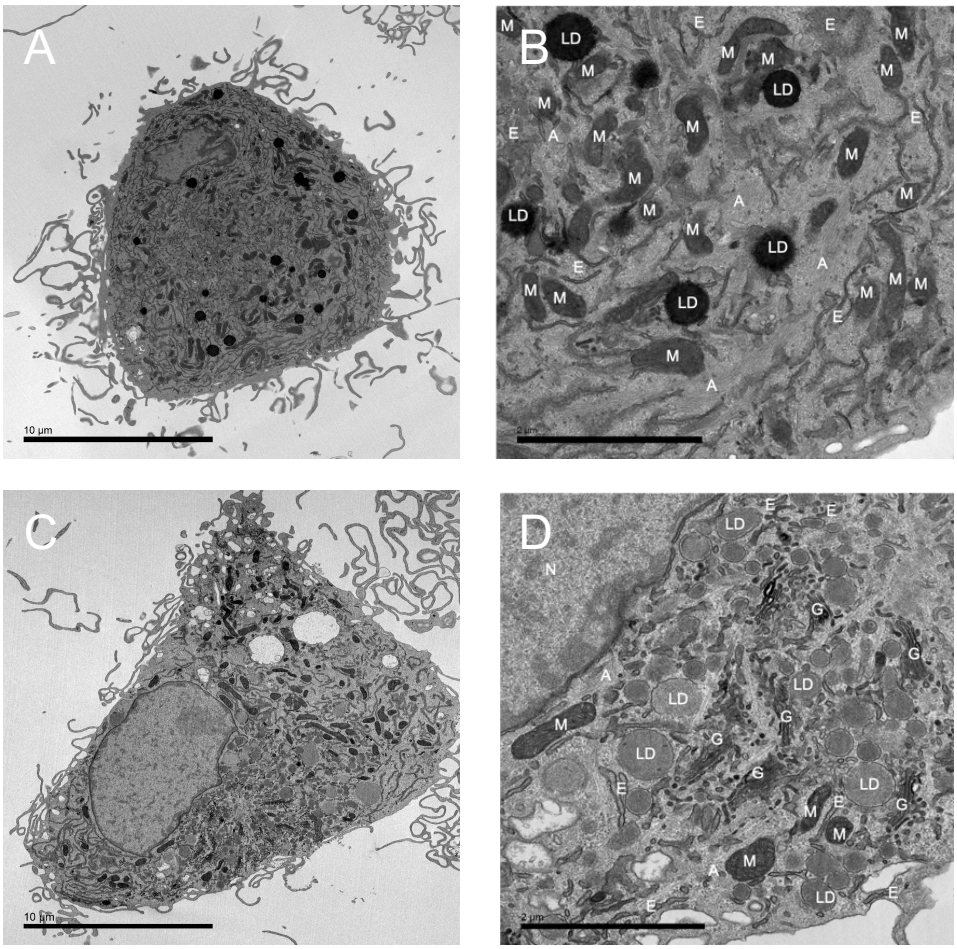

Figure 4.13: TEM of buffer-treated macrophages. (A-D) Transmission electron micrographs of primary human monocyte-derived macrophages after 48 hours of treatment with control buffer. Scale bars in (A) and (C) = 10 µm, and scale bars in (B) and (D) = 2 µm. LD – lipid droplet, M – mitochondria, E – endoplasmic reticulum, A – actin bundle, G – Golgi body, N – nucleus.

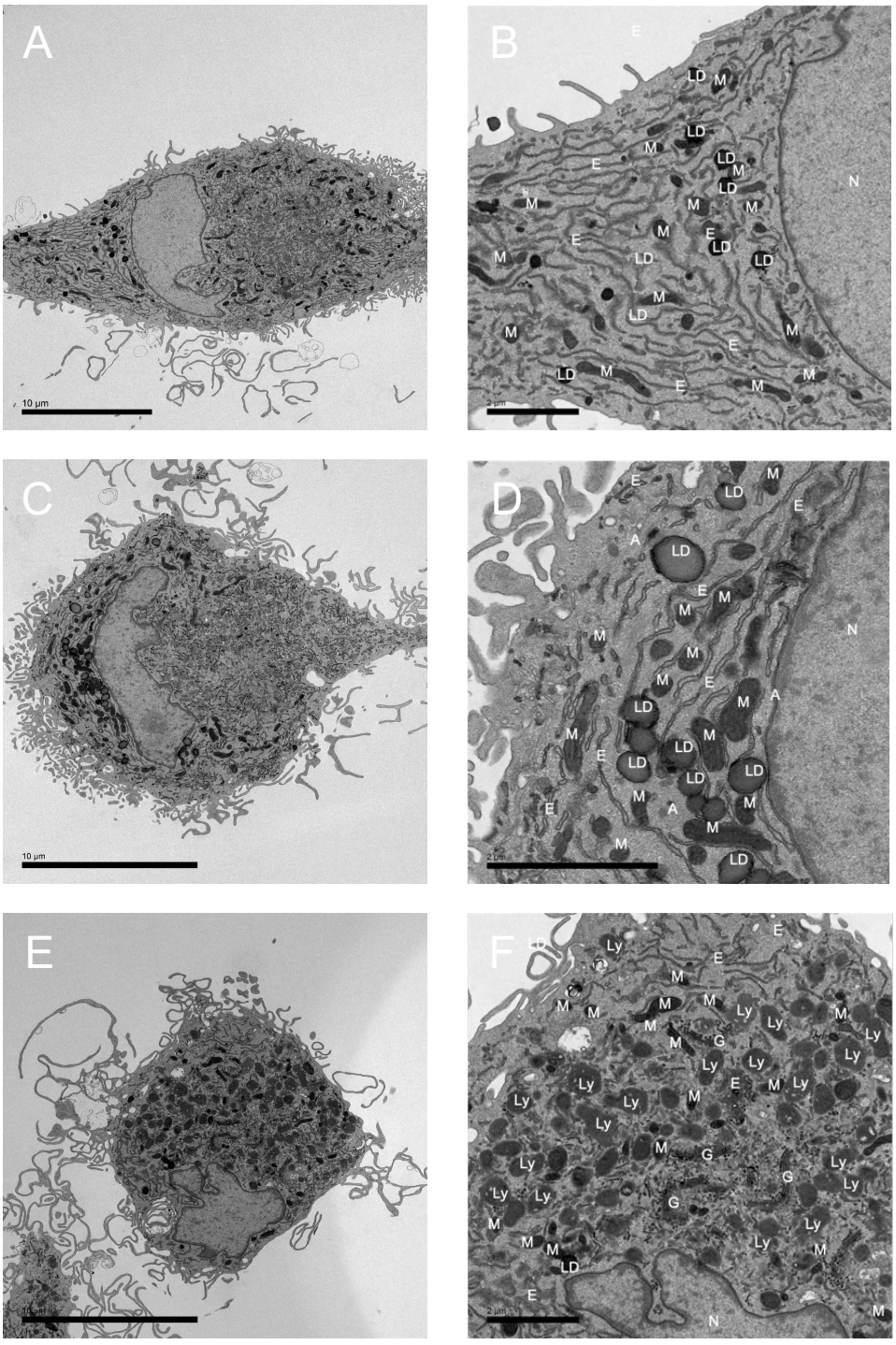

Figure 4.14: TEM of oxLDL-treated macrophages. (A-F) Transmission electron micrographs of primary human monocyte-derived macrophages after 48 hours of treatment with 50 µg/ml oxLDL. Scale bars in (A), (C) and (E) = 10 µm, and scale bars in (B), (D) and (F) = 2 µm. LD – lipid droplet, M – mitochondria, ER – endoplasmic reticulum, A – actin bundle, G – Golgi body, N – nucleus, Ly - lysosome.

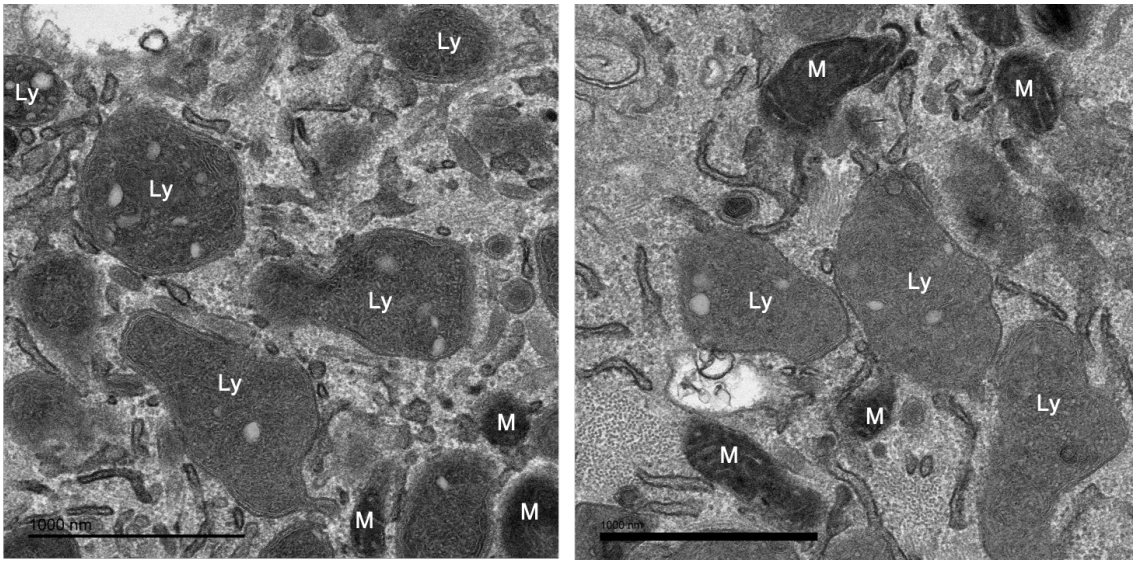

Figure 4.15: Intracellular lipid accumulation in oxLDL-treated macrophages. Multilamellar lipoid lysosomes were observed in primary human monocyte-derived macrophages after 48 hours of treatment with 50 µg/ml oxLDL. Scale bars = 1000 nm. M – mitochondria, Ly – lysosome.

4.10 OxLDL induces a shift from mitochondrial respiration to glycolysis in macrophages

To investigate the metabolic consequences of oxLDL for human macrophages, respiratory rates of buffer and oxLDL-treated cells were measured using an Agilent Seahorse XF Analyzer. This instrument indirectly measures mitochondrial respiration, as oxygen consumption rate (OCR), and glycolysis, as extracellular acidification rate (ECAR). Cells from 4 donors were treated sequentially with the ATP synthase inhibitor oligomycin, the uncoupler FCCP, and finally a combination of rotenone and antimycin A, which inhibit complexes I and III of the mitochondrial electron transport chain, respectively (Figure 4.16A). Each compound was injected after three measurements, with the final three measurements after injection of rotenone and antimycin A, and resulting changes in OCR over time were compared between buffer- and oxLDL-treated cells (Figure 4.16B). The profile of OCR change in response to the inhibitors was as expected, with a decrease following oligomycin injection, a pronounced increase after addition of

FCCP, and finally almost complete abolition of mitochondrial respiration with inhibition of the proton translocating complexes I and III by rotenone and antimycin A. OCR was higher for buffer-treated cells at all time points, indicating a higher basal rate of mitochondrial respiration compared with oxLDL-treated cells. Measurement of ECAR during the assay revealed a converse relationship, with a higher apparent rate of glycolysis in oxLDL-treated cells at all time points (Figure 4.16C). Mitochondrially-derived CO_2 can partially hydrate in the extracellular medium and contribute to extracellular acidification, making ECAR an inaccurate measure of glycolytic rate. Therefore, by inhibiting mitochondrial respiration with rotenone and antimycin A, the proton efflux rate (PER) attributable to lactate derived from glycolysis was calculated for buffer- and oxLDL-treated cells, which demonstrated a substantial increase in the glycolytic rate of macrophages exposed to oxLDL (Figure 4.16D).

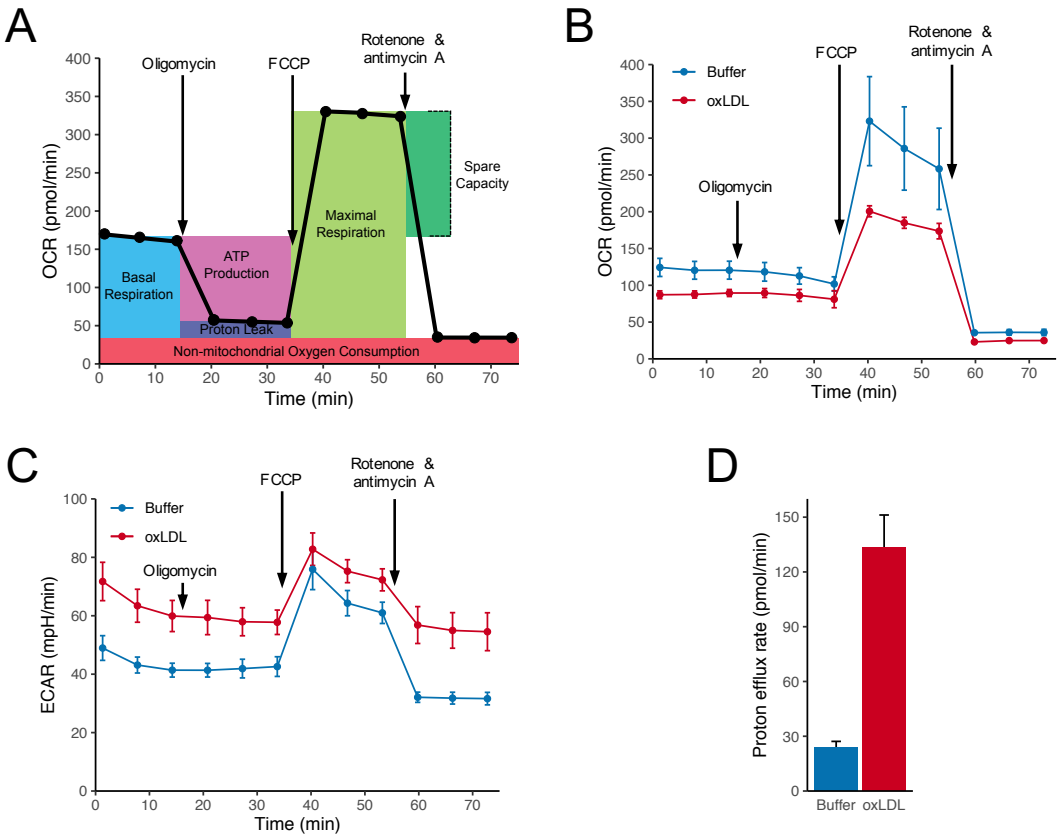

Figure 4.16: Metabolic reprogramming of oxLDL-treated macrophages. (A) Overview of the mito stress test assay. (B) Oxygen consumption rate (OCR) of buffer- and oxLDL-treated cells measured during a mito stress test assay on an Agilent Seahorse XF Analyzer. (C) Extracellular acidification rate (ECAR) measurements obtained during the same assay. (D) Glycolysis-specific proton efflux rate of buffer- and oxLDL-treated macrophages, measured using Agilent's glycolytic rate assay. Values are the mean of 5 technical replicates from 4 independent donors, and error bars show the standard deviation.

4.11 Reduced glutathione is depleted in oxLDL-treated macrophages

Glutathione is a major intracellular antioxidant and has been implicated in the development of atherosclerosis. The concentration of the reduced form of glutathione in cell lysates prepared from buffer- and oxLDL-treated macrophages was measured using a fluorometric assay, and a significant decrease in reduced glutathione concentration was observed with oxLDL treatment (Figure 4.17A). N-acetylcysteine (NAC) is a precursor of reduced glutathione synthesis *in vivo*, and cell culture treatment with NAC has been demonstrated to reduce oxidative stress from reactive oxygen species (ROS)[211,212].

Buffer- and oxLDL-treated macrophages were either pre-treated with NAC or left untreated, and then stimulated with pyocyanin, a potent inducer of ROS and superoxide in particular. The levels of superoxide were then measured using a fluorometric dye specific for superoxide. In buffer-treated macrophages, the level of superoxide was significantly increased by pyocyanin stimulation, but majorly reduced relative to untreated cells when macrophages were pre-treated with NAC (Figure 4.17B). This was presumably due to the enhanced levels of glutathione induced by NAC pre-treatment. The relative increase in superoxide levels upon stimulation with pyocyanin was comparable in buffer- and oxLDL-treated macrophages, but for oxLDL-treated macrophages the superoxide levels in cells pre-treated with NAC remained at the level of untreated cells. This indicated that while NAC pre-treatment still prevented pyocyanin-induced increase in superoxide in oxLDL-treated macrophages, it was unable to reduce the level relative to untreated cells, possibly due to the observed depletion of protective intracellular reserves of reduced glutathione by oxLDL.

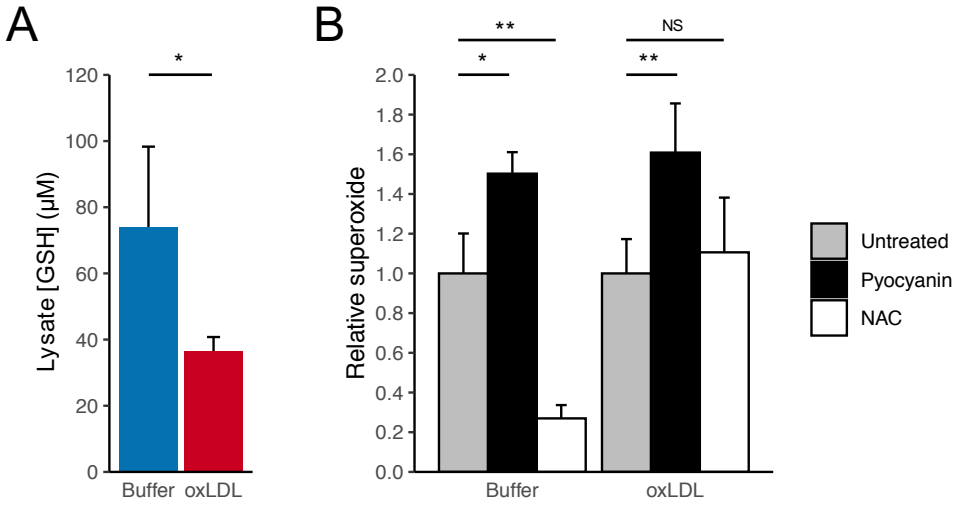

Figure 4.17: OxLDL depletes macrophage GSH. (A) Concentration of reduced glutathione (GSH) in 100 µl lysate prepared using 1 x 106 buffer- or oxLDL-treated cells. (B) Superoxide levels in buffer- and oxLDL-treated macrophages after 30 minutes of treatment with pyocyanin (200 µM), or 30 minutes of pre-treatment with N-acetyl-L-cysteine (5 mM) followed by 30 minutes of

pyocyanin treatment (200 μM), relative to unstimulated cells. Values shown are the mean of 4 independent donors, and error bars show the standard deviation. * $p < 0.05$, ** $p < 0.005$, NS - not significant.

4.12 Discussion

Macrophages display versatility and plasticity *in vivo*, which enables them to respond appropriately to a broad range of stimuli and adapt to different microenvironments[171]. Historically, researchers have applied a binary classification of observed macrophage phenotypes, designated M1 and M2, representing pro- and anti-inflammatory macrophage functions, respectively[196]. The present view is that description of macrophage 'polarisation' is an oversimplification, and in reality, macrophage populations contain cells with intermediate phenotypes readily adjusting to meet challenges from pathogens, damage associated cellular patterns, or immunometabolic modulators[195]. Specific macrophage subsets have been described as contributing to the development of atherosclerosis, based largely on murine disease models, but these subsets have not been observed in human lesions[200]. As such, studies of human macrophages in the context of atherosclerosis, and particularly the role of phenotypic and functional heterogeneity in disease progression, hold value in the quest to bridge the gap between observations from animal models and human disease-relevant biology, with the ultimate aim of translating these findings into therapies to mitigate the burden of atherosclerosis and resultant cardiovascular disease.

Transcriptomic and genomic methods have seen enormous development over the last two decades, and a recent boon has been the scalable application of single-cell RNA-sequencing to study the role of cellular heterogeneity in development and disease. A recent study sequenced single cells isolated from murine aortic lesions and identified transcriptomically segregated clusters of macrophages[209]. Separation of macrophages

based on gene expression was also achieved in this study, and several of the markers identified in human monocyte-derived macrophages *in vitro* were shared with the reported murine macrophage subpopulations. For example, expression of *F13a1, Txnip* and *Sepp1* marked the 'resident-like' macrophage cluster identified in murine lesions, and the human homologs *F13A1, TXNIP* and *SELENOP* marked cluster 0 in this study. Additionally, the researchers identified a smaller cluster of macrophages marked by relatively high expression of *Trem2* and *Cd9*, which was also present in the data obtained from human macrophages (cluster 2). Therefore, studies such as this one support the presence of disease-relevant macrophage subtypes in human-derived cell populations.

In addition to the populations identified in mouse models of atherosclerosis, a subpopulation of actively proliferating cells was detected in the human cells. Questions over the primary source of lesional macrophages have persisted in the literature, and while it seems likely that the consensus of monocyte recruitment and subsequent differentiation to macrophages is the driver of fatty streak development, there may exist a self-renewing population of macrophages within growing lesions[213–215]. In support of this hypothesis, proliferating macrophages were observed in both buffer- and oxLDL-treated samples, suggesting this may occur even in an atherogenic environment.

Other macrophage subtypes identified by scRNA-seq in this study are also of potential importance to atherogenesis. Cluster 1 was marked by relatively high expression of *SPP1*, encoding the integrin-binding glycoprotein Osteopontin, which has been associated with inflammation and atherosclerotic disease severity[216]. Osteopontin is expressed in atherosclerotic lesions, closely associated with macrophages and foam cells[217], and ablation of Osteopontin expression in *ApoE*$^{-/-}$ mice reduces macrophage infiltration and plaque size[218,219]. A dual role has been proposed for Osteopontin in atherogenesis, both

as a pro-inflammatory cytokine and an inhibitor of vascular calcification, marking it as a promising therapeutic target. Validation and further characterisation of the heterogeneity of macrophage Osteopontin expression *in vivo* could facilitate the development of therapeutic strategies targeting this gene and its product.

A small population of macrophages in this study were marked by increased expression of genes encoding MHC Class II molecules, as well as *CD74*, suggesting a role for these cells in the presentation of exogenous antigens. An adaptive immune response to atherogenic lipid has been reported, with both oxLDL-specific T lymphocytes and autoantibodies isolated from human plaque[220,221], and different lymphocyte subsets have been described as proatherogenic (e.g. T_H1 cells) or atheroprotective (e.g. T_{REG} cells)[222]. Relatively little is known about the contribution of macrophages to antigen presentation via MHC Class II in atherosclerosis, due in part to functional overlap with antigen-presenting dendritic cell populations found within developing lesions[223,224]. The results of this study support the existence of antigen-presenting macrophages in the presence of atherogenic oxLDL, and further characterisation of this population could improve our understanding of the origins of the adaptive immune response to atherogenic lipid.

The differentially expressed genes identified between buffer- and oxLDL-treated macrophages, from multiple donors, corresponded to previous observations from *in vitro* and *in vivo* studies. The response was dominated by genes involved in cholesterol storage and efflux, as well as scavenger receptors recognizing oxLDL, which implies a positive feedback loop that may contribute to a failure to adequately remove cholesterol from developing lesions[50]. Genes encoding inflammatory mediators were also among the differentially expressed genes, lending support to the characterisation of lipid-laden macrophages, known as foam cells, as activated, inflammatory cells[183]. Expression of

CCL2 in particular is strongly associated with atherogenesis, and the observation of differential regulation between macrophage subpopulations presents an intriguing avenue for further research[225]. Further advances in the generation and analysis of single-cell sequencing datasets are inevitable and likely imminent, and the current limits on gene detection within a single cell may eventually give way to deep transcriptomic profiling, allowing full characterisation of genetic responses at the level of an individual cell. Further to these advances, with the recent development of single-cell epigenomics, e.g. scATAC-seq, the epigenomic mechanisms underlying the observed heterogeneity may be determined, providing deeper insights into the perturbation of genetic mechanisms between individual cells.

While transcriptomic approaches supply a wealth of data and potential targets, additional methods are required to interpret the overlying biology of the observed changes in gene expression. A number of genes with roles in lipid storage were upregulated by oxLDL, including *PLIN2*, *FABP4* and *FABP5*. Electron microscopy was used to confirm the presence of intracellular lipid, and to visualise other ultrastructural effects of oxLDL treatment on macrophages. This resulted in high resolution electron micrographs of primary human macrophages, and the identification of extensive lipoid lysosomal structures occupying the cytoplasm of oxLDL-treated cells, although the spatiotemporal organisation of these structures remains an open question. Further electron microscopy approaches with the potential to address this include 3D-scanning electron microscopy, with the compilation of serial sections providing high resolution ultrastructural maps of entire cells. These will assist in the characterisation of possible nuclear deformity in oxLDL-treated cells, and the spatial distribution of lipoid lysosomes in lipid-laden macrophages. Further TEM studies of macrophages with varying durations of exposure

to oxLDL would also help to delineate the observed heterogeneity of the cells imaged during this study.

Other signatures of oxLDL-treated macrophages identified by scRNA-seq were expression changes of genes involved in metabolism and redox homeostasis, both of which represent areas of interest in the development of therapies for atherosclerosis. A number of genes encoding glycolytic enzymes were upregulated by oxLDL, such as *HK2*, *HK3* and *PKM*. Each of these genes encodes an enzyme catalysing one of the unidirectional regulatory steps in the glycolytic pathway, and their induction by oxLDL may drive glycolytic flux and metabolic reprogramming, which are associated with inflammatory activation of macrophages[141,226]. A potential mechanism for this switch was revealed by the reduction in cellular GSH concentration, which indicates increased oxidative stress in oxLDL-treated macrophages. Oxidative stress has been reported to induce expression of HIF-1α, a transcription factor controlling the switch to glycolytic metabolism during the cellular response to hypoxia[227,228]. HIF-1α has previously been implicated in atherosclerosis, but its involvement has been attributed to a hypoxic environment within an atherosclerotic lesion[229]. In this study, *HIF1A* was upregulated in a monolayer of oxLDL-treated macrophages, demonstrating that oxLDL activates HIF-1α in normoxic conditions. Further study to define the regulation of *HIF1A* by oxidative stress may provide opportunities for the development of therapies for atherosclerotic inflammation. In contrast to glycolysis, mitochondrial respiration has been associated with the 'M2', anti-inflammatory phenotype described in the literature[230], and oxLDL reduced the rate of mitochondrial respiration in human macrophages during this study. The link between metabolism and gene expression, particularly in immune cells, is an area of vigorous research and the integration of transcriptomic and metabolic datasets presents opportunities for the identification of functional changes in gene expression.

Overall, this work reveals that human blood-derived macrophages exhibit heterogeneity *in vitro* and can be separated by patterns of gene expression which recapitulate the range of known macrophage functions. It also shows that the macrophage response to oxLDL is largely conserved across cell types, but differences between them shed new light on processes central to atherosclerosis. The validation of human macrophage heterogeneity *in vivo*, coupled with the clarification of molecular mechanisms underlying oxLDL-induced gene expression, could lead to new therapies for atherosclerotic disease.

Chapter 5

Regulation of human endothelial gene expression by laminar shear stress

5.1 Introduction

Endothelial cells form a selectively permeable barrier lining the entire cardiovascular network, providing an interface between circulating blood components and all extravascular tissues. Major aspects of this barrier function include the containment of metabolites such as LDL in the vascular lumen, and the regulated transmigration of circulating immune cells at sites of inflammation. Dysregulation of these processes can lead to subendothelial deposition and accumulation of LDL, which is oxidised by various chemical and cellular factors and subsequently taken up by macrophages, which then release cytokines and other factors enhancing the recruitment and infiltration of circulating immune cells[23,231]. Although the major risk factors for atherosclerosis are systemic in nature (e.g. circulating LDL-cholesterol, age, gender, blood pressure, diabetes), plaque development is localised to specific parts of the vasculature, such as bifurcations, branches, and the inner curve of the aorta[5,34,232]. The key difference between these regions and more protected parts of the vascular tree is the haemodynamic environment, specifically the frictional force applied parallel to the endothelial surface by the flow of circulating blood, known as shear stress. Unidirectional, laminar flow in unbranched vessels generates relatively constant and high shear stress, which induces profound morphological changes in endothelial cells *in vitro*, and has been reported to promote an atheroprotective gene expression profile[32,36]. Conversely, low or oscillatory shear stress generated by disturbed flow is considered atherogenic, and endothelial cells cultured under these conditions display a range of functional differences likely to contribute to the development of atherosclerosis. These include increased proliferation and apoptosis, increased expression of adhesion molecules and inflammatory cytokines, and reduced production of vasodilatory nitric oxide (NO)[32].

The shear stress affecting the endothelium is sensed by a number of receptors at the endothelial cell membrane[40,233,234]. A range of responses to these signals have been described, and it is known that shear stress can directly influence expression of genes essential to normal endothelial function[37,42]. Little is known, however, about the factors mediating these changes at the level of the genome, and how mechanical forces in the endothelial environment can affect the epigenome. Previous studies have described an association between DNA methylation and atherosclerosis, and more specifically effects of oscillatory shear stress on endothelial gene methylation[44], however broader aspects of epigenetic regulation, such as chromatin accessibility and histone modifications, have only recently begun to be explored. The most recent study in this field reported the gain of H3K27Ac signal in HUVECs as early as 6 hours after the initiation of oscillatory shear stress, compared to static controls[92]. While H3K27Ac is associated with active regulatory elements, the factors governing the modulation of these marks are poorly understood. Additionally, studies employing deep transcriptomics and systems biology have identified relatively few differentially expressed genes between endothelial cells exposed to oscillatory shear stress versus static controls, whereas unidirectional laminar shear stress exerts a comparatively greater effect on endothelial gene expression when compared with either oscillatory shear stress or static controls[235]. This may therefore represent a more pertinent and informative comparison when studying the influence of endothelial mechanotransduction on atherosclerotic disease progression.

The development of coronary artery disease (CAD), of which atherosclerosis is the underlying cause, is multifactorial with both environmental and genetic influences. Many genomic loci comprising the heritable component of CAD risk have been identified by GWAS, and additional loci have been associated with CAD risk factors (e.g. blood pressure, LDL-cholesterol)[12]. Many of these variants lie outside of protein-coding

sequences, with a large number occurring in intergenic regions[13]. These variants are enriched in DHS from multiple cell types, and it is possible that some of these overlaps involve the disruption of sequences within regulatory elements (e.g. enhancers, promoters) affecting the expression of nearby genes. Transcription factor binding sites are concentrated in these elements, and binding affinity is dependent on the presence of specific sequence motifs. Therefore, some of the variation significantly associated with disease could result in disruption of transcriptional networks, and identification of the affected pathways may present opportunities for therapeutic intervention. Enhancer elements are cell-type and context specific, and determine the distinct gene expression profiles for all of an individual's different cell types[236]. As such, regulatory elements involved in mediating the response to environmental factors that contribute to disease progression could harbour disease-associated variants and explain the contribution of a specific locus to heritable CAD risk. Integration of epigenomics and transcriptomics enables the identification of such regulatory networks and has been used successfully to identify novel pathways affected by CAD-associated variation[53]. This approach can potentially be applied to a range of disease-specific environmental stimuli and individual cell types to identify causal mechanisms for GWAS variants.

The work presented in this chapter aims to establish an *in vitro* model of laminar shear stress and use this model to compare the transcriptional profile, determined using RNA-seq, of primary human aortic endothelial cells (HAECs) from three individual donors cultured under laminar shear stress with cells cultured in a static environment. It also aims, for the first time, to characterise genome-wide changes in HAEC chromatin accessibility in response to laminar shear stress using ATAC-seq, and to develop an understanding of how these changes may relate to changes in gene expression. Tools for transcription factor binding motif analysis will be deployed to identify potential

molecular mediators of any observed changes, and integration of RNA-seq and ATAC-seq datasets may reveal regulatory mechanisms of genes influencing endothelial cell biology relevant to atherosclerosis. Finally, overlapping disease-associated genetic variants identified by GWAS with genomic regions exhibiting differential activity between shear stress conditions may reveal pathways affected by heritable risk of cardiovascular disease, and ultimately lead to the development of therapeutic strategies for endothelial dysfunction in atherosclerosis.

5.2 *In vitro* application of high laminar shear stress to HAECs

HAECs were subjected to unidirectional laminar shear stress using a parallel-plate flow chamber connected to a flow circuit driven by a peristaltic pump (Figure 5.1A). The chamber was designed to allow modulation and definition of the shear stress exerted at the cell monolayer by altering the pump speed and medium flow rate (see methods). Cells cultured under high laminar shear stress (20 dynes/cm^2) for 48 hours showed pronounced morphological changes compared with control cells cultured in static medium. Statically cultured HAECs were densely packed and rounded, while those cultured under high laminar shear stress appeared less crowded and elongated along the axis of the direction of flow (Figure 5.1B). Cells cultured at lower shear stresses of 5 and 12.5 dynes/cm^2 showed a trend towards the same morphology, but the effect was greatest at 20 dynes/cm^2. This shear stress was selected for further experiments based on these observations, as well as data from haemodynamic studies suggesting an average shear stress of 15 dynes/cm^2 at the wall of the aorta[237]. Shear stress in large arteries can transiently reach as high as 60 dynes/cm^2, and in order to determine the effects of 'high' laminar shear stress on aortic endothelial cells a value greater than the average was chosen.

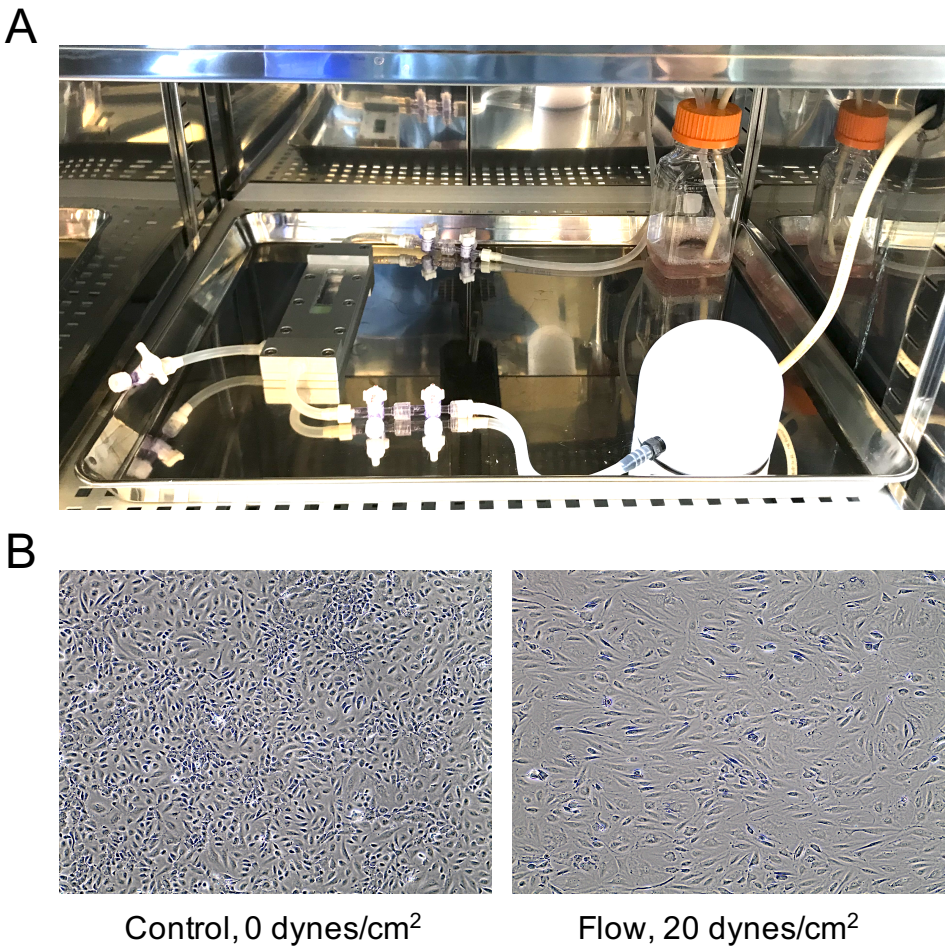

Figure 5.1: An in vitro model of laminar shear stress. (A) Assembled flow circuit with HAECs under laminar shear stress. Cells are cultured on glass slides which are loaded into the flow chamber (left) and medium is supplied from a reservoir (right, rear) to the peristaltic pump (not pictured). From the pump, the medium enters a pulse dampener (right, front) before reaching the flow chamber, after which it returns to the reservoir. 4-way stopcocks allow isolation of the flow chamber for removal and observation of the cells. (B) HAECs cultured in static medium (Control, left) or under high laminar shear stress (Flow, right) at 10X magnification.

5.3 Preparation and sequencing of RNA-seq libraries

HAECs from three donors were cultured under static and high laminar shear stress conditions for 48 hours, after which cells were lysed for purification of total RNA. Purity of samples was assessed using a NanoDrop ND-1000 spectrophotometer, and samples with a 260/280 ratio of ~2.0 and a 260/230 ratio between 2.0-2.2 were submitted to Oxford Genomics Centre, Wellcome Centre for Human Genetics, for Illumina TruSeq

library preparation. Sequencing of libraries prepared from polyA-selected transcripts on an Illumina HiSeq 4000 generated 75bp paired-end reads of high quality, as determined by comparison of the GC content of the sequenced reads with the theoretical GC content distribution for the human transcriptome (GRCh38/hg38), and analysis of Phred scores for quality of base calls. Read counts were similar between libraries (Figure 5.2A), with a mean count of 72,320,526 paired-end reads per sample, and % GC content was comparable across all libraries and closely followed the distribution generated from 75bp sequences randomly drawn from the reference transcriptome (Figure 5.2B). Average Phred scores were high (>30, corresponding to a probability of incorrect base call less than 10^{-3}) at all positions with a minor decline in quality at the end of mate 2 reads, which is a known feature of Illumina sequencing platforms (Figure 5.2C and D).

5.4 RNA-seq read alignment and quality control

Reads were aligned to the GRCh38/hg38 reference genome using the splice-aware STAR aligner at a rate of ~98% for all samples. Alignments were filtered to retain only uniquely mapped and properly paired reads, resulting in > 60 million quality-filtered alignments per sample (Table 5.1). Given that all remaining alignments consisted of a proper pair of paired-end reads, each pair was counted as an aligned fragment in downstream analyses (i.e. 1 fragment = 1 count). A gene-wise fragment count matrix for all samples was generated using GENCODE v29 gene annotations, and the majority of fragments were unambiguously assigned to genes in all samples (Figure 5.3A), with an average of 6.03% of fragments either ambiguously assigned (i.e. to multiple genes), or mapping outside of annotated regions. Counts were normalised using the trimmed mean of M-values (TMM) approach employed in the edgeR pipeline[103,104], accounting for composition bias which

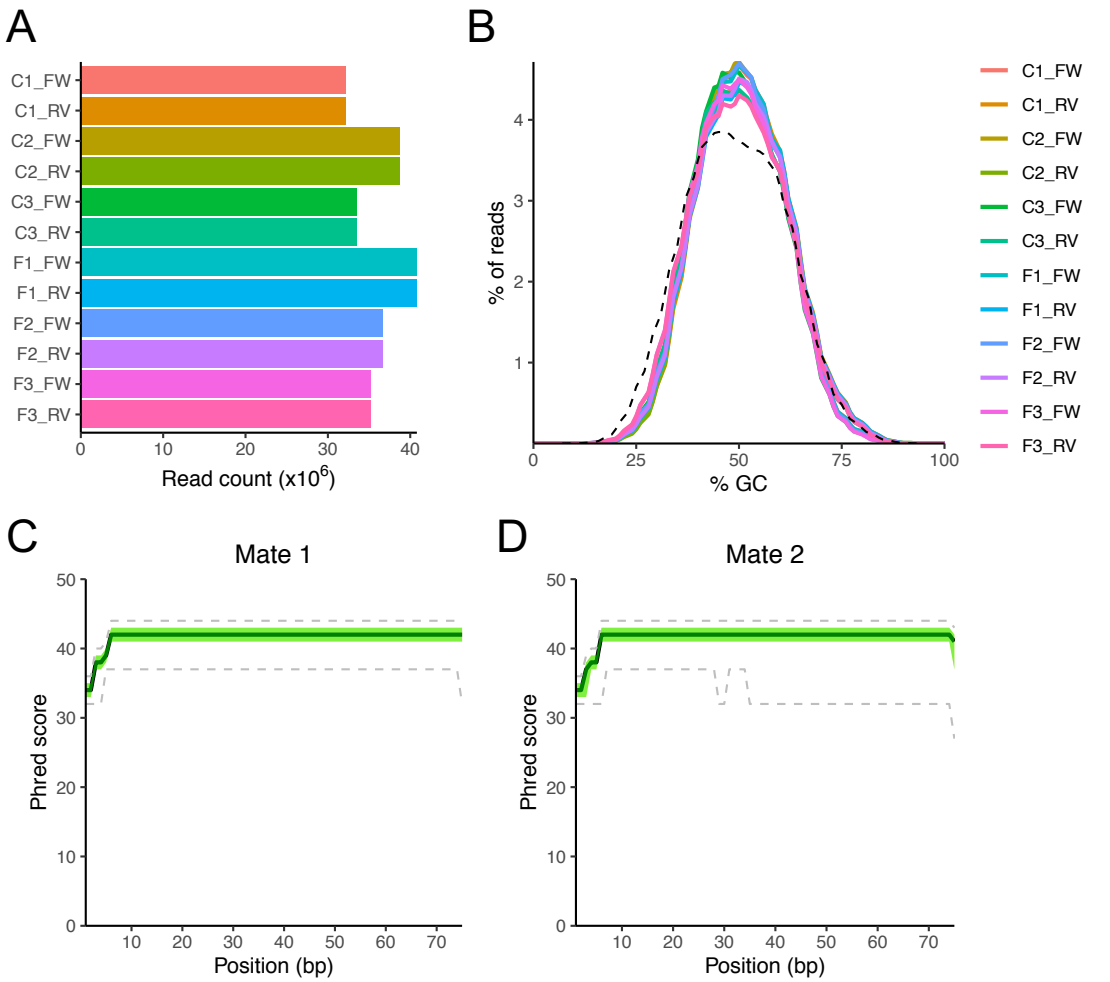

Figure 5.2: Quality control plots for RNA-seq fastq sequence files. (A) Read counts for each sample. Mate 1 and mate 2 paired-end reads are denoted by '_FW' and '_RV' following the sample name, respectively. (B) Percentage of G or C bases called in all reads in each sample. The theoretical %GC distribution for the human transcriptome (GRCh38/hg38) is shown by the dashed black line. (C and D) Per-base Phred scores for all reads in all fastq files containing mate 1 (C) and mate 2 (D) reads. The dark green line shows the median Phred score at each position, and the light green shaded area marks the lower and upper quartiles. 10th and 90th percentiles are marked by the dashed grey lines.

can cause systematic differences in average expression, to obtain an expression estimate for each gene expressed as counts per million mapped reads (CPM). Genes with low expression provide little evidence for differential expression between control and flow samples, and the low counts may interfere with downstream statistical analysis. Therefore, genes with low expression in all samples, defined here as a $\log_2$ CPM < 1, were filtered out of the normalised expression matrix. After filtering, 12,651 genes

remained in the expression matrix, and the overall distribution of normalised gene counts was consistent between samples (Figure 5.3B). In order to verify the concordance of samples within each treatment group, Pearson correlation coefficients for the gene-wise counts in each sample were calculated. All samples were highly correlated, but correlation was greater between samples within the same treatment group, suggesting distinct gene expression profiles for each treatment group (Figure 5.3C). Sample similarity was further examined using principal component analysis (PCA) and a plot of the first two principal components of the expression matrix showed clear separation of control and flow samples in the first principal component, which explained 80.5% of variance (Figure 5.3D). The second principal component appeared to correspond to donor-specific variation and explained a small proportion of total variance (11.1%). One of the samples, F2, clustered away from the other flow samples along the second principle component. This was found to be driven by a small number of highly expressed genes with high variance between flow samples, which also contributed to the first principal component, and as such was deemed to be a donor-specific effect rather than a systematic and potentially confounding source of variance.

Table 5.1: Summary statistics for alignment of RNA-seq reads. Numbers represent records contained in BAM files from mapping paired-end reads, and the total includes primary and secondary alignments, as well as unmapped reads. Proportions are shown as percentage of total records contained in the BAM output from the STAR aligner.

Sample	Total	Mapped	% Mapped	Filtered reads	% Passed
C1	70,955,006	69,419,274	97.84%	60,144,800	84.76%
C2	84,863,774	83,618,546	98.53%	72,914,542	85.92%
C3	73,099,152	72,083,980	98.61%	63,344,206	86.66%
F1	88,458,476	87,052,356	98.41%	77,327,418	87.42%
F2	79,677,634	78,443,806	98.45%	69,412,008	87.12%
F3	75,319,452	74,355,884	98.72%	67,030,430	88.99%

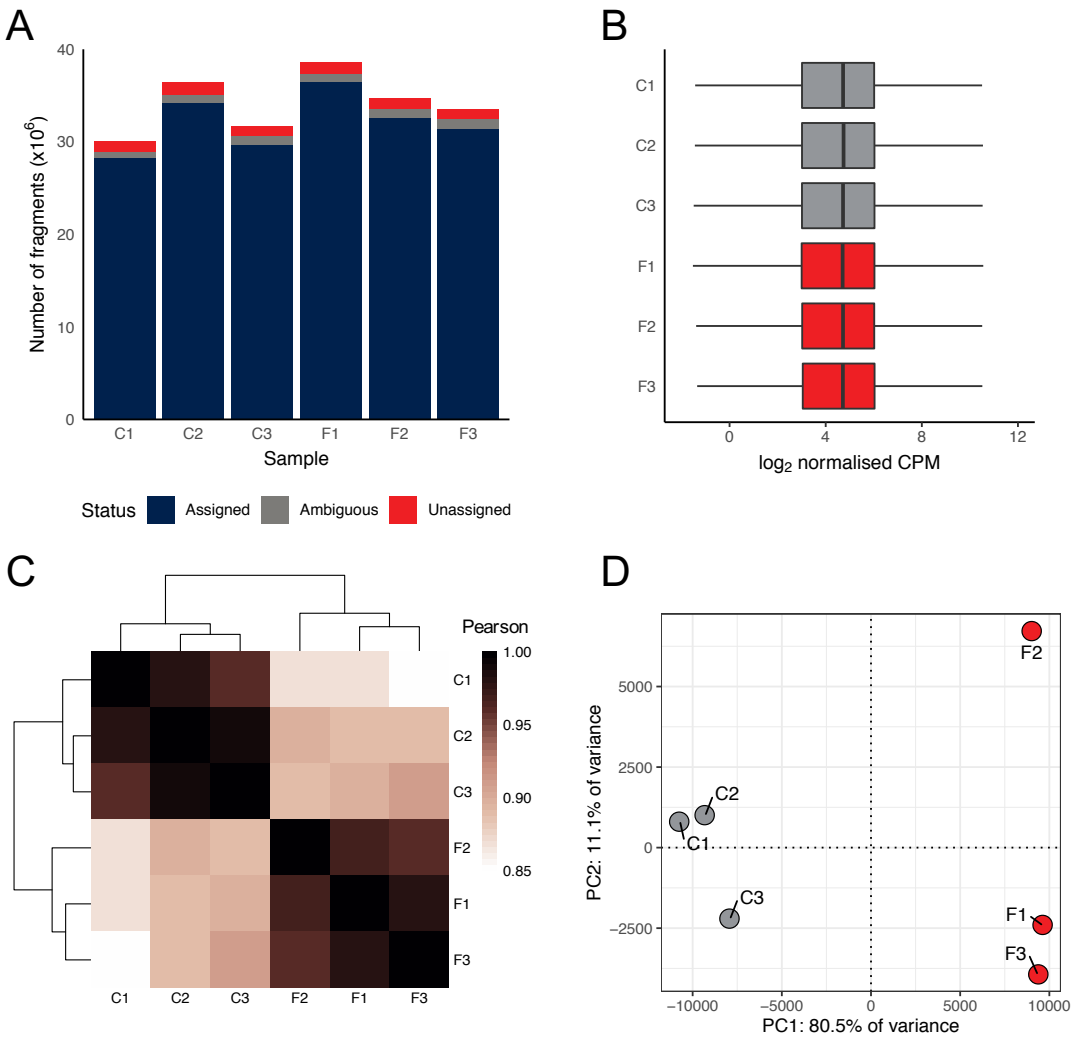

Figure 5.3: Quality control and preliminary analysis of normalised RNA-seq count data. (A) Number of fragments in each library assigned to a gene annotation, ambiguously assigned (and therefore not included in the final count table), or not assigned to a gene in the annotation file. (B) Boxplots of the distribution of normalised counts across all genes in each sample, expressed as the $\log_2$-transformed counts per million mapped reads (CPM). Samples in the control treatment group are shown in grey, and samples in the flow treatment group are shown in red. (C) Heatmap of Pearson correlation coefficients for comparison of normalised gene counts in each sample. Dendrograms show unsupervised clustering of gene counts based on Euclidean distances between samples (D) Principal component analysis of normalised gene counts. Samples are coloured as for (B).

5.5 Differential gene expression analysis

The quasi-likelihood (QL) framework[106] in the edgeR package was chosen for detection of differentially expressed genes, due to its sophisticated statistical modelling of technical and biological variability in count data generated by RNA-seq using the negative binomial (NB) distribution, and the rigorous control of type I error rate afforded by empirical Bayesian moderation of the QL dispersion estimates for each gene. This is especially important when the number of replicates is low, and thus the information available for stable estimation of the gene-wise QL dispersion parameter is limited. As part of the edgeR pipeline, NB dispersion estimates were obtained for all genes in the HAEC expression matrix, which exhibited the expected mean-variance trend for RNA-seq count data, with lower variance at higher average CPM (Figure 5.4A). Generalised linear models were fitted to expression values including coefficients for treatment group and donor of the original sample, reflecting the paired-sample experimental design and explicitly modelling the baseline differences in gene expression between donors. Estimates for the QL dispersion were calculated and the squeezed dispersion values visualised to determine the extent of shrinkage towards the trended dispersion (Figure 5.4B). The QL dispersion estimates for the majority of genes showed substantial shrinkage towards the trended dispersion, indicating low variability in the QL dispersions between genes. Analysis of differential gene expression between treatment groups was performed using QL F-tests and resulting p-values were adjusted using the Benjamini-Hochberg procedure to correct for multiple testing. At a threshold false discovery rate (FDR) of 5%, 6,982 genes were differentially expressed, with 3,426 genes increasing in expression, and 3,556 genes decreasing in expression in the flow treatment group compared with control (Figure 5.5). Of these, 559 and 384 genes were up- and downregulated by more than two-fold, respectively (Figure 5.6). Among the most

differentially expressed genes were several with known roles in immune processes and inflammation (e.g. *IL21R, TGFA, IL11, CXCL12, CCL14, CXCL6, CCL16, IL33, CXCL8*) as well as genes with previously described responses to shear stress (e.g. *CYP1B1, GPR68, KLF4, CYP1A1, KLF2, FABP4*). A number of key regulators of endothelial cell homeostasis were also differentially expressed between control and flow cells. Genes expressing vascular endothelial growth factors (*VEGFA*, *VEGFB*, *VEGFC*) were downregulated by high laminar shear stress, as was the angiopoietin-1 signalling antagonist *ANGPT2*. Interestingly, the genes encoding receptors for these signalling molecules (*FLT1*, *KDR, TEK*) were upregulated by high laminar shear stress, along with genes encoding molecules involved in establishing inter-endothelial cell junctions (*CDH5*, *PECAM1*), supporting the role of shear stress in regulating endothelial activation and permeability.

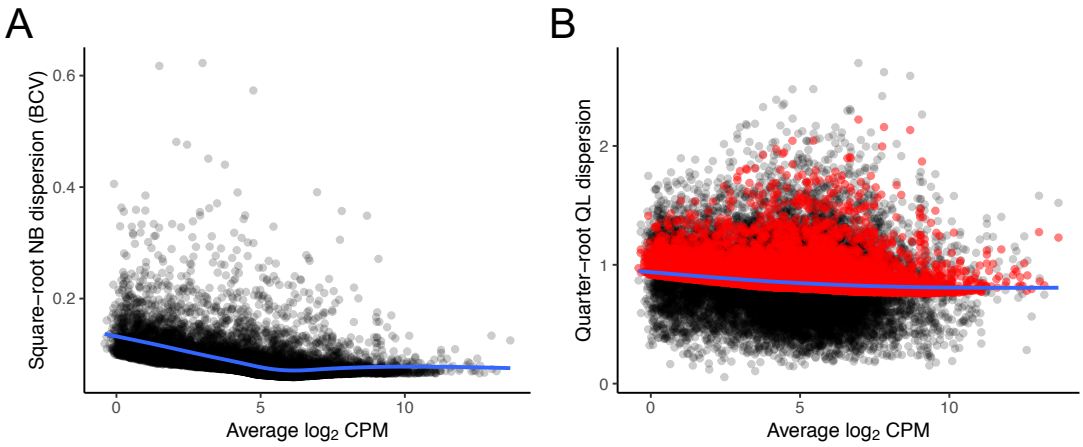

Figure 5.4: Dispersion estimates for edgeR differential gene expression analysis. (A) Square root of the negative binomial (NB) dispersion estimates (Biological coefficient of variance, BCV) for each gene in the expression matrix against average expression. (B) Quarter-root of the quasi-likelihood (QL) dispersion estimates against average expression. Prior estimates are shown in black, and squeezed estimates produced by empirical Bayesian moderation shown in red. The mean-variance trend in each plot is shown by the blue line.

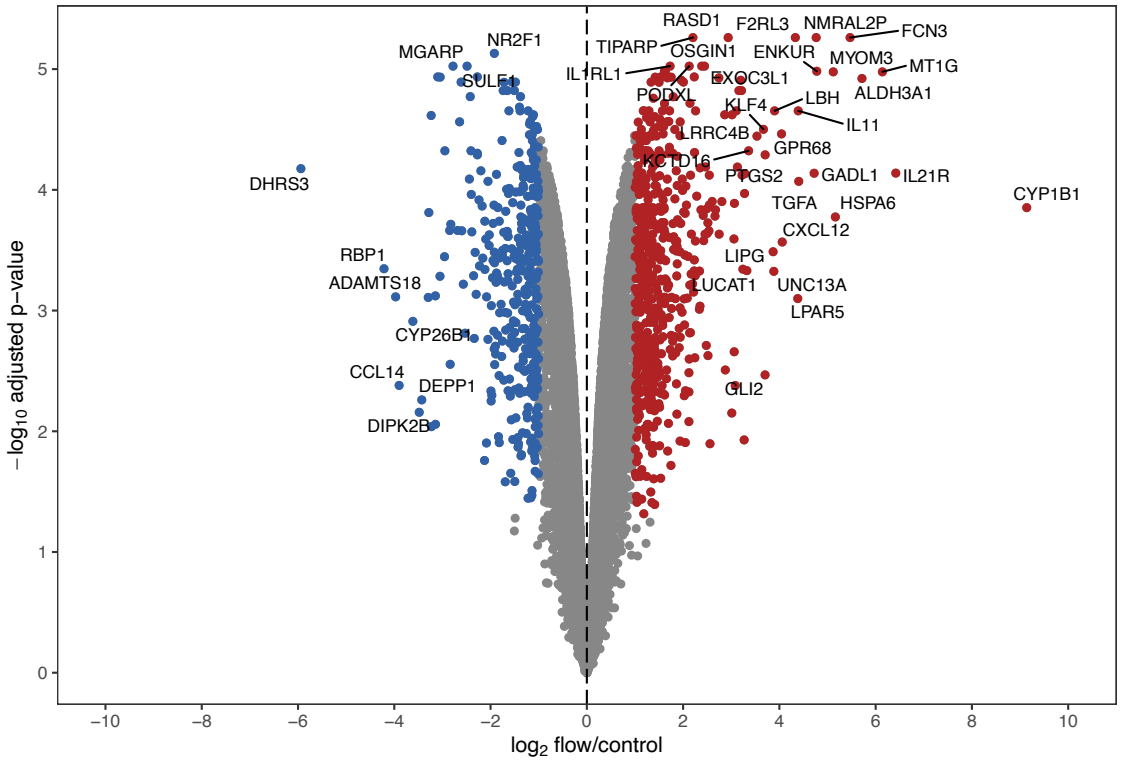

Figure 5.5: Volcano plot of differential expression analysis results. Red and blue points show genes upregulated and downregulated by at least 2-fold, respectively. Labelled genes either differ in expression between conditions by at least 10-fold or have an adjusted p-value < 10^{-5}.

Table 5.2: Top 10 most upregulated genes in HAECs under high laminar shear stress.

Symbol	**$\log_2$ flow/control**	**adj. p-value**	**Name**
CYP1B1	9.14	1.40E-04	Cytochrome P450 Family 1 Subfamily B Member 1
IL21R	6.42	7.27E-05	Interleukin 21 Receptor
MT1G	6.14	1.05E-05	Metallothionein 1G
ALDH3A1	5.72	1.20E-05	Aldehyde Dehydrogenase 3 Family Member A1
FCN3	5.47	5.48E-06	Ficolin 3
HSPA6	5.16	1.68E-04	Heat Shock Protein Family A (Hsp70) Member 6
MYOM3	5.12	1.05E-05	Myomesin 3
ENKUR	4.78	1.04E-05	Enkurin, TRPC Channel Interacting Protein
NMRAL2P	4.76	5.48E-06	NmrA Like Redox Sensor 2, Pseudogene
GADL1	4.72	7.29E-05	Glutamate Decarboxylase Like 1

Table 5.3: Top 10 most downregulated genes in HAECs under high laminar shear stress.

Symbol	log_2 flow/control	adj. p-value	Name
DHRS3	-5.94	6.68E-05	Dehydrogenase/Reductase 3
RBP1	-4.21	4.50E-04	Retinol Binding Protein 1
ADAMTS18	-3.97	7.70E-04	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 18
CCL14	-3.90	4.17E-03	C-C Motif Chemokine Ligand 14
CYP26B1	-3.61	1.23E-03	Cytochrome P450 Family 26 Subfamily B Member 1
DIPK2B	-3.48	6.96E-03	Divergent Protein Kinase Domain 2B
DEPP1	-3.43	5.50E-03	DEPP1 Autophagy Regulator
AQP1	-3.29	7.78E-04	Aquaporin 1
DPP4	-3.28	1.54E-04	Dipeptidyl Peptidase 4
CXCL6	-3.23	2.42E-05	C-X-C Motif Chemokine Ligand 6

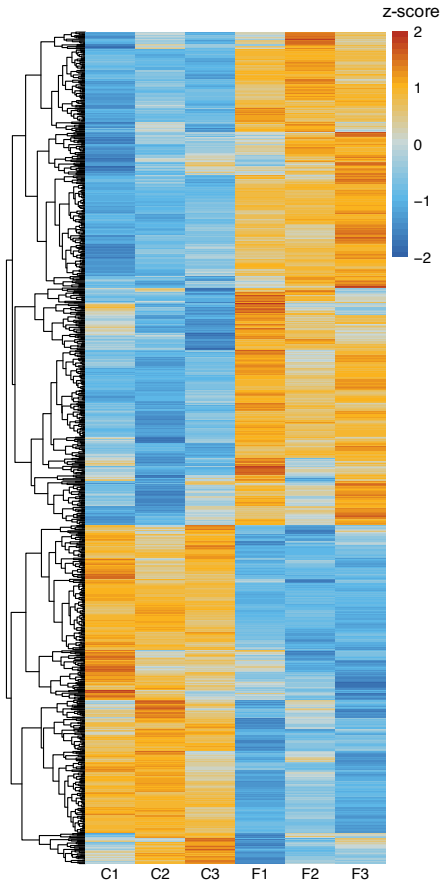

Figure 5.6: Shear stress-responsive gene expression. Heatmap of scaled log_2 counts per million mapped reads (CPM) of 943 differentially expressed genes with an absolute fold change greater than 2 and an adjusted p-value < 0.05. Each row shows a single gene, and values correspond to the z-score relative to the mean expression of each gene.

5.6 Gene ontology and pathway enrichment analysis

In order to interpret the wider consequences and biological context of differential gene expression in response to high laminar shear stress, the Gene Ontology (GO), Kyoto Encyclopaedia of Genes and Genomes (KEGG) and Reactome databases were used to test for enrichment of gene annotations relating to biological processes or pathways. Up- and downregulated genes were separately tested for overrepresentation of gene annotations in each database and resulting p-values were adjusted for multiple testing (Benjamini-Hochberg). For upregulated genes (Figure 5.7), the most strongly enriched terms from the 'GO: Biological Process' (BP) annotation database related to subcellular organisation, cell motility and signaling, and similarly the top terms from the 'GO: Molecular Function' (MF) database related to transmembrane ion transport, cytoskeletal rearrangement and signaling. Enriched KEGG pathways for upregulated genes also related to cytoskeletal organisation, membrane trafficking and signaling via Ras GTPase, and further identified enriched pathways involved in diseases such as cancer and cardiomyopathy. The Reactome annotation sets gave similar results, and also showed enrichment of terms relating to immune functions, extracellular matrix organisation, and platelet activation. For downregulated genes (Figure 5.8), there was very strong enrichment for terms relating to RNA processing, transcription, and translation in all annotation sets. Enriched KEGG pathways for downregulated genes also included 'DNA replication' and 'Cell cycle'. The GO and KEGG databases were found to contain terms relating to the cellular response to shear stress, although in the case of the GO terms only a small set of genes are annotated with such a term, with the largest set 'GO:0034405 response to fluid shear stress' containing 29 genes. This term, as well as the KEGG pathway 'hsa05418 Fluid shear stress and atherosclerosis', was significantly

overrepresented in the list of differentially expressed genes (adjusted p-value = 0.0263 and 0.00816, respectively).

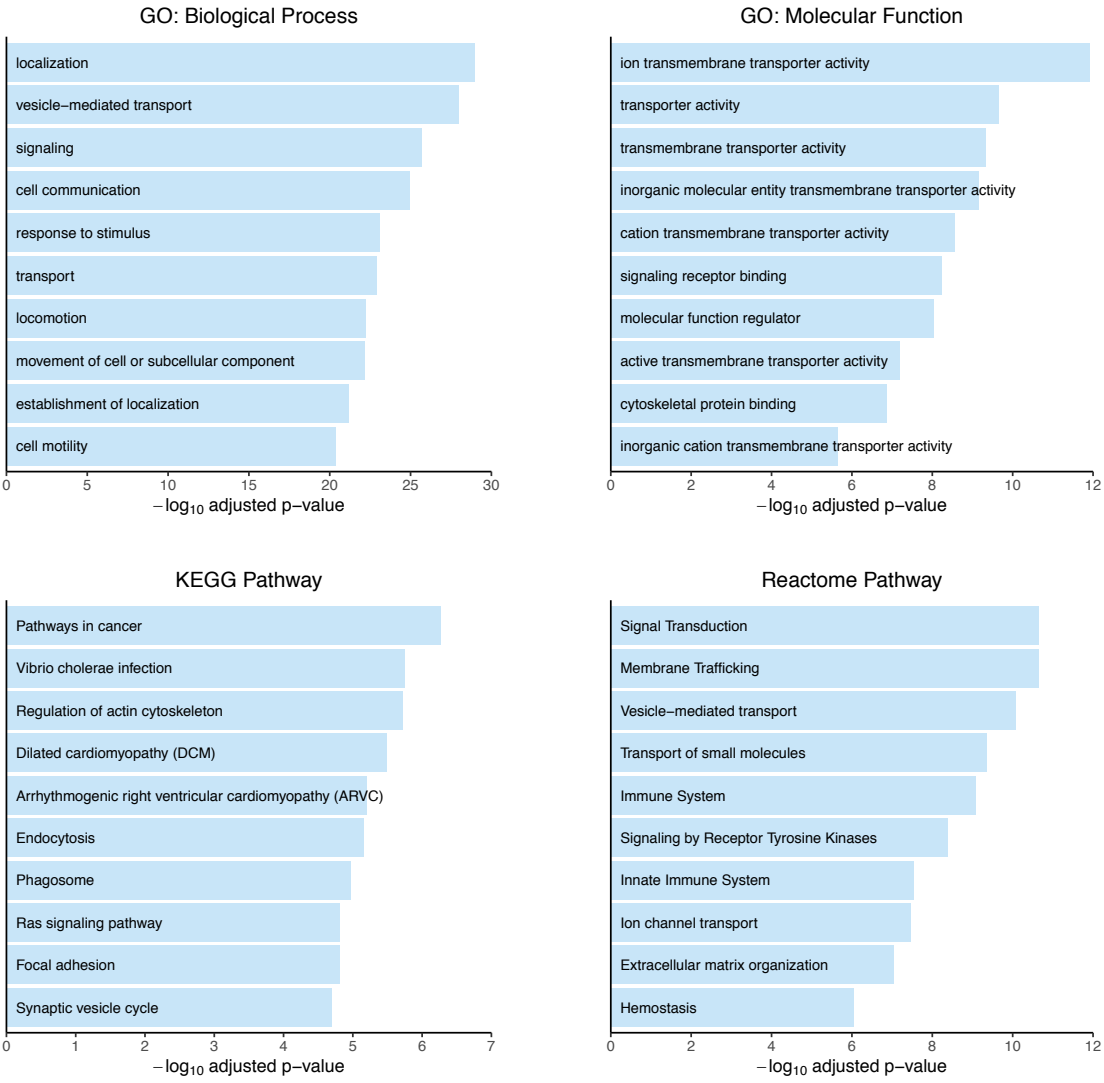

Figure 5.7: Top 10 enriched terms from each annotation database for genes upregulated by high laminar shear stress in HAECs.

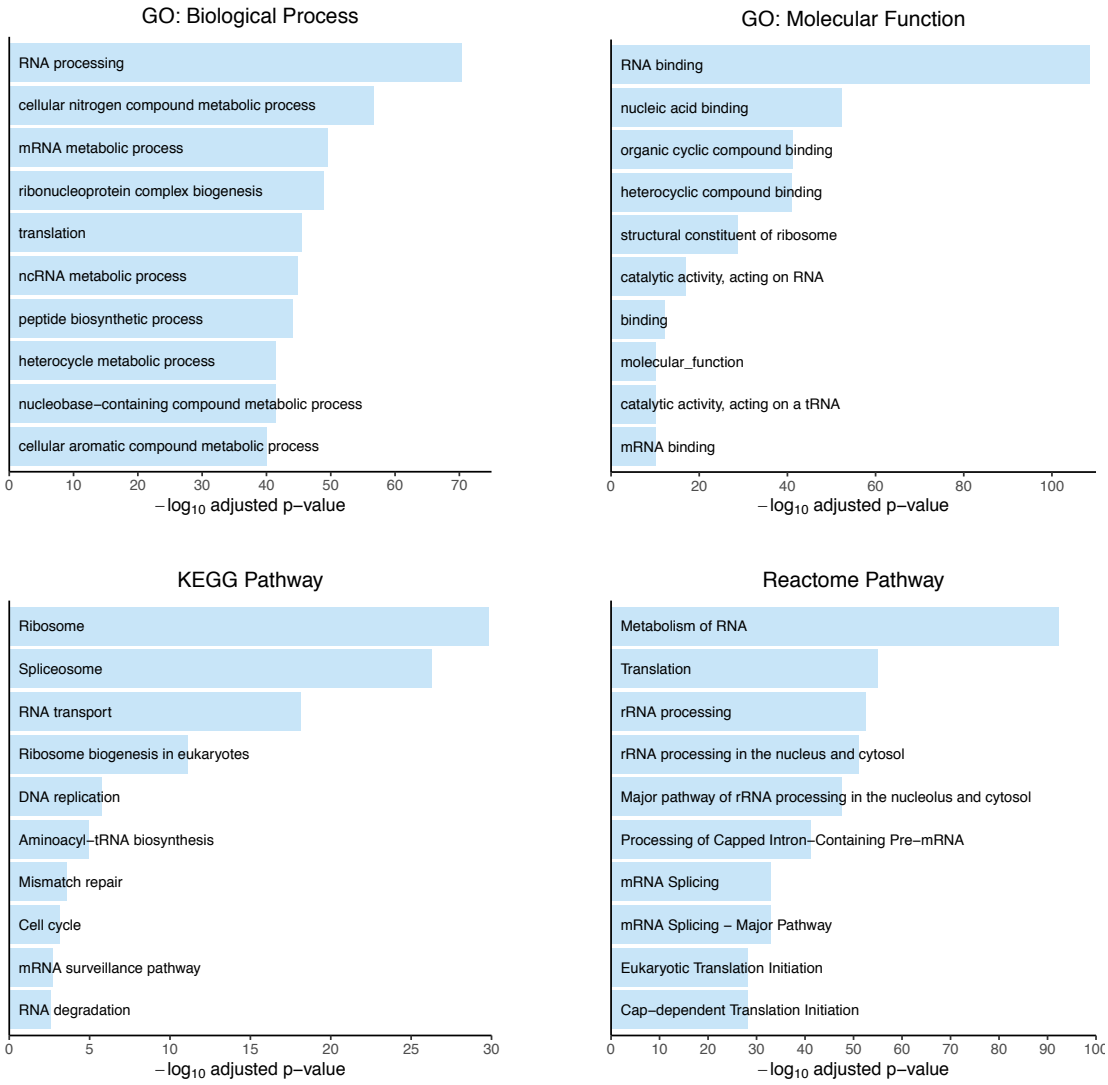

Figure 5.8: Top 10 enriched terms from each annotation database for genes downregulated by high laminar shear stress in HAECs.

5.7 Preparation, sequencing and alignment of ATAC-seq libraries

Transcriptomic analysis of HAECs under high laminar shear stress revealed extensive gene expression changes in response to high laminar shear stress, and there is substantial evidence in the literature of a link between chromatin accessibility and transcriptional regulation[77]. Therefore, in order to study regulatory mechanisms underlying these expression changes, cells from the same three donors were cultured in static and high laminar shear stress conditions for 48 hours, as described previously, and processed according to the original ATAC-seq protocol[107]. Quality of prepared libraries was assessed using the High Sensitivity D1000 assay on an Agilent TapeStation 2200, and successful chromatin transposition was associated with a characteristic periodicity of fragment size every ~150bp, corresponding to the length of DNA contained in a single nucleosome (Figure 5.9). Libraries of transposed DNA from accessible chromatin in the control and flow HAEC samples were indexed, quantified (fluorescence-based quantification) and pooled before submission to Oxford Genomics Centre and sequenced as 75bp paired-end reads on an Illumina HiSeq 4000. Sequenced reads were checked for sequencing depth and Phred scores of base calls, as for RNA-seq. %GC content was comparable across samples but differed from the theoretical distribution for the reference genome. This is likely a reflection of the non-random distribution of accessible regions at GC-rich elements, such as promoters[238]. Given the variable lengths of the fragments expected, reads were checked for adapter content and trimmed to remove sequences attributed to adapters added during library preparation. Trimmed reads were aligned to the GRCh38/hg38 reference genome and filtered to retain only uniquely mapped, properly paired reads.

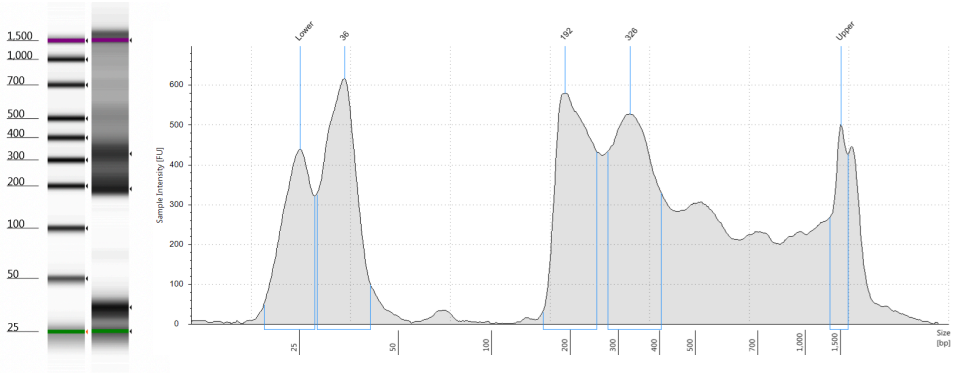

Figure 5.9: Representative TapeStation 2200 results for pooled ATAC-seq libraries. On the left is a reconstructed gel image showing the periodic bands characteristic of successful chromatin transposition, which is also shown in the electropherogram on the right. The peaks labelled 'Lower' and 'Upper' are markers of known length, and the peak at 36bp is due to the presence of primers used during PCR amplification of libraries.

5.8 Omni-ATAC-seq vastly reduces mitochondrial contamination

A known issue with the original ATAC-seq protocol is that a significant proportion of the sequenced reads align to the mitochondrial chromosome, and ~26-46% of reads in the HAEC samples aligned to chrM. This has implications for downstream analysis, as an abundance of reads mapping to chrM suggests reduced sequencing of fragments originating from the nuclear genome, and therefore a loss of autosomal signal. Since the publication of the original ATAC-seq protocol, a number of modified protocols designed to overcome the issue of mitochondrial contamination have been published, including a protocol known as Omni-ATAC[108]. Fresh samples of chromatin from control and flow HAECs were processed according to the Omni-ATAC protocol, and transposed DNA was assessed by TapeStation 2200, as described previously. Omni-ATAC libraries were sequenced as for the original ATAC-seq libraries, and reads were trimmed to remove adapter sequences, quality checked (Figure 5.10), and aligned to the reference genome. There was a dramatic reduction in the number of mitochondrial reads in the Omni-ATAC samples, with an average of 0.94% of reads aligning to chrM (Table 5.4). Given the

improvement in the mitochondrial read content, and the high quality of generated reads, the Omni-ATAC samples were taken forward for complete analysis in place of the original samples.

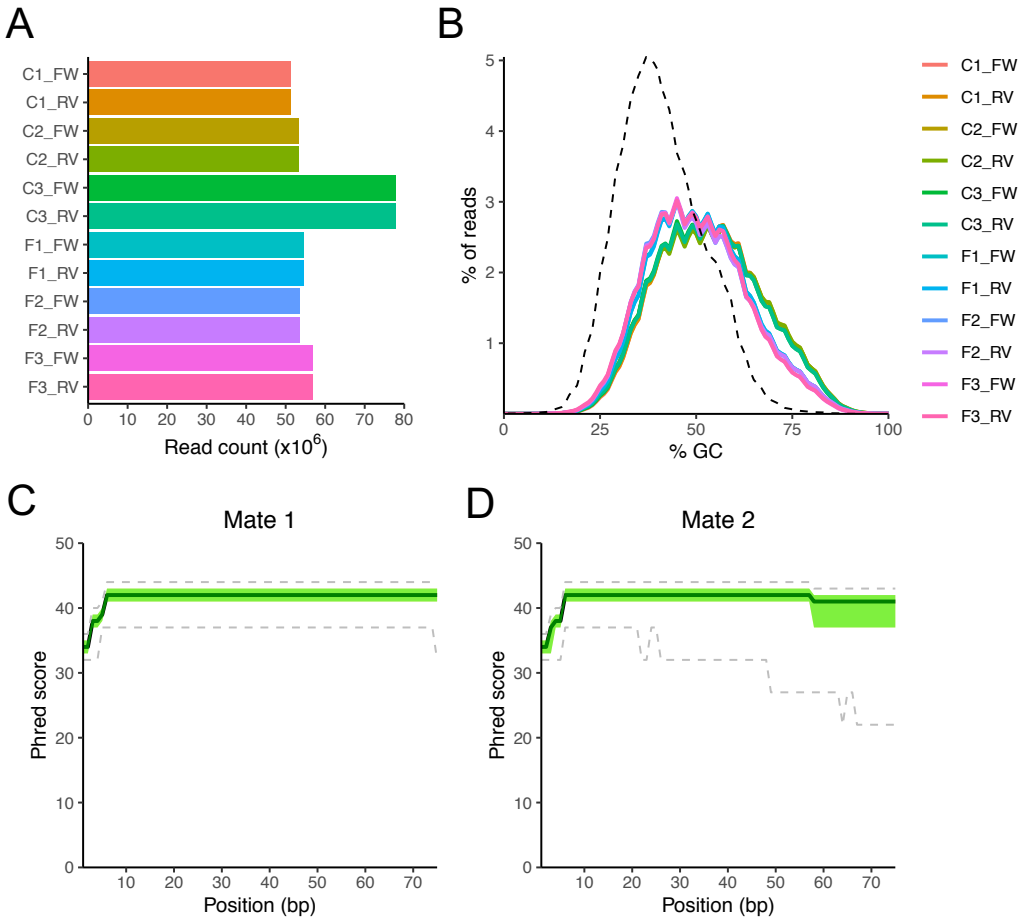

Figure 5.10: Quality control plots for adapter-trimmed Omni-ATAC-seq fastq sequence files. (A) Read counts for each sample. Mate 1 and mate 2 paired-end reads are denoted by '_FW' and '_RV' following the sample name, respectively. (B) Percentage of G or C bases called in all reads in each sample. The theoretical %GC distribution for the human genome (GRCh38/hg38) is shown by the dashed black line. (C and D) Per-base Phred scores for all reads in all fastq files containing mate 1 (C) and mate 2 (D) reads. The dark green line shows the median Phred score at each position, and the light green shaded area marks the lower and upper quartiles. 10th and 90th percentiles are marked by the dashed grey lines.

Table 5.4: Comparison of reads aligning to chrM in all samples generated using either the original ATAC-seq (ATAC) or the Omni-ATAC-seq (Omni) protocol.

Sample	Protocol	Mapped reads	chrM reads	% chrM
C1	ATAC	142,286,084	50,857,346	35.74
C2	ATAC	85,483,834	36,507,318	42.71
C3	ATAC	122,866,977	35,521,397	28.91
F1	ATAC	103,566,752	28,423,274	27.44
F2	ATAC	104,816,865	47,875,847	45.68
F3	ATAC	134,177,236	35,231,970	26.26
C1	Omni	95,795,219	887,489	0.93
C2	Omni	98,553,419	826,630	0.84
C3	Omni	144,013,935	1,530,103	1.06
F1	Omni	100,685,399	719,848	0.71
F2	Omni	98,799,867	1,130,777	1.14
F3	Omni	104,507,640	980,468	0.94

5.9 Quality control of Omni-ATAC-seq alignments

In addition to mitochondrial contamination, another known issue with ATAC-seq (and most DNA sequencing library preparation protocols) is the production of duplicate fragments during PCR amplification. A duplicate fragment is defined as a proper pair of sequenced reads in which the 5' positions of both reads are identical to those of another pair of reads. The duplication rate was high for the Omni-ATAC samples, with an average of 26.1% of reads marked as duplicates (Table 5.5). Since these are far more likely to be artefacts of PCR amplification than truly identical fragments produced by transposition of chromatin, they were removed. Alignments were also filtered to retain only uniquely mapping, properly paired reads. The sequenced fragments recapitulated the periodicity in size observed during library quality control, with peaks in the distribution every ~200bp (Figure 5.11). A large peak was observed for fragments between 0 and ~150bp, corresponding to nucleosome-free fragments produced by insertion of Tn5 transposases in accessible regions between nucleosomes. Subsequent peaks at ~200bp, ~400bp and

~600bp were deemed mono-, di- and tri-nucleosomal peaks produced by transposase insertion either side of the corresponding number of nucleosomes. A sequenced fragment in an ATAC-seq experiment represents a pair of Tn5 transposase insertion events located at either end (i.e. at the 5' ends of each read in the pair). Read start coordinates were shifted to represent the centre of each 9bp insertion event, with reads aligning to the + strand offset by +4bp and reads aligning to the - strand offset by -5bp. One of the quality metrics used by the ENCODE consortium to assess quality of ATAC-seq experiments is the transcription start site (TSS) enrichment score. This provides a measure of signal to noise ratio by forming an aggregate distribution of reads in 2000bp windows centred on a reference set of TSSs and determining the fold change in signal above the average coverage at each end of the window (i.e. in 100bp bins for 200bp of averaged coverage total). Since TSSs are transcriptionally active, they are expected to be accessible to proteins such as general transcription factors and RNA polymerase II, and therefore also accessible to the Tn5 transposase during ATAC-seq sample preparation. The distribution of the shifted reads from the control and flow samples around Refseq GRCh38/hg38 TSSs showed a clear peak at the TSS for all samples (Figure 5.12). ENCODE recommends that acceptable TSS enrichment scores, taken as the fold change above background at the centre of the distribution, are > 5, and ideal scores are > 7. All control samples had ideal TSS enrichment scores (> 7), while flow samples had slightly lower scores, but they were still above the acceptable minimum score of 5. The lower signal to noise ratio of the flow samples most likely represents a technical bias that will need to be addressed in downstream analyses. All samples showed a secondary peak ~200bp downstream of the TSS which indicates the position of the +1 nucleosome, demonstrating the sensitivity of ATAC-seq to chromatin structure across the genome.

Table 5.5: Summary statistics for alignment of Omni-ATAC-seq reads. Numbers represent records contained in BAM files from mapping paired-end reads, and the total includes primary and secondary alignments, as well as unmapped reads. Proportions are shown as percentage of total records contained in the BAM output from the Rsubread aligner.

Sample	Total	Mapped	% Mapped	Marked duplicates	% Marked	Filtered reads	% Passed
C1	102,554,442	95,795,219	93.41%	24,915,404	26.01%	66,181,708	64.53%
C2	106,414,586	98,553,419	92.61%	26,125,010	26.51%	67,576,866	63.50%
C3	155,551,108	144,013,935	92.58%	38,965,575	27.06%	97,979,752	62.99%
F1	109,121,628	100,685,399	92.27%	25,758,610	25.58%	69,748,578	63.92%
F2	107,087,038	98,799,867	92.26%	26,935,454	27.26%	66,688,098	62.27%
F3	113,741,876	104,507,640	91.88%	25,262,988	24.17%	73,746,236	64.84%

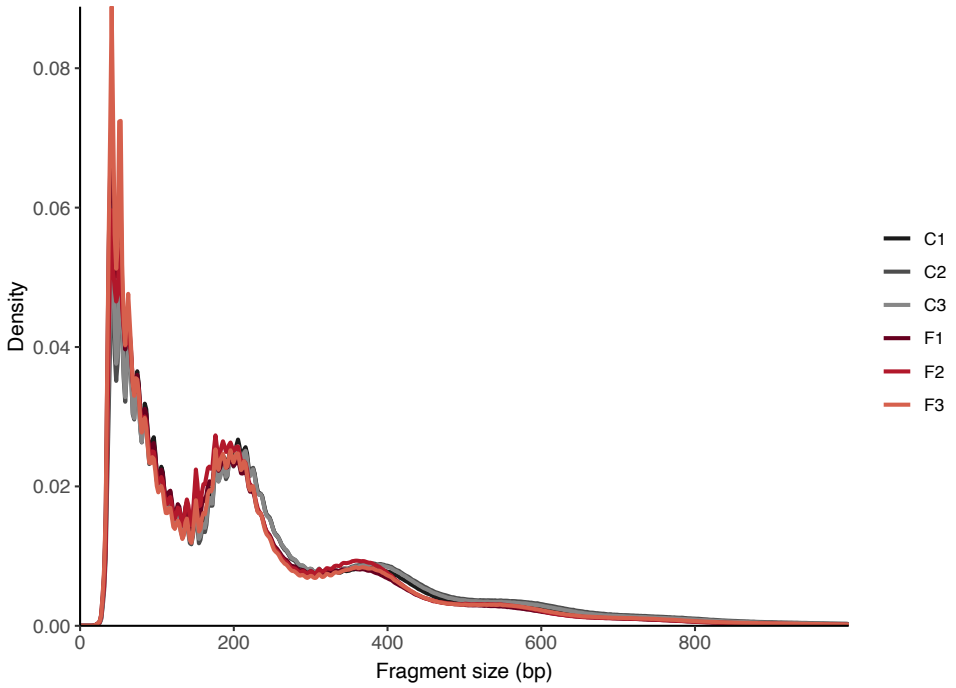

Figure 5.11: Distribution of sequenced fragment sizes for all Omni-ATAC-seq samples.

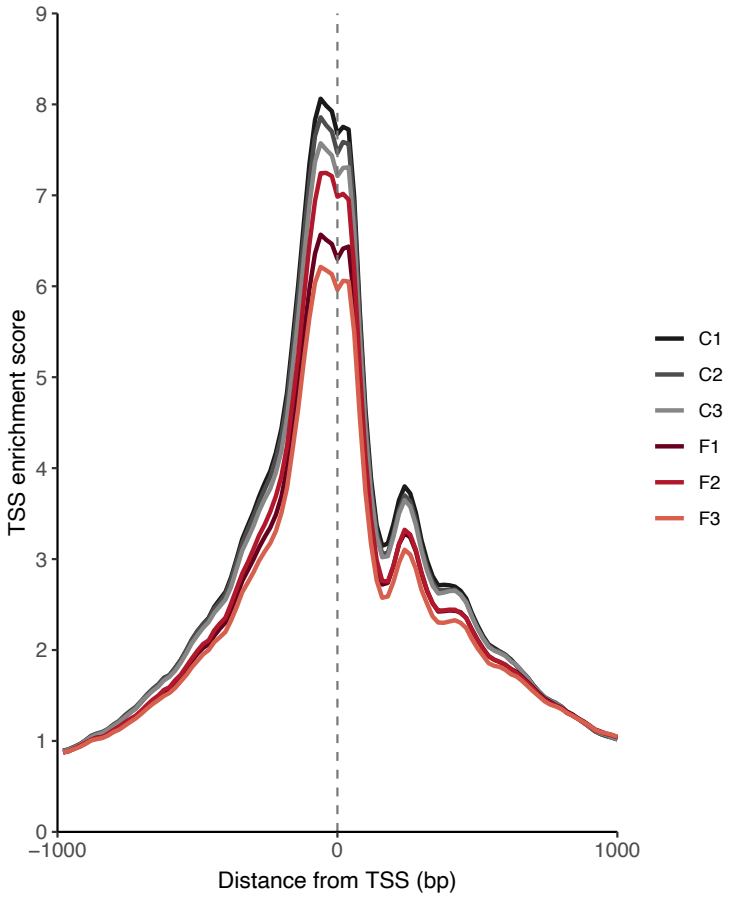

Figure 5.12: TSS enrichment plots for all Omni-ATAC-seq samples. The score at each position is calculated from the coverage in 2000bp windows centred on GRCh38/hg38 Refseq TSSs normalised to background signal in the terminal 100bp bins.

5.10 High laminar shear stress alters chromatin accessibility in HAECs

In order to identify genomic regions responsive to high laminar shear stress, peaks of signal enrichment were called using a merged set of the shifted Omni-ATAC-seq reads from all samples with MACS2. This resulted in a set of 345,519 peaks, which were filtered down to 264,466 peaks with an adjusted p-value < 10^{-5}. Peaks in unlocalised and unplaced scaffolds of the reference genome, in blacklisted regions, and on chrY were also removed, leaving a consensus set of 263,484 accessible chromatin peaks in HAECs cultured under control and flow conditions. Another quality metric used by the ENCODE

consortium to assess signal to noise in sequencing samples is the fraction of reads in peaks (FRiP), with an ideal minimum FRiP score of 0.3. Counts of shifted reads in each peak were obtained for each sample, and FRiP scores were determined to be > 0.3 for all samples, although differences were noted between treatment groups, as for TSS enrichment (control mean FRiP = 0.57, flow mean FRiP = 0.43). A count matrix, similar to the expression matrix generated for RNA-seq, was constructed and loaded into edgeR for differential analysis.

A 'first pass' differential analysis using the edgeR QL framework revealed that normalising peak counts by sequencing depth alone resulted in systematic downregulation of signal in peaks for flow samples. This was considered likely to be a consequence of the decreased signal to noise ratio in the flow samples relative to control samples, similar to efficiency bias in some ChIP-seq experiments. The approach chosen for normalisation of the signal, accounting for technical differences in sample quality, was a TMM approach analogous to that employed for normalisation of RNA-seq data. Under the assumption that accessibility at most peaks does not change in response to high laminar shear stress, counts were scaled to centre the distribution at a fold change of 1 (Figure 5.13). A number of peaks with low average counts showed relatively large fold changes, but visual inspection of the normalised signal at these peaks revealed only minor changes in accessibility (Figure 5.14). In order to limit the impact of these low abundance regions on differential analysis, peaks were filtered for a minimum $\log_2$ CPM of 0, corresponding to an average of ~30 counts per sample. After filtering, the remaining 133,338 peaks were tested for differential accessibility using edgeR. At a threshold FDR of 1%, 43,517 differentially accessible peaks were identified, of which 40,583 increased in accessibility with high laminar shear stress (30.4% of 133,338 filtered peaks) and 2,934

decreased in accessibility (2.2% of 133,338 filtered peaks). Accessibility increased or decreased more than 2-fold at 21,242 and 533 peaks, respectively (Figure 5.15).

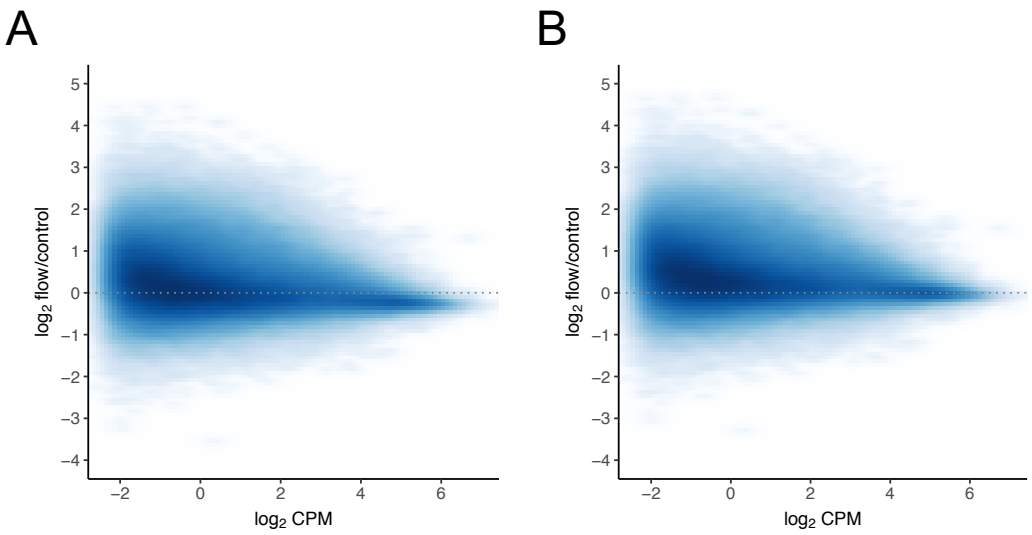

Figure 5.13: Effect of TMM normalisation on ATAC-seq signal. MA plots show the fold change in peak counts in flow samples relative to control, plotted against the average signal in counts per million mapped reads (CPM), after (A) normalising by library size, or (B) TMM normalisation. Shading corresponds to density of peaks, and the grey dotted line marks a fold change of 1 (i.e. no change in peak accessibility between conditions).

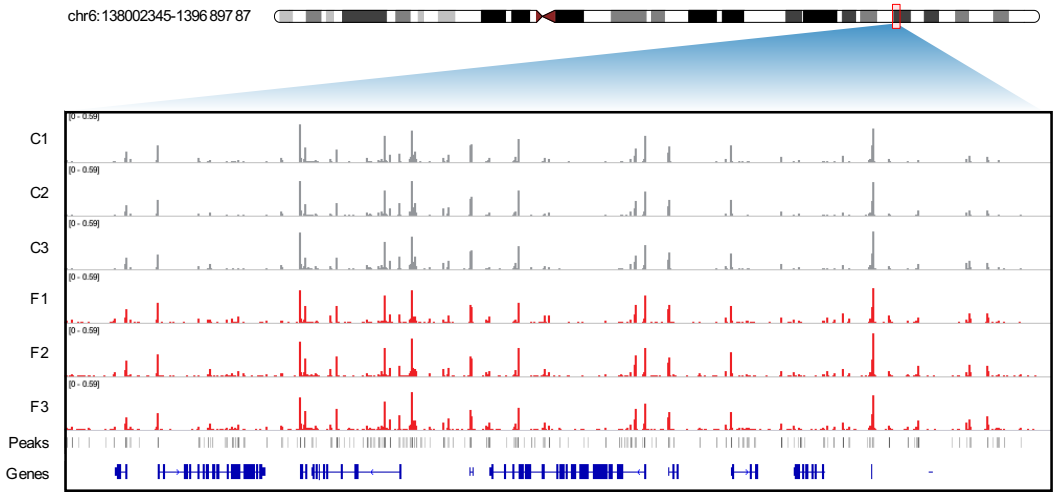

Figure 5.14: TMM-normalised insertion tracks for control and flow samples at a representative locus on chromosome 6 of the GRCh38/hg38 reference genome. Peaks called using MACS2 on merged samples are shown below insertion tracks. Gene annotations are from Refseq.

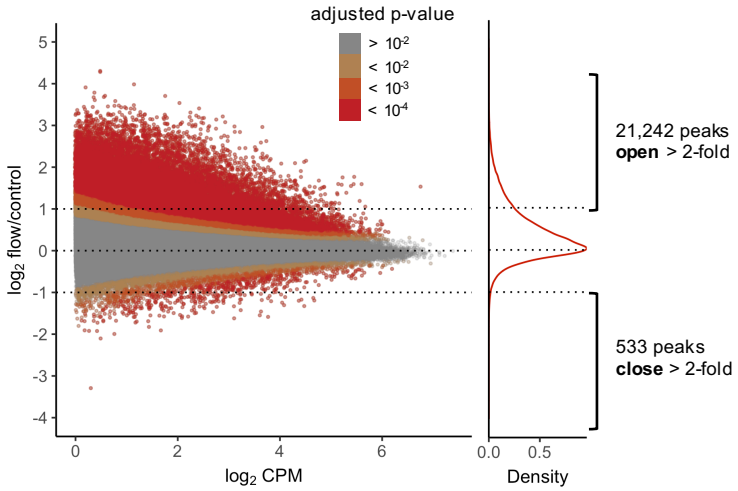

Figure 5.15: MA plot of all differentially accessible Omni-ATAC-seq peaks. Density plot shows the distribution of fold changes for the 133,338 peaks filtered for a minimum average $\log_2$ CPM of 0. P-values were adjusted using the Benjamini-Hochberg procedure.

5.11 Annotation of differentially accessible peaks

Shear stress-responsive peaks were annotated with their location relative to GENCODE v29 gene annotations, in order to determine the genomic context of accessible regions in HAECs exposed to high laminar shear stress. All peaks used for differential accessibility analysis were annotated with their genomic position, overlapping gene feature, and nearest gene. Gene promoters were defined as the region between -1000 and +200 bp around the TSS. The majority of peaks identified by Omni-ATAC-seq were located in introns or intergenic regions of the genome and were distal to gene promoters (Figure 5.16A). A greater percentage of peaks with increased accessibility were found in either introns or intergenic regions, compared with the full peak set or peaks with decreased accessibility. Promoter regions also appeared to be under-represented in the peaks with increased accessibility. To determine if differential peaks overlapped known regulatory regions of the genome, peaks were overlapped with chromatin states predicted by ENCODE ChromHMM in 9 cell types (GM12878, H1-hESC, K562, HepG2, HUVEC, HMEC, HSMM, NHEK, NHLF) (Figure 5.16B). Control peak sets were produced by

randomly shuffling the differentially accessible peaks along circularised chromosomes to maintain their order and distance, with a mask to exclude centromeric and blacklisted regions. Strong enrichment was observed for enhancer regions in the differentially accessible regions, and promoter and insulator elements were also significantly enriched ($p < 0.005$ by permutation). Given the substantial overlap with known regulatory elements, the differentially accessible peaks identified in HAECs were compared with DNase hypersensitive sites (DHS) in 52 types of primary cells, tissues and cell lines released by the Roadmap Epigenomics consortium (Figure 5.16D). More than 70% of the differentially accessible peaks overlapped with DHS in human umbilical vein endothelial cells (HUVECs), as well as primary skin and lung fibroblasts. There was also substantial overlap with DHS identified in human skeletal muscle myoblasts (HSMM) and astrocytes. Of the DHS sets tested, the least overlap was with DHS in embryonic stem cells and immune cells. The sharing of regulatory elements across the different cell types and tissues suggests common pathways likely to be involved in the HAEC response to high laminar shear stress.

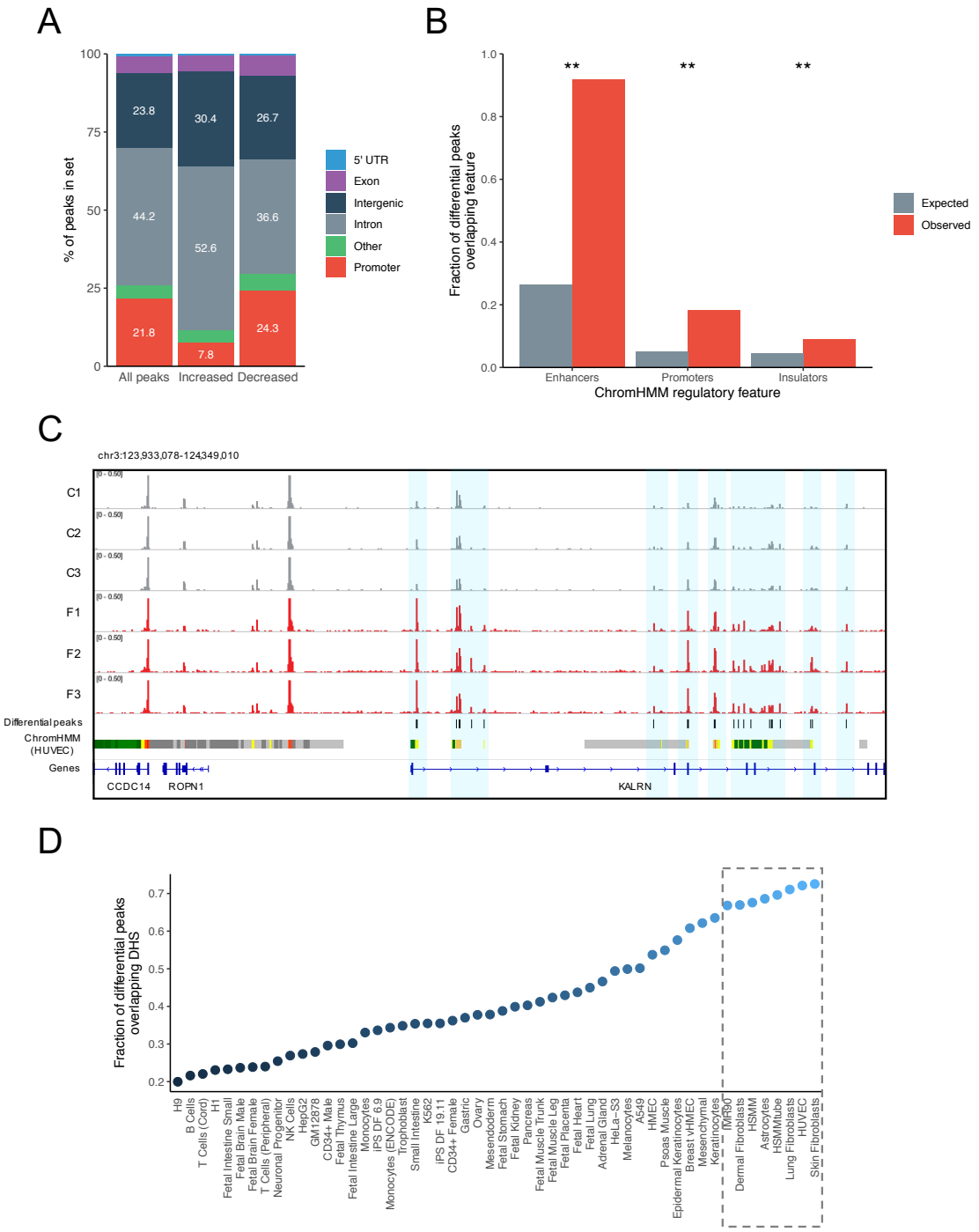

Figure 5.16: Laminar shear-responsive ATAC peaks overlap distal regulatory elements. (A) Percentage of peaks overlapping features of GENCODE v29 gene annotations. The set 'All peaks' consists of the 133,338 peaks used for differential accessibility analysis, and 'Increased' and 'Decreased' include all peaks with increased or decreased accessibility in response to high laminar shear stress, respectively (adjusted p-value < 0.01). (B) Overlap of differentially accessible peaks with chromatin states defined by ENCODE ChromHMM chromatin states for 9 cell types. Expected overlaps were calculated from 500 sets of shuffled peaks. P-values were calculated directly from permutation, ** $p < 0.005$. (C) Insertion track signal at differentially accessible peaks in the KALRN gene. Also shown are Roadmap Epigenomics ChromHMM chromatin states for HUVECs. Yellow/orange regions are enhancers, red regions are

TSSs/Promoters, green regions are transcriptionally active, and grey regions are repressed. (D) Overlap of differentially accessible peaks with Roadmap Epigenomics DNase Hypersensitive Sites (DHS) from 52 cell types/tissues.

5.12 Shear stress-induced chromatin changes are correlated with gene expression

Recent studies of chromatin accessibility have reported a correlation between accessibility to DNase or Tn5 transposase at a gene's promoter and its expression, likely reflecting increased accessibility to RNA polymerase II and general transcription factors[53,239,240]. To explore this relationship in the sequencing data from HAECs under high laminar shear stress, the complete set of accessibility peaks identified by MACS2[112] (263,484 peaks with adjusted p-value < 10^{-5}) was annotated, and normalised read counts were obtained at all peaks annotated as promoters (between -1000 bp and +200 bp of TSS). For genes with multiple promoters, the maximum normalised accessibility was used. For the 14,985 genes with accessible promoters, normalised expression in fragments per kilobase per million mapped reads (FPKM) was calculated from RNA-seq fragment counts. There was a positive correlation between promoter accessibility and gene expression (Pearson $r = 0.56$, $p < 2.2 \times 10^{-16}$), but also a number of genes with either high promoter accessibility and low expression, or high expression and low promoter accessibility (Figure 5.17A). In the former case, this is possibly a consequence of gene promoters acting as distal regulatory elements for other genes. In the latter case, this is likely due to the strict definition of promoters by proximity to the TSS. While this is generally the case, expression of some genes may be driven by promoter elements > 1000 bp upstream of the annotated TSS. Of the differentially accessible peaks identified (43,517, adjusted p-value < 0.01), 3,863 were located in gene promoters. Differentially accessible promoters were separated into those with increased or decreased accessibility

with high laminar shear stress, and the $\log_2$ fold changes in expression of the associated genes were compared (Figure 5.17B). As expected, an increase in promoter accessibility was associated with positive fold changes in gene expression, and a decrease in promoter accessibility with negative fold changes, on average ($p < 2.2 \times 10^{-16}$, Welch's t-test). Visual inspection of Tn5 insertion tracks revealed major chromatin rearrangements proximal to the TSS of some of the most up- (Figure 5.17C) and downregulated (Figure 5.17D) genes. A similar trend was observed for differential peaks distal to TSSs and overlapping ChromHMM enhancers. These peaks were associated with the nearest gene, and for genes with multiple enhancers the average fold change in accessibility was compared with the change in expression of the associated gene. On average, genes with an increase in enhancer accessibility increased in expression, and genes with decreased accessibility in enhancers decreased in expression ($p < 2.2 \times 10^{-16}$, Welch's t-test). The difference in mean $\log_2$ expression change was greater for peaks proximal to TSSs than for distal peaks (0.640 versus 0.451), which may be due to the fact that enhancer elements do not always regulate the expression of the nearest gene in terms of linear distance along a chromosome, but can act over distances of up to 500 kb, and regulatory interactions are often highly specific and cell-type dependent.

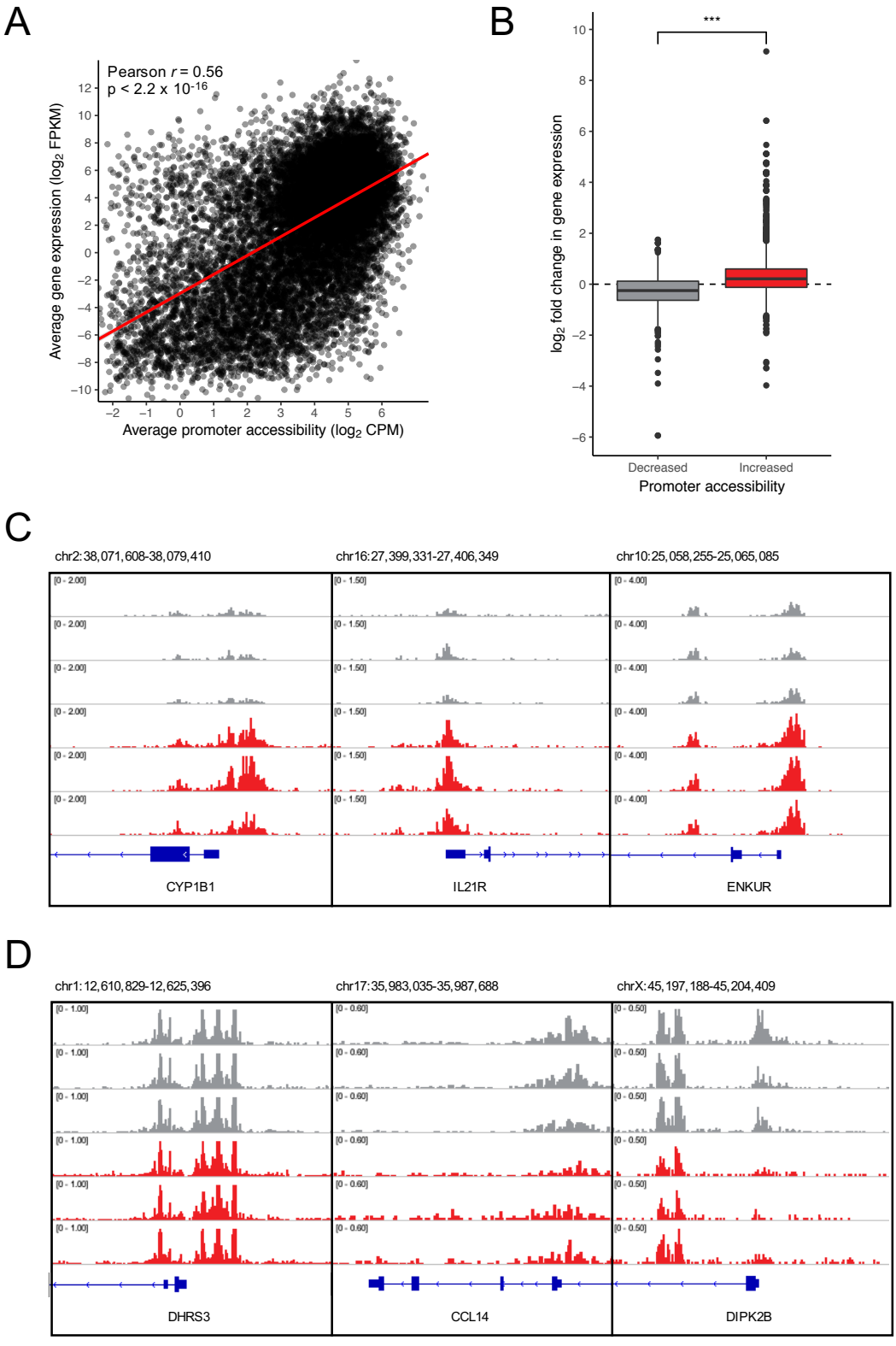

Figure 5.17: Laminar shear-induced changes in chromatin accessibility are correlated with gene expression. (A) Correlation of average promoter accessibility and average expression for 14,985 genes across all Control and Flow samples. Line shows relationship by simple linear regression. (B) Boxplots of $\log_2$ fold change in expression for genes with increased (2,145 genes, red) or

decreased (407 genes, grey) promoter accessibility. *** $p < 0.0005$, Welch's t-test. (C) and (D) Tracks of normalised Tn5 insertions at TSSs of differentially expressed genes for Control (grey) and Flow (red) samples. All tracks are plotted on the same scale for each gene.

5.13 Pathway and ontology enrichment analysis of differentially accessible peaks

Given the correlation between peak accessibility and expression of nearby genes, ontology enrichment analysis was performed based on the list of genes associated with differentially accessible peaks. Terms from the 'GO: Biological Process' database were tested for overrepresentation in the peak-associated genes and resulting p-values were corrected for multiple testing (Benjamini-Hochberg). Given the large number of genes used for overrepresentation analysis, terms with > 1000 genes in the annotation set were filtered out to remove uninformative, high-level ontologies such as 'biological process' and 'response to stimulus'. The remaining terms included several relating to cell motility, cytoskeletal rearrangement and signaling, which were also enriched for the differentially expressed genes identified by RNA-seq at a threshold FDR of 5%, as well as terms relating to vasculature and circulatory system development (Figure 5.18). Interestingly, terms relating to neuronal development were also enriched for the genes associated with differential peaks, possibly due to the degree of overlap between differentially accessible peaks in HAECs exposed to high laminar shear stress and DHS identified by the Roadmap Epigenomics consortium in astrocytes (Figure 5.16D). In the KEGG pathway database, strong enrichment was observed for the 'Pathways in cancer' term, which was also the most enriched pathway for upregulated genes. Additionally, there was enrichment of a number of signaling pathways with known roles in mediating angiogenesis, proliferation and response to shear stress in endothelial cells, such as signaling via GPCRs, the phosphatidylinositol 3-kinase (PI3K)-Akt pathway, MAPK

signaling, the Hippo and Wnt pathways, and TGF-β signaling[241] (Figure 5.19). Terms relating to processes involved in the development of cardiovascular disease, such as transendothelial migration of leukocytes, were also significantly enriched. Similar results were obtained using Reactome pathway annotation sets, with the additional enrichment of terms relating to growth factor signaling (specifically VEGF, PDGF and NGF), platelet activation, ion transport, and immune processes such as interleukin signaling (Figure 5.20). Overall, these results show concordance with the pathways enriched for differentially expressed genes and reinforce the association of chromatin accessibility with gene expression changes in HAECs exposed to high laminar shear stress. Furthermore, they highlight signaling cascades and biological pathways crucial to the transduction of mechanical signal at the surface of endothelial cells.

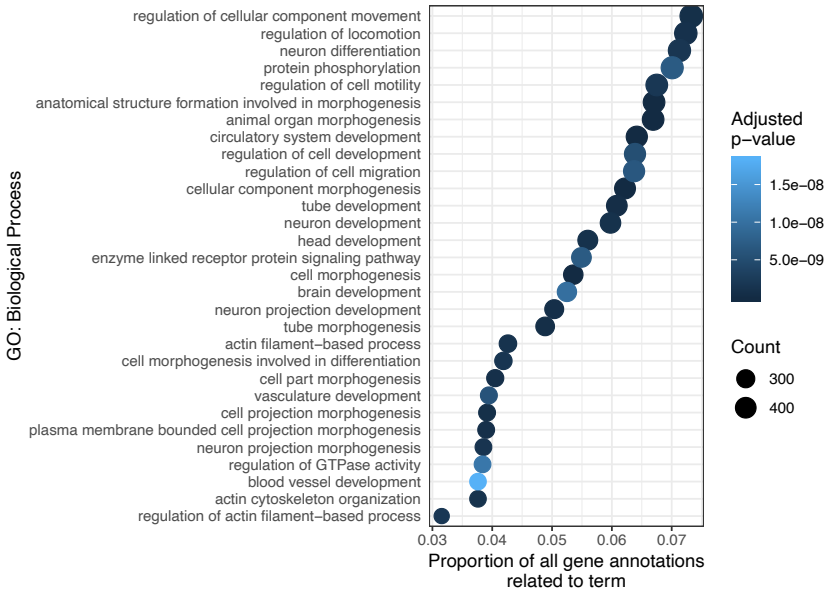

Figure 5.18: Top 30 enriched GO: Biological Process terms in the gene annotations for all differentially accessible peaks (adjusted p-value < 0.01). Terms were sorted by adjusted p-value. Each point represents a term, the size of the point is proportional to the number of peak-associated genes linked to the term, and the colour represents the adjusted p-value of enrichment. The x-axis shows the fraction of all peak-associated genes that relate to the term.

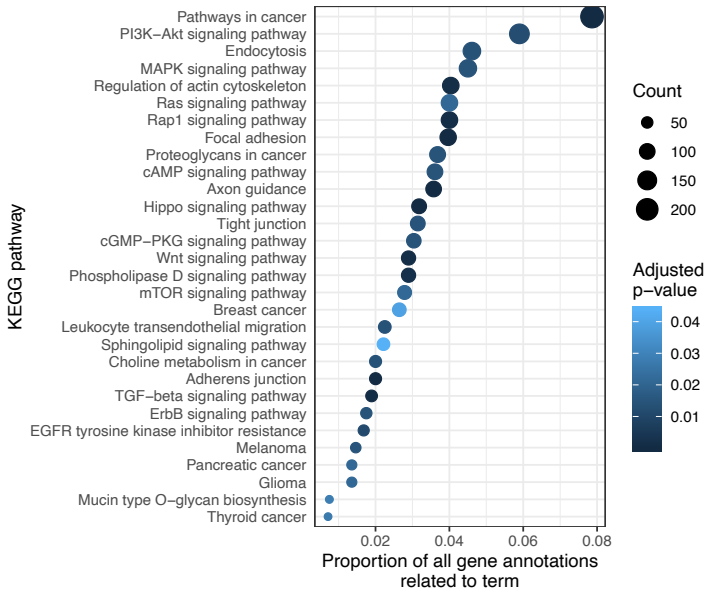

Figure 5.19: Top 30 enriched KEGG pathways in the gene annotations for all differentially accessible peaks (adjusted p-value < 0.01). Terms were sorted by adjusted p-value. Each point represents a term, the size of the point is proportional to the number of peak-associated genes linked to the term, and the colour represents the adjusted p-value of enrichment. The x-axis shows the fraction of all peak-associated genes that relate to the term.

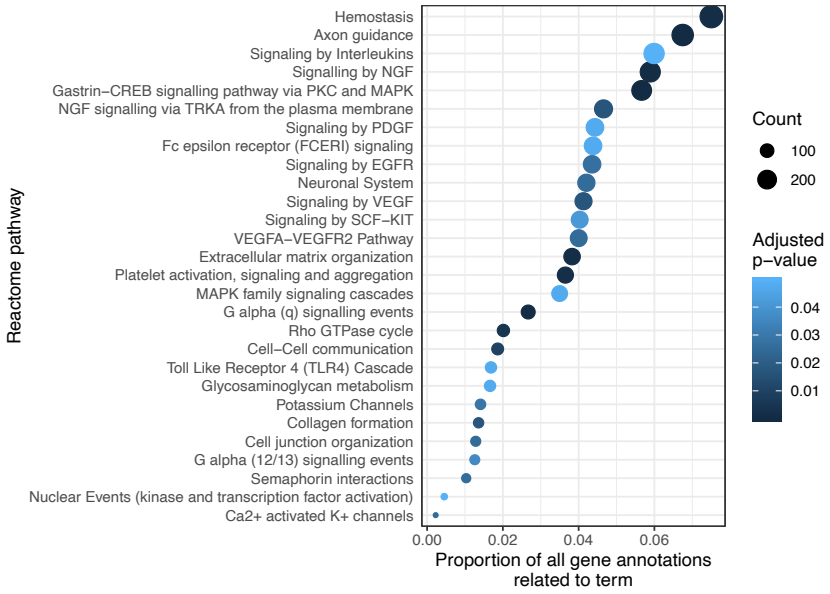

Figure 5.20: Top 30 enriched Reactome pathways in the gene annotations for all differentially accessible peaks (adjusted p-value < 0.01). Terms were sorted by adjusted p-value. Each point represents a term, the size of the point is proportional to the number of peak-associated genes linked to the term, and the colour represents the adjusted p-value of enrichment. The x-axis shows the fraction of all peak-associated genes that relate to the term.

5.14 Inference of transcription factor activity from differential accessibility

In order to identify potential molecular mediators of the chromatin accessibility changes induced by high laminar shear stress, differentially accessible peaks were tested for enrichment of transcription factor binding motifs using HOMER[118]. Peaks with no significant alteration of chromatin accessibility under high laminar shear stress (i.e. adjusted p-value > 0.01 and absolute $\log_2$ fold change < 1) were used as the genomic background to restrict the analysis to factors binding in dynamic regions only. Testing for the presence of 'known' motifs, derived from ChIP-seq datasets, in peaks with increased accessibility under high laminar shear stress revealed strong enrichment of motifs similar to that bound by the AP-1 transcription factor, as well as enrichment of motifs bound by SRY-related HMG-box (SOX) factors, the TGF-β induced factor homeobox 1 (Tgif1), Krüppel-like family (KLF) factors and the ETS family transcription factor ERG (Figure 5.21A). HOMER also identifies *de novo* motifs based on enriched sequences within the peaks relative to the background, and the results for the peaks with increased accessibility supported the results of the known motif enrichment analysis, with strong enrichment of the core AP-1 motif, and enrichment of motifs highly similar to SOX and KLF binding motifs (Figure 5.21B). Motif positions in each peak were determined using HOMER, and for the most enriched motifs the majority of peaks contained motif instances at the centre of the peak, particularly for AP-1 and ERG, indicating that transcription factor binding at these positions coincides with the position of greatest accessibility change. The same analysis was performed using peaks with decreased accessibility under high laminar shear stress, and for known motifs the strongest enrichment was observed for GATA family transcription factors, followed by

motifs for retinoic acid receptor (RAR)-α and nuclear transcription factor Y (NFY) (Figure 5.22A). The density of GATA and NFY motifs was greatest at the centre of peaks with decreased accessibility, but the RAR-α motif showed no clear positional preference.

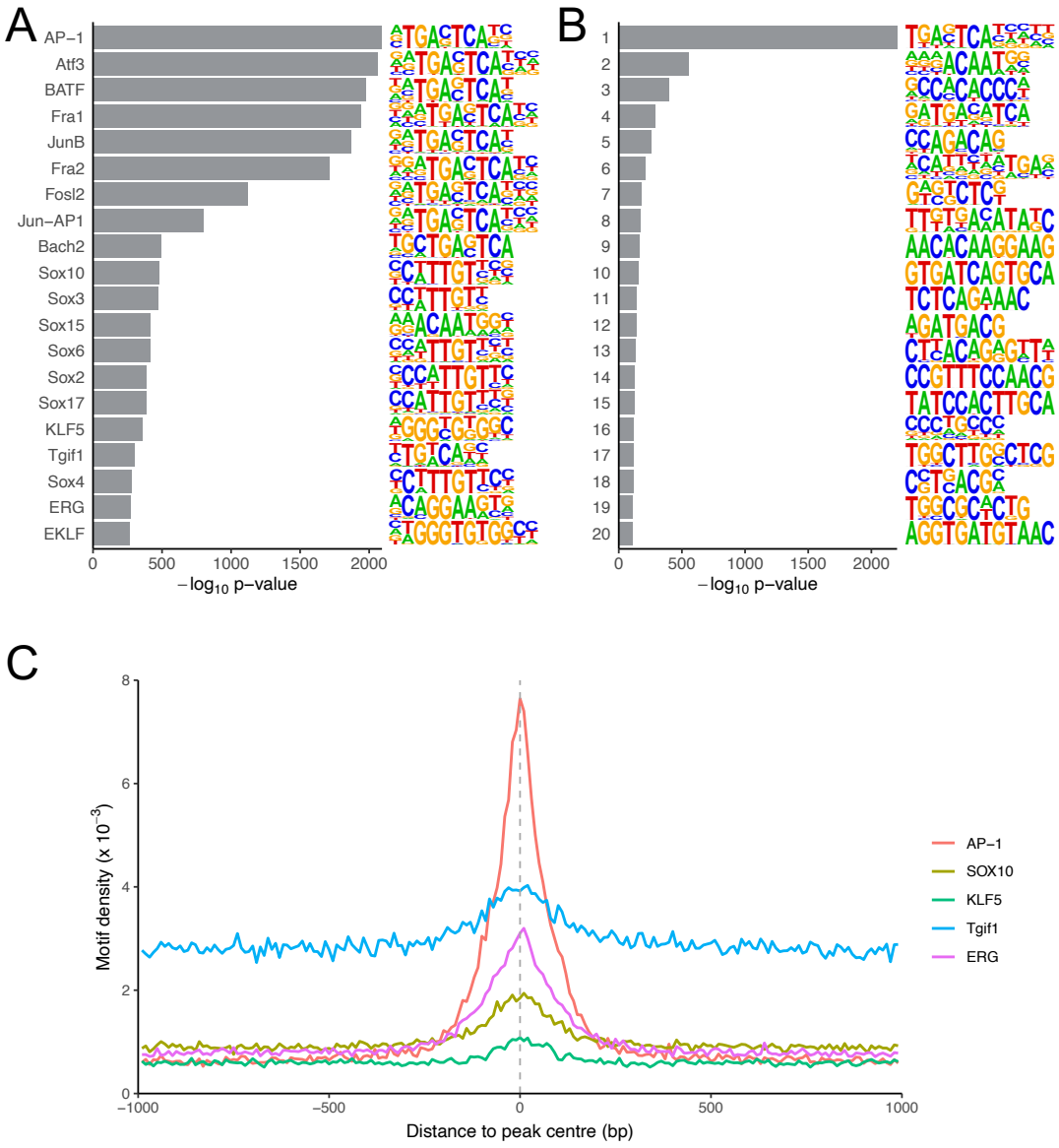

Figure 5.21: Motif enrichment in peaks with increased accessibility. Top 20 enriched known (A) and de novo (B) transcription factor binding motifs in peaks with increased accessibility (C) Density of top 5 enriched motifs after removal of similar motifs, relative to the centre of each peak. Motif density is expressed as motifs per bp per peak.

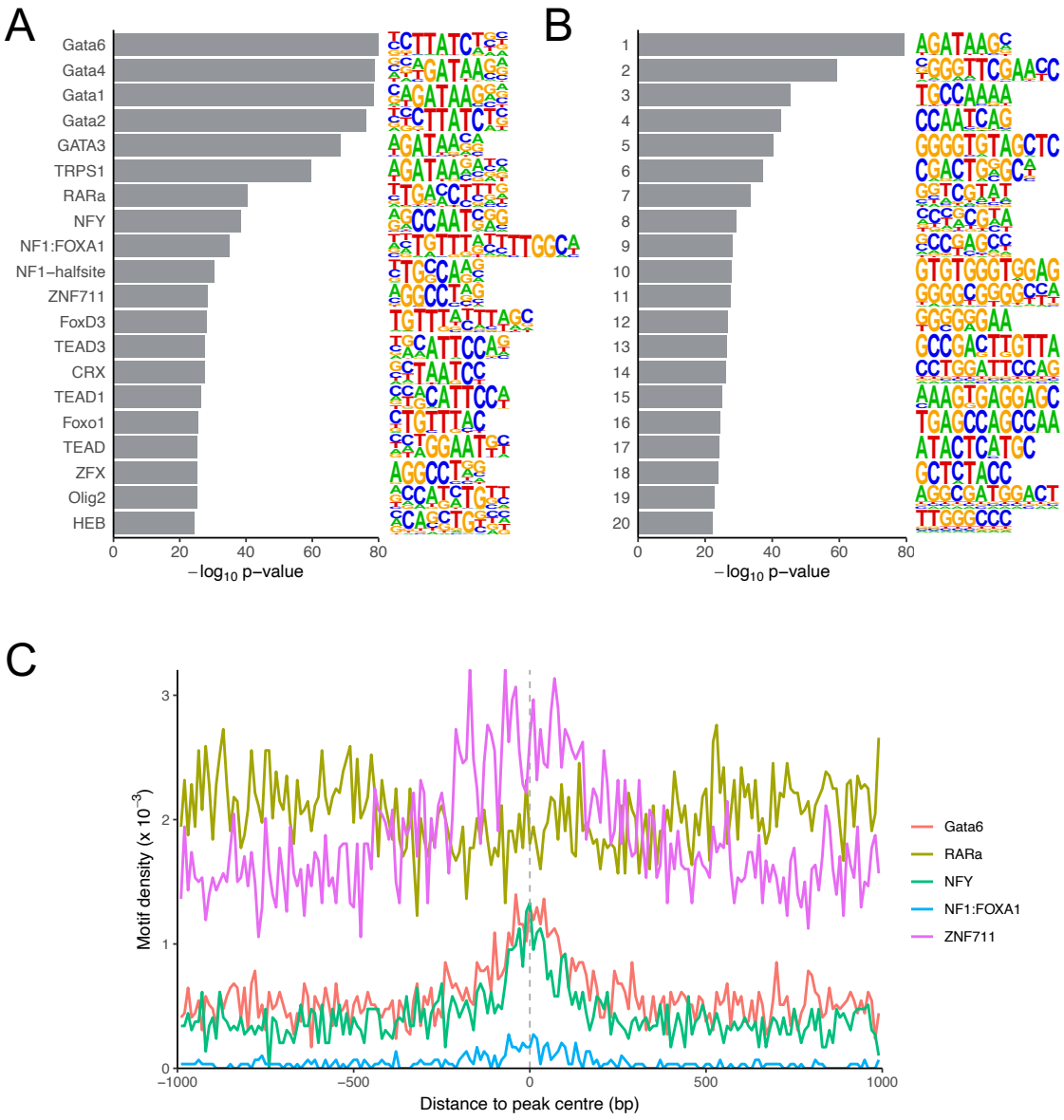

Figure 5.22: Motif enrichment in peaks with decreased accessibility. Top 20 enriched known (A) and de novo (B) transcription factor binding motifs in peaks with decreased accessibility. (C) Density of top 5 enriched motifs after removal of similar motifs, relative to the centre of each peak. Motif density is expressed as motifs per bp per peak.

To complement the motif enrichment analysis for differentially accessible peaks, the chromatin accessibility at motifs in all accessible regions of the genome was compared between control and flow HAECs using the R package chromVAR[119]. This package aggregates motifs within peaks and computes the deviation of chromatin accessibility from the expected fragment count based on the average across all samples. This is

corrected for known ATAC-seq technical biases, such as variability in signal to noise ratio and GC content, by comparing the accessibility deviation at peaks containing the motif of interest to a set of background peaks matched for GC content and average accessibility. Bias-corrected deviations and z-scores are calculated for each motif, allowing identification of factors potentially involved in mediating chromatin accessibility changes and hence influencing gene expression. All accessible regions identified by MACS2 (263,484 peaks with adjusted p-value < 10^{-5}) were resized to fixed-width 500 bp intervals and centred on the peak summit for use as input to chromVAR, along with aligned Omni-ATAC-seq fragments for HAECs in the control and flow treatment groups and position weight matrices (PWMs) for HOMER known motifs. The greatest variability in chromatin accessibility was observed at motifs bound by AP-1-related factors (e.g. Fra2, Fosl2, Atf3), as expected from motif enrichment analysis using HOMER (Figure 5.23A). There was also high variability at motifs bound by the transcriptional repressor and insulator factor CTCF and the highly similar BORIS (CTCFL). Other motifs with high variability included those bound by previously identified regulators of the endothelial response to shear stress, such as Nrf2[242], as well as motifs identified as enriched in differentially accessible peaks using HOMER, such as ERG, Sox15 and EKLF[41].

Bias-corrected deviation z-scores for all motifs with accessibility variance > 10 were visualised, and hierarchical clustering of z-scores across all samples identified 3 clusters of motifs with similar patterns of accessibility changes between treatment groups (Figure 5.23B). One of the clusters contained motifs at which accessibility decreased in the flow group versus control, such as motifs for GATA factors, as well as motifs for CTCF, NFY and E2F-family transcription factors. Another cluster contained motifs with relatively strong increases in accessibility in flow samples, and represented basic leucine zipper

(bZIP) transcription factors binding to motifs with high similarity to the AP-1 core motif TGA(C/G)TCA, including the transcriptional repressors MafK and Bach1/2. The remaining cluster included motifs bound by a number of different transcription factor families and all exhibited increased accessibility in the flow group relative to control. Among the represented transcription factor families in this cluster were SOX and KLF factors, ETS-family factors (including ERG), motifs for CREB/ATF family factors such as ATF2 and ATF7, SMAD proteins, and Runt-related factors. The clustering of motifs was recapitulated by t-Distributed Stochastic Neighbouring Embedding (t-SNE) of the bias-corrected deviations (Figure 5.23C). In order to confirm protein binding at differentially accessible motifs, transcription factor footprinting was explored. Tn5 insertions were pooled for samples in the control and flow treatment groups separately, and clear footprints were observed for aggregated CTCF (Figure 5.23D) and AP-1 (Figure 5.23E) motifs across the genome, suggesting protection of these sites from cleavage by the Tn5 transposase. The insertion signal for each treatment group was normalised by adjusting the insertion counts so that the values 200-500 bp away from the motif centre were equal. There was no difference in the average normalised insertion frequency at either motif between the control and flow treatment groups, but insertion frequency at the motif flanks differed in the expected direction (i.e. increased accessibility flanking AP-1 motifs and decreased accessibility flanking CTCF motifs in flow cells relative to control cells). The differences were minor, possibly due to only a subset of all accessible motifs displaying large changes in accessibility between treatment groups, which is masked by the averaged data.

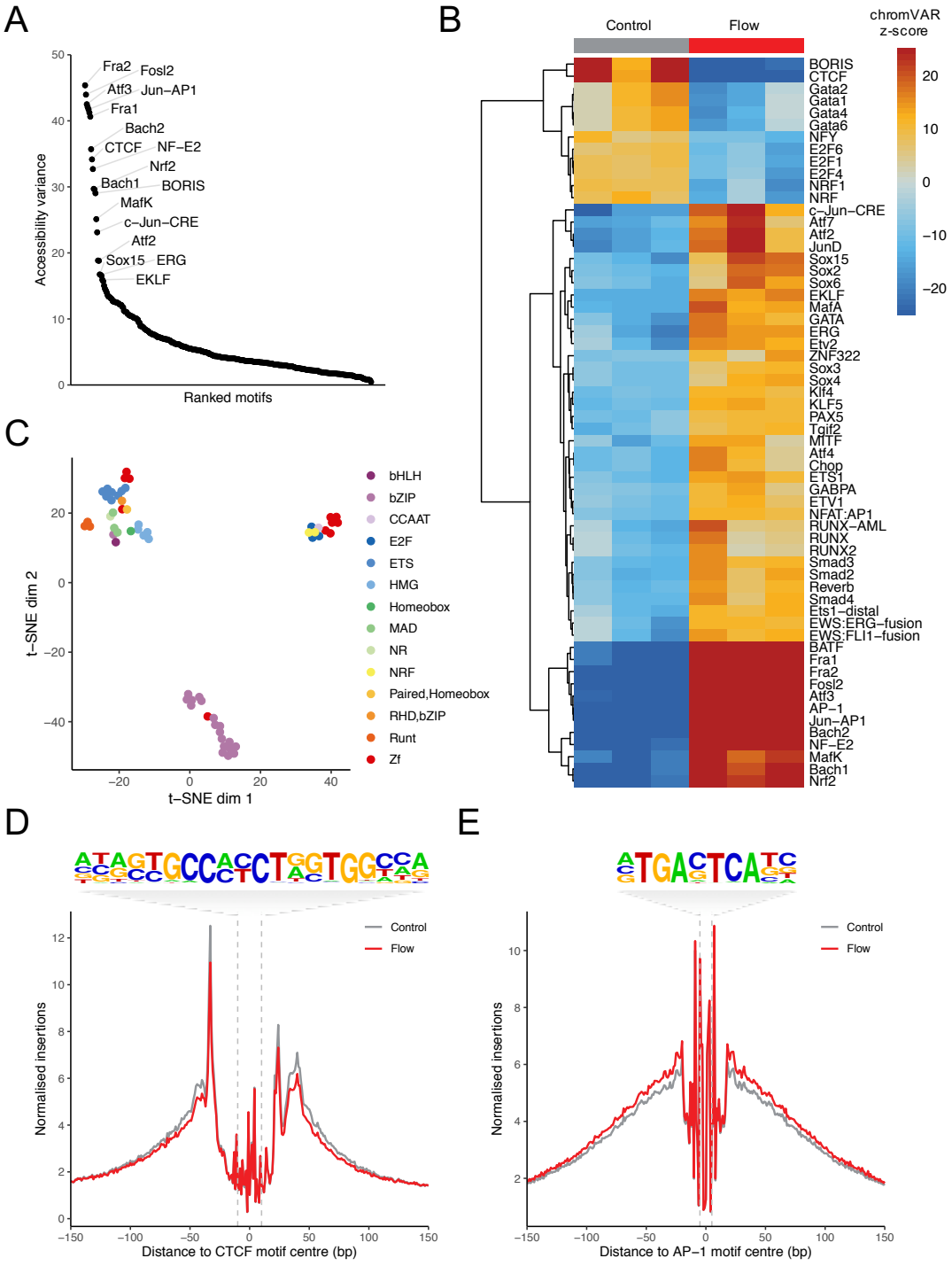

Figure 5.23: chromVAR analysis of accessibility at transcription factor binding motifs. (A) Variability in chromatin accessibility at ranked HOMER motifs. (B) Heatmap of chromVAR deviation z-scores for motifs (rows) with accessibility variance > 10 in control and flow samples (columns). Rows are grouped by unsupervised hierarchical clustering based on Euclidean distance between the z-scores. (C) tSNE plot based on bias-corrected deviations of motifs, coloured by transcription factor family. (D and E) Plots showing normalised Tn5 insertions at aggregated CTCF and AP-1 motifs in accessible regions across the genome.

A number of transcription factors represented by the enriched and differentially accessible motifs showed large changes in gene expression with high laminar shear stress (Figure 5.24), particularly bZIP factors such as FOSB and ATF3, and the Krüppel-like factors KLF2 and KLF4, which are known to be upregulated in endothelial cells by laminar shear stress.

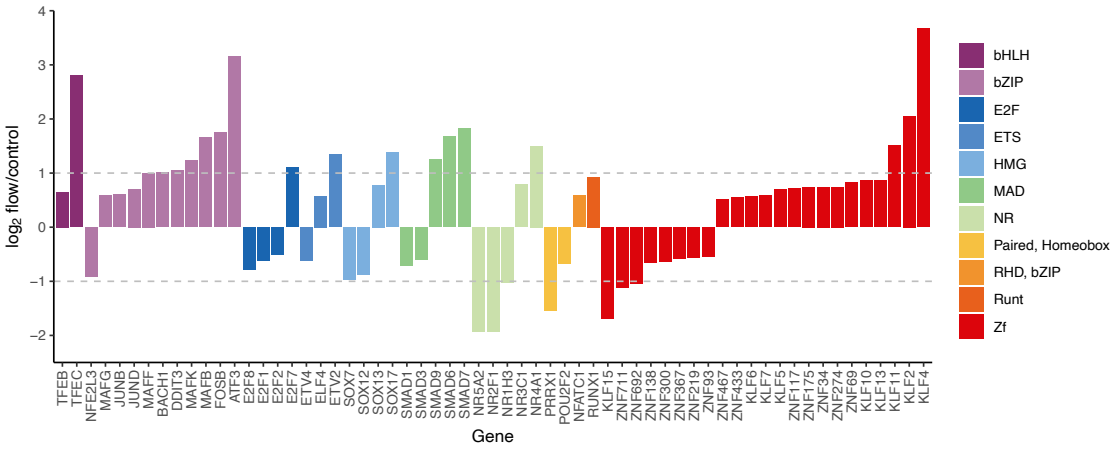

Figure 5.24: Transcription factor genes regulated by shear stress. Plot shows $\log_2$ fold changes in gene expression between control and flow treatment groups for genes encoding transcription factors, as determined by RNA-seq (absolute $\log_2$ fold change > 0.5 and adjusted p-value < 0.05).

The chromVAR package enables calculation of the 'synergy' between motifs with variable accessibility, which can indicate cooperative or competitive binding at these motifs. Synergy is defined as the excess variability in chromatin accessibility for peaks containing a pair of motifs compared to a random sub-sample of peaks containing only the motif with the greatest variability. Given that many of the motifs with variable accessibility can be bound by numerous transcription factors, a set of motifs were selected to represent the motif families with the greatest variability in chromatin accessibility between control and flow samples (Figure 5.25A). Of these, the greatest synergy was observed between the EKLF motif and several other motifs with increased accessibility in flow samples, namely Sox15, MafK and AP-1, and average accessibility in flow samples was greater for peaks with both motifs compared to peaks with AP-1 motifs only

(Figure 5.25B). Accessibility at the CTCF motif was strongly opposed to EKLF and Atf2 motifs, but not AP-1, and the motif with the strongest opposition to AP-1 was Gata1. Hierarchical clustering of motifs identified two major clusters of synergistic motifs, with one containing motifs with increased accessibility in flow samples and the other motifs with decreased accessibility, and synergy scores differed within each cluster identifying pairs of motifs with a greater ability to alter chromatin accessibility than either motif individually.

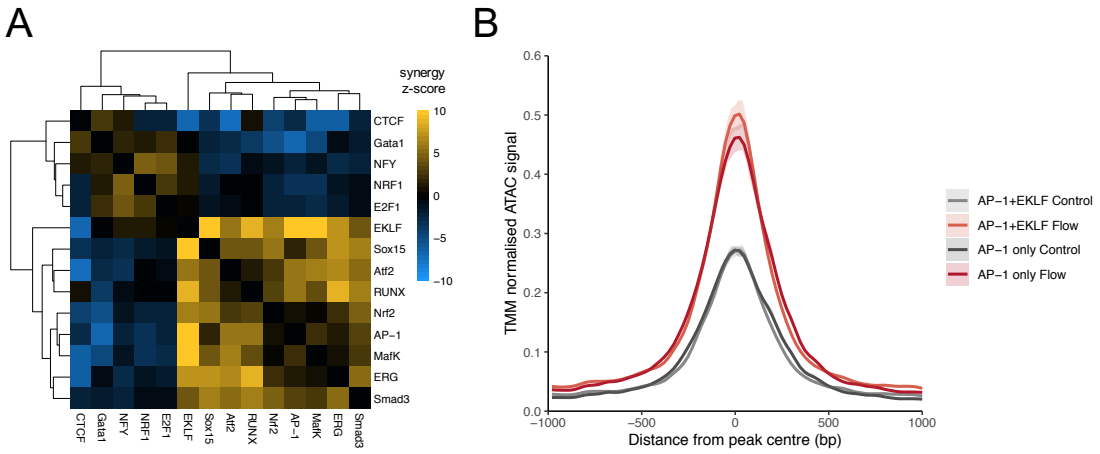

Figure 5.25: Inference of transcription factor 'synergy' with chromVAR. (A) Z-scores of chromVAR 'synergy' calculated for select representative motifs. Motifs are clustered based on Euclidean distance between z-scores of all comparisons for each motif. (B) Average TMM normalised ATAC signal for control and flow samples in peaks containing at least one AP-1 and one EKLF motif, compared with peaks containing AP-1 motifs only. The shaded area marks the mean ± 1 standard deviation.

5.15 Shear stress-responsive elements are enriched for disease-associated genetic variants

The co-localisation of disease-associated SNPs and differentially accessible genomic regions was explored to identify variants with the potential to alter the activity of shear-stress responsive regulatory elements, and hence transcription of nearby genes. SNPs in the NHGRI-EBI GWAS catalog reaching genome-wide significance ($p < 5 \times 10^{-8}$) for

association with coronary artery disease (CAD) were pruned to create a list of independent SNPs, defined as those with low linkage disequilibrium (LD) ($r^2 < 0.1$ in 1000G phase 3 EUR samples) and more than 500 kb from another CAD-associated SNP. This list of independent index SNPs was then used to identify all SNPs in high LD ($r^2 > 0.8$ in 1000G phase 3 EUR samples) to include all potential causal variants in further analyses. 6,206 CAD-associated SNPs were identified using this method, of which 74 overlapped the differentially accessible peaks in HAECs exposed to high laminar shear stress. To test the significance of this overlap relative to background expectation, 10,000 sets of random SNPs from 1000G phase 3 EUR samples, matched for minor allele frequency (MAF), gene density and number of other SNPs in LD ($r^2 > 0.8$), were overlapped with differentially accessible peaks, thereby generating a null distribution. Overlap of CAD SNPs with differentially accessible peaks was compared to this distribution using a binomial test and the resulting p-value was 3.1×10^{-6}, indicating that the set of CAD-associated SNPs showed significantly greater overlap than expected at a threshold of $p = 0.05$ (Figure 5.26A). This process was repeated for sets of SNPs associated with other traits listed in the GWAS catalog, and a number of these showed strong enrichment, including those associated with known risk factors for cardiovascular disease such as blood pressure, HDL cholesterol and triglyceride levels (Figure 5.26B).

CAD SNPs within differentially accessible peaks were overlapped with transcription factor binding motifs located within accessible regions, previously identified during chromVAR analysis, and several SNPs were found to alter the core sequences identified as likely transcription factor binding sites. One of these SNPs, rs17608766, altered the EKLF and KLF4 known motifs in the HOMER database (Figure 5.26C), which was confirmed using the SNP2TFBS tool (https://ccg.epfl.ch/snp2tfbs/snpselect.php). This tool identifies transcription factor binding motifs based on position weight matrices

(PWMs) from the Jaspar Core 2014 Vertebrate library and scores the motifs for the different alleles against the PWM. For rs17608766, the two motifs identified were Klf1 and Klf4, supporting the results of the HOMER motif matching by chromVAR, and scores for both motifs were reduced below the minimum threshold for the minor allele, indicating loss of binding affinity. The KLF motif was located in an intron of the GOSR2 gene, and the site showed H3K27Ac signal in ENCODE data from a number of cell lines, as well as a positive PhyloP conservation score, suggesting this region may act as an enhancer. Of the genes surrounding rs17608766 (± 500 kb), GOSR2 was significantly upregulated by high laminar shear stress, and CDC27 and ITGB3 were significantly downregulated (adjusted p-value < 0.05), which may be related to the increased accessibility and transcription factor activity at this locus, and present potential mechanisms for the disease risk associated with rs17608766.

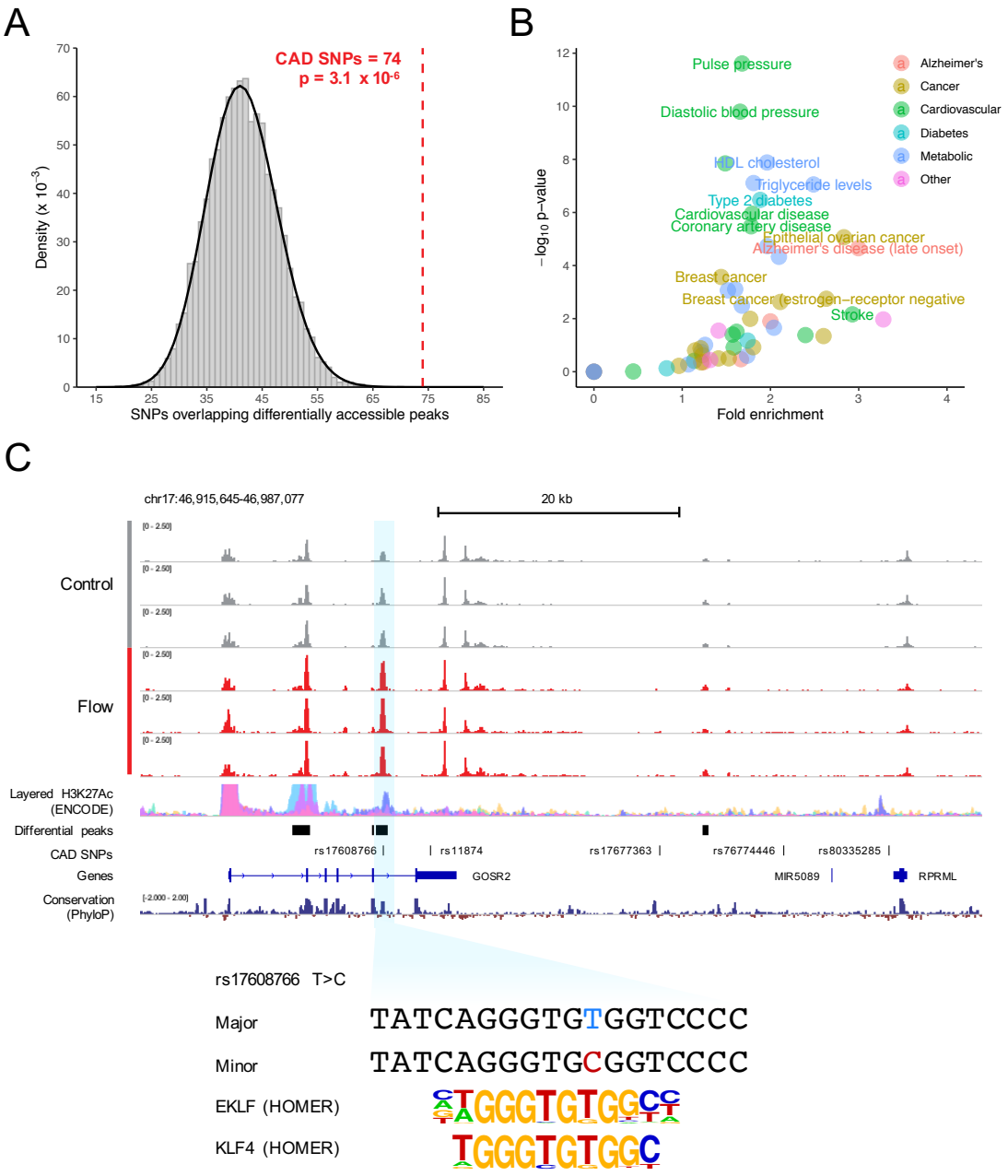

Figure 5.26: Overlap of GWAS SNPs with differentially accessible peaks. (A) Histogram showing the distribution of overlaps for 10,000 sets of control SNPs matched for MAF, gene density and 'LD buddies'. Overlaid is the binomial distribution used to model the background (black line), and the red dashed line marks the number of CAD-associated SNPs overlapping differentially accessible peaks. (B) Results of enrichment tests for 54 traits in the GWAS catalog related to cardiometabolic diseases, Alzheimer's, cancer, and traits related to endothelial cell function (e.g. Vascular endothelial growth factor levels). (C) Example of overlapping CAD-associated SNP, potentially altering gene regulation by disrupting a transcription factor binding motif. Also shown are other SNPs in high LD ($r^2 > 0.8$) with the index SNP rs17608766.

5.16 Discussion

The influence of haemodynamic forces on atheroma localisation is widely accepted, and studies of endothelial cells have reported atheroprotective transcriptomic profiles under conditions of laminar shear stress[34,37]. Molecular mediators of endothelial mechanotransduction have been identified, but studies have often focused on individual genes or specific regions of the genome, rather than genome-wide regulatory networks[40,234]. The work described in this chapter represents the first genome-wide interrogation of chromatin accessibility in human aortic endothelial cells cultured under high laminar shear stress and compares the resulting profile with cells cultured under static conditions. It also integrates transcriptomic profiles generated under the same conditions, and transcription factor binding motifs from public databases, to infer regulatory mechanisms relevant to endothelial cell biology. Furthermore, the overlap between genomic regions with dynamic accessibility and CAD-associated genetic variants identified by GWAS reveals potentially causal mechanisms of gene regulation, generating hypotheses for the functional consequences of non-coding GWAS SNPs, which largely remain unexplained.

Some of the previously studied *in vitro* models of shear stressed endothelial cells have utilised cone-and-plate apparatus to generate shear stress by rotation of fluid around a circular culture vessel[235]. Crucially, the resulting shear stress profile is non-uniform and forms a gradient along the radius of the culture surface. Precise definition of unidirectional laminar shear stress is therefore challenging using such a model. In addition, studies have attempted to generalise endothelial cell responses to shear stress using human umbilical vein endothelial cells (HUVECs)[234]. Several studies have demonstrated the heterogeneity of endothelial cells originating from different vessels,

and atheromatous plaque does not develop in veins, therefore the response of these cells to different shear stresses may not influence the progression of atherosclerosis *in vivo*[243,244]. The experiments detailed in this chapter employed a parallel-plate flow chamber to apply unidirectional laminar shear stress to HAECs, the design of which allowed precise definition of the applied shear stress in dynes/cm^2, simply by adjusting the flow rate of perfused culture medium. Primary aortic endothelial cells from human donors are derived from a large, atheroprone vessel, and as such the observed responses to laminar shear stress in these cells are directly relevant to human atherosclerotic disease. Similar experiments in other arterial endothelial cells, such as those derived from the coronary artery, may provide deeper insights into the genetic mechanisms underlying cardiovascular disease, as well as the contextual heterogeneity of endothelial cells at an epigenomic level.

Few of the previously described studies of the shear-stressed endothelial transcriptome have utilised RNA-seq, with many depending on microarray technology to identify differentially expressed transcripts using a restricted set of probes targeted at known transcript sequences[42,245,246]. RNA-seq has several advantages over the microarray, such as the acquisition of truly genome-wide information not limited by hybridisation to a chosen set of probes. Transcripts mapping outside of known gene annotations can also be discovered using RNA-seq, with novel transcript discovery limited only by the choice of library preparation and the reference genome. Other advantages of RNA-seq over microarray are the higher signal-to-noise ratio due to the absence of cross- or non-ideal-hybridisation kinetics, and the greater dynamic range in detectable expression levels, with no theoretical upper limit (but practically limited by sequencing depth). For these reasons, as well as the significant reduction in the cost of high-throughput sequencing over the last two decades, RNA-seq has been adopted as the method of choice for generating

transcriptomic profiles of cells and tissues. The RNA-seq data generated during this study identified 6,982 genes with significant differences in expression following exposure of HAECs to high laminar shear stress. Of these differentially expressed genes, 943 had a greater than two-fold change in either direction and a number of the changes in gene expression were concordant with results reported previously in similar experiments[92,235]. Among the most differentially expressed genes were several with known involvement in the development of atherosclerosis. For example, the *PTGS2* gene encoding the enzyme COX-2 was upregulated more than ten-fold by laminar shear stress. The increased risk of myocardial infarction or stroke with COX-2 inhibition is well known, most likely due to the reduced production of cardioprotective prostaglandins such as prostacyclin[247]. *CXCL8*, encoding interleukin-8 which is chemotactic for monocytes and T lymphocytes, was downregulated by high laminar shear stress, as were a number of other cytokines, suggesting that increased production of these factors at sites of low or oscillatory shear stress may contribute to increased leukocyte recruitment. Another example is *ANGPTL4*, which was upregulated by high laminar shear stress, and overexpression of this gene was found to suppress oxLDL uptake by macrophages, thus reducing foam cell formation[248].

Analysis of the gene ontology and pathway annotations for the differentially expressed genes identified enriched terms relevant to the biology of endothelial cells under laminar shear stress, which for upregulated genes related to signaling, cell motility, cytoskeletal rearrangement, extracellular matrix organisation, and ion transporter activity. An array of signaling pathways are known to be activated through stimulation of various mechanoreceptors at the endothelial cell surface, which include ion channels and GPCRs[36,233,241]. Phospholipase C-dependent increases in intracellular calcium concentration is well known in endothelial cells under flow conditions and signaling via G proteins has been implicated in this response[249]. A recently identified mechanosensitive

GPCR[38], encoded by *GPR68*, was upregulated by high laminar shear stress in this study, and this molecule may contribute to the observed effect on transmembrane ion transport and resulting gene expression changes in endothelial cells under shear stress. *RASD1,* encoding a Ras-family small GTPase which functions as a nucleotide exchange factor for heterotrimeric G proteins, was also among the genes most upregulated by laminar shear stress, and variants at this locus have been associated with CAD[250,251]. Rho family small GTPases, such as RhoA, RhoB, Rac and CDC42, have been shown to regulate the reorganisation of the actin cytoskeleton to alter cell shape and motility[252], and genes for all of these proteins were also upregulated by shear stress in HAECs. A mechanosensory complex at inter-endothelial cell junctions comprising PECAM-1, VE-cadherin and VEGFR2 has also been reported to mediate the observed effects of shear stress on cell alignment and gene expression[40], and the genes expressing each of these proteins were upregulated by shear stress.

Downregulated genes showed strong enrichment for terms relating to RNA processing, protein synthesis, DNA replication and cell proliferation. A number of known markers of cell proliferation were downregulated by shear stress, including *E2F1, PCNA, MCM2, MCM3, MCM4, MCM5, MCM6, MKI67* and *TOP2A*. Genes expressing cyclins, cyclin-dependent kinases and their inhibitors also showed expression changes consistent with reduced proliferation. The cyclin B genes *CCNB1* and *CCNB2*, required for transition from G2 to M phase of the cell cycle, were downregulated by shear stress. *CCND1* and *CCND3,* encoding homologues of cyclin D, were upregulated by shear stress, along with *CDK6*. These proteins regulate the transition through G1 phase, but G1/S transition is blocked by increased expression of *CDKN1A*, encoding the cyclin-dependent kinase inhibitor $p21^{Cip1}$, which was upregulated more than two-fold by laminar shear stress. There is evidence that increased $p21^{Cip1}$ expression causes cells to enter a G0/quiescent

state, and these expression changes may explain the observed reduction of proliferation in endothelial cells under flow conditions[253]. Increased proliferation at sites of low or oscillatory shear stress may disrupt endothelial integrity and increase permeability of inter-endothelial junctions to infiltrating immune cells or LDL particles. Overall, the RNA-seq data described in this chapter reveals biologically relevant signatures of the transcriptomic response to shear stress, and highlights atheroprotective pathways in primary HAECs.

Widespread adoption of ATAC-seq as the method of choice for assaying regions of accessible chromatin genome-wide, as an alternative to DNase-seq and FAIRE-seq, has led to the identification of technical concerns over the original protocol as published[107]. Chief among these is the issue of mitochondrial contamination of prepared nuclei, resulting in a significant proportion of sequenced reads mapping to the mitochondrial chromosome, and diminishing the signal derived from the nuclear genome. Several recently published protocols have addressed this issue directly, and one such protocol known as Omni-ATAC-seq[108] almost entirely mitigated the mitochondrial contamination of samples prepared during this study. Despite this improvement, fragment duplication during PCR amplification resulted in a large number of reads being discarded, which is another known issue of ATAC-seq library preparation. Further refinements of ATAC-seq library preparation, as well as advances in the sensitivity of sequencing platforms, may limit the need for PCR amplification and enhance the quality of data generated with these methods. The improved protocol generated high quality genome-wide chromatin accessibility data sensitive to fine details of chromatin structure. Calculation of both TSS enrichment and FRiP scores informed the selection of the normalisation strategy, as reduced signal-to-noise in the flow samples resulted in global depression of ATAC signal relative to control. The observed difference in sample quality between treatment groups

was likely due to differing transposition efficiencies, possibly resulting from slight differences in cell counts. Given these observations, and the greater likelihood of local alterations of chromatin accessibility by DNA binding proteins such as transcription factors, TMM normalisation was used to adjust the counts per peak from each sample and differentially accessible regions between control and flow samples were identified using statistical methods developed specifically for count-based high-throughput sequencing data.

Many of the accessible regions tested for differential accessibility between control and flow samples were more open in flow samples (30.4%), and a much smaller proportion were more open in control samples (2.2%), which possibly reflects the increased activity of the multiple signaling cascades involved in mechanotransduction converging on molecular mediators of transcriptional regulation. Relative to GENCODE annotations, the majority of accessible regions were located in gene introns or intergenic regions, with a smaller fraction located proximal to TSSs in putative promoter elements. For peaks with greater accessibility in flow samples, the proportions of intronic and intergenic peaks were increased, suggesting that newly open regions were predominantly distal elements that may include enhancers. Differentially accessible peaks co-localised with histone marks associated with active transcription, e.g. H3K27Ac, in ENCODE datasets and the majority overlapped genomic regions classified as enhancers by ChromHMM data from 9 cell lines. Shear-responsive chromatin changes were therefore deemed to be mostly occurring at enhancer elements. The overlap between differentially accessible peaks and DHS identified in primary cells and cell lines revealed commonly accessible sites in HAECs and several other cell types, particularly fibroblasts, HUVECs, skeletal muscle myoblasts and astrocytes. Some of these non-endothelial cell types display similar functions to endothelial cells *in vivo*, such as extracellular matrix deposition (fibroblasts)

and response to mechanical force (skeletal muscle myoblasts). This shared accessibility may result from common regulatory mechanisms at the epigenetic level, giving rise to similar patterns of functional gene expression in different cell types.

Previous studies have reported a correlation between chromatin accessibility and gene expression in a wide range of mammalian cells and tissues, and several have explored the relevance of this relationship to development and disease[53,239,240]. Integrating the RNA-seq and Omni-ATAC-seq data generated during this study supported this association, with a positive correlation between accessibility at gene promoters and expression. Accessibility changes in response to laminar shear stress were concordant with gene expression changes, but the average effect was relatively small, and some genes showed strong expression changes opposed to changes in promoter accessibility. These results suggest that while chromatin accessibility proximal to gene TSSs does correlate with gene expression, it does not fully explain the changes observed. Dynamic enhancers, distal to gene TSSs, may influence gene expression in response to laminar shear stress. A similar correlation to that between gene promoters and expression was observed for distal elements and expression of the nearest gene, however the average effect of chromatin accessibility changes was weaker than for promoters. Although many differentially accessible peaks overlapped chromatin states characteristic of enhancers, these may not all be active in this cell type and under these conditions. *Cis*-regulatory elements such as enhancers can act over large genomic distances, up to 500 kb in some cases, and not all enhancers directly regulate the closest gene. In addition, some genes may be regulated by multiple distal elements concurrently, with complex interplay between activation and repression. As such the precise linkage of enhancer activity and gene expression using correlational methods is a challenge, although some success has been reported for much larger datasets[239]. Additional methodologies such as high-

throughput chromosome conformation capture (Hi-C) can aid in the deconvolution of enhancer-gene relationships genome-wide, but publicly available data may not prove useful as physical interactions between loci are often cell-type and context specific[84]. Despite this limitation, gene ontology and pathway enrichment testing of the genes closest to differentially accessible peaks highlighted similar processes to those enriched for differentially expressed genes, such as terms related to cell motility, reorganisation of the actin cytoskeleton, extracellular matrix, and signaling via intracellular calcium and G proteins. Several other pathways implicated in endothelial mechanotransduction were also enriched, such as PI3K-Akt, MAPK, Wnt and Hippo signaling, as well as pathways relevant to endothelial homeostasis such as VEGF, PDGF and mTOR signaling[241]. Immune functions relevant to the role of endothelial cells in atherosclerosis were also enriched, including transendothelial leukocyte migration, interleukin signaling and the TLR4 cascade. A number of the enriched terms related to neuronal development and pathways, which in light of the substantial overlap of differentially accessible peaks with astrocyte DHS may point to signals and functions common to both endothelial cells and cells of neuronal origin.

In order to identify the molecular mediators of the flow-induced changes in HAECs, transcription factor motifs from the HOMER database were used to annotate the reference genome and their occurrence within differentially accessible peaks was analysed. Peaks with greater accessibility in flow samples were most strongly enriched for motifs similar to that bound by AP-1, known to play a role in various cellular processes and also regarded as a 'pioneer' transcription factor, capable of activating transcription by displacing nucleosomes to facilitate the binding of additional factors[254]. These motifs showed the greatest variability in accessibility between control and flow samples using chomVAR, and aggregation of Tn5 insertions around AP-1 motifs revealed a clear

footprint indicative of protection of DNA from cleavage by proteins bound at these sites. AP-1 is a heterodimeric transcription factor potentially comprising a number of proteins containing bZIP domains, including Fos, Jun, Maf and ATF proteins, and is subject to extensive regulation by post-translational modification[255]. As such, using these methods, the precise structure of the transcription factor (or factors) responsible for the widespread increase in chromatin accessibility in HAECs exposed to laminar shear stress cannot be determined with confidence. However, genes encoding bZIP proteins were strongly upregulated by laminar shear stress, particularly *ATF3*, which has been implicated in the cellular response to mechanical and ischemic stress[256], the protection of endothelial cells from TNFα-induced apoptosis[257], and the processing of inflammatory lipids by human macrophages[258]. The ATF-3 transcription factor has also been shown to bind enhancer regions and regulate the expression of nearby genes[259]. *ATF3* is therefore a promising candidate for further study, using methods such as ChIP-seq to identify binding events across the genome, and characterising the response to shear stress of cells with ablated *ATF3* expression to infer perturbed regulatory networks. Genome-wide maps of KLF2 and KLF4 binding, both strongly upregulated by laminar shear stress, would also be of significant value, considering the lack of publicly available data targeting these transcription factors and their significant role in endothelial mechanotransduction. Integration of transcription factor ChIP-seq with the data presented herein could also confirm the cooperative synergy between AP-1 and EKLF motifs indicated by analysis with chromVAR.

The contribution of genetic variation to the development of atherosclerosis and resulting cardiovascular disease remains mostly unexplained, despite the large number of loci significantly associated with CAD in genome-wide association studies. This is primarily due to the fact that most of the variants identified by GWAS lie outside of protein coding

genes, along with the limited resolution of SNP arrays typically utilised for association studies, causing ostensibly disease-associated SNPs to be reported only due to high linkage disequilibrium with a causal SNP. It is therefore not always possible to resolve the functional consequences of genetic variation from GWAS directly. CAD SNPs were significantly enriched in dynamically accessible genomic regions, relative to matched control SNPs identified by the 1000 genomes project, suggesting that CAD-associated variants may alter the sequence of regulatory elements targeted by endothelial mechanotransduction, and thereby influence a key determinant of plaque localisation and disease progression. Transcription factor motifs are concentrated in accessible regions of the genome, and it was predicted that some of the variants overlapping differentially accessible peaks might disrupt the core sequence necessary for the binding of particular transcription factors. This was true for a number of SNPs, with rs17608766 a key example. The minor allele of rs17608766 was found to disrupt the EKLF motif from the HOMER database, as well as the Klf1 and Klf4 motifs from the Jaspar Core 2014 Vertebrate motif library. Further work is required to confirm transcription factor binding at this motif, but strong upregulation of the genes coding for both KLF2 and KLF4 in HAECs under laminar shear stress makes these proteins the primary candidates. The GTEx consortium identified rs17608766 as an eQTL for the nearby genes *GOSR2*, *RPRML* and *CDC27*, but the association was not significant in arterial tissue. This may be due to the variant exerting its effect during the response to laminar shear stress or affecting different genes under these conditions. Chromosome conformation capture could be used to identify the spatial links between this region and other parts of the genome, and ultimately reveal pathways contributing to the heritable risk of CAD. SNPs associated with a number of other traits, some of which are intimately linked to atherogenesis (e.g. blood pressure, circulating lipids), were also enriched in the regions

responsive to laminar shear stress, and motif-oriented analysis of these may reveal further promising candidates for future study.

Overall, the work described in this chapter has provided high quality transcriptomic and epigenomic profiles of primary human endothelial cells cultured with and without high laminar shear stress, which is a known influence on endothelial activation, homeostasis, and atherogenesis. These profiles were deployed to characterise transcriptional responses highly relevant to endothelial cell biology and atherosclerotic disease progression, as well as identifying the regulatory mechanisms driving these responses. Integration with motif libraries and GWAS data identified genetic risk factors potentially influencing cellular responses to environmental risk factors. Further study and expansion of this model, along with validation and functional characterisation of prioritised variants, may identify pathways open to therapeutic intervention for the mitigation of cardiovascular disease risk.

Chapter 6

General discussion

The principal aim of this study was to investigate how environmental and genetic influences on atherosclerosis interact during key stages of pathogenesis. Defining the gene expression patterns and regulatory pathways underlying these processes, as well as the contribution of genetic variation, has the potential to yield targets for therapeutic intervention and ultimately mitigation of atherosclerotic disease burden. Lipid lowering and antihypertensive medications remain key strategies for the prevention of atherosclerotic disease, but beyond these the development of further therapeutic strategies has been limited due to ambiguity in the molecular mechanisms driving the disease. To address this, the present study applied high-throughput, so-called 'next generation', sequencing methodologies to determine the response of primary human cells, with known involvement in atherosclerosis, to environmental stimuli with demonstrated influence on disease progression. Specifically, these methods were applied to models of macrophages exposed to a hyperglycaemic environment, endothelial cells under laminar shear stress, and macrophages treated with inflammatory lipoprotein.

Accelerated atherogenesis is a major complication of diabetes mellitus types I and II, with cardiovascular disease being the predominant cause of mortality in diabetic patients[260]. An obvious hallmark of poorly controlled diabetes is elevated plasma glucose concentration, and fasting glucose levels have even been associated with coronary artery disease incidence in non-diabetic individuals[160], but diabetic dyslipidaemia and aberrant insulin signaling have also been posited as drivers of atherogenesis[161,261,262]. Numerous *in vitro* studies of cells found in vascular lesions, particularly macrophages, have reported a pro-atherogenic influence of high ambient glucose concentrations (usually ≥ 25 mM) on cell function. In the case of macrophages, this manifests as increased expression of scavenger receptors mediating the endocytosis of modified lipoproteins (e.g. oxLDL)[263,264], and increased expression of inflammatory cytokines[265,266]. Despite these

observations, if and how elevated glucose influences macrophage function in atherosclerosis remains unknown.

During the course of this study, primary human monocyte-derived macrophages were differentiated in media containing normal (5 mM) or high (25 mM) glucose concentrations over 7 days, and the transcriptome of cells in each condition was surveyed using RNA-seq. Only a limited number of genes showed differential expression between normal and high glucose macrophages, and the magnitude of expression change was small for the majority of these (i.e. < two-fold). In contrast to previous studies of prolonged glucose exposure in macrophages, there was no differential expression of genes encoding inflammatory cytokines, such as *IL1B*, or macrophage scavenger receptors, such as *CD36* and *MSR1*. The changes in gene expression that were observed did not appear to be mediated by changes in chromatin accessibility, as comparison of ATAC-seq profiles generated from normal and high glucose macrophages identified no significant differentially accessible regions. Among the genes with relatively large fold changes in expression was *CD28*, which encodes an immune receptor well studied for its co-stimulatory role in the activation of T lymphocytes[152], but its expression in macrophages has not previously been studied. Expression of *CD28* was confirmed on the surface of primary human macrophages and was shown to decrease in macrophages exposed to high glucose at both the mRNA and protein level. This supports the association between CD28 signaling and glucose metabolism established in T lymphocytes and extends this link to a potentially reciprocal relationship that also exists in human macrophages. The consequences of CD28 signaling for macrophage function represent an intriguing new area for future research. Overall, these results show that the direct effect of high ambient glucose concentration on human macrophages is minimal, and unlikely to provide an explanation for accelerated atherosclerosis in diabetes.

Modified lipoproteins have been considered central to the development of atherosclerosis for some time, with particularly widespread focus on oxLDL[267]. Macrophage uptake of oxLDL is well established as a mechanism for formation of the foam cells observed in early lesions[268], and there is a growing body of evidence that these are inflammatory cytokine-secreting cells[10,58,269,270]. OxLDL has been shown to drive epigenomic changes in human macrophages *in vitro*, and these changes were affected by SNPs associated with coronary artery disease[53]. Macrophage populations display heterogeneity both *in vitro* and *in vivo*, but the functional consequences of this heterogeneity for atherogenesis remain unknown[168]. The recent development of single-cell RNA-sequencing (scRNA-seq) has afforded the opportunity to study the patterns of gene expression within such populations, and to thereby infer functional separation of cell subtypes[271]. Several recently published studies used scRNA-seq to dissect the heterogeneity of lesional macrophages in murine models of atherosclerosis[209,272], but similar studies in human cells and tissues have yet to be reported.

In this study, primary human monocyte-derived macrophages from healthy donors were cultured with either atherogenic oxLDL or control buffer, and transcriptomic heterogeneity in macrophage samples was explored using scRNA-seq. Multiple clusters of human macrophages, separated by gene expression reflecting major macrophage functions such as cytokine secretion, antigen presentation and phagocytosis, were identified in each sample, and were conserved between buffer- and oxLDL-treated macrophages. The response to oxLDL was largely consistent across clusters, with strong upregulation of genes encoding scavenger receptors for oxLDL (*CD36*, *MSR1*), cholesterol efflux transporters (*ABCA1*, *ABCG1*) and inflammatory cytokines (*CCL2, IL1B*). There was also global upregulation of genes encoding key enzymes of glycolysis, such as *PKM* and *HK2*, and an increased rate of glycolysis was measured in oxLDL-

treated macrophages, with a concomitant decrease in oxidative phosphorylation. Intriguingly, macrophage expression of *HIF1A* was also upregulated by oxLDL, and the transcription factor encoded by this gene, HIF-1α, is known to target genes of the glycolytic pathway during the response to hypoxia[273]. Reactive oxygen species (ROS) have also been reported to regulate *HIF1A* expression[228], and depletion of the antioxidant reduced glutathione in oxLDL-treated macrophages was indicative of oxidative stress. This provides a potential mechanism for the metabolic reprogramming of macrophages during atherogenesis, which is associated with inflammatory activation[274], and suggests that this chain of events may occur in naïve macrophages during the early stages of atherogenesis. Cell cluster-specific responses to oxLDL were also detected by scRNA-seq, and further work to validate the existence of these macrophage subtype-unique responses *in vivo* could mark them as attractive targets for therapeutic immunomodulation.

As well as macrophages, endothelial cells play a central role in the development of atherosclerosis. The endothelium is a selective interface between circulating blood and tissues, and dysregulation of endothelial integrity is one of the earliest events in atherogenesis[19]. A range of environmental factors have been implicated in the subversion of normal endothelial function in atherosclerosis, including high glucose[275] and oxLDL[276], but there is clear evidence for the influence of haemodynamic forces, and endothelial shear stress in particular, on the formation of atheromatous plaque[34,232]. The distribution of shear stress throughout the cardiovascular network is non-uniform. Arterial branches, bifurcations and lesser curvatures are exposed to reduced, disturbed or oscillatory patterns of shear stress, in contrast to the unidirectional laminar shear stress acting on the endothelium in straight, unbranched regions of the vasculature. Atheromata are consistently localised to the former regions with low or multidirectional endothelial

shear, whereas high unidirectional shear stress protects vessels from atherosclerosis[5]. There have been great strides in elucidating the mechanosensory mechanisms at the endothelial cell surface, and the pathways involved in transduction of these signals[40,233]. As such, extensive changes in gene expression have been reported to induce a proatherogenic phenotype in endothelial cells subjected to low, oscillatory shear stress, with opposing changes under atheroprotective shear stress[42,245]. Despite these advances, there have been few studies of the role of epigenetic mechanisms in regulation of the endothelial response to shear stress[44,92], and the potential for vascular disease-associated genetic variation to influence this process.

Following on from these observations, primary human aortic endothelial cells were cultured in a parallel-plate flow chamber with perfusion of culture medium to simulate physiological shear stress. Genome-wide maps of changes in gene expression and chromatin accessibility, as determined by RNA-seq and ATAC-seq respectively, revealed profound regulation of endothelial cell transcription by laminar shear stress when compared to controls in static culture. Gene expression changes were consistent with an atheroprotective phenotype and were correlated with changes in chromatin accessibility. The dynamic chromatin regions identified by differential analysis were primarily located in introns or intergenic regions, and overlapped regulatory elements previously identified in other cell types. These regions were therefore deemed likely to primarily consist of enhancer elements, and comparisons with open chromatin profiles from other primary cells revealed regulatory similarities to other vascular cells, fibroblasts and, surprisingly, neuronal cells. The differences in chromatin accessibility between shear stress conditions also identified a number of transcription factor binding motifs which colocalised with dynamic regions, particularly AP-1 and KLF motifs, suggesting regulatory involvement of the factors binding these sequences. Integration

with RNA-seq data identified novel factors potentially mediating the endothelial response to shear stress at the genomic level, such as ATF3, which has been associated with atherosclerosis in a variety of contexts[257,258,277]. GWAS-identified genetic variants associated with coronary artery disease, as well as other traits known to influence atherosclerosis, were enriched in shear-responsive accessible chromatin regions. Analysis of the overlapping sequences revealed plausible mechanisms by which non-coding variation could influence gene regulation, such as the alteration of canonical binding motifs for shear-responsive transcription factors by SNPs. While the experiments described herein provide a restricted view of the epigenomic modifications induced by laminar shear stress in endothelial cells, they demonstrate the value of such an approach in discerning how heritable disease risk might influence disease progression. Expansion of this model by supplementation with additional epigenomic datasets, such as ChIP-seq for histone modifications and transcription factors, and chromatin conformation capture data to map the dynamics of enhancer-promoter associations, will provide a comprehensive view of the regulatory landscape defining the endothelial response to atheroprotective laminar shear stress.

In conclusion, these results repeatedly demonstrate the value of ‘omics’ approaches in understanding, at the molecular level, the influence of environmental factors on cellular processes. Striking transcriptomic and epigenomic changes were associated with responses to atherogenic and atheroprotective stimuli, and incorporation of public GWAS data enabled the identification of molecular mechanisms potentially causative for the genetic risk of coronary artery disease. In the post-GWAS era, studies such as this will fuel the discovery of therapies for complex disease tailored to individual genetics, and thus facilitate clinical translation in the emerging age of genomic medicine.

Appendix 1

Published work

Some of my work has appeared in peer-reviewed journal articles, which are attached in the following pages. This work is not featured in the main body of the thesis.

Intragenic transcriptional interference regulates the human immune ligand MICA. Lin D, **Hiron TK**, O'Callaghan CA *EMBO J* 2018 doi: 10.15252/embj.201797138

Purine nucleotide metabolism regulates expression of the human immune ligand MICA. McCarthy MT, Moncayo G, **Hiron TK**, Jakobsen NA, Valli A, Soga T, Adam J, O'Callaghan CA *J Biol Chem* 2018 Dec 26. doi: 10.1074/jbc.M117.809459

Published online: April 11, 2018

Article

TRANSPARENT PROCESS

THE EMBO JOURNAL

Intragenic transcriptional interference regulates the human immune ligand MICA

Da Lin, Thomas K Hiron & Christopher A O'Callaghan*

Abstract

Many human genes have tandem promoters driving overlapping transcription, but the value of this distributed promoter configuration is generally unclear. Here we show that *MICA*, a gene encoding a ligand for the activating immune receptor NKG2D, contains a conserved upstream promoter that expresses a noncoding transcript. Transcription from the upstream promoter represses the downstream standard promoter activity in *cis* through transcriptional interference. The effect of transcriptional interference depends on the strength of transcription from the upstream promoter and can be described quantitatively by a simple reciprocal repressor function. Transcriptional interference coincides with recruitment at the standard downstream promoter of the FACT histone chaperone complex, which is involved in nucleosomal remodelling during transcription. The mechanism is invoked in the regulation of MICA expression by the physiological inputs interferon-γ and interleukin-4 that act on the upstream promoter. Genome-wide analysis indicates that transcriptional interference between tandem intragenic promoters may constitute a general mechanism with widespread importance in human transcriptional regulation.

Keywords FACT; MICA; NKG2D; tandem promoter; transcriptional interference
Subject Categories Chromatin, Epigenetics, Genomics & Functional Genomics; Signal Transduction; Transcription
DOI 10.15252/embj.201797138 | Received 19 April 2017 | Revised 8 February 2018 | Accepted 9 February 2018 | Published online 11 April 2018
The EMBO Journal (2018) 37: e97138

Introduction

Genomic analyses demonstrate that many human genes have tandem promoter gene structure (Djebali *et al*, 2012; Fantom Consortium and the RIKEN PMI and CLST, 2014). A transcript initiated from an upstream promoter can overlap that from a downstream promoter; the transcripts will differ in their first exons, but may share similar downstream sequence. Additional promoters have been studied mainly in the context of the protein coding function of the different transcripts generated (Carninci *et al*, 2006; Davuluri *et al*, 2008; Wang *et al*, 2016). Less explored are the consequences of transcription from additional upstream promoters on gene expression.

Transcription can have an in *cis* influence on promoters that lie in the path of transcriptional elongation through a process known as transcriptional interference (Shearwin *et al*, 2005; Palmer *et al*, 2011). Transcriptional interference is thought to occur when a traversing RNA polymerase elongation complex, arising from one promoter, encounters and displaces a transcription initiation complex, which has formed transiently at a different promoter (Shearwin *et al*, 2005; Palmer *et al*, 2011). In some cases, epigenetic changes of transcriptional elongation have been associated with transcriptional interference (Houseley *et al*, 2008; Hainer *et al*, 2011; Ard & Allshire, 2016). Transcriptional interference is well characterized in lower eukaryotes with compact genomes where the transcriptional path of one gene runs on into the promoter of another gene (Greger *et al*, 2000; Martens *et al*, 2005; Hongay *et al*, 2006; Petruk *et al*, 2006; Gummalla *et al*, 2012; Ard *et al*, 2014). In mammalian species, reports of transcriptional interference are limited to a few examples of developmentally regulated genes that are subject to relatively stable epigenetic regulation following the cell differentiation process (Abarrategui & Krangel, 2007; Racanelli *et al*, 2008; Latos *et al*, 2012; MacIsaac *et al*, 2012). A necessary condition for transcriptional interference is overlapping transcription. Tandem promoter gene structure, as found in many mammalian genes, strictly implies overlapping transcription within the gene, such that the upstream promoter drives transcription through the downstream promoter. This raises the possibility that for some of the human genes with additional upstream promoters, the tandem promoter arrangement may have evolved to regulate gene expression through transcriptional interference.

MICA is a transmembrane protein with structural similarity to MHC class I molecules and is encoded by the *MICA* gene within the human MHC complex (Bahram *et al*, 1994; Li *et al*, 2001). MICA is a ligand for NKG2D, an activating immune receptor expressed on $CD8^+$ T cells, γδ T cells and natural killer cells (Bauer *et al*, 1999; Wu *et al*, 1999). Engagement of NKG2D on NK cells by MICA expressed on target cells promotes cytokine release and killing of the target cells and is implicated in cancer immunity, antiviral immunity and autoimmunity (Bauer *et al*, 1999; Ullrich *et al*, 2013; Lanier, 2015). Homologs of the *MICA* gene are present in all mammals studied, except rodents (Kasahara & Sutoh, 2015). The *MICA* gene is highly polymorphic, with over 80 coding alleles

Nuffield Department of Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
*Corresponding author. Tel: +44 1865 287794; E-mail: chris.ocallaghan@ndm.ox.ac.uk

known to date (Robinson *et al*, 2015). Genetic studies have associated the *MICA* locus with inflammatory diseases, and with the response to virus infection (Kumar *et al*, 2011; Zhou *et al*, 2014; Zhang *et al*, 2016), although strong linkage disequilibrium between *MICA* and other genes within the MHC, especially *HLA-B,* is a confounding factor in such studies (Le Clerc *et al*, 2014; Okada *et al*, 2014). MICA expression is upregulated in cancer tissues, and in response to a diverse range of stimuli including virus infection, heat shock, metabolic stress, cell proliferation and cytokines, such as TNFα and interferon-γ (Groh *et al*, 1996; Zou *et al*, 2005; Cerboni *et al*, 2007; Schwinn *et al*, 2009; Lin *et al*, 2012; McCarthy *et al*, 2017). However, the molecular mechanisms that govern MICA regulation remain enigmatic. Many stimuli are known to affect *MICA* transcription, but only a few specific transcription factors, such as NF-κB and HSF1, have been shown to directly regulate *MICA* transcription through binding to the *MICA* promoter (Venkataraman *et al*, 2007; Lin *et al*, 2012).

Here, we show that in *MICA* a conserved upstream additional promoter drives transcription of an unstable noncoding transcript and that this transcriptional activity inhibits transcription of the coding transcript from the standard downstream promoter in *cis* through transcriptional interference. This mode of transcriptional regulation is exemplified by the physiological signals interferon-γ and interleukin-4, which exert transcription factor-mediated positive and negative influences, respectively, on the upstream promoter, resulting in opposite effects on production of the *MICA* coding transcript from the standard downstream promoter. Our results demonstrate for the first time that intragenic transcriptional interference occurs in humans and constitutes a potentially powerful form of real-time transcriptional regulation. Evolutionary selection for upstream promoter components provides a means for reversing the polarity of an input signal on expression of a coding transcript, thus turning activators into repressors or *vice versa*.

Results

An alternative upstream MICA promoter drives expression of an unstable noncoding transcript

MICA encodes two major transcripts, a standard transcript (MICA-ST) that initiates from the standard *MICA* promoter, and an alternative upstream transcript (MICA-UT) that initiates from an alternative promoter 2.9 kb upstream (Fig 1A). The two transcripts have different first exons, but share common downstream exons. We found that both transcripts were widely expressed across a range of human cells and tissues, but with distinct expression patterns—in some cases there was abundant expression of the upstream transcript with little or no expression of the standard transcript (Figs 1B, and EV1A and B).

There are *MICA* homologs in the genomes of almost all mammals studied except rodents (Kasahara & Sutoh, 2015). We identified homologs of both the upstream and standard transcripts for pigs and cows (Fig EV1C) and alignment with the corresponding genomes revealed conservation of the tandem promoter gene structure and local sequence around the upstream promoter in these species and humans (Fig EV1C and D). RT–PCR analysis confirmed the expression of both transcripts in adult pigs and cows (Fig EV1E). This conservation suggested that the upstream promoter or transcript could have a biological function.

Cell surface expression of human MICA protein correlated with expression of the standard transcript, but not the upstream transcript (Fig 1C and D), consistent with the standard transcript encoding the MICA protein. Both transcripts were 5′ capped, and poly-A tailed (Fig 1E and F), indicating that the upstream transcript is transcribed by RNA polymerase II and processed in the same way as the standard transcript. Bioinformatic analysis of the upstream transcript sequence predicted that it would undergo nonsense-mediated decay due to the presence of an upstream open reading frame encoding a premature termination codon (Hug *et al*, 2016). Consistent with this, it displayed rapid decay kinetics compared to the standard *MICA* transcript, as measured by global transcriptional inhibition with actinomycin D (Fig 1G and H) or by 4-thiouridine metabolic labelling (Fig 1I). No protein product could be detected for the major open reading frame of this transcript (Fig EV1F). The standard transcript was enriched in RNA isolated from heavy polysome fractions similar to GAPDH and ITGB5, consistent with active translation of this transcript, but the upstream transcript was predominantly in the ribosome-free and monosome fractions (Fig 1J). This indicates that the upstream transcript is poorly translated, which is consistent with the prediction that it is noncoding and undergoes nonsense-mediated decay. Together, these data demonstrate that the upstream transcript of the *MICA* gene is an unstable noncoding transcript.

The upstream promoter represses expression of the standard coding MICA transcript in *cis*

Expression of the upstream transcript *in trans* or post-transcriptional siRNA-mediated downregulation of the upstream transcript had no effect on the expression of endogenous MICA (Fig 2A and B). These findings suggest that the transcript itself does not regulate MICA expression, leaving open the possibility that the alternative upstream promoter might itself regulate expression of the downstream coding transcript. To test this hypothesis, we used CRISPR genome editing tools to delete the upstream promoter in primary cells (Fig 2C). We mapped the core upstream and standard *MICA* promoters by serial deletion using reporter assays (Fig EV2A and B), and designed pairs of CRISPR guide RNAs for deletion of the core upstream promoter (Fig EV2C). In primary human fibroblasts and arterial endothelial cells, transient expression of the Cas9 nuclease and either of two different pairs of CRISPR guide RNAs for the upstream promoter caused significant upregulation of *MICA* expression compared to CRISPR guide RNAs targeting control loci and the effect was greater in the fibroblasts (Figs 2D, and EV2D and E). We also used a deactivated Cas9 (dCas9)-based programmable transcriptional activation system to specifically activate the upstream promoter and study its effect on *MICA* expression. This synergistic activation mediator (SAM) system consists of dCas9 fused to a VP64 transactivation domain and a locus-specific guide RNA that recruits multiple copies of a p65 and HSF1-derived multidomain transactivator (Konermann *et al*, 2015; Fig 2E). dCas9-based activation of the upstream promoter downregulated *MICA* expression from the standard downstream promoter, whereas activation of the standard promoter upregulated *MICA* expression (Fig 2F). These observations

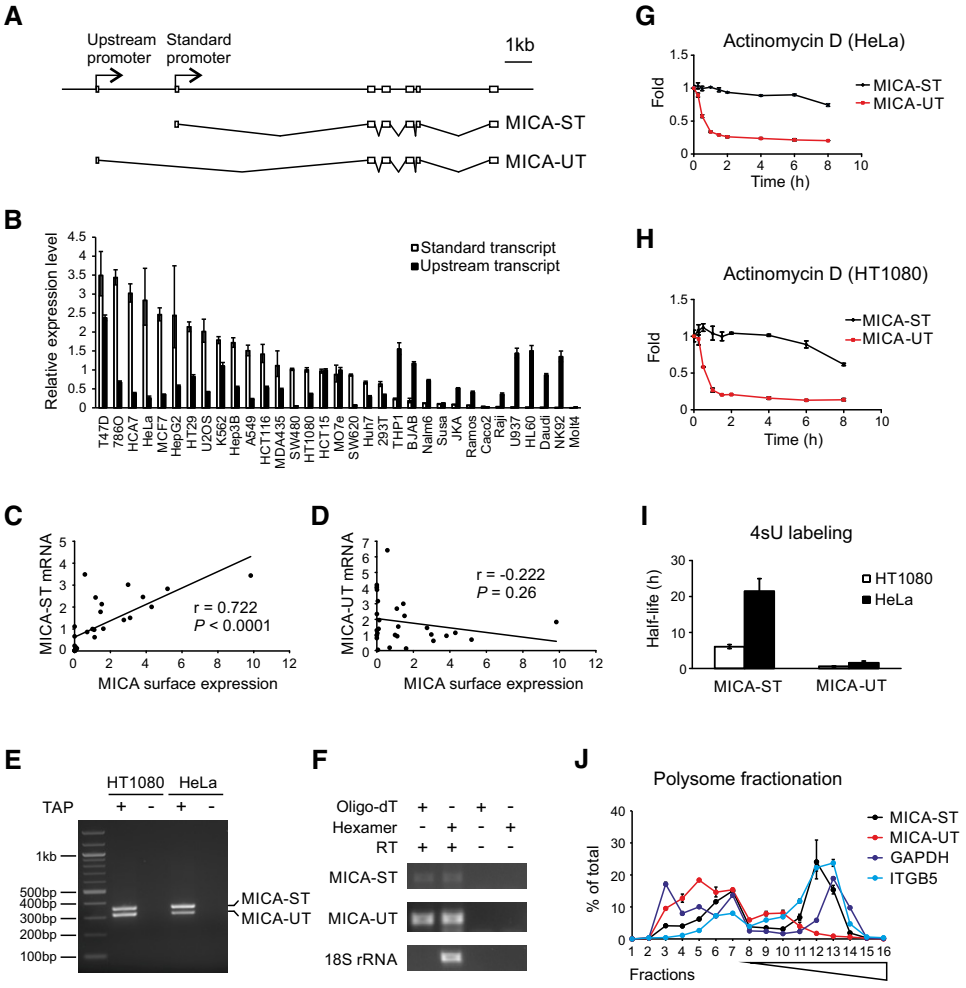

Figure 1. Properties of the MICA upstream transcript.

A Exon structure of the *MICA* gene, standard transcript (MICA-ST) and upstream transcript (MICA-UT). In the upstream transcript, an alternative upstream first exon is spliced to exon 2 of the gene. Both transcripts share common downstream exons.

B The upstream transcript and standard transcript levels measured by qPCR in different cells. Relative expression levels were normalized to that of the standard transcript in HT1080 cells and ranked according to expression of the standard transcript. Error bars represent standard deviations of three replicates.

C Positive correlation between the level of the standard transcript and cell surface MICA expression in 28 different cell types.

D No significant correlation was found between the level of the upstream transcript and cell surface MICA expression in these different cells. The Pearson's correlation coefficient *r* and associated *P*-values are shown.

E RLM-RACE analysis of the upstream and standard MICA transcripts using a common primer in exon 2 demonstrates that both transcripts are 5′ capped. Samples prepared without tobacco acid pyrophosphatase (TAP) treatment were used as negative controls.

F RT–PCR using hexamer or oligo-dT primed cDNA shows that both transcripts are 3′ polyadenylated. Nonpolyadenylated 18S rRNA and the RT–PCR without reverse transcriptase (RT) were used as negative controls.

G, H Stability of the upstream and standard MICA transcripts measured following actinomycin D treatment in HeLa cells (G) or HT1080 cells (H). Error bars represent standard deviations of three replicates.

I Stability of the upstream and standard MICA transcripts measured by 4sU metabolic labelling in HeLa or HT1080 cells. Error bars represent standard deviations of three replicates.

J Polysome profiling of the standard transcript (MICA-ST), upstream transcript (MICA-UT) and GAPDH and ITGB5 as controls in HT1080 cells. Distributions of mRNA across sucrose gradient fractions are shown. Numbers 1–6 represent ribosome-free fractions, number 7 represents the monosome fraction, and numbers 8–16 represent polysome fractions with increasing number of ribosomes. Error bars represent standard deviations of three replicates.

indicate that the upstream promoter negatively regulates the activity of the standard downstream promoter.

In one case, the CRISPR-deletion experiment was undertaken in primary human fibroblasts carrying one functional MICA allele (MICA*004) and one null allele (MICA*010) that does not reach the cell surface (Li *et al*, 2000). Exploiting this feature, we used a forward genetics approach to test whether the core upstream promoter regulates *MICA* expression in *cis* (Fig 2G). CRISPR deletion of the upstream promoter can result in four major genotypes: wild type, monoallelic deletion of either allele and biallelic deletion of the upstream promoter. If the upstream promoter only represses *MICA* expression in *cis*, then deletion of the upstream promoter of

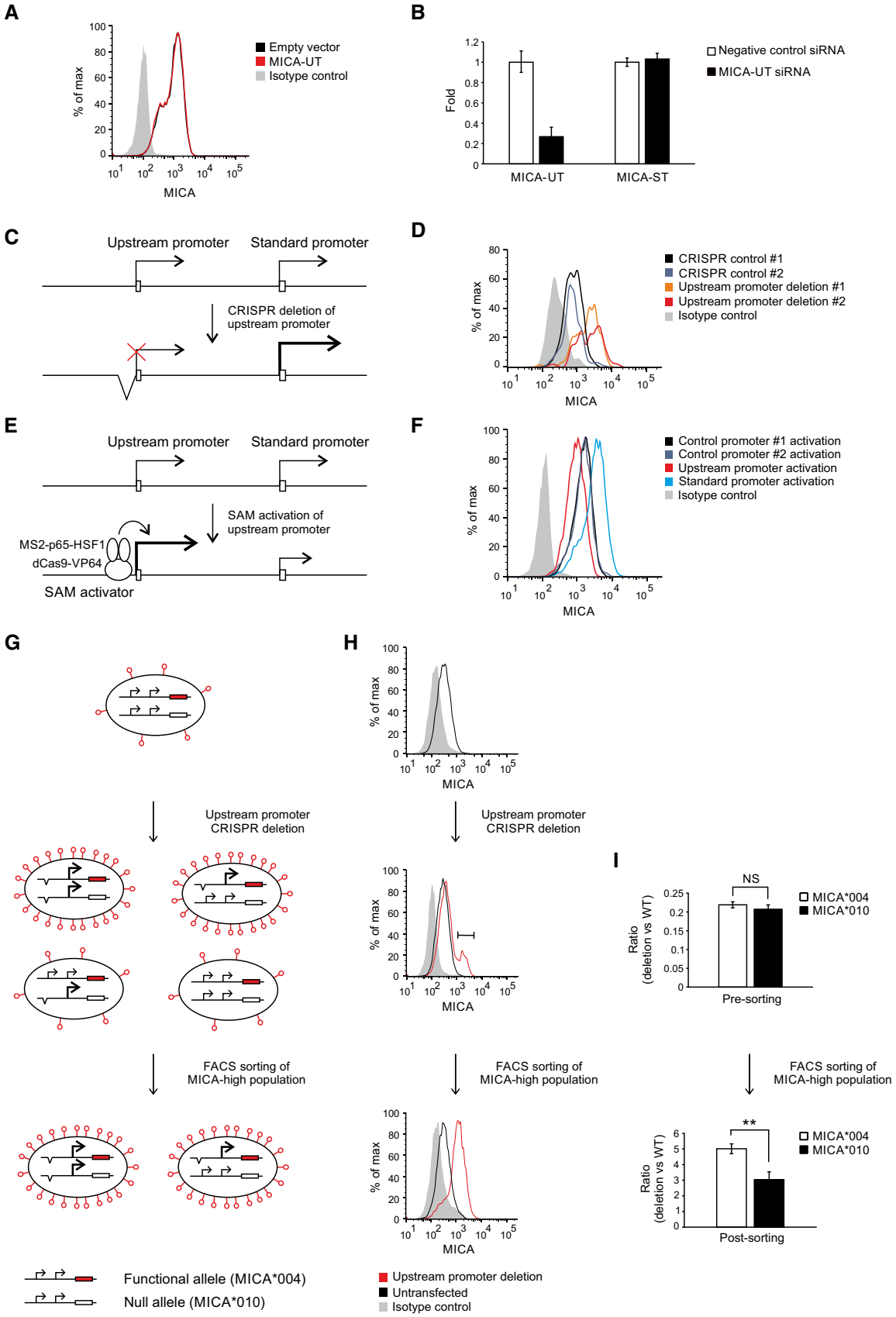

Figure 2.

Figure 2. The upstream promoter represses MICA expression in *cis*.

A Flow cytometric analysis of MICA surface expression in 293T cells transfected with plasmids expressing the full-length upstream transcript or empty vector control and the pEGFP-N1 plasmid as transfection control. Cells were gated for the GFP-positive population.

B qPCR analysis of the upstream transcript and standard transcript in 293T cells transfected with siRNA targeting the upstream transcript or with control siRNA. Error bars represent standard deviations of three replicates.

C Diagram of CRISPR deletion of the core upstream promoter.

D Flow cytometry of cell surface MICA expression in primary human fibroblasts following transfection with CRISPR plasmids targeting deletions of the MICA upstream promoter or control genes (*HLA-B* or *PDPN*). Cells were gated for the CRISPR Cas9 nuclease-transfected GFP-positive population.

E Diagram of dCas9-based transcriptional activation of the upstream promoter. The SAM transcription activator consists of dCas9 fused to a VP64 transcription activator and a guide RNA containing RNA aptamers that recruit multiple copies of the MS2 bacteriophage coat protein fused to a multidomain transcription activator derived from the p65 and HSF1 transcription factors.

F Flow cytometry of cell surface MICA expression in 293T cells following transfection with SAM plasmids activating the *MICA* upstream promoter, standard promoter or control promoters (*CD43* and *CD36*). Cells were gated for the transfected GFP-positive population.

G Schematic diagram representing the different possible genotypes that may arise and the associated phenotypes predicted by the in *cis* transcriptional interference hypothesis. Primary human fibroblasts were used which are heterozygous for the MICA*004 and MICA*010 alleles, of which only MICA*004 (red) reaches the cell surface. CRISPR deletion of the upstream promoter can result in three additional genotypes depending on whether one or both MICA alleles are affected. If, as we hypothesized, there is intragenic transcriptional interference, then deletion of the upstream promoter of the MICA*004 allele will result in upregulation of cell surface MICA expression, whereas deletion of the upstream promoter of the MICA*010 null allele will have no effect on cell surface MICA expression compared to wild-type cells. Accordingly, cells sorted for upregulated MICA surface expression should be enriched for cells with deletion of the MICA*004 allele if transcriptional interference occurs in *cis*.

H MICA surface expression of cells at different stages of the experiment. After CRISPR deletion of the upstream promoter, there is a population of cells with upregulation of MICA expression (middle panel) and this population was sorted for further analysis (lower panel).

I PCR analysis of the CRISPR-deletion genotype for each allele before and after sorting of cells with upregulated cell surface MICA expression. One of the differences between the MICA*004 and MICA*010 alleles is at SNP rs2596539 within the upstream promoter outside the deleted region, and this allows differentiation between the two alleles using restriction digestion prior to PCR amplification. As indicated by the genotype before sorting, CRISPR-mediated deletion of the upstream promoter arises with equal frequency for both alleles (upper panel). However, consistent with the in *cis* transcriptional interference hypothesis, the cells that have high MICA*004 surface expression are enriched for deletion of the in *cis* MICA*004 upstream promoter compared to the MICA*010 upstream promoter (lower panel). The difference in the enrichment of MICA*004 compared to MICA*010 upstream promoter deletion is modest as the majority of the sorted cells have biallelic deletion. Error bars represent standard deviations of four replicates. NS not significant; $^{**}P < 0.01$, Student's *t*-test.

the functional MICA*004 allele will upregulate cell surface MICA expression, but deletion of the upstream promoter of the MICA*010 null allele will not. Therefore, cells sorted for upregulated MICA surface expression would be enriched for deletion of the promoter of the functional MICA*004 allele (Fig 2G). In contrast, with *in trans* regulation, no allele-specific enrichment would be observed in cells sorted for upregulated MICA surface expression. As demonstrated in Fig 2H, deletion of the upstream promoter produced a population of cells with upregulation of MICA expression which is stable following cell sorting of this population. Although the upstream promoters of both alleles were deleted with equal efficiency in the pre-sorted population, cells with upregulation of cell surface MICA expression were preferentially enriched for deletion of the functional MICA*004 allele compared to deletion of the null MICA*010 allele (Fig 2I). This confirms that the upstream promoter represses *MICA* expression from the standard promoter in *cis*.

The upstream promoter represses MICA expression through transcriptional interference

We hypothesized that the process of transcription from the upstream promoter inhibited transcription from the standard promoter. However, in plasmid-based reporter assays we did not observe any inhibition of the standard promoter activity by the upstream promoter (Fig EV2F and G). Plasmid-based reporter systems lack distal genetic elements and the concomitant chromatin structure and environment. Therefore, we created a set of isogenic cell lines using recombinase-mediated cassette exchange to insert synthetic variants of the entire MICA locus into chromatin at the same genomic location (Fig 3A). These variants were constructed by BAC recombineering and included deletions of the core upstream or standard promoter or insertion of a transcription terminator between the two promoters (Fig 3B and Appendix Table S1). The endogenous *MICA* loci remained intact as controls for *in trans* effects. The host cell line was homozygous for the MICA*007 allele, which allows cell surface detection of the transgenic MICA*008 allele with an allele-specific monoclonal antibody (Fig EV3A).

Deletion of the upstream promoter in a *MICA* transgene containing the native ~20-kb flanking sequences downregulated expression of the transgenic upstream transcript and upregulated expression of the transgenic standard coding transcript as predicted (Fig 3C). Insertion of a transcription terminator between the upstream promoter and the standard promoter had a similar effect, which demonstrates that transcriptional elongation from the upstream promoter is required for repression of the standard promoter (Fig 3C). The slightly lesser effect of the upstream promoter deletion compared to the transcription terminator likely reflects residual low-level transcription from around the deleted region of the gene. Deletion of the standard promoter had no effect on expression of the upstream transcript, which excludes promoter competition as a mechanism for the inhibitory effect of the upstream promoter (Fig 3C). These different manoeuvres had no effect on the endogenous *MICA* alleles, confirming the in *cis* nature of the inhibitory effect of the upstream promoter on transcription from the standard promoter (Fig 3D). Changes in expression of the transgenic standard transcript were reflected in MICA protein expression at the cell surface (Fig 3E and F). Similar results were obtained using larger genomic inserts, which included ~75-kb flanking sequences on both sides of the *MICA* gene and incorporated the neighbouring *HLA-B* and *HCP5* loci (Fig EV3B and C). Further, we carried out ChIP for Ser5-phoshorylated RNA Pol II, which is a marker of transcription initiation activity, but decreases substantially during transcription elongation (Harlen & Churchman, 2017). Deletion of the standard downstream promoter abolished most of the Pol II phospho-Ser5

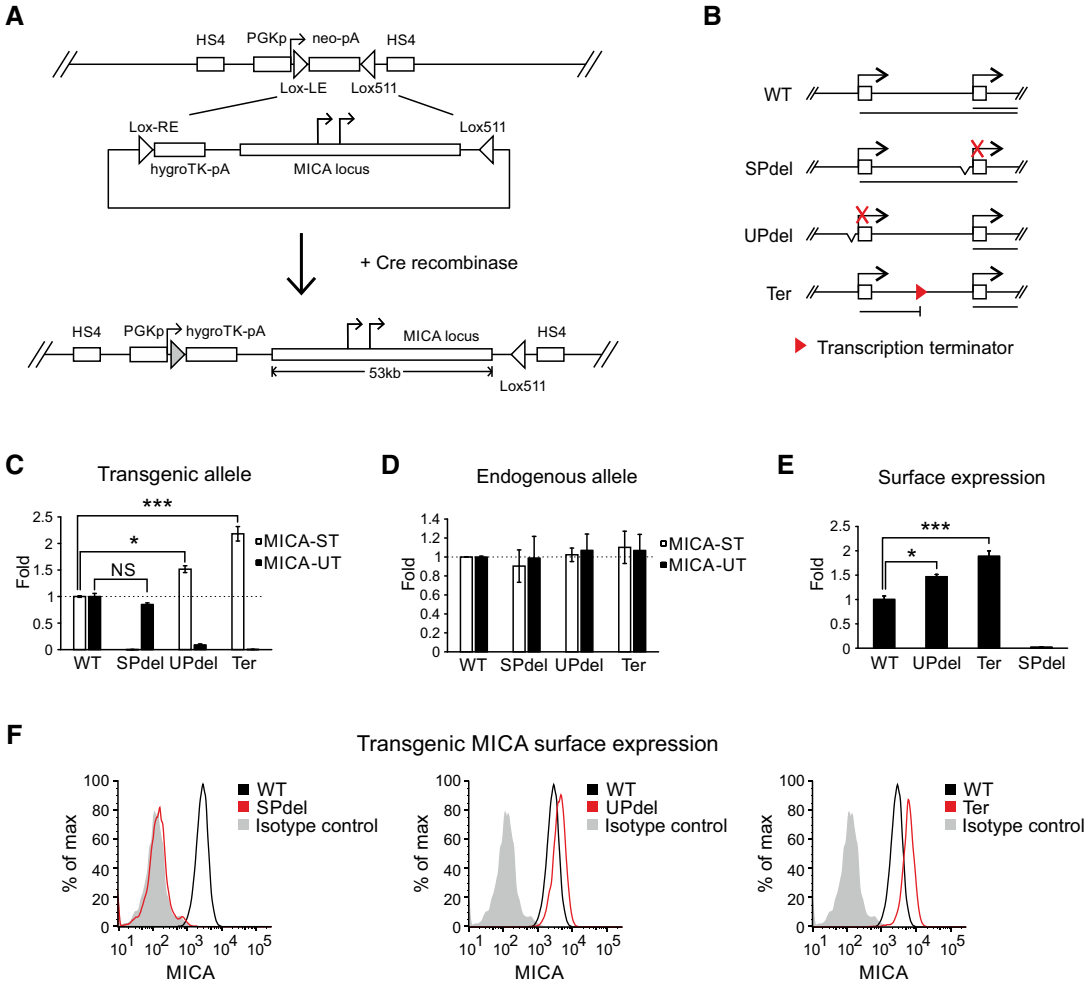

Figure 3. Transcriptional interference of the standard MICA promoter by the upstream promoter.

A Generation of the site-directed BAC reporter cell line. The acceptor cell line was generated such that it carries only a single copy of a landing site containing a neomycin-resistance cassette under the control of a PGK promoter. The neomycin cassette was flanked by variants of LoxP sites (Lox-LE and Lox511), and the entire inserted locus was insulated at both ends by HS4 insulators. The donor BAC constructs contain the *MICA* locus flanked on one side by a Lox-RE site followed by a promoterless hygromycin-resistance cassette, and on the other side by a Lox511 site compatible with the one in the landing site. Following co-transfection of the donor BAC construct and a Cre recombinase expression plasmid into the acceptor cell line, the ensuing Cre-Lox recombination results in irreversible exchange of the neomycin cassette in the cell with the insert containing the *MICA* locus from the BAC.

B Diagram of BAC constructs used to generate modified isogenic cell lines by recombinase-mediated cassette exchange. Four different constructs were created: wild type (WT), deletion of the core standard promoter (SPdel), deletion of the core upstream promoter (UPdel) and insertion of a transcription terminator between the two promoters (Ter). The primary transcripts generated are shown below the constructs.

C, D Transgenic (C) or endogenous (D) MICA upstream and standard transcript expression measured by qPCR in modified isogenic cell lines carrying a transgenic 53-kb *MICA* locus. Deletion of the upstream promoter or insertion of a transcription terminator between the promoters led to increased expression of the transgenic MICA standard transcript; no effect was seen on the endogenous transcripts. Error bars represent standard deviations of multiple independently generated clones (n = 2–4, Appendix Table S1). NS, not significant; *$P < 0.05$, ***$P < 0.001$, Student's t-test.

E, F Flow cytometric analysis of transgenic MICA surface expression in isogenic cell lines carrying a transgenic 53-kb MICA locus. Bar chart of mean fluorescent intensity is shown in (E), and histograms of representative clones in (F). Consistent with in *cis* transcriptional interference, expression of the transgenic MICA protein was upregulated by deletion of the upstream promoter or insertion of a transcription terminator between the two promoters. Error bars represent standard deviations of multiple independently generated clones (n = 2–4, Appendix Table S1). *$P < 0.05$, ***$P < 0.001$, Student's t-test.

signal in this region, demonstrating that the majority of this signal arose from transcription initiation from the downstream promoter (Fig EV3D). Interposition of a transcription terminator between the two promoters caused a significant increase in the ChIP signal seen at the downstream promoter, indicating that run-through transcription from the upstream promoter exerts an inhibitory effect on transcription initiation at the downstream promoter (Fig EV3D). Overall, these data together confirm that the upstream promoter represses *MICA* expression from the standard promoter in *cis* through transcriptional interference.

To define the input–output characteristics of the *MICA* dual-promoter system, we created further sets of isogenic cell lines with the *MICA* gene engineered to be tuneable at the upstream or standard promoter (Fig 4A, Appendix Table S1). When the core upstream

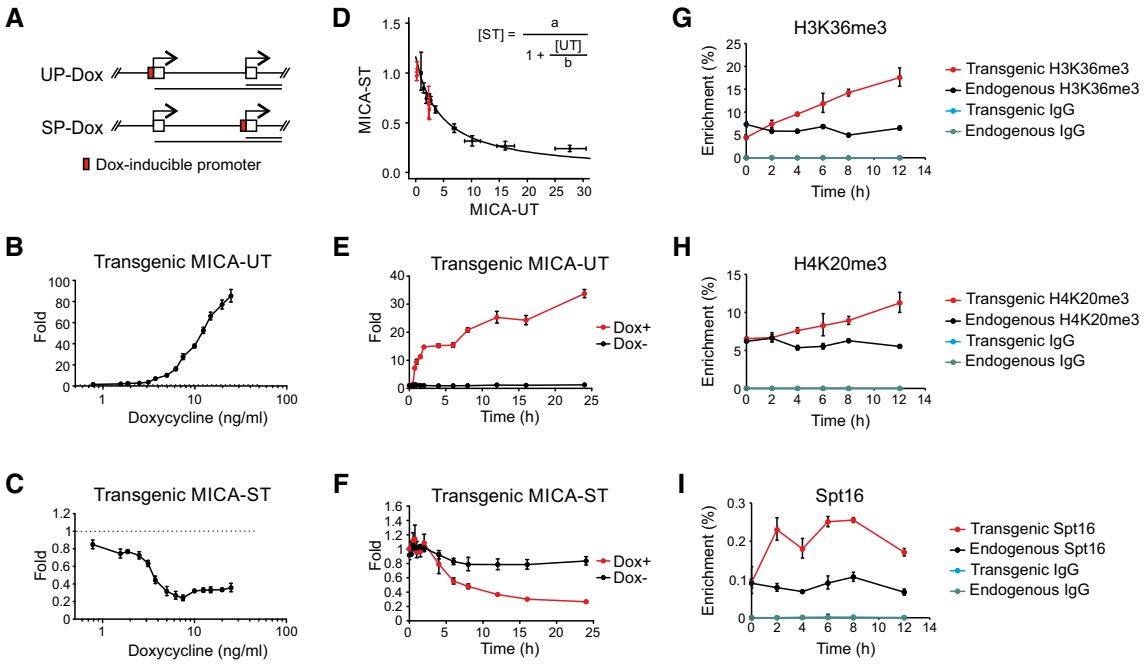

Figure 4. Characterization of intragenic transcriptional interference in MICA.

A Diagram of BAC constructs used to generate modified isogenic cell lines in which a doxycycline-inducible promoter was positioned to drive transcription of the upstream transcript (UP-Dox) or standard transcript (SP-Dox).

B, C Dose-dependent changes of transgenic upstream and standard transcript levels in isogenic cells in which expression of the upstream transcript are driven by doxycycline. Doxycycline induced expression of the transgenic upstream transcript (B) and caused a fourfold dose-dependent reduction in expression of the transgenic standard transcript (C). Error bars represent standard deviations of three replicates.

D Mathematical modelling of the dose-dependent regulation of expression of the standard transcript by the upstream promoter (UP-Dox). Data from doxycycline-treated cells (black) up to the dose with maximum transcriptional interference (0–7.5 ng/ml) fitted to a reciprocal function as shown. Data from wild-type clones and clones with deletion of the upstream promoter are represented in red. Error bars represent standard deviations of three replicates.

E, F Timecourse of change in expression of the transgenic upstream (E) or standard (F) transcript in an isogenic cell line in which expression of the upstream transcript is driven by doxycycline (7.5 ng/ml) or mock control. Error bars represent standard deviations of three replicates.

G–I ChIP analysis of H3K36me3 (G), H4K20me3 (H) and Spt16 (I) modifications at the transgenic and endogenous standard promoter over time following induction of the transgenic upstream transcript by doxycycline (7.5 ng/ml). Error bars represent standard deviations of three replicates.

promoter was replaced with a doxycycline-inducible promoter, doxycycline treatment induced dose-dependent upregulation of the transgenic upstream transcript, downregulation of the standard transcript in *cis* and downregulation of the encoded MICA protein (Figs 4B and C, and EV4B). As predicted, no *in trans* effect was observed on expression of the endogenous upstream or standard MICA transcripts (Fig EV4C). The wild-type upstream and standard promoters did not respond to doxycycline, confirming that repression of the transgenic standard promoter is caused directly by the induction of the upstream promoter with doxycycline (Fig EV4D). Mathematical analysis of the steady-state response curve shows that the transcriptional activity arising from the standard promoter is in a simple reciprocal relationship with the transcriptional activity arising from the upstream promoter (Fig 4D). Expression of the transgenic upstream or standard MICA transcripts in cells with the native upstream promoter or with upstream promoter deletions maps onto this response curve (Fig 4D). This shows that different upstream promoters, which drive quantitatively similar levels of transcription, give rise to similar levels of transcriptional interference; it is the strength of transcription from the upstream promoter, rather than the identity of the promoter, that drives the observed transcriptional interference. Timecourse analysis of upstream and standard transcript expression following doxycycline treatment demonstrates that transcription from the upstream promoter is rapidly followed by a shutdown of transcription from the standard promoter (Fig 4E and F). The decline in the steady-state level of the standard transcript is slower than the rise in the upstream transcript due to the longer half-life of the standard transcript. Replacement of the standard *MICA* promoter with a doxycycline-inducible promoter confers dose-dependent induction of the standard MICA transcript and cell surface MICA protein expression, but has no effect on expression of the upstream transcript (Fig EV4A, E and F). This confirms that transcriptional interference between the two promoters is unidirectional, and further excludes promoter competition as the mechanism for transcriptional interference. This observation reflects the unidirectional nature of the underlying transcription.

Transcriptional interference is synchronous with FACT recruitment

The process of transcription is associated with changes in histones along the path taken by the polymerase. We tested whether

deposition of candidate histone modifications was involved in transcriptional interference in the *MICA* gene. By combining *MICA* allele-specific restriction digestion and chromatin immunoprecipitation (ChIP), we were able to analyse independently a range of histone modifications at both the transgenic and endogenous *MICA* promoters in the sets of modified isogenic cell lines (Fig 3B). H3K36me3, a histone mark cotranscriptionally recruited by elongating polymerase and involved in silencing of cryptic transcription within gene bodies (Carrozza *et al*, 2005; Keogh *et al*, 2005), is decreased at the transgenic standard *MICA* promoter when transcriptional interference is inhibited by deletion of the upstream promoter or by interposition of a transcription terminator between the two promoters (Fig EV3E). The upstream promoter deletion and transcription terminator interposition both decrease the repressive mark H4K20me3 (Jorgensen *et al*, 2013; Fig EV3F) and increase the histone variant H2A.Z (Fig EV3J), but not pan-histone H3, H3K27ac or the promoter mark H3K4me3 (Fig EV3G–I). No changes were seen at the endogenous loci, demonstrating the in *cis* nature of the underlying mechanisms.

To further clarify the role of these markers in transcriptional interference in *MICA*, we studied dynamic changes in histone modifications over time following induction of transcriptional interference (Fig 4A). When transcriptional interference was induced by expression of the upstream transcript from a doxycycline-inducible promoter, there was progressive enrichment of H3K36me3 and H4K20me3 in *cis* at the transgenic standard *MICA* promoter (Fig 4G and H), but not of the other markers (Fig EV4G–J). However, the increases seen with H3K36me3 and H4K20me3 were gradual (Fig 4G and H), in contrast with the rapid induction of upstream transcript expression and the rapid onset of transcriptional interference (Fig 4E and F). This suggests that H3K36me3 and H4K20me3 deposition lags behind the transcriptional interference and, therefore, may not be acutely causative of the transcriptional interference.

The histone chaperone FACT plays a role in nucleosome remodelling during transcription (Belotserkovskaya *et al*, 2003), and ChIP analysis demonstrated that the signal for the Spt16 subunit of FACT at the transgenic standard downstream promoter was reduced by deletion of the upstream promoter or interposition of a transcription terminator between the two promoters (Fig EV3K). When expression of the upstream transcript was driven by the doxycycline promoter, the Spt16 signal rose rapidly at the downstream promoter (Fig 4I) with a timecourse that is synchronous with that of transcriptional interference itself and changes much faster than the changes in histone marks. This is consistent with a role in transcriptional interference for events involved in the nucleosome remodelling associated with elongation of the upstream transcript through the downstream standard promoter.

IFN-γ and IL-4 regulate MICA through transcriptional interference

Evolutionary conservation of the *MICA* tandem promoter arrangement suggests that transcriptional regulation of the upstream promoter and the associated transcriptional interference may have biological value. Bioinformatic analysis revealed a highly conserved binding site within the upstream promoter for the activating transcription factor interferon regulatory factor 1 (IRF1), and we analysed the possible regulatory function of this site (Fig 5A; Tanaka *et al*, 1993; Rettino & Clarke, 2013). ChIP assays demonstrated that IFN-γ treatment of primary human arterial endothelial cells induces binding of IRF1 to the upstream promoter and not to the standard promoter (Fig 5B). Specific binding to the IRF1 site was confirmed using supershift and competition electrophoretic mobility shift assays (EMSAs) for IRF1 (Fig 5C and Appendix Fig S1). Reporter assays confirmed that the upstream promoter, but not the standard promoter, was inducible by IFN-γ treatment of endothelial cells and that the induction was mediated through the IRF1 binding site (Fig 5D). Together, these data confirm that IFN-γ activates the upstream promoter through inducible binding of IRF1 to the conserved IRF1-binding site. Therefore, we hypothesized that activation of the upstream promoter through this site would act by in *cis* transcriptional interference to downregulate expression of the standard promoter and so of MICA protein expression. As predicted, IFN-γ treatment activated the upstream promoter leading to expression of the upstream transcript and downregulation of the standard MICA transcript and of protein expression, consistent with transcriptional interference (Fig 5E and F).

Conversely, further examination of the upstream promoter sequence revealed a conserved binding site for the transcription repressor E4BP4 (Fig 6A; Cowell *et al*, 1992). E4BP4 has been shown to play a role in the B-cell response to interleukin-4 (IL-4), which promotes immunoglobulin class switching in B cells (Tangye *et al*, 2002; Kashiwada *et al*, 2010). We hypothesized that induction of E4BP4 by IL-4 would downregulate activity of the upstream promoter and expression of the upstream transcript, so reducing transcriptional interference and increasing expression of the standard coding MICA transcript. Consistent with this, treatment of primary B cells with IL-4 increased binding of E4BP4 to the upstream promoter as evidenced by ChIP (Fig 6B), downregulated expression of the upstream transcript and induced expression of the standard coding *MICA* transcript and of MICA protein expression at the cell surface (Fig 6C–E). Reporter assays and EMSA studies using cells transfected with E4BP4 confirmed that E4BP4 represses the upstream promoter activity through specific binding to the E4BP4 site (Appendix Fig S2A and B). Across multiple individual donors, the magnitude of downregulation of the upstream transcript correlated reciprocally with the upregulation of the standard transcript (Fig 6F).

CAGE-seq data are consistent with transcriptional interference in other genes

The expression of transcripts from upstream promoters in many other human genes raises the possibility that transcriptional interference could be involved in transcriptional regulation in genes other than *MICA*. To survey this possibility, we analysed CAGE-seq data sets because CAGE-seq provides good clarity about which promoter a transcript arises from (Carninci *et al*, 2006). We identified promoters which had been experimentally validated (Dreos *et al*, 2013) and selected the subset of these promoters that formed tandem promoter systems. Data sets for analysis included CAGE-seq data sets from the FANTOM5 project of transcript expression at multiple time points following a stimulus (Arner *et al*, 2015). We determined the trajectory of expression over time in all transcripts that were expressed. For tandem promoter systems, we identified patterns of monotonic transcript expression that were consistent with

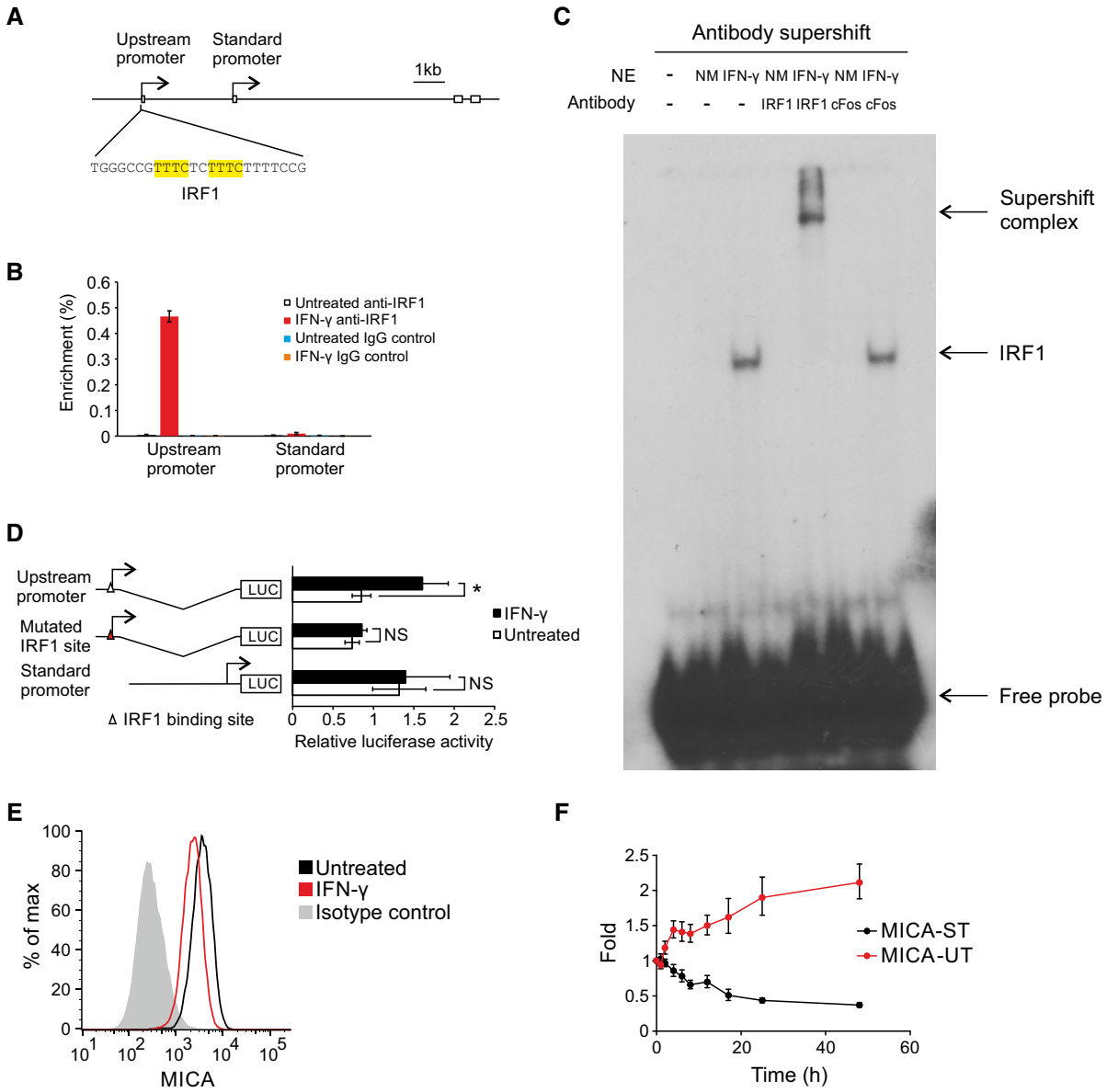

Figure 5. Regulation of MICA by IFN-γ through transcriptional interference.

A Diagram of the upstream promoter with the IRF1 binding site highlighted in yellow.

B ChIP analysis of the upstream promoter and standard promoter regions of the endogenous MICA locus for binding of the transcription factor IRF1 in primary human arterial endothelial cells following interferon-γ (IFN-γ) treatment. Interferon-γ causes substantial binding of IRF1 to the upstream promoter, but not to the standard promoter. Error bars represent standard deviations of three replicates.

C Electrophoretic mobility shift assay (EMSA) showing *in vitro* binding of IRF1 to the upstream promoter following interferon-γ treatment of primary human arterial endothelial cells. Nuclear extracts (NE) of cells treated with interferon-γ (IFN-γ) or untreated (NM) were pre-incubated with anti-IRF1 antibody or control anti-cFos antibody before the addition of ^{32}P-labelled probe containing the wild-type IRF1 binding site of the *MICA* upstream promoter.

D Reporter assays in primary human arterial endothelial cells demonstrate that the upstream promoter activity is increased by interferon-γ (IFN-γ) treatment and that this effect is abolished by mutation of the IRF1 binding site. The standard promoter does not contain any predicted IRF1 binding site and did not respond to interferon-γ. Error bars represent standard deviations of biological triplicates. NS, not significant; *$P < 0.05$, Student's *t*-test.

E Flow cytometry of MICA surface expression in primary human arterial endothelial cells treated with interferon-γ.

F Interferon-γ treatment of primary human arterial endothelial cells increased expression of the upstream transcript (red) over time and reduced expression of the standard transcript (black). Fold change over mock-treated control is shown. Error bars represent standard deviation of three replicates.

transcriptional interference—that is, an increase in the level of transcript arising from the upstream promoter associated with a fall in the level of transcript arising from the downstream promoter, or a fall in the level of transcript arising from the upstream promoter associated with an increase in the level of transcript arising from the downstream promoter (Table EV1). Across multiple human and mouse data sets, we found a substantial number of cases where there was a reciprocal pattern of transcript expression analogous to

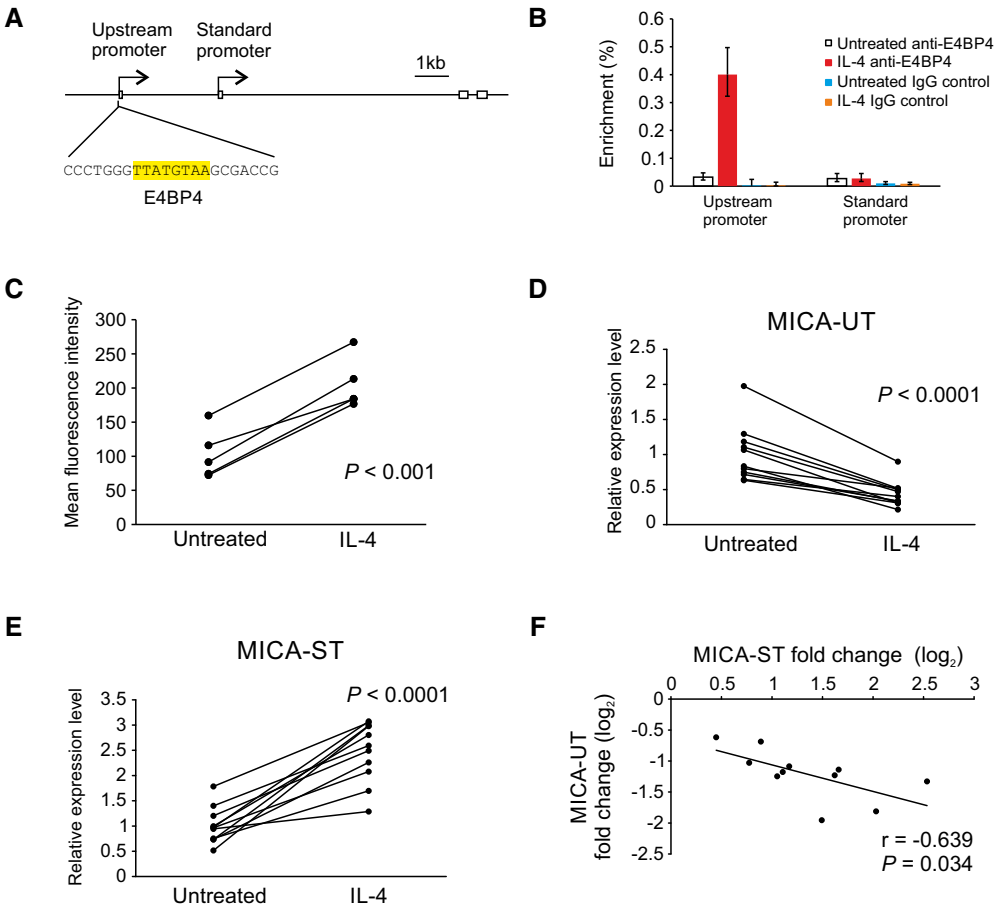

Figure 6. Regulation of MICA by IL-4 through transcriptional interference.

A Diagram of the upstream promoter with the E4BP4 binding site highlighted in yellow.

B ChIP analysis of the upstream promoter and standard promoter regions of the endogenous MICA locus for binding of the transcription factor E4BP4 in primary human B cells following interleukin-4 (IL-4) treatment. Interleukin-4 causes substantial binding of E4BP4 to the upstream promoter, but not to the standard promoter. Error bars represent standard deviations of three replicates.

C MICA surface expression on primary human $CD19^+$ B cells in PBMC from five different donors treated with interleukin-4 (IL-4) for 3 days. Cells were gated on the $CD3^-CD19^+$ population. Data shown are mean fluorescence intensity with the isotype control subtracted. *P*-value is from paired Student's *t*-test.

D–F Interleukin-4 (IL-4) treatment of primary human B cells from 11 donors reduced expression of the upstream transcript (D) and increased expression of the standard transcript (E). *P*-values are from paired Student's *t*-test. Linear regression (F) demonstrates a reciprocal correlation between changes in expression of the upstream transcript and of the standard transcript in response to interleukin-4 treatment. Pearson's correlation coefficient *r* and associated *P*-value are shown.

that seen with *MICA*. Examples of the transcript trajectories over time following a stimulus are illustrated in Fig 7. This analysis indicates that in higher eukaryotes there are multiple genes with tandem promoters displaying reciprocal patterns of transcript expression consistent with transcriptional interference.

Discussion

We studied the function of a set of tandem intragenic promoters in the transcriptional regulation of *MICA*. We found that transcription from the upstream promoter represses *MICA* expression through transcriptional interference. Transcriptional interference of *MICA* occurs in *cis* through transcriptional elongation from the upstream promoter over the downstream standard *MICA* promoter. The transcriptional interference is independent of the transcript generated from the upstream promoter. Deletion of the upstream core promoter or insertion of a transcription terminator downstream of the upstream promoter removes the transcriptional interference, but only in *cis*, demonstrating the requirement for overlapping transcription. Quantitative analysis, using an experimental system in which the upstream promoter activity is tuneable under the control of doxycycline, showed that the level of transcriptional interference observed was similar with either the native upstream promoter or with a heterologous promoter of equivalent strength. Therefore, transcription from an upstream promoter is sufficient for transcriptional interference; the degree of transcriptional interference depends on the strength of transcription from the upstream promoter, rather than the identity of the upstream promoter. Variation in the extent of transcriptional interference between different cell types will be influenced by differing transcription factor landscapes and so the strength of the transcriptional drive from each

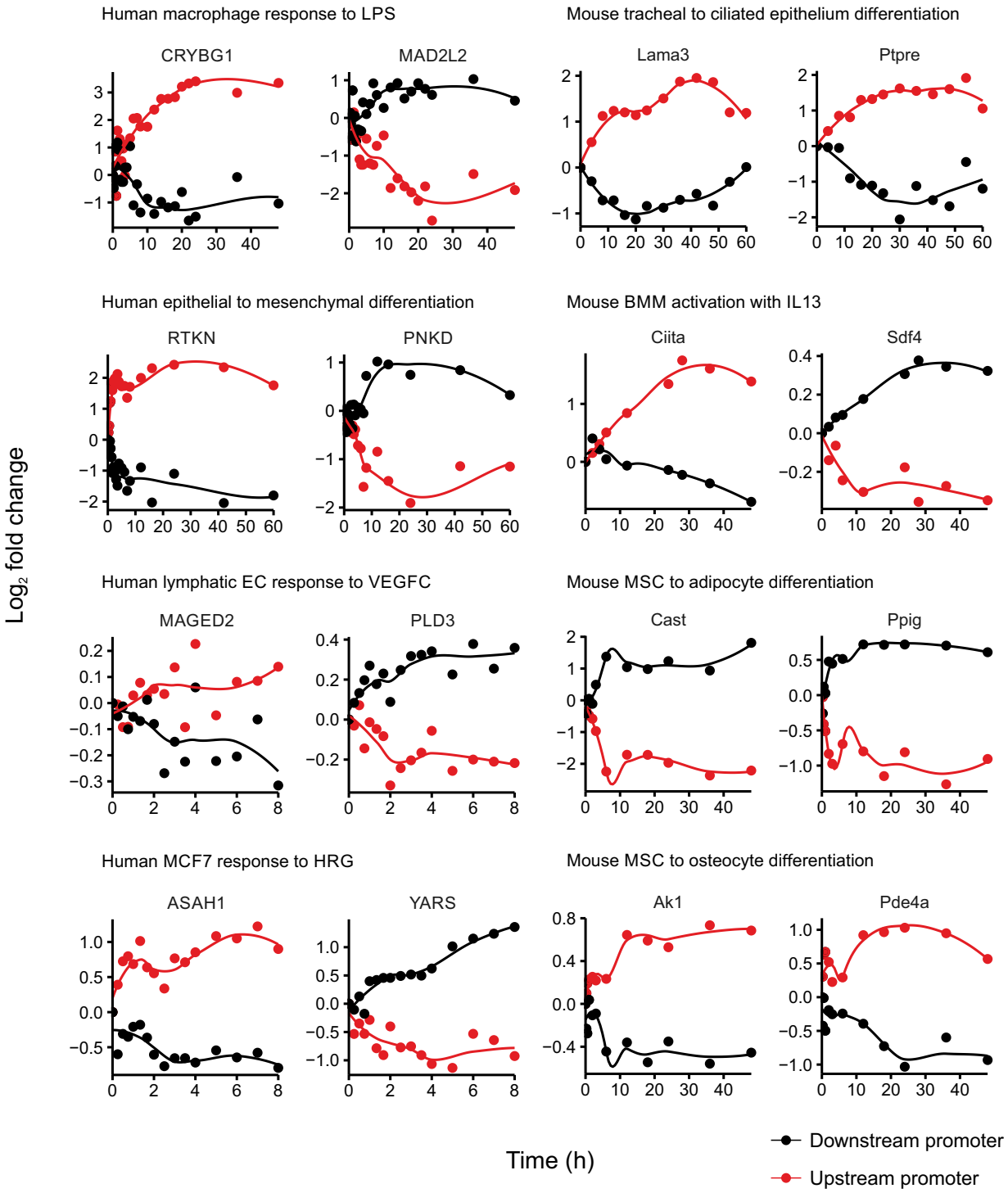

Figure 7. Tandem promoter transcript expression patterns consistent with transcriptional interference.

Transcript levels from tandem promoter pairs for genes from the FANTOM5 CAGE-seq data sets over time following the indicated stimuli are shown as $\log_2$ fold change, with the upstream promoter transcript in red and downstream promoter transcript in black. Coordinates of promoter windows used for counting expression levels are shown in Appendix Table S5. Reciprocal changes in expression from tandem promoter pairs are consistent with transcriptional interference.

promoter. Intragenic transcriptional interference thus constitutes a general mechanism available for regulation of human genes with tandem promoter configuration.

Current knowledge about the role of transcriptional interference in gene regulation in mammals is limited. The few known examples of mammalian transcriptional interference are developmentally or

ontologically regulated genes studied in model organisms. These genes are all controlled by tissue-specific promoters or enhancers that are subject to epigenetic modifications established during the cell differentiation programme, and some are special cases of genes that undergo developmental imprinting with irreversible epigenetic silencing due to antisense transcriptional interference (Abarrategui & Krangel, 2007; Racanelli *et al*, 2008; Latos *et al*, 2012; MacIsaac *et al*, 2012). In these examples, transcriptional interference has been demonstrated by genetic deletion of the interfering promoter or insertion of a transcription terminator. However, the long delay from genetic manipulation in stem cells to examining the effect of transcriptional interference in differentiated cells makes it difficult to establish the mechanism of transcriptional interference. In particular, it has been an open question whether overlapping transcription itself is sufficient for transcriptional interference without the involvement of the widespread epigenetic changes that occur during the cell differentiation process. It has also been unclear whether transcriptional interference can mediate real-time regulatory changes in gene expression.

Given the shortcoming of the chronic steady-state genetic approaches, we developed a cellular model with tuneable control of the interfering upstream promoter of *MICA* to study quantitative aspects of transcriptional interference in real time. Timecourse analysis of transcriptional activity from both upstream and standard downstream *MICA* promoters, as well as histone modification at the downstream promoter, show that transcriptional interference occurs rapidly without the involvement of histone modifications such as H3K36me3, previously known to be associated with transcriptional interference (Houseley *et al*, 2008). Therefore, overlapping transcription itself is sufficient to interfere with downstream promoter activity in real time for the *MICA* gene.

The timecourse of transcriptional interference is similar to that seen for changes in occupancy of the Spt16 subunit of the FACT histone chaperone at the downstream promoter. FACT plays a role in elongation of the polymerase through nucleosomes and, as the polymerase advances, is involved in both nucleosome destabilization and nucleosome reassembly (Belotserkovskaya *et al*, 2003). In yeast, FACT mutants are associated with cryptic transcription initiation, indicating a role for the FACT-mediated nucleosomal recovery, which follows polymerase nucleosomal navigation, in the inhibition of transcription initiation (Kaplan *et al*, 2003). During transcriptional interference, elongation of Pol II through the downstream standard promoter will be associated with nucleosomal reassembly in the wake of the polymerase and this reassembly may render the promoter relatively unfavourable for transcription initiation compared to the situation when there is no run-through transcription. This is consistent with a model of transcriptional interference in which transcription through the downstream promoter reduces the likelihood of initiation at that promoter (Fig 8).

We also used this tuneable system to study the stimulus–response relationship of transcriptional interference at equilibrium by varying the strength of the upstream promoter. The only previous attempt to quantitatively study eukaryotic transcriptional interference was in yeast and only assessed activity of the downstream promoter directly, leaving the quantitative relationship between downstream promoter activity and upstream promoter activity unclear (Buetti-Dinh *et al*, 2009). Our quantitative analysis demonstrates a simple reciprocal relationship between the activities of the two promoters, analogous to the classical repressor model that describes the repression of gene expression through binding of a transcription repressor to a single-copy binding site. This is consistent with transcriptional elongation over the downstream promoter as the physical basis for the transcriptional interference.

By virtue of its mechanism, the transcriptional interference that we have defined has distinct characteristics compared to other forms of gene regulation. Firstly, transcriptional interference is highly

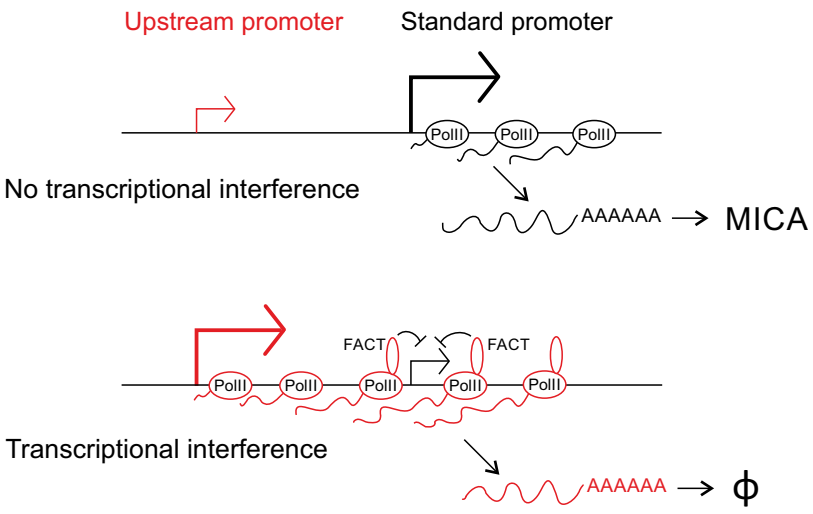

Figure 8. Model of transcription interference in *MICA*.

The upstream promoter of *MICA* encodes an unstable noncoding transcript which is subject to rapid degradation (φ), whereas the downstream promoter encodes the functional MICA protein. As illustrated in the lower panel, transcription from the upstream promoter inhibits transcription initiation at the downstream promoter and this transcriptional interference involves recruitment of the histone chaperone FACT.

specific in its effect. Information about upstream promoter activity is transmitted locally in *cis* through the elongating transcription complex directly to the downstream promoter. The specificity of transcriptional interference is endowed by the physical linkage and linear proximity of the two promoters and is thus restricted locally in *cis*. This use of the local linear gene structure does not require the transport—by diffusion or other means—of soluble trans-factors. In contrast, transcription factor or repressor or RNA-based gene regulation requires the remote production of the soluble mediator, its physical movement to the target site *in trans*, and then specific molecular interaction between the regulator and target molecules.

Secondly, transcriptional interference allows a very fast response to a regulatory input: information from the upstream promoter is transmitted rapidly along the DNA molecule at the speed of transcription. With trans-factors, there is the obligate time lag determined by the production or activation of the mediating factor and its transport to the site of action through either diffusion or active recruitment. Repressor-based transcription silencing may trigger long-term locuswide gene silencing through mechanisms such as heterochromatin formation with DNA methylation (Bintu *et al*, 2016). In contrast, the repressed state of transcriptional interference requires a strong and active upstream promoter within an active chromatin environment, which allows rapid upregulation of transcription from the downstream promoter when the transcriptional interference is reduced.

Thirdly, transcriptional interference allows the polarity of a regulatory input to be inverted, an input that increases transcription from the upstream promoter will reduce transcription from the downstream promoter. During evolution, a tandem promoter gene structure suitable for transcriptional interference may arise from a genetic duplication event or the insertion of transposable elements containing an upstream promoter. This can add sites for pre-existing transcription factors as new regulatory inputs to the gene and if transcriptional interference occurs, then the effect of these transcription factors on the downstream promoter—and so on transcription of the coding transcript—will be opposite to the effect on the upstream promoter itself. In this way, for a given gene, activating regulatory pathways can be converted into repressive pathways or *vice versa* without the need for new molecules to evolve. Consequently, the binding of a cis-acting transcription activator or repressor may have the opposite effect to that anticipated, if the binding affects the activity of an upstream promoter involved in transcriptional interference. This needs to be considered in bioinformatic and functional genomics studies of genetic variations associated with phenotypic traits or diseases in genomewide association studies (Hardy & Singleton, 2009; Price *et al*, 2015).

These characteristics of the transcriptional interference that we have identified are all pertinent to the evolution and function of *MICA*. Regulatory inputs which lower transcription from the upstream promoter can result in a rapid increase in transcription of the downstream coding transcript and so promote a prompt attack by immune cells expressing the NKG2D receptor (Bauer *et al*, 1999; Wu *et al*, 1999). A timely MICA response would be advantageous in defence against cancer or virus infection. The tandem promoter system in *MICA* allows for evolutionary selection within either promoter of binding sites for regulatory inputs, which allows expansion in the number and complexity of the physiological inputs that converge at the *MICA* promoter. Thus, a given regulatory input could drive transcription from either the upstream or the downstream promoter and so increase or decrease production of the downstream coding transcript. Such evolutionary selection at the *MICA* locus has no direct effects on the gene regulatory pathways themselves and so remains highly specific for *MICA*. Correct functional interpretation of the regulatory polymorphisms at the *MICA* locus is only possible if the polarity-reversing effect of transcriptional interference on regulatory inputs is incorporated.

We demonstrated the conversion of an activator into a repressor and *vice versa* in the regulation of *MICA* by IFN-γ and IL-4 through the IRF1 and E4BP4 transcription factors, respectively. Downregulation of *MICA* mRNA expression by IFN-γ has been observed previously (Zhang *et al*, 2008; Schwinn *et al*, 2009; Yadav *et al*, 2009), but the molecular mechanism was not clear. One report proposed that the miR-520b miRNA might mediate this effect by direct action on both the *MICA* promoter and the *MICA* 3′UTR, but blocking miR-520b failed to block IFN-γ induced MICA expression (Yadav *et al*, 2009). Our results demonstrate that IFN-γ acts via IRF1 to regulate *MICA* expression at the transcription level through transcriptional interference. Transcriptional interference alters the polarity of the IFN-γ effect from activator to repressor. Conversely, IL-4 acts via the transcriptional repressor E4BP4 to upregulate MICA. In this case, transcriptional interference inverts a negative gene regulation mechanism into one that positively upregulates expression of *MICA*.

The finding of intragenic transcriptional interference in *MICA* may lead to re-evaluation of the function of additional upstream promoters in other human genes. Previous studies have focused largely on the protein coding potential of the transcripts from these promoters, some of which may encode proteins with altered N-terminal sequence compared to the reference transcripts (Kimura *et al*, 2006). Our findings expand the scope of additional upstream promoter function. A tandem promoter gene structure results in promoter-specific exon 1 usage with potentially shared downstream sequence and strictly implies overlapping transcription. Overlapping transcription is sufficient to cause transcriptional interference in *MICA*, so tandem promoter gene structures in other genes may also have undergone evolutionary selection for their capacity to shape gene expression through transcriptional interference. For some of these genes, the spliced transcripts from the additional upstream promoter may be merely by-products of transcription. The production of these transcripts does impose an energetic cost, but their transcription is essential for transcriptional interference to occur. Early termination of the transcripts before they reach the downstream promoter would abolish the transcriptional interference. This requirement for their transcription will impose an evolutionary pressure to maintain them if the regulatory mechanisms that the upstream promoters confer have sufficient survival value. In *MICA*, the transcript arising from the upstream promoter is rapidly eliminated by the nonsense-mediated decay pathway, which will minimize any further unnecessary energetic expenditure, such as nonproductive protein translation.

Overall, these findings demonstrate that intragenic transcriptional interference plays a role in the transcriptional regulation of the *MICA* gene and is mediated by read-through transcription. These findings provide a roadmap for manipulating *MICA* expression for therapeutic benefit, especially in cancer immunotherapy. Intragenic transcriptional interference in an endogenous human gene represents a hitherto unrecognized and potentially important modality of

transcriptional regulation. A large number of human genes express transcripts from upstream promoters, so intragenic transcriptional interference may be widespread. The CAGE-seq analysis indicates that multiple human and mouse genes with tandem promoter systems display a pattern of reciprocal changes in transcript expression similar to that seen with transcriptional interference in *MICA*. This analysis demonstrates that transcriptional interference may be involved in the regulation of many genes in higher eukaryotes, which has implications for the study of transcriptional regulation in such genes. A tandem promoter system allows evolutionary variation in the *MICA* gene to capture a wide range of activating transcription factor pathways to upregulate standard MICA expression through the standard promoter or to downregulate it through the upstream promoter. Intragenic transcriptional interference actuates an integration of two opposing sets of inputs, one set from each promoter. The evolutionary conservation of the tandem promoter configuration in this gene and many other human genes attests to its likely survival value.

Materials and Methods

Reagents

Chemicals were obtained from Sigma, and enzymes for molecular biology from NEB, unless otherwise stated. Interferon-γ and interleukin-4 were from eBioscience. Primers were from Invitrogen or Integrated DNA Technologies. Primer sequences are listed in Appendix Table S2.

Cell lines and primary cells

HeLa, HT1080, 293T, T47D cell lines and human skin fibroblasts were maintained in Dulbecco's Modified Eagle Medium (Sigma) supplemented with 10% foetal calf serum (First Link) (D10). Human aortic arterial endothelial cells (Invitrogen) were maintained in Medium 200 with Low Serum Growth Supplement (Invitrogen). Human iPSC cells SFC840-03-03 (Fernandes *et al*, 2016) were cultured in mTeSR1 (StemCell Technologies) on plates coated with hESC-qualified Matrigel Matrix (BD Biosciences). Cell lines were confirmed to be free of mycoplasma contamination using MycoAlert Mycoplasma Detection Kit (Lonza). Peripheral blood mononuclear cells (PBMCs) were enriched from fresh human blood by Ficoll density centrifugation. Prior written consent was obtained from healthy blood donors (Ethical approval: South Central-Hampshire B, reference 13/SC/0392). Primary B cells were isolated from PBMC by MACS using CD19 MicroBeads (Miltenyi Biotec). The purity of B cells was consistently over 95% as determined by flow cytometry using anti-CD19-APC (eBioscience 17-0199-73) and anti-CD20-FITC (AbD Serotec MCA1822F) antibodies. Purified PBMC or B cells were cultured in RPMI1640 medium with 10% foetal calf serum. Primary monocytes and T cells were isolated from PBMC by MACS using CD14 or CD3 MicroBeads (Miltenyi Biotec), respectively.

Plasmid construction

BAC modification/retrofitting plasmids were constructed by standard cloning techniques using conditional replicating pSG80A plasmid in the pir116 strain. Plasmid pBS-Landing carrying the landing site for site-directed recombinase-mediated cassette exchange (RMCE) was constructed by sequential insertion of functional cassettes into the pBluescript plasmid. Luciferase reporter plasmids were constructed by cloning the MICA standard or upstream promoter into the pGL3 Basic or pGL4.10 plasmid. pHR-SIN-rtTA3 the lentiviral plasmid for expression of the rtTA3 doxycycline-responsive transactivator was constructed by cloning rtTA3 into pHR-SIN-BX-IRES-Emerald. CRISPR plasmids were constructed based on the PX458 plasmid (Ran *et al*, 2013). dCas9-based targeted transcription activation plasmids were constructed based on the UniSam plasmid with mCherry replaced with EGFP (Fidanza *et al*, 2017). The source of plasmids and detailed steps of plasmid construction are listed in Appendix Table S3. The inserts for all of the plasmids were verified by sequencing.

BAC modification

BAC modification was carried out by lambda red-based recombineering. Briefly, pRed/ET plasmid was first transformed into DH10B cells harbouring the target BAC to make the cells recombineering-proficient. Markerless mutations were then introduced into the BAC using an rpsL-neo-based positive/negative selection system, by first tagging the mutation site with an rpsL-neo cassette using a PCR product with 50-bp homologous arms and selection with kanamycin. The rpsL-neo cassette was then replaced using linear DNA fragments cut from plasmids that carry the intended mutations and 500- to 1,000-bp homology arms with streptomycin counterselection. BAC truncation, concatenation and retrofitting steps were carried out by recombineering using linearized plasmid or BAC fragments containing bacterial selection markers and over 50-bp homology arms. Following recombineering, temperature-sensitive pRed/ET plasmid was removed by plating cells at 37°C.

In total, reporters with six different 53-kb BAC inserts and four different 161-kb BAC inserts were constructed. The 53-kb BAC reporters were generated from BAC CH501-248L24, and the 161-kb BAC reporters from BAC CH501-248L24 and CH501-181B23 (BACPAC Resources Center, CHORI). Both CH501-248L24 and CH501-181B23 are derived from the PGF cell line, which carries a MICA*00804 haplotype. All the BAC constructs were engineered to contain synonymous mutations in the MICA exon 2 coding region to facilitate discrimination of transgenic and endogenous *MICA* by allele-specific PCR. Appendix Table S4 lists the detailed steps of the BAC modifications. Each step was verified by restriction digest and pulsed-field gel electrophoresis using the CHEF II system (Bio-Rad). Markerless mutations introduced into the BAC were verified by sequencing. Endotoxin-free BAC DNA was purified using the Phaseprep BAC DNA extraction kit (Sigma).

Generation of acceptor cell line for BAC RMCE

The acceptor cell line carrying a single-copy landing site was generated by transient transfection of ScaI-linearized pBS-Landing plasmid into HT1080 cells by nucleofection using Cell Line Nucleofector Kit V and Nucleofector II Device (Lonza), followed by selection with 500 μg/ml G418 (Calbiochem). Single clones were picked using cloning cylinders, expanded and screened first by PCR to check the integrity of the landing site, then by Southern blot using three

different restriction enzymes to check the copy number of the insert. One clone designated HT1080-L3N9 was confirmed to carry a single-copy landing site and was used for the subsequent generation of BAC RMCE clones. The stable HT1080-L3N9 acceptor cell line was maintained in D10 medium with 250 μg/ml G418.

Generation of site-directed BAC RMCE clones

BAC RMCE clones were generated by co-transfection of BAC and pCre-Pac plasmid into the HT1080-L3N9 acceptor cell line using GeneJuice transfection reagent (Merck) followed by selection with 250 μg/ml hygromycin (Calbiochem). Single clones were picked using cloning cylinders (Sigma), expanded, and genomic DNA was extracted using the ZR-96 Genomic DNA Kit (Zymo Research) for verification by PCR. PCR-verified clones were then screened by flow cytometry to exclude hyperploid clones based on forward and side-scattering properties and propidium iodide staining for DNA content. On average, 8% of picked clones successfully passed the whole screening process (Appendix Table S1). At least two independent BAC RMCE clones were generated for each BAC construct. Stable BAC RMCE clones were maintained in D10 medium with 125 μg/ml hygromycin. For functional assays, clones were cultured in hygromycin-free medium for at least 48 h before the start of the experiments. Transgenic MICA expression in BAC RMCE clones was stable and homogeneous without hygromycin selection for at least 2 months. BAC clones were confirmed to be free of mycoplasma contamination using MycoAlert Mycoplasma Detection Kit.

Lentivirus production and infection

Lentivirus was generated by co-transfection of the lentiviral expression plasmid with pMD2.G and psPax2 packaging plasmids into 293T cells. The supernatant was harvested, filtered through a 0.4-μm filter (Millipore) and titrated in HT1080 cells. For the generation of doxycycline-inducible cells, BAC clones were infected with pHR-SIN-rtTA3 lentivirus at an MOI of 10 and expanded and used between 5 and 7 days post-infection. The rtTA3 transactivator expression is stable with near 100% unimodal expression as indicated by GFP expression from the bicistronic construct.

RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE)

RLM-RACE was carried out using HeLa and HT1080 total RNA and the ExactStart Eukaryotic mRNA 5′- & 3′-RACE Kit (Epicentre) following the manufacture's protocol with the following modifications. A longer RNA ligation oligo GCUGAUGGCGAUGAAUGAA CACUGCGUUUGCUGGCUUUGAUGAAA was used in the ligation reaction, and reverse transcription was carried out using random hexamers (Qiagen) and BioScript reverse transcriptase (Bioline). The RACE product was amplified by nesting PCR using primers CO4210/606 followed by CO4211/1109, and cloned using the Strataclone PCR cloning kit (Stratagene) for sequencing.

RNA analysis

Total RNA was prepared using the Trizol plus kit (Invitrogen) with the on-column DNase digestion step to remove genomic DNA. Total bovine and porcine adult kidney RNA (Zyagen), and total human adult kidney RNA (Clontech) were cleaned up using the Purelink RNA Mini kit (Invitrogen) with DNase digestion. Total RNA was reverse transcribed into cDNA using BioScript reverse transcriptase and random hexamers. Where indicated, oligo-dT was used instead of random hexamers. Human adult and foetal tissue cDNA panels were from Clontech. qPCR was carried out using the Fast SYBR Green Master Mix and StepOne real-time PCR system (Applied Biosystems). qPCR results were analysed using the $\Delta\Delta C_t$ method and normalized to GAPDH expression for gene expression studies. Semi-quantitative RT–PCR was carried out using Biotaq polymerase (Bioline).

The following primers were used for SYBR green-based qPCR analysis: MICA-ST, CO3706/3707; MICA-UT, CO3705/3708; GAPDH, CO3744/3745; 18S rRNA, CO3746/3747; ITGB5, CO3740/3741. For qPCR analysis of endogenous and transgenic MICA transcripts in BAC RMCE clones, the following primers were used: transgenic MICA-ST, CO3526/3525; transgenic MICA-UT, CO3705/3525; endogenous MICA-ST, CO3526/4080; endogenous MICA-UT, CO3705/4080. For semi-quantitative RT–PCR analysis of tissue samples, the following primers were used: MICA-ST, CO1111/1109; MICA-UT, CO1099/1109; GAPDH, CO631/632; POLR2F, CO1831/1832. For semi-quantitative RT–PCR analysis of MIC homologs in human, pig and cow kidney tissues, the following primers were used: human MICA-ST, CO1111/1109; human MICA-UT, CO1099/1109; pig MIC2-ST, CO5217/5221; pig MIC2-UT, CO5219/5221; cow MIC1-ST, CO5218/5222; cow MIC1-UT, CO5220/5222; human, pig or cow GAPDH, CO5223/5224. Primer sequences are listed in Appendix Table S2.

mRNA stability measurement

mRNA stability was measured by either global transcription inhibition using actinomycin D, or metabolic labelling with 4-thiouridine (4sU, Carbosynth). For measurements based on actinomycin D treatment, cells were treated with 5 μM actinomycin D, and MICA-ST or MICA-UT mRNA was quantified independently by qPCR using cDNA prepared from the same amount of total RNA.

Measurement of mRNA half-life by 4sU metabolic labelling was based on the protocol described previously (Dolken *et al*, 2008). Briefly, total and 4sU-labelled RNA were isolated from proliferating cells pulsed with 500 μM 4sU for 2 h, and reverse transcribed into cDNA using BioScript reverse transcriptase and random hexamers. Enrichment of 4sU-labelled RNA was measured by qPCR. The half-life of mRNA measured by 4sU labelling is influenced by the combined effects of mRNA decay and dilution due to cell proliferation. The half-life was calculated using 18S rRNA as an internal control, assuming that mRNA decay follows first-order kinetics and that the half-life of stable 18S rRNA is dominated by cell proliferation.

Polysome fractionation

Polysome fractionation was carried out based on protocols described previously (Powley *et al*, 2009) with the following modifications: heparin in the sucrose gradient and cell lysis buffer was replaced with RNasin RNase inhibitor (Promega) at 40 or 100 U/ml, respectively. RNA was isolated using Trizol-LS (Invitrogen) and Purelink RNA Mini kit for RNA analysis.

Flow cytometry

Flow cytometry was carried out as described previously (Lin *et al*, 2012). The pan-allelic anti-MICA antibody clone 2C10 (Santa Cruz sc-23870) was used for measuring MICA surface expression in general. The allele-specific anti-MICA clone 159227 (R&D Systems MAB1300) was used to detect transgenic MICA surface expression in RMCE BAC clones. This antibody is an allele-specific antibody that recognizes the transgenic MICA*008 allele carried by the BAC, but not the endogenous MICA*007 allele of HT1080 cells. For flow cytometry of B cells, cells were blocked with Fc receptor-blocking reagent (Miltenyi Biotec) and stained with anti-CD19-APC (eBioscience 17-0199-73), anti-CD3-PE-Cy7 (BD Biosciences 557851), anti-MICA-PE (Santa Cruz sc-23870 PE) and the viability dye LIVE/DEAD® Fixable Near-IR stain (Invitrogen); the viable $CD3^{-}CD19^{+}$ B-cell population was gated on for analysis of MICA expression. Flow cytometry was performed using a BD FACSCanto system (Becton Dickinson), and data were analysed using FlowJo software (FlowJo. LLC).

Western blot

Cells were lysed in lysis buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% CHAPS) with 1× protease inhibitor cocktail (Roche), and proteins were separated by SDS–PAGE and transferred to Immobilon-P PVDF membrane (Millipore). Myc-tagged proteins were detected using anti-c-myc (clone 9E10) antibody and goat anti-mouse IgG-HRP secondary antibody (Abcam ab20043). HRP-conjugated anti-beta actin antibody (Abcam ab49900) was used as control. HRP signals were detected using ECL Prime reagent (GE Healthcare) and a ChemiDoc MP imaging system (Bio-Rad).

siRNA knockdown

Cells were transfected with Silencer Select siRNA targeting MICA-UT (target sequence GCAGUGGCGCCUAAAGUCU) or Silencer Select Negative Control siRNA #1 (Ambion) using Oligofectamine Reagent (Invitrogen).

Electrophoretic mobility shift assay (EMSA)

EMSAs were carried out as described previously (Lin *et al*, 2012). For EMSA of IRF1, nuclear extracts were prepared from HAECs treated with 20 ng/ml interferon-γ for 24 h. Wild-type probe containing the IRF-1 binding site of the MICA upstream promoter was prepared by annealing of oligonucleotides CO3611/3612 and mutant probe CO3615/3616. Cold competition assays were carried out by pre-incubation of nuclear extract with 100× excess of unlabelled cold probe. For supershift assays, nuclear extract was pre-incubated with 2 μg anti-IRF1 antibody (Santa Cruz sc-497X) or anti-c-Fos control antibody (Santa Cruz sc-52X) for 30 min on ice before adding ^{32}P-labelled probe. EMSA for E4BP4 was carried out using nuclear extracts from 293T cells transfected with plasmids expressing N- or C- terminal myc-tagged E4BP4 or control ATF2 transcription factor, and ^{32}P-labelled probes prepared by annealing oligonucleotides CO2717/2718.

Reporter assay

Reporter assays were carried out by co-transfection of firefly luciferase reporter plasmid carrying MICA promoter fragments and control pRL-SV Renilla luciferase plasmid, followed by cell lysis and luciferase assay using the Dual-Luciferase Reporter Assay System (Promega) and a TD-2020 luminometer (Turner Designs). 293T, HeLa and HT1080 cells were transfected using GeneJuice and primary human arterial endothelial cells were transfected using HCAEC Nucleofector Kit (Lonza), and cells were lysed 48 h post-transfection. Firefly luciferase reporters were constructed in pGL3-based plasmids, with the exception of the E4BP4 experiment, in which pGL4-based plasmids were used. This was due to concern about the presence of an E4BP4 site within the luciferase gene of the pGL3 plasmid. In the interferon-γ experiment, the upstream promoter reporters carry −78 bp upstream promoter (MICA-UT-P-78 bp and MICA-UT-P-78 bp-ISREmut) and the standard promoter reporter carries −2,779 bp standard promoter. In the E4BP4 experiment, the upstream promoter reporters carry −702 bp upstream promoter (MICA-UT-P-702 bp and MICA-UT-P-702 bp-E4BP4mut). Results represented as relative luciferase activity have been normalized to Renilla luciferase activity and pGL3P promoter control plasmid luciferase activity for pGL3-based reporters, or pGL4.23 control plasmid activity for pGL4-based reporters.

Determination of MICA promoter haplotypes

The haplotypes of the *MICA* promoter in selected cell lines were determined by PCR sequencing to facilitate identification of suitable restriction enzymes for allele-specific PCR analysis of the standard or upstream *MICA* promoters in ChIP assays and fluorescent PCR assay. For HT1080 cells homozygous for the MICA*007 allele, the promoter haplotype was assembled directly (deposited as GenBank KF724603). For the primary human fibroblasts heterozygous for MICA*004/010 alleles, the promoter haplotype linked to each allele was determined by sequencing of the −6-kb *MICA* promoter region of the fibroblasts as well as selected EBV-transformed B cells homozygous for MICA*004 or MICA*010 alleles (IHWG Cell and DNA Bank). The promoter haplotype linked to the MICA*004 allele in the fibroblasts was deposited as GenBank KF724624, and MICA*010 as KF724587.

Chromatin immunoprecipitation

Allele-specific ChIPs were carried out based on the Q2ChIP protocol (Dahl & Collas, 2007). The following antibodies were used: H3K36me3 (Abcam ab9050), H3K27ac (Abcam ab4729), H3K4me3 (Abcam ab8580), H4K20me3 (Abcam ab9053), H2A.Z (Abcam ab4174), H3 (Abcam ab1791), Spt16 (clone 8D2, BioLegend 607002), PolII phospho-Ser5 (clone 1H4B6, Millipore MABE954), and normal rabbit IgG control (Santa Cruz sc-2027X), mouse IgG2a isotype control (eBioscience 14-4724-85), and rat IgG2b isotype control (clone RTK4530, BioLegend 400601). Phosphatase inhibitor cocktail PhosSTOP (Roche) was included in the buffers for Pol II phospho-Ser5 ChIP. ChIP DNA was purified using the PCR purification kit (Qiagen). For allele-specific amplification of the *MICA* standard promoter region from the transgenic *MICA* allele in RMCE

clones, ChIP and input control samples were digested with BfaI to disrupt the endogenous MICA*007 allele promoter region that differs from the transgenic *MICA* at SNP rs116135464T, which is sensitive to BfaI. For allele-specific amplification of the same region from the endogenous allele, samples were treated with TspRI to disrupt the transgenic MICA*008 allele that differs from the endogenous *MICA* at the same SNP. The digestion efficiencies of BfaI and TspRI exceed 99% as assessed by qPCR analysis of ChIP input samples prepared from HT1080 cells that carry only the endogenous MICA*007 allele or 293T cells that carry only the transgenic MICA*008 allele, respectively. BfaI- or TspRI-treated samples were analysed by qPCR using primer pair CO6351/6358 for the *MICA* standard promoter.

ChIP for IRF1 was carried out similarly using anti-IRF1 antibody (Santa Cruz sc-497X) or rabbit IgG control (Santa Cruz sc-2027X) on human arterial endothelial cells treated with 20 ng/ml interferon-γ for 24 h. ChIP for E4BP4 was carried out using anti-E4BP4 antibody (Santa Cruz sc-9550X) or goat IgG control (Santa Cruz sc-2028) on primary human B cells treated with 20 ng/ml IL-4 for 24 h. ChIP and input control samples were digested with BseRI to disrupt homologous sequences in the genome similar to the *MICA* upstream promoter. BseRI-treated samples were then analysed by qPCR using primer pair CO6696/6699 for the *MICA* upstream promoter or CO6351/6358 for the standard promoter.

Transient CRISPR

Primary human arterial endothelial cells or fibroblasts were transfected with CRISPR plasmids targeting the *MICA* upstream promoter (pair 1: PX458-C40/C42 or pair 2: PX458-C41/C43), or *HLA-B* exon1 (PX458-C50/C52) or *PDPN* exon2 (PX458-C14) as negative controls by nucleofection using the HCAEC Nucleofector Kit or NHDF Nucleofector Kit, respectively (Lonza). Cells were cultured for 3 days before being lifted for flow cytometric analysis of surface MICA expression, or for cell sorting of GFP-positive cells directly into Trizol-LS using a MoFlo MLS sorter (Beckman Coulter) for RNA analysis.

Stable CRISPR and allele-specific DNA analysis

Primary fibroblasts were transfected by nucleofection with CRISPR pairs PX458-C41/43 targeting deletion of the *MICA* upstream promoter. Cells were expanded, stained with anti-MICA antibody (clone 2C10) followed by goat anti-mouse IgG Alexa Fluor 647 secondary antibody (Invitrogen A-21236), and sorted into MICA-high and MICA-low populations based on MICA surface expression using a MoFlo MLS sorter. Sorted cells were further expanded, and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). For analysis of allele-specific MICA upstream promoter deletion, genomic DNA was treated with BsrDI, specific for rs2596539T within the upstream promoter linked to the MICA*010 allele, or BsgI, specific for rs2596539C linked to the MICA*004 allele, for analysis of the MICA*004 and MICA*010 alleles, respectively. BsrDI- or BsgI-treated DNA samples were used for PCR amplification using CO6392 and 5′ 6-FAM-labelled CO6395, and 6-FAM-labelled PCR samples were mixed with GeneScan 500 LIZ size standard and analysed using a 3730xl DNA Analyzer and Peak Scanner software (Applied Biosystems).

dCas9-based transcriptional activation

293T cells were transfected with EGFP-based UniSam plasmids targeting the *MICA* upstream promoter (UniSamG-MICAUT), *MICA* standard promoter (UniSamG-MICAST), or *CD43* (UniSamG-CD43) or *CD36* (UniSamG-CD36) promoter as negative controls using GeneJuice. Cells were harvested 2 days post-transfection for flow cytometric analysis of surface MICA expression.

Genomewide analysis of CAGE-seq data

Sets of transcriptional start sites for the hg19 and mm9 reference genomes were downloaded from the Eukaryotic Promoter Database (Dreos *et al*, 2013) and extended to create 200-bp promoter windows centred on each transcriptional start site. Each promoter window was associated with the nearest UCSC knownGene annotated gene. Overlapping and bookended windows were merged using BEDTools (Quinlan & Hall, 2010), giving a total of 25,718 and 21,119 promoter windows covering 17,842 and 17,564 genes in the human and mouse genomes, respectively. FANTOM5 CAGE-seq timecourse data sets were downloaded from the DNA Data Bank of Japan (accession numbers DRA000991, DRA002711, DRA002747, and DRA002748) and analysed for promoter window tag counts using featureCounts (Liao *et al*, 2014; Arner *et al*, 2015). Raw counts were converted to relative log expression-normalized counts per million (CPM) using edgeR (Anders & Huber, 2010; Robinson *et al*, 2010), and libraries with a median normalized $\log_2$CPM < -1 were excluded from the analysis. Mean expression at each time point was calculated from three biological replicates in R, and only promoters with a minimum expression of 1 CPM were retained. The Kendall rank correlation coefficient was calculated for each promoter's expression over time. Promoters were paired by gene associations and then filtered to include only pairs with a positive correlation coefficient for one promoter and a negative coefficient for the other. Expression values were converted to $\log_2$ (fold change) compared to time 0, and diverging expression changes over time were visualized for each promoter pair passing the filter. All plots were generated in R (R Core Team, 2015).

Data availability

Sequences for *MICA* promoter haplotypes have been submitted to Genbank with accession numbers KF724603, KF724624 and KF724587.

Expanded View for this article is available online.

Acknowledgements

We are grateful to Dr Ben Davies for advice and assistance with BAC modification, to Dr Drew Worth for assistance with cell sorting, to Dr Matthew Cockman for advice and assistance with polysome fractionation, and to Professor Nicholas Proudfoot for providing the βΔ5-7 plasmid carrying the minimal beta-globin transcription terminator. This work was supported by the Medical Research Council (G116/165) and the Novo Nordisk Foundation (Grant Number NNF15CC0018346) and National Institute for Health Research Oxford Comprehensive Biomedical Research Centre Program.

Author contributions

DL and CAO designed the experiments. DL performed the experiments and TKH carried out CAGE-seq analysis. DL and CAO wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Abarrategui I, Krangel MS (2007) Noncoding transcription controls downstream promoters to regulate T-cell receptor alpha recombination. *EMBO J* 26: 4380 – 4390

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11: R106

Ard R, Tong P, Allshire RC (2014) Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. *Nat Commun* 5: 5576

Ard R, Allshire RC (2016) Transcription-coupled changes to chromatin underpin gene silencing by transcriptional interference. *Nucleic Acids Res* 44: 10619 – 10630

Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drablos F, Lennartsson A, Ronnerblad M, Hrydziuszko O, Vitezic M, Freeman TC, Alhendi AM, Arner P, Axton R, Baillie JK, Beckhouse A, Bodega B, Briggs J, Brombacher F, Davis M *et al* (2015) Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* 347: 1010 – 1014

Bahram S, Bresnahan M, Geraghty DE, Spies T (1994) A second lineage of mammalian major histocompatibility complex class I genes. *Proc Natl Acad Sci USA* 91: 6259 – 6263

Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, Spies T (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* 285: 727 – 729

Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D (2003) FACT facilitates transcription-dependent nucleosome alteration. *Science* 301: 1090 – 1093

Bintu L, Yong J, Antebi YE, McCue K, Kazuki Y, Uno N, Oshimura M, Elowitz MB (2016) Dynamics of epigenetic regulation at the single-cell level. *Science* 351: 720 – 724

Buetti-Dinh A, Ungricht R, Kelemen JZ, Shetty C, Ratna P, Becskei A (2009) Control and signal processing by transcriptional interference. *Mol Syst Biol* 5: 300

Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y *et al* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38: 626 – 635

Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ, Anderson S, Yates J, Washburn MP, Workman JL (2005) Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* 123: 581 – 592

Cerboni C, Zingoni A, Cippitelli M, Piccoli M, Frati L, Santoni A (2007) Antigen-activated human T lymphocytes express cell-surface NKG2D ligands via an ATM/ATR-dependent mechanism and become susceptible to autologous NK- cell lysis. *Blood* 110: 606 – 615

Cowell IG, Skinner A, Hurst HC (1992) Transcriptional repression by a novel member of the bZIP family of transcription factors. *Mol Cell Biol* 12: 3070 – 3077

Dahl JA, Collas P (2007) Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. *Stem Cells* 25: 1037 – 1046

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang TH (2008) The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet* 24: 167 – 177

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Roder M, Kokocinski F, Abdelhamid RF, Alioto T *et al* (2012) Landscape of transcription in human cells. *Nature* 489: 101 – 108

Dolken L, Ruzsics Z, Radle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, Koszinowski UH (2008) High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* 14: 1959 – 1972

Dreos R, Ambrosini G, Cavin Perier R, Bucher P (2013) EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Res* 41: D157 – D164

Fantom Consortium and the RIKEN PMI and CLST (2014) A promoter-level mammalian expression atlas. *Nature* 507: 462 – 470

Fernandes HJ, Hartfield EM, Christian HC, Emmanoulidou E, Zheng Y, Booth H, Bogetofte H, Lang C, Ryan BJ, Sardi SP, Badger J, Vowles J, Evetts S, Tofaris GK, Vekrellis K, Talbot K, Hu MT, James W, Cowley SA, Wade-Martins R (2016) ER stress and autophagic perturbations lead to elevated extracellular alpha-synuclein in GBA-N370S Parkinson's iPSC-derived dopamine neurons. *Stem Cell Reports* 6: 342 – 356

Fidanza A, Lopez-Yrigoyen M, Romano N, Jones R, Taylor AH, Forrester LM (2017) An all-in-one UniSam vector system for efficient gene activation. *Sci Rep* 7: 6394

Greger IH, Aranda A, Proudfoot N (2000) Balancing transcriptional interference and initiation on the GAL7 promoter of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 97: 8415 – 8420

Groh V, Bahram S, Bauer S, Herman A, Beauchamp M, Spies T (1996) Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc Natl Acad Sci USA* 93: 12445 – 12450

Gummalla M, Maeda RK, Castro Alvarez JJ, Gyurkovics H, Singari S, Edwards KA, Karch F, Bender W (2012) abd-A regulation by the iab-8 noncoding RNA. *PLoS Genet* 8: e1002720

Hainer SJ, Pruneski JA, Mitchell RD, Monteverde RM, Martens JA (2011) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* 25: 29 – 40

Hardy J, Singleton A (2009) Genomewide association studies and human disease. *N Engl J Med* 360: 1759 – 1768

Harlen KM, Churchman LS (2017) The code and beyond: transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nat Rev Mol Cell Biol* 18: 263 – 273

Hongay CF, Grisafi PL, Galitski T, Fink GR (2006) Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* 127: 735 – 745

Houseley J, Rubbi L, Grunstein M, Tollervey D, Vogelauer M (2008) A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. *Mol Cell* 32: 685 – 695

Hug N, Longman D, Caceres JF (2016) Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res* 44: 1483 – 1495

Jorgensen S, Schotta G, Sorensen CS (2013) Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res* 41: 2797 – 2806

Kaplan CD, Laprade L, Winston F (2003) Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301: 1096 – 1099

Kasahara M, Sutoh Y (2015) Comparative genomics of the NKG2D ligand gene family. *Immunol Rev* 267: 72 – 87

Kashiwada M, Levy DM, McKeag L, Murray K, Schroder AJ, Canfield SM, Traver G, Rothman PB (2010) IL-4-induced transcription factor NFIL3/E4BP4 controls IgE class switching. *Proc Natl Acad Sci USA* 107: 821 – 826

Keogh MC, Kurdistani SK, Morris SA, Ahn SH, Podolny V, Collins SR, Schuldiner M, Chin K, Punna T, Thompson NJ, Boone C, Emili A, Weissman JS, Hughes TR, Strahl BD, Grunstein M, Greenblatt JF, Buratowski S, Krogan NJ (2005) Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* 123: 593 – 605

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T *et al* (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16: 55 – 65

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517: 583 – 588

Kumar V, Kato N, Urabe Y, Takahashi A, Muroyama R, Hosono N, Otsuka M, Tateishi R, Omata M, Nakagawa H, Koike K, Kamatani N, Kubo M, Nakamura Y, Matsuda K (2011) Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat Genet* 43: 455 – 458

Lanier LL (2015) NKG2D receptor and its ligands in host defense. *Cancer Immunol Res* 3: 575 – 582

Latos PA, Pauler FM, Koerner MV, Senergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science* 338: 1469 – 1472

Le Clerc S, Delaneau O, Coulonges C, Spadoni JL, Labib T, Laville V, Ulveling D, Noirel J, Montes M, Schachter F, Caillat-Zucman S, Zagury JF (2014) Evidence after imputation for a role of MICA variants in nonprogression and elite control of HIV type 1 infection. *J Infect Dis* 210: 1946 – 1950

Li Z, Groh V, Strong RK, Spies T (2000) A single amino acid substitution causes loss of expression of a MICA allele. *Immunogenetics* 51: 246 – 248

Li P, Morris DL, Willcox BE, Steinle A, Spies T, Strong RK (2001) Complex structure of the activating immunoreceptor NKG2D and its MHC class I-like ligand MICA. *Nat Immunol* 2: 443 – 451

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30: 923 – 930

Lin D, Lavender H, Soilleux EJ, O'Callaghan CA (2012) NF-kappaB regulates MICA gene transcription in endothelial cell through a genetically inhibitable control site. *J Biol Chem* 287: 4299 – 4310

MacIsaac JL, Bogutz AB, Morrissy AS, Lefebvre L (2012) Tissue-specific alternative polyadenylation at the imprinted gene Mest regulates allelic usage at Copg2. *Nucleic Acids Res* 40: 1523 – 1535

Martens JA, Wu PY, Winston F (2005) Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. *Genes Dev* 19: 2695 – 2704

McCarthy MT, Moncayo G, Hiron TK, Jakobsen NA, Valli A, Soga T, Adam J, O'Callaghan CA (2017) Purine nucleotide metabolism regulates expression of the human immune ligand MICA. *J Biol Chem* 293: 3913 – 3924

Okada Y, Han B, Tsoi LC, Stuart PE, Ellinghaus E, Tejasvi T, Chandran V, Pellett F, Pollock R, Bowcock AM, Krueger GG, Weichenthal M, Voorhees JJ, Rahman P, Gregersen PK, Franke A, Nair RP, Abecasis GR, Gladman DD, Elder JT *et al* (2014) Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *Am J Hum Genet* 95: 162 – 172

Palmer AC, Egan JB, Shearwin KE (2011) Transcriptional interference by RNA polymerase pausing and dislodgement of transcription factors. *Transcription* 2: 9 – 14

Petruk S, Sedkov Y, Riley KM, Hodgson J, Schweisguth F, Hirose S, Jaynes JB, Brock HW, Mazo A (2006) Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* 127: 1209 – 1221

Powley IR, Kondrashov A, Young LA, Dobbyn HC, Hill K, Cannell IG, Stoneley M, Kong YW, Cotes JA, Smith GC, Wek R, Hayes C, Gant TW, Spriggs KA, Bushell M, Willis AE (2009) Translational reprogramming following UVB irradiation is mediated by DNA-PKcs and allows selective recruitment to the polysomes of mRNAs encoding DNA repair enzymes. *Genes Dev* 23: 1207 – 1220

Price AL, Spencer CC, Donnelly P (2015) Progress and promise in understanding the genetic basis of common diseases. *Proc Biol Sci* 282: 20151684

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841 – 842

R Core Team (2015) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing

Racanelli AC, Turner FB, Xie LY, Taylor SM, Moran RG (2008) A mouse gene that coordinates epigenetic controls and transcriptional interference to achieve tissue-specific expression. *Mol Cell Biol* 28: 836 – 848

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8: 2281 – 2308

Rettino A, Clarke NM (2013) Genome-wide identification of IRF1 binding sites reveals extensive occupancy at cell death associated genes. *J Carcinog Mutagen (Spec Iss Apoptosis)* pii: S6-009.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139 – 140

Robinson J, Halliwell JA, Hayhurst JD, Flicek P, Parham P, Marsh SG (2015) The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* 43: D423 – D431

Schwinn N, Vokhminova D, Sucker A, Textor S, Striegel S, Moll I, Nausch N, Tuettenberg J, Steinle A, Cerwenka A, Schadendorf D, Paschen A (2009) Interferon-gamma down-regulates NKG2D ligand expression and impairs the NKG2D-mediated cytolysis of MHC class I-deficient melanoma by natural killer cells. *Int J Cancer* 124: 1594 – 1604

Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference–a crash course. *Trends Genet* 21: 339 – 345

Tanaka N, Kawakami T, Taniguchi T (1993) Recognition DNA sequences of interferon regulatory factor 1 (IRF-1) and IRF-2, regulators of cell growth and the interferon system. *Mol Cell Biol* 13: 4531 – 4538

Tangye SG, Ferguson A, Avery DT, Ma CS, Hodgkin PD (2002) Isotype switching by human B cells is division-associated and regulated by cytokines. *J Immunol* 169: 4298 – 4306

Ullrich E, Koch J, Cerwenka A, Steinle A (2013) New prospects on the NKG2D/NKG2DL system for oncology. *Oncoimmunology* 2: e26097

Venkataraman GM, Suciu D, Groh V, Boss JM, Spies T (2007) Promoter region architecture and transcriptional regulation of the genes for the MHC class I-related chain A and B ligands of NKG2D. *J Immunol* 178: 961 – 969

Wang X, Hou J, Quedenau C, Chen W (2016) Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol Syst Biol* 12: 875

Wu J, Song Y, Bakker AB, Bauer S, Spies T, Lanier LL, Phillips JH (1999) An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* 285: 730 – 732

Yadav D, Ngolab J, Lim RS, Krishnamurthy S, Bui JD (2009) Cutting edge: down-regulation of MHC class I-related chain A on tumor cells by IFN-gamma-induced microRNA. *J Immunol* 182: 39 – 43

Zhang C, Niu J, Zhang J, Wang Y, Zhou Z, Zhang J, Tian Z (2008) Opposing effects of interferon-alpha and interferon-gamma on the expression of major histocompatibility complex class I chain-related A in tumors. *Cancer Sci* 99: 1279 – 1286

Zhang J, Liao D, Yang L, Hou S (2016) Association between functional MICA-TM and Behcet's disease: a systematic review and meta-analysis. *Sci Rep* 6: 21033

Zhou X, Wang J, Zou H, Ward MM, Weisman MH, Espitia MG, Xiao X, Petersdorf E, Mignot E, Martin J, Gensler LS, Scheet P, Reveille JD (2014) MICA, a gene contributing strong susceptibility to ankylosing spondylitis. *Ann Rheum Dis* 73: 1552 – 1557

Zou Y, Bresnahan W, Taylor RT, Stastny P (2005) Effect of human cytomegalovirus on expression of MHC class I-related chains A. *J Immunol* 174: 3098 – 3104

Purine nucleotide metabolism regulates expression of the human immune ligand MICA

Received for publication, August 3, 2017, and in revised form, December 12, 2017 Published, Papers in Press, December 26, 2017, DOI 10.1074/jbc.M117.809459

Michael T. McCarthy[‡], Gerard Moncayo[‡], Thomas K. Hiron[‡], Niels A. Jakobsen[‡], Alessandro Valli[§], Tomoyoshi Soga[¶], Julie Adam[‡§], and Christopher A. O'Callaghan[‡1]

From the [‡]Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, United Kingdom, the [§]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, United Kingdom, and the [¶]Institute for Advanced Biosciences, Keio University, 246-2 Mizukami, Kakuganji, Tsuruoka, Yamagata 997-0052, Japan

Edited by Luke O'Neill

Expression of the cell-surface glycoprotein MHC class I polypeptide-related sequence A (MICA) is induced in dangerous, abnormal, or "stressed" cells, including cancer cells, virus-infected cells, and rapidly proliferating cells. MICA is recognized by the activating immune cell receptor natural killer group 2D (NKG2D), providing a mechanism by which immune cells can identify and potentially eliminate pathological cells. Immune recognition through NKG2D is implicated in cancer, atherosclerosis, transplant rejection, and inflammatory diseases, such as rheumatoid arthritis. Despite the wide range of potential therapeutic applications of MICA manipulation, the factors that control MICA expression are unclear. Here we use metabolic interventions and metabolomic analyses to show that the transition from quiescent cellular metabolism to a "Warburg" or biosynthetic metabolic state induces MICA expression. Specifically, we show that glucose transport into the cell and active glycolytic metabolism are necessary to up-regulate MICA expression. Active purine synthesis is necessary to support this effect of glucose, and increases in purine nucleotide levels are sufficient to induce MICA expression. Metabolic induction of MICA expression directly influences NKG2D-dependent cytotoxicity by immune cells. These findings support a model of MICA regulation whereby the purine metabolic activity of individual cells is reflected by cell-surface MICA expression and is the subject of surveillance by NKG2D receptor-expressing immune cells.

The capacity to recognize and eliminate harmful elements while safeguarding healthy "self" cells is a key characteristic of the mature human immune system. This capacity arises, in large part, through the successful passage of lymphocytes through a series of immune self-tolerance mechanisms. Immune self-tolerance poses a challenge to the immune recognition of damaged or stressed self cells. The circumvention of such tolerance mechanisms using PD1-PDL1 inhibition has led to major recent clinical success with cancer immune therapies (1). The natural killer group 2D (NKG2D)[2] receptor-ligand system functions to overcome tolerance and eliminate stressed self cells even in the presence of functioning self-tolerance (2).

In humans, the NKG2D ligands are a group of eight cell-surface proteins, of which the highly polymorphic major histocompatibility class I chain-related protein A (MICA) is the most extensively studied. Generally, healthy quiescent cells do not express ligands for NKG2D (3, 4), but cell-surface expression of MICA is switched on by events such as malignant transformation (5), viral infection (6, 7), proliferative cell activation (8, 9), and some pro-inflammatory stimuli (10, 11). MICA-expressing cells are then targeted by immune cells (including natural killer (NK) cells, NKT cells, and cytotoxic $CD8^+$ T cells) that express the activating NKG2D receptor; this interaction can result in direct cytotoxicity (12), co-stimulation (6, 13, 14), or cytokine secretion (15), depending on the context.

Evidence from observational human studies and mouse models implicates immune recognition through NKG2D in a range of autoimmune and inflammatory conditions, including rheumatoid arthritis (16), inflammatory bowel disease (17), and atherosclerosis (11). The expression of MICA on human allogeneic organ transplants has been linked to transplant rejection (18), and advanced-stage cancers are associated with loss of MICA expression (5). Hence, the ability to manipulate MICA expression *in vivo* has many potential clinical applications: up-regulation of MICA could promote cancer immunity, and down-regulation could be beneficial in autoimmune disease or transplantation.

This work was supported by the Medical Research Council Grants G116/165 and G0901998, the National Institute for Health Research Oxford Comprehensive Biomedical Research Centre Program, and Novo Nordisk Foundation Grant NNF15CC0018346. The authors declare that they have no conflicts of interest with the contents of this article.

This article contains Figs. S1–S6.

[1] To whom correspondence should be addressed: Nuffield Dept. of Clinical Medicine, Henry Wellcome Bldg. for Molecular Physiology, University of Oxford, Roosevelt Dr., Oxford OX3 7BN, United Kingdom. E-mail: chris.ocallaghan@ndm.ox.ac.uk.

[2] The abbreviations used are: NKG2D, natural killer group 2D; NK, natural killer; DON, 6-diazo-oxonorleucine; MHC, major histocompatibility complex; MICA, MHC class I polypeptide-related sequence A; TLR, T cell receptor; TCA, trichloroacetic acid; HEK, human embryonic kidney; eGFP, enhanced green fluorescent protein; GIME, glucose-induced MICA expression; HAT, hypoxanthine, aminopterin, and thymidine; AICA-Rt, 5-aminoimidazole-4-carboxamide ribonucleotide; AICA-Rs, AICA ribonucleoside; CE-TOFMS, capillary electrophoresis time-of-flight mass spectrometry; LC-TOFMS, liquid chromatography time-of-flight mass spectrometry; CMV, cytomegalovirus; TCR, T cell receptor; RPMI, Roswell Park Memorial Institute 1640 culture medium; PI, propidium iodide; CFDA-SE, carboxyfluorescein diacetate succinimidyl ester; 2DG, 2-deoxyglucose; CFSE, carboxyfluorescein diacetate succinimidyl ester.

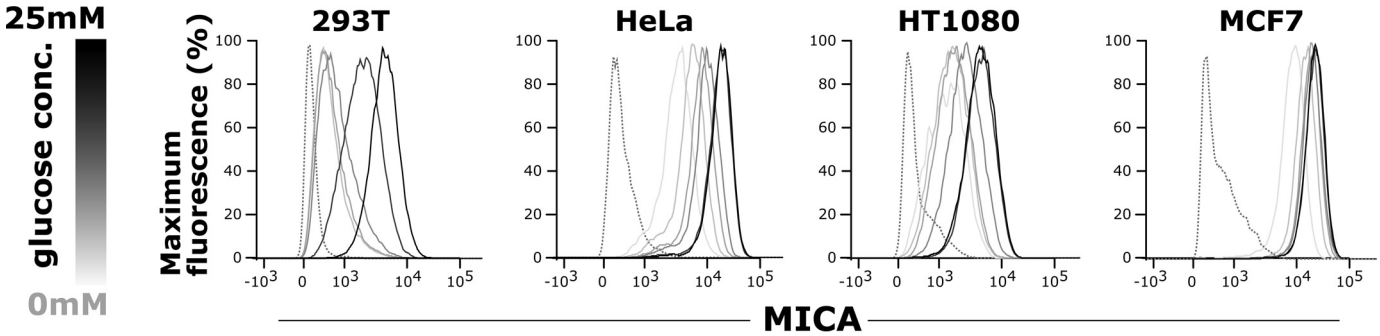

Figure 1. Glucose induces MICA expression. 293T (human embryonic kidney), HeLa (cervical cancer), HT1080 (fibrosarcoma), and MCF7 (breast cancer) cells were cultured for 48 h in medium containing 5 mM glucose that was then replaced with fresh medium containing either 0, 2.5, 5, 12.5, or 25 mM glucose. The cells were cultured for a further 48 h in these conditions before cell-surface MICA expression was measured by flow cytometry. MICA expression rose with the glucose concentration. The *dotted histogram* represents the isotype control sample, and the glucose concentration is indicated by the *gray scale*.

Multiple factors have been associated with changes in MICA expression, including activation of the DNA damage response pathway (19), Toll-like receptor (TLR) stimulation (10), histone deacetylation (20), heat shock transformation (21), ionizing radiation (22), growth factor pathway activation (23), cell-surface shedding (24), and microRNA expression (25). In addition, a number of gene-regulatory elements and transcription factors are known to play a role in MICA induction (11, 26). However, an integrated understanding of the mechanisms determining MICA expression remains elusive.

MICA expression in human primary cells or tissue samples is found in settings independently associated with high metabolic activity (increased glucose uptake, glycolysis, high lactate output, and proportionate reduction in TCA cycle metabolism, or "Warburg metabolism" (27–31)). This state of "activated metabolism" can be considered as a biosynthetic state, where enhanced glycolytic flux generates intermediate substrates for biomolecule synthesis (32). High-energy purine nucleotides, such as ATP, are among the downstream products.

Here, we show that glucose metabolism leading to the generation of high-energy purine nucleotides, a process at the core of the Warburg effect, induces cell-surface expression of MICA. We demonstrate that MICA induction by high-energy purine nucleotides is associated with increased NKG2D-dependent cellular immunogenicity and susceptibility to NK cell cytotoxicity, supporting our hypothesis that NKG2D provides a mechanism for immune oversight of metabolically activated cells.

Results

Glucose induces MICA expression

We hypothesized that the transition from quiescent to "activated" or Warburg metabolism plays an important role in NKG2D ligand induction. To test this hypothesis, we used glucose restriction to model quiescent *versus* activated metabolism and observed a direct correlation between the glucose concentration of culture medium and cell-surface expression of MICA in human embryonic kidney (HEK)-293T cells, cervical cancer cells (HeLa), fibrosarcoma cells (HT1080), and breast cancer cells (MCF7) (Fig. 1).

We undertook further characterization in HEK-293T cells (Fig. 2, *A*–*M*). In keeping with a Warburg phenotype, cells cultured in high (25 mM) glucose concentrations consumed more glucose and produced more lactate than cells cultured in low (5 mM) glucose (Fig. 2, *A* and *B*). Compared with cells cultured in 5 mM glucose, intracellular glucose concentrations increased 2.13-fold ($p < 0.05$) (Fig. 2*C*), and total RNA concentrations were higher ($p < 0.005$) in cells cultured in 25 mM glucose (Fig. 2*D*). The difference in cellular proliferation rate was ~2.2-fold across a range of glucose concentrations, but unlike MICA expression, the relationship was not directly proportional with maximum proliferation rates observed at intermediate glucose concentrations (Fig. 2*E*). The distribution of cell-cycle phases was not significantly altered by glucose concentration (Fig. 2, *F* and *G*). We used an inducible lentiviral eGFP construct to confirm that the capacity for gene expression remained intact in low glucose conditions (Fig. 2, *H* and *I*). Detected endogenous MICA levels varied directly with glucose concentration in permeabilized cells, suggesting that the observed effect of glucose is not due to a change in the distribution of MICA between different cellular compartments (Fig. 2, *J* and *K*). Changes in osmolality did not influence MICA cell-surface levels (Fig. 2, *L* and *M*). Glucose-induced MICA expression was observed at the cell surface in the presence of the matrix metalloprotease inhibitor MMPI III, demonstrating that the effects we observe are not due to changes in MICA shedding (Fig. S1, *A* and *B*). In addition, MICA mRNA transcript levels were increased over 4-fold in cells cultured in 25 mM glucose (Fig. S1*C*).

Glucose transport and metabolism are necessary for MICA induction

Next we investigated how glucose might alter MICA expression. Glucose is primarily transported into cells through the GLUT family of transporters (33). Two inhibitors of these transporters, cytochalasin B and phloretin, prevented glucose-induced MICA expression (GIME) in a dose-dependent manner (Fig. 3, *A* and *B*). Mannose, an alternative substrate for glycolysis transported via the same GLUT transporters, also caused MICA expression (Fig. 3*C*). A third glycolytic substrate, fructose, had a limited effect on MICA expression in untransfected HEK-293T cells (Fig. 3*C*), but fructose-induced MICA expression was observed in HEK-293T cells transfected with GLUT5, the fructose transporter (Fig. 3, *D* and *E*). 2-Deoxyglucose is a glucose analogue readily transported into the cell and phosphorylated by hexokinase to generate 2-deoxyglucose 6-phosphate. Whereas 2-deoxyglucose 6-phosphate inhibits

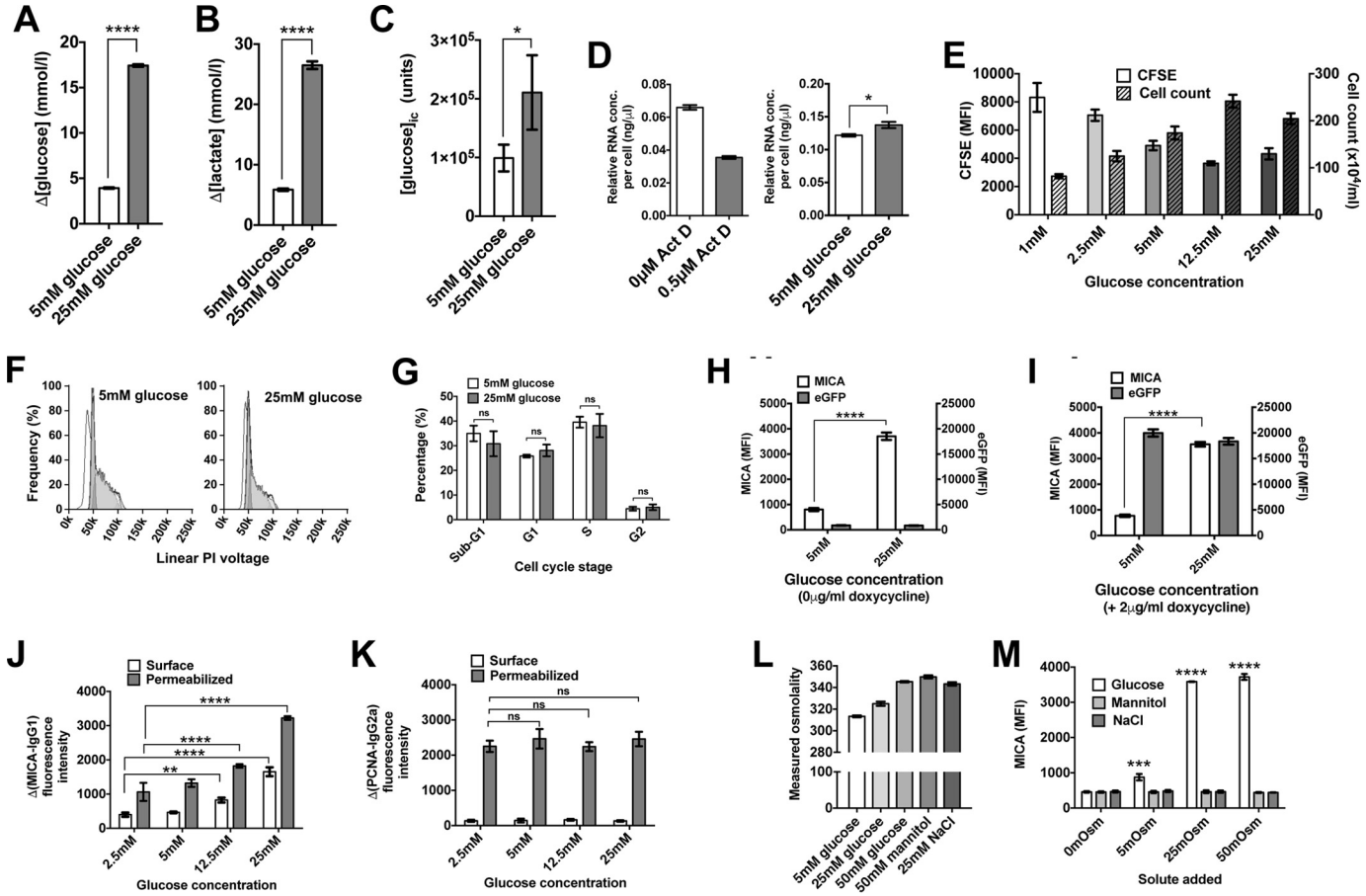

Figure 2. Glucose restriction replicates quiescent to activated metabolism transition. *A*, 293T cells cultured in low (5 mM) glucose consumed less glucose than cells cultured in high (25 mM) glucose ($p < 0.0001$). *B*, lactate production was ~4 times higher in cell cultured in high glucose compared with low glucose ($p < 0.0001$). *C*, intracellular glucose concentrations measured by nanoflow LC-MS were 2.13-fold higher in cells cultured in high glucose ($p < 0.05$). *D*, the RNA concentration per cell, reduced by culture with the RNA synthesis inhibitor actinomycin D (*ActD*), is 13% higher in cells cultured in 25 mM glucose compared with cells cultured in 5 mM glucose ($p < 0.005$). *E*, cellular proliferation measured by cell counting or CFSE dilution was maximal at intermediate glucose concentrations, unlike MICA expression. *F*, cell-cycle profiles were similar for cells cultured in 5 or 25 mM glucose. *G*, no significant difference was observed between cell-cycle phases in these conditions. *H*, MICA expression is induced by high (25 mM) glucose in cells transfected with inducible eGFP ($p < 0.0001$), but eGFP itself is not induced by high glucose. *I*, the induction of eGFP expression by doxycycline is equally efficient at low (5 mM) and high (25 mM) glucose. *J*, total cell MICA levels measured by flow cytometry of permeabilized cells is proportional to glucose concentration and cell-surface MICA. *K*, proliferating cell nuclear antigen (*PCNA*), used as a control for adequate permeabilization, is expressed independently of glucose concentration. *L*, glucose, mannitol, and sodium chloride each raise culture serum osmolality. *M*, change in culture medium osmolality does not affect MICA expression. *Error bars*, 95% confidence interval.

downstream enzymes in proximal glycolysis, it can induce the expression of genes directly regulated by glucose 6-phosphate (34). However, we found that unlike glucose, 2-deoxyglucose did not induce the expression of MICA (Fig. 3*F*) and inhibited its expression in high-glucose conditions (Fig. 3*G*), an effect that was not overcome by the addition of mannose (Fig. 3*H*). Together, these experiments suggest that the effect of glucose on MICA expression is dependent on the transport of glucose into the cell and its metabolism through glycolysis.

Purine nucleotides are necessary for glucose-induced MICA expression and sufficient to induce MICA expression

Glycolysis produces many intracellular metabolites. The proximal metabolites of glycolysis are essential for *de novo* nucleotide synthesis (Fig. 4). We hypothesized that nucleotide synthesis might mediate GIME. Because the *de novo* synthesis of the purine nucleobase is directly dependent on the supply of proximal glycolytic metabolites, we first tested this hypothesis by treating cells cultured in high glucose (25 mM) with two inhibitors of *de novo* purine synthesis, 6-diazo-oxo-norleucine (DON) and azaserine. Both compounds prevented GIME (Fig. 5, *A* and *B*). The specificity of this effect on *de novo* purine synthesis was tested using hypoxanthine, aminopterin, and thymidine (HAT)-selected cells. Whereas cells grown in standard culture medium depend on *de novo* purine synthesis, HAT-selected cells use the salvage pathway exclusively for new purine nucleotide synthesis. Azaserine inhibited GIME only in cells grown in standard culture medium and had no effect on HAT-selected cells (Fig. 5, *C* and *D*), consistent with a specific action on *de novo* purine synthesis. DON probably has additional off-target inhibitory effects. The addition of a purine salvage pathway substrate to azaserine-treated cells in high glucose caused dose-dependent MICA expression (Fig. 5*E*), confirming that purine synthesis is necessary for and can control GIME. This finding was reproducible with a range of salvage pathway substrates, including the purine nucleobases adenine and guanine, purine nucleosides including inosine, and deoxynucleosides (Fig. S2, *A*–*C*).

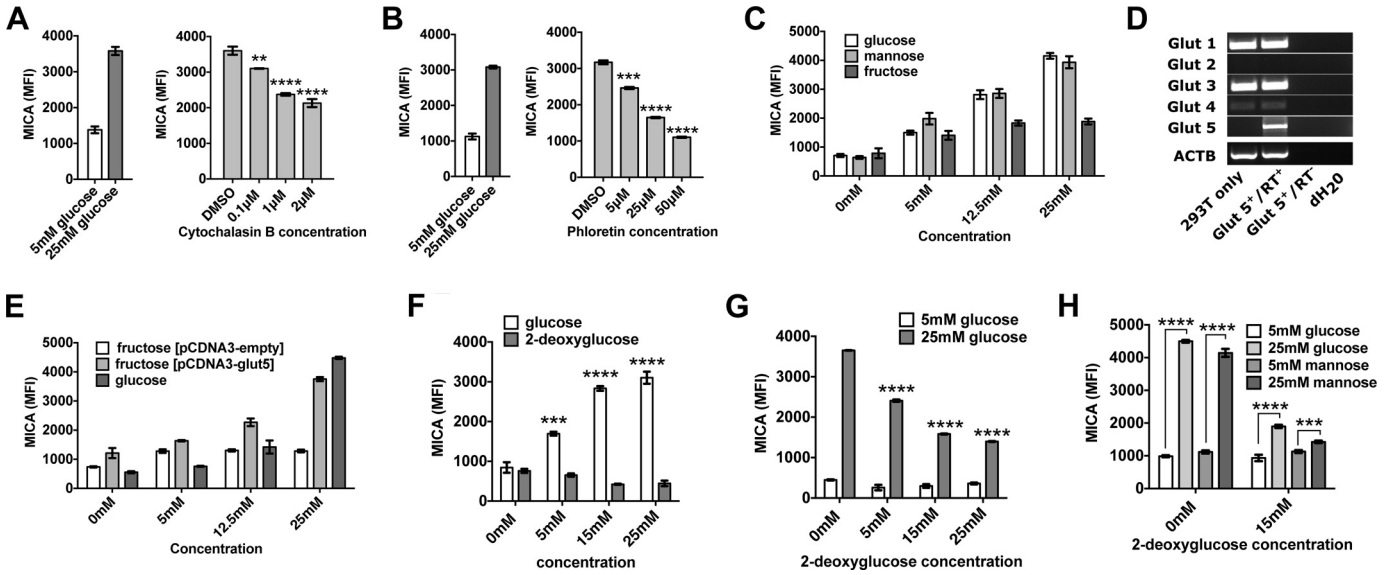

Figure 3. The transport of glycolytic substrates into the cell is necessary for their effect on MICA expression. *A*, cytochalasin B, an inhibitor of GLUT-mediated glucose transport, inhibits cell-surface MICA expression in HEK-293T cells cultured in 25 mM glucose. *B*, phloretin, a second GLUT transporter inhibitor, also reduces MICA expression in 25 mM glucose. *C*, mannose, an alternative substrate for glycolysis, induces MICA expression. The ability of fructose to induce MICA expression is limited in normal cells. *D*, mRNA for the fructose transporter GLUT5 is not detected in HEK-293T cells but is present in GLUT5-transfected cells. *E*, fructose-induced MICA expression is observed in GLUT5-transfected cells. *F*, the glucose analogue 2-deoxyglucose (2DG), which is not a substrate for glycolysis, does not induce MICA expression. *G*, 2DG inhibits MICA expression in 25 mM glucose. *H*, mannose does not overcome the inhibition of glucose-induced MICA expression by 2DG. RT^+, with reverse transcriptase; RT^-, without reverse transcriptase; dH_2O, deionized H_2O. *Error bars*, 95% confidence interval.

5-Aminoimidazole-4-carboxamide ribonucleotide (AICA-Rt) is a *de novo* pathway purine intermediate nucleotide composed of a phosphorylated ribose sugar and an incomplete purine nucleobase. The inhibitor azaserine blocks *de novo* purine synthesis before the point of AICA-Rt generation. AICA ribonucleoside (AICA-Rs) is the equivalent intermediate nucleoside, composed of an unphosphorylated ribose moiety and the same incomplete purine nucleobase. AICA-Rs can be transported into the cell and phosphorylated by adenosine kinase to generate AICA-Rt. We tested the ability of AICA-Rs to act as a salvage pathway substrate *in vitro* by culturing cells in the presence of high glucose, azaserine, and increasing amounts of AICA-Rs and observed AICA-Rs-dependent induction of MICA expression (Fig. 5*F*). To determine whether AICA-Rs was *sufficient* to induce MICA expression, cells were cultured in 5 mM glucose with increasing doses of AICA-Rs, and AICA-Rs–dependent MICA induction was observed (Fig. 5*G*). Similarly, the purine nucleosides adenosine and guanosine, but not pyrimidine nucleosides cytidine and thymidine, were sufficient to induce MICA expression in 5 mM glucose (Fig. S2*D*). The purine nucleosides adenosine and inosine, but not purine nucleobases hypoxanthine and adenine, were sufficient to induce MICA expression in low-glucose conditions (Fig. S2*E*). Whereas azaserine prevented GIME, it had no effect on AICA-Rs, adenosine, or inosine-induced MICA expression, consistent with the hypothesis that purine nucleosides induce MICA expression downstream and independently of *de novo* purine synthesis (Fig. S2*F*). To generate new purine nucleotides, AICA-Rs must be phosphorylated to the *de novo* pathway intermediate nucleotide AICA-Rt by adenosine kinase inside the cell. Inhibition of adenosine kinase prevented AICA-Rs–induced MICA expression, demonstrating that AICA-Rt, but not AICA-Rs, causes MICA induction (Fig. S2, *G–I*). There was no correlation between cell proliferation and MICA expression under these different metabolic conditions (Fig. S3).

Intracellular purine nucleotide and tricarboxylic acid cycle intermediates are associated with increased MICA expression

These observations suggested that cellular conditions associated with the production of energy-rich phosphorylated purine nucleotides were pivotal in MICA expression at the cell surface. We tested this association by measuring intracellular metabolites and cell-surface MICA expression in a range of conditions, using capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS), liquid chromatography time-of-flight mass spectrometry (LC-TOFMS), and flow cytometry (Fig. 6*A*). Increased intracellular concentrations of purine nucleotides were observed when the cell-surface MICA expression increased. Furthermore, concentrations of TCA intermediates, necessary for maintaining purine nucleotides in their phosphorylated state, were also associated with high MICA expression (Fig. 6, *B* and *C*). ATP demonstrated the strongest linear correlation with MICA expression (Table 1). Interrogation of the complete data sets also indicated a metabolic signature that is associated with MICA expression. Principal component analysis clearly separated a low-MICA expression cluster, and two high-MICA expression clusters: an active *de novo* or salvage purine synthesis cluster and a purine nucleoside cluster (Fig. S4*A*). Metabolite enrichment analysis for predicted metabolite sets showed significant enrichment in pathways supporting nucleotide and TCA cycle product synthesis, including citrate synthase, ATP synthase, nucleoside-diphosphatase, and glutamate transport (Fig. S4*B*). A heat map of the full metabolome data set demonstrated Ward clustering according to MICA expression

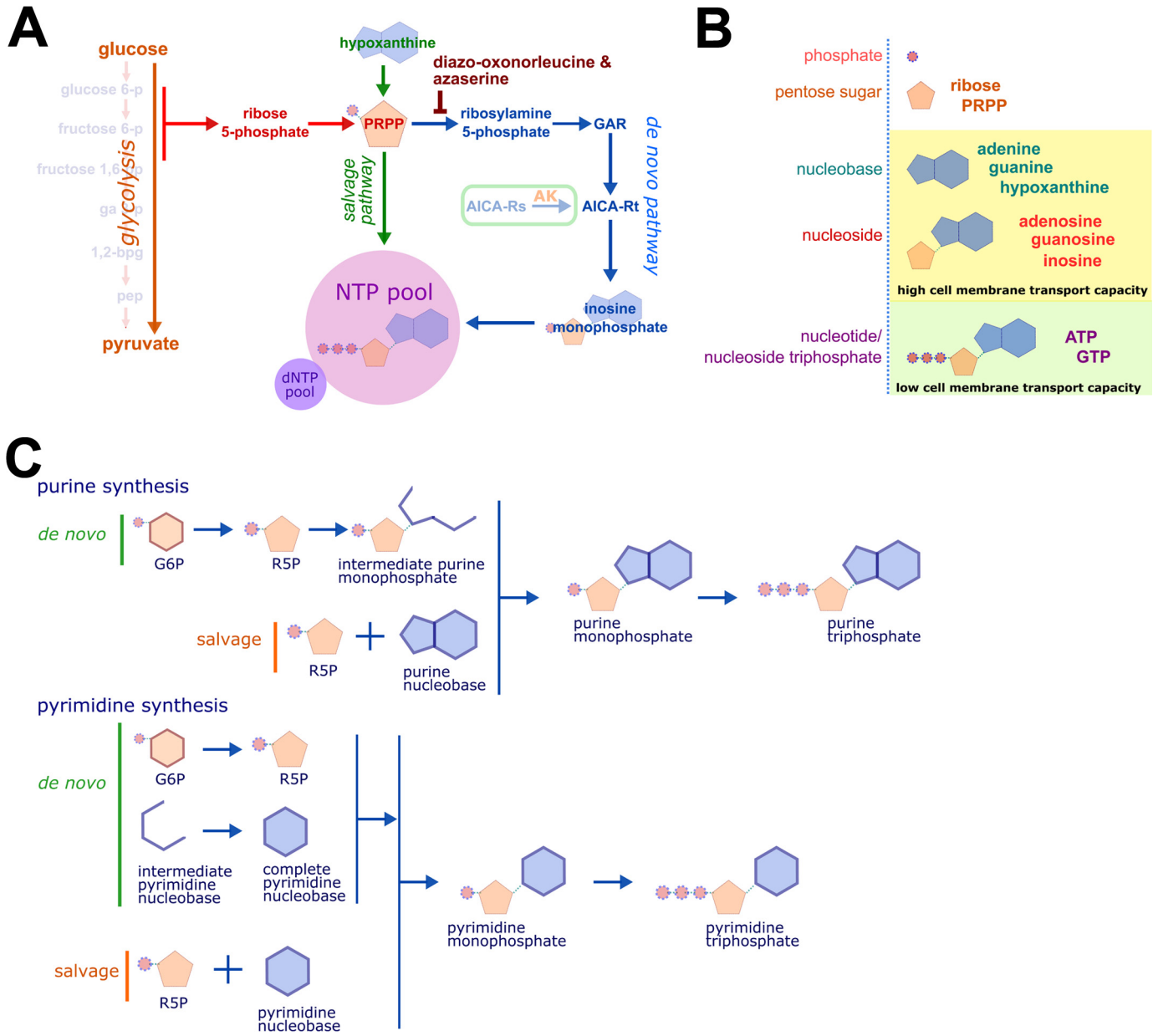

Figure 4. Nucleotide synthesis, structure, and nomenclature. *A*, ribose 5-phosphate (*R5P*) is produced by the metabolism of glucose 6-phosphate (*G6P*) through the pentose phosphate pathway (*PPP*). Phosphoribosyl pyrophosphate (*PRPP*) is essential for both salvage purine synthesis and *de novo* purine synthesis. The inhibitors DON and azaserine inhibit enzymes in the *de novo* purine synthesis pathway proximal to the intermediate AICA-Rt. The nucleoside AICA-Rs is readily transported across the cell membrane and is phosphorylated by adenosine kinase to the *de novo* synthesis pathway intermediate AICA-Rt. *B*, nucleobases and nucleosides are readily transportable across the cell membrane. Nucleotides (phosphonucleosides) have low cell membrane permeability. *C*, *de novo* and salvage purine synthesis both depend on PRPP. The carbon 5′-phosphate moiety of PRPP, originally added to glucose by hexokinase, defines purine synthesis pathway intermediates as nucleotides. In *de novo* synthesis, the purine ring is built directly onto PRPP. In contrast, the pyrimidine nucleobase is synthesized independently of PRPP, which is added to the complete nucleobase to form a pyrimidine nucleotide. *GAR*, glycinamide ribonucleotide; *NTP*, nucleotide triphosphate; *AK*, adenosine kinase; *arrows* represent multistep metabolic pathways.

across a diverse set of metabolic conditions (Fig. S5). MICA expression was associated with increased high-energy purine compounds (ATP, GTP) and TCA cycle metabolites.

Glucose- and purine-induced MICA expression increase NKG2D-dependent cellular immunogenicity

We performed chromium-release cytotoxicity assays to test the functional significance of the altered MICA expression seen with metabolic changes. However, first we measured the effect of glucose and purine nucleotides on the other six NKG2D ligands and confirmed that expression of MICB, ULBP1, ULBP2, and ULBP3 were similarly responsive to metabolic changes (Fig. 7, *A* and *B*). Chromium-release assays demonstrated a significant increase in NK cell cytotoxicity toward cells cultured in glucose, AICA-Rs, adenosine, or inosine (Fig. 7, *C–E*). This increased killing was significantly reduced by preincubation with an NKG2D receptor–blocking antibody (Fig. 7, *F–H*).

Glucose-induced MICA expression is observed in primary cells

The impact of changes in metabolism on NKG2D ligand expression was evaluated in primary human fibroblast cells

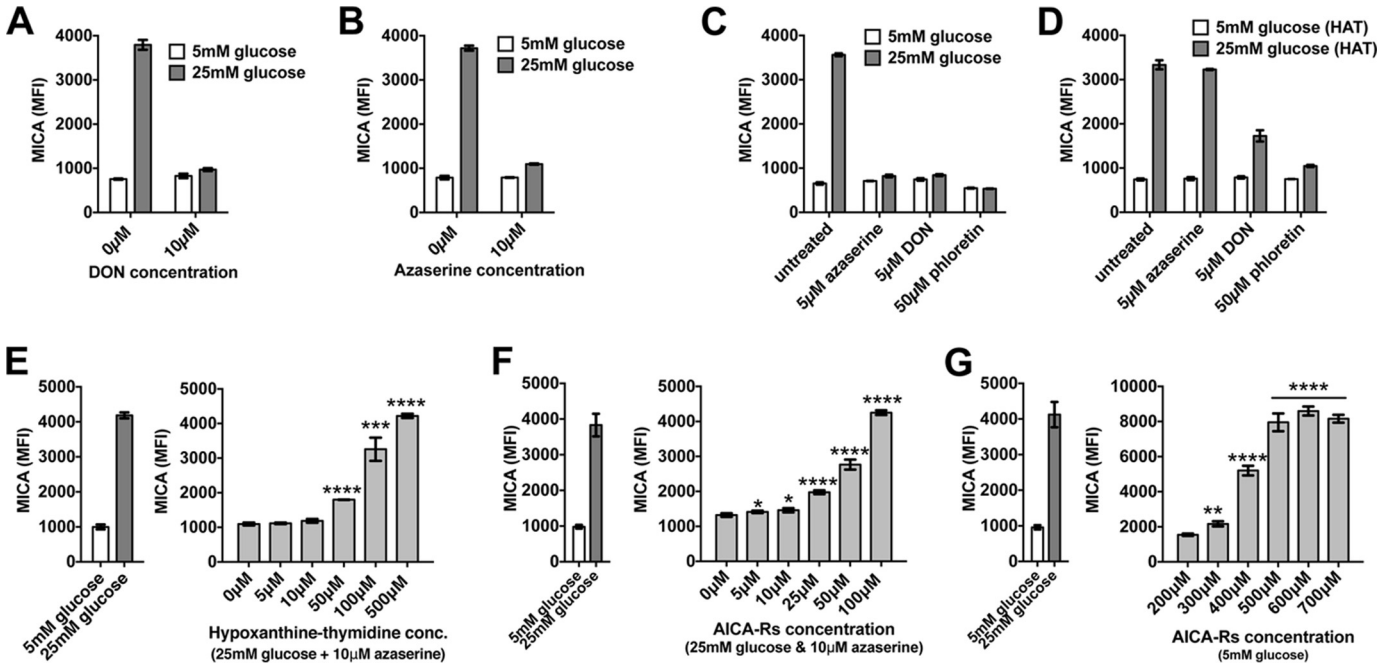

Figure 5. An intact purine synthesis pathway is necessary for MICA induction. *A*, DON, an inhibitor of proximal *de novo* purine synthesis, prevents GIME. *B*, azaserine, a second *de novo* pathway inhibitor, also blocks GIME. *C*, azaserine, DON, and the GLUT transporter inhibitor phloretin each prevent GIME in HEK-293T cells. *D*, HAT-selected HEK-293T cells produce purine nucleotides exclusively through the salvage pathway. Azaserine is unable to prevent GIME in HAT-selected cells, showing that inhibition of GIME by azaserine depends on its action in inhibiting *de novo* purine synthesis. DON, like phloretin, appears to have additional off-target actions. *E*, the salvage pathway substrate hypoxanthine rescues MICA expression in azaserine-treated cells in 25 mM glucose. *F*, similarly, AICA-Rs also rescues MICA expression in azaserine-treated cells. *G*, AICA-Rs, which can be converted to AICA-Rt by adenosine kinase, is sufficient to induce MICA expression even in conditions of low glucose. *Error bars*, 95% confidence interval.

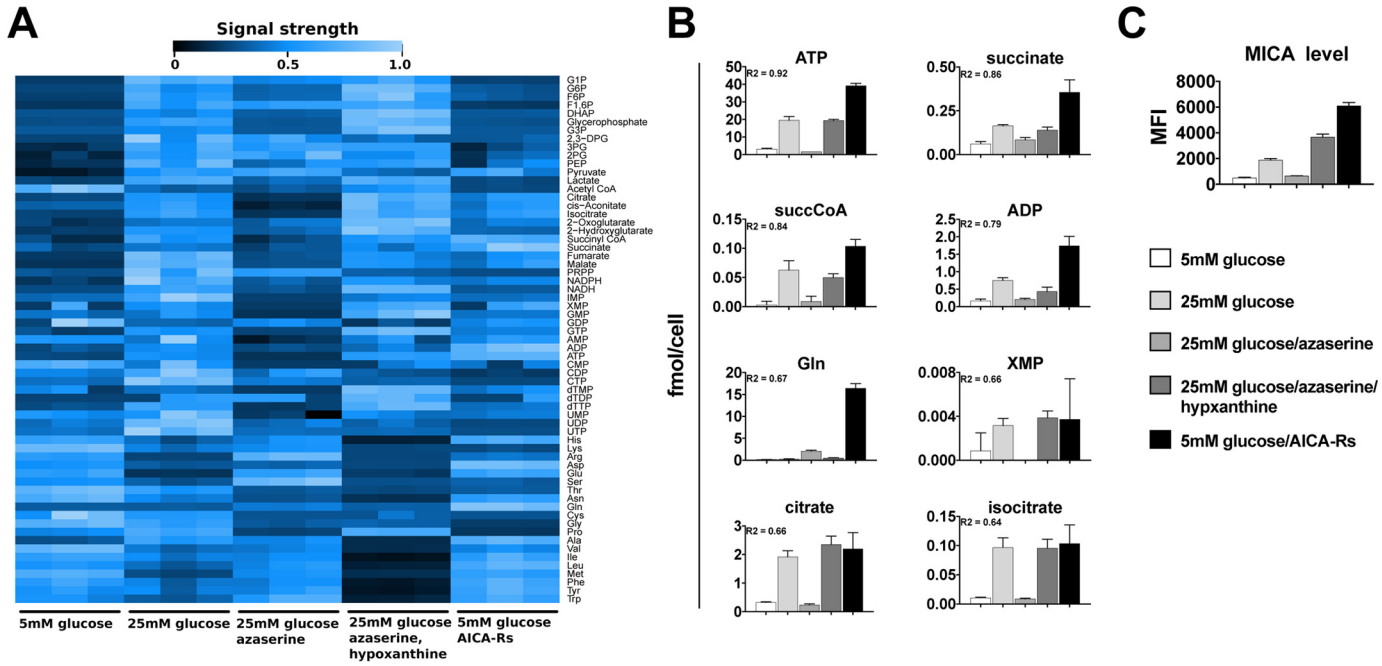

Figure 6. High-energy purine nucleotides and TCA cycle intermediates correlate strongly with MICA expression. *A*, intracellular metabolite concentrations, measured by CE-TOFMS and LC-TOFMS, in a range of metabolic conditions, are shown in this heat map. Signal strength is depicted as a percentage of the maximum for each compound. *B*, the strongest linear correlations between intracellular metabolite concentrations and cell-surface MICA expression were noted for purine nucleotides, including ATP and TCA cycle intermediates. *C*, cell-surface levels of MICA measured by flow cytometry. *Error bars*, 95% confidence interval.

during infection with human cytomegalovirus (CMV). CMV infection elicits increased glucose consumption and lactate production; these effects are prevented by azaserine and restored by the salvage pathway substrate hypoxanthine (Fig. S6, *A* and *B*). A raised glucose concentration was associated with an increase in the CMV-induced cell-surface expression of MICA, MICB, ULBP1, and ULBP2 (Fig. S6*C*). The induction of these NKG2D ligands was significantly higher in 25 mM glu-

Table 1
Linear correlation between intracellular metabolite concentrations and cell surface MICA expression

	5mM glucose	25mM glucose	25mM glucose + azaserine	25mM glucose + azaserine + hypoxanthine	5mM glucose + 1mM AICA-Rs		
MICA (MFI)	482.33	1877.00	642.33	3665.33	6096.00	R	R2
ATP	3.14	19.47	1.61	19.30	39.20	0.96	0.92
Succinate	0.06	0.16	0.08	0.14	0.36	0.93	0.86
SuccinylCoA	0.00	0.06	0.01	0.05	0.10	0.92	0.84
ADP	0.16	0.75	0.21	0.43	1.74	0.89	0.79
Glutamine	0.16	0.21	2.04	0.47	16.40	0.82	0.67
XMP	0.00	0.00	0.00	0.00	0.00	0.82	0.66
Citrate	0.32	1.92	0.23	2.35	2.19	0.81	0.66
Isocitrate	0.01	0.10	0.01	0.10	0.10	0.80	0.64
Lysine	4.01	3.30	3.10	1.13	1.82	-0.79	0.63
Cis-aconitate	0.04	0.12	0.02	0.13	0.13	0.79	0.62
Aspartic acid	3.57	2.36	2.34	0.91	16.74	0.76	0.58
α-Keto-glutarate	0.16	0.68	0.44	1.13	0.82	0.73	0.53
dTMP	0.00	0.00	0.00	0.01	0.01	0.72	0.51
AMP	0.08	0.18	0.05	0.11	0.19	0.70	0.49
GTP	1.44	12.07	1.90	16.92	10.66	0.66	0.43
GMP	0.03	0.08	0.01	0.12	0.07	0.64	0.41
Tyrosine	0.68	1.15	0.96	0.45	1.95	0.63	0.39
Pyruvate	0.16	1.10	0.90	0.61	1.34	0.61	0.38
Phenylalanine	0.81	1.16	0.97	0.42	2.10	0.61	0.37
Leucine	0.84	1.05	1.20	0.27	2.75	0.61	0.37
Alanine	4.46	21.62	11.89	2.57	32.29	0.60	0.37
Histidine	0.44	0.47	0.50	0.15	1.15	0.60	0.36
Ribose 5-phosphate	0.01	0.00	0.05	0.00	0.00	-0.58	0.33
tryptophan	0.10	0.15	0.18	0.07	0.30	0.56	0.32
Glutamine	33.05	20.04	47.35	18.20	80.86	0.55	0.31
dTTP	0.01	0.09	0.01	0.10	0.06	0.55	0.30
Malate	0.05	1.28	0.22	0.63	0.91	0.54	0.29
GDP	0.16	0.16	0.15	0.08	0.27	0.52	0.27
Asparagine	3.74	3.89	1.44	0.58	6.79	0.51	0.26
Isoleucine	0.87	1.27	0.99	0.26	2.10	0.51	0.26
Cysteine	0.05	0.05	0.00	0.01	0.00	-0.50	0.25
Phosphoenol pyruvate	0.02	0.12	0.06	0.09	0.09	0.48	0.23
Fumarate	0.02	0.39	0.06	0.17	0.24	0.44	0.20
Methionine	0.41	0.19	0.38	0.06	0.78	0.44	0.19
UMP	0.01	0.03	0.00	0.02	0.02	0.40	0.16
dTDP	0.00	0.00	0.00	0.00	0.00	0.34	0.12
AcetylCoA	0.02	0.01	0.00	0.02	0.00	-0.34	0.11
CDP	0.01	0.04	0.02	0.01	0.01	-0.32	0.10
Glycine	24.31	38.68	14.11	17.34	16.03	-0.29	0.08
Arginine	0.28	0.01	0.62	0.01	0.30	-0.28	0.08
Ribulose 5-phosphate	0.04	0.18	0.07	0.12	0.01	-0.27	0.07
Serine	0.68	0.73	1.70	0.52	0.96	-0.26	0.07
Dihydroxyacetone phosphate	0.00	0.31	0.05	0.73	0.04	0.23	0.05
Fructose 6-phosphate	0.00	0.20	0.05	0.23	0.06	0.23	0.05
UTP	0.57	4.69	0.73	0.75	0.28	-0.23	0.05
CTP	0.26	1.65	0.43	0.34	0.23	-0.22	0.05
UDP	0.04	0.25	0.05	0.05	0.03	-0.22	0.05
Glucose 6-phosphate	0.00	0.78	0.18	1.12	0.16	0.21	0.04
Valine	2.51	2.09	1.79	0.44	3.26	0.20	0.04
Glyceraldehyde 3-phosphate	0.00	0.04	0.00	0.10	0.00	0.20	0.04
Fructose 1,6-bisphosphate	0.01	0.89	0.58	0.74	0.05	-0.19	0.04
Threonine	20.31	27.74	8.57	7.34	15.46	-0.18	0.03
2-phosphglycerate	0.01	0.04	0.03	0.03	0.02	0.17	0.03
3-phosphoglycerate	0.04	0.25	0.17	0.19	0.13	0.15	0.02
IMP	0.01	0.08	0.00	0.02	0.02	0.15	0.02
6-Phospho-gluconate	0.00	0.03	0.01	0.08	0.00	0.13	0.02
Proline	21.65	67.95	8.79	56.71	7.63	-0.07	0.00
CMP	0.00	0.01	0.00	0.00	0.00	-0.01	0.00

cose, was limited by azaserine, and was rescued in the presence of azaserine by hypoxanthine ($p < 0.001$) (Fig. S6*D*).

Discussion

The focus of immunometabolism studies has been largely on the considerable influence of metabolism on the function of immune cells (35). Here we show that metabolism also affects cells that are potential targets for immune cells; glucose uptake and metabolism to purine nucleotides drives cell-surface MICA expression and NKG2D-dependent cytotoxic killing of these cells by immune cells.

These results are consistent with observational studies of NKG2D ligand expression in disparate settings from malignancy (5) to viral infection (6, 7) and physiological cell proliferation (8, 9). Each of these cellular processes is supported by enhanced biosynthesis, whether dysregulated, pathological, or physiological. Enhanced biosynthesis, the primary physiological outcome of "activated" or Warburg metabolism, can confer a survival advantage to individual cells by producing biological substrates (lipids, nucleotides, and amino acids) to generate the complex molecules (membranes, nucleic acids, and proteins) that support cell maintenance, function, and division. Dysregulation of biosynthesis is a challenge to the integrity and immunity of multicellular organisms, as it may support inappropriate cell survival or pathophysiological cellular processes. Thus, dysregulated enhanced biosynthesis may advantage a cancer cell, so detection and destruction of such cells by the immune system would be desirable.

Warburg metabolism describes a metabolic phenotype of increased glucose consumption, lactate production, biosynthesis, and an increased ratio of glycolysis to TCA cycle metabolism. This metabolic state supports viral replication (31) and proliferation of both healthy (36) and malignant (37) cells. The molecular events that control transition between quiescent and activated metabolic phenotypes remain unclear (38, 39). Our data demonstrate that glucose restriction replicates the biomolecular and metabolomic characteristics of quiescent and activated metabolic states, with increased glucose consumption, lactate production, intracellular ribonucleic acid levels, and TCA cycle intermediate and high-energy purine nucleotide concentrations. The inhibition of *de novo* purine synthesis replicates the glucose-restricted condition in both metabolic profile and MICA expression, whereas rescue of the purine salvage pathway restores both metabolic profile and MICA expression to the levels seen with high glucose. Supplying cells directly with the purine nucleotide substrate AICA-Rs induces MICA expression and increases cellular purine nucleotide concentrations without an increase in early glycolytic pathway intermediates, suggesting that the purine output from glycolysis is key to MICA induction. Whereas cell proliferation is observed in several settings in which MICA is expressed (malignant transformation, lymphocyte activation), our data suggest that MICA expression in response to metabolic change can occur in the absence of cellular proliferation.

In healthy individuals, interstitial glucose concentrations typically range from 4.2 to 9.3 mM over a 24-h period (40). In diabetic patients, plasma glucose concentrations can be higher, but glucose levels are generally controlled by diet, insulin, or other drugs because persistent changes substantially beyond the normal range can lead to diabetic ketoacidosis, which carries a high mortality. Raised glucose levels in diabetes have a limited effect on intracellular glucose concentration because of the lack of insulin, which is required to promote glucose uptake

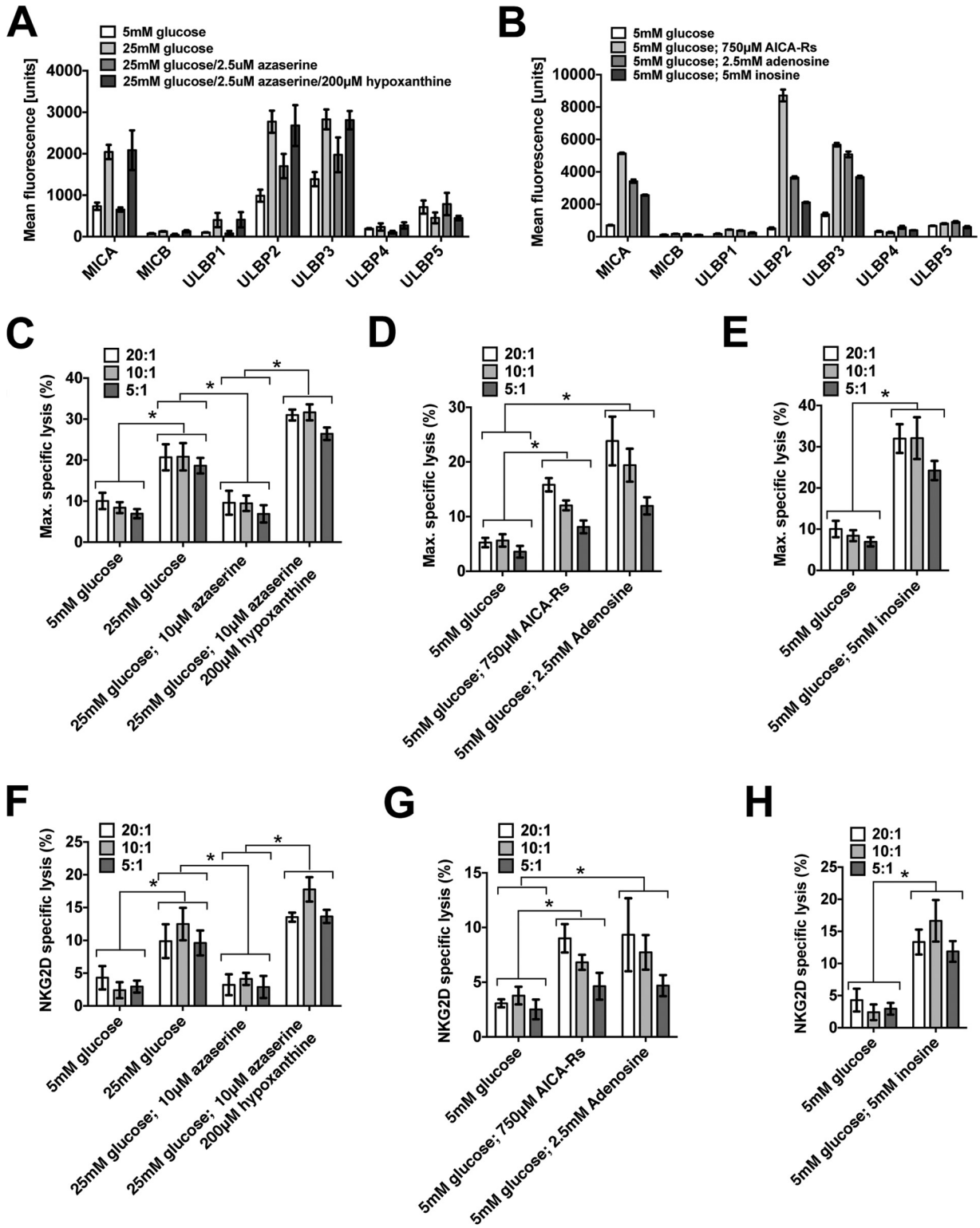

Figure 7. Cellular metabolism affects cellular immunogenicity. *A*, in addition to affecting MICA expression, glucose, azaserine, and azaserine/hypoxanthine have a similar effect on MICB, ULBP1, ULBP2, and ULBP3. *B*, AICA-Rs, adenosine, and inosine each induce the expression of ULBP2 and ULBP3 even in 5 mM glucose. As multiple NKG2D ligands are similarly affected, we used anti-NKG2D receptor for blocking assays. *C*, cells cultured in 25 mM glucose or in 25 mM glucose with azaserine and hypoxanthine were killed more effectively than cells cultured in 5 mM glucose or in 25 mM glucose with azaserine. *D*, cells cultured in 5 mM glucose with AICA-Rs or adenosine were more susceptible to killing than cells cultured in 5 mM glucose alone. *E*, similarly, cells cultured in 5 mM glucose with inosine were more susceptible to NK cell cytotoxicity than cells cultured in 5 mM glucose only. *F–H*, in conditions described in *C–E*, the addition of a blocking NKG2D antibody prevented a significant amount of induced killing, consistent with an NKG2D-dependent killing mechanism. *Error bars*, 95% confidence interval.

into cells. In contrast, glucose concentrations in the solid tumor microenvironment may be 10-fold lower than in healthy tissues (41). Overall, our findings indicate that extracellular glucose will only influence MICA levels in cells that can take up and metabolize the glucose through the purine synthesis pathway. Notably, CMV infection, which is associated with NKG2D ligand induction, is known to both up-regulate the GLUT4 glucose transporter (42) and induce Warburg metabolism (27).

The complexity of gene-regulatory networks means that a broad array of cellular interventions may alter the expression of a given gene. Previous studies of MICA regulation show that the DNA damage response pathway (19), TLR stimulation (10), histone deacetylation (20), heat shock transformation (21), ionizing radiation (22), growth factor pathway activation (23), and microRNA expression (25) can each induce MICA expression. Whereas TLR stimulation has been shown to induce a Warburg phenotype (28), the relationship between cellular metabolism and these other stimuli is not well-defined. Further characterization of the cellular metabolic state in these settings is necessary to assess the role of cellular nucleotide concentrations in NKG2D ligand induction under these conditions.

Patterns of NKG2D ligand expression differ between cell types. Whether individual NKG2D ligands are independently regulated for distinct immunological functions or there is redundancy in the ligand repertoire remains unclear. We show that where we observe MICA expression, it is dependent on intracellular purine nucleotide concentrations. Our data also show that the effect of cellular metabolism on MICA expression applies to other NKG2D ligands that are expressed in this cellular context.

T cell immunity, a relatively recent evolutionary development, provides adaptive immune oversight of the cellular proteome, marking somatic cells producing viral or abnormal self proteins as T cell targets (Fig. S7). In the cytosol, continuously sampled cellular proteins are degraded to peptides and displayed at the cell surface in the peptide-binding cleft of the major histocompatibility complex (MHC) class I molecule. Cells displaying abnormal peptide-MHC class I complexes, recognized by interaction with the T cell receptor (TCR), are susceptible to elimination by T cells through direct cytotoxicity or cytokine production. This potent adaptive T cell response is checked by an elaborate set of mechanisms collectively referred to as immune self-tolerance, preventing the activation of T cells by self protein-MHC class I complexes. In contrast to TCR surveillance of peptides, our data demonstrate that NKG2D-mediated immune recognition of MICA at the cell surface allows innate immune oversight of nucleotide biosynthesis. Thus, both TCR- and NKG2D- mediated immunity are triggered principally by distinct biosynthetic consequences of viral infection and malignant transformation. This functional homology is reflected in the structural homology of MHC class I molecules and MICA; MHC class I proteins bind peptides in a peptide-binding groove, but MICA does not, and structural studies have shown that the groove is narrowed and closed (43, 44). Immune oversight of nucleotide synthesis permits the early detection and elimination of transformed or excessively proliferating cells, which may demonstrate a "normal" or self proteome, toward which adaptive immunity is self-tolerant.

This insight into the molecular control of MICA underlines the importance of metabolic surveillance in cellular immunity and outlines a potential route to therapeutic advancement in cancer, autoimmune disease, and transplantation.

Experimental procedures

Cell lines and standard culture medium

293T, HeLa, MCF7, and HT1080 cell lines were cultured in glucose-free Dulbecco's modified Eagle's medium, supplemented with 10% fetal calf serum, 2 mM pyruvate, penicillin/streptomycin solution, and glucose at the concentration indicated. NK92 cells were cultured in Roswell Park Memorial Institute 1640 culture medium (RPMI) supplemented with 10% fetal calf serum, 2 mM pyruvate, penicillin/streptomycin solution, and interleukin-2 (200 units/ml). Primary human adult dermal fibroblasts were purchased from Life Technologies, Inc. and cultured in RPMI with 10% fetal calf serum, 2 mM glutamine, and penicillin/streptomycin solution.

Chemicals and reagents

Glucose (catalog no. 49152), fructose (catalog no. F3510), mannose (catalog no. M6020), mannitol (catalog no. 17311), azaserine (catalog no. 11430), 6 diazo-oxo-norleucine (catalog no. D2141), 2-deoxyglucose (catalog no. D8375), phloretin (catalog no. P7912), cytochalasin B (catalog no. C6762-1mg), carboxyfluorescein diacetate succinimidyl ester (CFDA-SE) (catalog no. 21888), all nucleosides and nucleobases, and 50× hypoxanthine-thymidine solution (catalog no. A9666) were purchased from Sigma-Aldrich. AICA-Rs (catalog no. 9944S) was purchased from New England Biolabs (Ipswich, MA). Adenosine kinase inhibitor (catalog no. 116890, CAS 214697-26-4) was purchased from Merck-Millipore (Darmstadt, Germany), and iodotubercidin was from Abcam (Cambridge, UK).

Glucose and lactate measurement

Measurements of glucose, lactate, and osmolality in cell culture media were performed by the Department of Clinical Biochemistry, John Radcliffe Hospital, Oxford. Glucose and lactate concentrations were determined using the glucose hexokinase II reagent kit (catalog no. 04903429, Bayer Healthcare, Leverkusen, Germany) and the lactate reagent kit (catalog no. 07109944, Bayer Healthcare), respectively. Osmolality was calculated on a VITECH 3320 micro-osmometer. Intracellular glucose and lactate concentrations were determined by a liquid chromatography/mass spectrometry quadrupole time-of-flight nanoflow method, as described previously (45).

Antibodies

Antibodies against MICA (2C10, IgG1), ULBP4 (6E6, IgG2B), and ULBP5 (6D10, IgM) were purchased from Santa Cruz Biotechnology (Heidelberg, Germany); MICB (MAB1599, IgG2b), ULBP1 (MAB1380, IgG2a), ULBP2 (MAB1298, IgG2a), and ULBP3 (MAB1517, IgG2a) were purchased from R&D Systems (Minneapolis, MN); podoplanin (NZ-1, IgG1) was purchased from AngioBio. HLA-ABC (W6/32, IgG2a), anti-PCNA (14-9910-80, IgG2a), and all isotype control antibodies were purchased from eBioscience (Hatfield, UK). The Alexa Fluor sec-

ondary antibodies 647-conjugated goat anti-mouse IgM (A-21238) and 647-conjugated goat anti-mouse IgM (A-21236) were purchased from Invitrogen.

Flow cytometry

Flow cytometry analysis was carried out on a BD Biosciences FacsCanto flow cytometer, using BD FacsDiva acquisition software, and results were analyzed using FlowJo (Ashland, OR).

For cell-surface staining, cells were washed with phosphate-buffered saline with 0.03% azide (PBSA). Primary and secondary antibody staining were carried out by standard protocols. Propidium iodide (PI; 4 μg/ml) was added to the final resuspension buffer to determine cell viability.

Total cell MICA was measured by flow cytometry of permeabilized cells. Cells were fixed by resuspension in 2% paraformaldehyde for 30 min and permeabilized with 0.05% saponin in PBSA for 30 min. The standard staining protocol was then followed, but with 0.05% saponin added to each of the buffers. Unless otherwise stated, experiments were conducted at least in triplicate. The histograms show average mean fluorescence intensity of three biological replicate samples, and *error bars* represent the 95% confidence interval.

Cell-cycle analysis

Cells were fixed in 1% paraformaldehyde for 60 min on ice, permeabilized in 70% ethanol, and resuspended in staining solution (40 μg/ml PI, 100 μg/ml RNase A in PBS) for 30 min at 37 °C. The cell-cycle analysis was conducted by flow cytometry with PI fluorescence measured in linear mode. Doublets were excluded by initial PI width-area gating. PI voltage was adjusted to center the G_1 population on the 50,000-V area level, and cell-cycle parameter modeling was performed using FlowJo software.

Carboxyfluorescein diacetate succinimidyl ester (CFSE) proliferation assay

For adherent cells, the culture medium was replaced with CFDA-SE (5 μM). Cells were incubated at 37 °C for 10 min, and the CFDA-SE solution was replaced with fresh culture medium, as indicated. Cells were cultured for a further 48 h, before measuring CFSE fluorescence by flow cytometry through a FITC filter.

Cloning and transfection

The Consensus Coding Sequence Database (CCDS) was used to identify the reference sequence for GLUT5 expression cloning. Primers were designed to include a Kozak sequence (forward, TTAATAAGCTTGCCATGGAGCAACAGGATCAGAG; reverse, TTAATCTCGAGTCACTGTTCCGAAGTGACAGGTG). The gene was cloned from MCF7 cell cDNA and ligated into the pCDNA3.1 expression vector. Cells were transfected using a standard polyethyleneimine protocol. The doxycycline-inducible eGFP lentiviral expression vector was generated by cloning eGFP from the expression vector pEGFP-N1 between the XhoI and MluI sites of a modified pTripZ vector. The modified pTripZ vector was made by excising Turboprop using the AgeI and MluI restriction sites and inserting the following modified multiple cloning site: forward primer, CCGGTATCGATGAATTCTTCGAACTCGAGATTAATAAGCTTA; reverse primer, CGCGTAAGCTTATTAATCTCGAGTTCGAAGAATTCATCGATA. This plasmid was used to generate a lentiviral vector using standard protocols. Infected cells were selected with puromycin and induced with doxycycline at a concentration of 2 μg/ml.

Reverse transcription-PCR

RNA extraction was carried out using TRIzol and the Purelink RNA minikit, including DNase I digestion using the Purelink DNase I kit (Life Technologies), according to the manufacturer's protocol. RNA concentration was measured by Nanodrop (Wilmington, DE). The following PCR primers were used: GLUT1, GTGCAGCAGCCTGTGTATGC and GGCCACGATGCTCAGATAGG; GLUT2, CTGTGCTGGGTTCCTTCCAG and AAGGGGTTGGTTTTGGGTTC; GLUT3, TACAGCGATGGGGACACAGAAG and CCAGAGAGACGTGAGCAGCAC; GLUT4, CCCTGGTCCTTGCTGTGTTC and AAAAGATGGCCACGGAGAGG; GLUT5, GAATTCATGGAAGACTT and GCCATCTACGTTTGCAA (46).

Chromium release cytotoxicity assay

1×10^6 target cells for each test condition were washed in fresh culture medium and pelleted. These target cell pellets were resuspended in 50 μl of chromium-51 (0.05 mCi)/sample and incubated for 1 h at 37 °C. Labeled cells were washed twice with fresh RPMI and transferred to a 96-well plate, with 5000 cells in 50 μl of fresh culture medium per well. NK92 cells were resuspended in fresh NK cell medium and serially diluted to generate the effector/target ratios described, at a volume of 100 μl/well. Background lysis was measured by adding 100 μl of NK92 medium only, and maximal lysis was measured by adding 100 μl of 5% Triton-X. Effector and target cells were mixed by gentle pipetting and co-incubated at 37 °C for 1 h. The reaction plate was then centrifuged, and 25 μl of supernatant from each well was added to 150 μl of scintillation fluid in a fresh 96-well plate. Scintillation counts were measured on a Microbeta TriLux liquid scintillation counter. To measure NKG2D-specific killing, effector NK92 cells were initially resuspended in 2 ml of NK92 culture medium, and an anti-NKG2D antibody (BD Biosciences, catalog no. 552866) was added at a concentration of 5 μg/ml. The NK92 cells were incubated at 37 °C for 30 min before being resuspended in RPMI and added to target cells as described previously.

Metabolite analysis

The concentrations of intracellular anionic and cationic metabolites were measured by CE-TOFMS and LC-TOFMS (47, 48). Samples were prepared as described previously (49).

Statistical analysis

Histogram bars represent mean values, and *error bars* represent the 95% confidence interval of the mean. Statistical significance and p values were calculated using t tests unless otherwise specified. Correlation between cell-surface MICA expression and intracellular metabolite concentration was calculated using the correlation coefficient and coefficient of

determination between mean values measured in parallel. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; and ****, $p < 0.0001$, unless otherwise specified. Metabolomic analysis was undertaken using R and algorithms implemented in Metaboanalyst version 3.0 (50).

Author contributions—M. T. M. and C. A. O. conceived the and wrote the paper. M. T. M., C. A. O., A. V., G. M., T. K. H., J. A., and T. S. designed the experiments. M. T. M., A. V., G. M., T. H., N. A. J., and T. S. performed the experiments. M. T. M. and C. A. O. wrote the original manuscript draft. M. T. M., C. A. O., A. V., G. M., J. A., and N. A. J. reviewed and edited the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments—We are grateful to Thue Schwartz and fellow members of the Novo Nordisk Foundation Immunometabolism consortium for advice and helpful discussions.

References

1. Topalian, S. L., Drake, C. G., and Pardoll, D. M. (2015) Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* **27,** 450–461 CrossRef Medline
2. Raulet, D. H., Gasser, S., Gowen, B. G., Deng, W., and Jung, H. (2013) Regulation of ligands for the NKG2D activating receptor. *Annu. Rev. Immunol.* **31,** 413–441 CrossRef Medline
3. Bauer, S., Groh, V., Wu, J., Steinle, A., Phillips, J. H., Lanier, L. L., and Spies, T. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285,** 727–729 CrossRef Medline
4. Wu, J., Song, Y., Bakker, A. B. H., Bauer, S., Spies, T., Lanier, L. L., and Phillips, J. H. (1999) An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* **285,** 730–732 CrossRef Medline
5. McGilvray, R. W., Eagle, R. A., Watson, N. F. S., Al-Attar, A., Ball, G., Jafferji, I., Trowsdale, J., and Durrant, L. G. (2009) NKG2D ligand expression in human colorectal cancer reveals associations with prognosis and evidence for immunoediting. *Clin. Cancer Res.* **15,** 6993–7002 CrossRef Medline
6. Groh, V., Rhinehart, R., Randolph-Habecker, J., Topp, M. S., Riddell, S. R., and Spies, T. (2001) Costimulation of CD8$\alpha\beta$ T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nat. Immunol.* **2,** 255–260 CrossRef Medline
7. Eagle, R. A., Traherne, J. A., Ashiru, O., Wills, M. R., and Trowsdale, J. (2006) Regulation of NKG2D ligand gene expression. *Hum. Immunol.* **67,** 159–169 CrossRef Medline
8. Cerboni, C., Zingoni, A., Cippitelli, M., Piccoli, M., Frati, L., and Santoni, A. (2007) Antigen-activated human T lymphocytes express cell-surface NKG2D ligands via an ATM/ATR-dependent mechanism and become susceptible to autologous NK-cell lysis. *Blood* **110,** 606–615 CrossRef Medline
9. Cerboni, C., Ardolino, M., Santoni, A., and Zingoni, A. (2009) Detuning CD8$^+$ T lymphocytes by down-regulation of the activating receptor NKG2D: role of NKG2D ligands released by activated T cells. *Blood* **113,** 2955–2964 CrossRef Medline
10. Kloss, M., Decker, P., Baltz, K. M., Baessler, T., Jung, G., Rammensee, H.-G., Steinle, A., Krusch, M., and Salih, H. R. (2008) Interaction of monocytes with NK cells upon Toll-like receptor-induced expression of the NKG2D ligand MICA. *J. Immunol.* **181,** 6711–6719 CrossRef Medline
11. Lin, D., Lavender, H., Soilleux, E. J., and O'Callaghan, C. A. (2012) NF-κB regulates MICA gene transcription in endothelial cell through a genetically inhibitable control site. *J. Biol. Chem.* **287,** 4299–4310 CrossRef Medline
12. Cosman, D., Müllberg, J., Sutherland, C. L., Chin, W., Armitage, R., Fanslow, W., Kubin, M., and Chalupny, N. J. (2001) ULBPs, novel MHC class I-related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. *Immunity* **14,** 123–133 CrossRef Medline
13. Ehrlich, L. I. R., Ogasawara, K., Hamerman, J. A., Takaki, R., Zingoni, A., Allison, J. P., and Lanier, L. L. (2005) Engagement of NKG2D by cognate ligand or antibody alone is insufficient to mediate costimulation of human and mouse CD8$^+$ T cells. *J. Immunol.* **174,** 1922–1931 CrossRef Medline
14. Rajasekaran, K., Xiong, V., Fong, L., Gorski, J., and Malarkannan, S. (2010) Functional dichotomy between NKG2D and CD28-mediated co-stimulation in human CD8$^+$ T cells. *PLoS One* **5,** e12635 CrossRef Medline
15. André, P., Castriconi, R., Espéli, M., Anfossi, N., Juarez, T., Hue, S., Conway, H., Romagné, F., Dondero, A., Nanni, M., Caillat-Zucman, S., Raulet, D. H., Bottino, C., Vivier, E., Moretta, A., and Paul, P. (2004) Comparative analysis of human NK cell activation induced by NKG2D and natural cytotoxicity receptors. *Eur. J. Immunol.* **34,** 961–971 CrossRef Medline
16. Andersson, A. K., Sumariwalla, P. F., McCann, F. E., Amjadi, P., Chang, C., McNamee, K., Tornehave, D., Haase, C., Agersø, H., Stennicke, V. W., Ahern, D., Ursø, B., Trowsdale, J., Feldmann, M., and Brennan, F. M. (2011) Blockade of NKG2D ameliorates disease in mice with collagen-induced arthritis: a potential pathogenic role in chronic inflammatory arthritis. *Arthritis Rheum.* **63,** 2617–2629 CrossRef Medline
17. Ito, Y., Kanai, T., Totsuka, T., Okamoto, R., Tsuchiya, K., Nemoto, Y., Yoshioka, A., Tomita, T., Nagaishi, T., Sakamoto, N., Sakanishi, T., Okumura, K., Yagita, H., and Watanabe, M. (2008) Blockade of NKG2D signaling prevents the development of murine CD4 + T cell-mediated colitis. *Am. J. Physiol. Gastrointest. Liver Physiol.* **294,** G199–G207 Medline
18. Suárez-Alvarez, B., López-Vázquez, A., Baltar, J. M., Ortega, F., and López-Larrea, C. (2009) Potential role of NKG2D and its ligands in organ transplantation: new target for immunointervention. *Am. J. Transplant.* **9,** 251–257 CrossRef Medline
19. Gasser, S., Orsulic, S., Brown, E. J., and Raulet, D. H. (2005) The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* **436,** 1186–1190 CrossRef Medline
20. Andresen, L., Jensen, H., Pedersen, M. T., Hansen, K. A., and Skov, S. (2007) Molecular regulation of MHC class I chain-related protein A expression after HDAC-inhibitor treatment of Jurkat T cells. *J. Immunol.* **179,** 8235–8242 CrossRef Medline
21. Groh, V., Bahram, S., Bauer, S., Herman, A., Beauchamp, M., and Spies, T. (1996) Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 12445–12450 CrossRef Medline
22. Kim, J.-Y., Son, Y.-O., Park, S.-W., Bae, J.-H., Chung, J. S., Kim, H. H., Chung, B.-S., Kim, S.-H., and Kang, C.-D. (2006) Increase of NKG2D ligands and sensitivity to NK cell-mediated cytotoxicity of tumor cells by heat shock and ionizing radiation. *Exp. Mol. Med.* **38,** 474–484 CrossRef Medline
23. Vantourout, P., Willcox, C., Turner, A., Swanson, C. M., Haque, Y., Sobolev, O., Grigoriadis, A., Tutt, A., and Hayday, A. (2014) Immunological visibility: posttranscriptional regulation of human NKG2D ligands by the EGF receptor pathway. *Sci. Transl. Med.* **6,** 231ra49–231ra49 CrossRef Medline
24. Waldhauer, I., Goehlsdorf, D., Gieseke, F., Weinschenk, T., Wittenbrink, M., Ludwig, A., Stevanovic, S., Rammensee, H.-G., and Steinle, A. (2008) Tumor-associated MICA is shed by ADAM proteases. *Cancer Res.* **68,** 6368–6376 CrossRef Medline
25. Stern-Ginossar, N., Gur, C., Biton, M., Horwitz, E., Elboim, M., Stanietsky, N., Mandelboim, M., and Mandelboim, O. (2008) Human microRNAs regulate stress-induced immune responses mediated by the receptor NKG2D. *Nat. Immunol.* **9,** 1065–1073 CrossRef Medline
26. Venkataraman, G. M., Suciu, D., Groh, V., Boss, J. M., and Spies, T. (2007) Promoter region architecture and transcriptional regulation of the genes for the MHC class I-related chain A and B ligands of NKG2D. *J. Immunol.* **178,** 961–969 CrossRef Medline
27. Landini, M. P. (1984) Early enhanced glucose uptake in human cytomegalovirus-infected cells. *J. Gen. Virol.* **65,** 1229–1232 CrossRef Medline
28. Krawczyk, C. M., Holowka, T., Sun, J., Blagih, J., Amiel, E., DeBerardinis, R. J., Cross, J. R., Jung, E., Thompson, C. B., Jones, R. G., and Pearce, E. J. (2010) Toll-like receptor-induced changes in glycolytic metabolism regulate dendritic cell activation. *Blood* **115,** 4742–4749 CrossRef Medline

29. Wang, T., Marquardt, C., and Foker, J. (1976) Aerobic glycolysis during lymphocyte proliferation. *Nature* **261,** 702–705 CrossRef Medline
30. Warburg, O., Wind, F., and Negelein, E. (1927) The metabolism of tumors in the body. *J. Gen. Physiol.* **8,** 519–530 CrossRef Medline
31. Yu, Y., Clippinger, A. J., and Alwine, J. C. (2011) Viral effects on metabolism: changes in glucose and glutamine utilization during human cytomegalovirus infection. *Trends Microbiol.* **19,** 360–367 CrossRef Medline
32. Cairns, R. A., Harris, I. S., and Mak, T. W. (2011) Regulation of cancer cell metabolism. *Nat. Rev. Cancer* **11,** 85–95 CrossRef Medline
33. Thorens, B., and Mueckler, M. (2010) Glucose transporters in the 21st Century. *Am. J. Physiol. Endocrinol. Metab.* **298,** E141–E145 CrossRef Medline
34. Li, M. V., Chen, W., Harmancey, R. N., Nuotio-Antar, A. M., Imamura, M., Saha, P., Taegtmeyer, H., and Chan, L. (2010) Glucose-6-phosphate mediates activation of the carbohydrate responsive binding protein (ChREBP). *Biochem. Biophys. Res. Commun.* **395,** 395–400 CrossRef Medline
35. Geiger, R., Rieckmann, J. C., Wolf, T., Basso, C., Feng, Y., Fuhrer, T., Kogadeeva, M., Picotti, P., Meissner, F., Mann, M., Zamboni, N., Sallusto, F., and Lanzavecchia, A. (2016) L-Arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell* **167,** 829–842.e13 CrossRef Medline
36. Frauwirth, K. A., Riley, J. L., Harris, M. H., Parry, R. V., Rathmell, J. C., Plas, D. R., Elstrom, R. L., June, C. H., and Thompson, C. B. (2002) The CD28 signaling pathway regulates glucose metabolism. *Immunity* **16,** 769–777 CrossRef Medline
37. Warburg, O. (1956) On the origin of cancer cells. *Science* **123,** 309–314 CrossRef Medline
38. Hirschey, M. D., DeBerardinis, R. J., Diehl, A. M. E., Drew, J. E., Frezza, C., Green, M. F., Jones, L. W., Ko, Y. H., Le, A., Lea, M. A., Locasale, J. W., Longo, V. D., Lyssiotis, C. A., McDonnell, E., Mehrmohamadi, M., *et al.* (2015) Dysregulated metabolism contributes to oncogenesis. *Semin. Cancer Biol.* **35,** S129–S150 CrossRef Medline
39. Vander Heiden, M. G., Cantley, L. C., and Thompson, C. B. (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324,** 1029–1033 CrossRef Medline
40. Freckmann, G., Hagenlocher, S., Baumstark, A., Jendrike, N., Gillen, R. C., Rössner, K., and Haug, C. (2007) Continuous glucose profiles in healthy subjects under everyday life conditions and after different meals. *J. Diabetes Sci. Technol.* **1,** 695–703 Medline
41. Hirayama, A., Kami, K., Sugimoto, M., Sugawara, M., Toki, N., Onozuka, H., Kinoshita, T., Saito, N., Ochiai, A., Tomita, M., Esumi, H., and Soga, T. (2009) Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry. *Cancer Res.* **69,** 4918–4925 CrossRef Medline
42. Yu, Y., Maguire, T. G., and Alwine, J. C. (2011) Human cytomegalovirus activates glucose transporter 4 expression to increase glucose uptake during infection. *J. Virol.* **85,** 1573–1580 CrossRef Medline
43. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., and Wiley, D. C. (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329,** 506–512 CrossRef Medline
44. Li, P., Willie, S. T., Bauer, S., Morris, D. L., Spies, T., and Strong, R. K. (1999) Crystal structure of the MHC class I homolog MIC-A, a γδ T cell ligand. *Immunity* **10,** 577–584 CrossRef Medline
45. Valli, A., Rodriguez, M., Moutsianas, L., Fischer, R., Fedele, V., Huang, H.-L., Van Stiphout, R., Jones, D., McCarthy, M., Vinaxia, M., Igarashi, K., Sato, M., Soga, T., Buffa, F., Mccullagh, J., *et al.* (2015) Hypoxia induces a lipogenic cancer cell phenotype via HIF1α-dependent and -independent pathways. *Oncotarget* **6,** 1920–1941 CrossRef Medline
46. Reinicke, K., Sotomayor, P., Cisterna, P., Delgado, C., Nualart, F., and Godoy, A. (2012) Cellular distribution of glut-1 and glut-5 in benign and malignant human prostate tissue. *J. Cell. Biochem.* **113,** 553–562 CrossRef Medline
47. Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., and Nishioka, T. (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Proteome Res.* **2,** 488–494 CrossRef Medline
48. Soga, T., Igarashi, K., Ito, C., Mizobuchi, K., Zimmermann, H.-P., and Tomita, M. (2009) Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry. *Anal. Chem.* **81,** 6165–6174 CrossRef Medline
49. Adam, J., Hatipoglu, E., O'Flaherty, L., Ternette, N., Sahgal, N., Lockstone, H., Baban, D., Nye, E., Stamp, G. W., Wolhuter, K., Stevens, M., Fischer, R., Carmeliet, P., Maxwell, P. H., Pugh, C. W., *et al.* (2011) Renal cyst formation in Fh1-deficient mice is independent of the Hif/Phd pathway: roles for fumarate in KEAP1 succination and Nrf2 signaling. *Cancer Cell* **20,** 524–537 CrossRef Medline
50. Xia, J., Sinelnikov, I. V., Han, B., and Wishart, D. S. (2015) MetaboAnalyst 3.0–making metabolomics more meaningful. *Nucleic Acids Res.* **43,** W251–W257 CrossRef Medline

Appendix 2

Participant information leaflet and consent form

On the following pages are the participant information leaflet and a blank copy of the consent form completed by participants in this study.

Participant information leaflet

Research study: **AtherOx1**

A study of the underlying causes of heart disease, strokes and other vascular diseases

We would like to invite you to take part in the AtherOx1 research study. This leaflet explains why the research is being done and what it would involve for you. Please read the following information carefully. Ask us if there is anything that is not clear or if you would like more information.

What is the purpose of the study?

The purpose of AtherOx1 is to perform research to increase our understanding of the causes of vascular diseases, such as heart attacks and strokes. This may lead to the development of new approaches to the treatment of vascular disease.

These vascular diseases are the leading cause of death in the UK and worldwide. In this study we will take blood from participants and use cells and other components of blood in order to simulate the processes that happen during the development of vascular disease.

The study is being undertaken as part of an educational project at the University of Oxford.

Why have I been invited?

You have been invited because we would like to recruit healthy volunteers from the general population.

Do I have to take part?

No. You do not have to take part. It is up to you to decide whether you wish to. If you decide to take part you are free to withdraw at any time, without giving a reason.

What will happen to me if I take part?

If you decide to take part then we would ask you for a blood sample of up to 50mls. This is up to about 3 tablespoons which is similar to the amount taken when people have blood tests for medical reasons. We may also ask you for a saliva sample which involves rinsing the mouth with a small amount of water and spitting it into a collection tube.

Overall, your involvement in the study would mean that you would typically have 2 appointments lasting 15 minutes each. One appointment is to discuss taking part in the study and the second appointment is for completing the consent form and having blood and/or saliva taken. Usually these appointments are one or two days apart. We can also combine both appointments into one appointment.

Sometimes it might be useful for us to take more blood at a later date. However, if you agree to the first blood test you are not under any obligation to provide another sample later on. You can simply say no without any reason.

Expenses and payments

We cannot provide expense or payment for participation in our study but will be able to provide travel expenses if you have made a special trip to take part in the study.

What is the drug, device or procedure that is being tested?

We are not testing any drug, device or procedure.

What are the possible disadvantages and risks of taking part?

Having blood taken can cause discomfort and can cause minor bruising. Blood is only taken by qualified doctors or trained nurses.

What are the possible benefits of taking part?

There are no direct benefits to you for taking part in this study.

However, we hope that the information gathered in this study will improve our understanding of heart disease and strokes and may eventually lead to new or improved treatments for these conditions.

What if there is a problem?

Given the nature of this study, it is highly unlikely that you will suffer harm by taking part. However, the University has arrangements in place to provide for harm arising from participation in the study for which the University is the Research Sponsor.

If you wish to complain about any aspect of the way in which you have been approached or treated during the course of this study, you should contact Professor Chris O'Callaghan on 01865 287794 or email chris.ocallaghan@ndm.ox.ac.uk or you may contact the University of

Oxford Clinical Trials and Research Governance (CTRG) office on 01865 572224 or the head of CTRG, email ctrg@admin.ox.ac.uk

Will my taking part in the study be kept confidential?

Yes. We will follow strict ethical and legal practice and all information about you will be handled in complete confidence.

Responsible members of the University of Oxford may be given access to data for monitoring and/or audit of the study to ensure we are complying with regulations.

What will happen to any samples / data I give?

Any samples or data that are collected will be labelled with a unique sample number, but not your name or other personal details. This data is therefore stored completely anonymously.

If you agree to us contacting you again for a future sample then your contact details will be stored securely (email and/or postal address).

All research will be conducted ethically and responsibly according to review and approval from the Research Ethics Committee.

Your blood sample will be processed to extract cells and components such as cholesterol. We may grow the cells in the laboratory and do experiments with them which may mimic the process of atherosclerosis. The cells may then be analysed in a variety of ways some of which give information about how DNA is used during atherosclerosis. Saliva provides an additional rich source of DNA.

We may deposit completely anonymous data in public databanks because it can be useful for other researchers but this data will be anonymised and no personal details will be associated with it. At the end of the study we will destroy any samples of your original cells. We may keep indefinitely and anonymously cells that we have derived from your original cells – these are called 'cell lines'. We may also keep indefinitely completely anonymous samples of your DNA that we have extracted from your cells.

Will any additional tests be done?

We may perform a range of laboratory tests on the blood and/or saliva sample you provide including of the blood cells and other components of blood and including sequencing of DNA.

Will you tell me if you find something wrong with me?

The researchers will not be able to identify you because the samples are anonymous so it will not be possible to give you any information. Also, we are not performing clinical tests and the experiments that we are doing, including DNA sequencing, are not done to the standard required for a clinical test. For example, laboratory DNA sequencing techniques produce many errors which do not affect our study results but make the data unsuitable for use as a diagnostic test.

What will happen to the results of the research study?

We will publish the results of our study in peer-reviewed journals.

Who is organising and funding the research?

This research study has been devised and will be organised and carried out by members of the research group of Professor Chris O'Callaghan in the Nuffield Department of Clinical Medicine in the University of Oxford.

This research is funded by the Wellcome Trust.

Who has reviewed the study?

Scientific review of this research has been carried out on behalf of the Wellcome Trust as part of the process of securing funding.

The research has been reviewed by an independent group of people, called a Research Ethics Committee who consider your safety, rights, wellbeing and dignity. This study has been reviewed and given favourable opinion by the South Central – Hampshire B Research Ethics Committee.

Contact information

This research is being carried out by the research group of Professor Chris O'Callaghan, who is Professor of Medicine at the University of Oxford in the Nuffield Department of Medicine. If you have any questions or queries please contact us.

Contact details of Dr Anil Chalisey – researcher:

Tel +44 (0)1865 287794
Email chalisey@well.ox.ac.uk

What will happen at the end of the study?

At the end of the study we will erase any contact details we have for you. We will also destroy any living cells from your blood samples which may have been stored during the study. Sometimes we may produce 'cell lines' which are laboratory strains of cells derived from your cells. We may continue to use these in research after this study ends. We will also keep anonymous samples of your DNA and protein. It is useful to keep these samples if, for example, results need to be checked or repeated at a later date.

What will happen if I do not wish to carry on with the study?

If you wish to withdraw from the study, you may do so at any stage without giving a reason and we will destroy all identifiable data if you wish us to do so.

Please keep this information sheet for your own records.

Thank you for taking the time to read this information.

UNIVERSITY OF OXFORD

Consent form: *AtherOx1*

Chief Investigator: Professor Christopher O'Callaghan

Thank you for reading the Participant Information Leaflet and asking any questions that you might have had. If you would like to participate please respond to each of the following questions by initialling 'I agree' and then sign the form.

1.	I have read and understand the Information Leaflet dated _____ (Version_____) for the above study. I have had the opportunity to consider the information, ask questions, and have had these answered satisfactorily.	I agree
2.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, and without affecting my medical care or legal rights.	I agree
3.	I agree to give a blood **and/or saliva** for research purposes. I consider this sample a gift to the University of Oxford and I understand I will not gain any direct personal benefit from this.	I agree
4.	I understand that some or all of the DNA in the blood **and/or saliva** that I give may be sequenced and stored indefinitely for use in future studies.	I agree
5.	I understand that completely anonymous data may be deposited in public databanks for use by other researchers.	I agree
6.	I understand that at the end of the study my original blood cells will be destroyed but the researchers may keep new cells which they have derived from my original cells anonymously.	I agree
7.	I understand that the blood **and/or saliva** samples I give will be fully anonymised so any results from testing this sample will not be linked to me.	I agree
8.	I understand that the researchers are not performing clinical tests and I will not receive any personal information about the tests performed on blood **and/or saliva** I donate.	I agree
9.	I agree to take part in the study.	I agree
10.	I agree to be approached to provide a further blood sample (optional).	I agree

_____	_____	_____
Name of participant	Date	Signature
_____	_____	_____
Name of person taking consent	Date	Signature

Bibliography

1. Herrington, W., Lacey, B., Sherliker, P., Armitage, J. & Lewington, S. Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease. *Circ. Res.* **118**, 535–546 (2016).

2. Naghavi, M. *et al.* Global, regional, and national age-sex specifc mortality for 264 causes of death, 1980-2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).

3. Stary, H. *et al.* A Definition of Initial , Fatty Streak , and Intermediate Lesions of Atherosclerosis. *Circulation* **89**, 2462–2478 (1994).

4. Bentzon, J. F., Otsuka, F., Virmani, R. & Falk, E. Mechanisms of plaque formation and rupture. *Circ. Res.* **114**, 1852–1866 (2014).

5. Chatzizisis, Y. S. *et al.* Role of Endothelial Shear Stress in the Natural History of Coronary Atherosclerosis and Vascular Remodeling. Molecular, Cellular, and Vascular Behavior. *J. Am. Coll. Cardiol.* **49**, 2379–2393 (2007).

6. Kwak, B. R. *et al.* Biomechanical factors in atherosclerosis: Mechanisms and clinical implications. *Eur. Heart J.* **35**, 3013–3020 (2014).

7. Ference, B. A. *et al.* Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement fromthe European Atherosclerosis Society Consensus Panel. *Eur. Heart J.* **38**, 2459–2472 (2017).

8. Pentikäinen, M. O., Öörni, K., Ala-Korpela, M. & Kovanen, P. T. Modified LDL - Trigger of atherosclerosis and inflammation in the arterial intima. *J. Intern. Med.* **247**, 359–370 (2000).

9. Yu, X.-H., Fu, Y.-C., Zhang, D.-W., Yin, K. & Tang, C.-K. Foam cells in atherosclerosis. *Clin. Chim. Acta* **424**, 245–252 (2013).

10. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).

11. Nikpay, M. *et al.* A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

12. Turner, A. W., Wong, D., Dreisbach, C. N. & Miller, C. L. GWAS Reveal Targets in Vessel Wall Pathways to Treat Coronary Artery Disease. *Front. Cardiovasc. Med.* **5**, 1–10 (2018).

13. Pim, V. der H. & Niek, V. Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. *Circ. Res.* **122**, 433–443 (2018).

14. Pi, X., Xie, L. & Patterson, C. Emerging roles of vascular endothelium in metabolic homeostasis. *Circ. Res.* **123**, 477–494 (2018).

15. Blankenberg, S., Barbaux, S. & Tiret, L. Adhesion molecules and atherosclerosis. *Atherosclerosis* **170**, 191–203 (2003).

16. Denninger, J. W. & Marletta, M. A. Guanylate cyclase and the .NO/cGMP signaling pathway. *Biochim. Biophys. Acta - Bioenerg.* **1411**, 334–350 (1999).

17. Yau, J. W., Teoh, H. & Verma, S. Endothelial cell control of thrombosis. *BMC Cardiovasc. Disord.* **15**, 1–11 (2015).

18. Förstermann, U. & Münzel, T. Endothelial nitric oxide synthase in vascular disease: From marvel to menace. *Circulation* **113**, 1708–1714 (2006).

19. Gimbrone, M. A. & García-Cardeña, G. Endothelial Cell Dysfunction and the Pathobiology of Atherosclerosis. *Circ. Res.* **118**, 620–636 (2016).

20. Liao, J. K. & Liao, J. K. Linking endothelial dysfunction with endothelial cell activation. **123**, 540–541 (2013).

21. Bonetti, P. O., Lerman, L. O. & Lerman, A. Endothelial dysfunction: A marker of atherosclerotic risk. *Arterioscler. Thromb. Vasc. Biol.* **23**, 168–175 (2003).

22. Dejana, E. Endothelial cell-cell junctions: Happy together. *Nat. Rev. Mol. Cell Biol.* **5**, 261–270 (2004).

23. Mundi, S. *et al.* Endothelial permeability, LDL deposition, and cardiovascular risk factors-A review. *Cardiovasc. Res.* **114**, 35–52 (2018).

24. Kraehling, J. R. *et al.* Genome-wide RNAi screen reveals ALK1 mediates LDL uptake and transcytosis in endothelial cells. *Nat. Commun.* **7**, (2016).

25. Lin, S. J., Jan, K. M. & Chien, S. Role of dying endothelial cells in transendothelial macromolecular transport. *Arterioscler. Thromb. Vasc. Biol.* **10**, 703–709 (1990).

26. Weinbaum, S., Tzeghai, G. & Ganatos, P. Effect of cell turnover and leaky junctions on arterial macromolecular transport. *Am. J. Physiol. - Hear. Circ. Physiol.* **17**, (1985).

27. Zhang, C. The role of inflammatory cytokines in endothelial dysfunction. *Basic Res. Cardiol.* **103**, 398–406 (2008).

28. Apostolakis, S., Vogiatzi, K., Amanatidou, V. & Spandidos, D. A. Interleukin 8 and cardiovascular disease. *Cardiovasc. Res.* **84**, 353–360 (2009).

29. Lin, D., Lavender, H., Soilleux, E. J. & Callaghan, C. A. O. NF-κB Regulates MICA Gene Transcription in Endothelial Cell through a Genetically Inhibitable Control Site. *J. Biol. Chem.* **287**, 4299–4310 (2012).

30. Groh, V. *et al.* Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 12445–12450 (1996).

31. Bauer, S. *et al.* Activation of NK Cells and T Cells by NKG2D, a Receptor for Stress-Inducible MICA. *Science (80-.).* **285**, 727–730 (1999).

32. Chistiakov, D. A., Orekhov, A. N. & Bobryshev, Y. V. Effects of shear stress on endothelial cells: go with the flow. *Acta Physiol.* **219**, 382–408 (2017).

33. Zarins, C. K. *et al.* Carotid Bifurcation Atherosclerosis: Quantative Correlation of Plaque Localization with Flow Velocity Profiles and Wall Shear Stress. *Circ. Res.* **53**, 502–514 (1983).

34. Chatzizisis, Y. S. *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* **117**, 993–1002 (2008).

35. Tarbell, J. M., Shi, Z.-D., Dunn, J. & Jo, H. Fluid Mechanics, Arterial Disease, and Gene Expression. *Annu. Rev. Fluid Mech.* **46**, 591–614 (2014).

36. Pan, S. Molecular mechanisms responsible for the atheroprotective effects of laminar shear stress. *Antioxidants Redox Signal.* **11**, 1669–1682 (2009).

37. Ajami, N. E. *et al.* Systems biology analysis of longitudinal functional response of endothelial cells to shear stress. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10990–10995 (2017).

38. Xu, J. *et al.* GPR68 Senses Flow and Is Essential for Vascular Physiology. *Cell* **173**, 762-775.e16 (2018).

39. Li, J. *et al.* Piezo1 integration of vascular architecture with physiological force. *Nature* **515**, 279–282 (2014).

40. Tzima, E. *et al.* A mechanosensory complex that mediates the endothelial cell response to fluid shear stress. *Nature* **437**, 426–431 (2005).

41. Nayak, L., Lin, Z. & Jain, M. K. 'Go with the flow': How Krüppel-like factor 2 regulates the vasoprotective effects of shear stress. *Antioxidants Redox Signal.* **15**, 1449–1461 (2011).

42. Dekker, R. J. *et al.* Prolonged fluid shear stress induces a distinct set of endothelial cell genes, most specifically lung Krüppel-like factor (KLF2). *Blood* **100**, 1689–1698 (2002).

43. Nakajima, H. & Mochizuki, N. Flow pattern-dependent endothelial cell responses through transcriptional regulation. *Cell Cycle* **16**, 1893–1901 (2017).

44. Dunn, J., Thabet, S. & Jo, H. Flow-dependent epigenetic DNA methylation in endothelial gene expression and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **35**, 1562–1569 (2015).

45. Qiao, J. H. *et al.* Role of macrophage colony-stimulating factor in atherosclerosis: Studies of osteopetrotic mice. *Am. J. Pathol.* **150**, 1687–1699 (1997).

46. Rajavashisth, T. *et al.* Heterozygous osteopetrotic (op) mutation reduces atherosclerosis in LDL receptor-deficient mice. *J. Clin. Invest.* **101**, 2702–2710 (1998).

47. Clinton, S. K. *et al.* Macrophage colony-stimulating factor gene expression in vascular cells and in experimental and human atherosclerosis. *Am. J. Pathol.* **140**, 301–316 (1992).

48. Xue, J. *et al.* Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. *Immunity* **40**, 274–288 (2014).

49. Murray, P. J. *et al.* Macrophage Activation and Polarization: Nomenclature and Experimental Guidelines. *Immunity* **41**, 14–20 (2014).

50. Zingg, Roberta Ricciarelli, Angelo, J.-M. Scavenger Receptors and Modified Lipoproteins: Fatal Attractions? *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **49**, 397–403 (2000).

51. Kennedy, D. J. *et al.* Dietary cholesterol plays a role in cd36-mediated atherogenesis in LDLR-knockout mice. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1481–1487 (2009).

52. Manning-Tobin, J. J. *et al.* Loss of SR-A and CD36 activity reduces atherosclerotic lesion complexity without abrogating foam cell formation in hyperlipidemic mice. *Arterioscler. Thromb. Vasc. Biol.* **29**, 19–26 (2009).

53. Reschen, M. E. *et al.* Lipid-Induced Epigenomic Changes in Human Macrophages Identify a Coronary Artery Disease-Associated Variant that Regulates PPAP2B Expression through Altered C/EBP-Beta Binding. *PLOS Genet.* **11**, e1005061 (2015).

54. Liu, W., Yin, Y., Zhou, Z., He, M. & Dai, Y. OxLDL-induced IL-1beta secretion promoting foam cells formation was mainly via CD36 mediated ROS production

leading to NLRP3 inflammasome activation. *Inflamm. Res.* **63**, 33–43 (2014).

55. Wang, N. *et al.* Interleukin 8 is induced by cholesterol loading of macrophages and expressed by macrophage foam cells in human atheroma. *J. Biol. Chem.* **271**, 8837–8842 (1996).

56. Trogan, E. *et al.* Gene expression changes in foam cells and the role of chemokine receptor CCR7 during atherosclerosis regression in ApoE-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3781–3786 (2006).

57. Merhi-Soussi, F. *et al.* Interleukin-1 plays a major role in vascular inflammation and atherosclerosis in male apolipoprotein E-knockout mice. *Cardiovasc. Res.* **66**, 583–593 (2005).

58. Boring, L., Gosling, J., Cleary, M. & Charo, I. F. Decreased lesion formation in CCR2−/− mice reveals a role for chemokines in the initiation of atherosclerosis. *Nature* **394**, 894–897 (1998).

59. Park, Y. M. CD36, a scavenger receptor implicated in atherosclerosis. *Exp. Mol. Med.* **46**, e99-7 (2014).

60. Han, J., Hajjar, D. P., Febbraio, M. & Nicholson, A. C. Native and modified low density lipoproteins increase the functional expression of the macrophage class B scavenger receptor, CD36. *J. Biol. Chem.* **272**, 21654–21659 (1997).

61. Feng, J. *et al.* Induction of CD36 expression by oxidized LDL and IL-4 by a common signaling pathway dependent on protein kinase C and PPAR-γ. *J. Lipid Res.* **41**, 688–698 (2000).

62. Johnson, J. L. Emerging regulators of vascular smooth muscle cell function in the development and progression of atherosclerosis. *Cardiovasc. Res.* **103**, 452–460 (2014).

63. Bennett, M. R., Sinha, S. & Owens, G. K. Vascular Smooth Muscle Cells in Atherosclerosis. *Circ. Res.* **118**, 692–702 (2016).

64. Kiyan, Y. *et al.* OxLDL induces inflammatory responses in vascular smooth muscle cells via urokinase receptor association with CD36 and TLR4. *J. Mol.*

Cell. Cardiol. **66**, 72–82 (2014).

65. Tse, K., Tse, H., Sidney, J., Sette, A. & Ley, K. T cells in atherosclerosis. *Int. Immunol.* **25**, 615–622 (2013).

66. Shoenfeld, Y., Wu, R., Dearing, L. D. & Matsuura, E. Are anti-oxidized low-density lipoprotein antibodies pathogenic or protective? *Circulation* **110**, 2552–2558 (2004).

67. Stemme, S. *et al.* T lymphocytes from human atherosclerotic plaques recognize oxidized low density lipoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3893–3897 (1995).

68. Harvey, E. J. & Ramji, D. P. Interferon-γ and atherosclerosis: Pro- or anti-atherogenic? *Cardiovasc. Res.* **67**, 11–20 (2005).

69. Mallat, Z. *et al.* Expression of interleukin-18 in human atherosclerotic plaques and relation to plaque instability. *Circulation* **104**, 1598–1603 (2001).

70. Lee, T. S., Yen, H. C., Pan, C. C. & Chau, L. Y. The role of interleukin 12 in the development of atherosclerosis in apoE-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **19**, 734–742 (1999).

71. Galkina, E. & Ley, K. Immune and Inflammatory Mechanisms of Atherosclerosis. *Annu. Rev. Immunol.* **27**, 165–197 (2009).

72. Schoenfelder, S. & Fraser, P. Long-range enhancer–promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).

73. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* (2019). doi:10.1038/s41576-019-0150-2

74. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

75. Venkatesh, S. & Workman, J. L. Histone exchange , chromatin structure and the regulation of transcription E i V. *Nat. Publ. Gr.* **16**, 178–189 (2015).

76. Happel, N. & Doenecke, D. Histone H1 and its isoforms: Contribution to

chromatin structure and function. *Gene* **431**, 1–12 (2009).

77. Tsompana, M. & Buck, M. J. Chromatin accessibility: A window into the genome. *Epigenetics and Chromatin* **7**, 1–16 (2014).

78. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

79. Adey, A. *et al.* Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol.* **11**, R119 (2010).

80. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381–395 (2011).

81. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).

82. Liang, G. *et al.* Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7357–7362 (2004).

83. Belton, J. M. *et al.* Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).

84. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369-1384.e19 (2016).

85. Qu, H. & Fang, X. A Brief Review on the Human Encyclopedia of DNA Elements (ENCODE) Project. *Genomics, Proteomics Bioinforma.* **11**, 135–141 (2013).

86. Satterlee, J. S. *et al.* The NIH Common Fund/Roadmap Epigenomics Program: Successes of a comprehensive consortium. *Sci. Adv.* **5**, 1–7 (2019).

87. Kodzius, R. *et al.* Cage: Cap analysis of gene expression. *Nat. Methods* **3**, 211 (2006).

88. Forrest, A. R. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

89. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).

90. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014).

91. Weinhold, N. *et al.* Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science (80-.).* **347**, 1010–1015 (2015).

92. Bondareva, O. *et al.* Identification of atheroprone shear stress responsive regulatory elements in endothelial cells. *Cardiovasc. Res.* **115**, 1487–1499 (2019).

93. Lee, D.-Y. & Chiu, J.-J. Atherosclerosis and flow: roles of epigenetic modulation in vascular endothelium. *J. Biomed. Sci.* **26**, 1–17 (2019).

94. Miller, C. L. *et al.* Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **7**, (2016).

95. Lane, W. O. *et al.* Parallel-plate Flow Chamber and Continuous Flow Circuit to Evaluate Endothelial Progenitor Cells under Laminar Flow Shear Stress. *J. Vis. Exp.* e3349 (2012). doi:10.3791/3349

96. Kaisers, W. seqTools: Analysis of nucleotide, sequence and quality content on fastq files. (2019).

97. Huber, W. *et al.* Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).

98. Babraham Bioinformatics. FastQC https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (2019).

99. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

100. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

101. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

102. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, (2013).

103. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).

104. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297 (2012).

105. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, (2010).

106. Lund, S. P., Dan, N., J, M. D. & K, S. G. Detecting Differential Expression in RNA-sequence Data Using Quasi-likelihood with Shrunken Dispersion Estimates. *Statistical Applications in Genetics and Molecular Biology* **11**, (2012).

107. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* 21.29.1-21.29.9 (2015). doi:10.1002/0471142727.mb2129s109

108. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

109. Li, Y. L. *et al.* PEAT: An intelligent and efficient paired-end sequencing adapter trimming algorithm. *BMC Bioinformatics* **16**, 1–11 (2015).

110. Broad Institute. Picard Toolkit. *GitHub Repository* http://broadinstitute.github.io/picard/ (2019).

111. Lun, A. T. L. & Smyth, G. K. Csaw: A Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res.* **44**, 1–10 (2015).

112. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

113. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).

114. Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

115. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).

116. James Kent, W. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

117. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

118. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol. Cell* **38**, 576–589 (2010).

119. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975 (2017).

120. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

121. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and

population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

122. Ambrosini, G. & Pe, R. C. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Res.* **41**, 157–164 (2013).

123. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

124. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).

125. Lun, A. T. L. *et al.* EmptyDrops: Distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* **20**, 63 (2019).

126. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

127. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv* 576827 (2019). doi:10.1101/576827

128. Shen, L. & Sinai, M. GeneOverlap: Test and visualize gene overlaps. (2019).

129. Chait, A. & Bornfeldt, K. E. Diabetes and atherosclerosis : is there a role for hyperglycemia ? *J. Lipid Res.* 335–339 (2009). doi:10.1194/jlr.R800059-JLR200

130. Patel, A., MacMahon, S., Chalmers, J. & Neal, B. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes. *N. Engl. J. Med.* **358**, 2560–2572 (2008).

131. Holman, R. R., Paul, S. K., Bethel, M. A., Matthews, D. R. & Neil, H. A. W. 10-Year Follow-up of Intensive Glucose Control in Type 2 Diabetes. *N. Engl. J. Med.* **359**, 1577–1589 (2008).

132. Effects of Intensive Glucose Lowering in Type 2 Diabetes. *N. Engl. J. Med.* **358**, 2545–2559 (2008).

133. Renard, C. B. *et al.* Diabetes and diabetes-associated lipid abnormalities have distinct effects on initiation and progression of atherosclerotic lesions. *J. Clin.*

Invest. **114**, 659–668 (2004).

134. Shanmugam, N., Reddy, M. A., Guha, M. & Natarajan, R. High glucose-induced expression of proinflammatory cytokine and chemokine genes in monocytic cells. *Diabetes* **52**, 1256–1264 (2003).

135. Miao, F. *et al.* RNA-sequencing analysis of high glucose-treated monocytes reveals novel transcriptome signatures and associated epigenetic profiles. *Physiol. Genomics* **45**, 287–299 (2013).

136. Griffin, E. *et al.* A link between diabetes and atherosclerosis: Glucose regulates expression of CD36 at the level of translation. *Nat. Med.* **7**, 840–846 (2001).

137. Li, L., Sawamura, T. & Renier, G. Glucose Enhances Human Macrophage LOX-1 Expression: Role for LOX-1 in Glucose-Induced Macrophage Foam Cell Formation. *Circ. Res.* **94**, 892–901 (2004).

138. Rahaman, S. O. *et al.* A CD36-dependent signaling cascade is necessary for macrophage foam cell formation. *Cell Metab.* **4**, 211–221 (2006).

139. Stein, S. *et al.* SIRT1 decreases Lox-1-mediated foam cell formation in atherogenesis. *Eur. Heart J.* **31**, 2301–2309 (2010).

140. Freemerman, A. J. *et al.* Metabolic reprogramming of macrophages: Glucose transporter 1 (GLUT1)-mediated glucose metabolism drives a proinflammatory phenotype. *J. Biol. Chem.* **289**, 7884–7896 (2014).

141. Cheng, S. C. *et al.* MTOR- and HIF-1α-mediated aerobic glycolysis as metabolic basis for trained immunity. *Science (80-.).* **345**, (2014).

142. Diskin, C. & Pålsson-McDermott, E. M. Metabolic modulation in macrophage effector function. *Front. Immunol.* **9**, 1–17 (2018).

143. Palsson-Mcdermott, E. M. *et al.* Pyruvate kinase M2 regulates Hif-1α activity and IL-1β induction and is a critical determinant of the warburg effect in LPS-activated macrophages. *Cell Metab.* **21**, 65–80 (2015).

144. McCarthy, M. T. *et al.* Purine nucleotide metabolism regulates expression of the

human immune ligand MICA. *J. Biol. Chem.* **293**, 3913–3924 (2018).

145. Tan, Z. *et al.* Pyruvate Dehydrogenase Kinase 1 Participates in Macrophage Polarization via Regulating Glucose Metabolism. *J. Immunol.* **194**, 6082–6089 (2015).

146. Yamashita, H. *et al.* A glucose-responsive transcription factor that regulates carbohydrate metabolism in the liver. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9116–9121 (2001).

147. Iizuka, K., Wu, W., Horikawa, Y. & Takeda, J. Role of glucose-6-phosphate and xylulose-5-phosphate in the regulation of glucose-stimulated gene expression in the pancreatic β cell line, INS-1E. *Endocr. J.* **60**, 473–482 (2013).

148. Dentin, R. *et al.* Glucose 6-phosphate, rather than xylulose 5-phosphate, is required for the activation of ChREBP in response to glucose in the liver. *J. Hepatol.* **56**, 199–209 (2012).

149. Ortega-Prieto, P. & Postic, C. Carbohydrate sensing through the transcription factor ChREBP. *Front. Genet.* **10**, 1–9 (2019).

150. Abdul-Wahed, A., Guilmeau, S. & Postic, C. Sweet Sixteenth for ChREBP: Established Roles and Future Goals. *Cell Metab.* **26**, 324–341 (2017).

151. Sarrazy, V. *et al.* Maintenance of Macrophage Redox Status by ChREBP Limits Inflammation and Apoptosis and Protects against Advanced Atherosclerotic Lesion Formation. *Cell Rep.* **13**, 132–144 (2015).

152. Esensten, J. H., Helou, Y. A., Chopra, G., Weiss, A. & Bluestone, J. A. CD28 Costimulation: From Mechanism to Therapy. *Immunity* **44**, 973–988 (2016).

153. Woerly, B. G. *et al.* Expression of CD28 and CD86 by Human Eosinophils and Role in the Secretion of Type 1 Cytokines (Interleukin 2 and Interferon γ): Inhibition by Immunoglobulin A Complexes. *J. Exp. Med.* **190**, 487–495 (1999).

154. Woerly, G. *et al.* Human eosinophils express and release IL-13 following CD28-dependent activation. *J. Leukoc. Biol.* **72**, (2002).

155. Woerly, G. *et al.* CD28 and secretory immunoglobulin A-dependent activation of eosinophils : inhibition of mediator release by the anti-allergic drug , suplatast tosilate. 1379–1387 (2004). doi:10.1111/j.1365-2222.2004.02036.x

156. Frauwirth, K. A. *et al.* The CD28 Signaling Pathway Regulates Glucose Metabolism. *Immunity* **16**, 769–777 (2002).

157. Jacobs, S. R. *et al.* Glucose Uptake Is Limiting in T Cell Activation and Requires CD28-Mediated Akt-Dependent and Independent Pathways. *J. Immunol.* **180**, 4476–4486 (2008).

158. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

159. Lun, A. T. L. & Smyth, G. K. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res.* **42**, (2014).

160. Sarwar, N. *et al.* Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: A collaborative meta-analysis of 102 prospective studies. *Lancet* **375**, 2215–2222 (2010).

161. Farmer, J. A. Diabetic dyslipidemia and atherosclerosis: Evidence from clinical trials. *Curr. Diab. Rep.* **8**, 71–77 (2008).

162. Gonzalez, Y. *et al.* High glucose concentrations induce TNF-α production through the down-regulation of CD33 in primary human monocytes. *BMC Immunol.* **13**, (2012).

163. Pan, Y. *et al.* Inhibition of high glucose-induced inflammatory response and macrophage infiltration by a novel curcumin derivative prevents renal injury in diabetic rats. *Br. J. Pharmacol.* **166**, 1169–1182 (2012).

164. Dasu, M. R., Devaraj, S. & Jialal, I. High glucose induces IL-1β expression in human monocytes: Mechanistic insights. *Am. J. Physiol. - Endocrinol. Metab.* **293**, 337–346 (2007).

165. Tedesco, S. *et al.* Convenience versus biological significance: Are PMA-

differentiated THP-1 cells a reliable substitute for blood-derived macrophages when studying in vitro polarization? *Front. Pharmacol.* **9**, 1–13 (2018).

166. Li, M.-F. *et al.* High Glucose Increases the Expression of Inflammatory Cytokine Genes in Macrophages Through H3K9 Methyltransferase Mechanism. *J. Interf. Cytokine Res.* **36**, 48–61 (2016).

167. Tabas, I. & Lichtman, A. H. Monocyte-Macrophages and T Cells in Atherosclerosis. *Immunity* **47**, 621–634 (2017).

168. Wilson, H. M. Macrophages heterogeneity in atherosclerosis - implications for therapy. *J. Cell. Mol. Med.* **14**, 2055–2065 (2010).

169. Gordon, S. & Taylor, P. R. Monocyte and macrophage heterogeneity. *Nat. Rev. Immunol.* **5**, 953–964 (2005).

170. Smythies, L. *et al.* Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J. Clin. Invest.* **115**, 66–75 (2005).

171. Sica, A. & Mantovani, A. Macrophage plasticity and polarization: in vivo veritas. *J. Clin. Invest.* **122**, 787–795 (2012).

172. Hamers, A. A. J. *et al.* Human Monocyte Heterogeneity as Revealed by High-Dimensional Mass Cytometry. *Arterioscler. Thromb. Vasc. Biol.* **39**, 25–36 (2019).

173. Wong, K. L. *et al.* Gene expression profiling reveals the defining features of the classical, intermediate,and nonclassical human monocyte subsets. *Blood* **118**, 16–32 (2011).

174. Patel, A. A. *et al.* The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).

175. Ziegler-Heitbrock, L. The CD14+ CD16+ blood monocytes: their role in infection and inflammation. *J. Leukoc. Biol.* **81**, 584–592 (2007).

176. Belge, K.-U. *et al.* The Proinflammatory CD14+ CD16+ DR++ Monocytes Are a

Major Source of TNF. *J. Immunol.* **168**, 3536–3542 (2002).

177. Hijdra, D., Vorselaars, A. D. M., Grutters, J. C., Claessen, A. M. E. & Rijkers, G. T. Phenotypic Characterization of Human Intermediate Monocytes. *Front. Immunol.* **4**, 4–6 (2013).

178. Thomas, G., Tacke, R., Hedrick, C. C. & Hanna, R. N. Nonclassical Patrolling Monocyte Function in the Vasculature. *Arterioscler. Thromb. Vasc. Biol.* **35**, 1306–1316 (2015).

179. Collison, J. L., Carlin, L. M., Eichmann, M., Geissmann, F. & Peakman, M. Heterogeneity in the Locomotory Behavior of Human Monocyte Subsets over Human Vascular Endothelium In Vitro. *J. Immunol.* **195**, 1162–1170 (2015).

180. Swirski, F. K. *et al.* Ly-6Chi monocytes dominate hypercholesterolemia-associated monocytosis and give rise to macrophages in atheromata. *J. Clin. Invest.* **117**, 195–205 (2007).

181. Combadière, C. *et al.* Combined inhibition of CCL2, CX3CR1, and CCR5 abrogates Ly6Chi and Ly6Clo monocytosis and almost abolishes atherosclerosis in hypercholesterolemic mice. *Circulation* **117**, 1649–1657 (2008).

182. Saederup, N., Chan, L., Lira, S. A. & Charo, I. F. Fractalkine deficiency markedly reduces macrophage accumulation and atherosclerotic lesion formation in CCR2-/- mice: Evidence for independent chemokine functions in atherogenesis. *Circulation* **117**, 1642–1648 (2008).

183. Moore, K. J., Sheedy, F. J. & Fisher, E. A. Macrophages in atherosclerosis: A dynamic balance. *Nat. Rev. Immunol.* **13**, 709–721 (2013).

184. Rogacev, K. S. *et al.* CD14++CD16+ monocytes independently predict cardiovascular events: A cohort study of 951 patients referred for elective coronary angiography. *J. Am. Coll. Cardiol.* **60**, 1512–1520 (2012).

185. Mills, C. D., Kincaid, K., Alt, J. M., Heilman, M. J. & Hill, A. M. M-1/M-2 Macrophages and the Th1/Th2 Paradigm. *J. Immunol.* **164**, 6166–6173 (2000).

186. Kawai, T. & Akira, S. Toll-like Receptors and Their Crosstalk with Other Innate

Receptors in Infection and Immunity. *Immunity* **34**, 637–650 (2011).

187. Hu, X., Chakravarty, S. D. & Ivashkiv, L. B. Regulation of interferon and Toll-like receptor signaling during macrophage activation by opposing feedforward and feedback inhibition mechanisms. *Immunol. Rev.* **226**, 41–56 (2008).

188. Nathan, C. F., Murray, H. W., Wlebe, M. E. & Rubin, B. Y. Identification of interferon-γ as the lymphokine that activates human macrophage oxidative metabolism and antimicrobial activity. *J. Exp. Med.* **158**, 670–689 (1983).

189. Wu, T. T., Chen, T. L. & Chen, R. M. Lipopolysaccharide triggers macrophage activation of inflammatory cytokine expression, chemotaxis, phagocytosis, and oxidative ability via a toll-like receptor 4-dependent pathway: Validated by RNA interference. *Toxicol. Lett.* **191**, 195–202 (2009).

190. Stein, M., Keshav, S., Harris, N. & Gordon, S. Interleukin 4 potently enhances murine macrophage mannose receptor activity: A marker of alternative immunologic macrophage activation. *J. Exp. Med.* **176**, 287–292 (1992).

191. Van Dyken, S. J. & Locksley, R. M. Interleukin-4- and Interleukin-13-Mediated Alternatively Activated Macrophages: Roles in Homeostasis and Disease. *Annu. Rev. Immunol.* **31**, 317–343 (2013).

192. Mantovani, A. *et al.* The chemokine system in diverse forms of macrophage activation and polarization. *Trends Immunol.* **25**, 677–686 (2004).

193. Van Linthout, S., Miteva, K. & Tschöpe, C. Crosstalk between fibroblasts and inflammatory cells. *Cardiovasc. Res.* **102**, 258–269 (2014).

194. Martinez, F. O. & Gordon, S. The M1 and M2 paradigm of macrophage activation: Time for reassessment. *F1000Prime Rep.* **6**, 1–13 (2014).

195. Mosser, D. M. & Edwards, J. P. Exploring the full spectrum of macrophage activation. *Nat. Rev. Immunol.* **8**, 958–969 (2008).

196. Tarique, A. A. *et al.* Phenotypic,functional,and plasticity features of classical and alternatively activated human macrophages. *Am. J. Respir. Cell Mol. Biol.* **53**, 676–688 (2015).

197. Davis, M. J. *et al.* Macrophage M1 / M2 Polarization Dynamically Adapts to Changes in Cytokine Microenvironments in Cryptococcus neoformans Infection. *MBio* **4**, 1–10 (2013).

198. Verreck, F. A. W. *et al.* Human IL-23-producing type 1 macrophages promote but IL-10-producing type 2 macrophages subvert immunity to (myco)bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4560–4565 (2004).

199. Lacey, D. C. *et al.* Defining GM-CSF– and Macrophage-CSF–Dependent Macrophage Responses by In Vitro Models. *J. Immunol.* **188**, 5752–5765 (2012).

200. Kadl, A. *et al.* Identification of a novel macrophage phenotype that develops in response to atherogenic phospholipids via Nrf2. *Circ. Res.* **107**, 737–746 (2010).

201. Waldo, S. W. *et al.* Heterogeneity of human macrophages in culture and in atherosclerotic plaques. *Am. J. Pathol.* **172**, 1112–1126 (2008).

202. Khallou-Laschet, J. *et al.* Macrophage plasticity in experimental atherosclerosis. *PLoS One* **5**, (2010).

203. Stöger, J. L. *et al.* Distribution of macrophage polarization markers in human atherosclerosis. *Atherosclerosis* **225**, 461–468 (2012).

204. Nagy, L., Tontonoz, P., Alvarez, J. G. A., Chen, H. & Evans, R. M. Oxidized LDL regulates macrophage gene expression through ligand activation of PPARγ. *Cell* **93**, 229–240 (1998).

205. Chawla, A. *et al.* A PPARγ-LXR-ABCA1 Pathway in Macrophages Is Involved in Cholesterol Efflux and Atherogenesis. *Mol. Cell* **7**, 161–171 (2001).

206. van Tits, L. J. H. *et al.* Oxidized LDL enhances pro-inflammatory responses of alternatively activated M2 macrophages: A crucial role for Krüppel-like factor 2. *Atherosclerosis* **214**, 345–349 (2011).

207. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* arXiv:1802.03426v2 (2018).

208. Aslanian, A. M. & Charo, I. F. Targeted disruption of the scavenger receptor and

chemokine CXCL16 accelerates atherosclerosis. *Circulation* **114**, 583–590 (2006).

209. Cochain, C. *et al.* Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.* **122**, 1661–1674 (2018).

210. Klinkner, A. M., Waites, C. R., Kerns, W. D. & Bugelski, P. J. Evidence of foam cell and cholesterol crystal formation in macrophages incubated with oxidized LDL by fluorescence and electron microscopy. *J. Histochem. Cytochem.* **43**, 1071–1078 (1995).

211. Victor, V. M. & De la Fuente, M. N-acetylcysteine improves in vitro the function of macrophages from mice with endotoxin-induced oxidative stress. *Free Radic. Res.* **36**, 33–45 (2002).

212. Aldini, G. *et al.* N-Acetylcysteine as an antioxidant and disulphide breaking agent: the reasons why. *Free Radic. Res.* **52**, 751–762 (2018).

213. Otero, K. *et al.* Macrophage colony-stimulating factor induces the proliferation and survival of macrophages via a pathway involving DAP12 and β-catenin. *Nat. Immunol.* **10**, 734–743 (2009).

214. Robbins, C. S. *et al.* Local proliferation dominates lesional macrophage accumulation in atherosclerosis. *Nat. Med.* **19**, 1166–1172 (2013).

215. Tang, J. *et al.* Inhibiting macrophage proliferation suppresses atherosclerotic plaque inflammation. *Sci. Adv.* **1**, (2015).

216. Wolak, T. Osteopontin - A multi-modal marker and mediator in atherosclerotic vascular disease. *Atherosclerosis* **236**, 327–337 (2014).

217. Isoda, K. *et al.* Osteopontin plays an important role in the development of medial thickening and neointimal formation. *Circ. Res.* **91**, 77–82 (2002).

218. Chiba, S. *et al.* Development of atherosclerosis in osteopontin transgenic mice. *Heart Vessels* **16**, 111–117 (2002).

219. Bruemmer, D. *et al.* Angiotensin II–accelerated atherosclerosis and aneurysm formation is attenuated in osteopontin-deficient mice. *J. Clin. Invest.* **112**, 1318–1331 (2003).

220. Stemme, S. *et al.* T lymphocytes from human atherosclerotic plaques recognize oxidized low density lipoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 3893–3897 (1995).

221. Shoenfeld, Y., Wu, R., Dearing, L. D. & Matsuura, E. Are anti-oxidized low-density lipoprotein antibodies pathogenic or protective? *Circulation* **110**, 2552–2558 (2004).

222. Lahoute, C., Herbin, O., Mallat, Z. & Tedgui, A. Adaptive immunity in atherosclerosis: Mechanisms and future therapeutic targets. *Nat. Rev. Cardiol.* **8**, 348–358 (2011).

223. Zernecke, A. Dendritic cells in atherosclerosis: Evidence in mice and humans. *Arterioscler. Thromb. Vasc. Biol.* **35**, 763–770 (2015).

224. Cybulsky, M. I., Cheong, C. & Robbins, C. S. Macrophages and Dendritic Cells: Partners in Atherogenesis. *Circ. Res.* **118**, 637–652 (2016).

225. Coll, B., Alonso-Villaverde, C. & Joven, J. Monocyte chemoattractant protein-1 and atherosclerosis: Is there room for an additional biomarker? *Clin. Chim. Acta* **383**, 21–29 (2007).

226. Palsson-Mcdermott, E. M. *et al.* Pyruvate kinase M2 regulates hif-1α activity and il-1β induction and is a critical determinant of the warburg effect in LPS-activated macrophages. *Cell Metab.* **21**, 65–80 (2015).

227. Pialoux, V. *et al.* Relationship between oxidative stress and HIF-1α mRNA during sustained hypoxia in humans. *Free Radic. Biol. Med.* **46**, 321–326 (2009).

228. Movafagh, S., Crook, S. & Vo, K. Regulation of hypoxia-inducible Factor-1a by reactive oxygen species: New developments in an old debate. *J. Cell. Biochem.* **116**, 696–703 (2015).

229. Gao, L., Chen, Q., Zhou, X. & Fan, L. The role of hypoxia-inducible factor 1 in

atherosclerosis. *J. Clin. Pathol.* **65**, 872–876 (2012).

230. Huang, S. C. C. *et al.* Cell-intrinsic lysosomal lipolysis is essential for alternative activation of macrophages. *Nat. Immunol.* **15**, 846–855 (2014).

231. Goldberg, I. J. & Bornfeldt, K. E. Lipids and the endothelium: bidirectional interactions. *Curr. Atheroscler. Rep.* **15**, 365 (2013).

232. VanderLaan, P. A., Reardon, C. A. & Getz, G. S. Site Specificity of Atherosclerosis: Site-Selective Responses to Atherosclerotic Modulators. *Arterioscler. Thromb. Vasc. Biol.* **24**, 12–22 (2004).

233. Chatterjee, S. & Fisher, A. B. Mechanotransduction in the endothelium: Role of membrane proteins and reactive oxygen species in sensing, transduction, and transmission of the signal with altered blood flow. *Antioxidants Redox Signal.* **20**, 899–913 (2014).

234. Huang, Y. *et al.* Integrin-YAP/TAZ-JNK cascade mediates atheroprotective effect of unidirectional shear flow. *Nature* **540**, 579–582 (2016).

235. Qiao, C. *et al.* Deep transcriptomic profiling reveals the similarity between endothelial cells cultured under static and oscillatory shear stress conditions. *Physiol. Genomics* **48**, 660–666 (2016).

236. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: Five essential questions. *Nat. Rev. Genet.* **14**, 288–295 (2013).

237. Malek, A. M. & Alper, S. L. Hemodynamic Shear Stress and Its Role in Atherosclerosis. *J. Am. Med. Assoc.* **282**, 2035–2042 (1999).

238. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).

239. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science (80-.).* **362**, (2018).

240. Ludwig, L. S. *et al.* Transcriptional States and Chromatin Accessibility

Underlying Human Erythropoiesis. *Cell Rep.* **27**, 3228-3240.e7 (2019).

241. Zhou, J., Li, Y. S. & Chien, S. Shear stress-initiated signaling and its regulation of endothelial function. *Arterioscler. Thromb. Vasc. Biol.* **34**, 2191–2198 (2014).

242. Warabi, E. *et al.* Shear stress stabilizes NF-E2-related factor 2 and induces antioxidant genes in endothelial cells: Role of reactive oxygen/nitrogen species. *Free Radic. Biol. Med.* **42**, 260–269 (2007).

243. Aird, W. C. Phenotypic heterogeneity of the endothelium: I. Structure, function, and mechanisms. *Circ. Res.* **100**, 158–173 (2007).

244. Aird, W. C. Phenotypic heterogeneity of the endothelium: II. Representative vascular beds. *Circ. Res.* **100**, 174–190 (2007).

245. Chen, B. P. C. *et al.* DNA microarray analysis of gene expression in endothelial cells in response to 24-h shear stress. *Physiol. Genomics* **7**, 55–63 (2001).

246. Zhang, J. & Friedman, M. H. Adaptive response of vascular endothelial cells to an acute increase in shear stress magnitude. *Am. J. Physiol. - Hear. Circ. Physiol.* **302**, 983–991 (2012).

247. Funk, C. D. & FitzGerald, G. A. COX-2 inhibitors and cardiovascular risk. *J. Cardiovasc. Pharmacol.* **50**, 470–479 (2007).

248. Georgiadi, A. *et al.* Overexpression of angiopoietin-like protein 4 protects against atherosclerosis development. *Arterioscler. Thromb. Vasc. Biol.* **33**, 1529–1537 (2013).

249. Ando, J. & Yamamoto, K. Flow detection and calcium signalling in vascular endothelial cells. *Cardiovasc. Res.* **99**, 260–268 (2013).

250. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–340 (2011).

251. Mega, J. L. *et al.* Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: An analysis of primary and secondary prevention trials. *Lancet* **385**, 2264–2271 (2015).

252. Tzima, E. Role of small GTPases in endothelial cytoskeletal dynamics and the shear stress response. *Circ. Res.* **98**, 176–185 (2006).

253. Perucca, P. *et al.* Loss of p21CDKN1A impairs entry to quiescence and activates a DNA damage response in normal fibroblasts induced to quiescence. *Cell Cycle* **8**, 105–114 (2009).

254. Madrigal, P. & Alasoo, K. AP-1 Takes Centre Stage in Enhancer Chromatin Dynamics. *Trends Cell Biol.* **28**, 509–511 (2018).

255. Tempé, D. *et al.* SUMOylation of the inducible (c-Fos:c-Jun)/AP-1 transcription complex occurs on target promoters to limit transcriptional activation. *Oncogene* **33**, 921–927 (2014).

256. Okamoto, Y. *et al.* Transgenic mice with cardiac-specific expression of activating transcription factor 3, a stress-inducible gene, have conduction abnormalities and contractile dysfunction. *Am. J. Pathol.* **159**, 639–650 (2001).

257. Kawauchi, J. *et al.* Transcriptional repressor activating transcription factor 3 protects human umbilical vein endothelial cells from tumor necrosis factor-α-induced apoptosis through down-regulation of p53 transcription. *J. Biol. Chem.* **277**, 39025–39034 (2002).

258. Gold, E. S. *et al.* ATF3 protects against atherosclerosis by suppressing 25-hydroxycholesterol – induced lipid body formation. *J. Exp. Med.* **209**, 807–817 (2012).

259. Zhao, J., Li, X., Guo, M., Yu, J. & Yan, C. The common stress responsive transcription factor ATF3 binds genomic sites enriched with p300 and H3K27ac for transcriptional regulation. *BMC Genomics* **17**, 1–14 (2016).

260. Saldanha, A. *et al.* Impact of Diabetes on Cardiovascular Disease : An Update. *Int. J. Hypertens.* **653789**, 1–15 (2013).

261. Johnson, L. A. *et al.* Apolipoprotein E4 exaggerates diabetic dyslipidemia and atherosclerosis in mice lacking the LDL receptor. *Diabetes* **60**, 2285–2294 (2011).

262. Rask-Madsen, C. *et al.* Loss of insulin signaling in vascular endothelial cells accelerates atherosclerosis in apolipoprotein e null mice. *Cell Metab.* **11**, 379–389 (2010).

263. Fukuhara-Takaki, K., Sakai, M., Sakamoto, Y. I., Takeya, M. & Horiuchi, S. Expression of class A scavenger receptor is enhanced by high glucose in Vitro and under diabetic conditions in Vivo: One mechanism for an increased rate of atherosclerosis in diabetes. *J. Biol. Chem.* **280**, 3355–3364 (2005).

264. Griffin, E. *et al.* A link between diabetes and atherosclerosis: Glucose regulates expression of CD36 at the level of translation. *Nat. Med.* **7**, 840–846 (2001).

265. Pavlou, S., Lindsay, J., Ingram, R., Xu, H. & Chen, M. Sustained high glucose exposure sensitizes macrophage responses to cytokine stimuli but reduces their phagocytic activity. *BMC Immunol.* **19**, 1–13 (2018).

266. Wen, Y. *et al.* Elevated glucose and diabetes promote interleukin-12 cytokine gene expression in mouse macrophages. *Endocrinology* **147**, 2518–2525 (2006).

267. Steinberg, D. & Witztum, J. L. History of discovery: Oxidized low-density lipoprotein and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2311–2316 (2010).

268. Glass, C. K. & Witztum, J. L. Atherosclerosis: The Road Ahead. *Cell* **104**, 503–516 (2001).

269. Gu, L. *et al.* Absence of monocyte chemoattractant protein-1 reduces atherosclerosis in low density lipoprotein receptor-deficient mice. *Mol. Cell* **2**, 275–281 (1998).

270. Kirii, H. *et al.* Lack of interleukin-1ß decreases the severity of atherosclerosis in apoE-deficient mice. *Arterioscler. Thromb. Vasc. Biol.* **23**, 656–660 (2003).

271. Schulthess, J. *et al.* The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages. *Immunity* **50**, 432-445.e7 (2019).

272. Lin, J. Da *et al.* Single-cell analysis of fate-mapped macrophages reveals heterogeneity, including stem-like properties, during atherosclerosis progression

and regression. *JCI insight* **4**, (2019).

273. Denko, N. C. Hypoxia, HIF and metabolism in the solid tumour. *Nat. Rev. Cancer* **8**, 705–713 (2008).

274. Corcoran, S. E. & Neill, L. A. J. O. HIF1a and metabolic reprogramming in inflammation. *J Clin Invest* **126**, 3699–3707 (2016).

275. Funk, S. D., Yurdagul, A. & Orr, A. W. Hyperglycemia and endothelial dysfunction in atherosclerosis: Lessons from type 1 diabetes. *Int. J. Vasc. Med.* **2012**, (2012).

276. Valente, A. J., Irimpen, A. M., Siebenlist, U. & Chandrasekar, B. OxLDL induces endothelial dysfunction and death via TRAF3IP2: Inhibition by HDL3 and AMPK activators. *Free Radic. Biol. Med.* **70**, 117–128 (2014).

277. Teasdale, J. E. *et al.* Cigarette smoke extract profoundly suppresses TNFα-mediated proinflammatory gene expression through upregulation of ATF3 in human coronary artery endothelial cells. *Sci. Rep.* **7**, 1–10 (2017).