Unmasking disease risk hidden in haematopoietic enhancers

Juliana Estefanía Arcila Galvis

Biosciences Institute
Newcastle University

Doctor of Philosophy
April 2024

Abstract

Histones are subject to a wide range of modifications, which are differentially regulated across cell types to modulate gene expression. The simultaneous presence of different sets of histone modifications in a genomic location is correlated with the function of the DNA wrapped around such histones. This information is especially relevant to understanding the role of non-coding regions of the genome, such as enhancers, which modulate the expression of target genes in specific conditions and a cell type-specific manner. Questions have been raised about the potential role of enhancers in the modulation of genes involved in complex immune-mediated diseases because most genetic variants associated with such diseases lie in non-coding regions. However, the exact localisation of enhancers in haematopoietic cell types is unclear. Different groups of researchers have made inconsistent efforts to annotate the epigenome of focused subgroups of cell types, making the comparison among genomic locations of enhancers across multiple haematopoietic cell types unfeasible. I produced the most complete collection of epigenomic annotations of haematopoietic cell types based on the presence of six different histone modifications across the epigenome of 107 samples from 31 different cell types of healthy individuals. I also identified around 200 diseases with associated non-coding genetic variants colocalising regions annotated as enhancers in different subsets of haematopoietic cell types, thus providing an invaluable resource for the interpretation of non-coding variants associated with complex diseases. I used Inflammatory Bowel disease and cardiovascular disease as examples to showcase our dataset's leverage for identifying complex diseases causing genes.

Acknowledgements

To all those who have contributed to my journey, your presence, love, and support have made this academic pursuit an enriching and transformative experience. I am humbled by the kindness and encouragement I have received, and I wholeheartedly acknowledge the role that each of you has played in shaping this chapter of my life.

I am deeply grateful to Daniel and Robert for believing in my potential and for giving me the incredible opportunity to pursue my PhD in the UK despite the challenges in securing funding due to my non-UK, non-European status. Your invaluable support and guidance have been instrumental in enabling me to progress in my research and continue my studies, even in the face of personal adversities. Your understanding, unconditional encouragement, and diligence have been a driving force throughout my academic journey.

Daniel, huge thanks for the extensive amount of time and effort you dedicated to supporting my training and research. Our brainstorming sessions on this project and the myriad of other side projects we embarked upon have been truly enjoyable and enlightening. Your approachability, wisdom, enthusiasm, and positivity have been a constant source of inspiration. I deeply admire your brilliance and boundless creativity, and I consider myself incredibly fortunate to have had you as my main supervisor. Your humanity and kindness have made this academic journey all the more enriching.

Robert, without your help, finishing this PhD on time would not have been possible. There were numerous moments when I felt lost in a sea of data and thoughts. Thanks to your broad knowledge across various fields and your exceptional ability to make connections between them, you helped me identify the true relevance of my data, structure it and synthesise complex ideas.

I also want to thank my co-supervisors, Professor Sophie Hambleton and Dr Chris Lamb, and my panel review for their scientific guidance.

To the wonderful people who belong to the Computational Epigenomics group, I express my gratitude for creating a supportive and collaborative work environment. Special thanks to Marco, Brigid, Junhua, and Athira for making my time at the office so enjoyable, and to Jenny for keeping me sane during the lockdowns, for making me fall in love with this city, for being my first friend in this country, for helping me proofread this dissertation and for continuing to be such an amazing friend during all these years.

To my husband Dmitrij Didenko, thanks for your love, for bringing happiness to my life and for your constant presence and support. From preparing meals during my busiest times to taking care of me when I fell ill, your care and understanding during my more stressful and sad times have made my life easier. Our companionship over these years has meant the world to me.

Thanks to my Dad, Julian Arcila, your unwavering emotional support and genuine interest in every aspect of my life, from personal to scientific, have been invaluable. Thanks for guiding me through important decisions and for encouraging me to look for opportunities in Europe and taking risks.

Lastly, I am indebted to my mom, Alicia Galvis, and my sister, Marcelita, for their love and guidance during my whole life and for working extra hard to help me accomplish my dreams. Mom, your daily words of love and kindness, even from afar, have cheered me up and given me strength. You have been my rock throughout these years. Your love has been a guiding light, inspiring me to reach for the stars.

Publications

Davidson, B.S.[*], **Arcila-Galvis, J.E.**[*], Trevisan-Herraz, M., Mikulasova, A., Brackley, C.A., Russell, L.J. & Rico, D. (2023) "The discovery of an evolutionarily conserved enhancer within the MYEOV locus suggests an unexpected role for this non-coding region in cancer," *bioRxiv*, p. 2023.09.18.558245.

* Equal contributions.

Table of contents

Abstract i
Acknowledgements ii
Publications iv
Table of contents v
List of figures vii
List of tables viii
Abbreviations ix
Chapter 1. Introduction 1
1.1. Genotype, phenotype and epigenetic mechanisms. 1
1.2. Histone post-translational modifications 3
1.3. The enhancers 9
1.4. Enhancers and disease. 11
1.4.1. Genetic factors underlying disease expression. 11
1.4.2. Disease risk and genetic variants at enhancers. 12
1.4.3. Complex disease risk: SNPs in enhancers and other potential genetic contributors. 15
1.5. Enhancer maps and enhancer-target gene links 20
1.6. Haematopoietic cells 25
1.7. Monocyte lineage. 27
1.7.1. Classical M1/M2 macrophage in vitro model. 29
1.8. Aims of my project. 30
1.8.1. General aim. 30
1.8.2. Specific aims. 31
Chapter 2. Materials and methods 33
2.1. Sources for ChIP-Seq data acquisition. 33
2.2. ChIP-Seq data and alignments. 34
2.3. Learning the haematopoietic chromatin state maps. 37
2.3.1. Authors contribution. 37
2.3.2. Building the chromatin-states model. 38
2.3.3. State labels, interpretation, and mnemonics. 40
2.4. Cell type consensus epigenomes. 41
2.5. Enhancer activity profiles. 42

2.6. GWAS enrichment analysis. ... 43
2.7. Identifying target genes of enhancers associated with GWAS traits. ... 45
2.8. Software used for data analysis. ... 46
Chapter 3. A resource of reference epigenomes of human haematopoietic cells ... 48
3.1. Introduction. ... 48
3.2. Results. ... 50
3.2.1. The largest dataset of chromatin state maps for individual primary haematopoietic samples. 50
3.2.2. A collection of consensus epigenomes for haematopoietic cell types. ... 52
3.2.3. Enhancer maps for haematopoietic cells ... 55
3.2.4. Enhancer activity profiles across haematopoietic cell types. ... 57
3.3. Discussion. ... 60
Chapter 4. Linking haematopoietic enhancers to complex disease genetic risk. ... 63
4.1. Introduction. ... 63
4.2. Results. ... 64
4.2.1. Prioritising high-impact enhancers by focusing on the most common haematopoietic activity profiles. ... 64
4.2.2. GWAS traits are linked to haematopoietic enhancers with specific activity profiles. 65
4.2.3. Activity profiles of enhancers match with relevant cell types in complex traits aetiology. 67
4.2.5. Ulcerative colitis-associated enhancers potentially target key players in NETosis. 72
4.3. Discussion. ... 75
Chapter 5. Macrophage enhancers are associated with CVD risk by regulating the expression of key genes in lipid metabolism. ... 79
5.1. Introduction. ... 79
5.1.1. Atherosclerosis. ... 79
5.1.2. Dyslipidaemia. ... 80
5.1.3. Macrophages' role in atherosclerosis. ... 82
5.1.4. Preliminary results link macrophage enhancers to CVD risk. ... 82
5.2. Results. ... 83
5.2.1. Exploring enhancer activity in cardiovascular disease risk. ... 83
5.2.2 Using eQTL data to identify the target genes of CVD-associated enhancers. ... 86
5.2.3 Expression patterns of enhancer's target genes. ... 88
5.3. Discussion ... 90

Chapter 6. Project contributions and future perspectives...................................... 98
6.1. A synopsis of project outcomes and scientific contributions. 98
6.2. Future perspectives. 99
6.2.1. Finding gene targets for all the identified disease-associated enhancers. 99
6.2.2. Enriching the dataset with other relevant cell types. 100
6.2.3. Experimental validation of genetic susceptibility...102
6.2.4. Considering the influence of genetic-environmental interactions.........................104
6.2.4.1. Identification of enhancers that may contribute to disease susceptibility only upon exposure to environmental triggers. 105
6.2.4.2. Using epidemiological studies to identify the Environmental factors influencing disease development. 106
6.3. Conclusion. 107
Appendix... 109
References... 126

List of figures

Figure 2. 1. Dataset description. 36
Figure 2. 2. Correlation between the emission probabilities of our model (12 states) with those of corresponding states from our previously published models. 39
Figure 2. 3. Consensus epigenomes and enhancer groups by activity profiles. 42
Figure 2. 4. GWAS enrichment analysis..45
Figure 3. 1. Epigenomic profile of haematopoietic cell types from the BLUEPRINT Consortium 51
Figure 3. 2. Using chromatin states to study the epigenomic landscape of *NR1H3*.................54
Figure 3. 3. Number and specificity of consensus AE per cell type.......................................56
Figure 3. 4. Enhancer Activity profiles in haematopoietic cells. 59
Figure 4. 1. The 100 activity profiles most frequently observed in hematopoietic AE. 67
Figure 4. 2. Interplay between genetic susceptibility, enhancer dynamics, and cellular contexts in complex diseases. 67
Figure 4. 3. Example of cell types associated with three groups of complex diseases. 70
Figure 4. 4. Exploring Enhancer Enrichment in IBD-associated loci. 71
Figure 4. 5. The epigenomic landscape of UC-associated genomic region enriched with enhancers. 74

Figure 5. 1. Macrophages' role in atherosclerosis. 83
Figure 5. 2. Enhancer activity profiles associated with CVD risk. 84
Figure 5. 3. Example of the *LPL* region enriched in macrophage enhancers and eQTLs. 87
Figure 5. 4.Expression patterns in haematopoietic cells of the potential gene targets of macrophage enhancers associated with CVD risk. 89
Figure 5. 5. Mapping macrophage enhancers target genes to specific steps in lipid metabolism. 92
Figure A. 1. Epigenomic landscape of the *ACP2*, *NR1H3* and *MADD* region from EpiMap.120
Figure A. 2. Using chromatin states to study the epigenomic landscape of *ARID5B*. 121
Figure A. 3. Visualization of the GTEX eQTL data for risk variants in UC-associated genomic region enriched with myeloid enhancers. 122
Figure A. 4. Epigenomic landscape from EpiMap of the UC-associated genomic regions enriched with neutrophil and myeloid enhancers. 125

List of tables

Table 2. 1.Best match (correlation coefficient > 0.75) between each state in our 12 states-model states from Carrillo et al. (2017) and Roadmap (2015) models. 40
Table 4. 1. Summary of results from the enrichment analysis. 68
Table 5. 1.Genomic regions associated with CVD Risk and enriched in macrophage enhancers. 85
Table 5. 2. GWAS traits associated with non-coding SNPs in a region enriched with macrophage enhancers and overlapping *LPL* gene. 88
Table A. 1. Sample metadata. 109
Table A. 2. Correlation of emission probabilities between Carrillo et al. (2027) model and ours. 114
Table A. 3. Correlation of emission probabilities between Roadmap (2015) model and ours. 115
Table A. 4. Expression patterns of potential enhancer gene targets enriched in CVD-associated regions according to the Human Protein Atlas database. 117

Abbreviations

3D: three-dimensional

ABCA1: ATP-binding cassette transporter A1

AconE: conventional active enhancer

ACT-seq: chromatin tagmentation sequencing

AE: active enhancers

AleloE: elongating active enhancer

AM: airway macrophages

Apo: apolipoprotein

ATRA: all-trans retinoic acid

CAD: coronary artery disease

CAGE: cap analysis of gene expression

ChIC-seq: chromatin immune cleavage sequencing

ChIL-seq: chromatin integration labelling followed by sequencing

ChIP: chromatin immunoprecipitation

ChIP-seq: chromatin immunoprecipitation followed by DNA sequencing

CNV: copy number variant

CR: chylomicron remnants

CUT & Tag: cleavage under targets and tagmentation

CUT & RUN: cleavage under targets & release using nuclease

CVD: cardiovascular disease

DAC: Data Analysis Centre of the BLUEPRINT Consortium

DNase-seq: DNase I hypersensitive site sequencing

ECs: endothelial cells

eGenes: genes with expression levels correlated with the eQTLs

eQTLs: expression quantitative trait loci

EpiRR: Epigenome Reference Registry

eryth: erythroblasts

EWAS: Epigenome-Wide Association Studies

FH: familial hypercholesterolemia

GWAS: Genome-Wide Association Studies

H3: histone 3

haQTLs: histone acetylation quantitative trait loci

HAWAS: Histone Acetylation Genome-Wide studies

HDL: high-density lipoprotein

Hi-C: high-throughput chromosome conformation capture

HMM: Hidden Markov Model

HSC: haematopoietic stem cells

HSM: haplotype-specific DNA methylation

IHEC: The International Human Epigenome Consortium

IPF: idiopathic pulmonary fibrosis

LD: linkage disequilibrium

LDL: low-density lipoprotein

LPA: lysophosphatidic acid

LPL: lipoprotein lipase

LPP3: Lipid Phosphate Phosphatase

LTRs: Long Terminal Repeats

LXRα: liver X receptor alpha

meos: mature eosinophils

Mnase: A-Micrococcal Nuclease

mRNA-Seq: messenger RNA sequencing

mTOR: mechanistic target of rapamycin

NETs: neutrophil extracellular traps

NIH: The National Institutes of Health

NK: natural killer cells

Num: number of

OxLDL: oxidised LDL

P: promoter regions

PLTP: Phospholipid Transfer Protein

POE: Parent-of-origin effect

RALDH2: retinaldehyde dehydrogenase 2

RCT reverse cholesterol transport pathway

ROS: reactive oxygen species

RXR: Retinoid X Receptor

SE: super-enhancer

SNP: single nucleotide polymorphism

SNVs: single nucleotide variants

SRBI: scavenger receptor class B type I

StAR: Steroidogenic Acute Regulatory protein

SVs: structural variants

STRs: short tandem repeats

TES: transcription end site

TFs: transcription factors

TSS: transcription start site

UC: ulcerative colitis

V2G2P: variant-to-gene-to-program

VLDL: very low-density lipoprotein

VSMC: vascular smooth muscle cells

Chapter 1. Introduction

1.1. Genotype, phenotype and epigenetic mechanisms.

All the potential traits that an organism may display throughout its life are encoded as a precise sequence of nucleotide bases in its DNA—The genome, which is inherited from the ancestors of the organism (Makałowski, 2001; Goldman & Landweber, 2016). Inside the nucleus, the genome is packaged in compact and dense structures —the chromosomes. These are composed of a mixture of DNA, RNA, and proteins, together forming the structure known as chromatin (Gilbert et al., 2004).

The chromatin is organised in a dynamic and hierarchical three-dimensional (3D) structure (Lanctôt et al., 2007), which defines the activity of different elements of the genome by controlling their accessibility. Very compacted regions of chromatin (heterochromatin) render genes inaccessible, effectively silencing their expression. Conversely, in open chromatin (euchromatin), genes become readily accessible for transcription, enabling their expression swiftly in response to environmental cues (Grandi et al., 2022; Klemm et al., 2019a) **(Fig. 1.1.).**

Multicellular organisms are composed of numerous types of cells that rise during development when pluripotent cells differentiate into distinct lineages, eventually forming various tissues and organs (Arendt et al., 2016; A. G. Fisher, 2002). These specialised cell types present different observable characteristics (phenotype), such as shape, size, and specific functions, and play several important roles that are crucial to the organism's fitness. Despite phenotypic differences, all cells within the same organism share the same genetic information (genotype), and their diversity is due to variations in gene expression patterns among them, which result from the selective activation or repression of specific sets of genes (Briggs et al., 2018; Cao et al., 2019).

In a given organism with a specific genotype, cell type- and context-specific gene expression is regulated by epigenetic mechanisms (Carter & Zhao, 2020). These include, among others, chemical modifications made to the chromatin (e.g., histone modifications) or the genome (e.g., DNA methylation), chromatin conformational and compositional changes (e.g. looping and nucleosome enrichment/depletion), and binding of transcription factors (TFs) to non-coding regulatory DNA sequences (e.g., promoters and enhancers) (Carter & Zhao, 2020; Schwartzman & Tanay, 2015; Fazzari & Greally, 2004). Epigenetic mechanisms are influenced

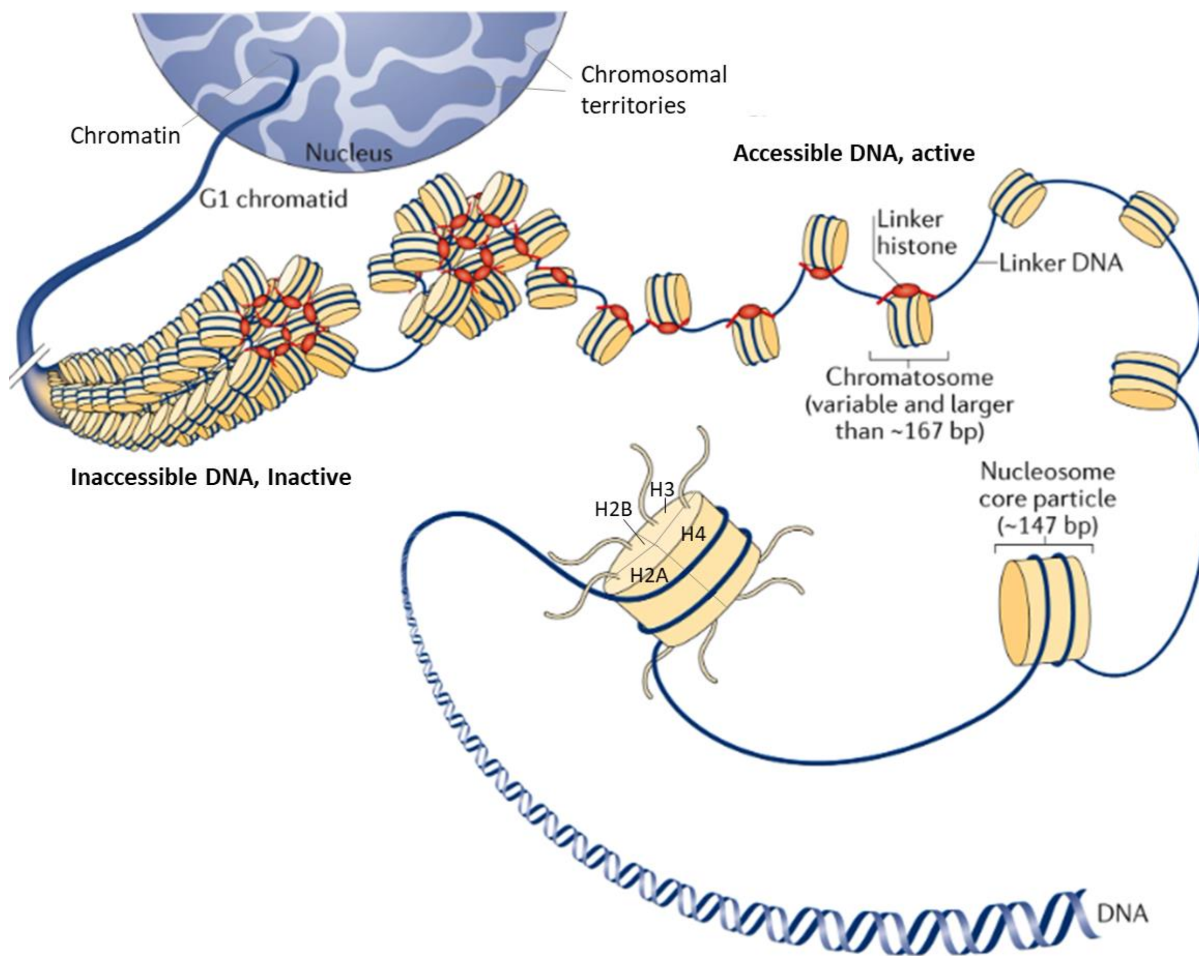


Figure 1. 1. Spatial organisation of the genome within the nucleus. The genome is distributed within chromosomes, each occupying delimited spaces within the nucleus known as chromosomal territories. Within chromosomes, DNA is wrapped around nucleosomes. Each nucleosome contains two copies of four core histones: H2A, H2B, H3, and H4. An additional linker histone (H1) can attach to the nucleosome, forming a chromatosome. These linker histones can form dimers, facilitating the formation of clutches, which are groups of chromatosomes. Together, this complex of DNA and proteins is referred to as chromatin. Chromatin exhibits a three-dimensional architecture characterised by chromatin loops. Condensed regions are characterised by transcriptional inactivity, while nucleosome-depleted areas allow accessibility for transcriptional processes. (Fyodorov et al., 2018).

by environmental cues such as nutrition, the composition of the microbiota, exposure to environmental toxicants, temperature changes, infections and other stresses (Wu et al., 2023; Rothschild et al., 2018; Gibson, 2008; Sanna et al., 2019) **(Fig. 1.2).**

The term “epigenome” refers to the complete set of epigenetic modifications across the genome of a cell type (Holtzman & Gersbach, 2018; Fazzari & Greally, 2004). Epigenomics is a field of study that seeks to define the location and nature of the genomic sequences that are epigenetically modified and their potential impact on gene regulation and cellular function (Schwartzman & Tanay, 2015; Stricker et al., 2016).

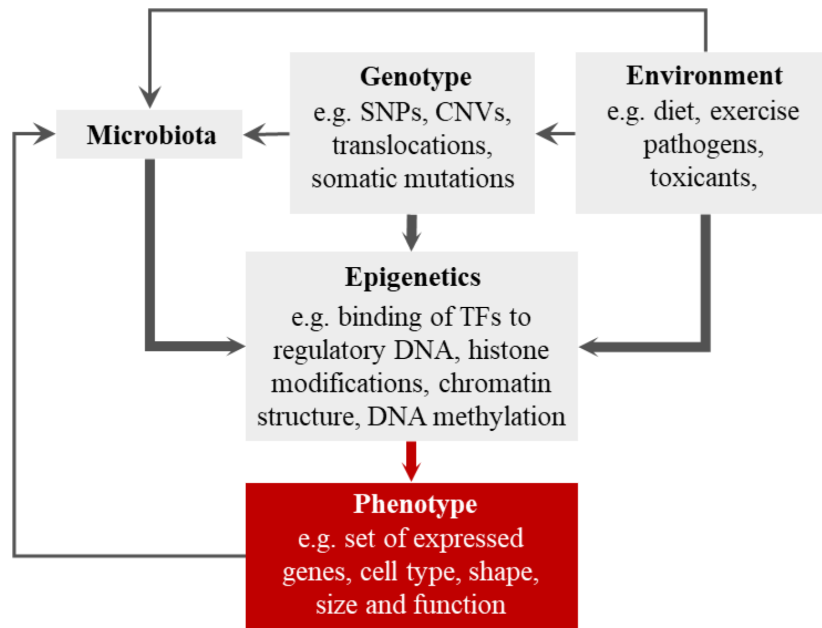


Figure 1. 2. Epigenetic mechanisms regulate the phenotype. The phenotype of a cell is influenced by the interplay between its genotype and exposure to external cues, such as toxicants, microorganisms (pathogens and microbiota), diet, etc. Epigenetic mechanisms mediate this influence by regulating gene expression required for an appropriate response to these exposures. The environment can also modify the genotype and the microbiota, for instance, through toxicants causing DNA damage. Additionally, diet and genotype influence the microbiota's composition: Abbreviations: Single Nucleotide Polymorphism (SNP), Copy Number Variant (CNV).

1.2. Histone post-translational modifications

The first level of DNA package and recurring structural unit of the chromatin is the nucleosome, which consists of 147 bp of DNA wrapped around a histone octamer (two copies each of the core histones H2A, H2B, H3, and H4) (Fyodorov et al., 2018; Gilbert et al., 2004). The next structural unit of DNA packaging is the chromatosome, which is formed by the binding of the linker histone (H1) to a nucleosome. By forming dimers, H1 facilitates the formation of higher-order structures, such as tetra nucleosome units, which are arranged in groups of heterogeneous sizes called “clutches”. Increased levels of H1 and denser clutches are found in regions with higher chromatin compaction. Therefore, H1 is thought to be an important modulator of heterochromatin dynamics and chromatin silencing. (Fyodorov et al., 2018; Zhou et al., 2015; Hizume et al., 2005).

The N-terminal amino acid tails of histone proteins protruding from nucleosomes and chromatosomes can be modified by the covalent binding of chemical groups such as methylation, acetylation, phosphorylation, deamination/citrullination and ubiquitination. These modifications are called histone marks. (Rando, 2012; Gates et al., 2017) **(Fig. 1.3.A)**. These posttranslational modifications are reversible, being added by enzymes called “writers” (e.g. MLL/set1 family of histone lysine methyltransferases, CBP/300 histone acetyltransferase) or removed by other enzymes called “erasers” (e.g. H3K4 demethylase LSD1) (Hyun et al., 2017; Kouzarides, 2007).

The localisation and number of chemical groups added to histone tails influence the likelihood of the DNA wrapped around them being involved in different cellular processes (Kouzarides, 2007). This influence can be exerted by histone modifications affecting DNA accessibility (Bannister & Kouzarides, 2011; Tropberger & Schneider, 2013) or by proteins known as ‘readers’, which recognise and bind particular histone modifications and recruit different proteins involved in specific biological processes (Tafessu & Banaszynski, 2020a; Hyun et al., 2017).

For instance, some histone marks influence transcriptional activity. Histone acetylation can influence the opening of the chromatin, making genes and non-coding regulatory DNA sequences more accessible for TFs and RNA polymerases (Bauer et al., 1994; Rando, 2012). Additionally, acetylated lysines are recognised by bromodomains present in a diverse array of transcriptional coactivators, such as CBP/300, which act as scaffolding to recruit additional transcriptional machinery (Tafessu & Banaszynski, 2020a). Moreover, methylation of the 27th lysine of histone 3 (H3K27me3) is implicated in the formation of facultative heterochromatin, which preserves cell-type identity by silencing alternative lineage genes and plays a role in dosage compensation between the sexes by inactivating one of the X chromosomes in female mammals (Allshire & Madhani, 2018; Wutz, 2011)

Histone marks can also influence the likelihood of DNA sequences participating in processes such as repair, recombination, and cell-cycle control. Methylation in the lysine 9 of histone 3 (H3K9me3) is implicated in the formation of constitutive heterochromatin (Boros et al., 2014a), mainly in telomeres and pericentromeric regions with repetitive sequences like satellite repeats and transposable elements (Allshire & Madhani, 2018; Nair et al., 2017). By suppressing these sequences' activity, constitutive heterochromatin decreases the chance of these regions participating in recombination or chromosomal rearrangements, hence maintaining genome stability. Additionally, constitutive heterochromatin at centromeric regions is important for the proper chromosome segregation during cell division by facilitating kinetochore assembly

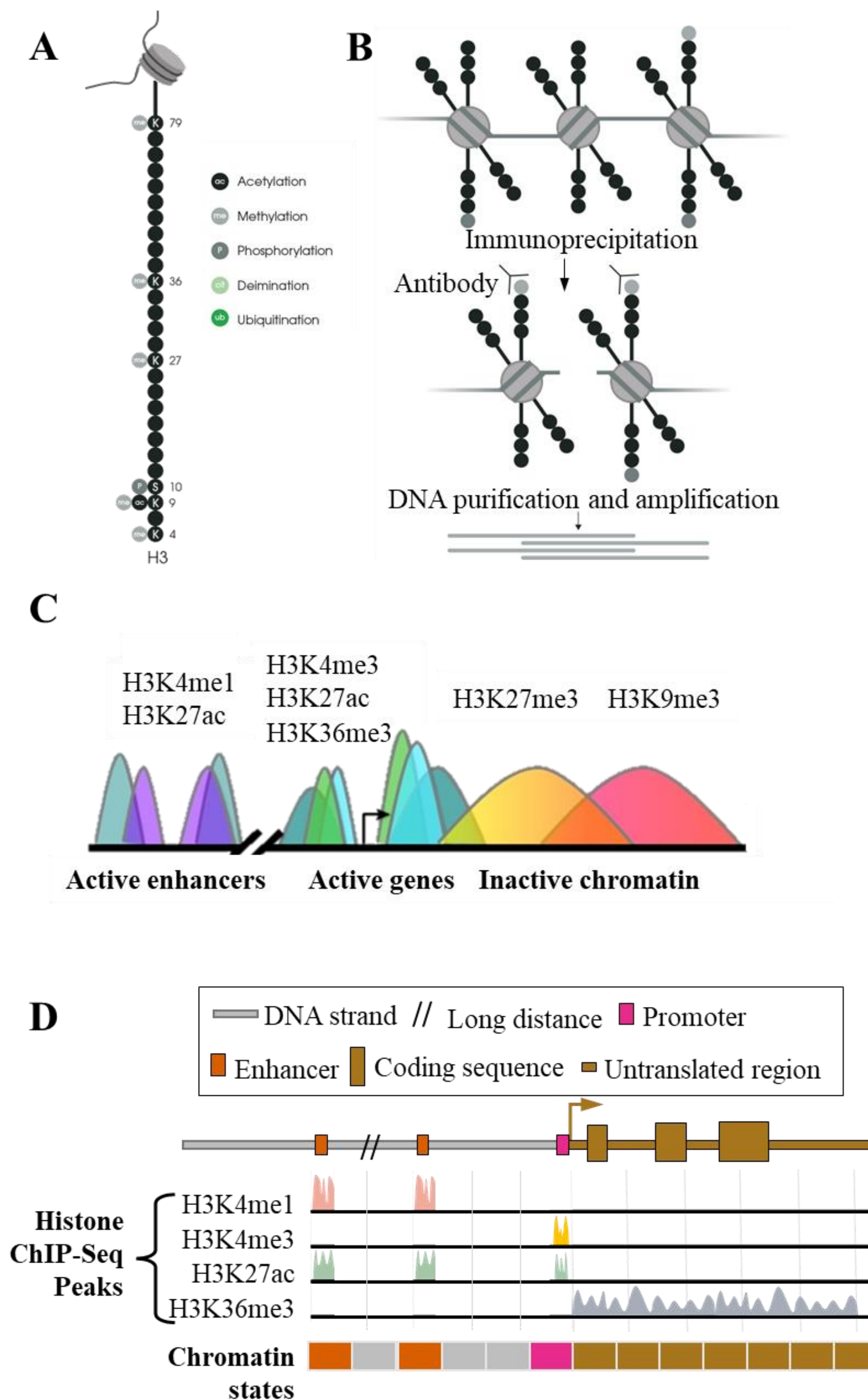


Figure 1. 3. Antibody-based identification of histone modifications. Illustration of a nucleosome showing positions of various histone modifications in the aminoacidic tail of the histone 3 (H3). **(B)** Diagram showing how the ChIP-seq assays work. Specific antibodies are used to identify histone modifications of interest (depicted in grey). Subsequently, the genome is fragmented, the targeted regions are isolated, and DNA is purified and sequenced. **(C)** Simplified illustration of the occupancy of histone modifications as determined by ChIP-seq. Coloured regions represent a typical ChIP-Seq signal for specific histone modifications at a given region: teal (H3K27ac), purple (H3K4me1), light blue (H3K36me3), green (H3K4me3), yellow (H3K27me3), red (H3K9me3). **(D)** At the top of the image is a schematic representation of a genome section, delineating the positions of various genomic elements. Directly beneath

are signals depicting the occupancy of four distinct histone modifications associated with these genomic elements. Further below, the figure presents a graphical representation of chromatin states. These states serve as annotations for genetic elements, organised in 200-base pair bins across the entire genome. The chromatin states are defined based on the combinatorial presence of the specified histone marks, providing a comprehensive view of the functional annotation of the genome. Figures A and B are adapted from (Abcam plc., 2023), and Figure C is adapted from (Gates et al., 2017).

(Allshire & Madhani, 2018; Ishii et al., 2008). This type of heterochromatin remains condensed and inactive throughout the cell cycle and across different cell types, in contrast with facultative heterochromatin, which can undergo reversible changes and become active in response to cellular signals and developmental cues (Allshire & Madhani, 2018; Trojer & Reinberg, 2007).

The position of histone modifications across the genome of individual cell types can be determined by using different methodologies, each with its own strengths and limitations. Chromatin immunoprecipitation (ChIP) followed by DNA sequencing (ChIP-Seq) (Robertson et al., 2007; Johnson et al., 2007) is the standard and most used method for chromatin profiling of histone marks, and it has been employed for profiling the epigenome of diverse samples as part of the ROADMAP (Roadmap Epigenomics Consortium et al., 2015), Encyclopedia of DNA Elements (ENCODE) (The ENCODE Project Consortium et al., 2012) and BLUEPRINT (Adams et al., 2012) epigenome consortia.

In the ChIP-Seq protocols, cells are first cross-linked or “fixed” using formaldehyde to stabilise the association of chromatin-bound proteins. The cells are lysed, and the whole free chromatin is fragmented into pieces. Target-specific antibodies are employed, and this binds preferentially to the fragments with the target histone mark. The antibody-bound chromatin is isolated from bulk chromatin using a magnet, and this is followed by a series of stringent washes with buffers that strip away off-target proteins. Immunoprecipitated DNA is then purified, sequenced, and mapped to the reference genome to determine the target histone mark’s genomic location (Landt et al., 2012) **(Fig. 1.3.B)**.

Specific steps in the ChIP-Seq protocol present inherent limitations that can result in low signals (low enrichment of target sequence reads) and high backgrounds (many reads of off-target sequences) (Landt et al., 2012; Park, 2009). For instance, the cross-linking of proteins to DNA can lead to epitope masking, where the epitope recognised by the antibody used for immunoprecipitation is inaccessible due to the cross-linked proteins. This can result in underrepresentation or failure to detect certain binding events (Landt et al., 2012; Park, 2009).

Fragmentation requires thorough optimisation, with a desired size range from 150-300 bp because long chromatin pieces (>600-700 bp) make it difficult to identify exactly where the histone mark is located, thus lowering resolution. In the immunoprecipitation step magnetic beads may pull down off-target fragments along with the chromatin containing the target (Landt et al., 2012; Park, 2009). Besides, immunoprecipitation also requires highly stringent washes, which can result in sample loss. Due to all the above, large numbers of cells (a minimum of 500,000 cells; however, in practice, most researchers use millions of cells per ChIP) and very deep sequencing (~20 million or more sequencing reads per reaction) are required to detect signal enrichment over background (Landt et al., 2012; Park, 2009)

More recent techniques have been developed to overcome some of these limitations, such as the Cleavage Under Targets & Release Using Nuclease (CUT & RUN) technique (Skene & Henikoff, 2017). In the CUT & RUN protocol, unfixed permeabilised cells are incubated with the target-specific antibody, then a protein A-Micrococcal Nuclease (MNase) fusion protein is bound via antibodies and selectively cleaves DNA near regions bound by the target protein, facilitating subsequent DNA release and purification for library preparation and sequencing (Skene & Henikoff, 2017). This method reduces the time and cost of the experiments by eliminating several steps from the protocol, such as cross-linking, cell lysis, fragmentation, and immunoprecipitation. Additionally, since the chromatin fragments released from the cell and available for sequencing are only those near the antibodies, the background levels are reduced, the resolution is significantly improved, and the starting number of cells for the experiment and the necessary sequencing depth to identify the binding events is reduced significantly.

Although CUT&RUN can produce high-quality data from as few as 100–1000 cells, it needs an additional step for adapter ligation prior to library preparation, thus increasing the time and effort required for the overall procedure. Additionally, the release of MNase-cleaved fragments into the supernatant complicates the adaptation of CUT&RUN to single-cell applications. A more recently developed technique, Cleavage Under Targets and Tagmentation (CUT&Tag) addresses these limitations (Kaya-Okur et al., 2019). In this method, an A-Tn5 transposase fusion protein, tethered to target chromatin proteins by antibodies, is employed. Instead of cleaving DNA, the Tn5 transposase inserts sequencing adapters directly into DNA at the sites bound by the target protein, eliminating the need for subsequent DNA release. With a starting point as low as 60 live cells, this technique generates amplified sequence-ready libraries within a day, thereby reducing experimental time and costs (Kaya-Okur et al., 2019).

In addition to the above-mentioned techniques, others have been developed, including chromatin integration labelling followed by sequencing (ChIL-seq) (Harada et al., 2019),

chromatin immunocleavage sequencing (ChIC-seq) (Ku et al., 2019), antibody-guided chromatin tagmentation sequencing (ACT-seq) (Carter et al., 2019) and *in situ* ChIP (Q. Wang et al., 2019). Each method has its merits and limitations, and the choice often depends on the specific research question, desired resolution, and available resources. However, since these methods are relatively new and the computational processing of their outputs differs from the used in ChIP-seq, comparing the results from these new methodologies to those from ChIP-seq experiments, which account for the vast majority of available epigenomic data, is challenging (Hu et al., 2023).

The International Human Epigenome Consortium (IHEC) (Stunnenberg et al., 2016a) describes the epigenome of a sample in practical and experimentally feasible terms as the collection of its methylome obtained by Whole Genome Bisulfite Sequencing (WGBS), transcriptome by RNA sequencing (RNA-Seq) or messenger RNA sequencing (mRNA-Seq), and the ChIP-Seq profiling of 6 histone marks **(Fig. 1.3.C)**, whose associations with different functional elements of the genome was defined previously by the Roadmap Epigenomics Consortium (2015) as:

- **H3K4me3**: trimethylation of lysine 4 on histone H3, associated with promoter regions.
- **H3K4me1**: H3 lysine 4 monomethylation, associated with enhancer regions.
- **H3K36me3**: H3 lysine 36 trimethylation, associated with transcribed regions.
- **H3K27ac**: H3 lysine 27 acetylation, associated with active enhancer and promoter regions.
- **H3K9me3**: H3 lysine 9 trimethylation, a mark of constitutive heterochromatin, associated with compact and transcriptionally silent regions. This mark guides the anchorage of HP1 proteins, which are responsible for the packaging of the chromatin in this condensed state.
- **H3K27me3**: H3 lysine 27 trimethylation, associated with repression of genes associated with alternative cell lineages by Polycomb repressor complex (facultative heterochromatin), can also be associated with constitutive heterochromatin since it contributes to HP1α stability through PRC2-mediated regulation (Boros et al., 2014b).

ChIP-Seq information can be summarised by building chromatin state models, which are statistical models that help us classify regions as functional elements of the genome based on the probability of the combinatorial patterns of histone modifications at each fixed window size across the genome (Ernst & Kellis, 2015). Using the combinatorial presence of these six reference histone marks, it is possible to infer the chromatin states describing the main functional elements of the genome, such as promoters, enhancers, transcribed regions, polycomb repressed regions, and heterochromatin **(Fig. 1.3.D)**.

1.3. The enhancers

The level and timing of expression of a gene in a cell depends on its function. Housekeeping genes that are essential for metabolism and maintenance of other fundamental processes in all cells are ubiquitously expressed in all cells and conditions (Eisenberg & Levanon, 2013; Zhu et al., 2008). The expression of these genes is regulated by interactions between the promoters of gene sets involved in the same biological processes. These genes often cluster in proximal regions of the genome, and this spatial proximity promotes a 3D spatial conformation that facilitates interactions between promoters (Dejosez et al., 2023). Genes that define cell type identity or that are important for an adequate response to environmental cues (external stimuli) are regulated by several epigenetic mechanisms (Jaenisch & Bird, 2003), and there are several actors involved in this regulation (Spitz & Furlong, 2012), but here I will be focusing on the enhancers.

Transcriptional enhancers are non-coding sequences of DNA which regulate which genes are expressed in a cell, the timing of their expression, and their expression levels in response to a variety of intrinsic and external signals (Karnuta & Scacheri, 2018). Approximately 1 million enhancer elements with gene regulatory potential have been identified in mammalian genomes (The ENCODE Project Consortium et al., 2012). They can overlap with gene open reading frames but are typically located in non-coding regions, particularly enriched in intronic areas (Rigau et al., 2019).

Enhancers may or may not be regulating their nearest genes; they can even be situated thousands to hundreds of thousands of bases away from their gene targets (Medina-Rivera et al., 2018; Benabdallah et al., 2019; Friman et al., 2023). Chromatin folding and genome accessibility allow the interaction among distal enhancers and promoters of potentially transcribed genes in decondensed areas of the genome (Popay & Dixon, 2022) **(Fig. 1.4.)**. This spatial arrangement can either enhance or repress gene expression, providing precision to cellular context-dependent responses (Feuerborn & Cook, 2015; Kolovos et al., 2012).

TFs mediate the interaction between enhancers and promoters through a process involving chromatin-remodelling complexes, coactivators, and RNA Polymerase II (Pol II) (Tafessu & Banaszynski, 2020; Karnuta & Scacheri, 2018). Enhancers contain a cluster of binding sites for sequence-specific transcription factors, thereby acting as a scaffold to recruit the elements necessary for the assembly of active transcriptional machinery at target core promoters **(Fig. 1.4.)** (Karnuta & Scacheri, 2018; Spitz & Furlong, 2012).

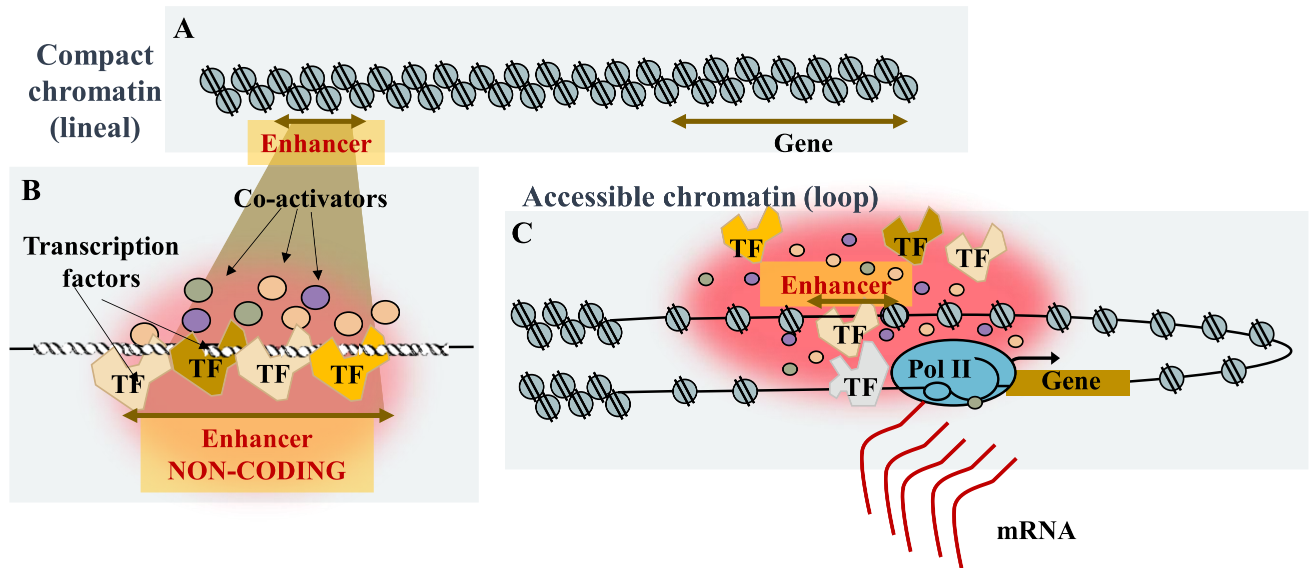


Figure 1. 4. Enhancer function and location in the chromatin context. The figure illustrates an example of a non-coding enhancer modulating the expression of a distal gene. **(A)** Enhancers and genes may be far apart in linear chromatin. **(B)** An enhancer located in a non-coding region serves as a recruiter of TFs with high affinity for its specific DNA sequence. These TFs recruit co-activators, creating a concentrated environment of transcription facilitators around the enhancer. **(C)** A 3D interaction between the enhancer and a distal promoter is mediated by chromatin looping and TFs. This interaction brings the cloud of co-activators and TFs recruited by the enhancer into proximity with the gene, increasing its transcription levels.

Chromatin structure determines whether enhancers and gene promoters are active in specific cell types (Lanctôt et al., 2007). Active enhancers are found in open chromatin regions and generally show bidirectional transcription, producing short-lived enhancer RNAs (eRNAs) (de Santa et al., 2010). Active enhancers are typically found surrounded by histone H3, which is covalently modified with monomethylation of lysine 4 (H3K4me1). Another important histone mark to identify enhancers is H3K4me3, which is generally used to detect active promoters (F. Cao et al., 2017; Thibodeau et al., 2017; Sharifi-Zarchi et al., 2017) but can also be found in hyperactive enhancers (Q. L. Li et al., 2019). H3K27ac histone mark is also present at active promoters (Tafessu & Banaszynski, 2020b).

Enhancers can exist in various states, including decommissioned (no histone modifications/inactive), poised (H3K4me1 and H3K27me3) (inactive, suppressed by the polycomb complex), primed (H3K4me1) (inactive, lacks H3K27ac), or active (H3K4me1 and H3K27ac) (Maurya, 2021). During development and cell differentiation, enhancers can switch between these states, correlating with changes in target gene expression and cell identity (Choi et al., 2021). Another type of enhancers, referred to in this dissertation as 'elongating enhancers,'

in addition to the typical enhancer mark (H3K4me1), possess the H3K36me3 mark associated with transcriptional elongation (Soldi et al., 2017). Some studies suggest that H3K36me3 is added to decommissioned enhancers that need to become active in response to environmental stimuli, and their transcription works as a priming signal that guides H3K4me1 deposition (Zentner et al., 2011; de Santa et al., 2010).

Enhancers play a crucial role in fine-tuning cell type-specific cellular phenotypes, and one gene can be regulated by multiple enhancers whose usage is differential in space and time throughout development (Kvon et al., 2021; Maurya, 2021; Osterwalder et al., 2018). Choi et al. (2021) found that most redundant enhancers cooperate in an additive way to regulate the transcription of assigned target genes. However, transcription of cell type-specific genes depends exponentially on the activity of its enhancers, indicating that these enhancers cooperate synergistically. Enhancer synergy appears to depend on cell type-specific transcription factors, suggesting that enhancer synergy contributes to cell fate determination (Choi et al., 2021).

1.4. Enhancers and disease.

1.4.1. Genetic factors underlying disease expression.

Some genetic differences among individuals (variants in their DNA) increase their likelihood of developing different diseases. For these diseases, we say there is genetic susceptibility. Genetic variants conferring susceptibility to complex diseases, whether inherited from parents or occurring spontaneously during gametogenesis, are typically found in the germinal DNA. However, susceptibility variants can also arise during early embryonic development or somatically throughout an individual's lifetime. These variants may not be found in the germinal DNA but in the DNA of specific affected tissues.

When the likelihood of developing the disease given that the individual possesses a specific genetic variant is very high, we say that the effect of the variant on the phenotype is *deterministic* (Strohman, 2002) (e.g., a recent study of Familial Aortic Disease showed that 100% of carriers of *ACTA2* missense mutation R118Q developed aortic disease above age 50 (Bobba et al., 2023), we call these variants *penetrant mutations* where penetrance refers to the percentage of carriers of the causative allele that develop the disease (Riordan & Nadeau, 2017; Cooper et al., 2013).

In some instances, a specific variant may confer susceptibility to the disease, but the likelihood of developing the disease will depend on the genetic background and other factors such as lifestyle and environmental exposures (Dipple & McCabe, 2000a). In those cases, we say the effect of the variant on phenotype is *probabilistic* (Strohman, 2002) and that its penetrance is *reduced* or *incomplete* (Cooper et al., 2013). For example, a study performed on family members carrying the p.V654L mutation in exon 19 of the *RB1* gene showed that the penetrance of this mutation was incomplete, with only 36% of the individuals developing unilateral retinoblastoma (Hung et al., 2011).

Genetic variants or environmental variables that affect the expression of the phenotype are called *modifiers* (Dipple & McCabe, 2000a; Riordan & Nadeau, 2017). Specifically, a *genetic modifier* is a locus where DNA variation can influence the expression of a phenotype primarily affected by variation at an independent locus. They can act non-additively and often do not affect the phenotype in absence of the primary genetic variant (Riordan & Nadeau, 2017; Cooper et al., 2013).

For instance, by studying multiple families with Hirschsprung disease, researchers demonstrated that although the presence of *RET* mutations is sufficient to explain susceptibility inheritance, specific alleles at a locus on 9q-31 are also required to cause the disease. This locus is a modifier of Hirschsprung disease penetrance (Bolk et al., 2000))

Besides assessing the penetrance of a given susceptibility allele, it is also important to study the effects of this allele on the severity of the symptoms once disease is manifested, as this directly affects the individual's quality of life. When a phenotype caused by an underlying genetic variant displays interindividual differences in severity (Dipple & McCabe, 2000a), we say it has different *expressivity* (Dipple & McCabe, 2000; Riordan & Nadeau, 2017; Cooper et al., 2013). The term *variable penetrance* can be used as a joint description of both variable expressivity (severity of phenotype) and penetrance (proportion of carriers with phenotype), and some factors underlying variable penetrance may be, for instance, environmental or genetic modifiers (Castel et al., 2018). For example, mutations in *CFTR* predispose to cystic fibrosis. However, a high degree of variability in the pulmonary phenotype has been observed between individuals with identical *CFTR* mutations, even within the same family. At least 11 different modifier loci have been described that alter the clinical phenotype (Paranjapye et al., 2020).

1.4.2. Disease risk and genetic variants at enhancers.

In some cases, germinal DNA mutations within enhancers can lead to significantly altered phenotypes and cause diseases. For example, certain congenital defects in humans can be attributed to mutations in enhancers associated with developmental genes, leading to conditions such as organ malformations or intellectual disabilities (Benko et al., 2009; De Vas et al., 2019; Smemo et al., 2012; Weedon et al., 2013). In some types of cancer, enhancers play a role in the overexpression of proto-oncogenes due to focal amplification or chromosomal translocations that result in the juxtaposition of proto-oncogenes and super-enhancers, a situation that is referred to as enhancer hijacking or due (Mikulasova et al., 2020; Kent et al., 2023). It's worth noting, however, that loss of function mutations in enhancers typically exhibit reduced penetrance, i.e., the probability of these mutations resulting in pronounced phenotypic effects is low (Osterwalder et al., 2018).

Variable penetrance of enhancer mutations could be partially explained by the fact that a single gene can be regulated by anywhere from zero to over a hundred enhancers (Schmidt et al., 2021). Hence, if the only enhancer of a non-functionally redundant gene has a loss-of-function mutation, there is a high probability that the phenotype associated with the gene function is altered (highly penetrant mutation). Conversely, if several enhancers regulate the expression of a gene, and one of them is affected, other redundant enhancers can still regulate the expression of the gene. (Kvon et al., 2021; X. Wang & Goldstein, 2020) **(Fig. 1.5.).**

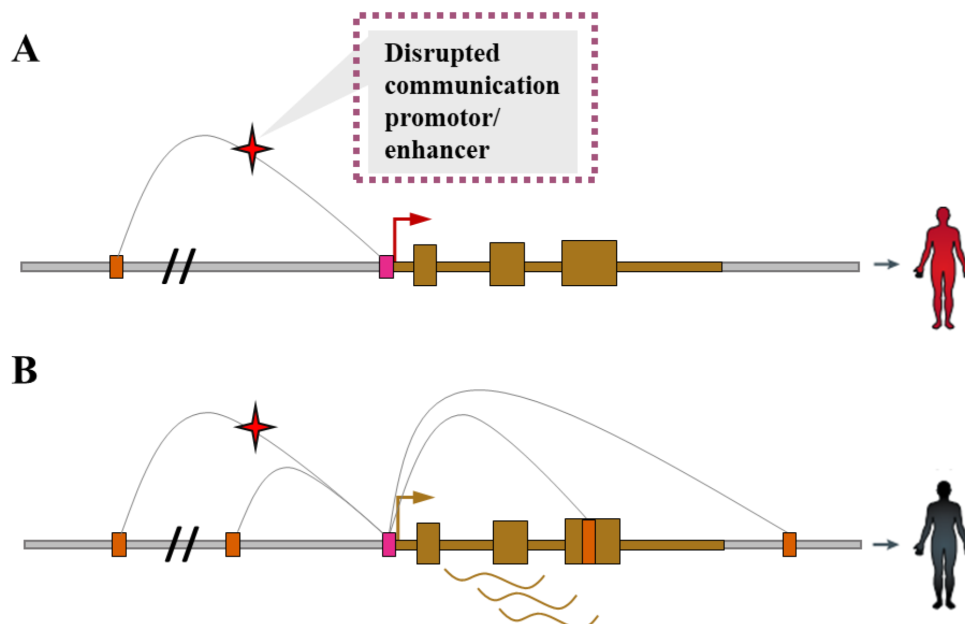


Figure 1. 5. Enhancer redundancy. The reason why mutations in coding regions are more likely to lead to disease than those in the non-coding areas is that the former can result in defective proteins, posing a greater chance of disrupting biological processes (high penetrance). This scenario also applies to mutations that affect TF binding sites in enhancers of genes

regulated by only one enhancer **(panel A)**. However, most of the human genes are regulated by multiple enhancers **(panel B)**. In these cases, if the connection between only one of the enhancers and the gene promoter is disrupted, the regulatory control can still be maintained by other redundant enhancers, and therefore, the phenotype will remain unaltered (low penetrance).

The prevailing model explaining the mechanism by which point mutations in enhancer sequences result in altered gene expression suggests that the mutations can modulate TF affinity, interfering with the normal formation of 3D loops that connect enhancers to their target genes/promoters (Karnuta & Scacheri, 2018). Various examples support this model; for instance, a specific enhancer indel can increase the affinity of RECQL, leading to increased *PARP1* expression, which in turn raises the risk for melanoma (Choi et al., 2017). Similarly, an SNP associated with prostate cancer risk can increase *RFX6* expression by enhancing HOXB13 binding at the *RFX6* enhancer (Huang et al., 2014). Additionally, SNPs in the *BLC11A* enhancer can affect foetal haemoglobin levels in human erythroblasts by modulating GATA1 and TAL1 binding(Bauer et al., 2013) **(Fig. 1.6)**.

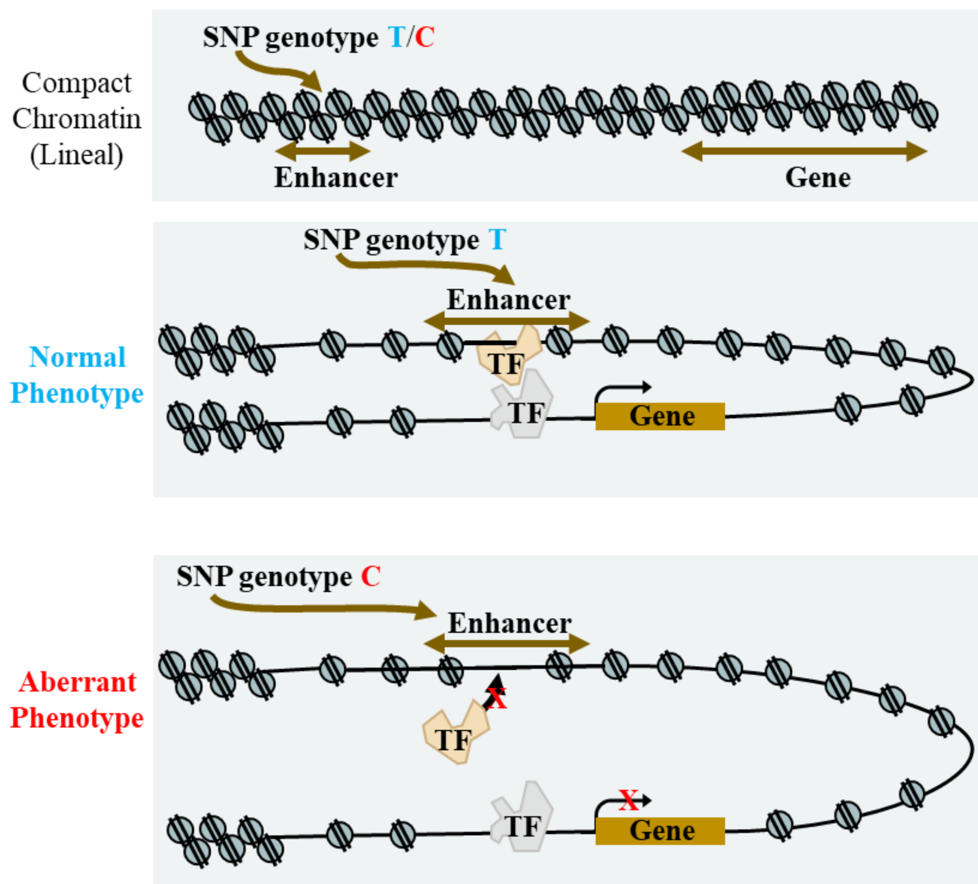


Figure 1. 6. Enhancers' role in disease. One prevalent hypothesis regarding the involvement of enhancers in disease development suggests that different alleles in TF binding sites may exhibit varying TF affinity. This variation can impact communication between enhancers and genes, resulting in altered gene expression, phenotype, and, in some instances, disease. The figure illustrates an example wherein the presence of the 'C' genotype at an SNP within an enhancer leads to the inability to recruit TFs, disrupting the enhancer's communication with its target gene.

1.4.3. Complex disease risk: SNPs in enhancers and other genetic contributors.

Common complex diseases are those that develop as a consequence of the interplay between an inherited genetic liability and exposure to certain environmental factors (Mitchell, 2012). Some examples of common complex diseases are bipolar disorder, coronary artery disease, Crohn's disease, hypertension, rheumatoid arthritis, and diabetes (Burton et al., 2007).

Genetic variants conferring susceptibility to complex diseases, whether inherited from parents or occurring spontaneously during gametogenesis, are typically found in the germinal DNA. However, susceptibility variants can also arise during early embryonic development or somatically throughout an individual's lifetime. These variants may not be found in the germinal DNA but in the DNA of specific affected tissues.

Aiming to uncover the genetic susceptibility to these diseases, researchers employed Genome-Wide Association Studies (GWAS) (Buniello et al., 2019; Visscher et al., 2017). GWAS follow a simple design: compare allele frequencies for hundreds of thousands of common SNP variants spread across the germline genome between large samples of disease cases and controls (Hirschhorn & Daly, 2005). Peripheral blood samples are commonly used for genotyping in GWAS due to their easy accessibility and high DNA yield after extraction (Abraham et al., 2012; Corvin et al., 2010) **(Fig. 1.7.)**

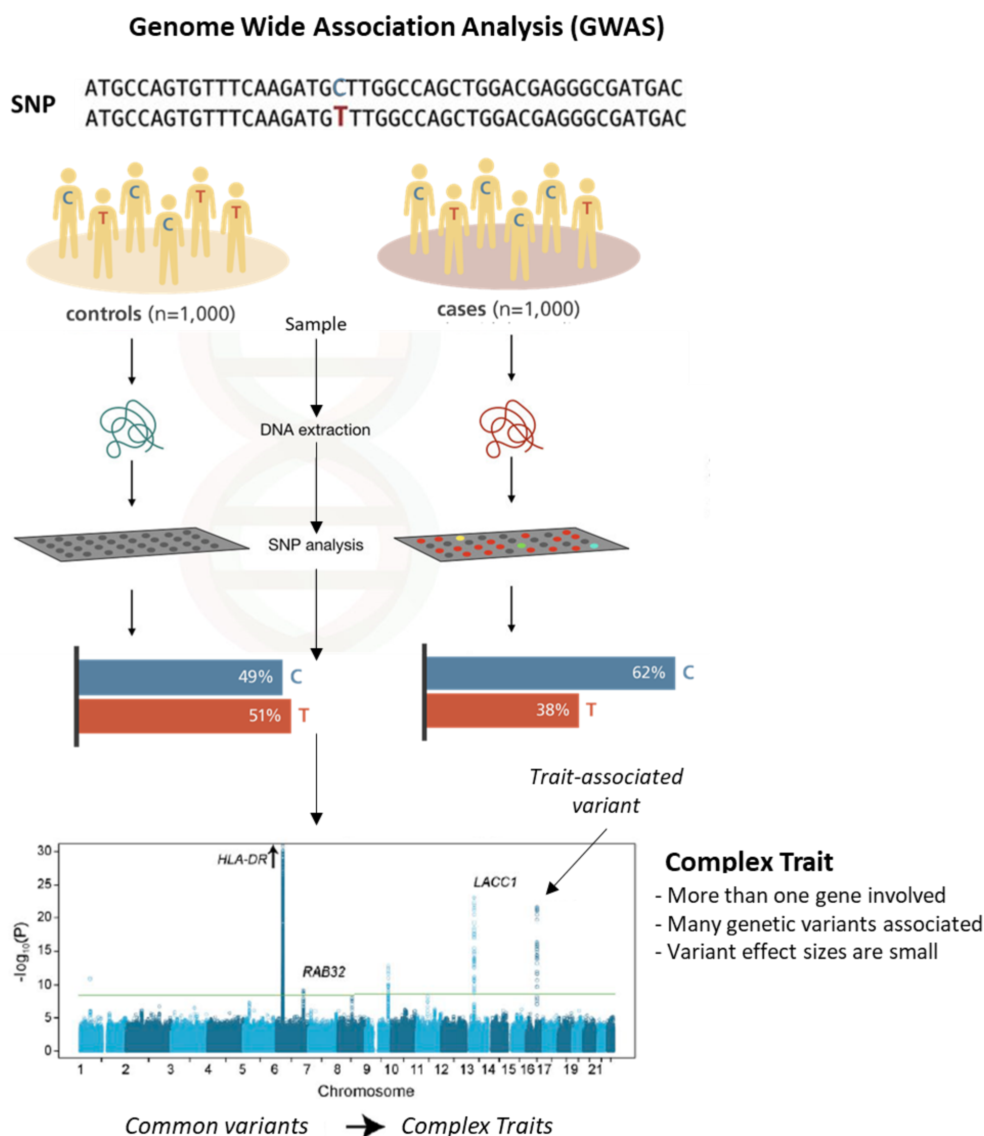


Figure 1. 7. Overview of Genome-Wide Association Studies (GWAS). The figure illustrates the general principles of GWAS analyses, where a control group is contrasted with a group of individuals with a given medical condition. At the top, two DNA sequences represent a genomic region with a single SNP highlighted in red. In this hypothetical example, all genomes of two distinct groups, each comprising 1000 unrelated individuals, are genotyped. One group comprises cases with a particular trait of interest, while the other serves as the control. For individuals in both case and control groups, the frequency of the genotype at each SNP is assessed within specific genomic regions. Statistical analysis determines whether specific genotypes are significantly overrepresented in the case group. The bottom of the figure illustrates the outcomes of the GWAS represented in a Manhattan plot; the x-axis shows the location of SNPs across the entire genome, while the significance of associations is represented on the y-axis. A green line represents the significance threshold. Notably, in the context of complex traits, numerous variants achieve significance. Once significant SNPs are identified, the next step is to examine their location in either coding or non-coding regions to gain insights.

GWAS assume that the genetic risk for common diseases will often be more prevalent in affected rather than in healthy individuals of the general population. The justification for using this study design has been founded on two observations: the non-mendelian segregation of the traits and their high prevalence (Becker, 2004; Mitchell, 2012). Since the inheritance of common complex traits aggregates in families but does not segregate in ways consistent with Mendelian inheritance, researchers assumed that these complex traits were polygenic in nature; that is, they arise due to the combined effects of the large number of genetic variants associated with the disease (Mitchell, 2012). As these diseases are common, researchers proposed the common disease/common variant (CD/CV) hypothesis, which states that susceptibility alleles, which are necessary but *not* sufficient to cause disease, should be found in the population at relatively high frequencies (>1%) (Becker, 2004).

Results of GWAS studies show that complex traits are associated with multiple variants with small effect sizes (Buniello et al., 2019) and that most of the variants associated with common complex diseases are not within genes but located in non-coding regions of the genome (X. Wang & Goldstein, 2020). From these observations, two concerns were raised: first, how can the functional impact of the non-coding variants be interpreted? and second, do GWAS results recover all genetic factors accounting for genetic susceptibility?

Regarding the first concern, researchers have shifted their focus to enhancers (Karnuta & Scacheri, 2018; Maurya, 2021; Claringbould & Zaugg, 2021). This interest is rooted in the fact that non-coding disease-associated variants exhibit a significant co-localization with enhancer features in epigenomic annotations (Javierre et al., 2016; Mumbach et al., 2017; J. Nasser et al., 2020, 2021). However, connecting enhancers to disease mechanisms is a complex task. This is because enhancers play a role in cell-type-specific and sometimes even condition-specific regulatory circuits. Therefore, it is essential to have enhancer maps for as many cell types and conditions as possible to interpret the effects of non-coding variants in disease.

The second concern was whether GWAS results recovered all genetic factors accounting for disease susceptibility. In this regard, it seems that, although GWAS results offered many robust and interesting associations, several other genetic sources of disease susceptibility are likely missed from this analysis. There are many possibilities about where to find the missed genetic factors behind disease predisposition (Trerotola et al., 2015; Eichler et al., 2010; López-Cortegano & Caballero, 2019), but here I would like to highlight three of them:

1. Due to the GWAS study's design, which focuses on common SNPs, rare single nucleotide variants (SNVs) and other genetic variants different from SNPs would be missed. For instance, *structural variants* (SVs), including deletions, mobile-element insertions, inversions, and copy-number variants, have been identified as contributors to a number of common diseases (Porubsky et al., 2022; Gonzalez et al., 2005; McCarroll & Altshuler, 2007; Yang et al., 2007; Antonacci et al., 2014; Mefford & Eichler, 2009; Büki et al., 2023; Girirajan et al., 2011; Billingsley et al., 2023) and are generally presumed to act through their effects on gene expression (Weischenfeldt et al., 2013). Since SVs are inherently more prone to mutations and rearrangements, the genotyping of these regions has been a challenge during the generation of reference genome assemblies. Additionally, the proportion of SNPs in linkage disequilibrium (LD) with SVs highly depends on the SV class, with duplications being currently under-ascertained for disease associations using tag-SNP-based approaches (Sudmant et al., 2015). Recent studies have shown that SVs are a major source of gene expression variation among humans and that their impact on gene expression may be larger than that of SNVs and indels (Chiang et al., 2017; Scott et al., 2021). Additionally, these studies suggest that most expression-altering SVs are noncoding and enriched at enhancers and other regulatory elements.

2. GWAS typically treat alleles inherited from the mother and the father as equivalent. However, identical inherited DNA sequences can have different effects based on the parental origin. *Parent-of-origin effects* (POEs) occur when the phenotypic effect of an allele depends on whether it is inherited from the mother or the father (Skaar et al., 2012). Recent evidence suggests that POEs can be important contributors to complex trait variation (Beaumont et al., 2023; Granot-Hershkovitz et al., 2020; Hofmeister et al., 2022; Hochner et al., 2015). A recent study performing a parent-of-origin GWAS for 21 quantitative phenotypes in a large Hutterite pedigree identified POEs with 11 phenotypes, including risk factors for cardiovascular disease (Mozaffari et al., 2019). Taking into account POE, they found twice as many genome-wide significant loci overall compared to standard GWAS of the same phenotypes in the same individuals (Mozaffari et al., 2019)

   Several phenomena can cause POEs, but the best characterised is genomic imprinting. Imprinted genes are a subset of genes that exhibit monoallelic expression controlled by parental-specific epigenetic marks established in gametogenesis and early embryonic

development and persisting in all somatic cells throughout life. These epigenetic marks include DNA methylation and histone modifications that regulate monoallelic expression by affecting promoter accessibility, chromatin structure, and chromatin configuration (Skaar et al., 2012).

Genetic defects that affect the monoallelic and parent-of-origin-specific expression of imprinted genes are the cause of several rare disorders (Eggermann et al., 2023). More recently, it has been suggested that some POEs associated with complex traits may be due to imprinting (Y. Zeng et al., 2019). Although previous studies estimated the number of imprinted expressed genes in the human genome at around 100 (Bartolomei & Ferguson-Smith, 2011), these imprinted regions also contain regulatory variants, and POEs can be spread to their genomic targets (Partida et al., 2018; Wolf et al., 2008). As such, imprinting-caused POEs on DNA methylation may have downstream effects on complex traits.

A recent study examined the entire genome of 5,101 individuals, focusing on CpG sites where methylation could be affected by POEs. They identified specific SNPs that act as modifiers of DNA methylation at these POE-influenced CpG sites, termed POE-mQTL SNPs. The study found that for 586 CpG sites, these POE-mQTL SNPs were located in regions of the genome known to be imprinted, providing strong evidence for the influence of imprinting on these CpG sites. Additionally, they discovered associations between some of these POE-influenced CpG sites and various traits, such as cardiovascular disease risk factors (Y. Zeng et al., 2019).

3. Linear models are commonly employed in GWAS, which assume an additive genetic architecture. In these models, the effects of individual genetic variants are assumed to act independently and additively (Hivert et al., 2021; Hill et al., 2008), overlooking the potential of these variants interacting synergistically due to *epistatic effects*. Although epistasis is difficult to identify in population studies (Wei et al., 2014), it is likely to be part of the genetic architecture of many traits, including complex diseases. Research using chromosome substitution strains in mice and rats to study the genetic architecture of blood, bone, and metabolic traits found that these complex traits tend to be highly polygenic and strong epistasis was found among the individual chromosomes (Shao et al., 2008). Additionally, "simple traits", traditionally thought to be mendelian and monogenic, can transform into complex traits with variable penetrance due to the action

of modifiers (Dipple & McCabe, 2000a, 2000b). In traditional GWAS, the effect size of some variants with epistatic effects in an individual would be underestimated or missed if, individually, the effect size of each variant is not high enough to achieve significance. Even in models accounting for epistasis, although the individual variants interacting could be common, the combination of them may be rare in the population (Eichler et al., 2010; Wei et al., 2014). More recently, new statistical models have been developed, and upon their application, they have found that epistasis is present in several common traits (Sheppard et al., 2021). Also, research has shown that enhancers regulating disease-associated genes may act on their targets in synergistic ways (Lin et al., 2022).

1.5. Enhancer maps and enhancer-target gene links

Identifying enhancers, determining their activity in different cell types, and pinpointing their target genes are critical steps in unravelling the genetic basis of complex diseases and phenotypic variation, as well as understanding how enhancer function can be disrupted by genetic variants. Therefore, to study the dynamics of enhancer gene expression regulation, we must overcome three challenges: first, we need to identify the localisation of enhancers across the linear genome; second, we need to know in which cell types these enhancers are active; and third, we need to determine which genes they are regulating.

To address the first and second challenges, epigenomic data from several cell types obtained from ChIP-Seq experiments (Roadmap Epigenomics Consortium et al., 2015; Nakato & Sakata, 2021), cap analysis of gene expression (CAGE) (Andersson et al., 2014; Melgar et al., 2011), or accessibility assays of chromatin - like DNase I hypersensitive site (DHSs) sequencing (DNase-seq) and Assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Grandi et al., 2022; Klemm et al., 2019a; Thurman et al., 2012) - can be employed **(Fig.1.8.A)**.

Various methodologies can be employed for inferring enhancer-gene regulatory interactions, including the utilisation of data on chromatin 3D contacts (Karbalayghareh et al., 2022; Avsec et al., 2021; Whalen et al., 2016), correlations between DNA accessibility and gene expression (Boix et al., 2021; Thurman et al., 2012; Sheffield et al., 2013), and the use of eQTL data (Y. Liu et al., 2017). Predictive models that combine multiple types of information and CRISPRi-based enhancer experimental data result in higher precision predictions than strategies relying solely on one type of information (Fulco et al., 2019; Luo et al., 2023).

For the third challenge, high-throughput chromosome conformation capture (Hi-C) from (Lieberman-Aiden et al., 2009) data **(Fig. 1.8.B)** and Expression Quantitative Trait Loci (eQTLs) data (Cano-Gamez & Trynka, 2020) **(Fig. 1.8.C)** can be utilised. Hi-C enables mapping of physical interactions between distant genome regions within the cell nucleus, providing insights into spatial proximity (Lieberman-Aiden et al., 2009), which is a key aspect of enhancer-gene communication **(see section 1.1.)**. eQTL analyses associate genomic variants with variations in gene expression levels across a population of individuals from the same species, revealing genetic determinants of gene expression variation (Cano-Gamez & Trynka, 2020).

Javierre et al. (2016) used promoter capture Hi-C to pinpoint genomic regions interacting with 31,253 gene promoters across 17 different human primary haematopoietic cell types, revealing the specificity of the interactions to the respective cell types and enrichment in connections between active promoters and enhancers (Javierre et al., 2016). Ongen et al. (2017) established the connection between noncoding genome regions to genes across multiple tissues from the GTEx project, enriching the interpretation of GWAS results by revealing specific tissues causally linked to given genetic associations (Ongen et al., 2017).

Other studies have provided landscapes of enhancer-promoter interactions in different cell contexts using the above-mentioned data types. For example, Mumbach et al. (2017) studied chromatin interactions centred on enhancers of primary human T cells using ATAC-seq and HiChip data for the H3K27ac histone mark, producing enhancer-promoter connectivity maps and chromatin accessibility maps across the genome of naïve T cells, regulatory T cells, and T helper cells at different differentiation steps (Mumbach et al., 2017).

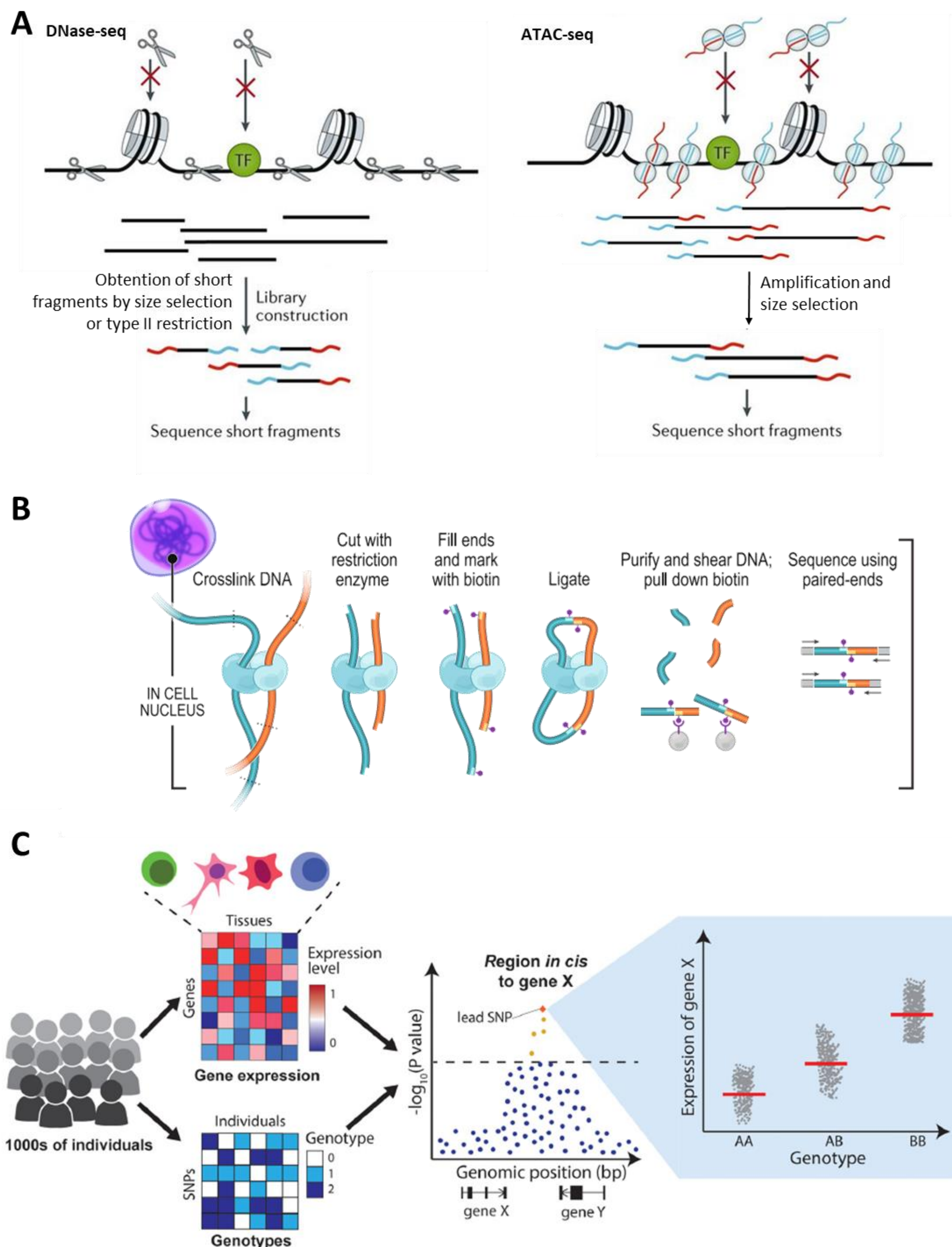


Figure 1. 8. Overview of enhancer target gene identification techniques. This figure outlines three standard experimental methods for elucidating enhancer target genes. **(A)** Outline of Chromatin accessibility profiling employing DNase-seq and ATAC-seq. These techniques identify open chromatin regions associated with regulatory elements like enhancers – reprinted from (Klemm et al., 2019) **(B)** The three-dimensional structure of the human genome using the Hi-C technique, revealing spatial interactions between genomic regions - reprinted from (Lieberman-Aiden et al., 2009) **(C)** Outline of the Analysis of eQTLs to identify genetic variants linked to changes in gene expression, providing insights into enhancer-target gene associations – reprinted from (Cano-Gamez & Trynka, 2020)

In another study, Engreitz et al. (2019) introduced the "Activity-by-Contact (ABC) model," a computational approach designed to predict gene targets of enhancers within human immune

cells. This innovative model identifies cell type-specific enhancers using data from chromatin accessibility measurements and histone modifications. It associates enhancer-gene pairs based on the proximity of enhancers to the transcription start site (TSS) of genes and the contact frequency derived from Hi-C matrices specific to each cell type (Fulco et al., 2019).

The ENCODE project Consortium (Abascal et al., 2020; The ENCODE Project Consortium 2012), The Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium et al., 2015; Bernstein et al., 2010), and The EPIMAP (Boix et al., 2021), FANTOM (Andersson et al., 2014) and BLUEPRINT (Adams et al., 2012) projects, have collectively contributed extensive data of the above-mentioned types for various cell types and tissues, significantly advancing our understanding of gene regulation and epigenetic mechanisms in human biology. However, considering the numerous cell types in humans (Regev et al., 2017), and given the cell-type and context-specific nature of epigenomic elements, the generation of data for a diverse range of cell types remains crucial, as current epigenomic landscapes are far from complete.

Moreover, considering the impact that genetic variation among individuals may have on the epigenome is also crucial. Some studies have found that the epigenomic changes are strongly influenced by genetic variability among European individuals (Hou et al., 2023; Chen et al., 2016; Bell et al., 2018). For instance, a study analysing regulatory elements with H3K27ac peaks that overlap with promoter or enhancer annotations in samples from brain, heart, muscle, and lung tissues found that a subset of them were genetically influenced by alleles at hundreds of genetic variants across tissues; these are termed "histone acetylation quantitative trait loci" (haQTLs). This result shows that SNVs can influence the activity of enhancers and promoters and, therefore, contribute to differential regulation of gene expression in various tissues across different genotypes (Hou et al., 2023).

Furthermore, another study analysing DNA methylation levels in regions in strong LD with GWAS SNPs, identified risk haplotype-specific DNA methylation (HSM) peaks, where methylation levels were influenced by the allelic count of the SNPs within haplotype blocks (Bell et al., 2018). These HSM peaks overlap with common non-coding SNPs in CpGs with the potential to affect CpG density and transcription factor binding sites. These HSM peaks also overlap with other common genetic variants such as CNVs, indels, short tandem repeats (STRs), Long Terminal Repeats (LTRs), and Alu repeats, of which a large proportion were functionally annotated as enhancers, open chromatin regions, or CTCF DNA motifs, suggesting their potential roles in gene expression regulation (Bell et al., 2018).

These studies serve as evidence that the epigenomic landscape can vary in the European population as a result of interindividual genetic variability. This suggests that genetic-driven epigenomic variability may be even more drastic among individuals with different ancestry. For that reason, acknowledging ancestry information when annotating and interpreting GWAS loci across diverse population backgrounds would also be crucial, as it has the potential to expedite research into disease risk factors and health disparities across populations with distinct genetic backgrounds.

As suggested by the analysis of Breeze et al. (2022) on the publicly available data from the International Human Epigenome Consortium (IHEC), current reference epigenome maps lack population diversity with significant discrepancies in the representation of various populations. Data from European populations is predominant, followed by a big disparity by that of African American populations, while the rest of the ancestries are notably underrepresented. This result emphasises the imperative need for further exploration to assess the diversity of epigenomic profiles within and between populations from different ancestry (Breeze, Beck, et al., 2022).

Recent studies suggested that the activity of enhancers varies across ancestries. Pettie et al. (2024) analysed the activity of regulatory regions, including enhancers and promoters (identified by ATAC-Seq and H3K27ac peaks) and the effect on gene expression of their target genes (identified by integrating HiC data in the ABC model) in lymphoblastoid cell lines from individuals across populations of African and European ancestries. They found that differentially expressed genes between ancestries were predicted by differential activity of enhancers rather than promoters and that the differential affinity of TF binding to these enhancers, given the allelic variability among the ancestries, was a contributor to these results (Pettie et al., 2024).

Another research study explored the relationship between genetic variation, DNA methylation, and gene expression from lymphoblastoid cell lines across individuals from five diverse human populations, including Yakut, Cambodian, Pathan, Mozabite and Mayan (Carja et al., 2017). They found a strong correlation between population-specific patterns of DNA methylation, gene expression, and genetic variation. Specifically, their results indicate a stronger genetic influence on DNA methylation than in gene expression patterns (Carja et al., 2017).

Similarly, a study that investigated the role of genetic variants in modulating genome-wide DNA methylation levels across the genome of individuals from various ancestries, including European, American, African, East Asian, and South Asian, identified polymorphic STRs associated with DNA methylation levels in several CpGs, termed “mSTRs” (Martin-Trujillo et

al., 2023). These mSTRs were enriched within introns, coding regions, enhancers and promoters compared with all STRs. To determine whether these STRs independently regulate DNA methylation or if they were in LD with other causal variants, they performed fine mapping analysis. Further analysis revealed that a subset of the fine-mapped mSTRs was well-tagged by nearby SNVs and showed potential associations with various health-related traits and diseases. Integration of gene expression data also indicated that some of these mSTRs modulate expression levels of nearby genes, highlighting their regulatory potential (Martin-Trujillo et al., 2023).

Therefore, considering how population genetic variability and ancestry influence the causal effect sizes (that is, the magnitude of change in phenotype per allele substitution) of disease risk variants, as well as the generating epigenome reference maps from diverse genetic backgrounds, would be of foremost importance when interpreting genetic variants associated with disease and is critical for the transferability of genetic risk knowledge across populations.

1.6. Haematopoietic cells

One of the best-studied collections of cell types in terms of their epigenome are those from the haematopoietic lineage (Adams et al., 2012; Astle et al., 2016a; Clien et al., 2014; Lara-Astiaso et al., 2014), comprising the lymphoid (e.g. T-cell and B-cells) and myeloid (e.g. neutrophils and macrophages) lineages. This is due to their ease of accessibility by blood sampling and ease of separation into different pure cell types (Vasquez et al., 2016). Besides, haematopoiesis is one of the best-characterised paradigms of cellular differentiation (Orkin & Zon, 2008). In addition to being relatively easier to study, haematopoietic lineage cell types are also associated with numerous complex diseases (Bao et al., 2019).

Haematopoietic cells, originating from the bone marrow's haematopoietic stem cells, play crucial roles in numerous essential bodily functions such as oxygen transport, immune surveillance, and haemostasis (Orkin & Zon, 2008). These cells can be categorised into two main groups: lymphoid and myeloid cells, each with distinct functions in both innate and adaptive immunity, as well as some non-immune roles (A. G. Fisher, 2002; Orkin & Zon, 2008).

Myeloid cells are primarily involved in innate immunity, offering rapid, non-specific defence against infections (Q. Zhang & Cao, 2019). Among them:

- *Neutrophils* respond quickly to bacterial and fungal threats. In addition to recruiting and activating other cells of the immune system (Vignali et al., 2008), neutrophils directly attack pathogens by ingestion (phagocytosis), release of soluble anti-microbials (degranulation), and generation and release of neutrophil extracellular traps (NETs) (Mayadas et al., 2014). NETs are web-like structures composed of chromatin and serine proteases, with a high local concentration of antimicrobial components that trap and kill extracellular microbes independent of phagocytic uptake (Papayannopoulos, 2018).

- *Monocytes*, with the ability to differentiate into tissue-resident macrophages, bridge the gap between innate and adaptive immunity by engaging in processes like phagocytosis and antigen presentation (Coillard & Segura, 2019; Patel et al., 2017; Zhao et al., 2018).

- *Macrophages* play critical roles in both the induction and resolution of sterile (damage-induced) and infection-induced inflammation. They exhibit considerable plasticity, showcasing a pro-inflammatory phenotype for immune activation and an anti-inflammatory reparative phenotype for tissue resolution (G. Y. Chen & Nuñez, 2010; E. A. Ross et al., 2021).

- *Dendritic cells* play a vital role in adaptive immunity by capturing, processing, and presenting antigens to T cells (Cabeza-Cabrerizo et al., 2021; Macri et al., 2018).

- *Eosinophils and basophils* participate in allergic responses and immune responses against parasitic infections (Iype & Fux, 2021; Obata-Ninomiya et al., 2020). *Basophils* also release histamine, which enhances blood flow, promotes healing, and facilitates the migration of other immune cells to infection sites. *Basophils* also release heparin which prevents blood clotting (Stone et al., 2010).

Additionally, some ***non-immune haematopoietic cells*** are also part of the myeloid lineage, such as:

- *Erythroblasts* are precursors of *erythrocytes* (red blood cells), which are essential for oxygen transport throughout the body (Helms et al., 2018; Moras et al., 2017).

- *Megakaryocytes* are responsible for giving rise to platelets (thrombocytes), which play a vital role in haemostasis, preventing excessive bleeding upon vascular injury (Jenne et al., 2013).

- *Osteoclasts***,** which adhere to the bone matrix, secreting acid and lytic enzymes that degrade old or damaged bone tissue. This process, in collaboration with the synthesis and deposition of new bone by osteoblasts, a non-haematopoietic cell type, ensures the continuous renewal and adaptation of the skeletal structure (Boyle et al., 2003)

In adaptive immunity, lymphoid cells take central stage. Lymphocytes include, among others:

- *T cells*, which recognise and eliminate infected or cancer cells after priming by antigen-presenting cells.

- *B cells*, which produce antibodies that target pathogens and foreign substances.

Some lymphoid cells, such as natural killer cells and innate lymphoid cells, ***are also involved in innate immunity*** (Sonnenberg & Hepworth, 2019). Briefly:

- *Natural killer cells* can swiftly identify and destroy virus-infected or cancer cells (Vivier et al., 2008).

- *Innate lymphoid cells*, which include three types, ILC1, ILC2 and ILC3, are involved in elimination of intracellular pathogens,activation of macrophage (ILC1), elimination of parasites, recruitment of eosinophils and basophils (ILC2), elimination of extracellular bacteria and fungi, and recruitment of neutrophils (ILC3) (Eberl et al., 2015).

In summary, these diverse haematopoietic cell types work collaboratively to maintain overall health, responding to various challenges through engagement in both innate and adaptive responses. Dysfunction in this interplay is implicated in multiple diseases.

1.7. Monocyte lineage.

Monocyte lineage are innate immune cells with immuno-modulatory, inflammatory, and tissue-repairing capabilities. Monocytes originate from haematopoietic stem cells (HSCs) in the adult bone marrow and from bone marrow-derived progenitor cells in the spleen. From there they are released to circulate into the peripheral blood (Gordon & Taylor, 2005) **(Fig. 1. 9.).**

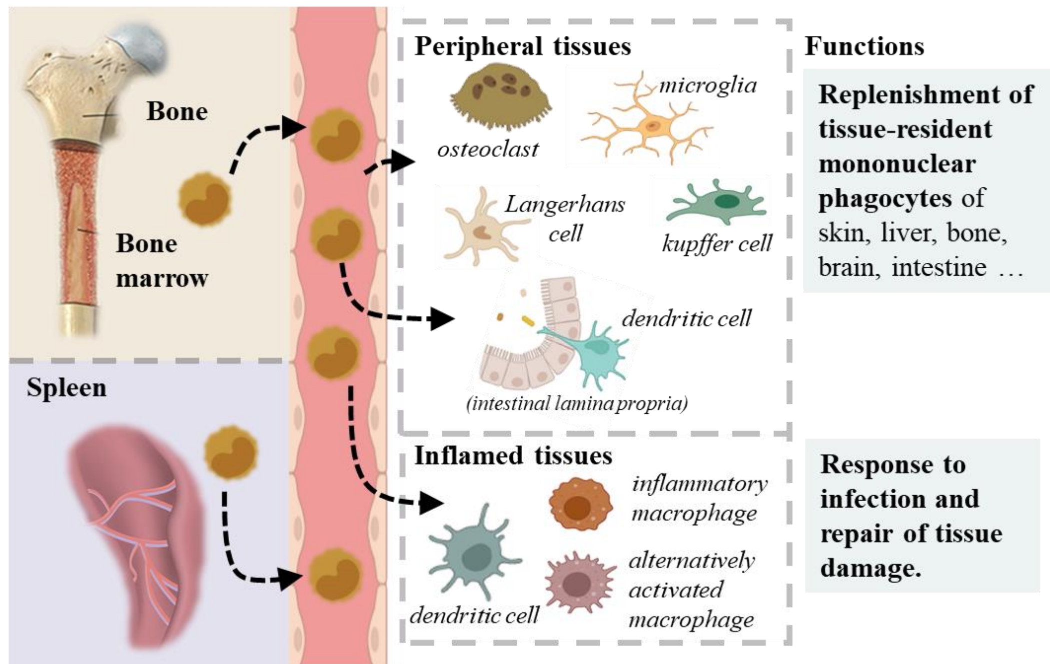


Figure 1. 9. Monocytes' role in inflammation and tissue homeostasis. Monocytes are derived from HSC in the bone marrow and progenitor cells in the spleen. They are then released to peripheral blood from where they can be recruited to inflamed tissues to differentiate into proinflammatory macrophages and DCs. Once inflammation is resolved, they differentiate into anti-inflammatory, pro-healing macrophages. Monocytes can also be recruited to peripheral tissues to replenish tissue-resident macrophages.

Monocytes can differentiate into macrophages, dendritic cells and osteoclasts (Coillard & Segura, 2019) **(Fig. 1. 10. A.)**. Although tissue-resident cell types may have a different developmental origin (Varol et al., 2015), during infection, monocytes rapidly migrate to inflamed tissues and differentiate into inflammatory DCs and inflammatory or anti-inflammatory macrophages, serving as the source of alternatives to tissue-resident macrophages (Italiani & Boraschi, 2015) **(Fig. 1.9.)**.

Macrophages are highly plastic innate immune cells that rapidly respond to diverse tissue-derived or environmental stimuli. They are involved in every stage of the acute immune response, as well as in the regulation of tissue homeostasis and in the orchestration of tissue repair processes. As phagocytes, they detect, engulf, and digest particles, microbes, and apoptotic cell debris. (Jain et al., 2019).

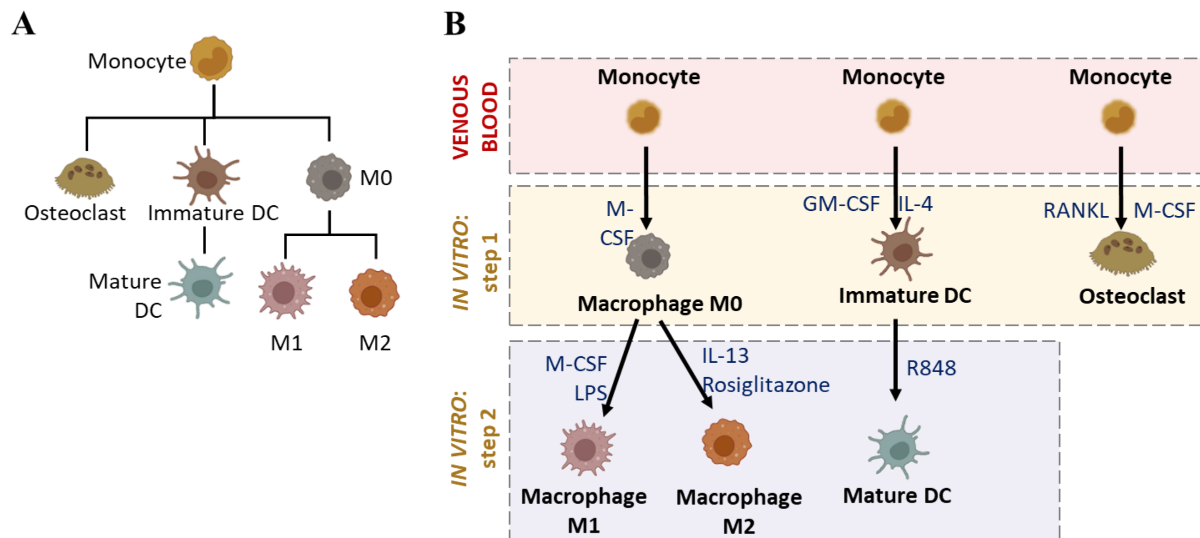


Figure 1. 10. *In vitro* models of monocyte differentiation. (A) Monocytes can undergo stepwise differentiation into osteoclasts, DCs, and macrophages. This process can be artificially induced *in vitro* by stimulating blood-derived monocytes with specific factors. **(B)** The figure depicts the stepwise differentiation process employed by the BLUEPRINT Consortium to generate *in vitro* models of macrophages, DCs, and osteoclasts (Adams et al., 2012).

The broad spectrum of macrophage functions depends on both the heterogeneity and plasticity of these cells (Remmerie & Scott, 2018; Viola et al., 2019). Macrophages are found in almost every tissue of the body, and monocyte-derived macrophages are quickly recruited to tissues upon loss of homeostasis (Gordon & Taylor, 2005) **(Fig. 1.9.)**. Macrophage phenotype and function are inherently tied to metabolic signals derived from their tissue environment (Buck et al., 2017; Van den Bossche et al., 2017). As a result, there are many different types of tissue-resident macrophages, such as osteoclasts (bone), alveolar macrophages (lung), microglial cells (brain), histiocytes (connective tissue), Kupffer cells (liver), Langerhans cells (LC) (skin), which are highly specialised in relevant functions to their particular tissue microenvironments (Italiani & Boraschi, 2015) **(Fig. 1.9.)**.

1.7.1. Classical M1/M2 macrophage in vitro model.

Although it is clear that macrophage phenotypes are difficult to categorise and should be seen as plastic and adaptable, the study of *in vivo*-generated macrophages is often difficult with laborious and cell-changing isolation procedures. (Luque-Martin et al., 2021).

The *in vitro* models using extreme macrophage phenotypes simplify their complexity and offer reproducible and robust platforms for studying human macrophages in health and

disease (Luque-Martin et al., 2021). The most extensively employed *in vitro* model of macrophage biology is based on the differentiation of primary blood monocytes into resting macrophages (M0) using colony-stimulating factor (M-CSF). This is followed by the subsequent stimuli-specific differentiation of M0 macrophages into two extreme phenotypes: M1 (pro-inflammatory) at one extreme and M2 (alternatively activated) at the other (Wculek et al., 2022) **(Fig. 1. 10. B)**. The *vitro* models using M1/M2 classification of extreme macrophage phenotypes simplify macrophage functions, providing a useful guide for reductionist approaches and standardisation in experiments.

Based on this definition, M1 equivalent macrophages *in vivo*, are those able to start and sustain inflammatory responses, secreting pro-inflammatory cytokines, activating endothelial cells, and inducing the recruitment of other immune cells into the inflamed tissue; on the other hand, *in vivo* macrophages equivalent to M2 are those that promote the resolution of inflammation, phagocytose apoptotic cells, drive collagen deposition, coordinate tissue integrity, and release anti-inflammatory mediators (Remmerie & Scott, 2018; Viola et al., 2019)

Current knowledge shows that the *in vivo* situation is more complex, and a wide range of molecules in the tissue microenvironment promote and impact monocyte to macrophage differentiation as well as activation. (Luque-Martin et al., 2021; Remmerie & Scott, 2018). Therefore, the most important limitation of the M1/M2 may not accurately represent macrophage behaviour *in vivo* (Nahrendorf & Swirski, 2016).

1.8. Aims of my project.

1.8.1. General aim.

Given that most of the genetic variants associated with complex diseases are non-coding, the functional annotation of non-coding regions in the genome becomes imperative for deciphering the mechanisms through which these variants modulate susceptibility to complex diseases.

The first assumption guiding my research is that susceptibility to these diseases arises from the dysregulation of gene expression in specific biological processes and tissues. This dysregulation is attributed to non-coding variants associated with the disease, impacting the function of regulatory regions that govern the expression of relevant genes.

Beyond the influence of genetic factors, the onset of complex diseases is profoundly shaped by environmental factors. Immune cells emerge as central players in this intricate interplay, serving

as primary sensors and responders to external stimuli. Their significance lies in their unique ability to act as both resident and circulating cell types, thereby extending their impact across all organs and tissues of the body. Given this pivotal role, my second assumption is that many complex diseases manifest as a result of defects in the phenotype of immune cells.

Prior research consistently underscores the enrichment of disease-associated non-coding variants in enhancers compared to other genomic elements. Furthermore, active enhancers play a pivotal role in controlling the context-dependent expression of genes, serving as a crucial link between genetic factors and the impact of environmental conditions on disease development. Consequently, my hypothesis posits that connecting haematopoietic enhancers with complex traits would significantly contribute to understanding susceptibility to complex diseases.

In light of these hypotheses, **the primary aim of this project is to propose connections between complex diseases and transcriptional enhancers active in haematopoietic cell types**. This connection will provide valuable hypotheses guiding future experimental research and enhancing our understanding of disease mechanisms.

To achieve this general aim, I proposed the following specific objectives:

1.8.2. Specific aims.

1. My first aim is **to generate the largest and most comprehensive collection of haematopoietic epigenomes, including the most common cell types of myeloid and lymphoid cells at various stages of differentiation**.
   This objective is motivated by the BLUEPRINT Consortium's foundational effort, which has generated the largest collection of profiles of histone marks for human haematopoietic primary cells so far. However, independent research groups have generated annotations for subsets of these cell types, employing diverse methodologies, thereby hindering direct comparability. In **Chapter 3**, I address this issue, building upon this dataset to generate a unified annotation of active enhancers across 107 samples from 31 different cell types.

2. My second aim is **to propose sets of haematopoietic cell types involved in susceptibility to complex traits**.

   Identifying the cell types impacted by defects in active enhancers is crucial to understanding the disease mechanism, designing or improving drugs to treat the disease

and understanding the response to treatment. In **Chapter 4**, I address this challenge by performing statistical analysis to test whether disease-associated loci were enriched with enhancers active in specific sets of haematopoietic cell types.

3. My third aim is **to suggest biological pathways potentially impacted by gene expression dysregulation due to disrupting enhancer activity in the context of Cardiovascular Disease.**

   Not only is it important to identify the cell types affected by genetic defects in enhancers (specific aim 2), but also the specific genes whose expression is being dysregulated in those cell types. This improves our knowledge of the biological processes being directly affected and provides a refined starting point to design experimental models that could be useful to understand the disease mechanism.

   Cardiovascular disease is the leading cause of death globally, and in **Chapter 5**, I used it as an example to showcase that by identifying genes whose regulation could be affected by disease-associated enhancers, we can generate information on the relevant pathways affected and propose disease mechanisms that guide the design of future experimental models. I focused on a set of complex traits, including diseases and risk factors related to cardiovascular risk, which in Chapter 4 were found to be predominantly associated with enhancers active in various types of macrophages. I investigated the gene targets of these enhancers and found them to be key players in all three lipid metabolism pathways. These genes are expressed in tissue-resident macrophages in different organs of the body, and these macrophages are known key actors in this process.

Chapter 2. Materials and methods

2.1. Sources for ChIP-Seq data acquisition.

The dataset utilised throughout the development of this PhD is sourced from public data. Specifically, the ChIP-Seq and RNA-Seq data employed in this study were generated by The Blueprint Consortium. The Blueprint Consortium was a five-year European project that ran between 2011 and 2016, aiming to advance our knowledge of gene expression regulation in healthy and diseased human blood cells (Adams et al., 2012). Data generated by the consortium is available in the European Genome-phenome Archive (EGA) (EGA Archive) (Freeberg et al., 2022).

EGA is a database that securely stores "personally identifiable" genetic and phenotypic information obtained from biomedical research projects. This means that the information in EGA can be traced back to specific individuals. Therefore, data within EGA is not readily accessible due to permission requirements that ensure the safeguarding, privacy and confidentiality of this sensitive data. Additionally, the datasets lack consistent identification matching the original donor and sample IDs assigned by BLUEPRINT and other contributing consortia. To address these challenges, a standardised and accessible resource known as The Epigenome Reference Registry (EpiRR) (EBI-EpiRR) was created.

EpiRR compiles references to raw data archived in public sequence repositories, such as EGA, ensuring both accessibility and traceability. EpiRR was created by The International Human Epigenome Consortium (IHEC) (Stunnenberg et al., 2016b). EpiRR serves as a centralised repository for cataloguing epigenomic datasets and relevant metadata generated by researchers from BLUEPRINT, The Canadian Epigenetics, Environment and Health Research Consortium (CEEHRC), The German Epigenome Programme (DEEP), The Encyclopedia of DNA Elements (ENCODE), The Korea Epigenome Project (KNIH), The NIH Roadmap Epigenomics, and The Japanese Agency for Medical Research in Development (AMED) in conjunction with The Strategic Basic Research Program of the Japanese Agency for Science and Technology (CREST) (AMED-CREST).

IHEC is a collaborative effort involving researchers from over ten countries and aims to provide high-quality reference epigenomic datasets for global research. Currently, IHEC is in the process of re-processing all available epigenomic data from various consortia. This

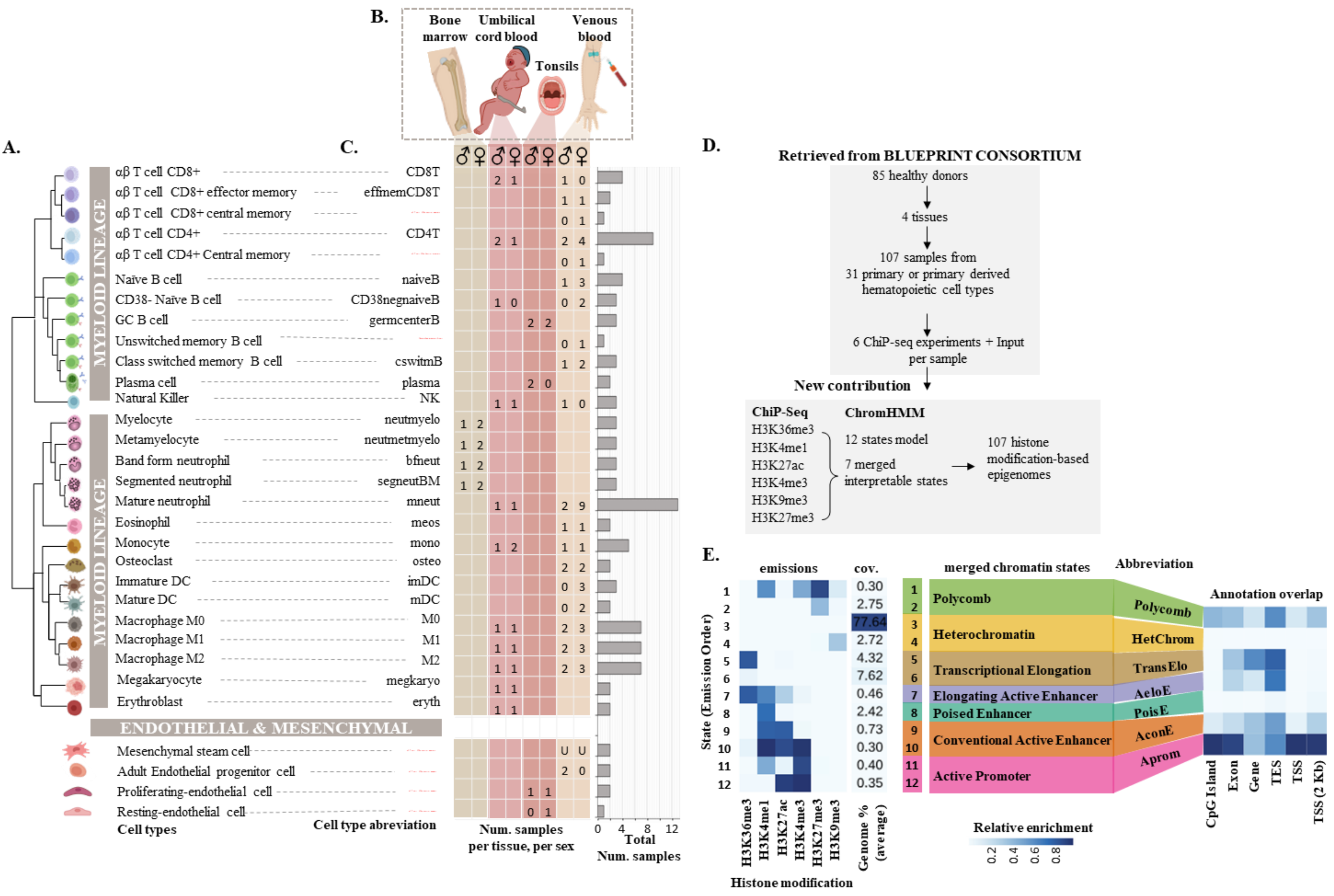
reprocessing adheres to standardised pipelines and quality control metrics established by IHEC based on the current understanding of epigenomic regulation in human cells and the state of epigenomic mapping technologies (IHEC Standards).

2.2. ChIP-Seq data and alignments.

For this PhD project, we did not use ChIP-Seq data fastq files; instead, we used the alignments of the fastq files against the human reference genome (assembly hg38/GRCh38), which were generated by the Data Analysis Centre (DAC) of the BLUEPRINT Consortium (Adams et al., 2012) and are available in EGA (study accession: EGAS00001000326) (EGA Archive).

Alignments of 642 ChIP-Seq experiments and 107 whole-cell extract sequenced controls (inputs) were retrieved from EGA as BAM files and converted to BED files. These alignments were performed by The Blueprint Consortium as described in (BLUEPRINT ChiP-Seq Analysis protocol). Briefly, alignments were performed using bwa 0.7.7 (H. Li & Durbin, 2009), Picard v2.8.1 (https://broadinstitute.github.io/picard/), and samtools v1.3.1 (Danecek et al., 2021). Duplicated reads were removed for all the experiments, and to assess the quality of the data, wiggle plots were generated using PhantomPeakQualTools (Kharchenko et al., 2008).

The dataset includes the whole-genome profiling of six histone marks (H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K27ac), as well as corresponding inputs/controls, for 107 human samples from 31 different blood-derived cell types belonging to the main haematopoietic lineages: myeloid (15 cell types) and lymphoid (12 cell types). These datasets also include endothelial (2 cell types) and bone marrow mesenchymal cells (2 cell types) **(Fig. 2.1.A-C).** Sample metadata, such as sex, tissue, biomaterial type (primary or primary derived), cell type, and EpiRR accessions, are included in **(Appendix Table A.1).** Data from most of the samples here included (78 out of 107) are part of the reference epigenomes that IHEC will release, and they conducted an extensive quality check. This information is expected to be published as part of the IHEC EpiAtlas paper in May 2024.



A.
B.
Bone marrow
Umbilical cord blood
Tonsils
Venous blood
C.
MYELOID LINEAGE
αβ T cell CD8+
αβ T cell CD8+ effector memory
αβ T cell CD8+ central memory
αβ T cell CD4+
αβ T cell CD4+ Central memory
Naïve B cell
CD38- Naïve B cell
GC B cell
Unswitched memory B cell
Class switched memory B cell
Plasma cell
Natural Killer
Myelocyte
Metamyelocyte
Band form neutrophil
Segmented neutrophil
Mature neutrophil
Eosinophil
Monocyte
Osteoclast
Immature DC
Mature DC
Macrophage M0
Macrophage M1
Macrophage M2
Megakaryocyte
Erythroblast
ENDOTHELIAL & MESENCHYMAL
Mesenchymal steam cell
Adult Endothelial progenitor cell
Proliferating-endothelial cell
Resting-endothelial cell
Cell types
CD8T
effmemCD8T
CD4T
naiveB
CD38negnaiveB
germcenterB
cswitmB
plasma
NK
neutmyelo
neutmetmyelo
bfneut
segneutBM
mneut
meos
mono
osteo
imDC
mDC
M0
M1
M2
megkaryo
eryth
Cell type abreviation
♂ ♀ ♂ ♀ ♂ ♀ ♂ ♀
Num. samples per tissue, per sex
0 4 8 12
Total Num. samples
D.
Retrieved from BLUEPRINT CONSORTIUM
85 healthy donors
4 tissues
107 samples from 31 primary or primary derived hematopoietic cell types
6 ChiP-seq experiments + Input per sample
New contribution
ChiP-Seq
H3K36me3
H3K4me1
H3K27ac
H3K4me3
H3K9me3
H3K27me3
ChromHMM
12 states model
7 merged interpretable states
107 histone modification-based epigenomes
E.
emissions
cov.
State (Emission Order)
0.30
2.75
77.64
2.72
4.32
7.62
0.46
2.42
0.73
0.30
0.40
0.35
Histone modification
Genome % (average)
merged chromatin states
Polycomb
Heterochromatin
Transcriptional Elongation
Elongating Active Enhancer
Poised Enhancer
Conventional Active Enhancer
Active Promoter
Abbreviation
Polycomb
HetChrom
TransElo
AeloE
PoisE
AconE
Aprom
Annotation overlap
CpG Island
Exon
Gene
TES
TSS
TSS (2 Kb)
Relative enrichment
0.2 0.4 0.6 0.8

Figure 2. 1. Dataset description. (A) The cell types of the haematopoietic lineage included in the analysis. **(B)** The tissues from which the cells were sampled. **(C)** The total number of samples of each cell type and the number of samples per tissue and sex. **(D)** A simplified diagram describing the data used to produce the 107 epigenomes, including ChiP-Seq of 6 histone modifications and control for each sample generated by the BLUEPRINT Consortium. **(E)** A heatmap in the left panel displays the emission probability signals from ChromHMM for our 12-states model. The middle panel describes the states based on literature and genomic annotation enrichments. Finally, the right panel shows the genomic annotation enrichment in each state after collapsing the 12 states model into seven interpretable states. Abbreviations: TSS (transcription start site) and TES (transcription end site).

Here we used the human reference genome assembly, hg38/GRCh38, which has limitations in accurately mapping repetitive regions such as those found in centromeres and telomeres. In 2022, a new reference assembly, the T2T-CHM13 assembly, was introduced to address these shortcomings (Nurk et al., 2022). The T2T-CHM13 assembly resolves gaps and corrects misassembled or patched regions present in GRCh38 in approximately 225 Mbp of genomic sequence, including entire acrocentric chromosome short arms, expansions of gene families, and a diverse array of repeat classes (Nurk et al., 2022).

This improvement has profound implications for epigenome annotation. A recent study aligning ChIP-Seq from ENCODE to the T2T-CHM13 demonstrates enhancement of the mapping of several histone marks, particularly those associated with constitutive heterochromatin such as H3K9me3 (19.4% more reads mapped compared to GRCh38) and facultative heterochromatin mark H3K27me3 (15.2% more reads mapped compared to GRCh38) (Gershman et al., 2022). Additionally, unresolved genes from high-copy-number gene families, such as GOLGA, NPIP, ZNF, and TBC1D3, in the hg38 assembly can now be accurately mapped using the new T2T-CHM13 reference assembly; in particular, precise localisation of H3K4me3 and H3K27ac peaks in 57 these regions previously unresolved genes was possible, and these include loci that have been associated with diseases such as cancer, autism and spinal muscular atrophy (Gershman et al., 2022).

Unfortunately, our project could not leverage the benefits of the new T2T-CHM13 assembly due to several reasons. Firstly, it was released after the initiation of our project (2020). Additionally, accessing the raw data required for the realignment posed challenges due to privacy concerns and access permissions. Furthermore, this reanalysis would have demanded significant computational resources and time, exceeding the constraints of our project.

Therefore, our version of BLUEPRINT epigenomes may not reflect the most up-to-date mapping of heterochromatin and repetitive regions, including genes in high-copy-number gene

families. However, the accuracy of other functional annotations outside repetitive regions, including enhancers, remains high.

2.3. Learning the haematopoietic chromatin state maps.

2.3.1. Authors contribution.

Different chromatin state maps generated and published by The BLUEPRINT Consortium are not comparable; this is because partial sets of the samples and different computational methodologies and/or reference genomes were used for their generation. Several research groups from BLUEPRINT generated ChromHMM models using ChIP-seq data aligned to the previous version of the human genome (GRCh37). For instance, a ChromHMM model (11 states) trained by Carrillo-de-Santa-Pau et al. (2017) with several cell types was applied to generate chromatin state maps for monocytes, neutrophils, CD4+ T-cells (Astle et al., 2016b; L. Chen et al., 2016; Ecker et al., 2017) and T-cell lineage (Cieslak et al., 2020). Javierre et al. (2016) used a different ChromHMM model for the epigenomic landscaping of a subset of samples from megakaryocytes, erythroblasts, neutrophils, monocytes, macrophages, endothelial precursors, B-cells, and T-cells (Javierre et al., 2016).Petersen et al. (2017) used a different software – IDEAS, instead of ChromHMM -to generate the chromatin state maps of megakaryocytes and erythroblasts (Petersen et al., 2017).

My supervisor, Dr. Daniel Rico, was a member of a research group associated with BLUEPRINT, led by Alfonso Valencia in Barcelona, Spain. In collaboration with Enrique Carrillo de Santa Pau and David Juan, who were also members of the group, they tackled the challenge of generating comparable chromatin state maps for BLUEPRINT samples.

They trained and applied a 12 states chromatin states model to a collection of 642 ChIP-Seq alignments to generate the 107 epigenomic maps described in **section 2.3.3**. This model has been applied in two published research papers to generate chromatin state maps of B-cell lineage (Beekman et al., 2018b) and neutrophil lineage (Grassi et al., 2018). However, chromatin state maps of the rest of BLUEPRINT cell types generated by applying this model have not been published yet. They will be included for the first time in the research paper that we plan to publish along with my dissertation results.

My contributions to this effort **(section 2)** were: to write a detailed description of the methodology implemented by my supervisor and collaborators (**sections 2.3.** and **2.4);** generate figures to accompany the description of the methodology; perform additional analysis to

justify/validate their chosen parameters, such as the number of states and the assignment of labels to the chromatin states; organise the data in a format that could be easy to use for the research community; and create a repository with the data to make it accessible to the research community. The work in the rest of the sections of this dissertation has been done by me, with the guidance

of supervisors and colleagues.

2.3.2. Building the chromatin-states model.

The ChromHMM software developed by Ernst and Kellis (2012) and based on multivariate Hidden Markov Models (HMM), facilitates the creation of chromatin state maps by generating models that capture the intricate combinatorial interactions between various chromatin marks across the genome of different samples (Ernst & Kellis, 2012).

In this study, ChromHMM version 1.10 was employed to generate a 12-state chromatin model **(Fig. 2.1.E)**. This model was trained on ChIP-Seq alignments for the core set of six histone modifications: H3K4me1, H3K4me3, H3K27me3, H3K36me3, H3K9me3, H3K27ac—as well as a corresponding input/control, recommended by IHEC guidelines, as described in **section 2.2**. The modelling process adhered to the guidelines recommended by ChromHMM developers, which included configuring prior parameters for training, such as the number of states.

Subsequently, the trained model was utilised to calculate the posterior probability of each state for every genomic bin within the epigenome of the 107 BLUEPRINT samples **(Fig. 2.1.D).** The genomic regions were then annotated by assigning the state with the maximum posterior probability to each bin, providing a comprehensive characterisation of the chromatin landscape.

Other ChromHMM models based on similar datasets to ours, that is, the same set of histone modifications and a similar number of samples but with different numbers of states, have been published before. Some examples are the 18-state and 15-state models by Roadmap Consortium (Roadmap Epigenomics Consortium et al., 2015) and the 11-state model by Carrillo et al. (Carrillo-de-Santa-Pau et al., 2017) **(Fig. 2.2.A).**

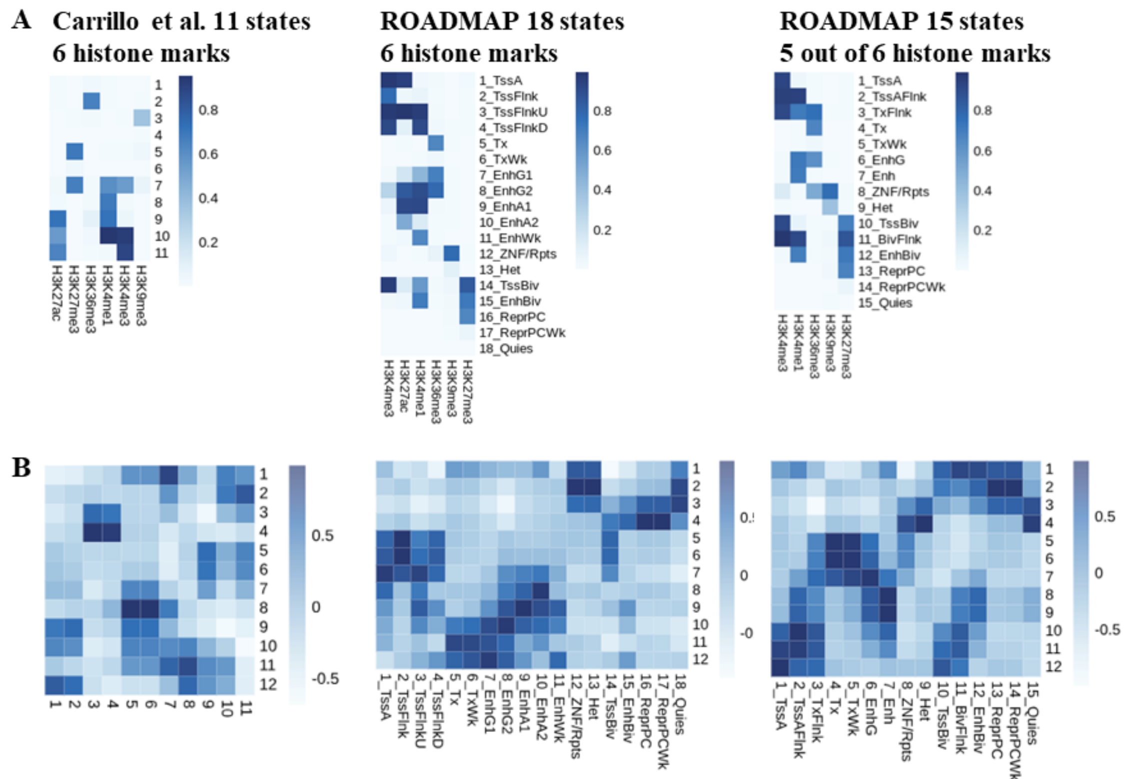


Figure 2. 2. Correlation between the emission probabilities of our model (12 states) with those of corresponding states from our previously published models. The left panel shows the emission probabilities of the 11-states model from Carrillo et al. (2017), while those from the 18 and 15-state models from Roadmap are presented in the middle and right panels, respectively. The figure displays the abbreviations of the chromatin state names used in the original papers; refer to **Appendix Tables A.1 and A2** for full names. **(B)** Heatmap displaying the coefficient values from the Pearson Correlation between our model and the 11-states model by Carrillo et al. (2017), as well as the 18-states and 15-states models by ROADMAP**.**

The choice of the number of states in a chromatin state model involves a trade-off between computational efficiency and interpretability. Simpler models have advantages in terms of interpretability and computational efficiency, but they may not capture all the complexity of chromatin dynamics. The previously published 11-state model (Carrillo-de-Santa-Pau et al., 2017) was obtained using an interim subset of the BLUEPRINT dataset and encapsulates key biologically interpretable states consistent with the larger 18-states and 15-state Roadmap models. It encompasses five primary regulatory states: active promoter, bivalent promoter, enhancer, elongation, and heterochromatin/low signal, with no functionally ambiguous chromatin states included. However, this model did not include the genic enhancer state previously described by Roadmap models **(Fig. 2.2.A).** Hence, our decision to employ a 12-state model, as the additional state revealed this type of elongated enhancers.

2.3.3. State labels, interpretation, and mnemonics.

To assign biologically meaningful mnemonics to the states of my 12-state chromatin model, we followed the methodology of Carrillo et al. (2017). I cross-correlated the emission probabilities of the states of our model with those from the 11 sates model by (Carrillo-de-Santa-Pau et al., 2017) **(see Appendix Table A.2)** and Roadmap models (with 18 states and 15 states) **(see Appendix Table A.3) (Fig 2.2.B).** I assessed which states from the 18 states and 11 states models exhibited correlation coefficients higher than 0.75 with the 12 states trained in our model **(Table 2.1.).** The 12 states model captures most of the states that were found in the larger models except by Flanking TSS Downstream, Weak Enhancer, Bivalent/Poised TSS, and Bivalent Enhancer.

Table 2. 1.Best match (correlation coefficient > 0.75) between each state in our 12 states-model states from Carrillo et al. (2017) and Roadmap (2015) models.

State from our model	Best match* from Carrillo et al. (2017) 11 states-model	Best match from Roadmap et al. (2015) 18 states-model	Best match from Roadmap et al. (2015) 15 states-model
1_Polycomb	7_Repressed Polycomb Promoter	""	11_BivFlnk, 12_EnhBiv
2_Polycomb	11_ Active TSS	12_ZNF/Rpts, 13_Het, 18_Quies	13_ReprPC, 14_ReprPCWk
3_HetChrom	""	18_Quies	15_Quies
4_HetChrom	3_Heterochromatin High H3K9me3, 4_Low signal	16_ReprPC, 17_ReprPCWk	8_ZNF/Rpts, 9_Het, 15_Quies
5_TransElo	""	2_TssFlnk	4_Tx, 5_TxWk
6_TransElo	""	2_TssFlnk	4_Tx, 5_TxWk
7_AeloE	""	1_TssA, 2_TssFlnk, 3_TssFlnkU	5_TxWk, 6_EnhG
8_PoisE	5_Heterochromatin High H3K27me3, 6_Heterochromatin Low H3K27me3	10_EnhA2	7_Enh
9_AconE	""	8_EnhG2, 9_EnhA1, 10_EnhA2	7_Enh
10_AconE	""	7_EnhG1, 8_EnhG2	2_TssAFlnk
11_Aprom	8_Active TSS	5_Tx, 6_TxWk, 7_EnhG1	1_TssA, 2_TssAFlnk
12_Aprom	""	6_TxWk, 7_EnhG1	1_TssA

*States with a coefficient value above 0.75 based on Pearson correlation between each state from our model and those from Roadmap et al. (2015) 15— and 18—states model.

Additionally, following Carrillo et al. (2017) methodology, I merged the labels of the 12-state model into seven interpretable labels, annotated using literature and genomic annotation enrichments (gene structures, CpG islands) **(Fig. 2.1.E).** This posterior collapse into seven chromatin states facilitates the functional interpretation of regulatory elements. Polycomb (states 1-2), Heterochromatin (HetChrom, states 3-4), Transcriptional Elongation (TransElo, states 5-6), Elongating Active Enhancer (AleloE, state 7), Poised Enhancer (PoiseE, state 8), Conventional Active Enhancer (AconE, states 9-10) and Active Promoter (Aprom, states 11-12) **(Fig. 2.1.E).** The a posteriori collapse of chromatin states for ease of interpretability is a methodology used in other published research such as (Carrillo-de-Santa-Pau et al., 2017; Cieslak et al., 2020; Ecker et al., 2017; Grassi et al., 2018).

To compare chromatin states between epigenomes, the genome coordinates were binned every 200bp, and the 23 chromosomes were concatenated into a single sequence. Then, I built a matrix with the bins as rows and samples as columns. The entries of the matrix are the chromatin state annotation at each sample for each bin. The matrix dimensions are 14,374,996 bins * 107 samples.

2.4. Cell type consensus epigenomes.

Previous studies, such as the work by Grassi et al. (2018) and Ecker et al. (2017), have defined cell-type consensus states by condensing annotations from individual samples. Grassi et al. (2018) defined consensus states for each region of the genome based on the consistent identification of such states in three replicates of the same cell type (Grassi et al., 2018). Ecker et al. (2017) adopted a similar approach but required a minimum of five biological replicates per cell type. In addition to this, they introduced the term "variable chromatin state" to indicate regions where annotations were not consistent in at least 80% of the biological replicates (Ecker et al., 2017).

Taking inspiration from these approaches, I defined consensus states for all cell types in our dataset with at least two biological replicates. I labelled bins as a "conserved state" if their annotation was consistent in at least 75% of the available biological replicates. When the annotation of a bin did not meet this level of conservation, I designated it as a "non-conserved state." (**Fig. 2.3.A**).

The decision to set a 75% conservation threshold aimed to strike a balance between accurately capturing chromatin states across diverse cell types and accommodating variations in sample

sizes. This approach ensures that most samples contribute to defining the chromatin state, avoiding unfair penalties for cell types with fewer replicates. In contrast, a 100% conservation threshold, while theoretically ideal in uniform sample conditions, would disproportionately penalise cell types with more samples. This could lead to the exclusion of valuable information and result in a biased representation of chromatin states in those tissues.

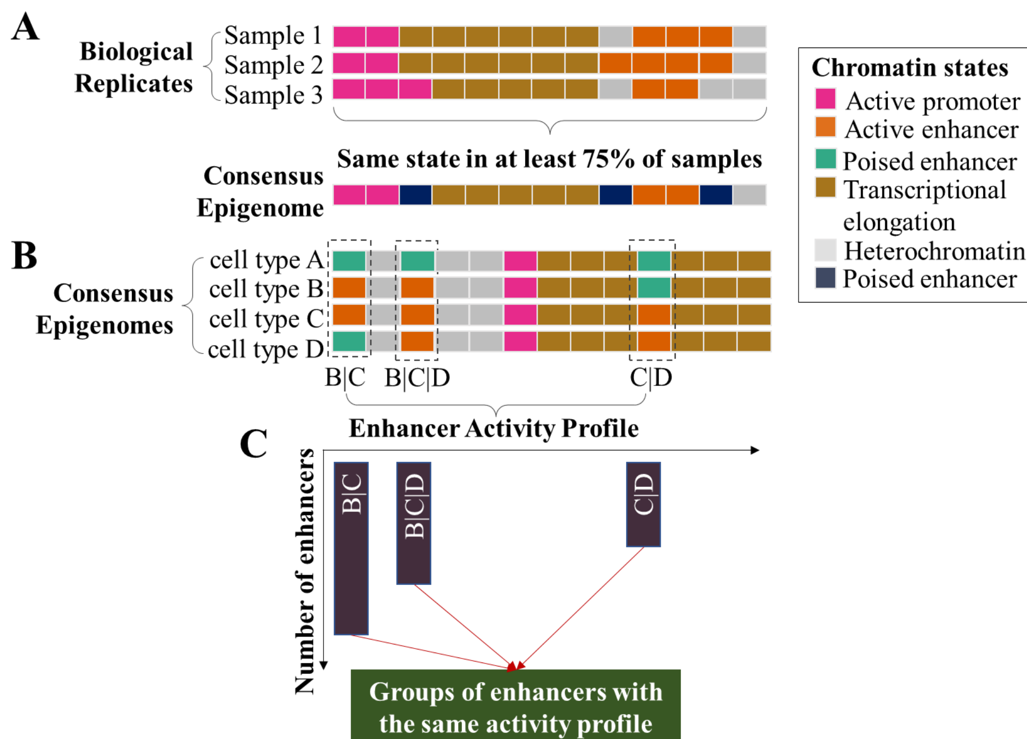


Figure 2. 3. Consensus epigenomes and enhancer groups by activity profiles. The figure shows hypothetical chromatin state annotations in a genome fragment, each rectangle represents one bin (200 bp) annotated using the chromatin states model. **(A)** For each cell type, I generated the consensus epigenome by assigning to each bin the states conserved in at least 75% of the samples available for that cell type. **(B)** For each enhancer bin across the genome, I assigned an activity profile, this is given by the set of cell types in which such enhancer is annotated as active enhancer. **(C)** I assigned enhancers to groups according to their activity profile. Some groups have more enhancers than others because the activity profiles are more common.

2.5. Enhancer activity profiles.

Epigenomic states are cell type-specific **(See sections 1.1 and 1.2),** which means that the same bin of the genome can be annotated as an active enhancer state in a specific cell type (or group of cell types) while having a completely different epigenomic annotation in the other cell types (Carrillo-de-Santa-Pau et al., 2017). To provide insights into the dynamics and specificity of

enhancer activity across different cell types, Ernst et al. (2011) defined “multicell activity profiles” by systematically analysing maps of chromatin histone marks associated with promoters and enhancers across nine cell types (Ernst et al., 2011). Taking their foundational work as a reference, I defined Active Enhancers (AE) bins (1bin = 200bp) as the collection of bins labelled with AeloE or AconE states in at least 2 out of the 107 epigenomes in our dataset (Total identified AE = 1,526,184 bins).

By defining AE bins based on individual epigenomes rather than using cell-type consensus epigenomes, my approach acknowledges variability in enhancer usage among cell types and different biological contexts. This individualised definition highlights functional elements that may be specific to certain epigenomic contexts.

I next defined the activity profile of an AE bin as the set of cell types in which that bin has been labelled as “conserved AE” in the consensus cell type annotations. Based on this definition and using the previously generated consensus epigenomes **(section 2.4)** for the 31 BLUEPRINT cell types, I found the activity profiles of each AE bin across the genome **(Fig. 2.3.B)**.

Considering all the possible combinations of cell types in which the enhancer can be active, I grouped AE bins with the same activity profile, no matter where in the genome they were located or if they were not consecutive. Distinct groups of enhancers were identified (107, 562). The groups based on enhancer activity pattern varied widely in number of enhancers **(Fig. 2.3.C)**, ranging from 1 to 75,814. Despite this broad range, most groups predominantly had fewer enhancer bins, with a mean of 10.5 bins (Q1 = 1 bin, Q3 = 2 bins), indicating a tight clustering of enhancer bin counts towards the lower end.

2.6. GWAS enrichment analysis.

To improve our understanding of the function of non-coding trait-associated variants, I combined information from two sources: data from the GWAS catalogue (Buniello et al., 2019), which is a database of genetic variants associated with complex traits, and the sets of AE bins identified in our dataset.

As we are interested in understanding non-coding regulatory regions, I excluded the SNPs whose genomic contexts were annotated in the GWAS catalog as "missense", "STOP-GAIN", and "frameshift", and kept only those annotated as ”intergenic", "", "ncRNA", "cds-synon", "intron", "UTR-5", "UTR-3", "splice-3", "nearGene-”, “nearGene-5”, or “splice-5". I restricted

the set of studies to those with more than 2000 individuals and the set of traits to those with more than 6 SNPs. This resulted in 11060 SNPs, 518 traits, and 985 PUBMED IDs.

I defined “trait-associated regions” across the genome as 10kb regions of the genome centred in GWAS SNPs (5 Kb on each side). To streamline the analysis, I focussed on enhancers with activity profiles among the top 100 most common across the 31 BLUEPRINT cell types, the frequency of these profiles ranged between 900 bins, up to 75,814 bins.

I found the overlaps between trait-associated regions and the sets of AE bins with the most common activity profiles. We wanted to test if there were significant enrichments of enhancers with specific activity profiles within the trait-associated regions of each GWAS study and trait. To that end, for each of the GWAS studies and each of the AE sets with specific activity profiles, I followed the step-by-step process described below **(Fig. 2.4.).**

1. Define the GWAS study of interest.
2. Define the specific enhancer activity profile of interest.
3. Identify the list of AE bins exhibiting this specific enhancer activity profile as the "test set."
4. Identify the list of AE bins with enhancer activity profiles different from the specific one under investigation. Designate this list as the "background set."
5. Conduct a Fisher test to assess the significance of the association between the identified AE bins (both test and background sets) and trait-associated regions.
   - Null Hypothesis (H0): There is no enrichment of enhancers with the activity profile of interest within the regions associated with the trait of interest.
   - Alternative Hypothesis (H1): There is a significant enrichment of enhancers with the activity profile of interest within the regions associated with the trait of interest.
6. Determine the number of overlaps between AE bins in both the test set and background set with the trait-associated regions.
7. Calculate raw p-values from the Fisher test.
8. Correct for multiple testing using the Bonferroni method to control the familywise error rate (adjusted p-value < 0.0001).

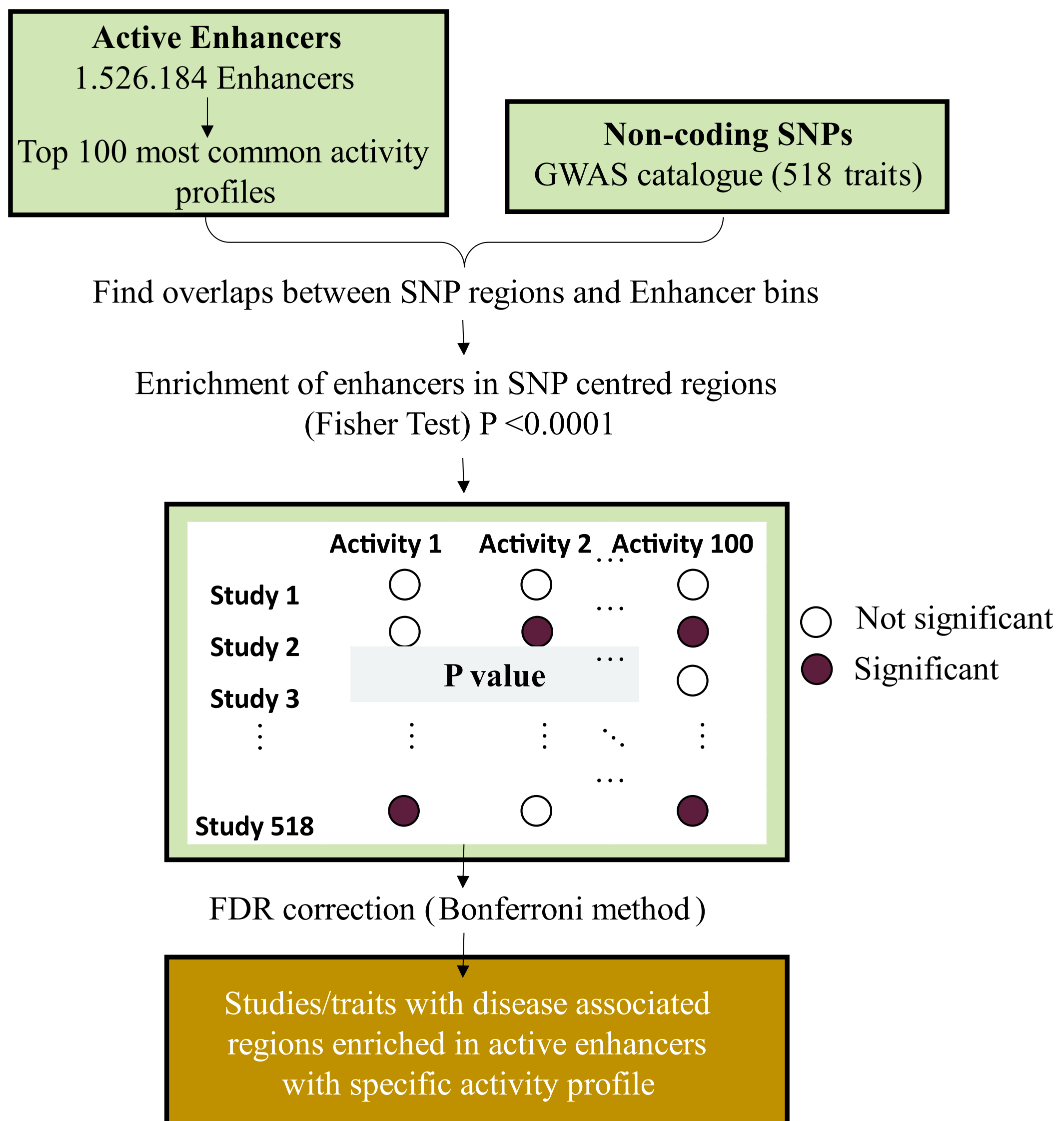


Figure 2. 4. GWAS enrichment analysis. The flow diagram illustrates the steps of the methodological approach employed to identify trait-associated regions enriched with active enhancer bins having activity profiles among the top 100 most common in our dataset.

2.7. Identifying target genes of enhancers associated with GWAS traits.

Once the significant enrichment of enhancers with specific activity profiles was identified for a certain trait following the methodology described in **section 2.6.**, I identified sets of genes whose expression could be regulated by trait-associated enhancers. The general process outlined below was used to identify target genes of macrophage enhancers associated with cardiovascular disease risk **(Chapter 5.)** and of neutrophil enhancers associated with Ulcerative Colitis **(section 4.2.5)**

1. I selected AE bins within the trait-associated regions exhibiting that activity profile.

2. I used eQTL data, which as mentioned in **Chapter 1. 1.5. Enhancer maps and enhancer-target gene links** is useful to identify target genes potentially regulated by those enhancer regions. Briefly,

    a. eQTL data assayed on relevant cell types/tissues matching the activity profiles of the enhancers, was obtained from the Genotype-Tissue Expression project (GTEx) (Lonsdale et al., 2013) and data available in the eQTL Catalogue (Kerimov et al., 2021a).

    b. I examined if the trait-associated SNPs in the enhancer regions also functioned as eQTLs in those tissues.

3. I assessed whether the expression patterns of the genes with expression levels correlated with the eQTLs (eGenes) matched the activity profiles of the enhancers. Briefly,

    a. I used RNA-seq data generated by the BLUEPRINT project to evaluate the mean expression of the eGenes across the 31 cell types in our dataset.

    b. I consulted the Human Protein Atlas database (Thul & Lindskog, 2018) to gain insights into the predominant cell types and tissues where these genes are active.

2.8. Software used for data analysis.

All scripts generated to perform the analysis described in this dissertation were written and ex ecuted in R version 3.6.3 (2020-02-29) (R Core Team, 2020). The scripts are available in a Git Hub repository (https://github.com/juearcilaga/Unmasking-disease-risk-in-haematopoietic-enhancers), and supplementary files to run the codes are available on Figshare (https://doi.org/10.6084/m9.figshare.25601301.v1). A detailed list of the R packages utilised is presented below i n the format (package, version, reference)

- dplyr 1.0.8 (Wickham et al., 2021)
- edgeR 3.28.1 (MD et al., 2010)
- forcats 0.5.1 (Wickham, 2022)
- ggpubr 0.4.0 (Kassambara, 2020)
- gridExtra 2.3 (Murrell, 2017)
- gtools 3.9.4 (Warnes, 2021)
- gwascat 2.18.0 (Magno & Corty, 2022)
- pheatmap 1.0.12 (Kolde, 2019)
- rtracklayer 1.46.0 (Lawrence, 2022)
- scales 1.2.1 (Wickham, 2020)
- superheat 1.0.0 (Tierney, 2019)
- tidyverse 1.3.1 (Wickham & Henry, 2021)
- treemap 2.4-3 (Tennekes, 2020)
- biomaRt 2.42.1 (Durinck & Huber, 2021)
- GenomicRanges 1.38.0 (Lawrence & Obenchain, 2022)
- ggplot2 3.3.6 (Villanueva et al., 2016)
- GO.db 3.10.0 (Bioconductor, 2021)
- plyr 1.8.6 (Package & Wickham, 2021)
- plyranges 1.6.10 (Shepherd & Lawrence, 2022)
- RColorBrewer 1.1-3 (Neuwirth, 2014)

Chapter 3. A resource of reference epigenomes of human haematopoietic cells

My main interest is to understand the role of enhancers in complex diseases, and for that, I need cell type-specific epigenomic maps. As the onset of these diseases is profoundly shaped by environmental factors, and immune cells are the first sensors and main responders to environmental stimuli in the human body, I hypothesise that they are an ideal set to focus on. Leveraging the BLUEPRINT Consortium's collection of profiles of histone marks for human primary haematopoietic cells, we trained a chromatin state model to learn the patterns of histone mark combinations across the genome and associate them with functional transcriptional states. With this model, we have generated the largest and most comprehensive collection of haematopoietic epigenomes to date. Additionally, we described the enhancer repertory of haematopoietic cells, characterised enhancers' activity profiles, and identified those that were more commonly observed in the dataset.

3.1. Introduction.

The human haematopoietic cells are descendants of a common progenitor and play vital roles in diverse processes such as immune defence, oxygen transport, and blood clotting (Clien et al., 2014; Jenne et al., 2013; Jensen, 2009; Varol et al., 2015). Haematopoietic cell differentiation is accompanied by dynamic changes in chromatin structure (Cedar & Bergman, 2011; Lara-Astiaso et al., 2014), involving stepwise modifications of histone marks at lineage-specific enhancers (Lara-Astiaso et al., 2014) and changes in chromatin interactions between enhancers and promoters (Javierre et al., 2016).

Epigenetic mechanisms not only play crucial roles during haematopoiesis but also orchestrate the rapid and reversible activation of specific genes essential for sterile inflammatory responses, as well as immune cell recruitment and pathogen elimination (Kondilis-Mangum & Wade, 2013; Medzhitov & Horng, 2009; Natoli & Ostuni, 2019; Q. Zhang & Cao, 2019). Histone modifications at enhancers exhibit dynamic changes between primed and activated states, and even *de novo* activation of non-primed enhancers is possible (Ghisletti et al., 2010; Kaikkonen et al., 2013a; Sciumè et al., 2020) demonstrating the immune system's adaptability to diverse challenges. Inadequate regulation of gene expression in haematopoietic cells can lead to or contribute to disease development and progression (Q. Zhang & Cao, 2019).

Following the US-funded efforts of the NIH Epigenomics Roadmap (Bernstein et al., 2010; Roadmap Epigenomics Consortium et al., 2015) and ENCODE consortia (Abascal et al., 2020; The ENCODE Project Consortium 2012), The EU-funded BLUEPRINT Consortium (Adams et al., 2012) generated one of the largest collections of histone mark profiles for human primary cells, including all six histone marks (H3K36me3, H3K4me1, H3K27ac, H3K4me3, H3K27me3 and H3K9me3) necessary for defining reference epigenomes according to IHEC guidelines (Stunnenberg et al., 2016b). Although the other consortia contributed data for a broader spectrum of cell types of the human body, they included a reduced number of primary haematopoietic cell types. In contrast, BLUEPRINT generated a comprehensive dataset for most main haematopoietic lineages, comprising cell types at several differentiation steps. The reference epigenomes provided by ENCODE and Roadmap were generated by compiling histone marks from multiple donors, and there are no biological replicates for each cell type (Abascal et al., 2020; Roadmap Epigenomics Consortium et al., 2015). In contrast, BLUEPRINT generated all six histone marks for the same donor for each cell type, and most cell types included multiple donors (between 1 and 12 donors per cell type, median = 3).

While BLUEPRINT Consortium has produced the most comprehensive dataset of histone modification maps for haematopoietic cells, there is currently a lack of comparable chromatin state maps available for comparing the epigenomes of the different haematopoietic cell types. This is because different groups of The BLUEPRINT Consortium have generated independent chromatin state models using partial sets of samples and different computational methodologies. Beekman et al. (2018) generated chromatin states only for the B-cell lineage using ChIP-seq aligned to genome version GRCh38 (Beekman et al., 2018a), while Grassi et al. (2018) generated a different ChromHMM (Ernst & Kellis, 2012) model exclusively for the neutrophil lineage, using GRCh38 too (Grassi et al., 2018). In contrast, Cieslak et al. (2020) segmented the ChIP-seq data of the T-cell lineage aligned to GRCh37 (Cieslak et al., 2020) using the previous ChromHMM model trained by Carrillo et al. 2017 with several cell types (Carrillo-de-Santa-Pau et al., 2017). Additional different ChromHMM models were generated with samples aligned to GRCh37 for some cell types with promoter-capture HiC data (Javierre et al., 2016) and for the annotation of regions of interest in population-based studies using monocytes, neutrophils and CD4 T-cells (Astle et al., 2016a; L. Chen et al., 2016; Ecker et al., 2017). Finally, Petersen et al. (2017) generated chromatin states for megakaryocytes (MKs) and erythroblasts (EBs) (Petersen et al., 2017) using the segmentation method IDEAS (Y. Zhang et al., 2016) instead of ChromHMM. Moreover, the chromatin states of some cell types with the complete set of six histone marks (and input control) have never been generated and described

before. These include natural killer (NK) cells, monocyte-derived mature and immature dendritic cells (DCs), monocyte-derived osteoclasts and eosinophils.

Using the BLUEPRINT Consortium dataset, here we provide the most complete collection of human haematopoietic epigenomes. We trained and applied a chromatin state model for generating 107 chromatin state maps spanning 31 cell types, including the most important mature cell types of the haematopoietic lineage. Each epigenome consists of a whole-genome annotation of 7 chromatin states representing the most relevant functional elements of the genome, including expressed regions and regulatory regions and their subclasses. I assessed the quantity of each type of enhancer, characterised their activity profiles, and identified the most frequently occurring profiles among enhancers.

3.2. Results.

3.2.1. The largest dataset of chromatin state maps for individual primary haematopoietic samples.

We built and applied a chromatin states model for generating epigenomes for 107 human primary human samples from 31 different cell types, including 27 haematopoietic, three endothelial and one bone marrow mesenchymal stem ceThe methodologylogy for building the model is outlined in **sections 2.1 and 2.3**. Samples were obtained by the BLUEPRINT Consortium (Adams et al., 2012) from healthy donors, encompassing both men and women. For detailed specifications on the samples, refer to **Fig. 2.1**.

The chromatin states model was trained on ChIP-Seq data generated by the BLUEPRINT Consortium, consisting of whole-genome maps for six histone marks: H3K36me3, H3K4me1, H3K27ac, H3K4me3, H3K27me3, and H3K9me3 for each sample. A detailed introduction to these histone marks is available in **section 1.3**.

The generated epigenomes encompass annotations for each 200 base pair segment across the entire genome (bin), classifying them into seven distinct chromatin states based on their transcriptional functions. These chromatin states comprise Polycomb, Heterochromatin (HetChrom), Transcriptional Elongation (TransElo), Active Elongating Enhancer (AleloE), Poised Enhancer (PoiseE), Conventional Active Enhancer (AconE), and Active Promoter (Aprom). A detailed introduction to these epigenomic elements is available in **section 1.3**.

As a first step in the characterisation of the 107 epigenomes, I calculated the proportion of the genome annotated with each chromatin state (**Fig 3.1.A**). As expected, I found that the HetChrom state covers most of the epigenome of all samples, followed by TransElo, while Polycomb, Aprom, and the three enhancer sub-groups (AconE, AeloE and PoisE) cover a small proportion of the epigenomes **(Fig. 3.1.A)**. When focusing on the three types of enhancers, PoisE, AconE and AeloE (see **section 1.3** for a detailed description of these enhancer types), I found that PoisE (H3K4me1 without H3K27ac) is the most abundant type of enhancer in most of the samples, followed by AconE (H3K4me1 plus H3K27ac) **(Fig. 3.1.B)**

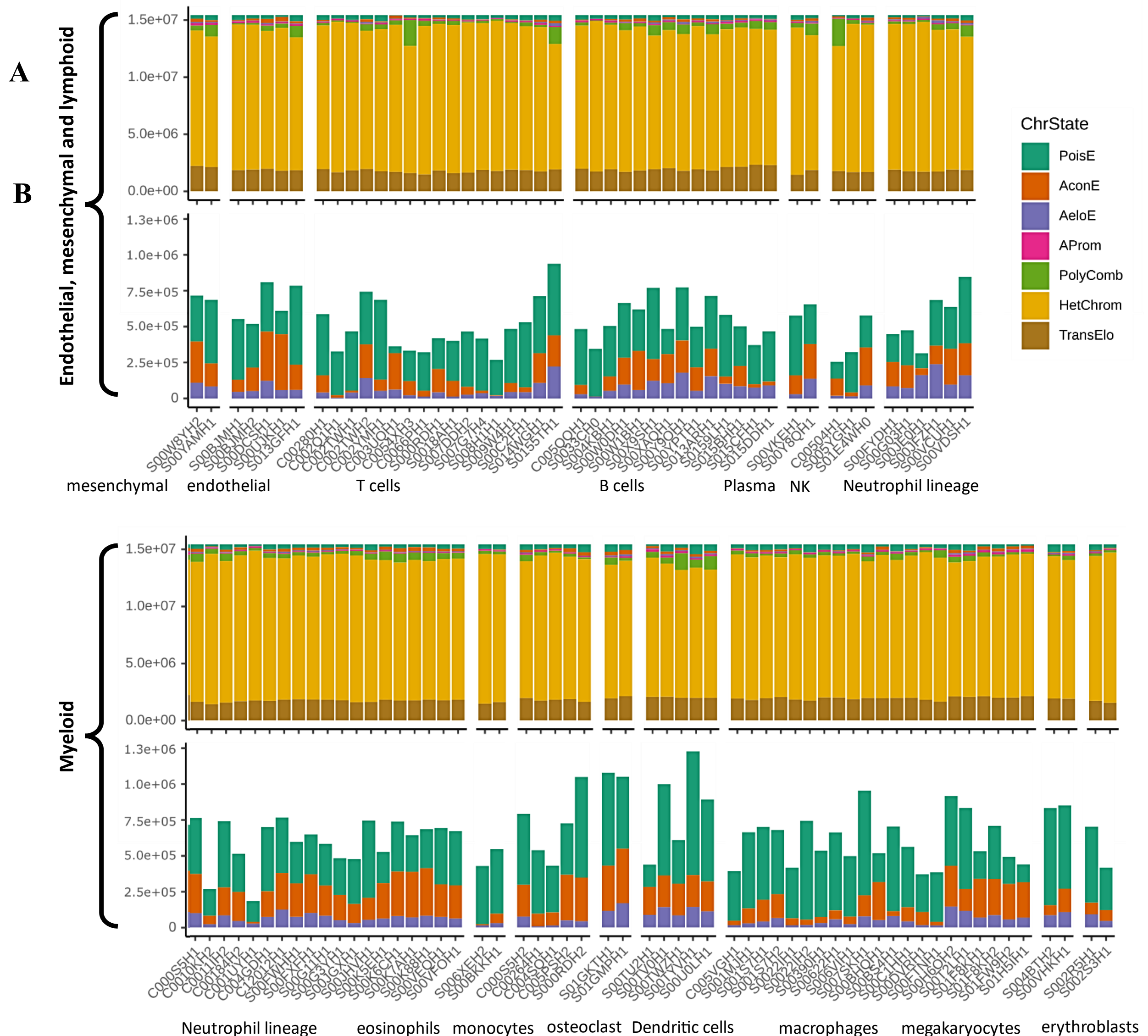


Figure 3. 1. **Epigenomic profile of haematopoietic cell types from the BLUEPRINT Consortium.** For each of the 107 haematopoietic epigenomes generated by us, the bar plots represent: The proportion of the epigenome length covered by each chromatin state (Top panel). The number of bins (1 bin = 200bp) annotated as each of the three types of enhancers (Bottom panel).

3.2.2. A collection of consensus epigenomes for haematopoietic cell types.

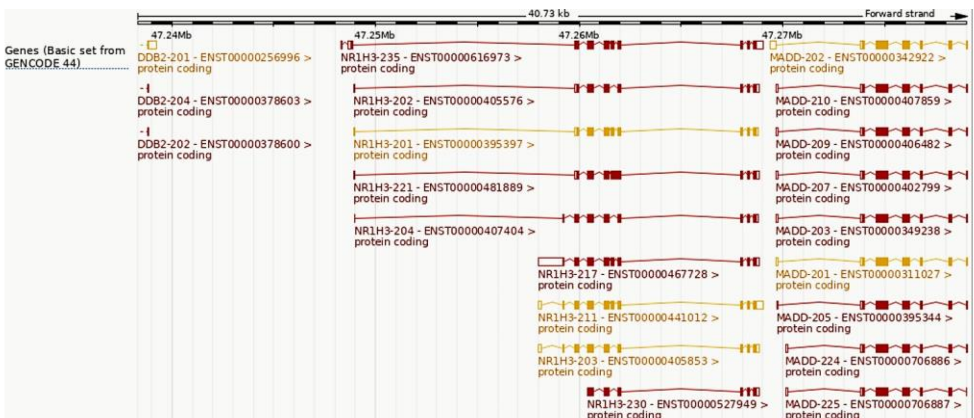
To facilitate the comparison of epigenomes across cell types, I generated a consensus epigenome per cell type for those cell types with at least two replicates (27 out of the 31 cell types). This was achieved by preserving states consistently annotated in the replicates; see the full method in **section 2.4**. In addition to the seven chromatin states defined in **section 3.2.1**., I introduced a new state called 'non-conserved state' for bins that lacked consistent annotation across replicates (see **Fig. 2.3.A**).

When comparing the entire consensus epigenome across different cell types, I observed that 46% of the bins across the genome have been consistently labelled with the same chromatin state across all cell types. Among these bins with conserved state, 93% are identified as HetChrom, comprising 43% of the genome. This indicates that 64% of the genome exhibits variable epigenomic annotations across the studied haematopoietic cell type. Using these consensus epigenomes, it is possible to analyse for any genomic region of interest, the dynamics of the epigenome across different cell types at a scale of 200bp. To illustrate the utility of this data, I present the epigenomic landscape of the *NR1H3* **(Fig. 3.2.) and** *ARID5B* **(Fig. A2**) genes.

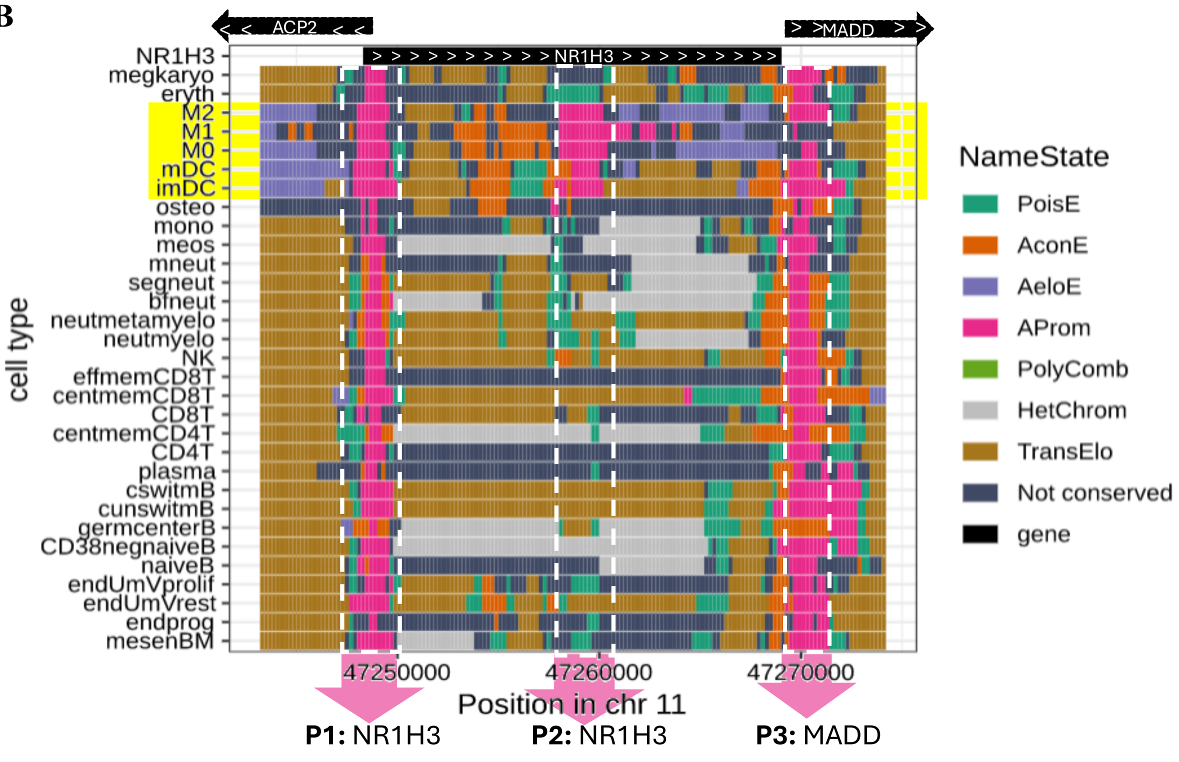
NR1H3 encodes a transcription factor known as liver X receptor alpha, or LXRα, which plays a critical role in coordinating lipid metabolism and immune responses (Repa & Mangelsdorf, 2000). The 5' region of *NR1H3* is proximal to the 3' region of *ACP2*, which codes for an enzyme involved in the degradation of glycolipids, glycoproteins, and sphingolipids in the lysosome (Thul & Lindskog, 2018). The *NR1H3* 3' region is proximal to the *MADD* 5' region, which encodes for an adaptor protein that plays a role in the regulation of TNFα-induced apoptosis by interacting with TNFα to activate MAPK pathway in response to apoptotic signals (Thul & Lindskog, 2018).

It can be observed from **Fig. 3.2. B.** that the genomic region overlapping *NR1H3* is in the TransElo state in most BLUEPRINT cell types, suggesting gene expression. I identified three active promoter regions in this epigenomic landscape: Promoter 1 (P1) on the 5' end, Promoter 2 (P2) in the middle, and Promoter 3 (P3) on the 3' end. P1, P2 and P3 overlap with the start of the transcript isoforms of *NR1H3,* and *MADD* (**Fig. 3.2.A)** reported in ENSEMBL release 111 (Martin et al., 2023). Additionally, we can learn from **Fig. 3.2. B** that P1 and P3 exhibit activity across all 31 cell types, whereas P2 is exclusive to mature DCs (mDC), immature DCs (imDC), and the three macrophage types, unstimulated macrophages (M0), anti-inflammatory

A



B



C

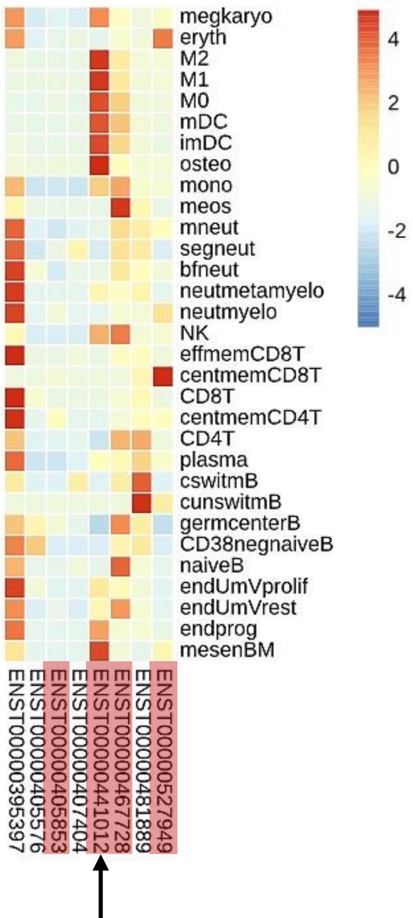


Figure 3. 2. **Using chromatin states to study the epigenomic landscape of *NR1H3*. (A)** Shows the annotation of the *NR1H3* transcripts as reported in ENSEMBL (Martin et al., 2023), based on the annotation from GENCODE release 44 (Frankish et al., 2021). **(B)** *NR1H3* gene region shows a complex expression regulatory pattern, with multiple promoters (P1, P2, P3) annotated as AProm state (pink colour) in different sets of cell types, and the presence of an intronic enhancer (E1) with a specific chromatin activity (AconE state in orange colour) in monocyte descendant cell types. Black arrows indicate the genome locations of *NR1H3* and its two neighbour genes, *ACP2* and *MADD*. Each row represents the conserved chromatin states in each of the 31 haematopoietic cell types. Promoter regions are indicated by dotted rectangles. In yellow, the box indicates the macrophage and dendritic cell subtypes that show activity of the alternative promoter P2. **(C)** P2 region overlaps the TSS of the shorter isoforms (highlighted in red) of *NR1H3*, from these shorter isoforms ENST00000467728 (indicated by a black arrow) is the transcript that shows the higher expression levels in the set of cell types that have P2 active. A zoomable version of the figure is available here.

macrophages (M2) and pro-inflammatory macrophages (M1). Data from the EpiMap repository (https://compbio.mit.edu/epimap/) (Boix et al., 2021) support these findings, where we can confirm that regions overlapping P1, and P3 are annotated as Active TSS state and Flanking TSS in most samples, while P2 displays these annotation only in a subset of the samples **(Fig. A1).**

Consequently, P2 emerges as a distinct alternative promoter for *NR1H3*, specific to macrophages and dendritic cells. P2 region overlaps the TSS of the shorter isoforms of *NR1H3* (**Fig. 3.2.A**). I explored the transcript expression data from BLUEPRINT to analyse the expression patterns of the different transcripts across the haematopoietic cell types and found that one of the shorter isoforms, ENST00000467728 is the transcript that shows the higher expression levels in the set of cell types that have P2 active (**Fig. 3.2.C**).

This information hints at the possibility that the shorter isoforms of *NR1H3* may exhibit specific expression in macrophages, although experimental verification is pending. Furthermore, the potential implications of this expression pattern in the regulation of lipid metabolism within macrophages, in comparison to other cell types, remain an intriguing avenue for further exploration.

Similar to the case of *NRIH3*, by analysing the BLUEPRINT dataset, I found support for two alternative promoters of a macrophage-relevant gene studied by McErlean et al. (2021). McErlean et al. investigated methylation patterns in airway macrophages (AM), which play a role in the pathogenesis of idiopathic pulmonary fibrosis (IPF), a fatal respiratory disease without a cure(McErlean et al., 2021). They discovered that *ARID5B*, a transcriptional cofactor regulating glucose metabolism (Okuno et al., 2013), exhibited differential methylation across 6

CpGs and showed chromatin enhancer signatures, specifically H3K4me1, in myeloid-derived macrophages.

McErlean et al. noted that the short isoform of *ARID5B* was more highly methylated in AMs from healthy controls, less in monocyte-derived macrophages, and displayed an intermediate methylation level in IPF AMs (McErlean et al., 2021). Additionally, they found that this gene was expressed at higher levels in M1 macrophages compared to M0 and M2, and its expression was elevated in IPF patients compared to healthy donors.

In my exploration of the epigenomic landscape surrounding *ARID5B*, I found evidence supporting the presence of an active promoter for the short isoform in M1 (**Fig. A1**). This promoter sequence is longer than those found in M2 and M0. No active promoter annotation was found in monocytes or other hematopoietic cells in my dataset, except in endothelial, mesenchymal, and class-switched B cells. Furthermore, the same region is annotated as a poised enhancer in some hematopoietic cells and as an active enhancer in others (**Fig. A1**). These findings align with the gene expression patterns described in McErlean et al. (2021), where the cell types with the longest promoter display higher expression levels (McErlean et al., 2021).

3.2.3. Enhancer maps for haematopoietic cells

I identified 1,526,184 total AE bins (AE =AeloE +AconE) in our dataset using the individual epigenomes described in **section 3.2.1.** and following the criteria described in **section 2.5**. This total count of AEs encompasses active enhancers operating in two distinct scenarios. Firstly, there are those conserved in at least 1 out of the 27 consensus epigenomes, constituting

1,128,959 AE bins, equivalent to 74% of all AE. Secondly, there are enhancers present solely in specific individual epigenomes from the dataset, totalling 397,225 AE bins, which represents 26% of all AE.

These results suggest that the majority of identified AE bins (74%) consistently activated in one or more cell types could be lineage-specific enhancers (Lara-Astiaso et al., 2014). However, it is also noteworthy that a subset of enhancers (26%) may function as stimuli-responsive elements, operating in a context-dependent fashion, for example, in response to environmental changes or pathogens (Ghisletti et al., 2010; Kaikkonen et al., 2013a; Sciumè et al., 2020).

To provide further information on the regulatory landscape of different cell types, I analysed the total number of AE bins per cell type. I found that the cell type with the largest number of

AE bins is germinal centre B-cells (germcenterB, 313,787 bins), followed by imDCs (285,986 bins) and neutrophilic metamyelocytes (neutmetmyelo, 252,185 bins) **(Fig. 3.3.A).** Interestingly, CD4+ T-cells (CD4T, 15,268 bins) and mature eosinophils (meos, 17,626) had a much lower number of active enhancer regions than any other cell type in our dataset.

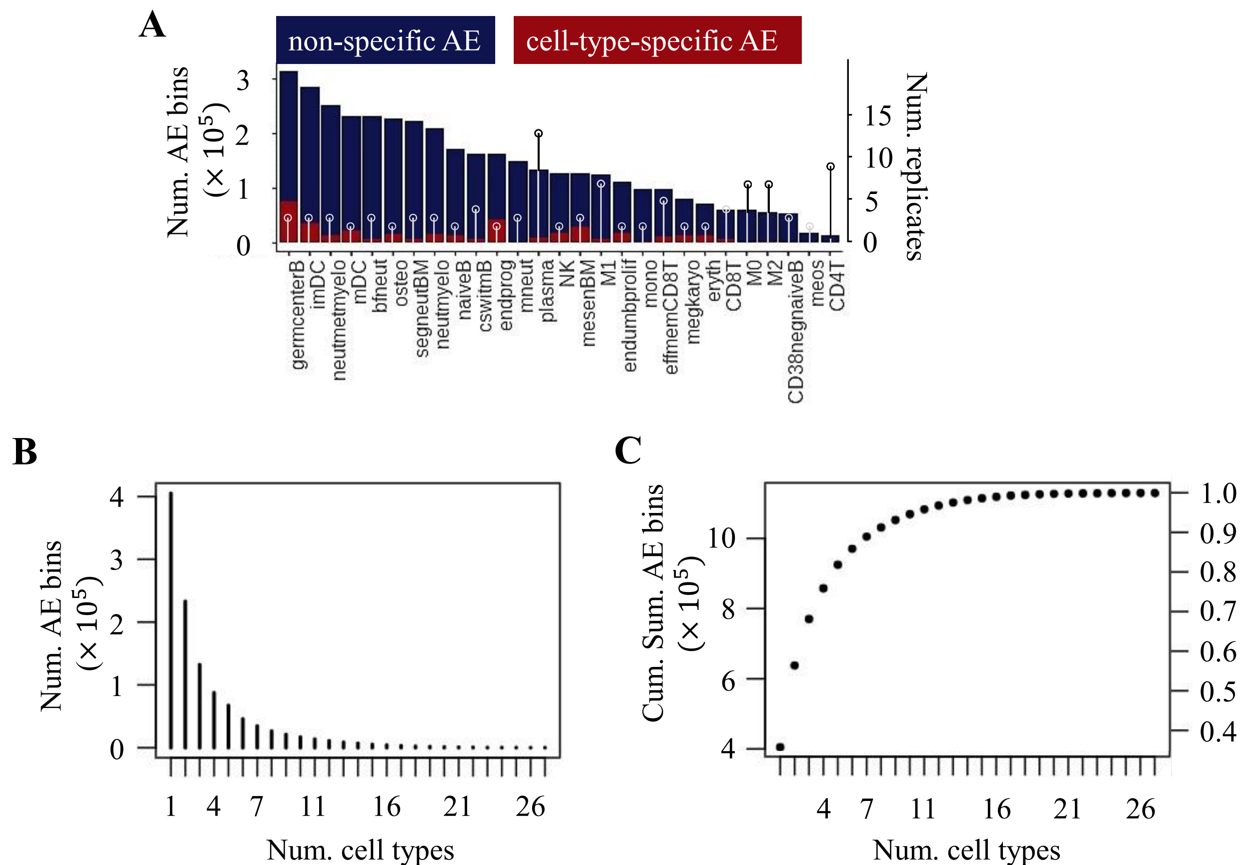


Figure 3. 3. **Number and specificity of consensus AE per cell type. (A)** Bar plot representing the number of AE bins (left y-axis) in each cell type consensus epigenome. The red colour represents cell type-specific enhancers, and the blue colour represents enhancers shared with other cell types. Lollipops represent the number of biological replicates (right y-axis) for each cell type (**B**) The frequency plot shows in the y-axis the number of AE bins that are unique to a single cell type x=1 or shared among two or more cell types, 2≤x≤27. **(C)** The graph represents in the Y axis the cumulative count of AE bins alongside the equivalent percentage from the total consensus AE that are unique to a single cell type x=1 or shared among two or more cell types, 2≤x≤27.

To assess the potential influence of variability in the number of replicates on the number of consensus enhancer regions detected, I evaluated the correlations between the number of consensus enhancer bins by cell type, and number of replicates per cell type. I found a weak correlation between these two variables (Pearson coefficient = 0.3), but a strong correlation

between the number of consensus enhancers and the mean number of enhancers per cell type (Pearson coefficient = 0.82). This suggests that the reduced number of enhancers in eosinophils and CD4+ T cells is not solely attributable to variability in the number of replicates.

3.2.4. Enhancer activity profiles across haematopoietic cell types.

Chromatin states are cell type-specific, which means that the same region of the genome can be in an active enhancer state in a specific cell type or group of cell types while having a completely different epigenomic annotation in the other cell types (Carrillo-de-Santa-Pau et al., 2017).

For instance, for the *NR1H3* gene shown in **Fig 3.2.B**, there is an enhancer region (E1) located 5´ upstream of P1, with an AeloE state (shown in purple) in M0, M1, M2, imDC and mDC cell types. There is a second enhancer (E2) in the AconE state (shown in orange) located between P1 and P2 that is specific to monocyte descendant cell types and only in M0 and M1 extends its length until P2. And there is a third enhancer (E3) in the AeloE state that is active only in macrophages (M0, M1, M2) (**Fig 3.2.B**).

Keeping this in mind, I defined chromatin activity profiles for active enhancer regions across the genome. These profiles are identified as the set of cell types in whose consensus epigenome the region has a conserved active enhancer state - refer to **section 2.4** and **Fig. 2.3.B** for methods.

Using this definition, the chromatin activity profiles of E1, E2, and E3 in the example described above would be M0|M1|M2|imDC|DC, M0|M1|M2|imDC|DC|oseo, and M0|M1|M2, respectively. However, to ensure comparability, since these regions vary in size across cell types, the activity profile definition is given for each 200 base pairs.

After distinguishing the consensus AE bins per cell type based on their activity profiles, I observed that in each cell type, only a small percentage of consensus AEs are cell type-specific, and most consensus AE are shared between two or more cell types (**Fig. 3.3.A**). For example, 76% of germcenterB AEs are shared with at least another cell type. On average, only 10% of active enhancer regions are specific to a single cell type (**Fig. 3.3.A**). Yet, looking at it collectively, when I sum the number of cell-type-specific AE bins across all cell types, it constitutes 40% of the overall count of haematopoietic enhancers (**Fig. 3.3.A-B**).

I assessed the number of distinct activity profiles (combinations of cell types) observed in the total consensus AEs and identified 107,562 unique combinations. Assessing the distribution of the frequency of each unique activity profile across all AE bins, I observed frequencies ranging from 1 to 75,814 bins, and the distribution was heavily left-skewed, with a mean of 10.5 bins (Q1 = 1 bin, Q3 = 2 bins) **(Fig. 3.4.A)**. Remarkably, including enhancers in the first 1,856 most frequent activity profiles is sufficient to cover 80% of consensus AE bins **(Fig. 3.4.B)**.

The four most frequently observed activity profiles across the AE bins were cell type-specific for germcenterB (75814 bins), endothelial progenitors (endprog, 43080 bins), immature DCs (imDC, 36598 bins) and mesenchymal cells (mesenBM, 30462 bins) **(Fig.3.4.C)**. Unsurprisingly, none of the top 100 most commonly observed enhancer activity profile patterns included CD4+ T-cells (514 bins) or mature eosinophils (514 bins), as they, in general, have a very low number of active enhancer regions compared to the other studied cell types **(Fig. 3.3.A).**

Among the top 100 most common activity profiles, the one including the maximum number of cell types corresponds to enhancers shared by seven myeloid cell types: monocytes (mono), osteoclasts (osteo), mature neutrophils (mneut) and the four neutrophil progenitors (segneutBM, neutmyelo, bfneut and neutmyelo) and there are 1025 genomic bins with this profile. I also found 2045 bins with enhancer activity in the five stages of B-cell differentiation, as well as 1228 bins with enhancer activity among natural killer (NK) cells, CD8+ (CD8T) and CD8+ effector memory T-cells (effmemCD8T) **(Fig. 3.4.C)**.

In general, enhancer activity profiles, including two or more haematopoietic cell types, tended to group into lymphoid or myeloid cells. The only exceptions to this general pattern correspond to sets of enhancers active in both germcenterB cells and some myeloid cells (some neutrophil progenitors, monocyte-derived DCs, osteoclasts, megakaryocytes and erythroblasts). This could be partially explained by the fact that germcenterB cells have the consensus epigenome with the highest number of active enhancer regions.

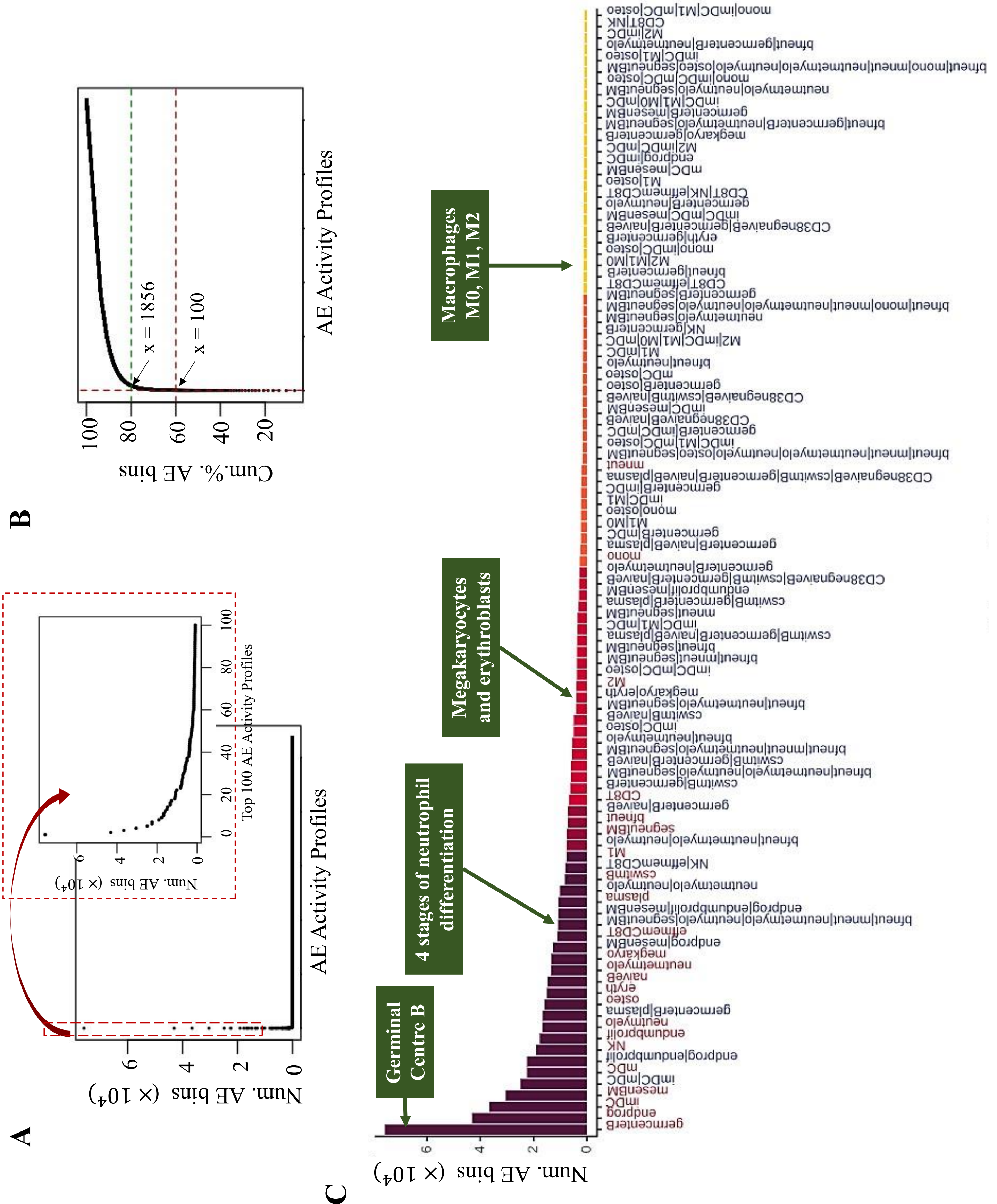


Figure 3. 4. **Enhancer Activity profiles in haematopoietic cells.** **(A)** The graph illustrates the frequency distribution of activity profiles within the entire set of active enhancers (AEs) identified across haematopoietic cells. The distribution is notably left-skewed, with the number of AE bins associated with each activity profile ranging from 1 to 75,814. The lower quantile

is 1, and the upper quantile is 2. The enclosed red dotted rectangle delineates the top 100 most frequent profiles, offering a closer view of their distribution. **(B)** The graph showcases the cumulative frequency of activity profiles, emphasising that the foremost 100 profiles, among the more than 100,000 total profiles, collectively represent 60% of the total AE bins. Intriguingly, a mere 1,856 profiles are required to encapsulate 80% of the total active enhancers. This highlights the concentrated nature of the most prevalent activity profiles and underscores their significant contribution to the overall enhancer landscape. **(C)** Top 100 most frequently observed active enhancer activity profiles.

3.3. Discussion.

The BLUEPRINT Consortium's epigenomic dataset has been pivotal in understanding epigenetics in haematopoietic differentiation (Cieslak et al., 2020; Grassi et al., 2018), differences between mature cell types (L. Chen et al., 2016; Javierre et al., 2016; Petersen et al., 2017), response to environmental stimuli (Ecker et al., 2017) and in complex diseases (Astle et al., 2016a; Beekman et al., 2018a). However, a gap in knowledge existed due to previous chromatin state analyses focusing on subsets of BLUEPRINT data and utilising different models, leading to non-comparable results.

Addressing prior limitations, I generated a comprehensive collection of epigenomic landscapes of human haematopoietic cells by training a chromatin states model with the entire BLUEPRINT dataset. This dataset consists of chromatin state maps for 107 individual primary haematopoietic samples, encompassing 31 cell types, with up to 12 replicates per cell type.

The chromatin state maps generated in this PhD project facilitate in-depth analysis of the epigenomic landscape of haematopoietic cells, providing accurate (200bp) localisation of repressed regions, heterochromatin, active genes, and promoters, as well as active and poised enhancers across the genome. To enable a standardised comparison of epigenomic features across diverse haematopoietic cell types, I also generated consensus epigenomes per cell type, for those with at least two biological replicates. This dataset offers an invaluable resource for researchers delving into the role of gene regulation in cell differentiation and disease.

Furthermore, there are different sets of cell types in our dataset whose chromatin state-based epigenomes have not been published yet and would be interesting to study—for example, eosinophils and erythroblasts, cord and adult endothelial cells and mesenchymal cells. In addition to studying the difference in the epigenomes between cell types, differences in epigenomes of the same cell type by sex or by sampled tissue could be studied, given the

availability of samples from men and women and some cell types being sampled from different tissues.

In this study, I observed a reduced number of enhancers in eosinophils and CD4+ T cells compared to other hematopoietic cell types. This observation could be attributed to both technical and biological factors. Technical challenges associated with performing ChIP-seq on these specific cell types might limit the accurate detection of enhancers. Eosinophils are relatively rare in peripheral blood (Kim & Jung, 2020), making their isolation and enrichment more challenging. This rarity could lead to lower input material for ChIP-seq and potentially hinder enhancer identification. Unfortunately, the BLUEPRINT Consortium did not make this information available.

One biological reason for these results could be the activation state of these cell types in blood. Eosinophils in blood are not typically activated by external stimuli stimuli (Gurtner et al., 2023; Barnig et al., 2015), which could be accompanied by a more quiescent chromatin state with fewer active enhancers. This idea is further supported by the fact that eosinophils at different activation states exhibit distinct transcriptional profiles (Gurtner et al., 2023).

Similarly, most circulating CD4+ T cells are in a resting state, meaning they are not actively proliferating or responding to antigens (Subbannayya et al., 2021). These resting T cells have a distinct epigenomic profile characterised by limited expression of gene sets associated with more differentiated states (Rose et al., 2023). Additionally, CD4+ T cells are known for their high heterogeneity due to various activation states and subsets (Subbannayya et al., 2021; Rose et al., 2023), which might contribute to the variability in active enhancer sets among replicates and a reduced number of consensus enhancers.

Lastly, it is important to note that the reported number of enhancer regions in this study does not necessarily reflect the count of distinct functional enhancer units. Instead, they indicate the overall extension of the epigenome covered by enhancers. A higher count of AE bins could imply either a greater number of individual enhancers or larger enhancer regions within the cell type's epigenome. Future analysis would be necessary to identify which of the two scenarios prevail. For instance, we could segment the genome into 5Kb windows and assess enhancer density within these windows. This approach could reveal whether enhancers are predominantly clustered in a few large regions or scattered across multiple smaller ones.

Quantifying AE bins revealed that a substantial majority (74%) are consensus AE in at least one cell type, indicating lineage-associated rather than stimuli-responsive enhancers (26%), which may be operating in a context-dependent manner. I also observed that from the set of

enhancers active in each cell type, on average, only 10% are cell type-specific. This low percentage of difference might be attributed to their close evolutionary relationship, as they share a common ancestor and embryological origin in the lateral plate mesoderm (Menegatti et al., 2019), therefore highlighting for future studies the importance of considering that defects in enhancers, even if discovered in a single haematopoietic cell type, may be affecting additional haematopoietic cell types as well.

While it is simpler to assess the impact of cell type-specific enhancers, shared elements present a more complex scenario. Investigating whether these enhancers act on the same target genes in different cell types would be invaluable. This idea is supported by Javierre et al., who demonstrated consistent connections between enhancers and gene promoters across haematopoietic lineages. However, experimental validation is essential to confirm these connections (Javierre et al., 2016).

Our contribution to this scientific question was performing detailed enhancer dynamics analysis and providing a comprehensive annotation for each active AE bin in the dataset, specifying the cell types where it exhibits an active chromatin state. I also examined the distribution of unique activity profiles across enhancers, highlighting that a relatively small subset covers a significant proportion (80%) of consensus AE bins. My analysis of enhancer dynamics establishes a solid groundwork for subsequent exploration and experimentation.

Chapter 4. Linking haematopoietic enhancers to complex disease genetic risk.

A vast collection of associations between genetic variants and complex diseases is available today. However, interpreting them is challenging due to the implication of non-coding regions in most cases. Only a handful of them have been experimentally validated. In an effort to contribute to more testable hypotheses, I integrated data from approximately 900 GWAS studies with our previously generated catalogue of haematopoietic enhancer annotations. I identified enhancer sets with specific activity profiles enriched in the disease-associated loci of 300 complex traits. These results provide a valuable collection of disease-enhancer-cell type associations that have the potential to guide the design of future experimental validations and contribute to the understanding of complex disease mechanisms.

4.1. Introduction.

The development of complex diseases depends on a combination of genetic, environmental, and lifestyle factors. (Jonkers & Wijmenga, 2017; Smith et al., 2005). Thanks to GWAS studies, many genomic loci have been associated with complex diseases, and it has been discovered that susceptibility to complex traits generally involves multiple genomic regions, each contributing a small effect (Visscher et al., 2017). A large collection of GWAS results has been compiled through the years and is available in prominent databases (Buniello et al., 2019; Watanabe et al., 2019).

The majority of the identified disease-associated variants are non-coding (Claringbould & Zaugg, 2021) with potential regulatory functions (Farh et al., 2015; Maurano et al., 2012; X. Wang & Goldstein, 2020), which complicates the assessment of the impact that genetic variation has in the phenotypes. A handful of these GWAS loci have already been validated as causal variants with direct insight into the underlying disease biology (Alsheikh et al., 2022). These validated non-coding disease variants are mostly within enhancers and at a distance of less than 10 Kb from the GWAS lead variant (Alsheikh et al., 2022).

To perform experimental validations, it is necessary to have information not only on the potentially relevant disease-associated loci but also on the specific cell type that would be relevant for the assays (Calderon et al., 2017). This is because complex diseases predominantly affect specific organs or tissues, as commonly observed through the signs and symptoms presented by patients; for instance, neurodegenerative disorders, cardiovascular diseases,

autoimmune diseases and systolic blood pressure (Hekselman & Yeger-Lotem, 2020). Once cell types are identified, they will likely be excellent candidates for cell type-specific functional investigations, including assays that will guide enhancer-gene linking or identification of active transcription factor–binding sites inside the trait-associated enhancers (Trynka et al., 2012).

A simple mechanism for tissue-specific susceptibility is the exclusive or preferential expression of the causal gene in susceptible tissues - that is, its overexpression relative to its levels in unaffected tissues (Hekselman & Yeger-Lotem, 2020). Variants that influence cell-type-specific gene regulation are indeed major contributors to common disease risk, and these variants tend to be enriched inside enhancer clusters (Corradin et al., 2014; Corradin & Scacheri, 2014; Hnisz et al., 2013; Parker et al., 2013; Whyte et al., 2013). Therefore, examining the activity of the enhancers enriched in GWAS variants is a common methodology used to identify relevant tissues for the disease (Boix et al., 2021; Dimas et al., 2009; Ernst et al., 2011; Markunas et al., 2017; Maurano et al., 2012; J. Nasser et al., 2021; Roadmap Epigenomics Consortium et al., 2015; Ward & Kellis, 2012).

All of these findings highlight the benefits of integrating enhancer maps with GWAS data to identify disease-relevant enhancers and employ their activity profiles to identify disease-relevant cell types. Therefore, I would leverage the unexplored haematopoietic enhancers map generated in Chapter 3, which includes activity profiles across 31 different cell types, to interpret the functional effect of non-coding variants associated with five hundred traits from the GWAS Catalog. I will present the results of an enrichment analysis of enhancers with specific activity profiles across the trait-associated GWAS loci, revealing several potential disease-relevant enhancers and cell types that can guide future experimental validation.

4.2. Results.

4.2.1. Prioritising high-impact enhancers by focusing on the most common haematopoietic activity profiles.

To understand how genetic variants within enhancers affect different cell types and their potential involvement in diseases, I will analyse the enrichment of AE bins with specific activity profiles in genomic regions associated with complex traits.

In the previous chapter (**section 3.2.4**), I discovered more than a hundred thousand unique activity profiles for consensus haematopoietic AE bins, and testing such a large number of groups is unfeasible in terms of computational time and resources. However, I also found that

by including enhancers in the first 1,856 most frequent activity profiles, I could cover 80% of consensus AE bins, which is still a significant number but approximately 55 times smaller. Moreover, I found that by including just the top 100 most common profiles, I include 60% of the total consensus, which is a significant proportion. Therefore, I narrowed down the focus to the top 100 most prevalent activity profiles within the AE set, which offer a representative and resource-efficient subset for further analysis.

Each of these top 100 AE activity profiles was observed across the genome in at least 900 AE bins and up to 75,814 bins. These profiles include cell type-specific AEs for most of the haematopoietic cell types in our dataset (23 out of the 27 cell types with consensus epigenomes), excluding only CD38- B-cells (CD38negB), mature eosinophil (meos), M0, and CD4T-specific AEs **(Fig. 4.1).**

4.2.2. GWAS traits are linked to haematopoietic enhancers with specific activity profiles.

In this section, my objective was to establish connections between haematopoietic enhancers and the non-coding genetic architecture of complex traits. I focused on AE bins, where AE represents bins annotated as either AeloE or AconE. Specifically, my attention was directed towards sets of AE bins displaying activity profiles within the top 100 most commonly observed in haematopoietic cells, as described in the previous section (**section 4.2.1**). I systematically assessed the enrichment of each set (100 sets) within 10 Kb windows centred on non-coding SNPs (11,060 SNPs) associated with complex traits (518 traits), 985 studies from the GWAS Catalogue (Buniello et al., 2019) were employed. Enrichment analyses were based on Fisher tests, and p-values were corrected for multiple testing using the Bonferroni method (adjusted p-value < 0.0001). For detailed methodology, please refer to **section 2.6**.

This analysis uncovered 1,113 significant activity profile-trait associations. Specifically, 172 out of 518 tested traits displayed significant associations with at least one and up to 15 different activity profiles. Most traits exhibited connections to numerous enhancer bins, showcasing multiple activity profiles across various cell types—typically up to 8 profiles involving up to 12 cell types **(Fig.4.2.)**.

This analysis sheds light on the intricate relationship between diverse enhancer activity profiles and complex traits, emphasising the widespread and cell type-specific regulatory roles within the non-coding genetic architecture of complex phenotypes.

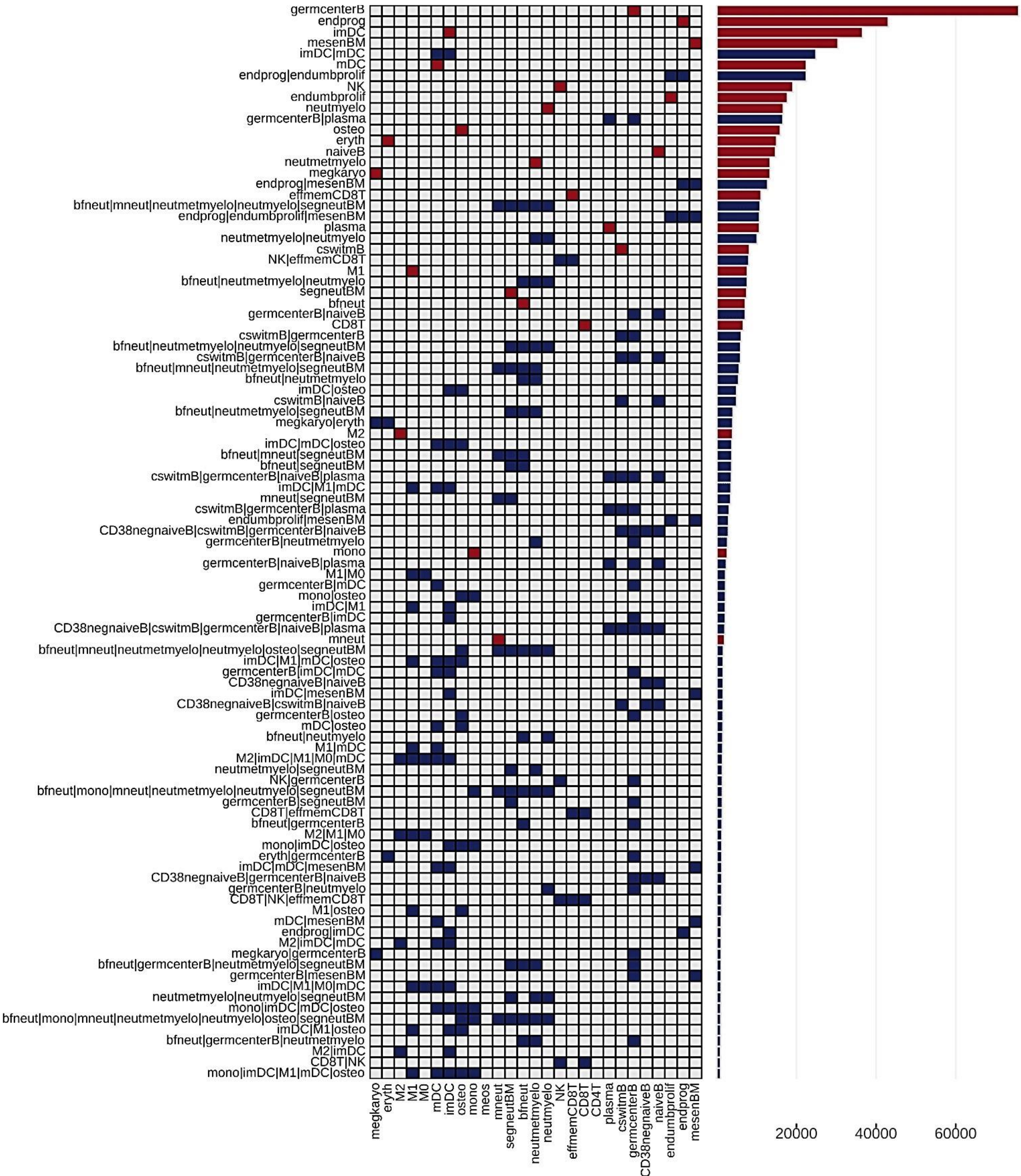

Cell type-specific
Multiple cell types
Activity profile
Cell types in profile
Number of AE bins
germcenterB
endprog
imDC
mesenBM
imDC|mDC
mDC
endprog|endumbprolif
NK
endumbprolif
neutmyelo
germcenterB|plasma
osteo
eryth
naiveB
neutmetmyelo
megkaryo
endprog|mesenBM
effmemCD8T
bfneut|mneut|neutmetmyelo|neutmyelo|segneutBM
endprog|endumbprolif|mesenBM
plasma
neutmetmyelo|neutmyelo
cswitmB
NK|effmemCD8T
M1
bfneut|neutmetmyelo|neutmyelo
segneutBM
bfneut
germcenterB|naiveB
CD8T
cswitmB|germcenterB
bfneut|neutmetmyelo|neutmyelo|segneutBM
cswitmB|germcenterB|naiveB
bfneut|mneut|neutmetmyelo|segneutBM
bfneut|neutmetmyelo
imDC|osteo
cswitmB|naiveB
bfneut|neutmetmyelo|segneutBM
megkaryo|eryth
M2
imDC|mDC|osteo
bfneut|mneut|segneutBM
bfneut|segneutBM
cswitmB|germcenterB|naiveB|plasma
imDC|M1|mDC
mneut|segneutBM
cswitmB|germcenterB|plasma
endumbprolif|mesenBM
CD38negnaiveB|cswitmB|germcenterB|naiveB
germcenterB|neutmetmyelo
mono
germcenterB|naiveB|plasma
M1|M0
germcenterB|mDC
mono|osteo
imDC|M1
germcenterB|imDC
CD38negnaiveB|cswitmB|germcenterB|naiveB|plasma
mneut
bfneut|mneut|neutmetmyelo|neutmyelo|osteo|segneutBM
imDC|M1|mDC|osteo
germcenterB|imDC|mDC
CD38negnaiveB|naiveB
imDC|mesenBM
CD38negnaiveB|cswitmB|naiveB
germcenterB|osteo
mDC|osteo
bfneut|neutmyelo
M1|mDC
M2|imDC|M1|M0|mDC
neutmetmyelo|segneutBM
NK|germcenterB
bfneut|mono|mneut|neutmetmyelo|neutmyelo|segneutBM
germcenterB|segneutBM
CD8T|effmemCD8T
bfneut|germcenterB
M2|M1|M0
mono|imDC|osteo
eryth|germcenterB
imDC|mDC|mesenBM
CD38negnaiveB|germcenterB|naiveB
germcenterB|neutmyelo
CD8T|NK|effmemCD8T
M1|osteo
mDC|mesenBM
endprog|imDC
M2|imDC|mDC
megkaryo|germcenterB
bfneut|germcenterB|neutmetmyelo|segneutBM
germcenterB|mesenBM
imDC|M1|M0|mDC
neutmetmyelo|neutmyelo|segneutBM
mono|imDC|mDC|osteo
bfneut|mono|mneut|neutmetmyelo|neutmyelo|osteo|segneutBM
imDC|M1|osteo
bfneut|germcenterB|neutmetmyelo
M2|imDC
CD8T|NK
mono|imDC|M1|mDC|osteo
megkaryo
eryth
M2
M1
M0
mDC
imDC
osteo
mono
meos
mneut
segneutBM
bfneut
neutmetmyelo
neutmyelo
NK
effmemCD8T
CD8T
CD4T
plasma
cswitmB
germcenterB
CD38negnaiveB
naiveB
endumbprolif
endprog
mesenBM
20000
40000
60000

Figure 4. 1. The 100 activity profiles most frequently observed in hematopoietic AE. The heatmap displays the cell types involved in each profile, while the bar plots illustrate the number of AE bins in which each profile was identified. Profiles shown in blue represent AE active in multiple cell types, whereas those in red are cell type-specific AE**.**

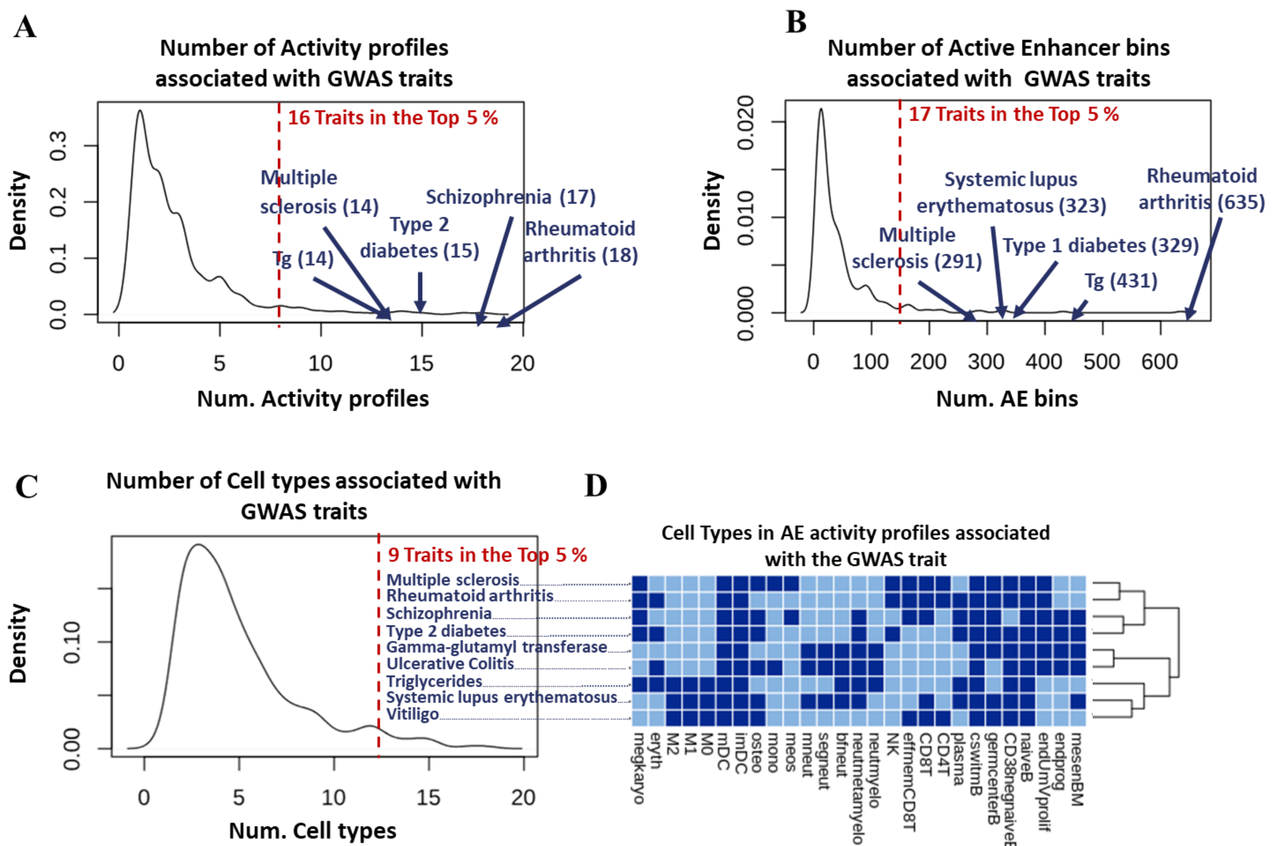


Figure 4. 2. Interplay between genetic susceptibility, enhancer dynamics, and cellular contexts in complex diseases. The figure illustrates that a single trait can be associated with numerous enhancers, each exhibiting unique activity profiles across various cell types. Density plots depict the distribution of different features linked to GWAS traits: **(A)** the number of activity profiles, **(B)** active enhancer bins, and **(C)** cell types. **(D)** Displays the cell types in AE activity profiles associated with the top 5% GWAS traits from the analysis on **panel C**. The red dotted line indicates the density point until 95% of the traits are covered. The number of traits within the upper 10% is highlighted in bold red font, emphasising a subset of 5 traits with the most extreme values, denoted in bold blue font.

4.2.3. Activity profiles of enhancers match with relevant cell types in complex traits aetiology.

With the aim of understanding how distinct cell types may be involved in the genetic regulation of specific complex traits, I analysed the cell types involved in the activity patterns of enhancers

enriched in genomic regions associated with different classes of traits, categorised based on their biological implications. Results are summarised in **Table 4.1**. The group of traits associated with autoimmune diseases and inflammation stands out by displaying the largest median value for AE bins and cell types associated with them.

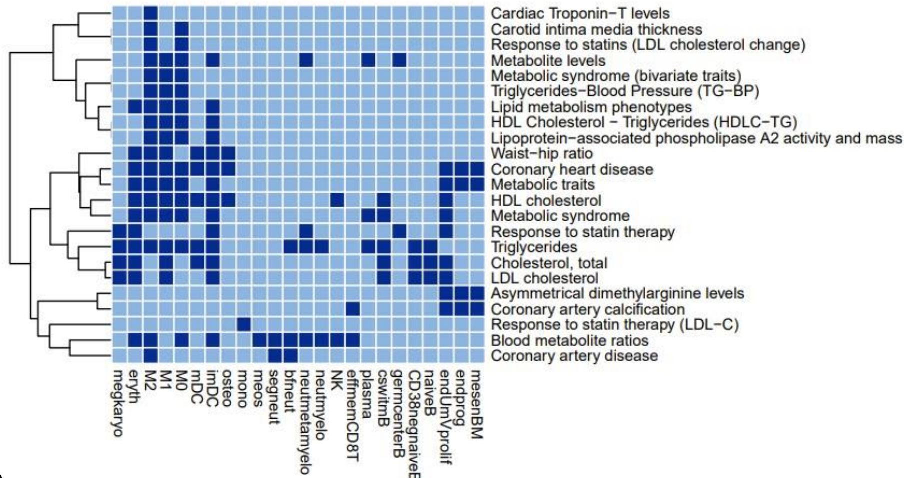
Table 4. 1. Summary of results from the enrichment analysis. The table outlines the characteristics of traits across different groups. The table includes details such as the number of traits with significant results, the number of studies per trait demonstrating significance, the count of activity profiles exhibiting substantial enrichment, the number of enhancers within those profiles overlapping with loci associated with traits, and the total count of cell types involved in the activity profiles significantly associated with traits.

Trait group	Num. GWAS traits	GWAS studies*	Activity profiles*	AE bins*	Cell types*
Anthropometric measures	25	1, 5, 1	1, 9, 2	5, 89, 19	2, 12, 3
Biochemical measures	22	1, 3, 1	1, 7, 2	8, 93, 30	2, 15, 4
Cancer and related traits	39	1, 8, 1	1, 10, 2	5, 193, 18	1, 12, 4
Cardiovascular/Glycaemic	25	1, 5, 1	1, 8, 2	5, 175, 33	2, 11, 4
Neuro-cognitive traits	48	1, 8, 1	1, 17, 1	4, 162, 15	2, 15, 3
Eye related traits	13	1, 2, 1	1, 4, 2	7, 52, 20	2, 10, 4
Haematological traits	18	1, 5, 1	1, 5, 2	6, 169, 44	2, 12, 4
Immune and Inflammatory	30	1, 12, 1	1, 18, 3	4, 635, 56	2, 18, 7
Lipids related traits	16	1, 2, 1	1, 5, 2	8, 132, 22	2, 7, 4
Cardiovascular Risk traits	23	1, 11, 2	1, 14, 4	5, 431, 48	2, 14, 6
Other traits	20	1, 3, 1	1, 5, 2	5, 69, 22	2, 13, 3
Other Medical Conditions	54	1, 7, 1	1, 10, 2	6, 162, 18	1, 13, 4
Substance use/dependence	7	1, 3, 1	1, 3, 1	3, 40, 13	1, 5, 2

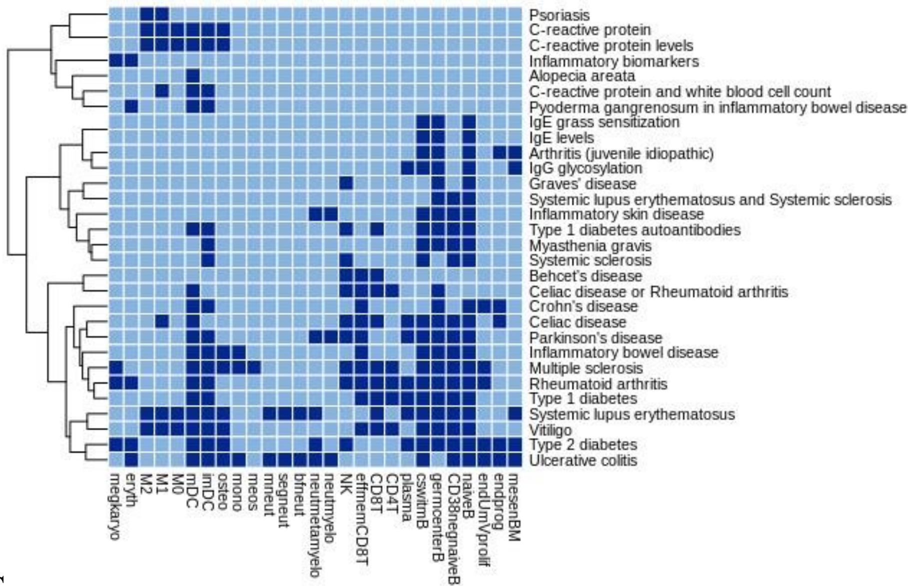
Abbreviations: number (num.). * values separated by a comma represent the minimum, maximum and median of the number of associated entities.

Figure 4.3 illustrates three representative trait groups—cardiovascular disease, cancer, and autoimmune diseases—depicting the cell types where enhancers associated with each trait are active. In cardiovascular disease, macrophages and DCs play a predominant role, with additional involvement of endothelial and mesenchymal cells **(Fig. 4.3.A).** Various stages of B cell differentiation, DCs, and effector memory CD8 T cells (effmemCD8T) exhibit a prevalent role in autoimmune conditions **(Fig. 4.3.B).** Interestingly, traits associated with inflammatory bowel diseases, such as ulcerative colitis and Crohn's disease, which are highly related conditions, do not exhibit clustering based on enhancer activity profiles **(Fig. 4.3.B).**

A **Enhancer-linked cell types in cardiovascular traits**



B **Enhancer-linked cell types in inflammation traits**



C **Enhancer-linked cell types in cancer traits**

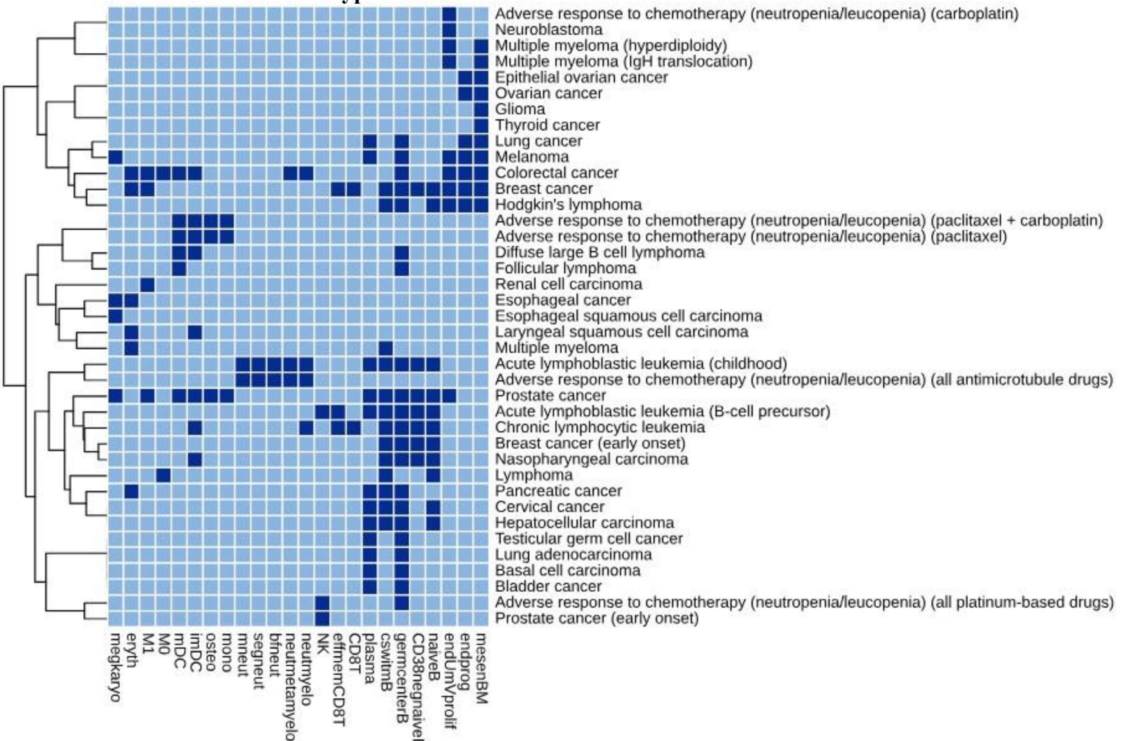


Figure 4. 3. Example of cell types associated with three groups of complex diseases. The heatmap illustrates the various cell types participating in the activity profiles of AE associated with groups of GWAS traits related to specific pathological conditions, including **(A)** cardiovascular disease, **(B)** cancer, and **(C)** autoimmune diseases. A zoomable version of the figure is available here.

A substantial group of traits such as prostate, testicular germ cell, breast, cervical, pancreatic and bladder cancer and some leukaemias and lymphomas indicate the involvement of different stages of B cell differentiation in enhancer-mediated susceptibility. Endothelial and mesenchymal cells seem to play a role in another set of cancer traits, among them melanoma, ovarian, epithelial ovarian, thyroid and lung cancer. Only a limited number of cancer traits are associated with myeloid cells,

such as acute lymphoblastic leukaemia, colorectal cancer and diffuse large B cell lymphoma. Enhancers linked to adverse responses to chemotherapy exhibit different activity profiles, with anti-microtubule drugs being associated with neutrophils, paclitaxel and paclitaxel + carboplatin with monocyte lineage cells, and carboplatin with endothelial cells **(Fig. 4.3.C).**

4.2.4. Non-coding SNPs in adaptive immune cells increase Crohn's disease risk, while those in innate immune cells increase the risk of ulcerative colitis.

Inflammatory Bowel Disease (IBD) is a complex and chronic inflammatory condition of the gastrointestinal tract, encompassing two major forms: Ulcerative Colitis (UC) and Crohn's Disease (CD). IBD presents a formidable challenge in clinical differentiation due to overlapping clinical presentations such as abdominal pain, diarrhoea, and fatigue, and shared inflammatory characteristics within the digestive system (Colombel et al., 2019). UC typically presents with inflammation limited to the colon, spreading proximally from the rectum in a continuous fashion, whereas CD often spares the rectum and can affect any part of the gastrointestinal tract from the mouth to the anus (Yu & Rodriguez, 2017). Diagnostic tests like endoscopy, imaging, and biopsies are often necessary to differentiate them, but often, these may not be definitive (Colombel et al., 2019).

The interplay of host genetics, immune dysregulation, and environmental factors contribute to the aetiology of these conditions (T.-C. Liu & Stappenbeck, 2016). The genetic complexity of IBD becomes evident when considering that 250 loci comprising both common and rare variants have been associated with the disease, and most identified susceptibility alleles are common variants located in the non-coding regions with modest effects in disease development (Anderson et al., 2011; De Lange et al., 2017; Ellinghaus et al., 2016; Hong et al., 2018; Julià

et al., 2014; Kakuta et al., 2018; Kenny et al., 2012; J. Z. Liu et al., 2015; Parkes et al., 2007; Yamazaki et al., 2013; Yang et al., 2014).

Since genetic factors contribute to the risk of developing IBD and are mostly non-coding variants (T.-C. Liu & Stappenbeck, 2016), understanding the molecular underpinnings of UC and CD requires exploration of the epigenomic landscape in these loci, especially the enhancer regions. Therefore, I investigated enhancer activity profiles associated with non-coding SNPs linked to IBD traits.

Our results revealed distinctions between UC and CD. Notably, enhancers associated with UC displayed activity in myeloid cell lineages, such as monocytes and neutrophils, as well as endothelial and mesenchymal cells and B cell lineage. On the other hand, enhancers linked to CD exhibited activity in lymphoid cell types, including effememCD8T cells and germcenterB and naïve B cells. Enrichment of enhancers active in DCs was found in loci associated with both diseases **(Fig. 4.4.)**.

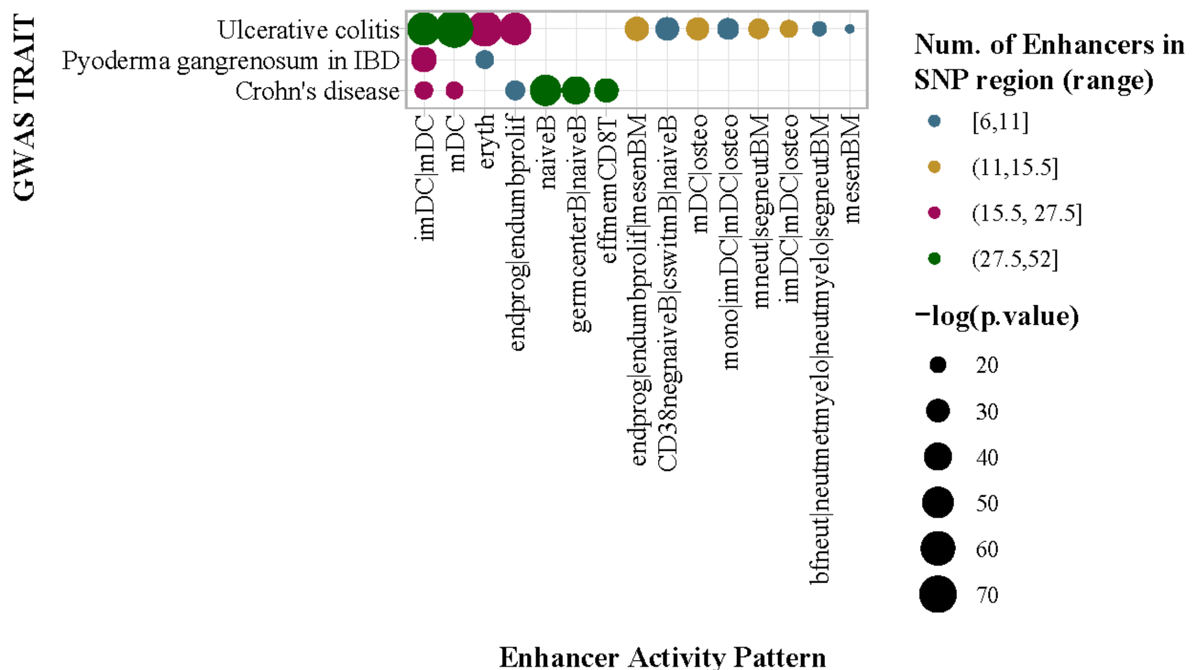


Figure 4. 4. Exploring Enhancer Enrichment in IBD-associated loci. The dot plot shows the activity profiles of enhancers enriched in associated loci associated with three IBD-related traits. Each dot intersecting a trait indicates that there is at least one significant p-adjusted value supporting the enrichment. The colour of the dot represents the number of enhancer bins in the enriched activity profile.

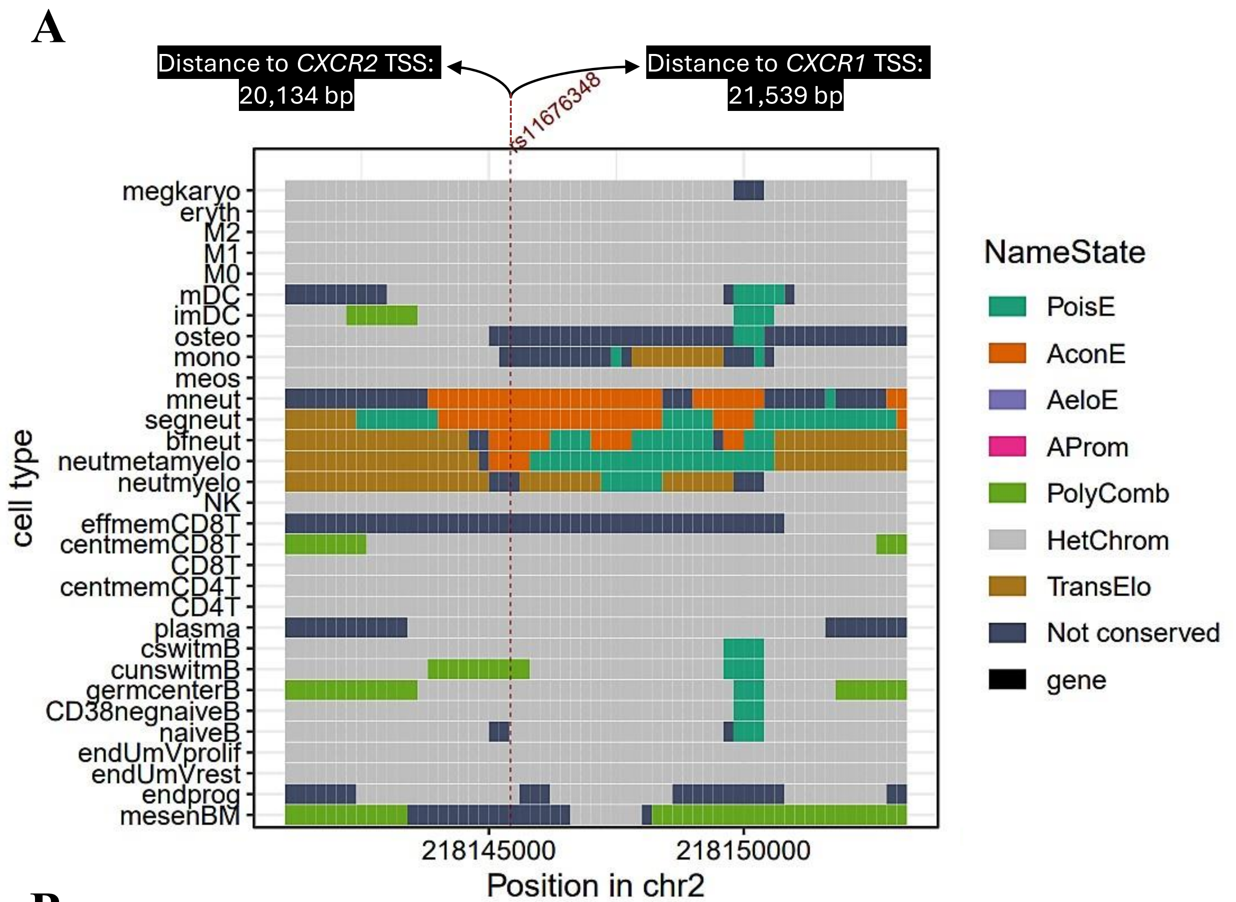
4.2.5. Ulcerative colitis-associated enhancers potentially target key players in NETosis.

Delving into the specific immune responses associated with UC and CD, it is crucial to consider the role of neutrophils in the disease mechanism. In the context of IBD, neutrophil infiltration in the intestine correlates strongly with disease activity, particularly in UC (Akpinar et al., 2018; Therrien et al., 2019; Zhou et al., 2018). This phenomenon is a reliable component of UC disease scoring systems (Jairath et al., 2019).

In a healthy intestine, damage to the intestinal barrier and bacteria-derived molecules triggers the recruitment of neutrophils from the circulation to the inflamed tissue (Drury et al., 2021; Ley et al., 2007). Recruited neutrophils participate in the elimination of microorganisms through phagocytosis, degranulation, reactive oxygen species (ROS) generation, and the release of neutrophil extracellular traps (NETs) (Mutua & Gershwin, 2020). Once their functions are completed in healthy tissues, neutrophils undergo apoptosis and efferocytosis, facilitating the resolution of inflammation, tissue repair, and a return to normal tissue homeostasis (McCracken & Allen, 2014; Scannell et al., 2007).

NETosis is a form of programmed cell death involving the release of DNA and antimicrobial proteins. Notably, increased NETs have been identified in the inflamed gut mucosa, stool, and blood of UC patients (M. Cao et al., 2017; Dinallo et al., 2019; He et al., 2016), and their accumulation in the colon is associated with tissue damage and inflammation (T. Li et al., 2020). Accumulation of NETs in the inflamed gut also boosts the production of neutrophil activation cytokines by macrophages, contributing to an amplification loop.

Here, I have discovered that UC-associated enhancers play a crucial role in regulating neutrophil activity. For instance, I have identified neutrophil enhancers in UC risk loci in gene-depleted regions containing SNPs rs11676348 and rs2310173 **(Fig. 4.5).** Using data from EpiMap (Boix et al., 2021), I found that these regions are annotated with the chromatin state associated with ZNF genes & repeats in most samples (**Fig. A3.A)**. This data is compatible with our findings because in our epigenomic maps, the enhancer is specific to neutrophils, a cell type not included in EpiMap. Similarly, the EpiMap data supports our finding of the regions overlapping with the enhancer on the right side of rs2310173. In their annotations, this region appears as a quiescent/low state in nearly half of the samples and as a weak enhancer (yellow) or active enhancer annotation in the rest (**Fig. A3. A).** This also fits with our findings, as the enhancer I identified is active in other BLUEPRINT myeloid cell types (such as dendritic cells,


A
Distance to CXCR2 TSS: 20,134 bp
Distance to CXCR1 TSS: 21,539 bp
rs11676348
megkaryo
eryth
M2
M1
M0
mDC
imDC
osteo
mono
meos
mneut
segneut
bfneut
neutmetamyelo
neutmyelo
NK
effmemCD8T
centmemCD8T
CD8T
centmemCD4T
CD4T
plasma
cswitmB
cunswitmB
germcenterB
CD38negnaiveB
naiveB
endUmVprolif
endUmVrest
endprog
mesenBM
cell type
218145000
218150000
Position in chr2
NameState
PoisE
AconE
AeloE
AProm
PolyComb
HetChrom
TransElo
Not conserved
gene

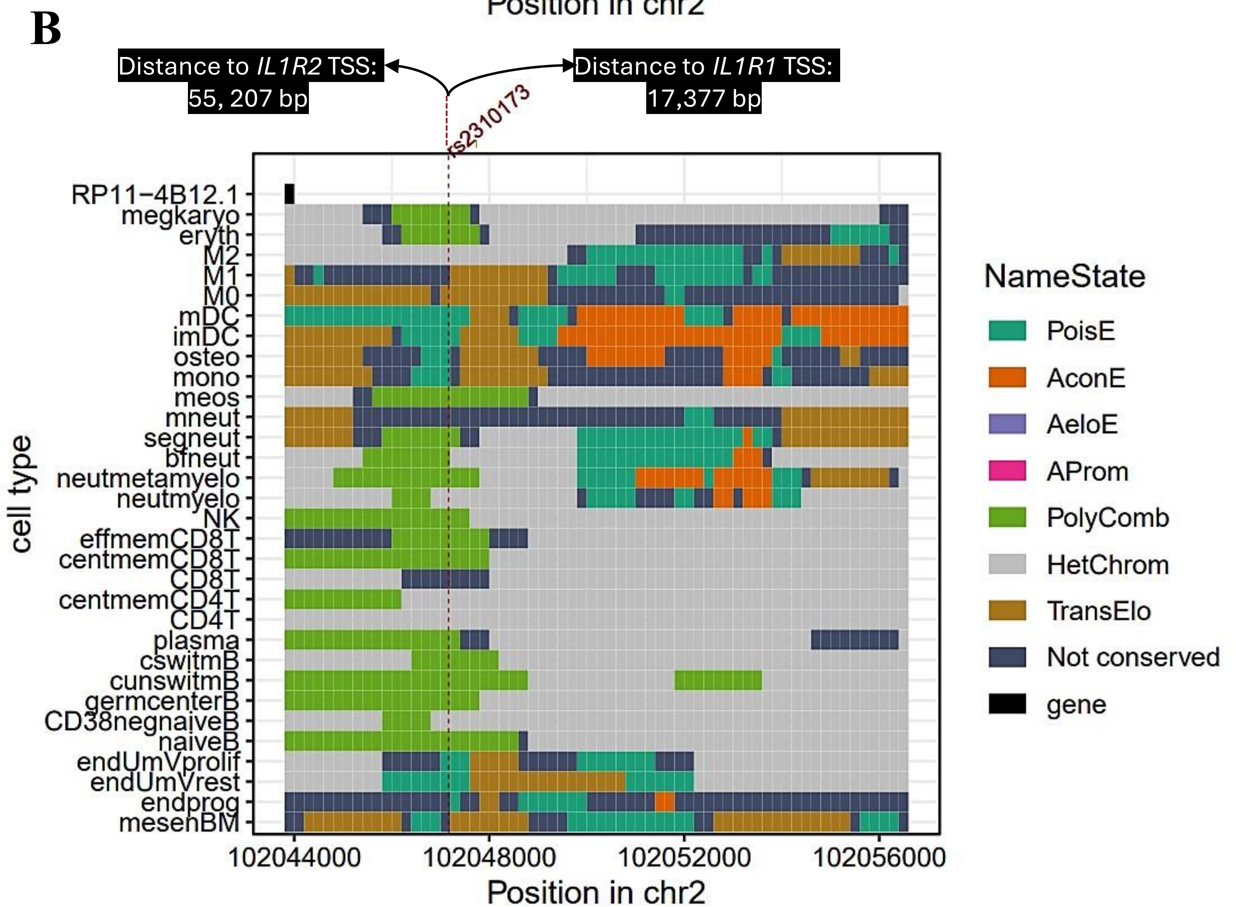

B
Distance to IL1R2 TSS: 55, 207 bp
Distance to IL1R1 TSS: 17,377 bp
rs2310173
RP11−4B12.1
megkaryo
eryth
M2
M1
M0
mDC
imDC
osteo
mono
meos
mneut
segneut
bfneut
neutmetamyelo
neutmyelo
NK
effmemCD8T
centmemCD8T
CD8T
centmemCD4T
CD4T
plasma
cswitmB
cunswitmB
germcenterB
CD38negnaiveB
naiveB
endUmVprolif
endUmVrest
endprog
mesenBM
cell type
102044000
102048000
102052000
102056000
Position in chr2
NameState
PoisE
AconE
AeloE
AProm
PolyComb
HetChrom
TransElo
Not conserved
gene

Figure 4. 5. The epigenomic landscape of UC-associated genomic region enriched with enhancers. Neutrophil enhancers enriched in a region associated with UC SNP rs11676348 in **panel A** and myeloid enhancers enriched in a region containing UC SNP (rs2310173) in **panel B**. Both regions are intergenic but the SNPs are eQTLs of genes (indicated with arrows) related to NETosis (see text for details). The SNP rs11676348 is an eQTL of *CXCR1* and *CXCR2,* while rs2310173 is an eQTL of *ILR1* and *ILR2*.

monocytes, and osteoclast), and the EpiMap included several samples of monocytes and dendritic cells.

By using eQTL data from the eQTL Catalogue (Kerimov et al., 2021b), I identified that the SNP rs11676348 is an eQTL in neutrophils of chemokine (CXCL-8) receptor genes *CXCR1* (study: CEDAR, p value:1.0 e-8, effect size: -0.123, distance to TSS: 21,539 bp) and *CXCR2* (study: BLUEPRINT, p value:1.1 e-8, effect size: -0.063, distance to TSS: 20,134) (**Fig. 4.5.A**). Furthermore, using eQTL data from The GTEX Consortium (Aguet et al., 2020), I have found that the SNP rs2310173, located upstream enhancers active in myeloid cells, including neutrophils (**Fig 4.5.B**), is an eQTL of two genes encoding for IL-1β receptors. These genes are *IL1R1* (study: GTEX, tissue: skin-sun-exposed, p-value: 5.3 e-5, variant effect in expression: increase, distance to TSS: 17,377 bp) and *IL1R2* (study: GTEX, tissue: cultured fibroblasts, p-value: 9.1 e-6, variant effect in expression: increase, distance to TSS: 55,207 bp). More details of the genomic context of these GTEX eQTLs, including genomic context and expression effects are presented in **Fig. A4**.

I also explored chromatin long-range interaction data for neutrophils generated by Javierre et al. (2016). However, there was no evidence of interaction between the gene promoters and any fragment containing the risk variants in any of the haematopoietic cell types included in their study. This could be due to the distance between the genes and the variants being lower than the typical resolution of Hi-C and PC Hi-C experiments. For instance, Javierre et al. reported a median linear distance of 331 Kb between promoters and their interacting regions in their study (Javierre et al., 2016). To the best of my knowledge, there is currently no high-resolution Hi-C dataset available specifically generated for neutrophils.

These results are significant because CXCL-8 and IL-1β are key cytokines and chemokines known to induce NETosis (Abrams et al., 2019; Gillot et al., 2021; Meher et al., 2018; Mitroulis et al., 2011) and CXCR1 and CXCR2 are the receptors for CXCL-8 (Baggiolini, 2000; Ishimoto et al., 2023; M. W. Nasser et al., 2009) while IL1R1 and IL1R2 receptors recognise IL-1β (Boraschi, 2022; Boraschi & Tagliabue, 2013). The specificity of these enhancers highlights the importance of neutrophils in the modulation of immune responses associated with UC. This

functional connection suggests that these enhancers play a role in the dysregulation of neutrophil activities, including the release of NETs. This contributes to the inflammatory processes observed in UC.

4.3. Discussion.

Complex diseases result from a combination of genetic, environmental, and lifestyle factors. Thanks to GWAS, many genomic loci have been associated with complex diseases, and it has been discovered that genetic associations with complex diseases generally involve multiple non-coding regions, each contributing a small effect (Jonkers & Wijmenga, 2017).

I leveraged information on non-coding GWAS variants from the GWAS Catalog (Buniello et al., 2019) to understand the potential involvement of enhancers in dysregulating the gene expression of disease-relevant genes, and identify the cell types in which those enhancers can induce a phenotypic change. To that end, I focused on the comprehensive catalogue of haematopoietic enhancers and enhancer activity profiles generated in Chapter 3. I narrowed down my analysis to the top 100 most commonly observed activity profiles among haematopoietic enhancers, prioritising the probability of relevance of phenotypic effects on disease variants and ensuring computational resource efficiency.

Our enrichment analysis of enhancers with specific activity profiles in proximity to non-coding SNPs associated with complex traits uncovered 1,113 significant activity profile-trait associations. I also provided quantification of the complexity of these diseases in terms of the number of enhancer bins potentially affected by enhancer dysfunction in specific sets of cell types. In particular, autoimmune diseases and inflammation-related traits stand out, with associations with multiple enhancer activity profiles.

Other researchers have integrated enhancer maps for various cell types and GWAS data to investigate whether disease-associated variants are more prevalent in enhancer regions compared to other genetic variants (Markunas et al., 2017; J. Nasser et al., 2021; Roadmap Epigenomics Consortium et al., 2015). Their focus is on the total set of enhancers active in a given cell type, without acknowledging the potential for enhancers to be either cell type-specific or active across multiple cell types. Additionally, their analysis requires fine mapping of variants, aiming to identify specific point mutations within enhancers with potential phenotypic effects.

In contrast, my enrichment analysis takes an enhancer-centric approach exploring whether enhancer bins of 200bp accumulate within a fixed-size window centred on disease-associated variants. I believe that enhancer-centric approaches, pinpointing the whole functionally relevant regions and considering their multicellular activity profiles, could be more appropriate for understanding disease mechanisms than variant-centric approaches aiming to identify specific point mutations within enhancers. This is because disease-associated tag variants may be located on clusters of multiple enhancers or longer enhancers, both of which are known to have greater impacts on gene expression than single or shorter enhancers (Li & Wunderlich, 2017; Khan et al., 2018; Quang et al., 2015; Parker et al., 2013; Joseph W. Blayney et al., 2023). Besides, they are more enriched in transcription factor binding sites than their counterparts and, therefore may contain multiple causal variants (Michida et al., 2020; Parker et al., 2013; Grosveld et al., 2021), whether already fine-mapped or yet to be discovered. Recently, Engritz et al., observed that the proximity between enhancers is associated with the size of their impact on gene expression, close by enhancers have a super-additive effect on gene expression changes and of one enhancer in the cluster can impact the function of other enhancers in its vicinity (Gschwind et al., 2023).

A recent study supports the idea that, although some specific points inside an enhancer may have a higher impact on gene expression, the whole region is important to achieve the enhancer's full physiological potential (Joseph W Blayney et al., 2023). Blayney et al. (2023) conducted a combinatorial reconstruction of the well-characterised mouse a-globin super-enhancer (SE), shedding light on the cooperative mechanisms among its constituent elements to impact target gene expression. All five SE components (R1, R2, R3, Rm, and R4) share enhancer chromatin signatures, but only two of them (R1 and R2) act as classical enhancers, synergistically increasing target gene expression by 450-fold when combined, while individually, they only increase it by 100-fold each. These additional elements (R3, Rm, and R4), termed "facilitators," enhance the activity of R1 and R2, but they do not increase target gene expression by themselves (Joseph W Blayney et al., 2023). Results from other studies suggest that this phenomenon is present in other enhancers across the genome (Hnisz et al., 2015; Sahu et al., 2022; Song et al., 2019).

It's worth noting that activity profiles may be incomplete, as enhancers could be active in additional cell types not included in our dataset. Despite this, the significance of my results is noteworthy. I utilised the most comprehensive set of enhancers for immune cells generated to date. These findings establish connections between more than 300 complex traits, including those related to cardiovascular disease, neurocognitive function, cancer, autoimmune disease,

and other diseases, with regions enriched in enhancers exhibiting specific activity profiles across haematopoietic cell types. This compilation serves as a valuable resource, offering insights for the design of future experiments to validate the phenotypic effects of enhancer disruption on disease risk. m

To exemplify the potential utility of my results, I used IBD as a prototypic example, highlighting the distinctions between activity profiles associated with UC and CD. Particularly, I found that UC is more associated with enhancers in myeloid cell types and CD with enhancers in lymphoid cell types. My findings also suggest that DC enhancers are associated with both UC and CD, corroborating previously published findings on the role of DC enhancers in IBD (Nasser et al., 2021). Another study which examined the enrichment of GWAS SNPs in DNase I hotspots using data from the Roadmap Epigenomics consortium discovered an enrichment of CD risk SNPs in DNase I hotspots of B and T cells, whereas UC risk SNPs where enriched in DNase I hotspots of monocytes (Breeze, Haugen, et al., 2022). These findings support our own results. However, Breeze et al. also observed that CD-associated GWAS SNPs were enriched in DNase I hotspots of NK and B cells, which we did not find in our analysis. This discrepancy might be due to DNase I hotspots encompassing other regions besides the active enhancers considered in our study.

I focused on neutrophil enhancers associated with UC and went one step further in identifying their potential gene targets, which appear to regulate key genes encoding proteins that play a role in NETosis (Abrams et al., 2019; Gillot et al., 2021; Meher et al., 2018; Mitroulis et al., 2011): specifically, CXCR1 and CXCR2 receptors of CXCL-8, and IL1R1 and IL1R2 receptors of IL-1β. Therefore, I proposed that some neutrophil enhancers are involved in the regulation of NETosis in UC. Previous studies have implicated these genes in UC (De Lange et al., 2017; Jostins et al., 2012), but have not shown the mapping of enhancers in the region.

A recent meta-analysis of GWAS data, which included IBD genetic associations for individuals of European and East Asian ancestry, found that genetic effects are generally consistent across ancestries (Liu et al., 2023). The analysis indicated that genetics underlying CD appear to be more ancestry-dependent than those for UC (Liu et al., 2023). It is worth noting that these interesting genetic associations were not included in my dissertation because they were not available in the GWAS Catalog at the time I performed my analysis.

Among their results on chromosome 2, they identified *CXCR1* as a gene associated with the IBD risk variant rs11669031. However, *CXCR2*, *IL1R1*, and *IL1R2* were not mentioned. According to information available on the Open Targets database (Mountjoy et al., 2021;

Ghoussaini et al., 2021), this rs11669031 is an eQTL and an sQTL (splicing Quantitative Trait Locus) for *CXCR2*, which is located 27,450 bp away. Additionally, Liu et al. reported an IBD risk variant within the intron of the *IL18R1* gene. *IL18R1* encodes a component of the IL18 receptor complex and is located upstream of *IL1R2* and *IL1R1*(Liu et al., 2023). I believe these gene associations could also support the hypothesis of non-coding variants regulating NETosis in UC. Some experimental evidence suggests that the expression of another component of the IL18 receptor complex, IL18RAP, is elevated in neutrophils of patients with Lupus erythematosus (Ma et al., 2021). These neutrophils showed enhanced IL18-mediated production of ROS species. The authors discuss that this may be attributed to high levels of IFN in serum, which have been shown to prime neutrophils to undergo NETosis (Ma et al., 2021).

In my study, I found that the distance between the NETosis-related SNP variants in the enhancer regions associated with UC and their potential enhancer gene targets (eGenes) ranges between 17,000 bp and 55,000 bp. These distances are in the expected range according to previous studies that, by using the ABC model (Fulco et al., 2019), identified target genes of IBD-associated enhancers in cell types different from neutrophils (Nasser et al., 2021). They found that the median distance between the SNPs associated with IBD in the enhancer regions and the TSS of the target genes is 18, 848 bp (IQR: 57,170 bp, MEAN: 60,781 bp) (Nasser et al., 2021).

These examples showcase the utility of our dataset for proposing links between enhancers and disease-related genes, information that could guide future experiments and contribute to our understanding of disease mechanisms.

Chapter 5. Macrophage enhancers are associated with cardiovascular disease risk by regulating the expression of key genes in lipid metabolism.

Building on the identification of cell types affected by genetic defects in enhancers (**Chapter 4**), it is equally crucial to pinpoint the specific genes whose expression is dysregulated in those cell types. This enhances our understanding of the directly affected biological processes and provides a refined starting point for designing experimental models crucial to understanding the disease mechanism. In **Chapter 4**, I identified enhancers associated with cardiovascular disease, the leading cause of death globally, which were predominantly active in macrophage subtypes. In this chapter, I investigated the genes targeted by those enhancers and mapped their functions to specific steps in the three distinct lipid metabolism pathways.

5.1. Introduction.

Cardiovascular disease (CVD) is a broad term used to describe a class of diseases that affect the heart (cardio) or blood vessels (vascular). CVDs cause around a quarter (27 %) of all deaths in the UK, more than 170000 deaths a year (British Heart Foundation, 2024). They are a major global health issue and a leading cause of morbidity and mortality worldwide (World Health Organization, 2021). As in other complex diseases, the development of CVDs is influenced by genetic predisposition, lifestyle, and environmental factors (Tada et al., 2022; Watkins & Farrall, 2006; Mozaffarian et al., 2008).

The underlying cause of several CVDs, including coronary artery disease, carotid artery disease, and peripheral artery disease, is atherosclerosis (Gisterå & Hansson, 2017). The development of atherosclerosis is influenced by factors such as sex, age, family history, hypertension, smoking, dyslipidemia, metabolic syndrome (Fruchart et al., 2004), obesity and type 2 diabetes (Lechner et al., 2020).

5.1.1. Atherosclerosis.

Atherosclerosis is a progressive inflammatory and lipid disorder. According to our current understanding of the pathology, it arises due to the accumulation of cholesterol-carrying low-density lipoproteins (LDLs) along the endothelial lining of blood vessels. Various modifications of the retained LDL, such as oxidation, mimic damage-associated molecular

patterns (DAMPs) and trigger an inflammatory response. This response leads to the activation of endothelial and vascular smooth muscle cells (SMCs) and the recruitment of immunocytes such as monocyte-derived macrophages, T cells, B cells, dendritic cells, and mast cells (Tabas et al., 2015). Immunocytes contribute to the local build-up and amplification loop of the inflammatory response. This build-up is called the "atherosclerotic plaque" and results in the narrowing of the area of the vessel for blood. Plaque formation at sites of vascular curvature and bifurcation makes them susceptible to rupture (Williams & Tabas, 1995; Skålén et al., 2002).

5.1.2. Dyslipidaemia.

Dyslipidaemia is defined as an abnormally high concentration of triglycerides and/or cholesterol in the blood (Stein et al., 2019). Lipids are transported through the bloodstream by lipoproteins such as the High-Density Lipoprotein (HDL), the Very Low-Density Lipoprotein (VLDL) and LDL; measuring their levels in plasma is used to assess dyslipidaemia and serves as established biomarkers for atherosclerotic CVD (Schunkert et al., 2011; Kathiresan & Srivastava, 2012; Tada et al., 2022). Researchers have explored genetic determinants of lipid blood levels in an effort to comprehend susceptibility to CVD.

It has been discovered that monogenic and oligogenic mutations inside the protein-coding genes cause severe cases of dyslipidaemia. For instance, familial hypercholesterolemia (FH), characterised by extremely elevated levels of cholesterol in plasma, is caused by mutations inside genes such as *LDLR*, *APOB* and *PCSK9* (Trinder et al., 2020). Disease manifestation in Europeans is mainly attributed (over 70% of instances) to mutations in the *LDLR* gene. *ApoB* gene mutations contribute to 2-5% of FH cases, while *PCSK9* gene mutations account for less than 1% (Henderson et al., 2016; Singh & Bittner, 2015; Vallejo-Vaz & Ray, 2018). About 15% of FH cases result from either polygenic factors or specific mutations like those in the *APOE* gene, but their prevalence remains unknown (Henderson et al., 2016). The prevalence of FH is typically 1 in 200–250, but in certain populations with founder effects, the prevalence is higher. For instance, in Ashkenazi Jews, it is as high as 1 in 67 (Henderson et al., 2016).

Severe cases of hypertriglyceridemia (extremely elevated levels of triglycerides in plasma) are caused by rare mutations in genes such as *LPL*, *APOC2*, *APOA5*, *LMF1*, *GPIHBP1*, and *GPD1* (Connelly et al., 1987; Dorfmeister et al., 2008; Emi et al., 1990; R. M. Fisher et al., 1995; Hata et al., 1990; Hegele et al., 2014; Henderson et al., 1991; Ishimura-Oka et al., 1992; Ma et al.,

1991; Mailly et al., 1995; Reymer et al., 1995; C. J. D. Ross et al., 2005; J. Wang et al., 2007, 2007, 2007). Finally, dyslipidaemia of HDL cholesterol levels can be presented when HDL levels are lower than the normal range or when they are extremely high (Rohatgi et al., 2021). Bi-allelic loss-of-function mutations in *APOA1* cause HDL deficiency (Zanoni & Von Eckardstein, 2020), while hyperalphalipoproteinaemia (extremely elevated levels of HDL in plasma) is caused by mutations in *CETP*, *LIPC*, and *APOC3* (Giammanco et al., 2021). These severe forms of dyslipidaemias confer a high risk of CVD and the development of premature atherosclerosis. Their study has contributed to the understanding of the pathways of lipid metabolism as well as the design of current treatments available for dyslipidaemia (Endo, 2010).

Most cases of atherosclerotic CVD develop in people without those underlying penetrant mutations; hence, understanding genetic susceptibility in these common cases remains an ongoing task. GWAS in multi-ethnic populations have identified numerous variants associated with plasma levels of HDL, LDL and triglycerides. For instance, the levels of HDL cholesterol (HDL- C) are associated with non-coding variants in loci containing *ABCA1, APOA1, APOE, LIPC, and STARD3* genes (Teslovich et al., 2010; Willer et al., 2013; Do et al., 2013); plasma triglycerides levels are associated with non-coding variants in loci containing *APOB*, *APOE*, *LPL* and *PLTP* (Keebler et al., 2010, 2009; Lanktree et al., 2009; Teslovich et al., 2010; Johansen et al., 2011); and LDL cholesterol levels are associated with non-coding variants in loci containing *APOA1, APOC1, APOB, LDLR, CYP7A1, VLDLR, LPA* loci (Do et al., 2013; Willer et al., 2013).

According to Alsheikh et al. (2022), who conducted a systematic review of the landscape of functionally validated non-coding GWAS variants across 130 human traits, only four non-coding GWAS variants associated with CVD-related lipid traits have been experimentally validated (Alsheikh et al., 2022). These include two cis-regulatory variants targeting the gene *GALNT2,* which are associated with HDL levels (Roman et al., 2015), one variant associated with total cholesterol levels in coronary artery disease that affects the function of miRNA targeting the *GOSR2* gene (Ghanbari et al., 2014), and a variant in the promoter of *LDLR* associated with hypercholesterolemia (De Castro-Orós et al., 2014).

While there is no evidence of the functional validation of the link between most non-coding GWAS variants and the plasma lipid levels, the fact that mutations inside some of the genes near those variants cause severe tryglyceridemia (e.g. *LPL*), hypercholesterolemia (e.g. *LDLR, APOB*) and hyperalphalipoproteinemia (e.g. *LIPC*) supports the hypothesis that GWAS non-

coding variants associated with plasma lipid levels could be inside regulatory elements controlling the expression of lipid metabolism genes.

5.1.3. Macrophages' role in atherosclerosis.

Macrophage phenotype and function are inherently tied to metabolic signals derived from their tissue environment (Buck et al., 2017; Van den Bossche et al., 2017). For example, Macrophages possess an efficient lipid-handling machinery; they can actively sense, engulf, store, export and burn lipids, and in lipid-rich tissues, they can process substantial lipid amounts (Yan & Horng, 2020; Remmerie & Scott, 2018; Vogel et al., 2022)

Lipid-dense macrophages, termed 'foam cells', are the most abundant type of cells in atherosclerotic plaques (Vogel et al., 2022; Bobryshev et al., 2017; Robbins et al., 2013)**.** At the initiation of atherosclerosis, monocytes attracted to the affected area will differentiate into tissue macrophages capable of taking up modified lipoproteins in an attempt to clear the neointima of hazardous material (Vogel et al., 2022; Remmerie & Scott, 2018).

Lipid metabolism by macrophages is marked by three different processes: cholesterol uptake, esterification and efflux. Uncontrolled uptake of oxidised low-density lipoprotein (ox-LDL), excessive cholesterol esterification and/or impaired cholesterol release result in the accumulation of cholesterol ester stored as cytoplasmic lipid droplets and subsequently trigger the formation of foam cells(Yu et al., 2013; Maguire et al., 2019). Over the course of atherosclerosis, these cells accumulate within the arterial lining and will become apparent along the arterial wall in early atherosclerotic lesions (Yu et al., 2013; Wculek et al., 2022) **(Fig. 5.1.)**

5.1.4. Preliminary results link macrophage enhancers to CVD risk.

All the evidence mentioned above suggests that the accumulation of lipids in blood vessels, the transformation of macrophages into lipid-engulfing foam cells, and the accumulation of immunocytes that amplify the inflammation loop contribute to the development of common cases of atherosclerosis. However, it also highlights that the explicit functional role of genetic regions conferring susceptibility to these events remains elusive.

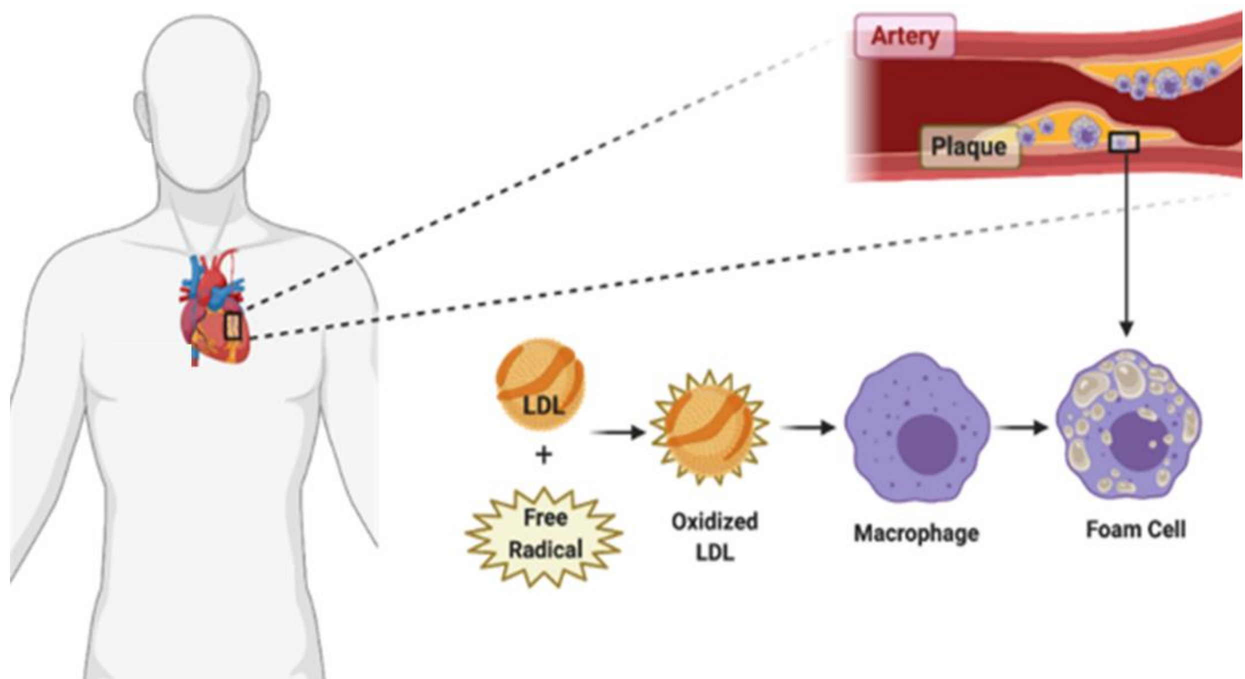


Figure 5. 1. Macrophages' role in atherosclerosis. Atherosclerosis is the main underlying cause of CVD and is characterised by the formation of plaques in arterial walls. Foam cells are the most abundant cells present in plaques. Accumulated foam cells in the plaques area actually lipid-laden macrophages that have ingested oxLDLs accumulated in the arterial walls. Figure adapted from (Eshghjoo et al., 2021)

In the previous chapter, I discovered that GWAS loci associated with plasma levels of HDL, LDL, and triglycerides, as well as those associated with metabolic syndrome and other risk factors for CVD, are enriched in enhancers. Moreover, these enhancers are selectively activated in subsets of haematopoietic cell types, with the majority of them being specific to M1 and/or M2 macrophages. Refer **to section 1.7** for a description of the M1/M2 macrophage *in vitro* model.

In this chapter, I investigated those macrophage enhancers (M1 and M2) associated with CVD risk traits, identified their potential gene targets, and explored their respective functions in the search for overlapping biological processes that could help us understand how the aggregate contribution of non-coding variants predisposes to CVD.

5.2. Results.

5.2.1. Exploring enhancer activity in cardiovascular disease risk.

In my GWAS enrichment analysis (**section 4.2.3.**), I made a noteworthy discovery regarding the relationship between non-coding SNPs associated with cardiovascular disease risk and

enhancers specifically active in macrophages. These CVD-associated traits encompassed levels of LDL cholesterol, HDL cholesterol and triglycerides in the blood, metabolic syndrome, coronary artery disease, waist-hip ratio, and various other lipid-related characteristics **(Fig. 5. 2)**. I observed a significant enrichment of enhancers displaying distinct activity patterns across different haematopoietic cells, totalling 18 specific enhancer activity profiles **(Fig. 5. 2).**

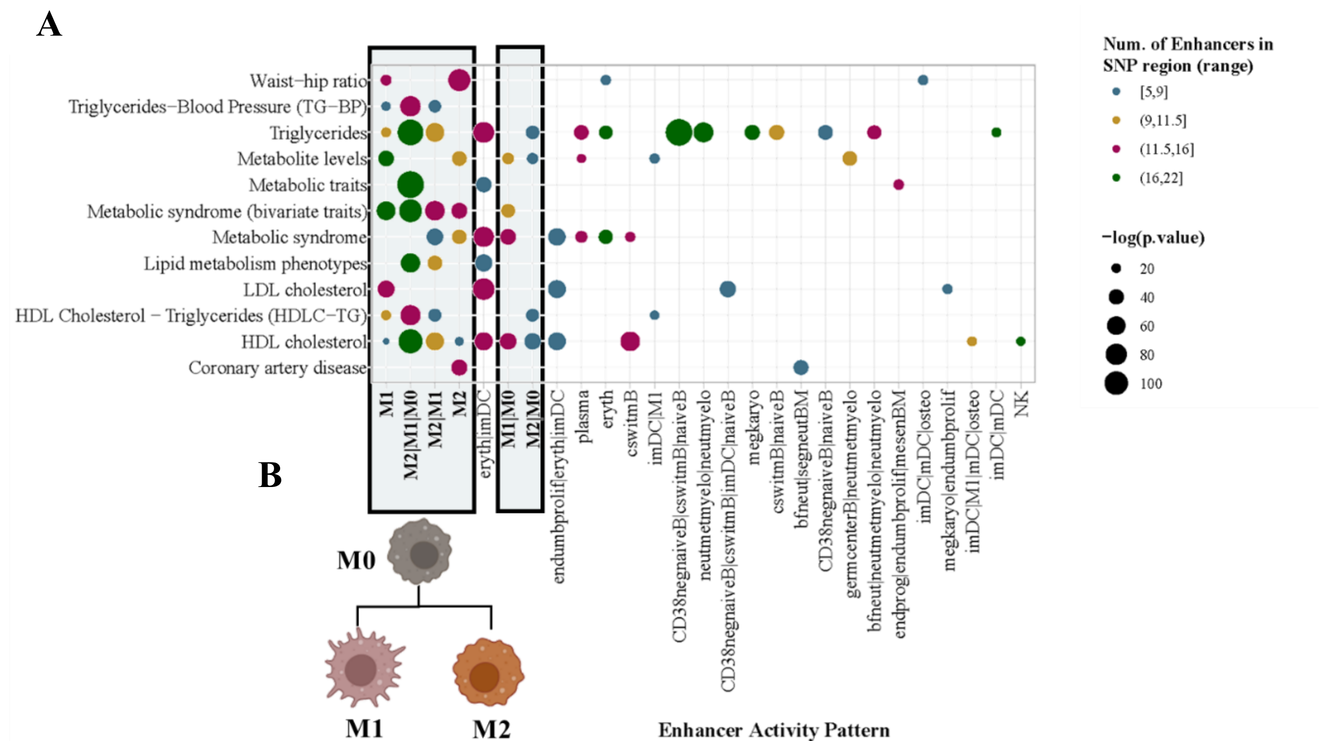


Figure 5. 2. Enhancer activity profiles associated with CVD risk. (A) The dot plot illustrates the activity profiles of enhancers enriched in loci associated with CVD risk-related traits. Each dot at the intersection of a trait and an activity profile indicates the presence of at least one GWAS study with a significant p-value (padj < 0.0001), supporting the enrichment for a trait. The colour of each dot corresponds to the minimum number of AE bins overlapping trait-associated loci among GWAS studies for that trait. The size of the dot reflects the level of significance (max padj among the GWAS studies). Macrophage-specific enhancers are highlighted in blue boxes, highlighting their higher significance and number of enhancers compared to other profiles. Notably, I observed that macrophage enhancer bins are overrepresented (52%) among loci associated with CVD. **(B)** Schematic representation of the developmental hierarchy of macrophages.

I identified 133 loci containing active enhancer (AE) bins with activity profiles associated with CVD traits. Notably, a substantial portion (69 loci, 52 %) was associated with macrophage activity profiles, including M1, M2, M0, and various combinations thereof. These macrophage-enriched enhancer regions spanned across 14 different chromosomes, as detailed in **Table 5.1.** I merged trait-associated regions that overlapped or were at a distance of less than 10 Kb,

Table 5. 1.Genomic regions associated with CVD Risk and enriched in macrophage enhancers. Overview of chromosomal locations, region sizes, trait-associated SNPs and their corresponding eQTL genes in 17 genomic regions associated with CVD risk that contain macrophage enhancers with activity profiles showing significance in the enrichment analysis.

	chr	Size (bp)	SNPs	Trait	Activity Profile	eGenes
1	chr1	10000	rs17114036	Coronary artery disease	M2	*PLPP3*
2	chr2	10400	rs10211524	Metabolite levels	M1|M0, M1	--
3	chr2	10000	rs895636	Metabolite levels	M1|M0	*SRBD1, PPM1B, PREPL, SIX3, LRPPRC, CAMKMT, PRKCE*
4	chr5	10000	rs6861681	Waist-hip ratio	M2	*CPEB4*
5	chr6	12200	rs1294410, rs1294421	Waist-hip ratio	M2	*RP1-80N2.2*
6	chr8	11900	rs1441756, rs2083637	Metabolic syndrome (bivariate traits), HDL cholesterol	M1, M1|M0	*PSD3, LPL*
7	chr8	26200	rs295, rs301, rs264, rs326, rs331, rs325, rs1059611, rs15285, rs13702, rs2197089, rs10096633, rs10105606, rs17482753	Metabolic syndrome, Metabolic syndrome (bivariate traits), Coronary artery disease, HDL cholesterol, Triglycerides, Lipid metabolism phenotypes, Triglycerides-Blood Pressure (TG-BP), HDL Cholesterol - Triglycerides (HDLC-TG)	M2, M2|M1, M2|M1|M0, M1, M1|M0	*LPL*
8	chr9	12900	rs4149310, rs2515629	Metabolite levels, HDL cholesterol	M1, M1|M0	*ABCA1*
9	chr10	10000	rs7081678	Waist-hip ratio	M1	*MACORIS*
10	chr11	19700	rs10838681, rs7120118	Metabolic syndrome, HDL cholesterol	M1|M0, M2|M0, M2	*NR1H3, MADD*
11	chr12	10000	rs718314	Waist-hip ratio	M1	*SSPN, RP11-283G6.4*
12	chr15	12400	rs10468017, rs2043085, rs1532085	Metabolic syndrome (bivariate traits), Metabolite levels	M1	*ALDH1A2*
13	chr15	12700	rs2306786	Metabolite levels	M1, M1|M0	*MYO1E*
14	chr19	34600	rs157580, rs157582, rs439401, rs445925, rs12721054	HDL cholesterol, Metabolic syndrome, HDL Cholesterol - Triglycerides (HDLC-TG), Triglycerides, Metabolite levels	M1|M0, M2|M0, M1, M2|M1	*TOMM40, APOE, APOC1*
15	chr20	11200	rs4810479	Metabolite levels	M1	*PLTP*
16	chr22	10000	rs12483959	Metabolite levels	M2	*PNPLA3*
17	chrX	10000	rs5031002	LDL cholesterol	M1	*OPHN1, STARD8, YIPF6*

resulting in 17 regions. Some of these regions contain full or partial open reading frames **Table 5.1.** Detailed methodology is in **section 2.7**.

5.2.2 Using eQTL data to identify the target genes of CVD-associated enhancers.

As previously explained in **section 1.5.**, eQTL (expression quantitative trait loci) analysis is a method used to identify genetic variants associated with gene expression levels. When applied to study enhancers, eQTL analysis can help identify genetic variants that influence the expression of genes targeted by enhancers. The genes (eGenes) whose expression is correlated with the genotype of CVD-associated non-coding SNPs (eQTLs) are the potential enhancer targets.

Using naïve macrophages eQTL data from Alasoo (2018) and Nedelec (2016), reprocessed by the eQTL Catalogue (Alasoo et al., 2018; Kerimov et al., 2021b; Nédélec et al., 2016), I discovered that 50 out of the 89 CVD-associated SNPs localised in macrophage-specific enhancer rich areas are eQTLs in macrophages. I confirmed that 28 of the genes that overlap with the 17 enhancer regions listed in **Table 5.1.** are eGenes of at least one of these eQTLs. eGenes are reported in **Table 5.1.** For instance, all GWAS SNPs at a region in chromosome 8, which is enriched with macrophage-specific active enhancers, are also eQTLs in macrophages, and their genotype is correlated with the expression of LPL gene **(Fig. 5.3, Table 5.2).**

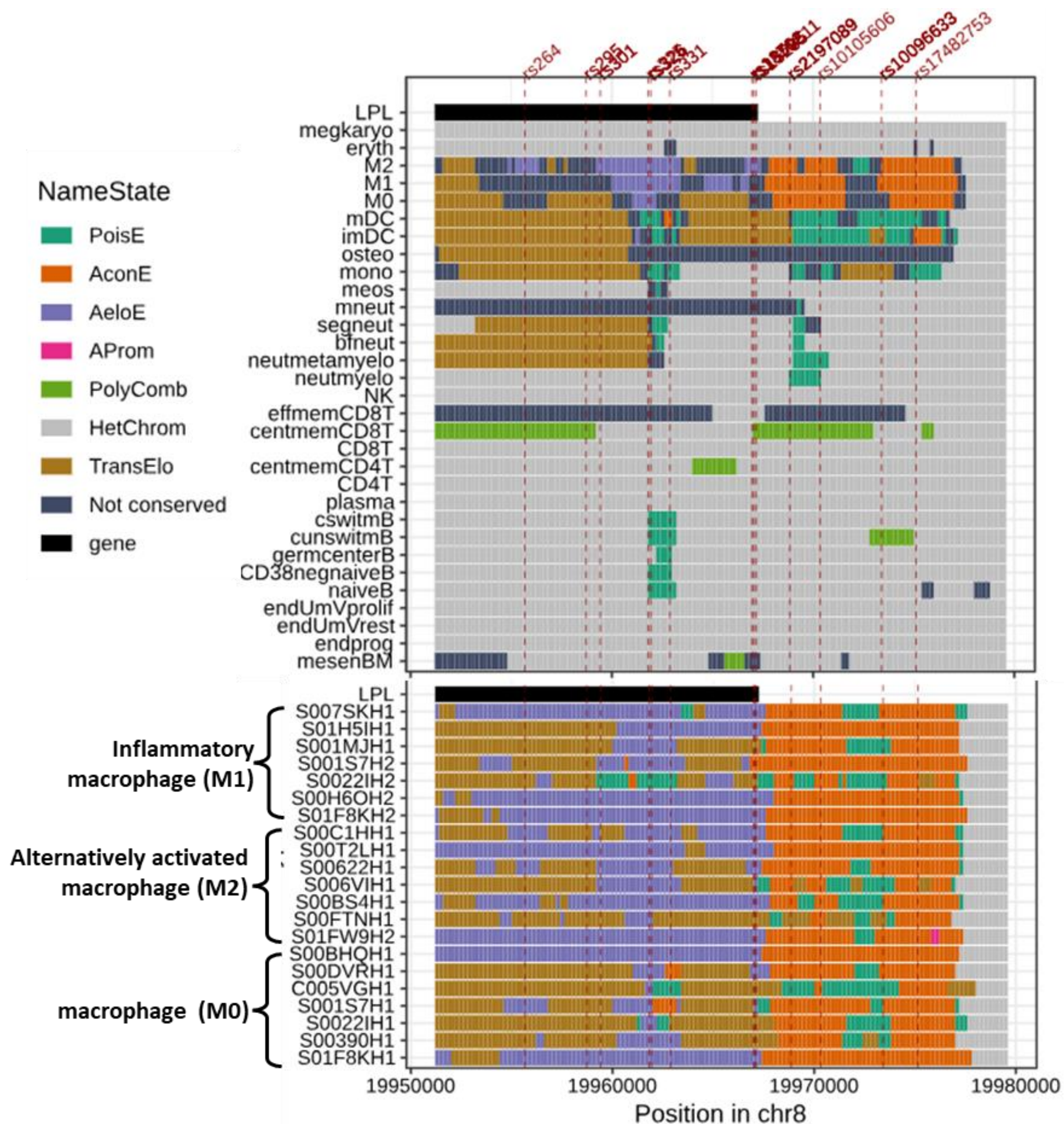


Figure 5. 3. Example of the *LPL* region enriched in macrophage enhancers and eQTLs. The figure displays the consensus chromatin states across 31 BLUEPRINT cell types (upper panel) and the individual biological replicates for macrophages (bottom panel) at a region on chromosome 8. This region contains 12 SNPs associated with lipid metabolism-related traits (see Table 5.2.), all of which overlap with active enhancers specific to macrophages (coloured orange or purple). The enhancer-rich region partially overlaps with *LPL*, and all GWAS SNPs are also eQTLs of *LPL* in macrophages.

Table 5. 2. GWAS traits associated with non-coding SNPs in a region enriched with macrophage enhancers and overlapping *LPL* gene.

rsID*	Variant	GWAS trait
rs264	chr8_19955669_G_A	Coronary artery disease
rs295	chr8_19958727_A_C	Metabolic syndrome
rs301	chr8_19959423_T_C	Metabolic syndrome (bivariate traits)
rs325	chr8_19961817_T_C	HDL cholesterol
rs326	chr8_19961928_A_G	HDL cholesterol, Triglycerides
rs331	chr8_19962894_G_A	Lipid metabolism phenotypes
rs13702	chr8_19966981_T_C	HDL Cholesterol - Triglycerides (HDLC-TG)
rs1059611	chr8_19967052_T_C	Lipid metabolism phenotypes
rs15285	chr8_19967156_C_T	Triglycerides-Blood Pressure (TG-BP)
rs2197089	chr8_19968862_G_A	Metabolic syndrome (bivariate traits)
rs10105606	chr8_19970337_C_A	Triglycerides
rs10096633	chr8_19973410_C_T	HDL cholesterol, Metabolic traits, Triglycerides
rs17482753	chr8_19975135_G_T	HDL cholesterol
rs2083637	chr8_20007664_A_G	HDL cholesterol
rs1441756	chr8_20010875_A_C	Metabolic syndrome (bivariate traits)

5.2.3 Expression patterns of enhancer's target genes.

In the previous sections, I found regions associated with CVD risk enriched in macrophage enhancers, and I used eQTL data to find their potential gene targets. In this section, I wanted to see if the expression of these genes was relevant in macrophages alone or in other cell types as well. Therefore, I used RNAseq data from BLUEPRINT (Adams et al., 2012) to analyse the expression patterns of these genes in the 31 haematopoietic cell types considered in our dataset. I identified a group of genes (*ABCA1, NR1H3, PNPLA3, ALDH1A2, STARD8, APOE, APOC1* and *LPL*) with higher expression in macrophages compared to other haematopoietic lineage cell types **(Fig. 5.4).** And a set of genes that show a high expression in macrophages but exhibit a higher expression in other cell types, such as *CPEB4* in M2 and erythroblasts (eryth) and *PLTP* in M1 and osteoclasts (osteo) and *PLP3P3* in mesenchymal cells.

I consulted the Human Protein Atlas (Thul & Lindskog, 2018) to corroborate the analysis of the expression patterns of eGenes associated with CVD risk-associated eQTL SNPs in regions containing macrophage enhancers. *NR1H3, APOE, APOC1, LPL, PLTP*, *STARD8* and *ABCA1* are prominently expressed in various types of macrophages, including blood and tissue residents such as Hoffbauer cells, Langerhans cells, and Kupffer cells. These genes exhibit robust expression in blood, adipose tissue or liver. *PNPLA3, PLPP3, CPEB4* and *ALDH1A2* are expressed in various tissue-resident macrophages. However, their expression in other cell

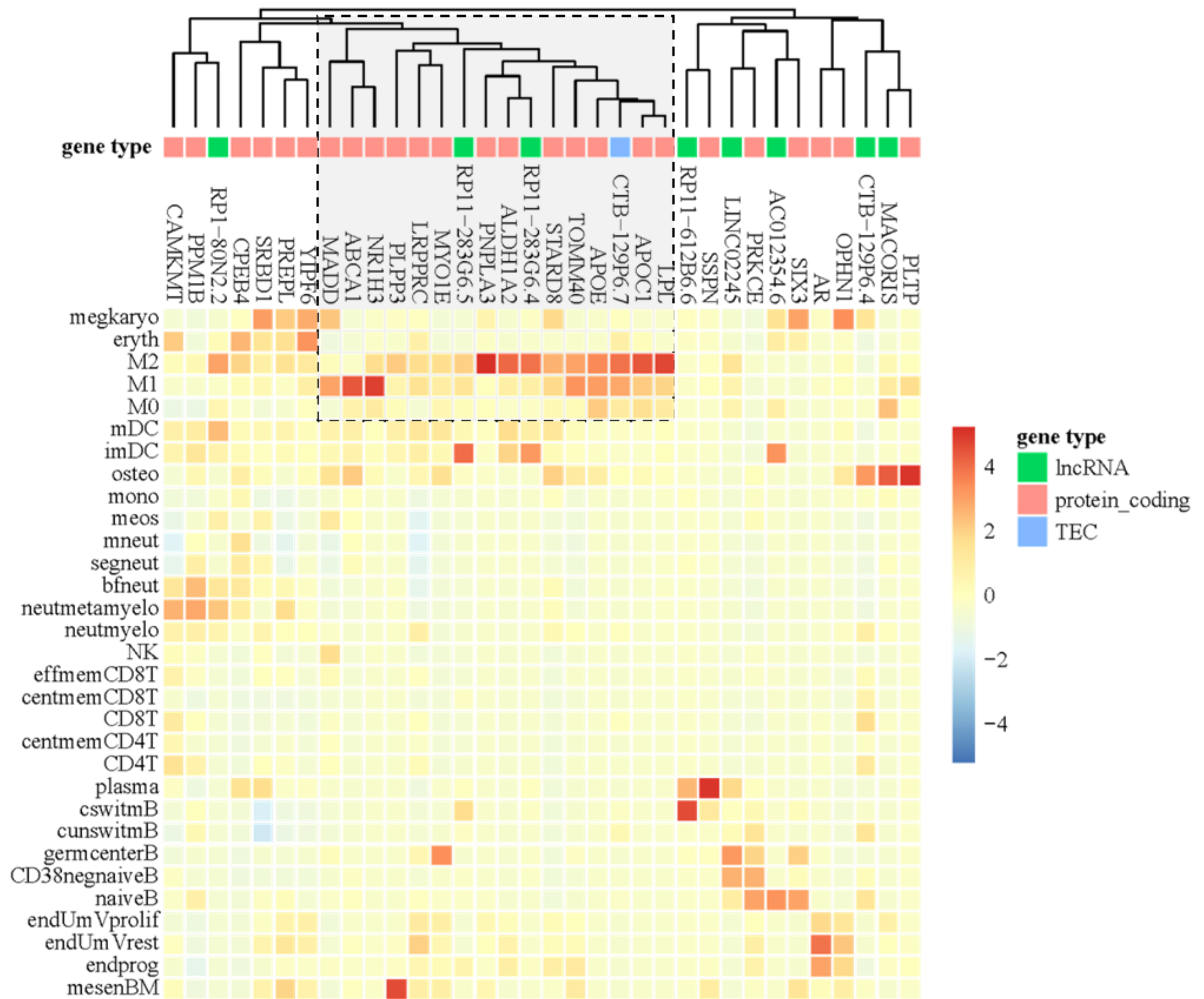


Figure 5. 4.Expression patterns in haematopoietic cells of the potential gene targets of macrophage enhancers associated with CVD risk. Heatmap shows the average of the expected counts among the cell type replicates for the 31 haematopoietic cell types from BLUEPRINT Consortium.

types surpasses the levels of expression observed in macrophages. *PNPLA3* is notably expressed in the liver whereas *PLPP3* expression is particularly associated with connective tissue, Leydig cells and fibroblasts, *ALDH1A2* is more highly expressed in tissues such as endometrium and fallopian tube, particularly in endometrial stromal cells and *CPEB4* is highly expressed in muscle tissues and liver **(Table A4)**.

In summary, my results emphasise that the potential target genes regulated by macrophage enhancers associated with CVD are expressed in relevant tissue and cell types such as

macrophages, different types of tissue-resident macrophages such as Hoffbauer cells, Langerhans cells, and Kupffer cells and in liver and adipose tissue.

5.3. Discussion

Lipid metabolism plays a fundamental role in maintaining the energy balance and providing essential structural components for various biological functions within the human body. There is an intricate choreography of genes, cell types and organs involved in each pathway of lipid metabolism, and genetic defects in key genes result in highly penetrant lipid disorders and high risk of developing premature atherosclerosis (Remmerie & Scott, 2018; Stein et al., 2019; Kathiresan & Srivastava, 2012; Tada et al., 2022). There is also evidence that common variants in non-coding regions of the genome are associated with risk factors of CVD, such as dyslipidaemia and metabolic syndrome (Teslovich et al., 2010; Keebler et al., 2010; Willer et al., 2013; Heller et al., 1993; Lanktree et al., 2009; Keebler et al., 2009) However the function of such variants remains elusive.

Our results from connecting enhancer activity profiles with susceptibility to complex traits **(Chapter 4)** revealed that non-coding genetic variants associated with CVD risk are enriched in enhancer regions active in macrophages. Leveraging macrophage eQTL and gene expression data, in this chapter, I identified genes that could be acting as potential targets of these enhancers (*APOE, APOC1, LPL, ABCA1, PLTP, NR1H3, STARD8, CPEB4, PLPP3, PNPLA3, and ALDH1A2*), by performing a literature review on the functions of all the genes identified in **sections 5.2.1** and **5.2.2** as potential enhancer targets in macrophages. I found that a set of them (*APOE, APOC1, LPL, ABCA1, PLTP*, *NR1H3)* encode well-known players in three pathways of lipid metabolism: *ApoE* and *LPL* play roles in the exogenous pathway, which manages the absorption and transport of dietary lipids; *ApoC1* and *LPL* participate in the endogenous pathway, responsible for the synthesis and transport of liver-produced lipids; *PLTP* and *ABCA1* are involved in the reverse cholesterol transport pathway (RCT), which facilitates the removal of excess cholesterol from peripheral tissues, returning it to the liver for recycling and clearance, and ultimately excreting the excess through faeces (Remmerie & Scott, 2018). Additionally, *NR1H3* is a transcription factor regulating the expression of all these five genes. The proteins encoded by *NR1H3, LPL, ABCA1, PLTP, APOE*, and APOC1 have established associations with CVD pathogenesis (Rouland et al., 2022; Marais, 2019; Endo-Umeda et al., 2022; Babaev et al., 1999; Shao et al., 2014; Lee-Rueckert et al., 2006).

The *LPL* gene encodes the lipoprotein lipase (LPL), a crucial enzyme responsible for breaking down triglycerides into fatty acids and glycerol, primarily within microvasculature walls (Remmerie & Scott, 2018; Merkel et al., 2002; Mead et al., 2002). To ensure a balanced distribution of lipids across the body, lipoprotein particles such as chylomicrons (exogenous pathway) and VLDL (endogenous pathway) travel through the bloodstream. Once they reach their target tissues, LPL hydrolyses their content, facilitating lipid uptake for storage or energy production (Remmerie & Scott, 2018; Mead et al., 2002). This step is crucial since chylomicrons and VLDL are too large to cross the capillary endothelium in most tissues (Mead et al., 2002; Remmerie & Scott, 2018) **(Fig. 5.5. A-B)**

The *APOE* and *APOC1* genes, encode for apolipoproteins A (ApoA) and C1 (ApoC1), which facilitate the transport of lipids through the bloodstream, aid lipid metabolism enzymes, and interact with cell receptors (Linton et al., 1998; Curtiss, 2000; Jong et al., 1999; Remmerie & Scott, 2018). In the bloodstream, chylomicrons (exogenous pathway) and VLDL particles (endogenous pathway) acquire ApoE and APOC1 apolipoproteins, respectively, among others. In the exogenous pathway, chylomicrons become smaller remnants after lipolysis by LPL. ApoE on the surface of these remnants is recognised and cleared from the bloodstream by liver cells through ApoE receptors (Remmerie & Scott, 2018; Linton et al., 1998) **(Fig.5.5. A)**. In the endogenous pathway, ApoC1 acts as an inhibitor of LPL enzyme activity, slowing down the breakdown of triglycerides and allowing for their transport to peripheral tissues (Verine et al., 1989; Remmerie & Scott, 2018)**(Fig.5.5. B)**.

The *PLTP* gene encodes the Phospholipid Transfer Protein (PLTP). PLTP transfers phospholipids between different lipoproteins and HDL particles in the bloodstream (Huuskonen et al., 2001; X. C. Jiang et al., 2012). In the RCT, excess cholesterol is effluxed from cells, primarily by macrophages, with HDL serving as their physiological receptors (Hutchins & Heinecke, 2015).

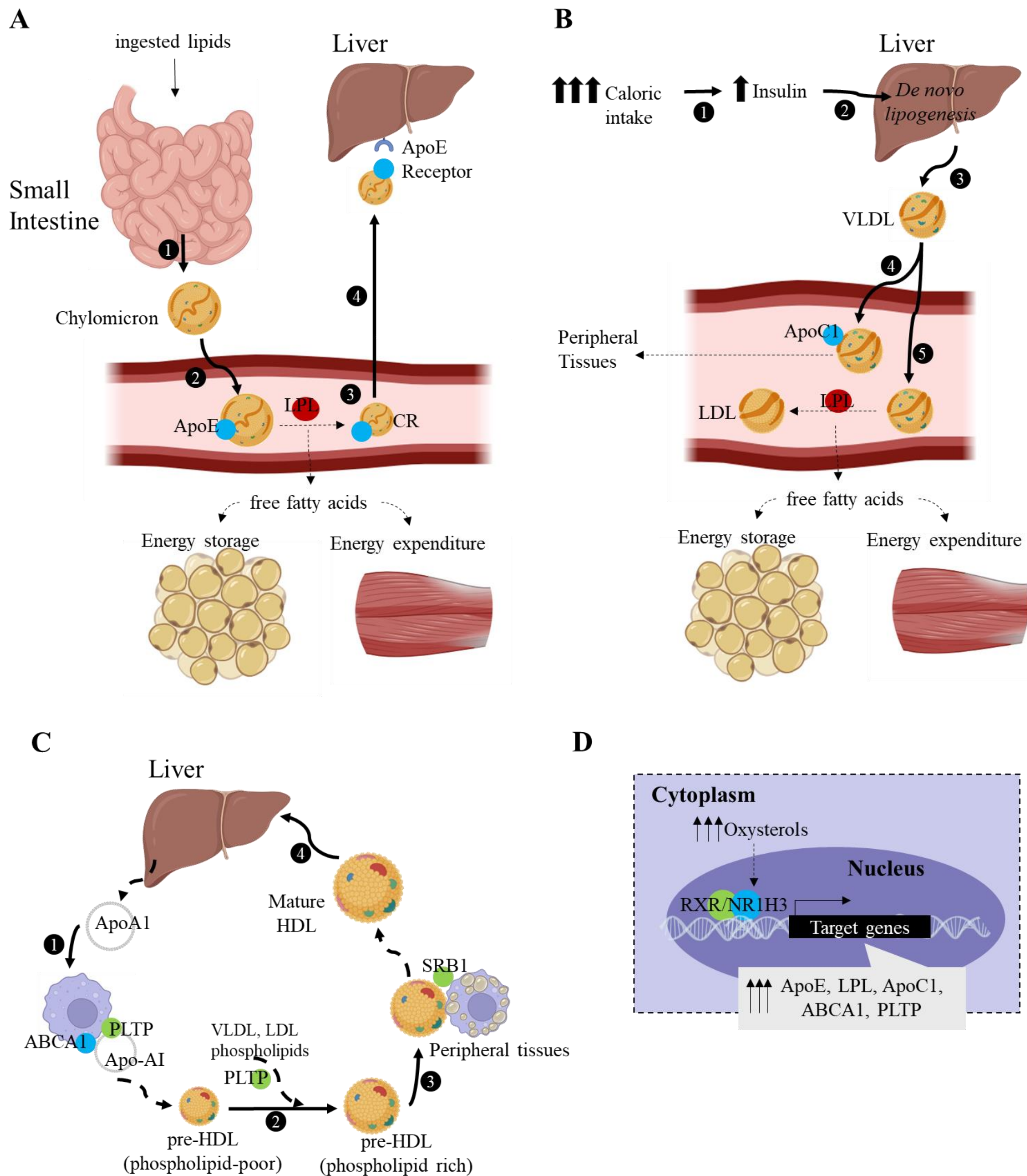


Figure 5. 5. Mapping macrophage enhancers target genes to specific steps in lipid metabolism. This figure shows a simplified version of the lipid metabolism pathways, highlighting the role of the proteins encoded by the genes potentially regulated by macrophage enhancers associated with CVD risk. **(A)** Exogenous Pathway: In the intestine, ingested lipids are packed into chylomicrons by enterocytes (Step 1). As these particles reach the bloodstream, they acquire ApoE (Step 2). Upon reaching target tissues, LPL hydrolyses their content, facilitating tissue uptake for storage or energy production (Step 3). Post lipolysis, chylomicrons become chylomicron remnants (CR) (Step 4). These remnants are recognised by hepatocytes via an ApoE receptor and cleared from the bloodstream (Step 5). **(B)** Endogenous Pathway: Triggered by excessive caloric intake, insulin concentration in the blood increases; this pathway begins with insulin signalling the liver to start lipogenesis (Step 1). Liver-synthesized lipids are packaged into VLDL (Step 2), which, in circulation, can undergo lipolysis in target tissues by LPL (Step 3) or acquire APOC1, inhibiting LPL and allowing VLDL to travel further to peripheral tissues (Step 4). **(C).** Reverse Cholesterol (RCT) Pathway: ApoAI synthesised by

the liver acquires phospholipids from macrophages via ABCA1 and PLTP (Step 1). Remodelled particles become preHDL, acquiring more phospholipids from other lipoproteins in circulation and increasing their size (Step 2). Phospholipids from PreHDL particles sequester cholesterol effluxed from macrophages via SRB1 (Step 3) and, after collecting cholesterol, become mature HDL particles that are cleared from the bloodstream by the liver (Step 4). D. Master Regulator of Lipid Metabolism: NHR1H3, also known as LXR alpha is the master regulator's role in orchestrating these pathways, this TF controls the transcription of *ABCA1, APOE, LPL, APOC1, PLTP* and other several genes that are key players in lipid metabolism.

The addition of phospholipids to HDL by PLTP in circulation results in the formation of large particles that can efficiently accept and transport more cholesterol from peripheral tissues, and that are preferentially taken up by the liver (Mulya et al., 2008) **(Fig.5.5. C).**

PLTP also plays a crucial role in the formation of Pre-β HDLs, which are precursors to mature HDL particles. This process involves an initial interaction between apoA-I and the cellular ATP-binding cassette transporter A1 (ABCA1) (Mulya et al., 2008). ABCA1, acting as a phospholipid translocase, facilitates the efflux of phospholipids from the cell (Mulya et al., 2008; Wang et al., 2001). PLTP and apoA-I bind to ABCA1 at closely related sites, and PLTP transports phospholipids from the cell membrane to apoA-I (Oram et al., 2003). The addition of phospholipids to apoA-I results in the formation of functional pre-β HDL particles capable of removing excess cholesterol from cells (Mulya et al., 2008). Phospholipids trap cholesterol released from the plasma membrane, and the transport of cholesterol from cells to HDL phospholipids is facilitated by a cell surface receptor called scavenger receptor B1 (Jian et al., 1998) **(Fig.5.5. C).**

Finally, the *NR1H3* gene encodes the transcription factor liver X receptor alpha (LXRα), which operates in a heterodimeric partnership with retinoid X receptor (RXR), to regulate the expression of genes related to lipid metabolism (Repa et al., 2000). LXRα is activated by oxysterols, which are endogenous metabolites of cholesterol (Janowski et al., 1996). All of the genes described above (*ABCA1, APOE, LPL, APOC1, PLTP*) are targets of LXRα (G. Cao et al., 2002; Mak et al., 2002) **(Fig.5.5. D).** LXRα also regulates the expression of other genes that are crucial to lipid metabolism, such as ABCG1, fatty acid synthase and ApoC2 (Mak et al., 2002) and cholesteryl ester transfer protein (G. Cao et al., 2002). Besides regulating lipid metabolism, LXRα also plays a role in immunity regulation, exhibiting anti-inflammatory effects (Endo-Umeda et al., 2022; Matalonga et al., 2017). LXRα also acts as an important regulator of adrenal cholesterol homeostasis through its ability to modulate the transcription of genes that govern the three major pathways of adrenal cholesterol utilisation. These pathways

include cholesterol efflux (*ABCA1, ABCG1*), storage *(*apoE, SREBP-1c), and conversion to steroid hormones (Steroidogenic Acute Regulatory protein, StAR) (Cummins et al., 2006).

Despite their significant roles, the enhancers regulating the expression of these genes and their activity patterns are unknown. Here, I identified a specific set of regions that hold the potential to function as enhancers for these genes, specifically in macrophages. Further validation through experimental testing in cell cultures and animal models is essential to confirm the influence of these regions on the expression of these genes, and to explore the potential effects of enhancer dysfunction in lipid metabolism. However, my contribution serves as the initial stepping stone in the crucial investigative process of unravelling the regulatory mechanisms governing these key players in cardiovascular health.

Our analysis also identified a set of genes (*PNPLA3, PLPP3, CPEB4* and *ALDH1A2, STARD8*) that have not been directly linked with macrophage metabolism; they could be involved in CVD as indirect actors in lipid metabolism or as modulators of inflammation, but their specific roles are unclear.

The *PNPLA3* gene encodes an enzyme that converts TG into fatty acids in liver cells and is tightly associated with membranes and lipid droplets (Bruschi et al., 2017). A mutant of this protein (PNPLA3 I148M) reduces cholesterol efflux in human hepatic stellate cells (Bruschi et al., 2019) and increases inflammatory responses in M1 macrophages derived from primary cultures of human monocytes (Dixon et al., 2023). These effects are associated with compromised signalling of PPARγ and LXRα, leading to the downregulation of ABCA1 (Bruschi et al., 2019).

The *STARD8* gene encodes the protein Deleted in Liver Cancer 3 (DLC3), a Rho-specific GTPase-activating protein involved in the coordination of endocytic vesicle trafficking by associating with the small GTPase Rab8 (Braun et al., 2015). This protein has been implicated in gonadal dysgenesis and steroidogenesis (Sotillos et al., 2022; Ilaslan et al., 2018). STARD8 shows the highest homology (82% protein sequence similarity) with StAR, also known as STARD1. STARD1 controls intracellular cholesterol movement from the outer to the inner mitochondrial membranes (Christenson & Strauss, 2000; Stocco, 2000), a crucial step in the synthesis of pregnenolone and oxysterols(Larsen et al., 2020). STARD1 is a target of LXRα (Cummins et al., 2006). Dramatic accumulation of cholesterol in lipid droplets has been identified after STARD1 deletion (Ishii et al., 2002)

The *ALDH1A2* gene encodes the enzyme retinaldehyde dehydrogenase 2 (RALDH2) that catalyses the oxidation of retinaldehyde to all-trans-retinoic acid (ATRA) (Kedishvili, 2016).

ATRA functions as a ligand for the retinoic acid receptor (RAR) (Allenby et al., 1993). RAR, in heterodimeric partnership with RXR, can activate the ABCA1 in human macrophages via the same promoter element as LXR/RXR (Costet et al., 2003). In THP1 cells and primary monocytes, ATRA induces the increased expression of scavenger receptor CD36, mediated by RAR signalling (Wuttge et al., 2001). Oral administration of ATRA significantly decreased serum total cholesterol, LDL-cholesterol levels and the size of atherosclerotic plaques in a rabbit model of fat-induced atherosclerosis (Zarei et al., 2018).

The *PLPP3* gene codifies the membrane protein Lipid Phosphate Phosphatase (LPP3) enzyme (Mueller et al., 2019). Among its substrates is lysophosphatidic acid (LPA); LPP3 transforms LPA into monoglycerides (L. Chen et al., 2017; Pyne et al., 2004). A study using porcine and murine experimental models of atherosclerosis showed that LPP3 levels were increased in atherosclerosis and LPP3 absence in smooth muscle cells heightened plaque formation, inflammation, and LPA levels in atheroma (Mueller et al., 2019). Another study showed that treatment with LPA of human primary monocytes and mice monocytes from the spleen and bone marrow induces differentiation into macrophages with a more proinflammatory phenotype upon exposure to LPS than MCSF-induced macrophages (Ray & Rai, 2017). Chen et al. (2017) revealed that LPA treatment of the murine macrophage cell line increased the uptake of ox-LDL by upregulating the scavenger receptor A, which, in turn, promoted the formation of foam cells. Moreover, LPA treatment led to the downregulation of scavenger receptor class B type I (SRBI), a protein essential for cholesterol efflux (L. Chen et al., 2017)

Finally, the CPEB4 gene encodes an mRNA-binding protein that controls the translation of several mRNAs via binding to their mRNA and stabilising them in the cytoplasm (Pell et al., 2021; Cui et al., 2020). Suñer et al. (2022) showed that CPEB4 is involved in the temporal control of inflammation by helping to stabilise anti-inflammatory transcripts, allowing for their sustained expression in late times following macrophage LPS stimulation. They also showed that CPEB4 depletion in the sepsis model impairs the inflammation resolution of macrophages (Suñer et al., 2022).

In my study, I utilised eQTL data to establish a link between genes and GWAS variants located in enhancers, a strategy that led to the identification of relevant and meaningful target genes. However, it is crucial to note that while my findings are significant, the use of eQTL data may not always be the most appropriate approach. There are alternative methods for enhancer-gene linking, which I discuss in detail in **Section 6.2.1**.

According to a recent study by Mostafavi et al. (2023), only 43% of GWAS hits are colocalised with eQTLs. The authors discussed that the remaining GWAS hits may act as eQTLs only in specific contexts or may act on different mechanisms than gene expression regulation, such as splicing and polyadenylation. For GWAS SNPs potentially impacting gene regulation that are not detected by eQTL analysis, they investigated the differences between eQTLs and GWAS nearest genes. They found that eQTLs predominantly (43%) cluster within a median distance of 13 kb to the nearest TSS, while GWAS SNPs often (78%) lie at greater distances from TSSs (36 kb). In terms of enhancer regions, GWAS loci overlap with longer enhancer regions than eQTLs loci. Moreover, GWAS-proximal genes use more alternative TSS across different cell types, averaging 6.4 TSSs per gene, compared to 4.4 for eQTL genes (Mostafavi et al., 2023).

In my study, I chose to focus on macrophages because their enhancers represent a substantial proportion (52%) of the haematopoietic enhancers that I found associated with CVD. While macrophages are clearly relevant, it is essential to broaden our investigation to other cell types. For example, we observed that endothelial cells are implicated in several other enhancer activity profiles and these genomic regions warrant further exploration to understand the genetic associations with CVD fully. Additionally, other cells not included in BLUEPRINT, such as vascular smooth muscle cells, may also be worth exploring (more details on their role can be found in **section 6.2.2**) .

In a recent study, Schnitzler et al. employed a variant-to-gene-to-program (V2G2P) approach to link non-coding, non-lipid-related GWAS variants associated with coronary artery disease (CAD) to relevant biological pathways (Schnitzler et al., 2024). They utilised bulk RNA-seq, ATAC-seq, and H3K27ac ChIP-seq to identify active genes and candidate enhancers in endothelial cells (ECs) with CAD variants. They used the ABC model (Fulco et al., 2019) to identify the target genes of the EC enhancers. Subsequently, they employed a CRISPRi-based method to systematically silence the ABC-identified candidate genes, using single-cell RNA sequencing as a readout (Perturbseq). Furthermore, they used unsupervised machine learning to identify disease-associated gene 'programs' or coexpressed gene modules. Using this comprehensive approach, the study identified 41 CAD-associated genes in ECs that participate in five gene 'programmes' related to extracellular matrix organisation, cell migration, and angiogenesis, among other functions. Notably, the *PLPP3* gene, previously identified in your analysis as a cardiovascular risk gene targeted by macrophage enhancers, is among these 41 CAD-associated genes and participates in two of the five identified programs in the study (Schnitzler et al., 2024).

In conclusion, by providing a comprehensive set of enhancer regions associated with CVD risk, my work establishes a foundation for hypothesis testing, which is necessary to advance our understanding of the regulatory mechanisms governing the expression of genes involved in cardiovascular health.

Chapter 6. Project contributions and future perspectives

6.1. A synopsis of project outcomes and scientific contributions.

This thesis offers novel insights into the connections between complex traits and transcriptional enhancers active in specific sets of haematopoietic cell types. Building upon the foundational effort of the BLUEPRINT Consortium, which generated the largest collection of profiles of histone marks for human haematopoietic primary cells to date (Adams et al., 2012), I have produced the largest and most comprehensive collection of haematopoietic epigenomes, including the most common cell types of myeloid and lymphoid lineages at various stages of differentiation. This dataset will facilitate the exploration of epigenomic differences during differentiation and can be used to complement other types of information for the study of the relationship between regulatory elements across the genome and specific biological processes in health and disease.

Using these epigenomic maps and the extensive information connecting genetic variants with complex traits from the GWAS catalogue, I identified sets of cell types potentially affected by defects in enhancer activity associated with around 175 complex traits. Analysed traits include cardiometabolic, vascular, immune-related traits, and cancers, among others. These results serve as a stepping stone for understanding the role of enhancers in disease by providing useful information to guide the selection of cell types to test hypotheses of disease mechanisms experimentally.

Our results pointed out a set of traits, including cardiovascular-associated diseases and risk factors, that were predominantly associated with enhancers active in macrophages. I investigated the gene targets of these enhancers and revealed a meaningful biological relationship between these genetic risk factors and specific steps in the three distinct lipid metabolism pathways. I investigated the cell types' expression of these genes and found that, in agreement with my results, the genes are expressed in tissue-resident macrophages across different tissues that participate in the lipid metabolism pathways. This example showcases how identifying specific cell types and genes affected by genetic defects in enhancers could improve our understanding of the biological processes underlying the disease, providing a refined starting point for designing experimental models crucial to understanding disease mechanisms.

Our findings deepen our comprehension of the role of enhancers in susceptibility to complex diseases by guiding the selection of cell types in experimental model design. However, there is ample room for future improvement. In this chapter, I will discuss our limitations and propose

future research that could build upon our findings to elevate their impact, with the ultimate goal of translating these insights into potential cures or treatments for complex diseases.

6.2. Future perspectives.

6.2.1. Finding gene targets for all the identified disease-associated enhancers.

In Chapter 4, I identified cell types enriched in enhancers in genome regions associated with several complex traits. In the final section of Chapter 4 and in Chapter 5, I used cell type-specific eQTL data to identify potential target genes regulated by enhancers active in neutrophils and macrophages associated with UC and CVD risk, respectively. However, I did not search for the gene targets of the rest of the enhancers associated with the diseases. The logical progression following the outcomes elucidated in this PhD project would be the identification of potential genes regulated by all disease-associated enhancers in the relevant cell types pinpointed by my analysis.

Various methodologies can be employed for inferring enhancer-gene regulatory interactions, including the utilisation of data on chromatin 3D contacts (Avsec et al., 2021; Karbalayghareh et al., 2022; Whalen et al., 2016), correlations between DNA accessibility and gene expression (Boix et al., 2021; Sheffield et al., 2013; Thurman et al., 2012), and the use of eQTL data (Y. Liu et al., 2017). Predictive models that combine multiple types of information and CRISPRi-based enhancer experimental data result in higher precision predictions than strategies relying solely on one type of information (Fulco et al., 2019; Luo et al., 2023).

The "Activity by Contact" (ABC) model, a highly precise machine learning model trained on CRISPRi-based enhancer screening data and designed to infer enhancer-gene regulatory links, has identified two essential features crucial for accurate predictions: the strength of activating chromatin marks at an element, also known as 'enhancer activity,' detected through DNase-seq experiments; and the frequency at which an enhancer establishes physical contact with a promoter, referred to as '3D interaction frequency.' This frequency is determined through Hi-C experiments (Fulco et al., 2019). These methods assume that enhancers have additive effects on gene expression and that their distance to genes doesn't significantly impact these effects.

More recently, Engreitz et al. (2023) introduced the ENCODE-rE2G models, which are also supervised machine learning models trained on large-scale genetic perturbation experiments. These models challenge the assumptions made by the ABC model by revealing that, in addition to 'enhancer activity' and '3D interaction frequency', factors such as promoter class and

enhancer-enhancer synergy play crucial roles in determining the impact of enhancers on the expression of their target genes (Gschwind et al., 2023). Concerning promoter class, they found that promoters of genes that are expressed ubiquitously tend to be less responsive to distal enhancers. They also observed that the proximity between enhancers is associated with the size of their impact on gene expression. Enhancers located within 5 Kb of each other have a super-additive effect on gene expression. Moreover, changes in the activity of one enhancer can impact the activity of other enhancers in its vicinity (Gschwind et al., 2023).

Engritz et al.'s findings align seamlessly with my strategy of identifying trait-related regions enriched in enhancer regions rather than focusing on regions enriched in GWAS variants. This approach is grounded in the idea that if enhancer regions act super-additively, collectively impacting gene expression, disruptions in enhancer-rich regions could indeed have more profound effects on phenotype changes.

Applying Engreitz et al.'s models to the haematopoietic cell types in the BLUEPRINT dataset could significantly contribute to uncovering the target genes of the disease-associated enhancers identified. However, as the ENCODE dataset, where the ENCODE-rE2G model was applied, does not cover the primary immune cell types studied by the BLUEPRINT Consortium, the absence of DNase-seq data poses an obstacle, hindering the immediate replication of the ENCODE-rE2G model in this specific context. The future generation of this data for the cell types in the BLUEPRINT dataset could address this challenge.

6.2.2. Enriching the dataset with other relevant cell types.

In this PhD project, my focus was on studying the epigenomes of haematopoietic cells to gain insights into the sets of cell types associated with complex diseases such as cardiometabolic, vascular, immune-related, neurocognitive, and cancers. However, it is crucial to acknowledge that many of these diseases exhibit pathogenic phenotypes in cell types that are not represented in our dataset. Thus, while haematopoietic cells may contribute to the development of these diseases, it is possible that in other cell types, biological processes equally or even more critical than those occurring in haematopoietic cell types may take place.

For instance, in Chapter 4, I identified neutrophil enhancers associated with UC and my results highlighted the involvement of their target genes in NETosis—a process extensively associated with UC—underscoring the significance of my findings. Nevertheless, it is imperative to acknowledge that prior studies have also unequivocally indicated the involvement of other cell types in UC aetiology, such as intestinal innate lymphoid type 3 cells (ILC3) (Malysheva et al.,

2022; B. Zeng et al., 2019) and intestinal epithelial cells (Kong et al., 2023; van Unen et al., 2022), which are not included in our dataset.

In Chapter 5, I discussed another example—CVD. My investigation into enhancers associated with Cardiovascular risk emphasised their predominant activity in macrophages, a cell type that has been previously implicated in both the progression and regression of atherosclerosis—the main underlying cause of CVDs (Gisterå & Hansson, 2017). Additionally, I mapped the function of the target genes for these enhancers to specific steps in the three different lipid metabolism pathways. My findings contribute valuable information within the context of the included cell types. However, it is worth noting that other critical cell types in the development of atherosclerosis, such as vascular smooth muscle cells (VSMC) (Harman & Jørgensen, 2019; Xue et al., 2022), are not included in our dataset, emphasising the need for future studies to encompass a broader range of cellular contributors for a comprehensive understanding of complex disease pathogenesis.

Integrating into my analysis the chromatin state profiles of VSMCs from aortic tissue, as well as ILC3 and epithelial cells from intestinal tissue, may uncover the potential roles of their enhancers in the aetiology of CVDs and IBD respectively. However, performing additional ChIP-Seq experiments on these cell types is difficult in practice. Obtaining samples from aortic and intestinal tissues requires invasive procedures, and both tissues are composed of heterogeneous cell populations, making it challenging to isolate specific cell types. Moreover, ILC3 is a rare population (Malysheva et al., 2022), as is the VSMC clone directly linked to atherosclerosis progression (Dobnikar et al., 2018). The scarcity of these cell types poses a significant hurdle in obtaining a sufficient number of cells for ChIP-Seq assays. In light of these complexities, employing single-cell methodologies (Grosselin et al., 2019) and developing novel analysis strategies may become necessary for this type of research.

Meanwhile, many published histone mark profiles exist for tissues relevant to complex diseases. The most straightforward next step in my research would be to extend my strategy to identify disease-associated enhancers in those cell types and tissues.

In the coming year, the IHEC consortium (Stunnenberg et al., 2016b) aims to provide the research community with a comprehensive set of reference epigenomes, that will be called the EpiAtlas. This initiative involves collecting data originally published by independent epigenomic consortiums worldwide and reprocessing these data using unified pipelines to ensure the highest quality standards in the processing steps and facilitate data comparability.

Within IHEC, our laboratory is part of the Integrative Analysis Group, contributing to the EpiAtlas collaborative effort.

We anticipate the IHEC Consortium's release of a comprehensive collection of chromatin state maps as part of the EpiAtlas, holding great potential for enhancing my analyses. For instance, they are reprocessing ChIP-Seq data from a range of tissues highly relevant to studying the aetiology of UC, such as colon epithelial cells, mucosa in the rectum, large intestine, and duodenum, as well as tissues from the small and large intestine, sigmoid, ascending, and transversal colon, colonic muscle layer, rectal smooth muscle tissue, and Peyer's patch tissue from healthy individuals. Additionally, the consortium will release data on smooth muscle cells, hepatocytes, and tissues from the liver, thoracic aorta, right lobe of the liver, tibial artery, and ascending aorta—tissues highly relevant to lipid metabolism and, consequently, CVD.

The implications of extending my research strategy to these alternative cellular contexts are monumental. This approach has the potential to reveal genes and regulatory mechanisms underlying complex diseases in cell types that may be more relevant to the aetiology, and we may even discover crucial cell types for disease pathogenesis that were previously obscured.

6.2.3. Experimental validation of genetic susceptibility.

The findings of this PhD project provide valuable insights into potential enhancer activity profiles associated with various complex diseases. This aids in unravelling disease mechanisms by guiding the selection of relevant cell types for experimentation. For instance, in Chapter 5, I discovered that enhancers linked to cardiovascular risk are predominantly active in macrophages. Consequently, conducting experiments in this specific cell type will be essential to validate the role of these enhancers in the aetiology of CVD.

To that end, *in vitro* experiments in cell cultures offer a convenient starting point before progressing to more complex models. For instance, macrophages derived from THP1, a human leukaemia monocytic cell line, or macrophages derived from primary monocytes isolated from the blood of human donors, could be used. The advantage of using cell lines is that they have acquired the non-physiological ability to proliferate indefinitely due to accumulated genetic mutations. However, experimental results may be misleading since the cell line does not reflect entirely the original physiological properties of the cell type. Alternatively, primary cells better resemble the natural state of the cell type in the organism; they maintain their biological identity and can only propagate for a few generations *in vitro* (Arango et al., 2013; Kaur & Dufour, 2012).

In addition to identifying that macrophage enhancers were associated with cardiovascular CVD risk, I took an extra step by identifying their potential gene targets and establishing their links with the disease based on information from the existing literature **(See sections 5.2.4 and 5.2.5)**. Existing literature supports a relationship to CVD for a subset of these enhancer target genes, namely, *APOE, APOC1, LPL, ABCA1, PLTP,* and *NR1H3.*

For genes with unclear roles in CVD, such as *PNPLA3, PLPP3, CPEB4, ALDH1A2*, and *STARD8*, assessing phenotypic changes when their function is directly disrupted could provide insights into their roles in CVD. Techniques such as CRISPR/Cas9 or the use of inhibitors targeting the encoded enzyme could be employed for this purpose (H. Li et al., 2020; Tsai et al., 2022; H. X. Wang et al., 2017).

For the identified genes with well-known links to CVD, disrupting the function of enhancers that potentially influence their expression might offer more informative insights. This can be accomplished, for instance, by using CRISPR/Cas9 technology to target TF binding sites within enhancers or by deleting entire enhancer regions with guide RNAs that target specific sequences on both ends of the enhancer (T. Jiang et al., 2021; Osterwalder et al., 2018; Spicuglia et al., 2017).

Editing primary cells is feasible, but it can be rather challenging, especially in macrophages, since they have an innate mechanism to resist foreign genetic material (perceived as signs of infection). The resulting degradation of CRISPR components can result in low editing efficiencies. Maintaining cell viability and function during genetic manipulations is also a challenge, as they are freshly isolated from donor tissue and stressed during this process, and are highly sensitive to any changes in their growth conditions. Lastly, the limited availability of samples from donors may restrict the number of attempts to perform genetic manipulations (Distler et al., 2005; Dudek & Porteus, 2021; Freund et al., 2020)

Further validation in more physiologically relevant tissues, such as macrophages from arterial walls, gut, liver, and adipose tissues, is imperative for comprehending the tissue-specific effects of macrophage enhancer activity in atherosclerosis, but obtaining tissue-resident macrophages from relevant organs is challenging in humans.

Mouse models could be employed to investigate the *in vivo* effects of enhancer disruption, enabling the assessment of tissue-specific responses and observation of systemic effects on the immune system and overall health (Bhatia et al., 2021; Hollingsworth et al., 2023), thereby enhancing the translational relevance of the findings. However, these experiments will require

a previous step for identifying the orthologues of the human enhancers in the model species; fortunately, there are already catalogues of mice macrophage enhancers that could be explored to that end (Denisenko et al., 2017).

In summary, the findings presented in this dissertation hold therapeutic potential, as the feasibility of targeting identified enhancers or their target genes for therapeutic interventions could lead to novel treatment strategies. However, verifying if these enhancers, cell types, and genes play a sufficiently important role in CVD development or progression is an essential early step towards translation.

6.2.4. Considering the influence of genetic-environmental interactions

Throughout this PhD project, my primary focus has been uncovering genetic susceptibility to diseases within haematopoietic enhancers. Nevertheless, I must acknowledge that the development of complex diseases is not solely dependent on genetic factors.

Numerous environmental factors encountered in daily life, such as the microbiome, pollutants, viral infections, and climate change, have been associated with disease risk (Virolainen et al., 2022). For instance, risk factors for CVDs, such as blood levels of lipoproteins and apolipoproteins, are influenced by various lifestyle factors, including diet (fish oil, saturated fatty acids, waist-hip ratio), physical activity (frequency, duration, and intensity), and alcohol consumption (Hartiala et al., 2021; Laville et al., 2022; Lee et al., 2011). Mode of birth, diet, childhood antibiotic use, and alterations in gut microbiota are closely associated with the initiation or progression of IBD (Ananthakrishnan et al., 2017; A. Z. Yang & Jostins-Dean, 2022). Smoking (both current and previous) elevates the risk for CD, while a previous appendectomy provides protection against UC (Alan Z Yang, 2022).

The combined effect of genetic factors and environmental influences is often more significant than either factor alone in the development of complex diseases (Hunter, 2005). Hence, further research into the influence of environmental factors on susceptibility to diseases mediated by enhancers is necessary for devising effective strategies for prevention and treatment. In the subsequent sections, I will outline two key initiatives that could enhance our understanding of this critical question.

6.2.4.1. Identification of enhancers that may contribute to disease susceptibility only upon exposure to environmental triggers.

As immune cells mediate the body's response to environmental stimuli, our dataset is invaluable for studying the intricate interplay between genetics and the environment. When immune cells are exposed to external stimuli (environmental factors), extensive gene expression changes take place. Transcriptional states differ between unstimulated and stimulated cells, with some sets of genes being expressed transiently at specific time points. These changes shape the functional properties of cells that are necessary for an adequate response to the stimuli. (Margraf & Perretti, 2022; Soskic et al., 2022; X. Zhang et al., 2004).

The dynamic expression of specific genes in particular contexts (e.g cells under different external stimuli) is regulated by enhancers. In this PhD project, I studied the relationship between 'active enhancers' and complex diseases. However, since our dataset mainly consists of unstimulated cell types, with the exception of anti-inflammatory and pro-inflammatory macrophages, it's essential to recognise that other types of enhancers that will become active only upon environmental stimuli, referred to as 'stimulation-responsive enhancers,' escape my analysis.

These enhancers could potentially serve as crucial mediators in disease development by regulating appropriate gene expression in contexts yet to be determined. Consequently, they might harbour susceptibility that becomes apparent only when the organism is exposed to environmental factors triggering the disease (Kaikkonen et al., 2013b; Maurya, 2021)

Our research should advance to investigate stimuli-responsive enhancers in disease. Our dataset includes annotations of poised enhancers, holding characteristic histone mark profiles with repressive marks that are erased and replaced by activation marks in response to specific stimuli; these are one type of stimulation-responsive enhancers (Maurya, 2021). It would be straightforward to extend the same approach used to link active enhancers to cell types and complex traits to this poised enhancer type.

However, environmental exposure can reshape the pre-existing enhancer repertoire in differentiated cells. Some enhancers, known as "latent enhancers", lack TF binding and enhancer-specific marks in terminally differentiated cells. However, when the cells are exposed to environmental stimuli, sequential TF binding occurs, leading to the acquisition of enhancer-specific marks. These epigenetic signatures may persist even after the stimulation has ended and result in a faster and more robust response during subsequent environmental challenges

(Ostuni et al., 2013). Since latent enhancers lack typical enhancer marks in unstimulated cells, they are absent from our epigenomic annotations, and I would require additional experimental data for their identification.

For this purpose, previous methodologies could be leveraged. For example, Calderon et al. (2019) conducted ATAC-seq and RNA sequencing across various cell types before and after exposure to different stimuli, mapping stimulation-responsive elements based on the correlations between gene expression dynamics and remodelling of chromatin accessibility (Calderon et al., 2019). Additionally, Simeonov et al. (2017) developed a CRISPRa-based platform that artificially recruits potent transcriptional activators to genomic regions without enhancer marks, aiming to discover stimulation-responsive immune enhancers capable of driving target gene expression independently of actual stimulus exposure (Simeonov et al., 2017)

In summary, expanding my research to explore the effects of genetic variation on enhancer dynamics during stimulation can provide a more comprehensive understanding of the genetic susceptibility to complex diseases, bridging the gap between genetics and environmental influences.

6.2.4.2. Using epidemiological studies to identify the Environmental factors influencing disease development.

Epidemiological studies play a crucial role in understanding how environmental exposures impact disease risk, providing actionable information for public health strategies and personalised interventions. This is particularly important since modifying environmental exposures is more feasible than altering the genome itself (Virolainen et al., 2022).

To identify environmental factors influencing disease development, understanding gene–environment interactions in complex diseases is essential. This requires a comprehensive examination of both genetic and environmental components. Therefore, collecting high-quality information on environmental factors and lifestyles, along with genetic data in association studies, is imperative to explore the interactions between these elements (Hunter, 2005).

Identifying gene-environment effects in disease is challenging, partly due to the unknown or difficult-to-record nature of many environmental exposures (i.e., microbiome composition), the expected small effect sizes of these exposures, the polygenic nature of most human traits, and

the considerable multiple hypothesis burden, requiring correspondingly large sample sizes (Di Scipio et al., 2023; H. Wang et al., 2019)

Using case-only studies as an alternative to the classic case–control studies will benefit studies of gene–environment interactions, as inclusion criteria limit sample selection to individuals with the outcome of interest. A limitation of the case-only design is that a priori knowledge of causal regions of the genome is required (Flowers et al., 2012). However, if we integrate findings from previously proposed experiments in cellular and animal models, this could be feasible.

Epigenome-Wide Association Studies (EWAS) conducted on large populations offer a powerful and cost-effective tool for assessing the environmental influence on genetic regulation. Given the stability of DNA methylation marks and their reliable detection in blood samples, DNA methylation-based EWAS are particularly well-suited for large-scale studies (Yousefi et al., 2022). These studies allow for comprehensive analysis of DNA methylation across the genome of blood cells (Battram et al., 2022). Integrating the findings from methylation-based EWAS with reference epigenomes can enhance our understanding of epigenome–environment interactions. For instance, focusing on DNA methylation patterns at promoters and enhancers, EWAS can indicate whether these regulatory regions are activated or silenced in blood cells in response to environmental exposures.

Histone Acetylation Genome-Wide studies (HAWAS) are another type of EWAS based on ChIP-seq analysis of H3K27ac (Sun et al., 2016) that allow more acurate detection of differential promoter and enhancer activity r than methylation-based EWAS but are more technically challenging. However, recent technological advances in profiling chromatin marks, such as those described in **section 1.2.**, or single-cell approaches, like single-cell ATAC-seq (Lareau et al., 2019), are opening new opportunities for chromatin-based EWAS studies in the future that could help us better understand environmental-genetic interactions mediated by the epigenome.

6.3. Conclusion.

The outcomes of this PhD project advance our understanding of the intricate relationship between active enhancers and complex diseases. I successfully identified potential enhancers and haematopoietic cell types linked to many complex traits, as well as target genes for macrophage enhancers associated with CVD. Future efforts, such as extending my analysis to

include other disease-relevant cell types and tissues, identifying potential gene targets for all disease-related enhancers, and performing experimental validation in physiologically relevant tissues and animal models, will deepen our insights for potential therapeutic interventions. Furthermore, I emphasise the importance of considering genetic-environmental interactions in complex diseases by exploring enhancers responsive to environmental stimuli and utilising epidemiological data to identify environmental factors influencing disease development.

Appendix

Table A. 1. Sample metadata.

BIO-MATERIAL TYPE	SAMPLE NAME	CELL TYPE	TISSUE TYPE	DONOR SEX	DONOR ID	EPIRR
Primary Cell Culture	S00622H1	alternatively activated macrophage	venous blood	Male	S00622	IHECRE00000129.3
Primary Cell Culture	S006VIH1	alternatively activated macrophage	venous blood	Male	S006VI	IHECRE00000059.3
Primary Cell Culture	S00BS4H1	alternatively activated macrophage	venous blood	Female	S00BS4	IHECRE00000013.3
Primary Cell Culture	S00C1HH1	alternatively activated macrophage	cord blood	Female	S00C1H	IHECRE00000055.3
Primary Cell Culture	S00FTNH1	alternatively activated macrophage	venous blood	Female	S00FTN	IHECRE00000071.3
Primary Cell Culture	S00T2LH1	alternatively activated macrophage	cord blood	Male	S00T2L	IHECRE00000269.2
Primary Cell Culture	S01FW9H2	alternatively activated macrophage	venous blood	Male	S01FW9	IHECRE00001391.1
Primary Cell Culture	S004BTH2	CD34-negative, CD41-positive, CD42-positive megakaryocyte cell	cord blood	Female	S004BT	IHECRE00000105.3
Primary Cell Culture	S00VHKH1	CD34-negative, CD41-positive, CD42-positive megakaryocyte cell	cord blood	Male	S00VHK	IHECRE00000257.2
Primary Cell Culture	S00BJMH1	endothelial cell of umbilical vein (proliferating)	cord blood	Male	S00BJM	IHECRE00000184.3
Primary Cell Culture	S00DCSH1	endothelial cell of umbilical vein (proliferating)	cord blood	Female	S00DCS	IHECRE00000126.3
Primary Cell Culture	S00BJMH2	endothelial cell of umbilical	cord blood	Male	S00BJM	IHECRE00000099.3

		vein (resting)				
Primary Cell Culture	S002R5H1	erythroblast	cord blood	Male	S002R5	IHECRE00000193.3
Primary Cell Culture	S002S3H1	erythroblast	cord blood	Female	S002S3	IHECRE00000112.3
Primary Cell Culture	S00TU2H1	immature conventional dendritic cell	venous blood	Male	B270	IHECRE00001548.1
Primary Cell Culture	S00TV0H1	immature conventional dendritic cell	venous blood	Male	B271	IHECRE00001457.1
Primary Cell Culture	S00TWZH1	immature conventional dendritic cell	venous blood	Male	B272	IHECRE00001387.1
Primary Cell Culture	S001MJH1	inflammatory macrophage	venous blood	Male	S001MJ	IHECRE00000174.3
Primary Cell Culture	S001S7H2	inflammatory macrophage	venous blood	Female	S001S7	IHECRE00000043.3
Primary Cell Culture	S0022IH2	inflammatory macrophage	venous blood	Female	S0022I	IHECRE00000161.3
Primary Cell Culture	S007SKH1	inflammatory macrophage	cord blood	Male	S007SK	IHECRE00000195.3
Primary Cell Culture	S00H6OH2	inflammatory macrophage	venous blood	Male	S00H6O	IHECRE00000318.2
Primary Cell Culture	S01F8KH2	inflammatory macrophage	venous blood	Male	S01F8K	IHECRE00001293.1
Primary Cell Culture	S01H5IH1	inflammatory macrophage	cord blood	Female	S01H5I	IHECRE00001289.1
Primary Cell Culture	S001S7H1	macrophage	venous blood	Female	S001S7	IHECRE00000084.3
Primary Cell Culture	S0022IH1	macrophage	venous blood	Female	S0022I	IHECRE00000177.3
Primary Cell Culture	S00390H1	macrophage	venous blood	Male	S00390	IHECRE00000008.3
Primary Cell Culture	S00BHQH1	macrophage	cord blood	Female	S00BHQ	IHECRE00000121.3
Primary Cell Culture	S00DVRH1	macrophage	cord blood	Male	S00DVR	IHECRE00000060.3
Primary Cell Culture	S01F8KH1	macrophage	venous blood	Male	S01F8K	IHECRE00001398.1
Primary Cell Culture	C005VGH1	macrophage	venous blood	Male	C005VG	
Primary Cell Culture	S00TYVH1	mature conventional dendritic cell	venous blood	Male	B271	IHECRE00001546.1
Primary Cell Culture	S00U0LH1	mature conventional dendritic cell	venous blood	Male	B272	IHECRE00001490.1
Primary Cell Culture	S01GKTH1	osteoclast	venous blood	Male	BC2_0	IHECRE00001249.1
Primary Cell Culture	S01GMPH1	osteoclast	venous blood	Male	BC2_10	IHECRE00001471.1
Primary Cell	S00UJKH1	adult endothelial	venous blood	Female	S00UJK	IHECRE00000303.2

		progenitor cell				
Primary Cell	S013GFH1	adult endothelial progenitor cell	venous blood	Female	S013GF	IHECRE00001510.1
Primary Cell	S00VEQH1	band form neutrophil	bone marrow	Female	BM060814	IHECRE00000285.2
Primary Cell	S00G11H1	band form neutrophil	bone marrow	Male	BM220513	
Primary Cell	S00JGXH1	band form neutrophil	bone marrow	Male	BM030613	
Primary Cell	C000S5H2	CD14-positive, CD16-negative classical monocyte	venous blood	Male	C000S5	IHECRE00000027.3
Primary Cell	C00264H1	CD14-positive, CD16-negative classical monocyte	cord blood	Male	C00264	IHECRE00000135.3
Primary Cell	C004SQH1	CD14-positive, CD16-negative classical monocyte	venous blood	Female	C004SQ	IHECRE00000101.3
Primary Cell	C005PSH2	CD14-positive, CD16-negative classical monocyte	cord blood	Female	C005PS	IHECRE00000155.3
Primary Cell	S000RDH2	CD14-positive, CD16-negative classical monocyte	cord blood	Male	S000RD	IHECRE00000048.3
Primary Cell	S004KBH1	CD38-negative naive B cell	venous blood	Male	S004KB	IHECRE00000125.3
Primary Cell	C005QQH1	CD38-negative naive B cell	cord blood	Female	C005QQ	
Primary Cell	S0033CH0	CD38-negative naive B cell	venous blood	Male	C003JB,C003RW,C003N3,C003QY	
Primary Cell	C00280H1	CD4-positive, alpha-beta T cell	cord blood	Female	C00280	IHECRE00001251.1
Primary Cell	C002Q1H1	CD4-positive, alpha-beta T cell	venous blood	Male	C002Q1	IHECRE00000009.3
Primary Cell	C002TWH1	CD4-positive,	venous blood	Male	C002TW	IHECRE00000075.3

		alpha-beta T cell				
Primary Cell	S000RDH1	CD4-positive, alpha-beta T cell	cord blood	Male	S000RD	IHECRE00000140.3
Primary Cell	S0018AH1	CD4-positive, alpha-beta T cell	cord blood	Female	S0018A	IHECRE00000191.3
Primary Cell	S008H1H1	CD4-positive, alpha-beta T cell	venous blood	Male	S008H1	IHECRE00000160.3
Primary Cell	S009W4H1	CD4-positive, alpha-beta T cell	venous blood	Female	S009W4	IHECRE00000194.3
Primary Cell	S007DDH2	CD4-positive, alpha-beta T cell	venous blood	Female	S007DD	
Primary Cell	S007G7H4	CD4-positive, alpha-beta T cell	venous blood	Male	S007G7	
Primary Cell	C002YMH1	CD8-positive, alpha-beta T cell	cord blood	Female	C002YM	IHECRE00000035.3
Primary Cell	C0066PH1	CD8-positive, alpha-beta T cell	cord blood	Female	C0066P	IHECRE00000076.3
Primary Cell	S00C2FH1	CD8-positive, alpha-beta T cell	cord blood	Male	S00C2F	IHECRE00000022.3
Primary Cell	S014WGH1	CD8-positive, alpha-beta T cell	venous blood	Female	S014WG	
Primary Cell	C002TWH2	central memory CD4-positive, alpha-beta T cell	venous blood	Male	C002TW	IHECRE00000102.3
Primary Cell	S0155TH1	central memory CD8-positive, alpha-beta T cell	venous blood	Male	S0155T	
Primary Cell	S00YPTH1	class switched memory B cell	venous blood	Male	S00YPT	IHECRE00001540.1
Primary Cell	S015BHH1	class switched memory B cell	venous blood	Male	csMBC pool 2	IHECRE00001255.1

Primary Cell	S015CFH1	class switched memory B cell	venous blood	Female	csMBC pool 8	
Primary Cell	S005YGH1	cytotoxic CD56-dim natural killer cell	cord blood	Male	S005YG	IHECRE00000049.3
Primary Cell	C00504H1	cytotoxic CD56-dim natural killer cell	venous blood	Female	C00504	
Primary Cell	S01E4WH0	cytotoxic CD56-dim natural killer cell	cord blood	Female	S01DWH,S01DWH,S01DWH,S01DWH,S01DWH	
Primary Cell	C003UQH1	effector memory CD8-positive, alpha-beta T cell	venous blood	Male	C003UQ	IHECRE00000010.3
Primary Cell	C0054XH3	effector memory CD8-positive, alpha-beta T cell	venous blood	Female	C0054X	IHECRE00000017.3
Primary Cell	S00Y9OH1	germinal center B cell	tonsil	Female	T14_10	IHECRE00000332.2
Primary Cell	S013ARH1	germinal center B cell	tonsil	Male	T14_11	IHECRE00001375.1
Primary Cell	S00W0DH1	germinal center B cell	tonsil	Female	T14_5	
Primary Cell	S00BKKH1	mature eosinophil	venous blood	Female	S00BKK	IHECRE00000114.3
Primary Cell	S006XEH2	mature eosinophil	venous blood	Male	S006XE	
Primary Cell	C000S5H1	mature neutrophil	venous blood	Male	C000S5	IHECRE00000109.3
Primary Cell	C0010KH2	mature neutrophil	venous blood	Female	C0010K	IHECRE00000004.3
Primary Cell	C0011IH2	mature neutrophil	venous blood	Female	C0011I	IHECRE00000159.3
Primary Cell	C00184H2	mature neutrophil	cord blood	Male	C00184	IHECRE00000095.3
Primary Cell	C001UYH1	mature neutrophil	venous blood	Male	C001UY	IHECRE00000094.3
Primary Cell	C004GDH1	mature neutrophil	cord blood	Female	C004GD	IHECRE00000124.3
Primary Cell	C12012H1	mature neutrophil	venous blood	Male	C12012	IHECRE00000178.3
Primary Cell	S00FWHH1	mature neutrophil	venous blood	Male	PB130513	
Primary Cell	S00FXFH1	mature neutrophil	venous blood	Male	PB130513	
Primary Cell	S00K5EH1	mature neutrophil	venous blood	Male	PB100713	
Primary Cell	S00K6CH1	mature neutrophil	venous blood	Male	PB100713	
Primary Cell	S00K7AH1	mature neutrophil	venous blood	Male	PB270313	

Primary Cell	S00K88H1	mature neutrophil	venous blood	Male	PB270313	
Primary Cell	S00W8YH2	mesenchymal stem cell of the bone marrow	venous blood	Unknown	S00W8Y	IHECRE00001335.1
Primary Cell	S00YAMH1	mesenchymal stem cell of the bone marrow	venous blood	Unknown	S00YAM	IHECRE00000250.2
Primary Cell	S00X9SH1	naive B cell	venous blood	Male	NC14_42	IHECRE00000280.2
Primary Cell	S00XAQH1	naive B cell	venous blood	Male	NC14_47	IHECRE00000258.2
Primary Cell	S00W1BH1	naive B cell	venous blood	Male	NC14_5	
Primary Cell	S0159LH1	naive B cell	venous blood	Female	B15_50	
Primary Cell	S00G03H1	neutrophilic metamyelocyte	bone marrow	Male	BM220513	IHECRE00000262.2
Primary Cell	S00VDSH1	neutrophilic metamyelocyte	bone marrow	Female	BM060814	IHECRE00000248.2
Primary Cell	S00JFZH1	neutrophilic metamyelocyte	bone marrow	Male	BM030613	
Primary Cell	S00VCUH1	neutrophilic myelocyte	bone marrow	Female	BM060814	IHECRE00000330.2
Primary Cell	S00FYDH1	neutrophilic myelocyte	bone marrow	Male	BM220513	
Primary Cell	S00JE0H1	neutrophilic myelocyte	bone marrow	Male	BM030613	
Primary Cell	S00Y8QH1	plasma cell	tonsil	Female	T14_10	IHECRE00000347.2
Primary Cell	S00VKEH1	plasma cell	tonsil	Female	T14_5	
Primary Cell	S00VFOH1	segmented neutrophil of bone marrow	bone marrow	Female	BM060814	IHECRE00000317.2
Primary Cell	S00G3YH1	segmented neutrophil of bone marrow	bone marrow	Male	BM220513	
Primary Cell	S00JHVH1	segmented neutrophil of bone marrow	bone marrow	Male	BM030613	
Primary Cell	S015DDH1	unswitched memory B cell	venous blood	Male	Pool_9	

* Rows highlighted in yellow indicate samples that have not been included in the EPIRR.

Table A. 2. Correlations of emission probabilities between Carrillo et al. (2027) model and ours.

*Top value for the Pearson correlation coefficient between each state (11 states) in Carrillo et al. (2027) model and our model (12 states). States with correlation coefficients below 0.75 are

State from Carrillo et.al (2017)	**State mnemonic (as in the original paper)**	**State label description (as in the original paper)**	**Best match* (correlation coefficient)**	**Best match* (state from our model)**
1	Transcription	Transcription Low Signal H3K36me3	0.74	12
2	Transcription	Transcription High Signal H3K36me3	0.63	9
3	Heterochromatin	Heterochromatin High Signal H3K9me3	1	4
4	Heterochromatin	Low signal	0.95	4
5	Heterochromatin	Heterochromatin High Signal H3K27me3	1	8
6	Heterochromatin	Heterochromatin Low Signal H3K27me3	1	8
7	Repressed Promoter	Repressed Polycomb Promoter High Signal H3K4me3, H3K4me1 and H3K27me3	0.89	1
8	Enhancer	Enhancer High Signal H3K4me1	0.85	11
9	Enhancer	Active Enhancer High Signal H3K4me1 & H3K27Ac	0.62	5
10	Active Promoter	Distal Active Promoter (2Kb) High Signal H3K4me3 & H3K27Ac & H3K4me1	0.54	2
11	Active Promoter	Active TSS High Signal H3K4me3 & H3K27Ac	0.76	2

highlighted in red.

Table A. 3. Correlation of emission probabilities between the NIH Epigenomic Roadmap (2015) model and ours.

State from Roadmap (2015)	State mnemonic (as in original paper)	State label description (as in original paper)	Sest match* (correlation coefficient)	Best match* (state from our model)
1	TssA	Active TSS	0.94	7
2	TssFlnk	Flanking TSS	1	6
3	TssFlnkU	Flanking TSS Upstream	0.89	7
4	TssFlnkD	Flanking TSS Downstream	0.69	6
5	Tx	Strong transcription	0.87	11
6	TxWk	Weak transcription	0.86	11
7	EnhG1	Genic enhancer1	0.94	12
8	EnhG2	Genic enhancer2	0.97	10
9	EnhA1	Active Enhancer 1	1	9
10	EnhA2	Active Enhancer 2	0.97	8
11	EnhWk	Weak Enhancer	0.68	9
12	ZNF/Rpts	ZNF genes & repeats	0.99	2
13	Het	Heterochromatin	1	2
14	TssBiv	Bivalent/Poised TSS	0.63	5
15	EnhBiv	Bivalent Enhancer	0.63	4
16	ReprPC	Repressed PolyComb	1	4
17	ReprPCWk	Weak Repressed PolyComb	1	4
18	Quies	Quiescent/Low	0.88	3

*Top value for the Pearson correlation coefficient between each state (18 states) in the NIH Epigenomic Roadmap model and our model (12 states). States with correlation coefficients below 0.75 are highlighted in red.

Table A. 4. Expression patterns of potential enhancer gene targets enriched in CVD-associated regions according to the Human Protein Atlas database.

Gene	Tissue expression cluster	RNA tissue cell type enrichment	RNA single cell type specific nTPM
PLPP3	Cluster 89: Fibroblasts - ECM organization	Kidney - Endothelial cells, Prostate - Fibroblasts, Skin - Fibroblast_2, Thyroid - Thyroid glandular cells	Astrocytes: 362.5; Fibroblasts: 403.1; Leydig cells: 881.4; Peritubular cells: 311.7
LPL	Cluster 82: **Adipose tissue** - ECM organization	Adipose subcutaneous - Adipocytes (Subcutaneous), Adipose visceral - Adipocytes (Visceral), Breast - Adipocytes (Breast), Skin - Adipocytes (Skin), Testis - Endothelial cells	Adipocytes: 409.8; Cardiomyocytes: 664.2; Granulosa cells: 513.8; Schwann cells: 156.9
NR1H3	Cluster 64: **Macrophages** - Immune response	Breast - Adipocytes (Breast), Testis - Early spermatids, Testis - Late spermatids	Hepatocytes: 58.9; **Hofbauer cells**: 47.6; Late spermatids: 48.2; Proximal enterocytes: 68.3
PLTP	Cluster 82: **Adipose tissue** - ECM organization	**Adipose subcutaneous - Macrophages, Adipose visceral - Macrophages, Colon - Macrophages**, Lung - Fibroblast_2, Prostate - Fibroblasts, **Skeletal muscle - Macrophages, Thyroid - Macrophages**	**Hofbauer cells**: 1473.5
APOC1	Cluster 62: **Liver** - Hemostasis	Liver - Hepatocytes, **Lung - Macrophages**, Skin - Sebaceous gland cells	Hepatocytes: 34731.7
ALDH1A2	Cluster 89: Fibroblasts - ECM organization	Adipose visceral - Mesothelial cells, Heart muscle - Fibroblasts, Skeletal muscle - Fibroblasts, Testis - Early spermatids	Early spermatids: 122.1; Endometrial stromal cells: 137.8; Late spermatids: 46.6; **Microglial cells**: 41.0; Spermatocytes: 62.2; Thymic epithelial cells: 82.1
ABCA1	Cluster 15: **Liver** - Metabolism	Breast - Adipocytes (Breast)	granulocytes: 120.9; Hepatocytes: 136.6; **Langerhans cells**: 127.5; **Macrophages**: 110.5
PNPLA3	Cluster 85: **Liver** - Metabolism	Adipose visceral - Adipocytes (Visceral), Kidney - Proximal tubular cells, Kidney - Proximal tubular cells, Skin - Keratinocyte (granular)	Bipolar cells: 33.8; Hepatocytes: 36.1; Rod photoreceptor cells: 35.9
APOE	Cluster 15: **Liver** - Metabolism	Heart muscle - Fibroblasts, Kidney - Proximal tubular cells, Kidney - Proximal tubular cells, Liver - Hepatocytes, **Lung - Macrophages, Pancreas - Macrophages**, Skeletal muscle - Fibroblasts	Hepatocytes: 8379.2; **Hofbauer cells**: 5636.7; Leydig cells: 2655.2; Melanocytes: 3312.6; Muller glia cells: 8525.9; Peritubular cells: 4801.9; Proximal tubular cells: 6176.1; Theca cells: 8535.1
STARD8	Cluster 7: **Adipose tissue** - Mixed function	Breast - Endothelial cells, Testis - Endothelial cells	Adipocytes: 23.9; Endothelial cells: 24.8; **Kupffer cells**: 10.2; **Langerhans cells**: 17.0;

			Macrophages: 14.6; monocytes: 29.5; Schwann cells: 17.6
CPEB4	Cluster 56: Non-specific - Unknown function	NA	Oligodendrocyte precursor cells: 302.3

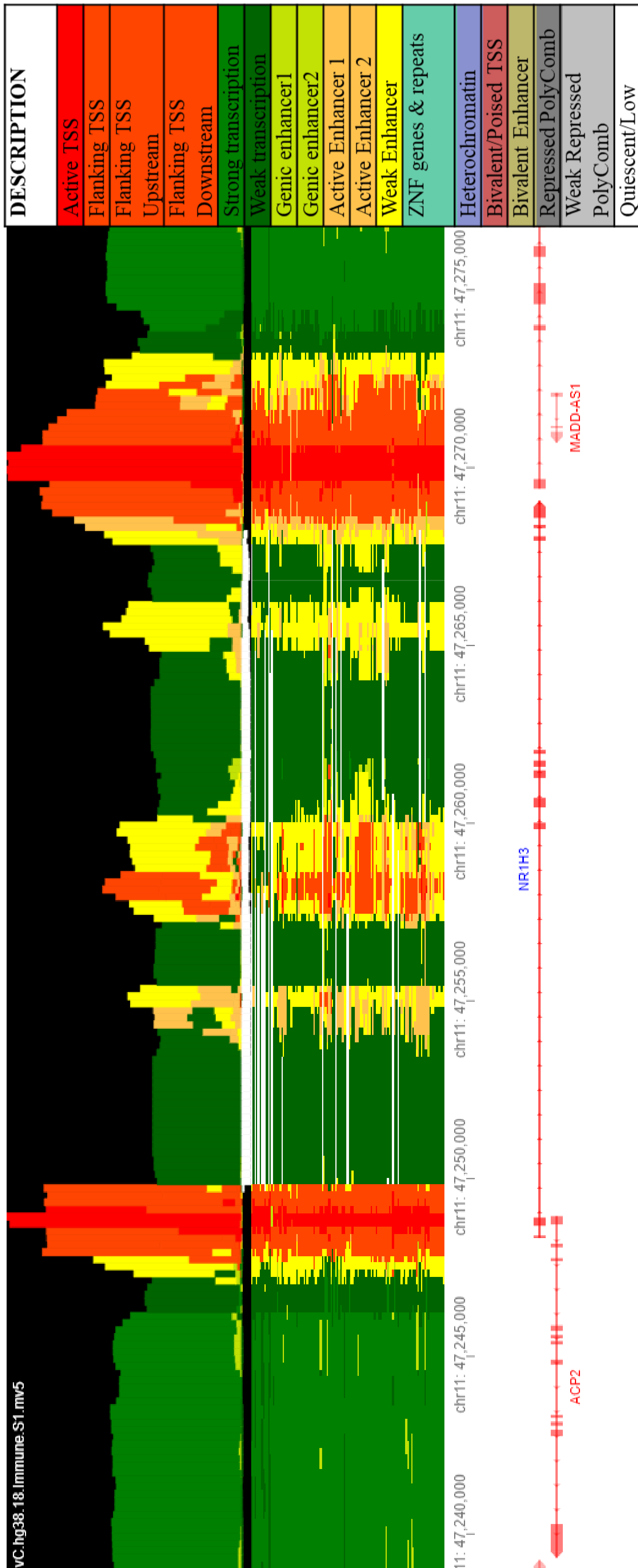

vC.hg38.18.Immune.S1.mv5
DESCRIPTION
Active TSS
Flanking TSS
Flanking TSS Upstream
Flanking TSS Downstream
Strong transcription
Weak transcription
Genic enhancer1
Genic enhancer2
Active Enhancer 1
Active Enhancer 2
Weak Enhancer
ZNF genes & repeats
Heterochromatin
Bivalent/Poised TSS
Bivalent Enhancer
Repressed PolyComb
Weak Repressed PolyComb
Quiescent/Low
chr11: 47,240,000
chr11: 47,245,000
chr11: 47,250,000
chr11: 47,255,000
chr11: 47,260,000
chr11: 47,265,000
chr11: 47,270,000
chr11: 47,275,000
ACP2
NR1H3
MADD-AS1

Figure A. 1. Epigenomic landscape of the *ACP2*, *NR1H3* and *MADD* region from EpiMap. Epilogos visualisation of ChromHMM chromatin states from the immune cells included in EpiMap analysis (Boix et al., 2021), according to the 18-states model described in **section 2.3.2**. Regions overlapping P1 and P3 are annotated as Active TSS state (Red) and Flanking TSS (Orange red) in most samples, while P2 displays that annotation only in a subset of them. These regions are surrounded by weak enhancer states (yellow), and the gene bodies show strong transcription annotation (green) in most samples.

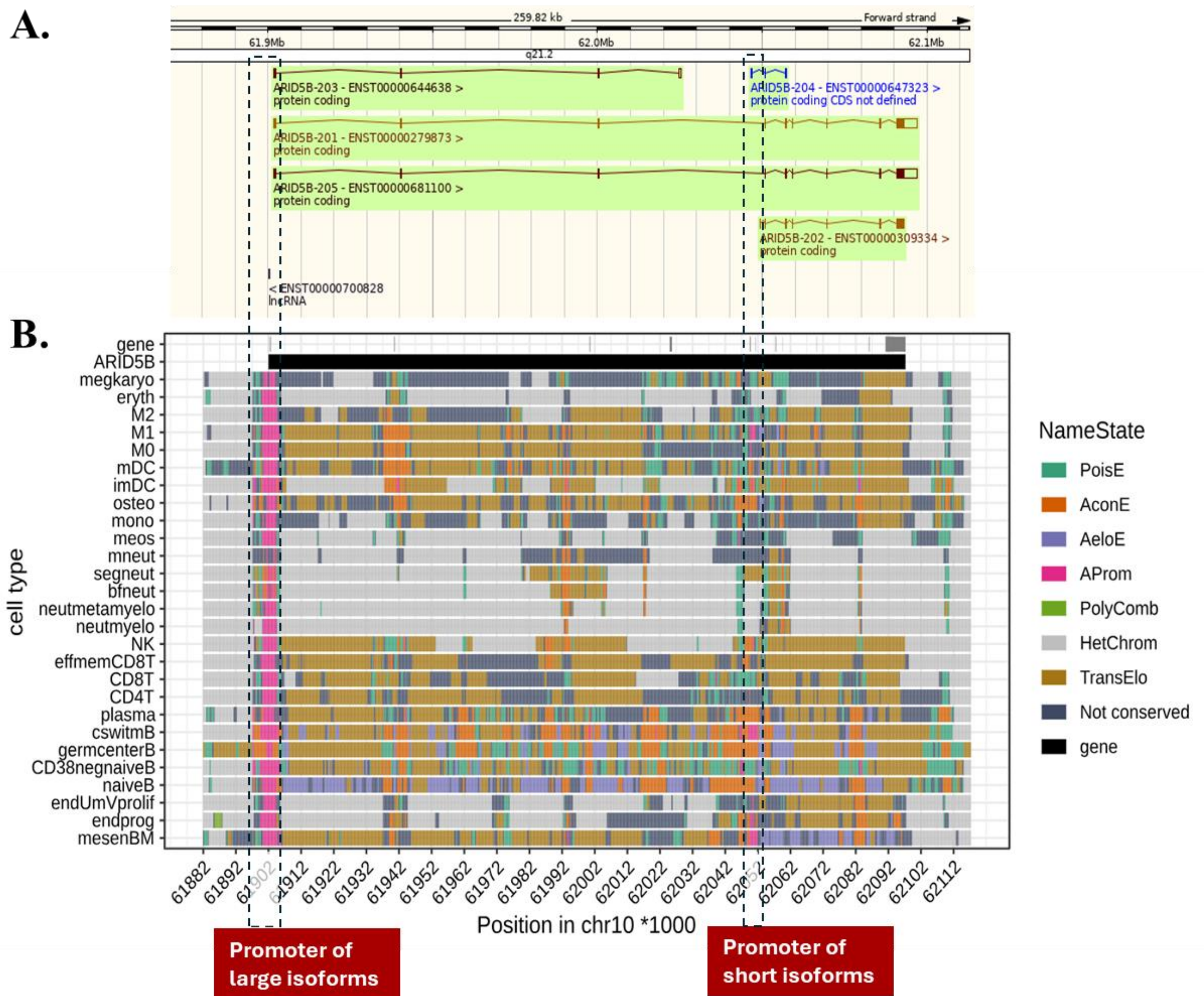


Figure A. 2. Using BLUEPRINT chromatin states to study the epigenomic landscape of *ARID5B*. (A) Annotation of the *ARID5B* transcripts as reported in ENSEMBL release 111 (Martin et al., 2023), based on the annotation from GENCODE release 44 (Frankish et al., 2021). **(B)** *ARID5B* gene region presents two promoters annotated as AProm state (pink colour) in different sets of cell types. The promoter of the larger isoforms is consistently active across the cell types, while the promoter of the shorter isoforms is present at macrophages, endothelial, mesenchymal and class-switched B cells and displays enhancer annotations in the rest of the cell types (AconE state in orange colour and PoisE in teal colours).

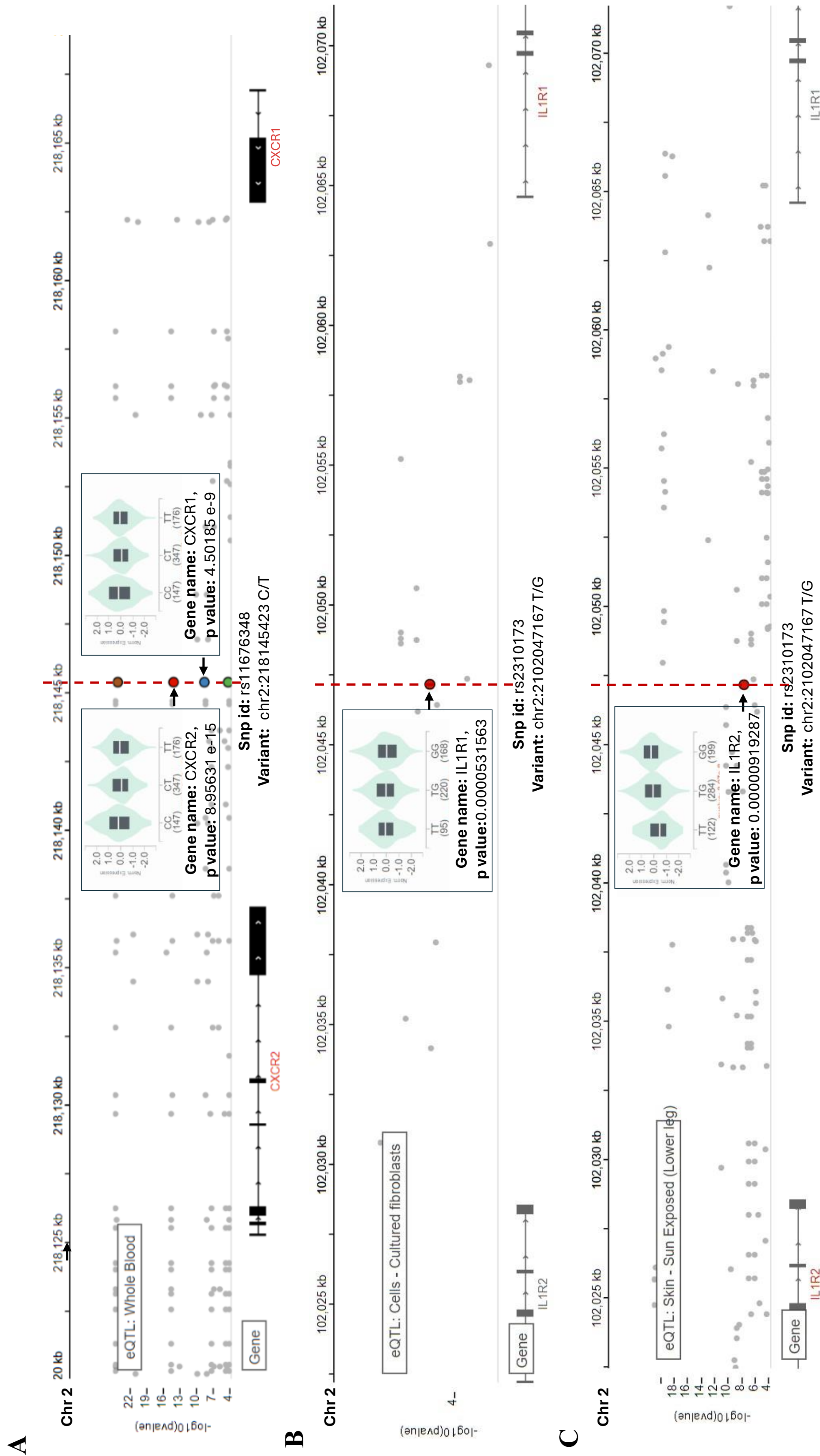
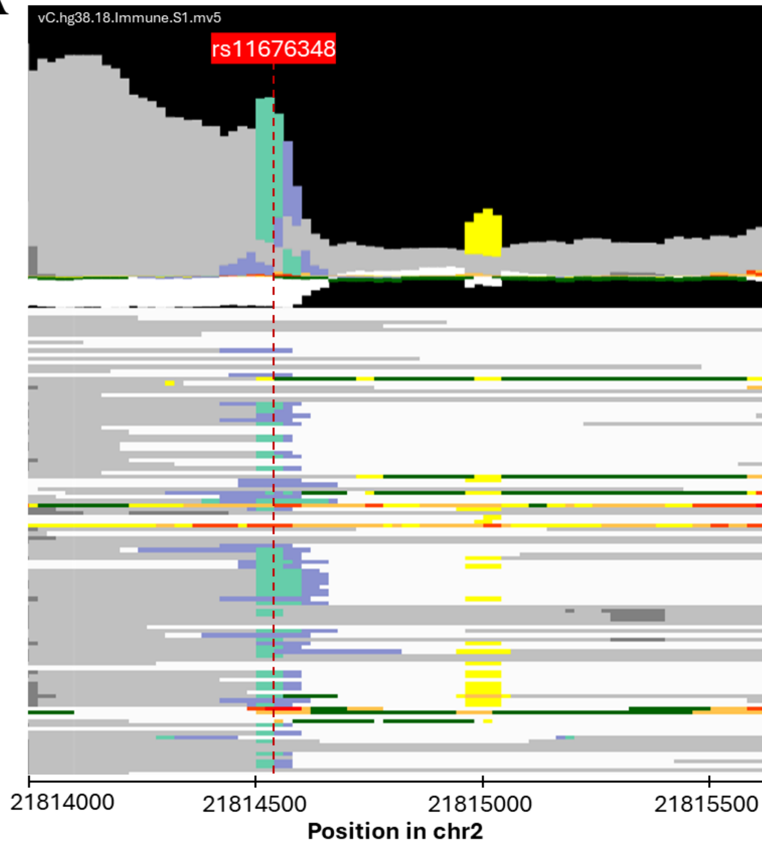


Figure A. 3. Visualisation of the GTEX eQTL data for risk variants in UC-associated genomic region enriched with myeloid enhancers. (A) rs11676348 SNP eQTL in blood of

CXCR1 and CXCR2. **(B)** rs2310173 SNP eQTL of *ILR1* in cultured fibroblasts and of *ILR2* in sun-exposed skin from the lower leg. The figure displays the genomic context of eQTL variants in the x-axis and the –log (p-value) from the eQTL analysis in the y-axis. Points corresponding to UC risk variants are highlighted as red and blue dots, and the eQTL violin plots adjacent to them show the gene expression levels for the three genotypes. These figures are customised screenshots taken from the GTEX Portal (https://gtexportal.org).

A



DESCRIPTION
Active TSS
Flanking TSS
Flanking TSS Upstream
Flanking TSS Downstream
Strong transcription
Weak transcription
Genic enhancer1
Genic enhancer2
Active Enhancer 1
Active Enhancer 2
Weak Enhancer
ZNF genes & repeats
Heterochromatin
Bivalent/Poised TSS
Bivalent Enhancer
Repressed PolyComb
Weak Repressed PolyComb
Quiescent/Low

B

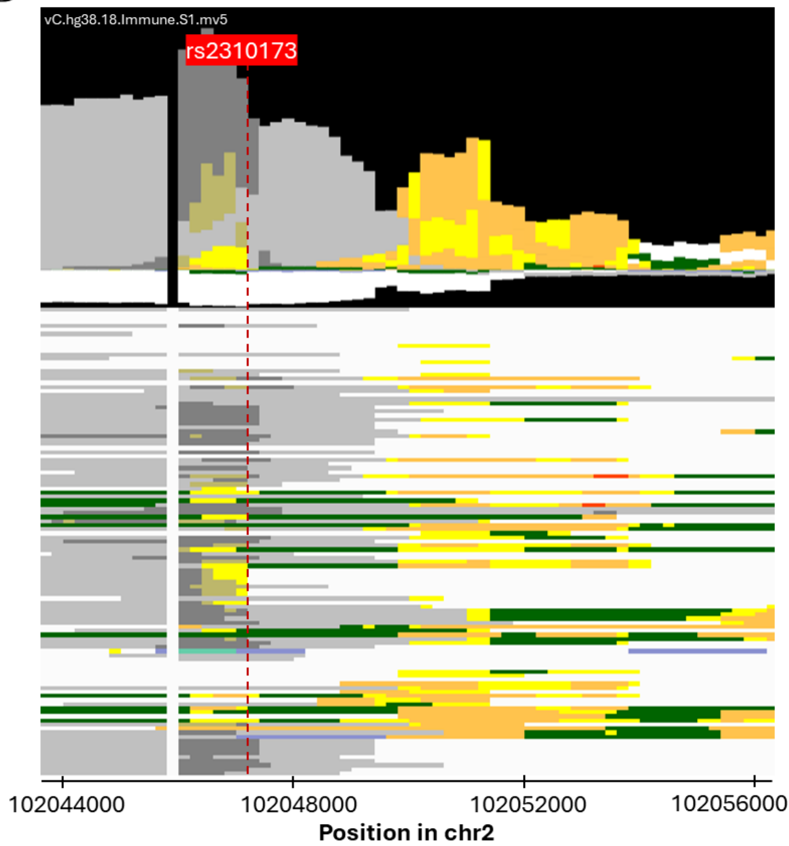


Figure A. 4. Epigenomic landscape of the UC-associated genomic regions enriched with neutrophil and myeloid enhancers in EpiMap (that lacks neutrophil data). This figure shows the ChromHMM state visualisation from Epilogos for the immune cells included in EpiMap analysis (Boix et al., 2021), according to the 18-states model described in section 2.3.2. **(A)** Regions overlapping rs11676348, are annotated as ZNF genes & repeats state (medium aquamarine) surrounded by heterochromatin (pale turquoise) in most samples. **(B)** Regions overlapping with the enhancer on the right side of rs2310173 display quiescent/low state (white) annotation in nearly half of the samples and weak enhancer (yellow) or active enhancer (orange) annotation in the rest. In the samples where the enhancer annotation is present, the region extending on the flanking sides appears annotated as weak transcription annotation (dark green).

References

Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Aken, B., Akiyama, J.A., Jammal, O. Al, Amrhein, H., Anderson, S.M., Andrews, G.R., Antoshechkin, I., Ardlie, K.G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A.A., Barnes, I.H.A., Barozzi, I., et al. (2020) 'Expanded encyclopaedias of DNA elements in the human and mouse genomes', *Nature 2020 583:7818*, 583(7818), pp. 699–710.

Abraham, J.E., Maranian, M.J., Spiteri, I., Russell, R., Ingle, S., Luccarini, C., Earl, H.M., Pharoah, P.P.D., Dunning, A.M. & Caldas, C. (2012) 'Saliva samples are a viable alternative to blood samples as a source of DNA for high throughput genotyping', *BMC Medical Genomics*, 5.

Abrams, S.T., Morton, B., Alhamdi, Y., Alsabani, M., Lane, S., Welters, I.D., Wang, G. & Toh, C.H. (2019) 'A novel assay for neutrophil extracellular trap formation independently predicts disseminated intravascular coagulation and mortality in critically ill patients', *American Journal of Respiratory and Critical Care Medicine*, 200(7), pp. 869–880.

Adams, D., Altucci, L., Antonarakis, S.E., Ballesteros, J., Beck, S., Bird, A., Bock, C., Boehm, B., Campo, E., Caricasole, A., Dahl, F., Dermitzakis, E.T., Enver, T., Esteller, M., Estivill, X., Ferguson-Smith, A., Fitzgibbon, J., Flicek, P., Giehl, C., et al. (2012) 'BLUEPRINT to decode the epigenetic signature written in blood'. Nature Biotechnology 30 (3).

Aguet, F., Barbeira, A.N., Bonazzola, R., Brown, A., Castel, S.E., Jo, B., Kasela, S., Kim-Hellmuth, S., Liang, Y., Oliva, M., Flynn, E.D., Parsana, P., Fresard, L., Gamazon, E.R., Hamel, A.R., He, Y., Hormozdiari, F., Mohammadi, P., Muñoz-Aguirre, M., et al. (2020) 'The GTEx Consortium atlas of genetic regulatory effects across human tissues', *Science*, 369(6509), .

Akpinar, M.Y., Ozin, Y.O., Kaplan, M., Ates, I., Kalkan, I.H., Kilic, Z.M.Y., Yuksel, M. & Kayacetin, E. (2018) 'Platelet-to-lymphocyte Ratio and Neutrophil-to-lymphocyte Ratio Predict Mucosal Disease Severity in Ulcerative Colitis', *Journal of medical biochemistry*, 37(2), pp. 155–162.

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G. & Gaffney, D.J. (2018) 'Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response', *Nature genetics*, 50(3), p. 424.

Allenby, G., Bocquelt, M.-T., Saunderst, M., Kazmer, S., Speck, J., Rosenbergerf, M., Lovey, A., Kastnert, P., Grippo, J.F., Chambont, P. & Levin, A.A. (1993) 'Retinoic acid receptors and retinoid X receptors: Interactions with endogenous retinoic acids (afl-ftns-retinoic acid/9-cis-retinoic acid/3,4-didehydroretinoic add/binding affinity/transcriptional activation)', *Proc. Nati. Acad. Sci. USA*, 90pp. 30–34.

Allshire, R.C. & Madhani, H.D. (2018) 'Ten principles of heterochromatin formation and function'. Nature Reviews Molecular Cell Biology 19 (4).

Alsheikh, A.J., Wollenhaupt, S., King, E.A., Reeb, J., Ghosh, S., Stolzenburg, L.R., Tamim, S., Lazar, J., Davis, J.W. & Jacob, H.J. (2022) 'The landscape of GWAS validation; systematic

review identifying 309 validated non-coding variants across 130 human diseases', *BMC Medical Genomics*, 15(1), .

Ananthakrishnan, A.N., Bernstein, C.N., Iliopoulos, D., Macpherson, A., Neurath, M.F., Ali, R.A.R., Vavricka, S.R. & Fiocchi, C. (2017) 'Environmental triggers in IBD: a review of progress and evidence', *Nature Reviews Gastroenterology & Hepatology 2017 15:1*, 15(1), pp. 39–49.

Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A., Lagacé, C., Scott, R., Amininejad, L., Bumpstead, S., Baidoo, L., Baldassano, R.N., Barclay, M., Bayless, T.M., Brand, S., et al. (2011) 'Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47'. Nature Genetics 43 (3) p.pp. 246–252.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., et al. (2014) 'An atlas of active enhancers across human cell types and tissues', *Nature 2014 507:7493*, 507(7493), pp. 455–461.

Antonacci, F., Dennis, M.Y., Huddleston, J., Sudmant, P.H., Steinberg, K.M., Rosenfeld, J.A., Miroballo, M., Graves, T.A., Vives, L., Malig, M., Denman, L., Raja, A., Stuart, A., Tang, J., Munson, B., Shaffer, L.G., Amemiya, C.T., Wilson, R.K. & Eichler, E.E. (2014) 'Palindromic GOLGA8 core duplicons promote chromosome 15q13.3 microdeletion and evolutionary instability', *Nature Genetics*, 46(12), .

Arango, M.-T., Quintero-Ronderos, P., Castiblanco, J. & Montoya-Ortíz, G. (2013) *Cell culture and cell analysis*,

Arendt, D., Musser, J.M., Baker, C.V.H., Bergman, A., Cepko, C., Erwin, D.H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M.D. & Wagner, G.P. (2016) 'The origin and evolution of cell types', *Nature Reviews Genetics 2016 17:12*, 17(12), pp. 744–757.

Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., Lambourne, J.J., Sivapalaratnam, S., Downes, K., Kundu, K., Bomba, L., Berentsen, K., Bradley, J.R., Daugherty, L.C., Delaneau, O., et al. (2016a) 'The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease', *Cell*, 167(5), pp. 1415-1429.e19.

Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., Lambourne, J.J., Sivapalaratnam, S., Downes, K., Kundu, K., Bomba, L., Berentsen, K., Bradley, J.R., Daugherty, L.C., Delaneau, O., et al. (2016b) 'The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease', *Cell*, 167(5), p. 1415.

Avsec, Ž., Agarwal, V., Visentin, D., Ledsam, J.R., Grabska-Barwinska, A., Taylor, K.R., Assael, Y., Jumper, J., Kohli, P. & Kelley, D.R. (2021) 'Effective gene expression prediction from sequence by integrating long-range interactions', *Nature Methods 2021 18:10*, 18(10), pp. 1196–1203.

Babaev, V.R., Fazio, S., Gleaves, L.A., Carter, K.J., Semenkovich, C.F. & Linton, M.F. (1999) ‘Macrophage lipoprotein lipase promotes foam cell formation and atherosclerosis in vivo’, *The Journal of Clinical Investigation*, 103(12), pp. 1697–1705.

Baggiolini, M. (2000) ‘Reflections on chemokines’, *Immunological Reviews*, 177(1), pp. 5–7.

Bao, E.L., Cheng, A.N. & Sankaran, V.G. (2019) ‘The genetics of human hematopoiesis and its disruption in disease’, *EMBO Molecular Medicine*, 11(8), .

Barnig, C., Alsaleh, G., Jung, N., Dembélé, D., Paul, N., Poirot, A., Uring-Lambert, B., Georgel, P., De Blay, F. & Bahram, S. (2015) ‘Circulating human eosinophils share a similar transcriptional profile in asthma and other hypereosinophilic disorders’, *PLoS ONE*, 10(11), .

Bartolomei, M.S. & Ferguson-Smith, A.C. (2011) ‘Mammalian genomic imprinting’, *Cold Spring Harbor Perspectives in Biology*, 3(7), .

Battram, T., Yousefi, P., Crawford, G., Prince, C., Sheikhali Babaei, M., Sharp, G., Hatcher, C., Vega-Salas, M.J., Khodabakhsh, S., Whitehurst, O., Langdon, R., Mahoney, L., Elliott, H.R., Mancano, G., Lee, M.A., Watkins, S.H., Lay, A.C., Hemani, G., Gaunt, T.R., et al. (2022) ‘The EWAS Catalog: A database of epigenome-wide association studies’, *Wellcome Open Research*, 7.

Bauer, D.E., Kamran, S.C., Lessard, S., Xu, J., Fujiwara, Y., Lin, C., Shao, Z., Canver, M.C., Smith, E.C., Pinello, L., Sabo, P.J., Vierstra, J., Voit, R.A., Yuan, G.C., Porteus, M.H., Stamatoyannopoulos, J.A., Lettre, G. & Orkin, S.H. (2013) ‘An erythroid enhancer of BCL11A subject to genetic variation determines fetal hemoglobin level’, *Science*, 342(6155), .

Bauer, W.R., Hayes, J.J., White, J.H. & Wolffe, A.P. (1994) ‘Nucleosome Structural Changes Due to Acetylation’, *Journal of Molecular Biology*, 236(3), pp. 685–690.

Beaumont, R.N., Flatley, C., Vaudel, M., Wu, X., Chen, J., Moen, G.H., Skotte, L., Helgeland, Ø., Solé-Navais, P., Banasik, K., Albiñana, C., Ronkainen, J., Fadista, J., Stinson, S.E., Trajanoska, K., Wang, C.A., Westergaard, D., Srinivasan, S., Sánchez-Soriano, C., et al. (2023) ‘Genome-wide association study of placental weight identifies distinct and shared genetic influences between placental and fetal growth’, *Nature Genetics 2023 55:11*, 55(11), pp. 1807–1819.

Becker, K.G. (2004) ‘The common variants/multiple disease hypothesis of common complex genetic disorders’, *Medical Hypotheses*, 62(2), .

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, Núria, A Martens, J.H., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M., Wingett, S.W., Clot, G., Queirós, A.C., Castellano, G., Blanc, J., Gut, M., Merkel, A., Heath, S., Vlasova, A., et al. (2018) *The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia*,

Beekman, R., Chapaprieta, V., Russiñol, N., Vilarrasa-Blasi, R., Verdaguer-Dot, Núria, Martens, J.H.A., Duran-Ferrer, M., Kulis, M., Serra, F., Javierre, B.M., Wingett, S.W., Clot, G., Queirós, A.C., Castellano, G., Blanc, J., Gut, M., Merkel, A., Heath, S., Vlasova,

A., et al. (2018) ‘The reference epigenome and regulatory chromatin landscape of chronic lymphocytic leukemia’, *Nature medicine*, 24(6), pp. 868–880.

Bell, C.G., Gao, F., Yuan, W., Roos, L., Acton, R.J., Xia, Y., Bell, J., Ward, K., Mangino, M., Hysi, P.G., Wang, J. & Spector, T.D. (2018) ‘Obligatory and facilitative allelic variation in the DNA methylome within common disease-associated loci’, *Nature Communications*, 9(1), .

Benabdallah, N.S., Williamson, I., Illingworth, R.S., Kane, L., Boyle, S., Sengupta, D., Grimes, G.R., Therizols, P. & Bickmore, W.A. (2019) ‘Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation’, *Molecular cell*, 76(3), pp. 473-484.e7.

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., Mcbride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M.M., et al. (2009) *Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence*,

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., Farnham, P.J., Hirst, M., Lander, E.S., Mikkelsen, T.S. & Thomson, J.A. (2010) ‘The NIH Roadmap Epigenomics Mapping Consortium’, *Nature biotechnology*, 28(10), p. 1045.

Bhatia, S., Kleinjan, D.J., Uttley, K., Mann, A., Dellepiane, N. & Bickmore, W.A. (2021) ‘Quantitative spatial and temporal assessment of regulatory element activity in zebrafish’, *eLife*, 10.

Billingsley, K.J., Ding, J., Jerez, P.A., Illarionova, A., Levine, K., Grenn, F.P., Makarious, M.B., Moore, A., Vitale, D., Reed, X., Hernandez, D., Torkamani, A., Ryten, M., Hardy, J., Chia, R., Scholz, S.W., Traynor, B.J., Dalgard, C.L., Ehrlich, D.J., et al. (2023) ‘Genome-Wide Analysis of Structural Variants in Parkinson Disease’, *Annals of Neurology*, 93(5), .

Bioconductor (2021) *GO.db: A set of annotation maps describing the entire Gene Ontology*.

Blayney, Joseph W, Francis, H., Rampasekova, A., Boeke, J.D., Higgs, D.R., Correspondence, M.K. & Org (, J.B. (2023) ‘Super-enhancers include classical enhancers and facilitators to fully activate gene expression’, *Cell*, 186pp. 5826-5839.e18.

Blayney, Joseph W., Francis, H., Rampasekova, A., Camellato, B., Mitchell, L., Stolper, R., Cornell, L., Babbs, C., Boeke, J.D., Higgs, D.R. & Kassouf, M. (2023) ‘Super-enhancers include classical enhancers and facilitators to fully activate gene expression’, *Cell*, 186(26), pp. 5826-5839.e18.

Bobba, C.M., Azarrafiy, R., Spratt, J.R., Hendrickson, J., Martin, T.D., Arnaoutakis, G.J., Jeng, E.I. & Beaver, T.M. (2023) ‘A highly penetrant ACTA2 mutation of thoracic aortic disease’, *Journal of Cardiothoracic Surgery*, 18(1), pp. 1–4.

Bobryshev, Y. V., Nikiforov, N.G., Elizova, N. V. & Orekhov, A.N. (2017) ‘Macrophages and their contribution to the development of atherosclerosis’, *Results and Problems in Cell Differentiation*, 62pp. 273–298.

Boix, C.A., James, B.T., Park, Y.P., Meuleman, W. & Kellis, M. (2021) 'Regulatory genomic circuitry of human disease loci by integrative epigenomics', *Nature 2021 590:7845*, 590(7845), pp. 300–307.

Bolk, S., Pelet, A., Hofstra, R.M.W., Angrist, M., Salomon, R., Croaker, D., Buys, C.H.C.M., Lyonnet, S. & Chakravarti, A. (2000) 'A human model for multigenic inheritance: Phenotypic expression in Hirschsprung disease requires both the RET gene and a new 9q31 locus', *Proceedings of the National Academy of Sciences of the United States of America*, 97(1), .

Boraschi, D. (2022) 'What Is IL-1 for? The Functions of Interleukin-1 Across Evolution', *Frontiers in Immunology*, 13.

Boraschi, D. & Tagliabue, A. (2013) 'The interleukin-1 receptor family', *Seminars in Immunology*, 25(6), pp. 394–407.

Boros, J., Arnoult, N., Stroobant, V., Collet, J.-F. & Decottignies, A. (2014a) 'Polycomb Repressive Complex 2 and H3K27me3 Cooperate with H3K9 Methylation To Maintain Heterochromatin Protein 1α at Chromatin', *Molecular and Cellular Biology*, 34(19), p. 3662.

Boros, J., Arnoult, N., Stroobant, V., Collet, J.-F. & Decottignies, A. (2014b) 'Polycomb Repressive Complex 2 and H3K27me3 Cooperate with H3K9 Methylation To Maintain Heterochromatin Protein 1α at Chromatin', *Molecular and Cellular Biology*, 34(19), p. 3662.

Van den Bossche, J., O'Neill, L.A. & Menon, D. (2017) 'Macrophage Immunometabolism: Where Are We (Going)?' Trends in Immunology 38 (6).

Boyle, W.J., Simonet, W.S. & Lacey, D.L. (2003) 'Osteoclast differentiation and activation'. Nature 423 (6937).

Braun, A.C., Hendrick, J., Eisler, S.A., Schmid, S., Hausser, A. & Olayioye, M.A. (2015) 'The Rho-specific GAP protein DLC3 coordinates endocytic membrane trafficking', *Journal of Cell Science*, 128(7), pp. 1386–1399.

Breeze, C.E., Beck, S., Berndt, S.I. & Franceschini, N. (2022) 'The missing diversity in human epigenomic studies'. Nature Genetics 54 (6).

Breeze, C.E., Haugen, E., Reynolds, A., Teschendorff, A., van Dongen, J., Lan, Q., Rothman, N., Bourque, G., Dunham, I., Beck, S., Stamatoyannopoulos, J., Franceschini, N. & Berndt, S.I. (2022) 'Integrative analysis of 3604 GWAS reveals multiple novel cell type-specific regulatory associations', *Genome Biology*, 23(1), .

Briggs, J.A., Weinreb, C., Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W. & Klein, A.M. (2018) 'The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution', *Science*, 360(6392), .

British Heart Foundation (2024) *Heart Statistics*. [Online]

Bruschi, F.V., Claudel, T., Tardelli, M., Starlinger, P., Marra, F. & Trauner, M. (2019) ‘PNPLA3 I148M Variant Impairs Liver X Receptor Signaling and Cholesterol Homeostasis in Human Hepatic Stellate Cells’, *Hepatology communications*, 3(9), pp. 1191–1204.

Bruschi, F.V., Tardelli, M., Claudel, T. & Trauner, M. (2017) ‘PNPLA3 expression and its impact on the liver: current perspectives’, *Hepatic Medicine: Evidence and Research*, 9pp. 55–66.

Buck, M.D., Sowell, R.T., Kaech, S.M. & Pearce, E.L. (2017) ‘Metabolic Instruction of Immunity’. Cell 169 (4).

Büki, G., Hadzsiev, K. & Bene, J. (2023) ‘Copy Number Variations in Neuropsychiatric Disorders’. International Journal of Molecular Sciences 24 (18).

Buniello, A., Macarthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P.L., Amode, R., Guillen, J.A., Riat, H.S., Trevanion, S.J., Hall, P., Junkins, H., et al. (2019) ‘The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019’, *Nucleic Acids Research*, 47(D1), .

Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J., Todd, J.A., Donnelly, P., Barrett, J.C., Davison, D., Easton, D., Evans, D., Leung, H.T., Marchini, J.L., Morris, A.P., et al. (2007) ‘Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls’, *Nature 2007 447:7145*, 447(7145), pp. 661–678.

Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C.M., Pereira Da Costa, M. & Reis E Sousa, C. (2021) ‘Dendritic Cells Revisited’. Annual Review of Immunology 39.

Calderon, D., Bhaskar, A., Knowles, D.A., Golan, D., Raj, T., Fu, A.Q. & Pritchard, J.K. (2017) *Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression*,

Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J. V., Knowles, D.A., Gao, Z., Blaeschke, F., Parent, A. V., Burt, T.D., Anderson, M.S., Criswell, L.A., Greenleaf, W.J., Marson, A., et al. (2019) ‘Landscape of stimulation-responsive chromatin across diverse human immune cells’, *Nature genetics*, 51(10), pp. 1494–1505.

Cano-Gamez, E. & Trynka, G. (2020) ‘From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases’, *Frontiers in Genetics*, 11p. 505357.

Cao, F., Fang, Y., Tan, H.K., Goh, Y., Choy, J.Y.H., Koh, B.T.H., Hao Tan, J., Bertin, N., Ramadass, A., Hunter, E., Green, J., Salter, M., Akoulitchev, A., Wang, W., Chng, W.J., Tenen, D.G. & Fullwood, M.J. (2017) ‘Super-enhancers and broad h3k4me3 domains form complex gene regulatory circuits involving chromatin interactions’, *Scientific Reports*, 7(1), pp. 1–14.

Cao, G., Beyer, T.P., Yang, X.P., Schmidt, R.J., Zhang, Y., Bensch, W.R., Kauffman, R.F., Gao, H., Ryan, T.P., Liang, Y., Eacho, P.I. & Jiang, X.-C. (2002) ‘Phospholipid transfer protein

is regulated by liver X receptors in vivo.', *The Journal of biological chemistry*, 277(42), pp. 39561–5.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., Trapnell, C. & Shendure, J. (2019) 'The single cell transcriptional landscape of mammalian organogenesis', *Nature*, 566(7745), p. 496.

Cao, M., Yu, M., Zhang, Y., Tong, D., Guo, L., Wang, C., Li, T., Yang, X., Zhang, C., Kou, J., Bi, Y., Zhou, J. & Shi, J. (2017) 'Neutrophil Extracellular Traps Exacerbate Inflammatory Responses and Thrombotic Tendency in Both a Murine Colitis Model and Patients with Inflammatory Bowel Disease', *Blood*, 130(Suppl_1), p. 994.

Carja, O., MacIsaac, J.L., Mah, S.M., Henn, B.M., Kobor, M.S., Feldman, M.W. & Fraser, H.B. (2017) 'Worldwide patterns of human epigenetic variation', *Nature Ecology and Evolution*, 1(10), .

Carrillo-de-Santa-Pau, E., Juan, D., Pancaldi, V., Were, F., Martin-Subero, I., Rico, D. & Valencia, A. (2017) 'Automatic identification of informative regions with epigenomic changes associated to hematopoiesis', *Nucleic Acids Research*, 45(16), pp. 9244–9259.

Carter, B., Ku, W.L., Kang, J.Y., Hu, G., Perrie, J., Tang, Q. & Zhao, K. (2019) 'Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq)', *Nature Communications*, 10(1), .

Carter, B. & Zhao, K. (2020) 'The epigenetic basis of cellular heterogeneity', *Nature Reviews Genetics 2020 22:4*, 22(4), pp. 235–250.

Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov, I., Vasileva, A. & Lappalainen, T. (2018) 'Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk', *Nature Genetics*, 50(9), .

De Castro-Orós, I., Pérez-López, J., Mateo-Gallego, R., Rebollar, S., Ledesma, M., León, M., Cofán, M., Casasnovas, J.A., Ros, E., Rodríguez-Rey, J.C., Civeira, F. & Pocoví, M. (2014) 'A genetic variant in the LDLR promoter is responsible for part of the LDL-cholesterol variability in primary hypercholesterolemia', *BMC Medical Genomics*, 7(1), .

Cedar, H. & Bergman, Y. (2011) 'Epigenetics of haematopoietic cell development', *Nature Reviews Immunology 2011 11:7*, 11(7), pp. 478–488.

Chen, G.Y. & Nuñez, G. (2010) 'Sterile inflammation: Sensing and reacting to damage'. Nature Reviews Immunology 10 (12).

Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., Datta, A., Richardson, D., Burden, F., Mead, D., Mann, A.L., Fernandez, J.M., Rowlston, S., Wilder, S.P., Farrow, S., et al. (2016) 'Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells', *Cell*, 167(5), pp. 1398-1414.e24.

Chen, L., Zhang, J., Deng, X., Liu, Y., Yang, X., Wu, Q. & Yu, C. (2017) 'Lysophosphatidic acid directly induces macrophage-derived foam cell formation by blocking the expression of SRBI', *Biochemical and Biophysical Research Communications*, 491(3), pp. 587–594.

Chiang, C., Scott, A.J., Davis, J.R., Tsang, E.K., Li, X., Kim, Y., Hadzic, T., Damani, F.N., Ganel, L., Montgomery, S.B., Battle, A., Conrad, D.F. & Hall, I.M. (2017) 'The impact of structural variation on human gene expression', *Nature Genetics*, 49(5), .

Choi, J., Lysakovskaia, K., Stik, G., Demel, C., Söding, J., Tian, T. V., Graf, T. & Cramer, P. (2021) 'Evidence for additive and synergistic action of Mammalian enhancers during cell fate determination', *eLife*, 10.

Choi, J., Xu, M., Makowski, M.M., Zhang, T., Law, M.H., Kovacs, M.A., Granzhan, A., Kim, W.J., Parikh, H., Gartside, M., Trent, J.M., Teulade-Fichou, M.P., Iles, M.M., Newton-Bishop, J.A., Bishop, D.T., Macgregor, S., Hayward, N.K., Vermeulen, M. & Brown, K.M. (2017) 'A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF', *Nature Genetics*, 49(9), .

Christenson, L.K. & Strauss, J.F. (2000) 'Steroidogenic acute regulatory protein (StAR) and the intramitochondrial translocation of cholesterol', *Biochimica et biophysica acta*, 1529(1–3), pp. 175–187.

Cieslak, A., Charbonnier, G., Tesio, M., Mathieu, E.L., Belhocine, M., Touzart, A., Smith, C., Hypolite, G., Andrieu, G.P., Martens, J.H.A., Janssen-Megens, E., Gut, M., Gut, I., Boissel, N., Petit, A., Puthier, D., Macintyre, E., Stunnenberg, H.G., Spicuglia, S., et al. (2020) 'Blueprint of human thymopoiesis reveals molecular mechanisms of stage-specific TCR enhancer activation', *Journal of Experimental Medicine*, 217(9), .

Claringbould, A. & Zaugg, J.B. (2021) 'Enhancers in disease: molecular basis and emerging treatment strategies', *Trends in Molecular Medicine*, 27(11), pp. 1060–1073.

Clien, L., Kostadraia, M., Martens, J.H.A., Canu, G., Garcia, S.P., Torro, E., Downes, K., Macaolay, L.C., Bielczyk-Maezynska, E., Coe, S., Farrow, S., Poudel, P., Borden, F., Jansen, S.B.G., Astle, W.J., Attwood, A., Bariana, T., De Bono, B., Bresechi, A., et al. (2014) 'Transcriptional diversity during lineage commitment of human blood progenitors', *Science (New York, N.Y.)*, 345(6204), .

Coillard, A. & Segura, E. (2019) 'In vivo differentiation of human monocytes'. Frontiers in Immunology 10 (AUG).

Colombel, J.F., Shin, A. & Gibson, P.R. (2019) 'AGA Clinical Practice Update on Functional Gastrointestinal Symptoms in Patients With Inflammatory Bowel Disease: Expert Review', *Clinical Gastroenterology and Hepatology*, 17(3), pp. 380-390.e1.

Connelly, P.W., Maguire, G.F. & Little, J.A. (1987) 'Apolipoprotein CII(St. Michael). Familial apolipoprotein CII deficiency associated with premature vascular disease', *Journal of Clinical Investigation*, 80(6), .

Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. (2013) 'Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease'. Human Genetics 132 (10).

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sallari, R., Lupien, M., Markowitz, S. & Scacheri, P.C. (2014) ‘Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits’, *Genome Research*, 24(1), .

Corradin, O. & Scacheri, P.C. (2014) ‘Enhancer variants: Evaluating functions in common disease’. Genome Medicine 6 (10).

Corvin, A., Craddock, N. & Sullivan, P.F. (2010) ‘Genome-wide association studies: A primer’. Psychological Medicine 40 (7).

Costet, P., Lalanne, F., Gerbod-Giannone, M.C., Molina, J.R., Fu, X., Lund, E.G., Gudas, L.J. & Tall, A.R. (2003) ‘Retinoic Acid Receptor-Mediated Induction of ABCA1 in Macrophages’, *Molecular and Cellular Biology*, 23(21), pp. 7756–7766.

Cui, H.S., Joo, S.Y., Cho, Y.S., Kim, J.B. & Seo, C.H. (2020) ‘CPEB1 or CPEB4 knockdown suppresses the TAK1 and Smad signalings in THP-1 macrophage-like cells and dermal fibroblasts’, *Archives of biochemistry and biophysics*, 683.

Cummins, C.L., Volle, D.H., Zhang, Y., Mcdonald, J.G., Sion, B., Lefrançois-Martinez, A.-M., Caira, F., Veyssière, G., Mangelsdorf, D.J. & Lobaccaro, J.-M.A. (2006) ‘Liver X receptors regulate adrenal cholesterol balance’, *The Journal of Clinical Investigation*, 116(7), .

Curtiss, L.K. (2000) ‘ApoE in atherosclerosis: A protein with multiple hats’, *Arteriosclerosis, Thrombosis, and Vascular Biology*, 20(8), pp. 1852–1853.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A. & Davies, R.M. (2021) ‘SAMtools’, *GigaScience*, 10(2), .

Dejosez, M., Dall’agnese, A., Ramamoorthy, M., Abraham, B.J., Young, R.A., Zwaka, T.P., Platt, J., Yin, X., Hogan, M., Brosh, R., Weintraub, A.S. & Hnisz, D. (2023) *Regulatory architecture of housekeeping genes is driven by promoter assemblies*,

Denisenko, E., Guler, R., Mhlanga, M.M., Suzuki, H., Brombacher, F. & Schmeier, S. (2017) ‘Genome-wide profiling of transcribed enhancers during macrophage activation’, *Epigenetics and Chromatin*, 10(1), pp. 1–17.

Dimas, A.S., Deutsch, S., Stranger, B.E., Montgomery, S.B., Borel, C., Attar-Cohen, H., Ingle, C., Beazley, C., Arcelus, M.G., Sekowska, M., Gagnebin, M., Nisbett, J., Deloukas, P., Dermitzakis, E.T. & Antonarakis, S.E. (2009) ‘Common regulatory variation impacts gene expression in a cell type-dependent manner’, *Science*, 325(5945), .

Dinallo, V., Marafini, I., Fusco, D. Di, Laudisi, F., Franzè, E., Grazia, A. Di, Figliuzzi, M.M., Caprioli, F., Stolfi, C., Monteleone, I. & Monteleone, G. (2019) ‘Neutrophil Extracellular Traps Sustain Inflammatory Signals in Ulcerative Colitis’, *Journal of Crohn’s & colitis*, 13(6), pp. 772–784.

Dipple, K.M. & McCabe, E.R.B. (2000) ‘Modifier genes convert “simple” Mendelian disorders to complex traits’, *Molecular Genetics and Metabolism*, 71(1–2), .

Distler, J.H.W., Jüngel, A., Kurowska-Stolarska, M., Michel, B.A., Gay, R.E., Gay, S. & Distler, O. (2005) 'Nucleofection: a new, highly efficient transfection method for primary human keratinocytes*', *Experimental Dermatology*, 14(4), pp. 315–320.

Dixon, E.D., Genger, J.-W., Zhu, C., Stadlmayer, S., Paolini, E., Nardo, A.D., Mlitz, V., Fuchs, C., Bergthaler, A., Dongiovanni, P., Ellmeier, W. & Trauner, M. (2023) 'The PNPLA3 I148M variant aggravates inflammation through dysfunctional LXR and PPAR gamma signaling', *EASL Congress 2023*,

Do, R., Willer, C.J., Schmidt, E.M., Sengupta, S., Gao, C., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., Beckmann, J.S., Bragg-Gresham, J.L., Chang, H.Y., Demirkan, A., Den Hertog, H.M., Donnelly, L.A., Ehret, G.B., et al. (2013) 'Common variants associated with plasma triglycerides and risk for coronary artery disease'. Nature Genetics 45 (11).

Dobnikar, L., Taylor, A.L., Chappell, J., Oldach, P., Harman, J.L., Oerton, E., Dzierzak, E., Bennett, M.R., Spivakov, M. & Jørgensen, H.F. (2018) 'Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels', *Nature Communications 2018 9:1*, 9(1), pp. 1–17.

Dorfmeister, B., Zeng, W.W., Dichlberger, A., Nilsson, S.K., Schaap, F.G., Hubacek, J.A., Merkel, M., Cooper, J.A., Lookene, A., Putt, W., Whittall, R., Lee, P.J., Lins, L., Delsaux, N., Nierman, M., Kuivenhoven, J.A., Kastelein, J.J.P., Vrablik, M., Olivecrona, G., et al. (2008) 'Effects of six APOA5 variants, identified in patients with severe hypertriglyceridemia, on in vitro lipoprotein lipase activity and receptor binding', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 28(10), .

Drury, B., Hardisty, G., Gray, R.D. & Ho, G. tzer (2021) 'Neutrophil Extracellular Traps in Inflammatory Bowel Disease: Pathogenic Mechanisms and Clinical Translation', *Cellular and Molecular Gastroenterology and Hepatology*, 12(1), p. 321.

Dudek, A.M. & Porteus, M.H. (2021) 'Answered and Unanswered Questions in Early-Stage Viral Vector Transduction Biology and Innate Primary Cell Toxicity for Ex-Vivo Gene Editing', *Frontiers in Immunology*, 12p. 660302.

Durinck, S. & Huber, W. (2021) *biomaRt: Interface to BioMart databases (e.g. Ensembl, COSMIC, WormBase) with additional features*.

Eberl, G., Colonna, M., Santo, J.P.D. & McKenzie, A.N.J. (2015) 'Innate lymphoid cells: A new paradigm in immunology'. Science 348 (6237).

Ecker, S., Chen, L., Pancaldi, V., Bagger, F.O., Fernández, J.M., Carrillo de Santa Pau, E., Juan, D., Mann, A.L., Watt, S., Casale, F.P., Sidiropoulos, N., Rapin, N., Merkel, A., Stunnenberg, H.G., Stegle, O., Frontini, M., Downes, K., Kuijpers, T.W., Kuijpers, T.W., et al. (2017) 'Genome-wide analysis of differential transcriptional and epigenetic variability across human immune cell types', *Genome Biology*, 18(1), pp. 1–17.

Eggermann, T., Monk, D., de Nanclares, G.P., Kagami, M., Giabicani, E., Riccio, A., Tümer, Z., Kalish, J.M., Tauber, M., Duis, J., Weksberg, R., Maher, E.R., Begemann, M. & Elbracht, M. (2023) 'Imprinting disorders', *Nature Reviews Disease Primers*, 9(1), .

Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H. & Nadeau, J.H. (2010) 'Missing heritability and strategies for finding the underlying causes of complex disease'. Nature Reviews Genetics 11 (6).

Eisenberg, E. & Levanon, E.Y. (2013) 'Human housekeeping genes, revisited', *Trends in Genetics*, 29(10), pp. 569–574.

Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hübenthal, M., Folseraas, T., Wang, Y., Esko, T., Metspalu, A., Westra, H.J., Franke, L., Pers, T.H., Weersma, R.K., Collij, V., et al. (2016) 'Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci', *Nature Genetics*, 48(5), pp. 510–518.

Emi, M., Wilson, D.E., Iverius, P.H., Wu, L., Hata, A., Regele, R., Williams, R.R. & Lalouel, J.M. (1990) 'Missense mutation (Gly → Glu188) of human lipoprotein lipase imparting functional deficiency', *Journal of Biological Chemistry*, 265(10), .

Endo, A. (2010) 'A historical perspective on the discovery of statins'. Proceedings of the Japan Academy Series B: Physical and Biological Sciences 86 (5).

Endo-Umeda, K., Kim, E., Thomas, D.G., Liu, W., Dou, H., Yalcinkaya, M., Abramowicz, S., Xiao, T., Antonson, P., Gustafsson, J.-Å., Makishima, M., Reilly, M.P., Wang, N. & Tall, A.R. (2022) 'Myeloid LXR (Liver X Receptor) Deficiency Induces Inflammatory Gene Expression in Foamy Macrophages and Accelerates Atherosclerosis.', *Arteriosclerosis, thrombosis, and vascular biology*, 42(6), pp. 719–731.

Ernst, J. & Kellis, M. (2012) 'ChromHMM: automating chromatin-state discovery and characterization', *Nature methods*, 9(3), pp. 215–216.

Ernst, J. & Kellis, M. (2015) 'Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues', *Nature Biotechnology 2015 33:4*, 33(4), pp. 364–376.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. & Bernstein, B.E. (2011) 'Mapping and analysis of chromatin state dynamics in nine human cell types', *Nature 2011 473:7345*, 473(7345), pp. 43–49.

Eshghjoo, S., Jayaraman, A., Sun, Y. & Alaniz, R.C. (2021) 'Microbiota-Mediated Immune Regulation in Atherosclerosis', *Molecules 2021, Vol. 26, Page 179*, 26(1), p. 179.

Farh, K.K.H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., De Jager, P.L., Kuchroo, V.K., Epstein, C.B., Daly, M.J., et al. (2015) 'Genetic and epigenetic fine mapping of causal autoimmune disease variants', *Nature*, 518(7539), pp. 337–343.

Fazzari, M.J. & Greally, J.M. (2004) ‘Epigenomics: beyond CpG islands’, *Nature Reviews Genetics 2004 5:6*, 5(6), pp. 446–455.

Feuerborn, A. & Cook, P.R. (2015) ‘Why the activity of a gene depends on its neighbors’, *Trends in Genetics*, 31(9), pp. 483–490.

Fisher, A.G. (2002) ‘Cellular identity and lineage choice’, *Nature Reviews Immunology 2002 2:12*, 2(12), pp. 977–982.

Fisher, R.M., Mailly, F., Peacock, R.E., Hamsten, A., Seed, M., Yudkin, J.S., Beisiegel, U., Feussner, G., Miller, G., Humphries, S.E. & Talmud, P.J. (1995) ‘Interaction of the lipoprotein lipase asparagine 291 → serine mutation with body mass index determines elevated plasma triacylglycerol concentrations: A study in hyperlipidemic subjects, myocardial infarction survivors, and healthy adults’, *Journal of Lipid Research*, 36(10), .

Flowers, E., Froelicher, E.S. & Aouizerat, B.E. (2012) ‘Gene-environment interactions in cardiovascular disease’, *European Journal of Cardiovascular Nursing*, 11(4), pp. 472–478.

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J.E., Mudge, J.M., Sisu, C., Wright, J.C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Sala, S.C., Cunningham, F., Domenico, T. Di, Donaldson, S., Fiddes, I.T., Girón, C.G., et al. (2021) ‘GENCODE 2021’, *Nucleic Acids Research*, 49(D1), pp. D916–D923.

Freeberg, M.A., Fromont, L.A., D’Altri, T., Romero, A.F., Ciges, J.I., Jene, A., Kerry, G., Moldes, M., Ariosa, R., Bahena, S., Barrowdale, D., Barbero, M.C., Fernandez-Orth, D., Garcia-Linares, C., Garcia-Rios, E., Haziza, F., Juhasz, B., Llobet, O.M., Milla, G., et al. (2022) ‘The European Genome-phenome Archive in 2021’, *Nucleic Acids Research*, 50(D1), pp. D980–D987.

Freund, E.C., Lock, J.Y., Oh, J., Maculins, T., Delamarre, L., Bohlen, C.J., Haley, B. & Murthy, A. (2020) ‘Efficient gene knockout in primary human and murine myeloid cells by non-viral delivery of crispr-cas9’, *Journal of Experimental Medicine*, 217(7), .

Friman, E.T., Flyamer, I.M., Marenduzzo, D., Boyle, S. & Bickmore, W.A. (2023) ‘Ultra-long-range interactions between active regulatory elements’, *Genome Research*, 33(8), pp. 1269–1283.

Fruchart, J.C., Nierman, M.C., Stroes, E.S.G., Kastelein, J.J.P. & Duriez, P. (2004) ‘New risk factors for atherosclerosis and patient risk assessment’. Circulation 109 (23 SUPPL.).

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., Nguyen, T.H., Kane, M., Perez, E.M., Durand, N.C., Lareau, C.A., Stamenova, E.K., Aiden, E.L., Lander, E.S. & Engreitz, J.M. (2019) ‘Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations’. Nature Genetics 51 (12) p.pp. 1664–1669.

Fyodorov, D. V., Zhou, B.R., Skoultchi, A.I. & Bai, Y. (2018) ‘Emerging roles of linker histones in regulating chromatin structure and function’. Nature Reviews Molecular Cell Biology 19 (3).

Gates, L.A., Foulds, C.E. & O'Malley, B.W. (2017) 'Histone Marks in the "Driver's Seat": Functional Roles in Steering the Transcription Cycle', *Trends in biochemical sciences*, 42(12), pp. 977–989.

Gershman, A., Sauria, M.E.G., Guitart, X., Vollger, M.R., Hook, P.W., Hoyt, S.J., Jain, M., Shumate, A., Razaghi, R., Koren, S., Altemose, N., Caldas, G. V., Logsdon, G.A., Rhie, A., Eichler, E.E., Schatz, M.C., O'Neill, R.J., Phillippy, A.M., Miga, K.H., et al. (2022) 'Epigenetic patterns in a complete human genome', *Science*, 376(6588), .

Ghanbari, M., de Vries, P.S., de Looper, H., Peters, M.J., Schurmann, C., Yaghootkar, H., Dörr, M., Frayling, T.M., Uitterlinden, A.G., Hofman, A., van Meurs, J.B.J., Erkeland, S.J., Franco, O.H. & Dehghan, A. (2014) 'A Genetic variant in the seed region of miR-4513 shows pleiotropic effects on lipid and glucose homeostasis, blood pressure, and coronary artery disease', *Human Mutation*, 35(12), .

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L., Ragoussis, J. & Natoli, G. (2010) 'Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages', *Immunity*, 32(3), pp. 317–328.

Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., Burdett, T., Hayhurst, J., Baker, J., Ferrer, J., Gonzalez-Uriarte, A., Jupp, S., Karim, M.A., Koscielny, G., MacHlitt-Northen, S., et al. (2021) 'Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics', *Nucleic Acids Research*, 49(D1), .

Giammanco, A., Noto, D., Barbagallo, C.M., Nardi, E., Caldarella, R., Ciaccio, M., Averna, M.R. & Cefalù, A.B. (2021) 'Hyperalphalipoproteinemia and beyond: The role of HDL in cardiovascular diseases'. Life 11 (6).

Gibson, G. (2008) 'The environmental contribution to gene expression profiles', *Nature Reviews Genetics 2008 9:8*, 9(8), pp. 575–581.

Gilbert, N., Gilchrist, S. & Bickmore, W.A. (2004) 'Chromatin Organization in the Mammalian Nucleus', *International Review of Cytology*, 242pp. 283–336.

Gillot, C., Favresse, J., Mullier, F., Lecompte, T., Dogné, J.M. & Douxfils, J. (2021) 'NETosis and the Immune System in COVID-19: Mechanisms and Potential Treatments', *Frontiers in Pharmacology*, 12.

Girirajan, S., Campbell, C.D. & Eichler, E.E. (2011) 'Human copy number variation and complex genetic disease', *Annual Review of Genetics*, 45.

Gisterå, A. & Hansson, G.K. (2017) 'The immunology of atherosclerosis', *Nature Reviews Nephrology 2017 13:6*, 13(6), pp. 368–380.

Goldman, A.D. & Landweber, L.F. (2016) 'What Is a Genome?', *PLOS Genetics*, 12(7), p. e1006181.

Gonzalez, E., Kulkarni, H., Bolivar, H., Mangano, A., Sanchez, R., Catano, G., Nibbs, R.J., Freedman, B.I., Quinones, M.P., Bamshad, M.J., Murthy, K.K., Rovin, B.H., Bradley, W.,

Clark, R.A., Anderson, S.A., O'Connell, R.J., Agan, B.K., Ahuja, S.S., Bologna, R., et al. (2005) 'The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility', *Science*, 307(5714), .

Gordon, S. & Taylor, P.R. (2005) 'Monocyte and macrophage heterogeneity'. Nature Reviews Immunology 5 (12).

Grandi, F.C., Modi, H., Kampman, L. & Corces, M.R. (2022) 'Chromatin accessibility profiling by ATAC-seq', *Nature Protocols 2022 17:6*, 17(6), pp. 1518–1552.

Granot-Hershkovitz, E., Wu, P., Karasik, D., Peter, I., Peloso, G.M., Levy, D., Vasan, R.S., Adrienne Cupples, L., Liu, C.T., Meigs, J.B., Siscovick, D.S., Dupuis, J., Friedlander, Y. & Hochner, H. (2020) 'Searching for parent-of-origin effects on cardiometabolic traits in imprinted genomic regions', *European Journal of Human Genetics*, 28(5), .

Grassi, L., Pourfarzad, F., Ullrich, S., Merkel, A., Were, F., Carrillo-de-Santa-Pau, E., Yi, G., Hiemstra, I.H., Tool, A.T.J., Mul, E., Perner, J., Janssen-Megens, E., Berentsen, K., Kerstens, H., Habibi, E., Gut, M., Yaspo, M.L., Linser, M., Lowy, E., et al. (2018) 'Dynamics of Transcription Regulation in Human Bone Marrow Myeloid Differentiation to Mature Blood Neutrophils', *Cell Reports*, 24(10), .

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O., Pousse, Y., Reichen, M., Woolfe, A., Brenan, C., Griffiths, A.D., Vallot, C. & Gérard, A. (2019) 'High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer', *Nature Genetics 2019 51:6*, 51(6), pp. 1060–1066.

Grosveld, F., Van Staalduinen, J. & Stadhouders, R. (2021) 'Transcriptional Regulation by (Super)Enhancers: From Discovery to Mechanisms', *https://doi.org/10.1146/annurev-genom-122220-093818*, 22pp. 127–146.

Gschwind, A.R., Mualim, K.S., Karbalayghareh, A., Sheth, M.U., Dey, K.K., Jagoda, E., Nurtdinov, R.N., Xi, W., Tan, A.S., Jones, H., Ma, X.R., Yao, D., Nasser, J., Avsec, Ž., James, B.T., Shamim, M.S., Durand, N.C., Rao, S.S.P., Mahajan, R., et al. (2023) 'An encyclopedia of enhancer-gene regulatory interactions in the human genome', *bioRxiv*, 17p. 2023.11.09.563812.

Gurtner, A., Borrelli, C., Gonzalez-Perez, I., Bach, K., Acar, I.E., Núñez, N.G., Crepaz, D., Handler, K., Vu, V.P., Lafzi, A., Stirm, K., Raju, D., Gschwend, J., Basler, K., Schneider, C., Slack, E., Valenta, T., Becher, B., Krebs, P., et al. (2023) 'Active eosinophils regulate host defence and immune responses in colitis', *Nature*, 615(7950), .

Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H. & Ohkawa, Y. (2019) 'A chromatin integration labelling method enables epigenomic profiling with lower input', *Nature Cell Biology*, 21(2), .

Harman, J.L. & Jørgensen, H.F. (2019) 'The role of smooth muscle cells in plaque stability: Therapeutic targeting potential', *British journal of pharmacology*, 176(19), pp. 3741–3753.

Hartiala, J.A., Hilser, J.R., Biswas, S., Lusis, A.J. & Allayee, H. (2021) 'Gene-Environment Interactions for Cardiovascular Disease', *Current atherosclerosis reports*, 23(12), p. 75.

Hata, A., Robertson, M., Emi, M. & Lalouel, J.M. (1990) 'Direct detection and automated sequencing of individual alleles after electrophoretic strand separation: Identification of a common nonsense mutation in exon 9 of the human lipoprotein lipase gene', *Nucleic Acids Research*, 18(18), .

He, Z., Si, Y., Jiang, T., Ma, R., Zhang, Y., Cao, M., Li, T., Yao, Z., Zhao, L., Fang, S., Yu, B., Dong, Z., Thatte, H.S., Bi, Y., Kou, J., Yang, S., Piao, D., Hao, L., Zhou, J., et al. (2016) 'Phosphotidylserine exposure and neutrophil extracellular traps enhance procoagulant activity in patients with inflammatory bowel disease', *Thrombosis and haemostasis*, 115(4), pp. 738–751.

Hegele, R.A., Ginsberg, H.N., Chapman, M.J., Nordestgaard, B.G., Kuivenhoven, J.A., Averna, M., Borén, J., Bruckert, E., Catapano, A.L., Descamps, O.S., Hovingh, G.K., Humphries, S.E., Kovanen, P.T., Masana, L., Pajukanta, P., Parhofer, K.G., Raal, F.J., Ray, K.K., Santos, R.D., et al. (2014) 'The polygenic nature of hypertriglyceridaemia: Implications for definition, diagnosis, and management'. The Lancet Diabetes and Endocrinology 2 (8).

Hekselman, I. & Yeger-Lotem, E. (2020) 'Mechanisms of tissue and cell-type specificity in heritable traits and diseases', *Nature Reviews Genetics 2020 21:3*, 21(3), pp. 137–150.

Heller, D.A., de Faire, U., Pedersen, N.L., Dahlen, G. & McClearn, G.E. (1993) 'Genetic and Environmental Influences on Serum Lipid Levels in Twins', *New England Journal of Medicine*, 328(16), .

Helms, C.C., Gladwin, M.T. & Kim-Shapiro, D.B. (2018) 'Erythrocytes and vascular function: Oxygen and nitric oxide'. Frontiers in Physiology 9 (FEB).

Henderson, H.B., Ma, Y., Hassan, M.F., Monsalve, M. V., Marais, A.D., Winkler, F., Gubernator, K., Peterson, J., Brunzell, J.D. & Hayden, M.R. (1991) 'Amino acid substitution (Ile194 → Thr) in exon 5 of the lipoprotein lipase gene causes lipoprotein lipase deficiency in three unrelated probands: Support for a multicentric origin', *Journal of Clinical Investigation*, 87(6), .

Henderson, R., O'Kane, M., McGilligan, V. & Watterson, S. (2016) 'The genetics and screening of familial hypercholesterolaemia'. Journal of Biomedical Science 23 (1).

Hill, W.G., Goddard, M.E. & Visscher, P.M. (2008) 'Data and theory point to mainly additive genetic variance for complex traits', *PLoS Genetics*, 4(2), .

Hirschhorn, J.N. & Daly, M.J. (2005) 'Genome-wide association studies for common diseases and complex traits'. Nature Reviews Genetics 6 (2).

Hivert, V., Sidorenko, J., Rohart, F., Goddard, M.E., Yang, J., Wray, N.R., Yengo, L. & Visscher, P.M. (2021) 'Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals', *American Journal of Human Genetics*, 108(5), .

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A. & Young, R.A. (2013) 'Super-enhancers in the control of cell identity and disease', *Cell*, 155(4), p. 934.

Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E. & Young, R.A. (2015) 'Convergence of Developmental and Oncogenic Signaling Pathways at Transcriptional Super-Enhancers', *Molecular Cell*, 58(2), .

Hochner, H., Allard, C., Granot-Hershkovitz, E., Chen, J., Sitlani, C.M., Sazdovska, S., Lumley, T., McKnight, B., Rice, K., Enquobahrie, D.A., Meigs, J.B., Kwok, P., Hivert, M.F., Borecki, I.B., Gomez, F., Wang, T., van Duijn, C., Amin, N., Rotter, J.I., et al. (2015) 'Parent-of-Origin Effects of the APOB Gene on Adiposity in Young Adults', *PLoS Genetics*, 11(10), .

Hofmeister, R.J., Rubinacci, S., Ribeiro, D.M., Buil, A., Kutalik, Z. & Delaneau, O. (2022) 'Parent-of-Origin inference for biobanks', *Nature Communications*, 13(1), .

Hollingsworth, E.W., Liu, T.A., Jacinto, S.H., Chen, C.X., Alcantara, J.A. & Kvon, E.Z. (2023) 'Rapid and Quantitative Functional Interrogation of Human Enhancer Variant Activity in Live Mice', *bioRxiv*, p. 2023.12.10.570890.

Holtzman, L. & Gersbach, C.A. (2018) 'Editing the Epigenome: Reshaping the Genomic Landscape', *Annu. Rev. Genom. Hum. Genet*, 19pp. 43–71.

Hong, M., Duk Ye, B., Yang, S.-K., Jung, S., Lee, H.-S., Kim, B.M., Lee, S. Bin, Hong, J., Baek, J., Park, S.H., Han, B., Li, Y., Liu, W., Haritunians, T., Taylor, K.D., Rotter, J.I., Bang, S.-Y., Kim, T.-H., Mcgovern, D.P.B., et al. (2018) 'Immunochip Meta-Analysis of Inflammatory Bowel Disease Identifies Three Novel Loci and Four Novel Associations in Previously Reported Loci', *Journal of Crohn's and Colitis*, pp. 730–741.

Hou, L., Xiong, X., Park, Y., Boix, C., James, B., Sun, N., He, L., Patel, A., Zhang, Z., Molinie, B., Van Wittenberghe, N., Steelman, S., Nusbaum, C., Aguet, F., Ardlie, K.G. & Kellis, M. (2023) 'Multitissue H3K27ac profiling of GTEx samples links epigenomic variation to disease', *Nature Genetics*, 55(10), .

Hu, D., Abbasova, L., Schilder, B.M., Nott, A., Skene, N.G. & Marzi, S.J. (2023) 'CUT&Tag recovers up to half of ENCODE ChIP-seq peaks in modifications of H3K27', *bioRxiv*,

Huang, Q., Whitington, T., Gao, P., Lindberg, J.F., Yang, Y., Sun, J., Väisänen, M.R., Szulkin, R., Annala, M., Yan, J., Egevad, L.A., Zhang, K., Lin, R., Jolma, A., Nykter, M., Manninen, A., Wiklund, F., Vaarala, M.H., Visakorpi, T., et al. (2014) 'A prostate cancer susceptibility allele at 6q22 increases RFX6 expression by modulating HOXB13 chromatin binding', *Nature Genetics*, 46(2), .

Hung, C.C., Lin, S.Y., Lee, C.N., Chen, C.P., Lin, S.P., Chao, M.C., Chiou, S.S. & Su, Y.N. (2011) 'Low penetrance of retinoblastoma for p.V654L mutation of the RB1 gene', *BMC Medical Genetics*, 12.

Hunter, D.J. (2005) 'Gene–environment interactions in human diseases', *Nature Reviews Genetics 2005 6:4*, 6(4), pp. 287–298.

Hutchins, P.M. & Heinecke, J.W. (2015) ‘Cholesterol efflux capacity, macrophage reverse cholesterol transport and cardioprotective HDL.’, *Current opinion in lipidology*, 26(5), pp. 388–93.

Huuskonen, J., Olkkonen, V.M., Jauhiainen, M. & Ehnholm, C. (2001) ‘The impact of phospholipid transfer protein (PLTP) on HDL metabolism.’, *Atherosclerosis*, 155(2), pp. 269–81.

Hyun, K., Jeon, J., Park, K. & Kim, J. (2017) ‘Writing, erasing and reading histone lysine methylations’. Experimental and Molecular Medicine 49 (4).

Ilaslan, E., Calvel, P., Nowak, D., Szarras-Czapnik, M., Slowikowska-Hilczer, J., Spik, A., Sararols, P., Nef, S., Jaruzelska, J. & Kusz-Zamelczyk, K. (2018) ‘A Case of Two Sisters Suffering from 46,XY Gonadal Dysgenesis and Carrying a Mutation of a Novel Candidate Sex-Determining Gene STARD8 on the X Chromosome’, *Sexual development : genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation*, 12(4), pp. 191–195.

Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y. & Takahashi, K. (2008) ‘Heterochromatin integrity affects chromosome reorganization after centromere dysfunction’, *Science*, 321(5892), .

Ishii, T., Hasegawa, T., Pai, C.I., Yvgi-Ohana, N., Timberg, R., Zhao, L., Majdic, G., Chung, B.C., Orly, J. & Parker, K.L. (2002) ‘The roles of circulating high-density lipoproteins and trophic hormones in the phenotype of knockout mice lacking the steroidogenic acute regulatory protein’, *Molecular endocrinology (Baltimore, Md.)*, 16(10), pp. 2297–2309.

Ishimoto, N., Park, J.H., Kawakami, K., Tajiri, M., Mizutani, K., Akashi, S., Tame, J.R.H., Inoue, A. & Park, S.Y. (2023) ‘Structural basis of CXC chemokine receptor 1 ligand binding and activation’, *Nature Communications 2023 14:1*, 14(1), pp. 1–11.

Ishimura-Oka, K., Faustinella, F., Kihara, S., Smith, L.C., Oka, K. & Chan, L. (1992) ‘A missense mutation (Trp86→Arg) in exon 3 of the lipoprotein lipase gene: A cause of familial chylomicronemia’, *American Journal of Human Genetics*, 50(6), .

Italiani, P. & Boraschi, D. (2015) ‘New Insights Into Tissue Macrophages: From Their Origin to the Development of Memory’, *Immune Network*, 15(4), .

Iype, J. & Fux, M. (2021) ‘Basophils orchestrating eosinophils’ chemotaxis and function in allergic inflammation’. Cells 10 (4).

Jaenisch, R. & Bird, A. (2003) ‘Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals’, *Nature Genetics 2003 33:3*, 33(3), pp. 245–254.

Jain, N., Moeller, J. & Vogel, V. (2019) ‘Mechanobiology of Macrophages: How Physical Factors Coregulate Macrophage Plasticity and Phagocytosis’. Annual Review of Biomedical Engineering 21.

Jairath, V., Peyrin-Biroulet, L., Zou, G., Mosli, M., Vande Casteele, N., Pai, Reetesh K., Valasek, M.A., Marchal-Bressenot, A., Stitt, L.W., Shackelton, L.M., Khanna, R.,

D'Haens, G.R., Sandborn, W.J., Olson, A., Feagan, B.G. & Pai, Rish K. (2019) 'Responsiveness of histological disease activity indices in ulcerative colitis: a post hoc analysis using data from the TOUCHSTONE randomised controlled trial', *Gut*, 68(7), pp. 1162–1168.

Janowski, B.A., Willy, P.J., Devi, T.R., Falck, J.R. & Mangelsdorf, D.J. (1996) 'An oxysterol signalling pathway mediated by the nuclear receptor LXR alpha', *Nature*, 383(6602), pp. 728–731.

Javierre, B.M., Sewitz, S., Cairns, J., Wingett, S.W., Várnai, C., Thiecke, M.J., Freire-Pritchett, P., Spivakov, M., Fraser, P., Burren, O.S., Cutler, A.J., Todd, J.A., Wallace, C., Wilder, S.P., Kreuzhuber, R., Kostadima, M., Zerbino, D.R., Stegle, O., Kreuzhuber, R., et al. (2016) 'Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters', *Cell*, 167(5), pp. 1369-1384.e19.

Jenne, C.N., Urrutia, R., Kubes, P, Phoebe, C., Snyder, J., Kubes, Paul & of Physiology, D. (2013) 'Platelets: bridging hemostasis, inflammation, and immunity', *International Journal of Laboratory Hematology*, 35(3), pp. 254–261.

Jensen, F.B. (2009) 'The dual roles of red blood cells in tissue oxygen delivery: oxygen carriers and regulators of local blood flow', *Journal of Experimental Biology*, 212(21), pp. 3387–3393.

Jian, B., De la Llera-Moya, M., Ji, Y., Wang, N., Phillips, M.C., Swaney, J.B., Tall, A.R. & Rothblat, G.H. (1998) 'Scavenger receptor class B type I as a mediator of cellular cholesterol efflux to lipoproteins and phospholipid acceptors', *Journal of Biological Chemistry*, 273(10), pp. 5599–5606.

Jiang, T., Zhang, X.O., Weng, Z. & Xue, W. (2021) 'Deletion and replacement of long genomic sequences using prime editing', *Nature Biotechnology 2021 40:2*, 40(2), pp. 227–234.

Jiang, X.C., Jin, W. & Hussain, M.M. (2012) 'The impact of phospholipid transfer protein (PLTP) on lipoprotein metabolism', *Nutrition & metabolism*, 9(1), .

Johansen, C.T., Kathiresan, S. & Hegele, R.A. (2011) 'Genetic determinants of plasma triglycerides'. Journal of Lipid Research 52 (2).

Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. (2007) 'Genome-wide mapping of in vivo protein-DNA interactions', *Science*, 316(5830), .

Jong, M.C., Hofker, M.H. & Havekes, L.M. (1999) 'Role of ApoCs in Lipoprotein Metabolism', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 19(3), pp. 472–484.

Jonkers, I.H. & Wijmenga, C. (2017) 'Context-specific effects of genetic variants associated with autoimmune disease'. Human Molecular Genetics 26 (R2) p.pp. R185–R192.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Philip Schumm, L., Sharma, Y., Anderson, C.A., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S.L., Raychaudhuri, S., Goyette, P., Wei, Z., et al. (2012) 'Host-microbe

interactions have shaped the genetic architecture of inflammatory bowel disease', *Nature*, 491(7422), pp. 119–124.

Julià, A., Domè nech, E., Chaparro, M., García-Sá nchez, V., Gomolló, F., Pané, J., Mañ osa, M., Barreiro-De Acosta, M., Gutié rrez, A., Garcia-Planella, E., Aguas, M., Muñ oz, F., Esteve, M., Mendoza, J.L., Vera, M., Má rquez, L., Tortosa, R., Ló pez-Lasanta, M., Alonso, A., et al. (2014) 'A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative colitis', *Human Molecular Genetics*, pp. 1–8.

Kaikkonen, Minna U, Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C. & Glass, C.K. (2013) *Article Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription*,

Kaikkonen, Minna U., Spann, N.J., Heinz, S., Romanoski, C.E., Allison, K.A., Stender, J.D., Chun, H.B., Tough, D.F., Prinjha, R.K., Benner, C. & Glass, C.K. (2013) 'Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription', *Molecular cell*, 51(3), p. 310.

Kakuta, Y., Kawai, Y., Naito, T., Hirano, A., Umeno, J., Fuyuno, Y., Liu, Z., Li, D., Nakano, T., Izumiyama, Y., Ichikawa, R., Okamoto, D., Nagai, H., Matsumoto, S., Yamamoto, K., Yokoyama, N., Chiba, H., Shimoyama, Y., Onodera, M., et al. (2018) 'A Genome-wide Association Study Identifying RAP1A as a Novel Susceptibility Gene for Crohn's Disease in Japanese Individuals', *Journal of Crohn's and Colitis*, 13(5), pp. 648–658.

Karbalayghareh, A., Sahin, M. & Leslie, C.S. (2022) 'Chromatin interaction-aware gene regulatory modeling with graph attention networks', *Genome research*, 32(5), pp. 930–944.

Karnuta, J.M. & Scacheri, P.C. (2018) 'Enhancers: bridging the gap between gene control and human disease', *Human Molecular Genetics*, 27(R2), pp. R219–R227.

Kassambara, A. (2020) 'ggpubr: "ggplot2" Based Publication Ready Plots. R package version 0.4.0'. https://CRAN.R-project.org/package=ggpubr

Kathiresan, S. & Srivastava, D. (2012) 'Genetics of Human Cardiovascular Disease', *Cell*, 148(6), pp. 1242–1257.

Kaur, G. & Dufour, J.M. (2012) 'Cell lines: Valuable tools or useless artifacts', *Spermatogenesis*, 2(1), p. 1.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K. & Henikoff, S. (2019) 'CUT&Tag for efficient epigenomic profiling of small samples and single cells', *Nature Communications*, 10(1), .

Kedishvili, N.Y. (2016) 'RETINOIC ACID SYNTHESIS AND DEGRADATION', *Sub-cellular biochemistry*, 81p. 127.

Keebler, M.E., Deo, R.C., Surti, A., Konieczkowski, D., Guiducci, C., Burtt, N., Buxbaum, S.G., Sarpong, D.F., Steffes, M.W., Wilson, J.G., Taylor, H.A. & Kathiresan, S. (2010)

‘Fine-mapping in African Americans of 8 recently discovered genetic loci for plasma lipids: The Jackson Heart Study’, *Circulation: Cardiovascular Genetics*, 3(4), .

Keebler, M.E., Sanders, C.L., Surti, A., Guiducci, C., Burtt, N.P. & Kathiresan, S. (2009) ‘Association of blood lipids with common DNA sequence variants at 19 genetic loci in the multiethnic United States national health and nutrition examination survey III’, *Circulation: Cardiovascular Genetics*, 2(3), .

Kenny, E.E., Pe’er, I., Karban, A., Ozelius, L., Mitchell, A.A., Ng, S.M., Erazo, M., Ostrer, H., Abraham, C., Abreu, M.T., Atzmon, G., Barzilai, N., Brant, S.R., Bressman, S., Burns, E.R., Chowers, Y., Clark, L.N., Darvasi, A., Doheny, D., et al. (2012) ‘A genome-wide scan of ashkenazi jewish crohn’s disease suggests novel susceptibility loci’, *PLoS Genetics*, 8(3), .

Kent, D., Marchetti, L., Mikulasova, A., Russell, L.J. & Rico, D. (2023) ‘Broad H3K4me3 domains: Maintaining cellular identity and their implication in super-enhancer hijacking’. BioEssays 45 (10).

Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., Burdett, T., Jupp, S., Parkinson, H., Papatheodorou, I., Yates, A.D., Zerbino, D.R. & Alasoo, K. (2021a) ‘A compendium of uniformly processed human gene expression and splicing quantitative trait loci’, *Nature Genetics 2021 53:9*, 53(9), pp. 1290–1299.

Kerimov, N., Hayhurst, J.D., Peikova, K., Manning, J.R., Walter, P., Kolberg, L., Samoviča, M., Sakthivel, M.P., Kuzmin, I., Trevanion, S.J., Burdett, T., Jupp, S., Parkinson, H., Papatheodorou, I., Yates, A.D., Zerbino, D.R. & Alasoo, K. (2021b) ‘A compendium of uniformly processed human gene expression and splicing quantitative trait loci’, *Nature Genetics 2021 53:9*, 53(9), pp. 1290–1299.

Khan, A., Mathelier, A. & Zhang, X. (2018) ‘Super-enhancers are transcriptionally more active and cell type-specific than stretch enhancers’, *Epigenetics*, 13(9), pp. 910–922.

Kharchenko, P. V., Tolstorukov, M.Y. & Park, P.J. (2008) ‘Design and analysis of ChIP-seq experiments for DNA-binding proteins’, *Nature Biotechnology*, 26(12), .

Kim, H.J. & Jung, Y. (2020) ‘The emerging role of eosinophils as multifunctional leukocytes in health and disease’. Immune Network 20 (3).

Klemm, S.L., Shipony, Z. & Greenleaf, W.J. (2019) ‘Chromatin accessibility and the regulatory epigenome’, *Nature Reviews Genetics 2018 20:4*, 20(4), pp. 207–220.

Kolde, R. (2019) *pheatmap: Pretty Heatmaps*.

Kolovos, P., Knoch, T.A., Grosveld, F.G., Cook, P.R. & Papantonis, A. (2012) ‘Enhancers and silencers: An integrated and simple model for their function’, *Epigenetics and Chromatin*, 5(1), pp. 1–8.

Kondilis-Mangum, H.D. & Wade, P.A. (2013) ‘Epigenetics and the adaptive immune response’, *Molecular Aspects of Medicine*, 34(4), pp. 813–825.

Kong, L., Pokatayev, V., Lefkovith, A., Carter, G.T., Creasey, E.A., Krishna, C., Subramanian, S., Kochar, B., Ashenberg, O., Lau, H., Ananthakrishnan, A.N., Graham, D.B., Deguine, J. & Xavier, R.J. (2023) 'The landscape of immune dysregulation in Crohn's disease revealed through single-cell transcriptomic profiling in the ileum and colon', *Immunity*, 56(2), pp. 444-458.e5.

Kouzarides, T. (2007) 'Chromatin Modifications and Their Function'. Cell 128 (4).

Ku, W.L., Nakamura, K., Gao, W., Cui, K., Hu, G., Tang, Q., Ni, B. & Zhao, K. (2019) 'Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification', *Nature Methods*, 16(4), .

Kvon, E.Z., Waymack, R., Gad, M. & Wunderlich, Z. (2021) 'Enhancer redundancy in development and disease', *Nature reviews. Genetics*, 22(5), p. 324.

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G. & Cremer, T. (2007) 'Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions', *Nature Reviews Genetics 2007 8:2*, 8(2), pp. 104–115.

Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K.I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A.J., Hoffman, M.M., et al. (2012) 'ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia'. Genome Research 22 (9).

De Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G., Heap, G., Nimmo, E.R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D.C., et al. (2017) 'Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease', *Nature Genetics*, 49(2), pp. 256–261.

Lanktree, M.B., Anand, S.S., Yusuf, S. & Hegele, R.A. (2009) 'Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample', *Journal of Lipid Research*, 50(7), .

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., Friedman, N. & Amit, I. (2014) 'Chromatin state dynamics during blood formation.', *Science*, 345(6199), .

Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R. & Buenrostro, J.D. (2019) 'Droplet-based combinatorial indexing for massive-scale single-cell chromatin accessibility', *Nature Biotechnology*, 37(8), .

Larsen, M.C., Lee, J., Jorgensen, J.S. & Jefcoate, C.R. (2020) 'STARD1 Functions in Mitochondrial Cholesterol Metabolism and Nascent HDL Formation. Gene Expression and Molecular mRNA Imaging Show Novel Splicing and a 1:1 Mitochondrial Association', *Frontiers in endocrinology*, 11.

Laville, V., Majarian, T., Sung, Y.J., Schwander, K., Feitosa, M.F., Chasman, D.I., Bentley, A.R., Rotimi, C.N., Cupples, L.A., de Vries, P.S., Brown, M.R., Morrison, A.C., Kraja, A.T.,

Province, M., Gu, C.C., Gauderman, W.J., Laville, V., Majarian, T., Sung, Y.J., et al. (2022) ‘Gene-lifestyle interactions in the genomics of human complex traits’, *European Journal of Human Genetics 2022 30:6*, 30(6), pp. 730–739.

Lawrence, M. (2022) *rtracklayer: R interface to genome browsers and their annotation tracks*.

Lawrence, M. & Obenchain, V. (2022) *GenomicRanges: Representation and Manipulation of Genomic Ranges*.

Lechner, K., von Schacky, C., McKenzie, A.L., Worm, N., Nixdorff, U., Lechner, B., Kränkel, N., Halle, M., Krauss, R.M. & Scherr, J. (2020) ‘Lifestyle factors and high-risk atherosclerosis: Pathways and mechanisms beyond traditional risk factors’, *European Journal of Preventive Cardiology*, 27(4), .

Lee, Y.-C., Lai, C.-Q., Ordovas, J.M. & Parnell, L.D. (2011) ‘A Database of Gene-Environment Interactions Pertaining to Blood Lipid Traits, Cardiovascular Disease and Type 2 Diabetes’, *Journal of data mining in genomics & proteomics*, 2(1), .

Lee-Rueckert, M., Vikstedt, R., Metso, J., Ehnholm, C., Kovanen, P.T. & Jauhiainen, M. (2006) ‘Absence of endogenous phospholipid transfer protein impairs ABCA1-dependent efflux of cholesterol from macrophage foam cells’, *Journal of Lipid Research*, 47pp. 1725–1732.

Ley, K., Laudanna, C., Cybulsky, M.I. & Nourshargh, S. (2007) ‘Getting to the site of inflammation: the leukocyte adhesion cascade updated’, *Nature Reviews Immunology 2007 7:9*, 7(9), pp. 678–689.

Li, H. & Durbin, R. (2009) ‘Fast and accurate short read alignment with Burrows-Wheeler transform’, *Bioinformatics*, 25(14), .

Li, H., Yang, Y., Hong, W., Huang, M., Wu, M. & Zhao, X. (2020) ‘Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects’, *Signal Transduction and Targeted Therapy 2020 5:1*, 5(1), pp. 1–23.

Li, L. & Wunderlich, Z. (2017) ‘An enhancer’s length and composition are shaped by its regulatory task’, *Frontiers in Genetics*, 8(MAY), p. 63.

Li, Q.L., Wang, D.Y., Ju, L.G., Yao, J., Gao, C., Lei, P.J., Li, L.Y., Zhao, X.L. & Wu, M. (2019) ‘The hyper-activation of transcriptional enhancers in breast cancer’, *Clinical Epigenetics*, 11(1), p. 48.

Li, T., Wang, C., Liu, Y., Li, B., Zhang, W., Wang, L., Yu, M., Zhao, X., Du, J., Zhang, J., Dong, Z., Jiang, T., Xie, R., Ma, R., Fang, S., Zhou, J. & Shi, J. (2020) ‘Neutrophil Extracellular Traps Induce Intestinal Damage and Thrombotic Tendency in Inflammatory Bowel Disease’, *Journal of Crohn’s & colitis*, 14(2), pp. 240–253.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S. & Dekker, J. (2009) ‘Comprehensive mapping of long-range interactions reveals folding principles of the human genome’, *Science (New York, N.Y.)*, 326(5950), pp. 289–293.

Lin, X., Liu, Y., Liu, S., Zhu, X., Wu, L., Zhu, Y., Zhao, D., Xu, X., Chemparathy, A., Wang, H., Cao, Y., Nakamura, M., Noordermeer, J.N., Russa, M. La, Hung Wong, W., Zhao, K. & Qi, L.S. (2022) 'Nested epistasis enhancer networks for robust genome regulation', *Science*, 377(6610), .

Linton, M.F., Hasty, A.H., Babaev, V.R. & Fazio, S. (1998) 'Hepatic apo E expression is required for remnant lipoprotein clearance in the absence of the low density lipoprotein receptor', *The Journal of clinical investigation*, 101(8), pp. 1726–1736.

Liu, J.Z., Van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., Abedian, S., Cheon, J.H., Cho, J., Daryani, N.E., Franke, L., Fuyuno, Y., Hart, A., Juyal, R.C., Juyal, G., et al. (2015) 'Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations', *Nature Genetics*, 47(9), pp. 979–986.

Liu, T.-C. & Stappenbeck, T.S. (2016) 'Genetics and Pathogenesis of Inflammatory Bowel Disease', *Annual Review of Pathology: Mechanisms of Disease*, 11(1), pp. 127–148.

Liu, Y., Sarkar, A., Kheradpour, P., Ernst, J. & Kellis, M. (2017) 'Evidence of reduced recombination rate in human regulatory domains', *Genome Biology*, 18(1), pp. 1–11.

Liu, Z., Liu, R., Gao, H., Jung, S., Gao, X., Sun, R., Liu, X., Kim, Y., Lee, H.S., Kawai, Y., Nagasaki, M., Umeno, J., Tokunaga, K., Kinouchi, Y., Masamune, A., Shi, W., Shen, C., Guo, Z., Yuan, K., et al. (2023) 'Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries', *Nature Genetics*, 55(5), .

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., Foster, B., Moser, M., Karasik, E., Gillard, B., Ramsey, K., Sullivan, S., Bridge, J., Magazine, H., Syron, J., et al. (2013) 'The Genotype-Tissue Expression (GTEx) project', *Nature Genetics 2013 45:6*, 45(6), pp. 580–585.

López-Cortegano, E. & Caballero, A. (2019) 'Inferring the nature of missing heritability in human traits using data from the GWAS catalog', *Genetics*, 212(3), .

Luo, R., Yan, J., Oh, J.W., Xi, W., Shigaki, D., Wong, W., Cho, H., Murphy, D., Cutler, R., Rosen, B.P., Pulecio, J., Yang, D., Glenn, R., Chen, T., Li, Q. V, Vierbuchen, T., Sidoli, S., Apostolou, E., Huangfu, D., et al. (2023) 'Dynamic network-guided CRISPRi screen reveals CTCF loop-constrained nonlinear enhancer-gene regulatory activity in cell state transitions', *bioRxiv : the preprint server for biology*,

Luque-Martin, R., Mander, P.K., Leenen, P.J.M. & Winther, M.P.J. (2021) 'Classic and new mediators for in vitro modelling of human macrophages'. Journal of Leukocyte Biology 109 (3).

Ma, J., Lam, I.K.Y., Lau, C.S. & Chan, V.S.F. (2021) 'Elevated interleukin-18 receptor accessory protein mediates enhancement in reactive oxygen species production in neutrophils of systemic lupus erythematosus patients', *Cells*, 10(5), .

Ma, Y., Henderson, H.E., Ven Murthy, M.R., Roederer, G., Monsalve, M. V., Clarke, L.A., Normand, T., Julien, P., Gagné, C., Lambert, M., Davignon, J., Lupien, P.J., Brunzell, J. & Hayden, M.R. (1991) 'A Mutation in the Human Lipoprotein Lipase Gene as the Most

Common Cause of Familial Chylomicronemia in French Canadians', *New England Journal of Medicine*, 324(25), .

Macri, C., Pang, E.S., Patton, T. & O'Keeffe, M. (2018) 'Dendritic cell subsets'. Seminars in Cell and Developmental Biology 84.

Magno, R. & Corty, R.W. (2022) *gwascat: Tools for Querying Human GWAS Catalog*.

Maguire, E.M., Pearce, S.W.A. & Xiao, Q. (2019) 'Foam cell formation: A new target for fighting atherosclerosis and cardiovascular disease'. Vascular Pharmacology 112.

Mailly, F., Tugrul, Y., Reymer, P.W.A., Bruin, T., Seed, M., Groenemeyer, B.F., Asplund-Carlson, A., Vallance, D., Winder, A.F., Miller, G.J., Kastelein, J.J.P., Hamsten, A., Olivecrona, G., Humphries, S.E. & Talmud, P.J. (1995) 'A Common Variant in the Gene for Lipoprotein Lipase (Asp9→Asn)', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 15(4), .

Mak, P.A., Laffitte, B.A., Desrumaux, C., Joseph, S.B., Curtiss, L.K., Mangelsdorf, D.J., Tontonoz, P. & Edwards, P.A. (2002) 'Regulated expression of the apolipoprotein E/C-I/C-IV/C-II gene cluster in murine and human macrophages. A critical role for nuclear liver X receptors alpha and beta.', *The Journal of biological chemistry*, 277(35), pp. 31900–8.

Makałowski, W. (2001) 'The human genome structure and organization.', *Acta Biochimica Polonica*, 48(3), pp. 587–598.

Malysheva, V., Ray-Jones, H., Cazares, T.A., Clay, O., Ohayon, D., Artemov, P., Wayman, J.A., Rosa, M. Della, Petitjean, C., Booth, C., Ellaway, J.I.J., Orchard, W.R., Chen, X., Parameswaran, S., Nagano, T., Fraser, P., Schoenfelder, S., Weirauch, M.T., Kottyan, L.C., et al. (2022) 'High-resolution promoter interaction analysis in Type 3 Innate Lymphoid Cells implicates Batten Disease gene CLN3 in Crohn's Disease aetiology', *bioRxiv*, 6p. 2022.10.19.512842.

Marais, A.D. (2019) 'Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease', *Pathology*, 51(2), pp. 165–176.

Margraf, A. & Perretti, M. (2022) 'Immune Cell Plasticity in Inflammation: Insights into Description and Regulation of Immune Cell Phenotypes', *Cells*, 11(11), .

Markunas, C.A., Johnson, E.O. & Hancock, D.B. (2017) 'Comprehensive evaluation of disease- and trait-specific enrichment for eight functional elements among GWAS-identified variants', *Hum Genet*, 136pp. 911–919.

Martin, F.J., Amode, M.R., Aneja, A., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Becker, A., Bennett, R., Berry, A., Bhai, J., Bhurji, S.K., Bignell, A., Boddu, S., Branco Lins, P.R., Brooks, L., Ramaraju, S.B., Charkhchi, M., Cockburn, A., Da Rin Fiorretto, L., et al. (2023) 'Ensembl 2023', *Nucleic Acids Research*, 51(D1), pp. D933–D941.

Martin-Trujillo, A., Garg, P., Patel, N., Jadhav, B. & Sharp, A.J. (2023) 'Genome-wide evaluation of the effect of short tandem repeat variation on local DNA methylation', *Genome Research*, 33(2), .

Matalonga, J., Glaria, E., Bresque, M., Escande, C., Carbó, J.M., Kiefer, K., Vicente, R., León, T.E., Beceiro, S., Pascual-García, M., Serret, J., Sanjurjo, L., Morón-Ros, S., Riera, A., Paytubi, S., Juarez, A., Sotillo, F., Lindbom, L., Caelles, C., et al. (2017) 'The Nuclear Receptor LXR Limits Bacterial Infection of Host Macrophages through a Mechanism that Impacts Cellular NAD Metabolism.', *Cell reports*, 18(5), pp. 1241–1255.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., et al. (2012) 'Systematic Localization of Common Disease-Associated Variation in Regulatory DNA', *Science*, 337(6099), pp. 1190–1195.

Maurya, S.S. (2021) 'Role of Enhancers in Development and Diseases', *Epigenomes*, 5(4), .

Mayadas, T.N., Cullere, X. & Lowell, C.A. (2014) 'The multifaceted functions of neutrophils', *Annual Review of Pathology: Mechanisms of Disease*, 9.

McCarroll, S.A. & Altshuler, D.M. (2007) 'Copy-number variation and association studies of human disease', *Nature Genetics*, 39(7S), .

McCracken, J.M. & Allen, L.A.H. (2014) 'Regulation of human neutrophil apoptosis and lifespan in health and disease', *Journal of cell death*, 7(1), pp. 15–23.

McErlean, P., Bell, C.G., Hewitt, R.J., Busharat, Z., Ogger, P.P., Ghai, P., Albers, G.J., Calamita, E., Kingston, S., Molyneaux, P.L., Beck, S., Lloyd, C.M., Maher, T.M. & Byrne, A.J. (2021) 'DNA methylome alterations are associated with airway macrophage differentiation and phenotype during lung fibrosis', *American Journal of Respiratory and Critical Care Medicine*, 204(8), .

MD, R., DJ, M. & GK, S. (2010) 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics (Oxford, England)*, 26(1), pp. 139–140.

Mead, J.R., Irvine, S.A. & Ramji, D.P. (2002) 'Lipoprotein lipase: Structure, function, regulation, and role in disease', *Journal of Molecular Medicine*, 80(12), pp. 753–769.

Medina-Rivera, A., Santiago-Algarra, D., Puthier, D. & Spicuglia, S. (2018) 'Widespread Enhancer Activity from Core Promoters', *Trends in Biochemical Sciences*, 43(6), pp. 452–468.

Medzhitov, R. & Horng, T. (2009) 'Transcriptional control of the inflammatory response', *Nature Reviews Immunology 2009 9:10*, 9(10), pp. 692–703.

Mefford, H.C. & Eichler, E.E. (2009) 'Duplication hotspots, rare genomic disorders, and common disease'. Current Opinion in Genetics and Development 19 (3).

Meher, A.K., Spinosa, M., Davis, J.P., Pope, N., Laubach, V.E., Su, G., Serbulea, V., Leitinger, N., Ailawadi, G. & Upchurch, G.R. (2018) 'A novel role of IL-1β in NETosis and

Abdominal Aortic Aneurysms', *Arteriosclerosis, thrombosis, and vascular biology*, 38(4), p. 843.

Melgar, M.F., Collins, F.S. & Sethupathy, P. (2011) 'Discovery of active enhancers through bidirectional expression of short transcripts', *Genome Biology*, 12(11), pp. 1–11.

Menegatti, S., de Kruijf, M., Garcia-Alegria, E., Lacaud, G. & Kouskoff, V. (2019) 'Transcriptional control of blood cell emergence'. FEBS Letters 593 (23).

Merkel, M., Heeren, J., Dudeck, W., Rinninger, F., Radner, H., Breslow, J.L., Goldberg, I.J., Zechner, R. & Greten, H. (2002) 'Inactive lipoprotein lipase (LPL) alone increases selective cholesterol ester uptake in vivo, whereas in the presence of active LPL it also increases triglyceride hydrolysis and whole particle lipoprotein uptake', *Journal of Biological Chemistry*, 277(9), pp. 7405–7411.

Michida, H., Imoto, H., Shinohara, H., Yumoto, N., Seki, M., Umeda, M., Hayashi, T., Nikaido, I., Kasukawa, T., Suzuki, Y. & Okada-Hatakeyama, M. (2020) 'The Number of Transcription Factors at an Enhancer Determines Switch-like Gene Expression', *Cell Reports*, 31(9), p. 107724.

Mikulasova, A., Trevisan-Herraz, M., Fung, K., Ashby, C., Karataraki, N., Morgan, G., Corcoran, A., Hambleton, S., Walker, B., Rico, D. & Russell, L. (2020) *Epigenomic translocation of H3K4me3 broad domains following super-enhancer hijacking*,

Mitchell, K.J. (2012) 'What is complex about complex disorders?' Genome Biology 13 (1).

Mitroulis, I., Kambas, K., Chrysanthopoulou, A., Skendros, P., Apostolidou, E., Kourtzelis, I., Drosos, G.I., Boumpas, D.T. & Ritis, K. (2011) 'Neutrophil Extracellular Trap Formation Is Associated with IL-1β and Autophagy-Related Signaling in Gout', *PLoS ONE*, 6(12), .

Moras, M., Lefevre, S.D. & Ostuni, M.A. (2017) 'From erythroblasts to mature red blood cells: Organelle clearance in mammals'. Frontiers in Physiology 8 (DEC).

Mostafavi, H., Spence, J.P., Naqvi, S. & Pritchard, J.K. (2023) 'Systematic differences in discovery of genetic effects on gene expression and complex traits', *Nature Genetics*, 55(11), .

Mountjoy, E., Schmidt, E.M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., Fumis, L., Hayhurst, J., Buniello, A., Karim, M.A., Wright, D., Hercules, A., Papa, E., Fauman, E.B., Barrett, J.C., Todd, J.A., Ochoa, D., Dunham, I. & Ghoussaini, M. (2021) 'An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci', *Nature Genetics*, 53(11), .

Mozaffari, S. V., DeCara, J.M., Shah, S.J., Sidore, C., Fiorillo, E., Cucca, F., Lang, R.M., Nicolae, D.L. & Ober, C. (2019) 'Parent-of-origin effects on quantitative phenotypes in a large Hutterite pedigree', *Communications Biology*, 2(1), .

Mozaffarian, D., Wilson, P.W.F. & Kannel, W.B. (2008) 'Beyond Established and Novel Risk Factors', *Circulation*, 117(23), pp. 3031–3038.

Mueller, P.A., Yang, L., Ubele, M., Mao, G., Brandon, J., Vandra, J., Nichols, T.C., Escalante-Alcalde, D., Morris, A.J. & Smyth, S.S. (2019) ‘Coronary Artery Disease Risk-Associated Plpp3 Gene and Its Product Lipid Phosphate Phosphatase 3 Regulate Experimental Atherosclerosis’, *Arteriosclerosis, thrombosis, and vascular biology*, 39(11), pp. 2261–2272.

Mulya, A., Lee, J.Y., Gebre, A.K., Boudyguina, E.Y., Chung, S.K., Smith, T.L., Colvin, P.L., Jiang, X.C. & Parks, J.S. (2008) ‘Initial interaction of apoA-I with ABCA1 impacts in vivo metabolic fate of nascent HDL’, *Journal of Lipid Research*, 49(11), p. 2390.

Mumbach, M.R., Satpathy, A.T., Boyle, E.A., Dai, C., Gowen, B.G., Cho, S.W., Nguyen, M.L., Rubin, A.J., Granja, J.M., Kazane, K.R., Wei, Y., Nguyen, T., Greenside, P.G., Corces, M.R., Tycko, J., Simeonov, D.R., Suliman, N., Li, R., Xu, J., et al. (2017) ‘Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements’, *Nature Genetics*, 49(11), pp. 1602–1612.

Murrell, P. (2017) *gridExtra: Miscellaneous Functions for ‘Grid’ Graphics*.

Mutua, V. & Gershwin, L.J. (2020) ‘A Review of Neutrophil Extracellular Traps (NETs) in Disease: Potential Anti-NETs Therapeutics’, *Clinical Reviews in Allergy & Immunology 2020 61:2*, 61(2), pp. 194–211.

Nahrendorf, M. & Swirski, F.K. (2016) ‘Abandoning M1/M2 for a network model of macrophage function’. Circulation Research 119 (3).

Nakato, R. & Sakata, T. (2021) ‘Methods for ChIP-seq analysis: A practical workflow and advanced applications’, *Methods*, 187pp. 44–53.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., Mualim, K., Natri, H.M., Weeks, E.M., Munson, G., Kane, M., Kang, H.Y., Cui, A., Ray, J.P., Eisenhaure, T.M., et al. (2021) ‘Genome-wide enhancer maps link risk variants to disease genes’, *Nature 2021 593:7858*, 593(7858), pp. 238–243.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Natri, H.M., Weeks, E.M., Munson, G., Kane, M., Kang, H.Y., Cui, A., Ray, J.P., Eisenhaure, T.M., Mualim, K., Collins, R.L., et al. (2020) ‘Genome-wide maps of enhancer regulation connect risk variants to disease genes’, *bioRxiv*, p. 2020.09.01.278093.

Nasser, M.W., Raghuwanshi, S.K., Grant, D.J., Jala, V.R., Rajarathnam, K. & Richardson, R.M. (2009) ‘Differential Activation and Regulation of CXCR1 and CXCR2 by CXCL8 monomer and dimer’, *Journal of immunology (Baltimore, Md. : 1950)*, 183(5), p. 3425.

Natoli, G. & Ostuni, R. (2019) ‘Adaptation and memory in immune responses’, *Nature Immunology 2019 20:7*, 20(7), pp. 783–792.

Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., Pagé Sabourin, A., Luca, F., Blekhman, R., Hernandez, R.D., Pique-Regi, R., Tung, J., Yotova, V. & Barreiro, L.B. (2016) ‘Genetic Ancestry and

Natural Selection Drive Population Differences in Immune Responses to Pathogens', *Cell*, 167(3), pp. 657-669.e21.

Neuwirth, E. (2014) *RColorBrewer: ColorBrewer Palettes*.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., et al. (2022) 'The complete sequence of a human genome', *Science*, 376(6588), .

Obata-Ninomiya, K., Domeier, P.P. & Ziegler, S.F. (2020) 'Basophils and Eosinophils in Nematode Infections'. Frontiers in Immunology 11.

Okuno, Y., Inoue, K. & Imai, Y. (2013) 'Novel insights into histone modifiers in adipogenesis', *Adipocyte*, 2(4), .

Ongen, H., Brown, A.A., Delaneau, O., Panousis, N.I., Nica, A.C. & Dermitzakis, E.T. (2017) 'Estimating the causal tissues for complex traits and diseases', *Nature genetics*, 49(12), pp. 1676–1683.

Oram, J.F., Wolfbauer, G., Vaughan, A.M., Tang, C. & Albers, J.J. (2003) *Phospholipid Transfer Protein Interacts with and Stabilizes ATP-binding Cassette Transporter A1 and Enhances Cholesterol Efflux from Cells**,

Orkin, S.H. & Zon, L.I. (2008) 'Hematopoiesis: An Evolving Paradigm for Stem Cell Biology'. Cell 132 (4).

Osterwalder, M., Barozzi, I., Tissiéres, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., Kato, M., Garvin, T.H., Pham, Q.T., Harrington, A.N., Akiyama, J.A., Afzal, V., Lopez-Rios, J., Dickel, D.E., Visel, A., et al. (2018) 'Enhancer Redundancy Allows for Phenotypic Robustness in Mammalian Development', *Nature*, 554(7691), p. 239.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S. & Natoli, G. (2013) 'Latent enhancers activated by stimulation in differentiated cells', *Cell*, 152(1–2), pp. 157–171.

Package, T. & Wickham, H. (2021) 'plyr: Tools for splitting, applying and combining data', *R package version 1.8.6*, 9.

Papayannopoulos, V. (2018) 'Neutrophil extracellular traps in immunity and disease'. Nature Reviews Immunology 18 (2).

Paranjapye, A., Ruffin, M., Harris, A. & Corvol, H. (2020) 'Genetic variation in CFTR and modifier loci may modulate cystic fibrosis disease severity', *Journal of Cystic Fibrosis*, 19.

Park, P.J. (2009) 'ChIP-seq: Advantages and challenges of a maturing technology'. Nature Reviews Genetics 10 (10).

Parker, S.C.J., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., Van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., Axel, V., Pennacchio, L.A. & Collins, F.S. (2013) ‘Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(44), pp. 17921–17926.

Parkes, M., Barrett, J.C., Prescott, N.J., Tremelling, M., Anderson, C.A., Fisher, S.A., Roberts, R.G., Nimmo, E.R., Cummings, F.R., Soars, D., Drummond, H., Lees, C.W., Khawaja, S.A., Bagnall, R., Burke, D.A., Todhunter, C.E., Ahmad, T., Onnie, C.M., McArdle, W., et al. (2007) ‘Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn’s disease susceptibility’, *Nature Genetics*, 39(7), pp. 830–832.

Partida, G.C., Laurin, C., Ring, S.M., Gaunt, T.R., McRae, A.F., Visscher, P.M., Montgomery, G.W., Martin, N.G., Hemani, G., Suderman, M., Relton, C.L., Smith, G.D. & Evans, D.M. (2018) ‘Genome-wide survey of parent-of-origin effects on DNA methylation identifies candidate imprinted loci in humans’, *Human Molecular Genetics*, 27(16), .

Patel, A.A., Zhang, Y., Fullerton, J.N., Boelen, L., Rongvaux, A., Maini, A.A., Bigley, V., Flavell, R.A., Gilroy, D.W., Asquith, B., Macallan, D. & Yona, S. (2017) ‘The fate and lifespan of human monocyte subsets in steady state and systemic inflammation’, *Journal of Experimental Medicine*, 214(7), .

Pell, N., Garcia-Pras, E., Gallego, J., Naranjo-Suarez, S., Balvey, A., Suñer, C., Fernandez-Alfara, M., Chanes, V., Carbo, J., Ramirez-Pedraza, M., Reina, O., Thingholm, L., Bang, C., Rühlemann, M., Franke, A., Schierwagen, R., Rheinwalt, K.P., Trebicka, J., Mendez, R., et al. (2021) ‘Targeting the cytoplasmic polyadenylation element-binding protein CPEB4 protects against diet-induced obesity and microbiome dysbiosis’, *Molecular metabolism*, 54.

Petersen, R., Lambourne, J.J., Javierre, B.M., Grassi, L., Kreuzhuber, R., Ruklisa, D., Rosa, I.M., Tomé, A.R., Elding, H., Van Geffen, J.P., Jiang, T., Farrow, S., Cairns, J., Al-Subaie, A.M., Ashford, S., Attwood, A., Batista, J., Bouman, H., Burden, F., et al. (2017) ‘Platelet function is modified by common sequence variation in megakaryocyte super enhancers’, *Nature communications*, 8.

Pettie, K.P., Mumbach, M., Lea, A.J., Ayroles, J., Chang, H.Y., Kasowski, M. & Fraser, H.B. (2024) ‘Chromatin activity identifies differential gene regulation across human ancestries’, *Genome Biology*, 25(1), pp. 1–23.

Popay, T.M. & Dixon, J.R. (2022) *Coming full circle: On the origin and evolution of the looping model for enhancer-promoter communication*,

Porubsky, D., Höps, W., Ashraf, H., Hsieh, P.H., Rodriguez-Martin, B., Yilmaz, F., Ebler, J., Hallast, P., Maria Maggiolini, F.A., Harvey, W.T., Henning, B., Audano, P.A., Gordon, D.S., Ebert, P., Hasenfeld, P., Benito, E., Zhu, Q., Lee, C., Antonacci, F., et al. (2022) ‘Recurrent inversion polymorphisms in humans associate with genetic instability and genomic disorders’, *Cell*, 185(11), .

Pyne, S., Kong, K.C. & Darroch, P.I. (2004) ‘Lysophosphatidic acid and sphingosine 1-phosphate biology: the role of lipid phosphate phosphatases’, *Seminars in cell & developmental biology*, 15(5), pp. 491–501.

Quang, D.X., Erdos, M.R., Parker, S.C.J. & Collins, F.S. (2015) ‘Motif signatures in stretch enhancers are enriched for disease-associated genetic variants’, *Epigenetics and Chromatin*, 8(1), pp. 1–14.

Rando, O.J. (2012) ‘Combinatorial complexity in chromatin structure and function: revisiting the histone code’, *Current Opinion in Genetics & Development*, 22(2), pp. 148–155.

Ray, R. & Rai, V. (2017) ‘Lysophosphatidic acid converts monocytes into macrophages in both mice and humans’, *Blood*, 129(9), pp. 1177–1183.

Regev, A., Teichmann, S., Lander, E., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P. & Enard, W. (2017) ‘Science Forum: The Human Cell Atlas’, *eLife*,

Remmerie, A. & Scott, C.L. (2018) ‘Macrophages and lipid metabolism.’, *Cellular immunology*, 330pp. 27–42.

Repa, J.J., Liang, G., Ou, J., Bashmakov, Y., Lobaccaro, J.M.A., Shimomura, I., Shan, B., Brown, M.S., Goldstein, J.L. & Mangelsdorf, D.J. (2000) ‘Regulation of mouse sterol regulatory element-binding protein-1c gene (SREBP-1c) by oxysterol receptors, LXRalpha and LXRbeta’, *Genes & development*, 14(22), pp. 2819–2830.

Reymer, P.W.A., Gagné, E., Groenemeyer, B.E., Zhang, H., Forsyth, I., Jansen, H., Seidell, J.C., Kromhout, D., Lie, K.E., Kastelein, J. & Hayden, M.R. (1995) ‘A lipoprotein lipase mutation (Asn291Ser) is associated with reduced HDL cholesterol levels in premature atherosclerosis’, *Nature Genetics*, 10(1), .

Rigau, M., Juan, D., Valencia, A. & Rico, D. (2019) ‘Intronic CNVs and gene expression variation in human populations’, *PLoS Genetics*, 15(1), pp. 1–23.

Riordan, J.D. & Nadeau, J.H. (2017) ‘From Peas to Disease: Modifier Genes, Network Resilience, and the Genetics of Health’. American Journal of Human Genetics 101 (2).

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., et al. (2015) ‘Integrative analysis of 111 reference human epigenomes’, *Nature*, 518(7539), .

Robbins, C.S., Hilgendorf, I., Weber, G.F., Theurl, I., Iwamoto, Y., Figueiredo, J.L., Gorbatov, R., Sukhova, G.K., Gerhardt, L.M.S., Smyth, D., Zavitz, C.C.J., Shikatani, E.A., Parsons, M., Van Rooijen, N., Lin, H.Y., Husain, M., Libby, P., Nahrendorf, M., Weissleder, R., et al. (2013) ‘Local proliferation dominates lesional macrophage accumulation in atherosclerosis’, *Nature medicine*, 19(9), pp. 1166–1172.

Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O.L., He, A., Marra, M., Snyder, M. & Jones, S. (2007) ‘Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing’, *Nature Methods*, 4(8), .

Rohatgi, A., Westerterp, M., von Eckardstein, A., Remaley, A. & Rye, K.-A. (2021) ‘HDL in the 21st Century: A Multifunctional Roadmap for Future HDL Research’, *Circulation*, 143(23), .

Roman, T.S., Marvelle, A.F., Fogarty, M.P., Vadlamudi, S., Gonzalez, A.J., Buchkovich, M.L., Huyghe, J.R., Fuchsberger, C., Jackson, A.U., Wu, Y., Civelek, M., Lusis, A.J., Gaulton, K.J., Sethupathy, P., Kangas, A.J., Soininen, P., Ala-Korpela, M., Kuusisto, J., Collins, F.S., et al. (2015) ‘Multiple Hepatic Regulatory Variants at the GALNT2 GWAS Locus Associated with High-Density Lipoprotein Cholesterol’, *American Journal of Human Genetics*, 97(6), .

Rose, J.R., Akdogan-Ozdilek, B., Rahmberg, A.R., Powell, M.D., Hicks, S.L., Scharer, C.D. & Boss, J.M. (2023) ‘Distinct transcriptomic and epigenomic modalities underpin human memory T cell subsets and their activation potential’, *Communications Biology*, 6(1), .

Ross, C.J.D., Liu, G., Kuivenhoven, J.A., Twisk, J., Rip, J., Van Dop, W., Excoffon, K.J.D.A., Lewis, S.M.E., Kastelein, J.J. & Hayden, M.R. (2005) ‘Complete rescue of lipoprotein lipase-deficient mice by somatic gene transfer of the naturally occurring LPLS447X beneficial mutation’, *Arteriosclerosis, Thrombosis, and Vascular Biology*, 25(10), .

Ross, E.A., Devitt, A. & Johnson, J.R. (2021) ‘Macrophages: The Good, the Bad, and the Gluttony’. Frontiers in Immunology 12.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., Shilo, S., Lador, D., Vila, A.V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B.C., et al. (2018) ‘Environment dominates over host genetics in shaping human gut microbiota’, *Nature 2018 555:7695*, 555(7695), pp. 210–215.

Rouland, A., Masson, D., Lagrost, L., Vergès, B., Gautier, T. & Bouillet, B. (2022) ‘Role of apolipoprotein C1 in lipoprotein metabolism, atherosclerosis and diabetes: a systematic review’, *Cardiovascular Diabetology 2022 21:1*, 21(1), pp. 1–18.

Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., Cramer, P., Kivioja, T. & Taipale, J. (2022) ‘Sequence determinants of human gene regulatory elements’, *Nature Genetics*, 54(3), .

Sanna, S., van Zuydam, N.R., Mahajan, A., Kurilshikov, A., Vich Vila, A., Võsa, U., Mujagic, Z., Masclee, A.A.M., Jonkers, D.M.A.E., Oosting, M., Joosten, L.A.B., Netea, M.G., Franke, L., Zhernakova, A., Fu, J., Wijmenga, C. & McCarthy, M.I. (2019) ‘Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases’, *Nature Genetics 2019 51:4*, 51(4), pp. 600–605.

de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L. & Natoli, G. (2010) ‘A large fraction of extragenic RNA Pol II transcription sites overlap enhancers’, *PLoS Biology*, 8(5), .

Scannell, M., Flanagan, M.B., deStefani, A., Wynne, K.J., Cagney, G., Godson, C. & Maderna, P. (2007) ‘Annexin-1 and peptide derivatives are released by apoptotic cells and stimulate phagocytosis of apoptotic neutrophils by macrophages’, *Journal of immunology (Baltimore, Md. : 1950)*, 178(7), pp. 4595–4605.

Schmidt, F., Marx, A., Baumgarten, N., Hebel, M., Wegner, M., Kaulich, M., Leisegang, M.S., Brandes, R.P., Göke, J., Vreeken, J. & Schulz, M.H. (2021) ‘Integrative analysis of epigenetics data identifies gene-specific regulatory elements’, *Nucleic Acids Research*, 49(18), pp. 10397–10418.

Schnitzler, G.R., Kang, H., Fang, S., Angom, R.S., Lee-Kim, V.S., Ma, X.R., Zhou, R., Zeng, T., Guo, K., Taylor, M.S., Vellarikkal, S.K., Barry, A.E., Sias-Garcia, O., Bloemendal, A., Munson, G., Guckelberger, P., Nguyen, T.H., Bergman, D.T., Hinshaw, S., et al. (2024) ‘Convergence of coronary artery disease genes onto endothelial cell programs’, *Nature*, 626(8000), .

Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C., Absher, D., Aherrahrou, Z., Allayee, H., Altshuler, D., Anand, S.S., Andersen, K., Anderson, J.L., Ardissino, D., Ball, S.G., et al. (2011) ‘Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease’, *Nature Genetics 2011 43:4*, 43(4), pp. 333–338.

Schwartzman, O. & Tanay, A. (2015) ‘Single-cell epigenomics: techniques and emerging applications’, *Nature Reviews Genetics 2015 16:12*, 16(12), pp. 716–726.

Di Scipio, M., Khan, M., Mao, S., Chong, M., Judge, C., Pathan, N., Perrot, N., Nelson, W., Lali, R., Di, S., Morton, R., Petch, J. & Paré, G. (2023) ‘A versatile, fast and unbiased method for estimation of gene-by-environment interaction effects on biobank-scale datasets’, *Nature Communications 2023 14:1*, 14(1), pp. 1–15.

Sciumè, G., Mikami, Y., Jankovic, D., Kanno, Y., O’shea, J.J., Shih, H.-Y., Nagashima, H., Villarino, A. V, Morrison, T., Yao, C., Signorella, S., Sun, H.-W., Brooks, S.R., Fang, D., Sartorelli, V., Nakayamada, S., Hirahara, K., Zitti, B. & Davis, F.P. (2020) *Rapid Enhancer Remodeling and Transcription Factor Repurposing Enable High Magnitude Gene Induction upon Acute Activation of NK Cells*,

Scott, A.J., Chiang, C. & Hall, I.M. (2021) ‘Structural variants are a major source of gene expression differences in humans and often affect multiple nearby genes’, *Genome Research*, 31(12), .

Shao, B., Tang, C., Sinha, A., Mayer, P.S., Davenport, G.D., Brot, N., Oda, M.N., Zhao, X.Q. & Heinecke, J.W. (2014) ‘Humans with atherosclerosis have impaired ABCA1 cholesterol efflux and enhanced high-density lipoprotein oxidation by myeloperoxidase’, *Circulation Research*, 114(11), pp. 1733–1742.

Shao, H., Burrage, L.C., Sinasac, D.S., Hill, A.E., Ernest, S.R., O'Brien, W., Courtland, H.W., Jepsen, K.J., Kirby, A., Kulbokase, E.J., Dalye, M.J., Broman, K.W., Lander, E.S. & Nadeau, J.H. (2008) 'Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis', *Proceedings of the National Academy of Sciences of the United States of America*, 105(50), .

Sharifi-Zarchi, A., Gerovska, D., Adachi, K., Totonchi, M., Pezeshk, H., Taft, R.J., Schöler, H.R., Chitsaz, H., Sadeghi, M., Baharvand, H. & Araúzo-Bravo, M.J. (2017) 'DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism', *BMC Genomics*, 18(1), pp. 1–21.

Sheffield, N.C., Thurman, R.E., Song, L., Safi, A., Stamatoyannopoulos, J.A., Lenhard, B., Crawford, G.E. & Furey, T.S. (2013) 'Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions', *Genome research*, 23(5), pp. 777–788.

Shepherd, L. & Lawrence, M. (2022) *plyranges: A fluent interface for manipulating GenomicRanges*.

Sheppard, B., Rappoport, N., Loh, P.R., Sanders, S.J., Zaitlen, N. & Dahl, A. (2021) 'A model and test for coordinated polygenic epistasis in complex traits', *Proceedings of the National Academy of Sciences of the United States of America*, 118(15), .

Simeonov, D.R., Gowen, B.G., Boontanrart, M., Roth, T.L., Gagnon, J.D., Mumbach, M.R., Satpathy, A.T., Lee, Y., Bray, N.L., Chan, A.Y., Lituiev, D.S., Nguyen, M.L., Gate, R.E., Subramaniam, M., Li, Z., Woo, J.M., Mitros, T., Ray, G.J., Curie, G.L., et al. (2017) 'Discovery of stimulation-responsive immune enhancers with CRISPR activation', *Nature 2017 549:7670*, 549(7670), pp. 111–115.

Singh, S. & Bittner, V. (2015) 'Familial Hypercholesterolemia—Epidemiology, Diagnosis, and Screening'. Current Atherosclerosis Reports 17 (2).

Skaar, D.A., Li, Y., Bernal, A.J., Hoyo, C., Murphy, S.K. & Jirtle, R.L. (2012) 'The human imprintome: regulatory mechanisms, methods of ascertainment, and roles in disease susceptibility.' ILAR journal / National Research Council, Institute of Laboratory Animal Resources 53 (3–4).

Skålén, K., Gustafsson, M., Knutsen Rydberg, E., Hultén, L.M., Wiklund, O., Innerarity, T.L. & Boren, J. (2002) 'Subendothelial retention of atherogenic lipoproteins in early atherosclerosis', *Nature 2002 417:6890*, 417(6890), pp. 750–754.

Skene, P.J. & Henikoff, S. (2017) 'An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites', *eLife*, 6.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C. & Nobrega, M.A. (2012) 'Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease', *Human Molecular Genetics*, 21(14), pp. 3255–3263.

Smith, G.D., Ebrahim, S., Lewis, S., Hansell, A.L., Palmer, L.J. & Burton, P.R. (2005) 'Genetic epidemiology and public health: hope, hype, and future prospects', *The Lancet*, 366(9495), pp. 1484–1498.

Soldi, M., Mari, T., Nicosia, L., Musiani, D., Sigismondo, G., Cuomo, A., Pavesi, G. & Bonaldi, T. (2017) ‘Chromatin proteomics reveals novel combinatorial histone modification signatures that mark distinct subpopulations of macrophage enhancers’, *Nucleic Acids Research*, 45(21), .

Song, W., Sharan, R. & Ovcharenko, I. (2019) ‘The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance’, *Genome Biology*, 20(1), .

Sonnenberg, G.F. & Hepworth, M.R. (2019) ‘Functional interactions between innate lymphoid cells and adaptive immunity’. Nature Reviews Immunology 19 (10).

Soskic, B., Cano-Gamez, E., Smyth, D.J., Ambridge, K., Ke, Z., Matte, J.C., Bossini-Castillo, L., Kaplanis, J., Ramirez-Navarro, L., Lorenc, A., Nakic, N., Esparza-Gordillo, J., Rowan, W., Wille, D., Tough, D.F., Bronson, P.G. & Trynka, G. (2022) ‘Immune disease risk variants regulate gene expression dynamics during CD4+ T cell activation’, *Nature genetics*, 54(6), pp. 817–826.

Sotillos, S., von der Decken, I., Mercadé, I.D., Srinivasan, S., Sirokha, D., Livshits, L., Vanni, S., Nef, S., Biason-Lauber, A., Gutiérrez, D.R. & Hombría, J.C.G. (2022) ‘A conserved function of Human DLC3 and Drosophila Cv-c in testis development’, *eLife*, 11.

Spicuglia, S., Santiago-Algarra, D., Dao, L.T.M., Pradel, L. & España, A. (2017) ‘Recent advances in high-throughput approaches to dissect enhancer function’, *F1000Research*, 6p. 939.

Spitz, F. & Furlong, E.E.M. (2012) ‘Transcription factors: from enhancer binding to developmental control’, *Nature Reviews Genetics 2012 13:9*, 13(9), pp. 613–626.

Stein, R., Ferrari, F. & Scolari, F. (2019) ‘Genetics, Dyslipidemia, and Cardiovascular Disease: New Insights’. Current Cardiology Reports 21 (8).

Stocco, D.M. (2000) ‘Intramitochondrial cholesterol transfer’, *Biochimica et biophysica acta*, 1486(1), pp. 184–197.

Stone, K.D., Prussin, C. & Metcalfe, D.D. (2010) ‘IgE, mast cells, basophils, and eosinophils’, *Journal of Allergy and Clinical Immunology*, 125(2 SUPPL. 2), .

Stricker, S.H., Köferle, A. & Beck, S. (2016) ‘From profiles to function in epigenomics’, *Nature Reviews Genetics 2016 18:1*, 18(1), pp. 51–66.

Strohman, R. (2002) ‘Maneuvering in the complex path from genotype to phenotype’. Science 296 (5568).

Stunnenberg, H.G., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., Amit, I., Antonarakis, S.E., Aparicio, S., Arima, T., Arrigoni, L., Arts, R., Asnafi, V., Badosa, M.E., Bae, J.B., Bassler, K., Beck, S., Berkman, B., Bernstein, B.E., et al. (2016a) ‘The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery’, *Cell*, 167(5), pp. 1145–1149.

Stunnenberg, H.G., Abrignani, S., Adams, D., de Almeida, M., Altucci, L., Amin, V., Amit, I., Antonarakis, S.E., Aparicio, S., Arima, T., Arrigoni, L., Arts, R., Asnafi, V., Badosa, M.E., Bae, J.B., Bassler, K., Beck, S., Berkman, B., Bernstein, B.E., et al. (2016b) 'The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery', *Cell*, 167(5), pp. 1145–1149.

Subbannayya, Y., Haug, M., Pinto, S.M., Mohanty, V., Meås, H.Z., Flo, T.H., Keshava Prasad, T.S. & Kandasamy, R.K. (2021) 'The proteomic landscape of resting and activated cd4+ t cells reveal insights into cell differentiation and function', *International Journal of Molecular Sciences*, 22(1), .

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.Y., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Casale, F.P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., et al. (2015) 'An integrated map of structural variation in 2,504 human genomes', *Nature*, 526(7571), .

Sun, W., Poschmann, J., Cruz-Herrera del Rosario, R., Parikshak, N.N., Hajan, H.S., Kumar, V., Ramasamy, R., Belgard, T.G., Elanggovan, B., Wong, C.C.Y., Mill, J., Geschwind, D.H. & Prabhakar, S. (2016) 'Histone Acetylome-wide Association Study of Autism Spectrum Disorder', *Cell*, 167(5), .

Suñer, C., Sibilio, A., Martín, J., Castellazzi, C.L., Reina, O., Dotu, I., Caballé, A., Rivas, E., Calderone, V., Díez, J., Nebreda, A.R. & Méndez, R. (2022) 'Macrophage inflammation resolution requires CPEB4-directed offsetting of mRNA degradation', *eLife*, 11.

Tabas, I., García-Cardeña, G. & Owens, G.K. (2015) 'Recent insights into the cellular biology of atherosclerosis'. Journal of Cell Biology 209 (1).

Tada, H., Fujino, N., Hayashi, K., Kawashiri, M. aki & Takamura, M. (2022) 'Human genetics and its impact on cardiovascular disease', *Journal of Cardiology*, 79(2), pp. 233–239.

Tafessu, A. & Banaszynski, L.A. (2020a) 'Establishment and function of chromatin modification at enhancers', *Open biology*, 10(10), p. 200255.

Tafessu, A. & Banaszynski, L.A. (2020b) 'Establishment and function of chromatin modification at enhancers', *Open biology*, 10(10), p. 200255.

Tennekes, M. (2020) *treemap: Treemap Visualization*.

Teslovich, T.M., Musunuru, K., Smith, A. V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., Johansen, C.T., Fouchier, S.W., Isaacs, A., Peloso, G.M., Barbalic, M., Ricketts, S.L., Bis, J.C., Aulchenko, Y.S., Thorleifsson, G., et al. (2010) 'Biological, clinical and population relevance of 95 loci for blood lipids', *Nature*, 466(7307), .

The ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B.R., Landt, S.G., Lee, B.K., Pauli, F., Rosenbloom, K.R., Sabo, P., Safi, A., et al. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature 2012 489:7414*, 489(7414), pp. 57–74.

Therrien, A., Chapuy, L., Bsat, M., Rubio, M., Bernard, G., Arslanian, E., Orlicka, K., Weber, A., Panzini, B.P., Dorais, J., Bernard, E.J., Soucy, G., Bouin, M. & Sarfati, M. (2019) 'Recruitment of activated neutrophils correlates with disease severity in adult Crohn's disease', *Clinical and experimental immunology*, 195(2), pp. 251–264.

Thibodeau, A., Márquez, E.J., Shin, D.G., Vera-Licona, P. & Ucar, D. (2017) 'Chromatin interaction networks revealed unique connectivity patterns of broad H3K4me3 domains and super enhancers in 3D chromatin', *Scientific Reports*, 7(1), pp. 1–12.

Thul, P.J. & Lindskog, C. (2018) 'The human protein atlas: A spatial map of the human proteome', *Protein Science : A Publication of the Protein Society*, 27(1), p. 233.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., et al. (2012) 'The accessible chromatin landscape of the human genome', *Nature 2012 489:7414*, 489(7414), pp. 75–82.

Tierney, N. (2019) *superheat: Heatmap Visualisation with 'ggplot2'.*

Trerotola, M., Relli, V., Simeone, P. & Alberti, S. (2015) 'Epigenetic inheritance and the missing heritability', *Human genomics*, 9(1), .

Trinder, M., Francis, G.A. & Brunham, L.R. (2020) 'Association of Monogenic vs Polygenic Hypercholesterolemia with Risk of Atherosclerotic Cardiovascular Disease', *JAMA Cardiology*, 5(4), .

Trojer, P. & Reinberg, D. (2007) 'Facultative Heterochromatin: Is There a Distinctive Molecular Signature?' Molecular Cell 28 (1).

Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. & Raychaudhuri, S. (2012) 'Chromatin marks identify critical cell types for fine mapping complex trait variants', *Nature Genetics 2012 45:2*, 45(2), pp. 124–130.

Tsai, H.C., Pietrobon, V., Peng, M., Wang, S., Zhao, L., Marincola, F.M. & Cai, Q. (2022) 'Current strategies employed in the manipulation of gene expression for clinical purposes', *Journal of Translational Medicine 2022 20:1*, 20(1), pp. 1–21.

van Unen, V., Ouboter, L.F., Li, N., Schreurs, M., Abdelaal, T., Kooy-Winkelaar, Y., Beyrend, G., Höllt, T., Maljaars, P.W.J., Mearin, M.L., Mahfouz, A., Witte, A.M.C., Clemens, C.H.M., Abraham, S., Escher, J.C., Lelieveldt, B.P.F., Pascutti, M.F., van der Meulen – de Jong, A.E. & Koning, F. (2022) 'Identification of a Disease-Associated Network of Intestinal Immune Cells in Treatment-Naive Inflammatory Bowel Disease', *Frontiers in Immunology*, 13p. 893803.

Vallejo-Vaz, A.J. & Ray, K.K. (2018) 'Epidemiology of familial hypercholesterolaemia: Community and clinical'. Atherosclerosis 277.

Varol, C., Mildner, A. & Jung, S. (2015) 'Macrophages: Development and Tissue Specialization', *https://doi.org/10.1146/annurev-immunol-032414-112220*, 33pp. 643–675.

De Vas, M.G., Garstang, M.G., Joshi, S.S., Khan, T.N., Atla, G., Parry, D., Moore, D., Cebola, I., Zhang, S., Cui, W., Lampe, A.K., Lam, W.W., FitzPatrick, D.R., Ferrer, J., Pradeepa, M.M. & Atanur, S.S. (2019) 'De novo variants in population constrained fetal brain enhancers and intellectual disability', *bioRxiv*,

Vasquez, L.J., Mann, A.L., Chen, L. & Soranzo, N. (2016) ' From GWAS to function: lessons from blood cells ', *ISBT Science Series*, 11(S1), pp. 211–219.

Verine, A., Benkirane, M., Meignen, J.M. & Boyer, J. (1989) 'Lipoprotein lipase in rat heart--II. Influence of apolipoproteins and nutritional factors on tri-, di- and monoacylglycerol lipase activities in post-heparin effluents', *Comparative biochemistry and physiology. B, Comparative biochemistry*, 94(1), pp. 19–25.

Vignali, D.A.A., Collison, L.W. & Workman, C.J. (2008) 'How regulatory T cells work'. Nature Reviews Immunology 8 (7).

Villanueva, R.A.M., Chen, Z.J. & Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis Using the Grammar of Graphics*. p.pp. 160–167.

Viola, A., Munari, F., Sánchez-Rodríguez, R., Scolaro, T. & Castegna, A. (2019) 'The metabolic signature of macrophage responses'. Frontiers in Immunology 10 (JULY).

Virolainen, S.J., VonHandorf, A., Viel, K.C.M.F., Weirauch, M.T. & Kottyan, L.C. (2022) 'Gene–environment interactions and their impact on human health', *Genes & Immunity 2022 24:1*, 24(1), pp. 1–11.

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A. & Yang, J. (2017) '10 Years of GWAS Discovery: Biology, Function, and Translation', *The American Journal of Human Genetics*, 101(1), pp. 5–22.

Vivier, E., Tomasello, E., Baratin, M., Walzer, T. & Ugolini, S. (2008) 'Functions of natural killer cells'. Nature Immunology 9 (5).

Vogel, A., Brunner, J.S., Hajto, A., Sharif, O. & Schabbauer, G. (2022) 'Lipid scavenging macrophages and inflammation.', *Biochimica et biophysica acta. Molecular and cell biology of lipids*, 1867(1), p. 159066.

Wang, H., Zhang, F., Zeng, J., Wu, Y., Kemper, K.E., Xue, A., Zhang, M., Powell, J.E., Goddard, M.E., Wray, N.R., Visscher, P.M., McRae, A.F. & Yang, J. (2019) 'Genotype-by-environment interactions inferred from genetic effects on phenotypic variability in the UK Biobank', *Science Advances*, 5(8), .

Wang, H.X., Li, M., Lee, C.M., Chakraborty, S., Kim, H.W., Bao, G. & Leong, K.W. (2017) 'CRISPR/Cas9-Based Genome Editing for Disease Modeling and Therapy: Challenges and Opportunities for Nonviral Delivery', *Chemical reviews*, 117(15), pp. 9874–9906.

Wang, J., Cao, H., Ban, M.R., Kennedy, B.A., Zhu, S., Anand, S., Yusuf, S., Pollex, R.L. & Hegele, R.A. (2007) 'Resequencing genomic DNA of patients with severe

hypertriglyceridemia (MIM 144650)', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 27(11), .

Wang, N., Silver, D.L., Thiele, C. & Tall, A.R. (2001) *ATP-binding Cassette Transporter A1 (ABCA1) Functions as a Cholesterol Efflux Regulatory Protein**,

Wang, Q., Xiong, H., Ai, S., Yu, X., Liu, Y., Zhang, J. & He, A. (2019) 'CoBATCH for High-Throughput Single-Cell Epigenomic Profiling', *Molecular Cell*, 76(1), .

Wang, X. & Goldstein, D.B. (2020) 'Enhancer Domains Predict Gene Pathogenicity and Inform Gene Discovery in Complex Disease', *American Journal of Human Genetics*, 106(2), pp. 215–233.

Ward, L.D. & Kellis, M. (2012) 'Interpreting noncoding genetic variation in complex traits and human disease'. Nature Biotechnology 30 (11).

Warnes, G.R. (2021) *gtools: Various R Programming Tools*.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M. & Posthuma, D. (2019) 'A global overview of pleiotropy and genetic architecture in complex traits', *Nature Genetics*, 51(9), .

Watkins, H. & Farrall, M. (2006) 'Genetic susceptibility to coronary artery disease: from promise to progress', *Nature Reviews Genetics 2006 7:3*, 7(3), pp. 163–173.

Wculek, S.K., Dunphy, G., Heras-Murillo, I., Mastrangelo, A. & Sancho, D. (2022) 'Metabolism of tissue macrophages in homeostasis and pathology'. Cellular and Molecular Immunology 19 (3).

Weedon, M.N., Cebola, I., Patch, A.M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H.H., Allen, H.L., Houghton, J.A.L., Roth, C.L., Chen, R., Hussain, K., Marsh, P., Vallier, L., Murray, A., Ellard, S., Ferrer, J., et al. (2013) 'Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis', *Nature Genetics 2013 46:1*, 46(1), pp. 61–64.

Wei, W.H., Hemani, G. & Haley, C.S. (2014) 'Detecting epistasis in human complex traits'. Nature Reviews Genetics 15 (11).

Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J.O. (2013) 'Phenotypic impact of genomic structural variation: Insights from and for human disease'. Nature Reviews Genetics 14 (2).

Whalen, S., Truty, R.M. & Pollard, K.S. (2016) 'Enhancer–promoter interactions are encoded by complex genomic signatures on looping chromatin', *Nature Genetics 2016 48:5*, 48(5), pp. 488–496.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. & Young, R.A. (2013) 'Master transcription factors and mediator establish super-enhancers at key cell identity genes', *Cell*, 153(2), pp. 307–319.

Wickham, H. (2022) *forcats: Tools for Working with Categorical Variables (Factors)*.

Wickham, H. (2020) *scales: Scale Functions for Visualization*.

Wickham, H., François, R., Henry, L. & Müller, K. (2021) 'dplyr: A Grammar of Data Manipulation. R package version 1.0.8.' Media p.pp. 1–88.

Wickham, H. & Henry, L. (2021) 'tidyr: Tidy Messy Data', *R package version 1.1.3*,

Willer, C.J., Schmidt, E.M., Sengupta, S., Peloso, G.M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M.L., Mora, S., Beckmann, J.S., Bragg-Gresham, J.L., Chang, H.Y., Demirkan, A., Den Hertog, H.M., Do, R., Donnelly, L.A., Ehret, G.B., Esko, T., et al. (2013) 'Discovery and refinement of loci associated with lipid levels', *Nature Genetics*, 45(11), .

Williams, K.J. & Tabas, I. (1995) 'The Response-to-Retention Hypothesis of Early Atherogenesis', *Arteriosclerosis, Thrombosis, and Vascular Biology*, 15(5), pp. 551–562.

Wolf, J.B., Cheverud, J.M., Roseman, C. & Hager, R. (2008) 'Genome-wide analysis reveals a complex pattern of genomic imprinting in mice', *PLoS Genetics*, 4(6), .

World Health Organization (2021) *Cardiovascular diseases (CVDs)*. [Online]

Wu, H., Eckhardt, C.M. & Baccarelli, A.A. (2023) 'Molecular mechanisms of environmental exposures and human disease', *Nature Reviews Genetics 2023 24:5*, 24(5), pp. 332–344.

Wuttge, D.M., Romert, A., Eriksson, U., Törmä, H., Hansson, G.K. & Sirsjö, A. (2001) 'Induction of CD36 by all-trans retinoic acid: retinoic acid receptor signaling in the pathogenesis of atherosclerosis', *The FASEB Journal*, 15(7), pp. 1221–1223.

Wutz, A. (2011) 'Gene silencing in X-chromosome inactivation: Advances in understanding facultative heterochromatin formation'. Nature Reviews Genetics 12 (8).

Xue, Y., Luo, M., Hu, X., Li, X., Shen, J., Zhu, W., Huang, L., Hu, Y., Guo, Y., Liu, L., Wang, L. & Luo, S. (2022) 'Macrophages regulate vascular smooth muscle cell function during atherosclerosis progression through IL-1β/STAT3 signaling', *Communications Biology 2022 5:1*, 5(1), pp. 1–16.

Yamazaki, K., Umeno, J., Takahashi, A., Hirano, A., Johnson, T.A., Kumasaka, N., Morizono, T., Hosono, N., Kawaguchi, T., Takazoe, M., Yamada, T., Suzuki, Y., Tanaka, H., Motoya, S., Hosokawa, M., Arimura, Y., Shinomura, Y., Matsui, T., Matsumoto, T., et al. (2013) 'A genome-wide association study identifies 2 susceptibility loci for crohn's disease in a Japanese population', *Gastroenterology*, 144(4), pp. 781–788.

Yan, J. & Horng, T. (2020) 'Lipid Metabolism in Regulation of Macrophage Functions'. Trends in Cell Biology 30 (12).

Yang, A.Z. & Jostins-Dean, L. (2022) 'Environmental variables and genome-environment interactions predicting IBD diagnosis in large UK cohort', *Scientific Reports 2022 12:1*, 12(1), pp. 1–13.

Yang, S.K., Hong, M., Zhao, W., Jung, Y., Baek, J., Tayebi, N., Kim, K.M., Ye, B.D., Kim, K.J., Park, S.H., Lee, I., Lee, E.J., Kim, W.H., Cheon, J.H., Kim, Y.H., Jang, B.I., Kim, H.S.,

Choi, J.H., Koo, J.S., et al. (2014) ‘Genome-wide association study of Crohn’s disease in Koreans revealed three new susceptibility loci and common attributes of genetic susceptibility across ethnic populations’, *Gut*, 63(1), pp. 80–87.

Yang, Y., Chung, E.K., Yee, L.W., Savelli, S.L., Nagaraja, H.N., Zhou, B., Hebert, M., Jones, K.N., Shu, Y., Kitzmiller, K., Blanchong, C.A., McBride, K.L., Higgins, G.C., Rennebohm, R.M., Rice, R.R., Hackshaw, K. V., Roubey, R.A.S., Grossman, J.M., Tsao, B.P., et al. (2007) ‘Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): Low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans’, *American Journal of Human Genetics*, 80(6), .

Yousefi, P.D., Suderman, M., Langdon, R., Whitehurst, O., Davey Smith, G. & Relton, C.L. (2022) ‘DNA methylation-based predictors of health: applications and statistical considerations’. Nature Reviews Genetics 23 (6).

Yu, X.H., Fu, Y.C., Zhang, D.W., Yin, K. & Tang, C.K. (2013) ‘Foam cells in atherosclerosis’. Clinica Chimica Acta 424.

Yu, Y.R. & Rodriguez, J.R. (2017) ‘Clinical presentation of Crohn’s, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes’, *Seminars in Pediatric Surgery*, 26(6), .

Zanoni, P. & Von Eckardstein, A. (2020) ‘Inborn errors of apolipoprotein A-I metabolism: implications for disease, research and development’, *Current Opinion in Lipidology*, 31(2), .

Zarei, L., Bahrami, M., Farhad, N., Froushani, S.M.A. & Abbasi, A. (2018) ‘All-trans retinoic acid effectively reduces atheroma plaque size in a rabbit model of high-fat-induced atherosclerosis’, *Advances in Clinical and Experimental Medicine*, 27(12), pp. 1631–1636.

Zeng, B., Shi, S., Ashworth, G., Dong, C., Liu, J. & Xing, F. (2019) ‘ILC3 function as a double-edged sword in inflammatory bowel diseases’, *Cell Death & Disease 2019 10:4*, 10(4), pp. 1–12.

Zeng, Y., Amador, C., Xia, C., Marioni, R., Sproul, D., Walker, R.M., Morris, S.W., Bretherick, A., Canela-Xandri, O., Boutin, T.S., Clark, D.W., Campbell, A., Rawlik, K., Hayward, C., Nagy, R., Tenesa, A., Porteous, D.J., Wilson, J.F., Deary, I.J., et al. (2019) ‘Parent of origin genetic effects on methylation in humans are common and influence complex trait variation’, *Nature Communications*, 10(1), .

Zentner, G.E., Tesar, P.J. & Scacheri, P.C. (2011) ‘Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions’, *Genome Research*, 21(8), .

Zhang, Q. & Cao, X. (2019) ‘Epigenetic regulation of the innate immune response to infection’, *Nature Reviews Immunology 2019 19:7*, 19(7), pp. 417–432.

Zhang, X., Kluger, Y., Nakayama, Y., Poddar, R., Whitney, C., DeTora, A., Weissman, S.M. & Newburger, P.E. (2004) 'Gene expression in mature neutrophils: early responses to inflammatory stimuli', *Journal of leukocyte biology*, 75(2), pp. 358–372.

Zhang, Y., An, L., Yue, F. & Hardison, R.C. (2016) 'Jointly characterizing epigenetic dynamics across multiple human cell types', *Nucleic acids research*, 44(14), pp. 6721–6731.

Zhao, Yang, Zou, W., Du, J. & Zhao, Yong (2018) 'The origins and homeostasis of monocytes and tissue-resident macrophages in physiological situation'. Journal of Cellular Physiology 233 (10).

Zhou, G., Yu, L., Fang, L., Yang, W., Yu, T., Miao, Y., Chen, M., Wu, K., Chen, F., Cong, Y. & Liu, Z. (2018) 'CD177+ neutrophils as functionally activated neutrophils negatively regulate IBD', *Gut*, 67(6), pp. 1052–1063.

Zhu, J., He, F., Hu, S. & Yu, J. (2008) 'On the nature of human housekeeping genes', *Trends in Genetics*, 24(10), pp. 481–484.