FUNCTIONAL CHARACTERISATION OF THE THREE-DIMENSIONAL EPIGENETIC LANDSCAPE IN ANKYLOSING SPONDYLITIS-LINKED GENOME-WIDE ASSOCIATION STUDY LOCI

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Musculoskeletal Sciences at the University of Oxford, UK

59,912 words

CONNOR IAN DAVIDSON

Wolfson College

&

The Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS)

MICHAELMAS TERM, 2024

Supervisors: Prof. Julian Knight, Dr. Matteo Vecellio, Dr. Carla Cohen, Prof. Bryan Paul Wordsworth

ABSTRACT

FUNCTIONAL CHARACTERISATION OF THE THREE-DIMENSIONAL EPIGENETIC LANDSCAPE IN ANKYLOSING SPONDYLITIS-LINKED GENOME-WIDE ASSOCIATION STUDY LOCI

Connor Ian Davidson, Wolfson College, Michaelmas Term 2024

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford, UK

Ankylosing spondylitis (AS) is a highly heritable chronic arthritic condition characterised by inflammation of the sacroiliac joints and spinal entheses. Irregularities in bone erosion and deposition are seen at the site of inflammation, which can lead to spinal fusion and severe disability. The first genetic association was described in the MHC class I associated gene HLA-B27 and subsequent genetic studies have identified over 100 disease-associated loci. The majority of these associations are found outside protein-coding regions, and studies functionally characterising them remain scarce. This thesis examines the epigenetic changes that underlie AS-associated genetic variants in immune cell subsets from the perspective of multi-omics, focusing primarily on integrative modelling of non-MHC loci with chromatin hidden Markov models (ChromHMM) and detecting looping interactions between genetic elements with chromosome conformation capture (3C)-based methods.

In this thesis, 16 chromatin states were defined by ChromHMM upon integration of open chromatin, histone post-translational modification, and gene expression datasets from $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells obtained from AS patients and healthy controls. AS-associated SNPs were found to be enriched in *cis*-regulatory elements including active enhancers and promoters and globally there were significant differences in the epigenetic landscape between AS patients and healthy controls. Capture-C was then used to characterise looping interactions between 44 lead GWAS SNPs and putative target genes in AS patient and healthy control immune cells. Nine interactions were observed between the targeted AS lead SNP and cognate gene promoters. Tiled-C was then used to compare the interaction strength at 21 loci from healthy control $CD4^+$ and $CD8^+$ T cells in response to immunogenic stimulation, and from AS patient cells obtained before and after initiation of anti-TNFα therapy. T cell activation was associated with strengthened chromatin interactions between AS-associated genetic variants and 11 genes in $CD4^+$ T cells and 5 genes in $CD8^+$ T cells across 6 AS-associated GWAS loci; with strongest evidence seen for the genes *BACH2*, *PTGER4*, and *DEXI*.

Collectively, the work presented in this thesis has advanced our understanding of how AS-associated genetic variants contribute to disease pathogenesis. The biological implications of these findings highlight potential novel therapeutic targets for AS treatment outside of conventional anti-TNFα or anti-IL17 biologics.

ACKNOWLEDGEMENTS

"At least for a while,' said Elrond. 'The road must be trod, but it will be very hard. And neither strength nor wisdom will carry us far upon it. This quest may be attempted by the weak with as much hope as the strong. Yet such is oft the course of deeds that move the wheels of the world: small hands do them because they must, while the eyes of the great are elsewhere."

- J.R.R. Tolkien, Fellowship of the Ring (1954)

Firstly, I would like to extend my sincerest appreciation to my four supervisors Prof. Julian Knight, Dr. Matteo Vecellio, Dr. Carla Cohen, and Prof. Bryan Paul Wordsworth for the immense amount of support they have shown me over the course of my DPhil studies.

Julian has provided one of the most stable and encouraging laboratory environments I have ever experienced, and has lent me great support to the many ideas I have keenly sought to explore. His additional financial support has been nothing short of crucial to the production of this thesis, and I feel I will never truly be able to repay the kindnesses he has shown me.

Matteo and Carla have been my rocks, meeting with me every week to make sure I am on track both academically and personally meeting my academic goals. Their conversations were a very warm comfort over the course of the Covid-19 pandemic, where many of my experiments took the backseat during the numerous lockdowns.

I cannot thank Paul enough for all of the in-depth clinical chats and his encyclopaedic knowledge of ankylosing spondylitis. Both very much allowed me to place my genomic studies within the right context of the disease. I must also show my gratitude for his unwavering support towards my experiments, never having hesitated to aid me in the blood draw of both patients and healthy controls on request.

Dr. Andrew Brown has been a fast friend over my DPhil studies. I have learnt more computational biology from Andy than I ever could hope for over an undergraduate bioinformatics course. His patience in regards to my constant pestering has been saint-like, and his dedication to science has proven a constant source of inspiration for my own experiments.

I must also thank Dr. Olga Mielczarek and Dr. Gabriele Migliorini for all they had taught be in regards to chromosome conformation capture, which has quickly become my favourite genomic assay to perform.

Every member in the Knight lab has contributed to making the DPhil experience a source of enjoyment. Many of whom have been directly involved within the AS project, with their contributions highlighted in the declarations and throughout this thesis.

I would like to also acknowledge the support of my friends outside of Oxford over the course of the last few years. Charles Day, Christian Lewes, Conor Howard, Corey Dickson, Jamie Chadwick, James Ashford, Matthew Turner, Maxine Ransom, Myles Bijsterbosch, and Oliver Farrell have kept me grounded and sane in times I thought I might have lost it all.

Last, but certainly not least, I would not have succeeded in the completion of my DPhil studies without the undying love and support shown by my family. In particular, my parents Sean and Karen Davidson, and my grandparents, Douglas and Mary Duncan, are my true sources of strength. If not for them, I would not have achieved half of what I previously thought impossible.

The work carried out in this thesis was generously funded by Versus Arthritis through their Prize PhD Studentship 2019 (Grant reference: 22198).

DECLARATIONS

In signing this declaration:

[1] I hereby declare that the work embodied in this doctoral thesis – *"*Functional characterisation of the three-dimensional epigenetic landscape in ankylosing spondylitis-linked genome-wide association study loci*"* – is the result of my own independent research and investigation, and that any findings and/or conclusions that have been drawn from the presented data are also of my own, unless otherwise stated.

[2] I confirm that the main text of this thesis consists of 59,912 words and is under the approved extended 60,000-word limit exclusive of those words found in figures, tables, legends, appendices, footnotes, and the bibliography.

[3] I further declare that this work has not been previously submitted for any prior degree, and is not concurrently being submitted in candidature for further qualification, at either the University of Oxford or any other higher education institution.

[4] The data presented in this thesis has contributed more widely to the Knight group's investigation into the genetic basis of ankylosing spondylitis. As such, this thesis has at times analysed and relied upon previously generated data and/or aid from collaborators in the form of scripts or labour at several points throughout. This includes the following:

Bulk RNA/ATAC-seq: RNA and ATAC-seq samples from recruitment cohorts one through five were processed at various points in-house by Dr. Andrew Brown, Dr. Anna Sanniti, Dr. Alicia Lledó Lara, Dr. Julie Osgood, Dr. Olga Mielczarek, and Ms. Alice Martin of the Centre for Human Genetics (CHG), University of Oxford. RNA/ATAC-seq samples from cohort six were isolated and prepared by myself, and sent to Azenta© Life Sciences and Novogene® for library preparation and/or sequencing. Modified scripts used for limma normalisation and DESeq2 differential analysis of RNA/ATAC counts were initially provided by Dr. Andrew Brown and Dr. Gabriele Migliorini whilst the latter was a research scientist at the CHG; now of the Institute for Cancer Research at the University of Cambridge.

Bulk ChIP-seq: Material for H3K27me3 ChIPmentation-seq analyses were collected by Dr. Andrew Brown, Dr. Alicia Lledó Lara, and Dr. Olga Mielczarek. I processed these samples with the aid of Dr. Andrew Brown and Dr. Conor Short of the Nuffield Department of Medicine, University of Oxford. Previous H3K4me3 and H3K27ac ChIPmentation-seq experiments were performed by the researchers above. The scripts for differential analysis of histone peaks using MACS2 were modified to accept H3K27me3 broad-peak data from those provided by Dr. Andrew Brown and Dr. Gabriele Migliorini.

Variant enrichment analysis: GARFIELD and Fisher permutation scripts used for functional enrichment analyses of the 16-emission state ChromHMM in chapter three were modified from those produced by Prof. Felicie Constantino whilst a visiting fellow at the CHG; now of the Ambroise Paré University Hospital.

Capture-C: Capture-C samples were processed either by (healthy volunteer and AS patient $CD4^+$ and $CD8^+$ T cells; AS patient $CD14^+$ monocytes) or under the direct tutelage of (healthy volunteer $CD14^+$ monocytes) Dr. Olga Mielczarek. Design of oligonucleotide baits and analysis of bait interaction profiles with peaky was performed by Dr. Gabriele Migliorini. Oligonucleotide baits were synthesised by Sigma-Aldrich®.

Tiled-C: Tiled-C oligonucleotide tiled bait sets were designed with guidance from Dr. Olga Mielczarek and Dr. Gabriele Migliorini. Tiled-C baits were synthesised by Twist Biosciences®. I performed analysis of subsequent interaction matrices using CapCruncher with aid from Dr. Andrew Brown and Dr. Alastair Smith of the Weatherhall Institute for Molecular Medicine, University of Oxford.

[5] I additionally certify that, to the best of my knowledge, this dissertation has offended neither the proprietary nor copyright laws of any cited parties, and that any findings, ideas, and/or techniques presented in their research have duly been acknowledged according to standardised referencing practices.

Submitting doctoral student:

Connor Ian Davidson



Submitted 12th of October 2024

Supervisor names and contact details:

Prof. Julian Knight (julian@well.ox.ac.uk) – CHG

Dr. Matteo Vecellio (matteo.vecellio@ndorms.ox.ac.uk) – NDORMS

Dr. Carla Cohen (carla.cohen@ndorms.ox.ac.uk) – NDORMS

Prof. Bryan Paul Wordsworth (paul.wordsworth@ndorms.ox.ac.uk) – NDORMS

ASSOCIATED PUBLICATIONS

[1] Vecellio M., Rodolfi S., **Davidson C**., Simone D. (2024) "Role of regulatory T cells in pathogenesis and therapeutics of ankylosing spondylitis." *Regulatory T cells and Autoimmune Diseases: Chapter 10*. doi: 10.1016/B978-0-443-13947-5.00021-X.

[2] Rodolfi S., **Davidson C.**, Vecellio M. (2024) "Regulatory T cells in spondyloarthropathies: genetic evidence, functional role, and therapeutic possibilities." *Front. Immunol*. 14(1): e. 1303640. doi: 10.3389/fimmu.2023.1303640.

[3] Brown A. C., Cohen C. J., Mielczarek O., Migliorini G., Costantino F., Allcock A., **Davidson C.**, Elliott K. S., Fang H., Lledó Lara A., Martin A. C., Osgood J. A., Sanniti A., Scozzafava G., Vecellio M., Zhang P., Black M. H., Li S., Truong D., Molineros J., Howe T., Wordsworth B. P., Bowness P., Knight J. C. (2023) "Comprehensive epigenomic profiling reveals the extent of disease-specific chromatin states and informs target discovery in ankylosing spondylitis." *Cell Genom*. 3(6): e. 100306. doi: 10.1016/j.xgen.2023.100306.

[4] Davidson C., Wordsworth B. P., Cohen C. J., Knight J. C., Vecellio M. (2023) "Chromosome conformation capture approaches to investigate 3D genome architecture in Ankylosing Spondylitis." *Front. Genet*. 14(1): e. 1129207. doi: 10.3389/fgene.2023.1129207.

[5] Cohen C. J., **Davidson C.**, Selmi C., Bowness P., Knight J. C., Wordsworth B. P., Vecellio M. (2022) "Disruption of c-MYC Binding and Chromosomal Looping Involving Genetic Variants Associated with Ankylosing Spondylitis Upstream of the *RUNX3* Promoter." *Front. Genet*. 12(1): e. 741867. doi: 10.3389/fgene.2021.741867.

[6] Wordsworth B. P., Cohen C. J., **Davidson C.**, Vecellio M. (2021) "Perspectives on the Genetic Associations of Ankylosing Spondylitis." *Front. Immunol*. 12(1): e. 603726. doi: 10.3389/fimmu.2021.603726.

[7] Vecellio M., Hake V. X., **Davidson C.**, Carena M. C., Wordsworth B. P., Selmi C. (2021) "The IL-17/IL-23 Axis and Its Genetic Contribution to Psoriatic Arthritis." *Front. Immunol*. 11(1): e. 596086. doi: 10.3389/fimmu.2020.596086.

TABLE OF CONTENTS

SECTION **PAGE**

Abstract …… I

Acknowledgements ……..… II

Declarations ……...…… III

Associated publications …….……...….... V

Table of contents …….……... VI

List of figures …….…….. XV

List of tables ……. XXIII

Abbreviations ……...…… XXIV

CHAPTER 1: GENERAL INTRODUCTION

1.1. Thesis objectives ……... 1

1.2. Overview of ankylosing spondylitis …….. 2

1.2.1. Introduction to ankylosing spondylitis ……... 2

1.2.2. Epidemiology of ankylosing spondylitis ……..... 4

1.2.3. Diagnosis of ankylosing spondylitis ……... 5

1.2.4. Contemporary and future biomarkers for ankylosing spondylitis …… 10

1.2.5. Therapeutic strategies for treating ankylosing spondylitis …….……..… 11

1.3. Genetics of ankylosing spondylitis …….. 14

1.3.1. Heritability of ankylosing spondylitis …….. 14

1.3.2. Genetic variation and protein expression ……. 15

1.3.3. Genome-wide association studies ……. 18

1.3.4. Linkage disequilibrium and statistical fine-mapping ……..….. 19

1.3.5. GWAS in ankylosing spondylitis ……. 21

1.4. Pathophysiology of ankylosing spondylitis ……. 26

1.4.1. MHC class I and class II molecules and T cells in AS …… 26

1.4.2. The MHC class I molecule HLA-B27 and ankylosing spondylitis ……. 27

1.4.2.1. The arthritogenic peptide hypothesis ……. 28

1.4.2.2. The unfolded protein response hypothesis …… 30

1.4.2.3. The homodimer hypothesis ……... 31

TABLE OF CONTENTS

1.4.2.4. HLA-B27 homodimers, the UPR, and Th17 cells 31
1.4.3. Non-HLA-B27 MHC contributions to AS 34
1.4.4. The contribution of monocytes and macrophages to AS 34
1.4.5. The contribution of B cells to AS 36
1.4.6. Other implicated cell types in AS 37
1.4.7. The gut microbiome and AS 38
1.4.8. Non-MHC AS-associated genes and AS pathogenesis 39
1.4.8.1. *ERAP1* and *ERAP2*: MHC class I processing 39
1.4.8.2. *IL-23R*: Modifying the IL-17/IL-23 signalling axis 40
1.4.8.3. *RUNX3*: Immune cell development 41
1.4.8.4. *TNFRSF1A*: TNFα signalling 43
1.5. Gene regulation, genetic variants, and disease 43
1.5.1. Non-coding genetic variants and eQTLs 44
1.5.2. *Cis*-regulatory genetic variants in the non-coding genome 45
1.5.3. The structure of *cis*-regulatory region 45
1.5.3.1. Associated histone marks for promoter regions 48
1.5.3.2. Associated histone marks for enhancer regions 49
1.5.3.3. lncRNA is associated with chromatin remodelling complexes . 50
1.5.3.4. Associated marks for silencer regions 51
1.5.3.5. Bivalency of histone marks at promoter regions 52
1.5.3.6. The role of insulators in maintaining looping interactions 54
1.6. Functional genomics and integrative analyses 55
1.6.1. Experimental techniques for the detection of open chromatin regions 55
1.6.2. Experimental techniques for the detection of histone marks 58
1.6.3. Experimental techniques to assay differential gene expression 60
1.6.4. Experimental techniques to define chromatin interactions 61
1.7. Specific aims and objectives 64
1.7.1. Aims: Chapter 3 64
1.7.2. Aims: Chapter 4 64
1.7.3. Aims: Chapter 5 65

CHAPTER 2: METHODS & MATERIALS

2.1. Study recruitment and ethical approval 66

2.1.1. Patient and healthy control sampling 66

2.1.2. Eligibility criteria and recruitment demographics 67

2.2. Sample processing and cell culture 69

2.2.1. Peripheral blood mononuclear cell purification 69

2.2.2. Magnetic-activated cell sorting 72

2.2.3. PBMC culture and cell preparation 73

2.2.3.1. $CD4^+$ and $CD8^+$ T cells 73

2.2.3.2. $CD14^+$ monocytes and macrophages 73

2.3. Capture-C and Tiled-C 74

2.3.1. Fixation of cell for Capture-C and Tiled-C 74

2.3.2. Digestion of 3C libraries with DpnII 75

2.3.3. 3C library ligation and de-crosslinking 76

2.3.4. 3C library purification 76

2.3.5. Sonication of 3C libraries 79

2.3.6. 3C library indexing 80

2.3.7. Primary capture of material for Capture-C libraries 82

2.3.8. Washing of captured material for Capture-C 89

2.3.9. Amplification and second capture of Capture-C libraries 90

2.3.10. Test qPCR and final amplification of Capture-C libraries 91

2.3.11. Hybridisation and capture baits for Tiled-C 92

2.3.12. Washing of captured Tiled-C libraries 94

2.3.13. Amplification and sequencing of Tiled-C libraries 95

2.4. ChIPmentation-seq 96

2.4.1. Preparing cells for ChIPmentation-seq 96

2.4.2. Reverse cross-linking and tagmentation of ChIPmentation inputs 97

2.4.3. Antibody binding and immunoprecipitation 98

2.4.4. Washing of immunoprecipitated bead-DNA complexes 99

2.4.5. Indexing and sequencing of ChIPmentation samples 100

2.5. ATAC-seq 102

2.5.1. Cell preparation for ATAC-seq and transposition 102

2.5.2. Indexing and sequencing of ATAC-seq material 102

2.6. qRT-PCR and RNA-seq 103

2.6.1. Purification and reverse transcription of total RNA for qPCR 103

2.6.2. qRT-PCR reaction and cycling conditions 104

2.6.3. Processing of cell for RNA-seq and sample sequencing 105

2.7. Computational and statistical analyses 105

2.7.1. Analysis of Capture-C data 105

2.7.2. Analysis of Tiled-C data 107

2.7.3. Analysis of ChIPmentation-seq data 108

2.7.4. Analysis of ATAC-seq data 109

2.7.5. Analysis of RNA-seq data 110

2.7.6. ChromHMM 111

2.7.7. Enrichment of genetic variants in ChromHMM emission states 113

2.7.8. Script and data availability 114

CHAPTER 3: INTEGRATIVE MODELLING OF THE EPIGENETIC LANDSCAPE IN AS-ASSOCIATED GENETIC LOCI SHOWS ENRICHMENT OF GWAS SNPs FOR *CIS*-REGULATORY ELEMENTS

3.1. Chapter introduction 115

3.1.1. Identification of disease-specific gene expression networks 115

3.1.2. Nitric oxide synthase 2 as an AS-associated risk locus 117

3.1.3. Using ChromHMM to identify cell-type specific genomic annotations . 118

3.1.4. Enrichment of disease-associated genetic variants at regulatory genomic regions 121

3.1.5. Chapter aims 122

3.2. Functional characterisation of the *NOS2* locus 123

3.2.1. ChromHMM analysis of the *NOS2* locus in T cells and monocytes 123

3.2.2. NOS2 is epigenetically silenced in monocyte-derived macrophages 125

3.3. Enrichment of AS-associated SNPs in regulatory regions 130

3.3.1. An updated epigenomic map in AS incorporating Polycomb repression-associated post-translational histone modifications 130

3.3.2. Global and state-specific differences in chromatin conformation between AS patients and healthy volunteers 137

3.3.3. $CD14^+$ monocytes and $CD19^+$ B cells show quantitative differences in ChromHMM emission states in AS patients compared to healthy volunteers 140

3.3.4. Permutation analysis reveals significant differences in the identity of quiescent and Polycomb-repressed emission states within AS compared to healthy volunteers 149

3.3.5. Analysis of transition profiles of significantly permuted ChromHMM emission states reveals bias in emission state transformations 150

3.3.6. Enrichment of AS-associated GWAS variants in *cis*-regulatory and Polycomb-repressed regions 157

3.4. Chapter Discussion 161

3.4.1. The NOS2 association signal is silenced by H3K27me3 in T cells, B cells, and monocytes and likely does not contribute towards AS pathogenesis in these cells 161

3.4.2. Genome-wide H3K27me3 signals were not found to be differential in AS patients and healthy volunteers and suggested consistent Polycomb-associated gene repression between the two states 163

3.4.3. Incorporation of H3K27me3 into ChromHMM identifies regions of Polycomb repression that can help identify inactive genes in T cells, B cells, and monocytes 165

3.4.4. AS-associated GWAS SNPs are enriched for cis-regulatory ChromHMM chromatin states that may differ between AS patients and healthy volunteers 167

3.4.5. AS-associated SNPs enrich for cis-regulatory elements but only MHC-associated AS SNPs enrich for Polycomb-repressed chromatin states .. 170

3.4.6. Concluding remarks 172

TABLE OF CONTENTS

CHAPTER 4: AS-ASSOCIATED GWAS LEAD SNPs SHOW *CIS*-REGULATORY LOOPING TO INFLAMMATION-RELATED GENE PROMOTERS WITH CAPTURE-C

4.1. Chapter introduction ……..…… 173

4.1.1. Chromosome conformation capture can link SNPs in regulatory regions with cognate genes …… 173

4.1.2. Capture-C has previously been used successfully in inflammatory conditions to show that differential chromatin interactions may be involved in disease pathogenesis ……. 174

4.1.3. Selecting the appropriate chromosome conformation capture assay for the investigation of AS-associated lead SNP interactions …… 175

4.1.4. Chapter aims ……... 177

4.2. Experimental design and analysis of known interactions at control loci for method validation …….……. 178

4.2.1. Experimental design ……... 178

4.2.2. Interactions at control loci provide confidence in data quality …… 178

4.3. SNP-promoter interactions at AS-associated loci ……..… 184

4.3.1. Summary of interactions and intersection of Capture-C and epigenetic data …… 184

4.3.2. The chr17q23.3 locus shows clear reciprocal looping interactions between the promoter of ERN1 and rs196941 in $CD4^+$, $CD8^+$, and $CD14^+$ cells ……... 186

4.3.3. Cell-type specific interactions are found between two independent AS-associated SNPs and the promoter of *PTGER4* …….…… 190

4.3.4. The AS-associated SNP rs2836883 shows context-specific looping onto the ETS2 promoter in CD14+ monocytes but not in $CD4^+$ or $CD8^+$ T cells . 194

4.4. Complex and context-specific interaction profiles …… 196

4.4.1. The interactions at the chr1p36.11 locus shows super-enhancer-like structure between the promoter of *RUNX3* and distal regulatory elements in T cells and monocytes …….. 198

4.4.2. The lead SNP *rs12615545* in the chr2q31.3 locus loops onto the promoter of *ITGA4* instead of the most proximal gene *UBE2E3* in $CD4^+$ and $CD8^+$ T cells and $CD14^+$ monocytes ……... 200

4.4.3. The promoter of ETS1 loops onto an active enhancer element outside of the lead-SNP-containing LD block in $CD4^+$ and $CD8^+$ T cells but not in monocytes …….. 203

4.4.4. The 7p21.1 locus shows a lack of interaction between the lead SNP and any gene promoters in the region …….. 206

4.5. Chapter discussion ……. 208

4.5.1. A subset of the AS-associated GWAS risk loci show strong looping interactions between ChromHMM-predicted *cis*-regulatory regions … 208

4.5.2. The majority of loci showed no interactions for either the promoter viewpoint or the lead SNP viewpoint …… 211

4.5.3. Other cell types and certain T cell and monocyte subtypes may show different interaction profiles between AS-associated lead SNPs and their target genes ……... 214

4.5.4. The interaction of cis-regulatory elements with the *RUNX3* promoter were complex and resembled super-enhancer-like structure …….. 215

4.5.5. Some regions may show differential interaction between AS patients and healthy volunteers …….. 216

4.5.6. Tiled-C can be used to investigate the interactions between differential epigenetic marks at each locus not captured in the Capture-C experiments ……... 218

4.5.7. Concluding remarks …….. 219

CHAPTER 5: APPLICATION OF TILED-C IN $CD4^+$ AND $CD8^+$ T CELLS TO INTERROGATE CHROMATIN CONTACTS IN AS-ASSOCIATED GWAS LOCI

5.1. Chapter introduction …….. 220

5.1.1. Tiled-C can be used to comprehensively map chromosome looping interactions across TADs ……. 220

5.1.2. Experimental design of Tiled-C experiments ……. 221

5.1.2.1. Prioritisation of genetic loci ……. 221

5.1.2.2. Choice of cell types and stimulation conditions …….. 222

5.1.2.3. Choice of patient samples for Tiled-C experiments …… 223

5.1.3. Defining areas of chromatin interaction in Tiled-C data …… 224

5.1.4. Intersection of Tiled-C data with other genomic modalities can facilitate

interpretation of disease-associated interactions 226

5.1.5 Chapter aims 228

5.2. Tiled-C at control loci identifies known interactions 230

5.2.1. Assessment of data quality at the control locus *MYC* 230

5.2.2. Tiled-C data quality varied across cell types 231

5.2.3.The regulatory landscape of *MYC* is altered by stimulation of $CD4^+$ and $CD8^+$ T cells with CD3/CD28 T cell activation beads 234

5.2.4. AS patients pre and post-treatment with adalimumab lack the CD3/CD28-associated enhancer contacts observed in stimulated healthy volunteer T cells 238

5.3. Genome-wide epigenetic changes in untreated, treated, and stimulated cells 243

5.3.1. The gene expression and chromatin accessibility profiles of AS patient $CD4^+$ and $CD8^+$ T cells pre and post-adalimumab treatment more closely resemble unstimulated healthy volunteer cells over those stimulated with CD3/CD28 beads 243

5.3.2. AS-assocaited SNPs are found in regulatory elements that respond to T cell activation and stimulation with CD3/CD28 beads 247

5.4. Example Tiled-C regions 249

5.4.1. Upregulation of *PTGER4* in $CD4^+$ T cells on CD3/CD28 activation requires the coordination of multiple stimulation-specific enhancers 249

5.4.2. *rs17765610* forms CD3/CD28-stimulation-associated contacts with the *BACH2* promoter that are absent in AS patient cells pre and post-3 month treatment with adalimumab 254

5.4.3. An active enhancer element in close proximity to rs367569 interacts more strongly with the promoter of *DEXI* on CD3/CD28 stimulation 263

5.5. Chapter discussion 267

5.5.1. Generation of comprehensive looping interaction maps at AS-associated GWAS risk loci 269

5.5.1.1. Data quality 269

5.5.1.2 Comparison of conditions 271

5.5.2. Considerations for AS pathogenesis 273

5.5.2.1. PTGER4 expression may be mediated by the SNPs *rs1992661* and *rs9283753* in activated T cells 273

5.5.2.2. BACH2 expression may be modulated by rs17765610 in activated T cells 275

5.5.2.3. Expression of DEXI may be controlled in activated T cells by differential AS-associated cis-regulatory elements 277

5.5.3. Study limitations and future work 280

5.5.3.1. Cellular epigenetics may differ between AS flares and patients in remission and at the site of inflammation 280

5.5.3.2. Machine-learning for the identification of differential interactions in AS 282

5.5.4. Concluding remarks 283

CHAPTER 6: GENERAL DISCUSSION

6.1. Summary of findings 285

6.2. Modelling the epigenetic landscape of AS 287

6.2.1. *NOS2* was not a tractable locus for further study 287

6.2.2. AS-associated genetic variants enriched primarily for cis-regulatory elements and Polycomb-repressed DNA 288

6.3. T cell activation and AS-associated genetic variants 291

6.3.1. Capture-C was used to link AS-associated lead SNPs to cognate genes . 291

6.3.2. Tiled-C was used to map the complete interaction profiles of AS-associated TADs in multiple contexts 292

6.3.3. T cell activation was associated with increased chromatin interactions at multiple inflammation-associated loci in $CD4^+$ and $CD8^+$ T cells 294

6.4. Limitations and future work 295

6.4.1. Sample numbers and extended recruitment 295

6.4.2. Additional cell types and tissue sampling 297

6.4.3. Incorporating more epigenetic datasets and stimulation conditions into ChromHMM 297

6.4.4. Functional follow-up of identified variants 299

6.4.5. Pathway to new therapies 300

6.5. Concluding remarks 301

Appendix 303

Bibliography 352

LIST OF FIGURES

CHAPTER 1: GENERAL INTRODUCTION **PAGE**

Fig. 1.1. ASAS classification criteria for the diagnosis of axial spondyloarthropathy ... 7

Fig. 1.2. HLA-B27 contributes to ankylosing spondylitis by upregulating the IL-23/IL-17 signalling axis in $CD4^+$ and $CD8^+$ T cells through three models …….…. 33

Fig. 1.3. Schematic overview of regulatory interaction between enhancer and promoter elements and their histone post-translational modification and associated proteins ……...…… 46

Fig. 1.4. Integrative modelling of epigenetic datasets in AS to determine the effects of disease-associated SNPs on *cis*-regulatory function and gene regulation in PBMCs used by the Knight group ……... 56

CHAPTER 2: METHODS & MATERIALS

Fig. 2.1. Resolution of sacroiliac inflammation within an AS patient following a three-month course of anti-TNFα therapy …… 71

Fig. 2.2. Quality control of Capture-C libraries using undigested and digested controls 74

Fig. 2.3. Quality control of 3C libraries through the processing stages of sonication, pre-indexing, and post-indexing ……. 78

CHAPTER 3: INTEGRATIVE MODELLING OF THE EPIGENETIC LANDSCAPE IN AS-ASSOCIATED GENETIC LOCI SHOWS ENRICHMENT OF GWAS SNPs FOR *CIS*-REGULATORY ELEMENTS

Fig. 3.1. ChromHMM emission states from the Knight group …….. 120

Fig. 3.2. The epigenetic landscape surrounding the NOS2 AS-associated GWAS risk locus in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes ……... 124

Fig. 3.3. Linkage disequilibrium at the *NOS2* locus ……. 126

Fig. 3.4. No induction of NOS2 is seen upon stimulation with either MDP or GMDP in $CD14^+$ monocyte-derived macrophages ……. 128

Fig. 3.5. The expression of NOS2 is repressed within $CD4^+$ and $CD8^+$ T cells and $CD14^+$ monocytes through an intragenic H3K27me3-rich silencing element …… 129

Fig. 3.6. Differential genes linked to H3K27me3 peaks within all investigated PBMC cell types ……... 131

Fig. 3.7. Correlation of emission parameters reveals the 16-emission state ChromHMM to be the most appropriate model size upon incorporation of H3K27me3 ChIP-seq data …….. 133

Fig.3.8. Identifying the biology behind each emission state within the 16-state

ChromHMM ……...…….. 134

Fig. 3.9. The 16-emission state ChromHMM reveals a number of repression-associated regulatory regions that were hidden in the 14-emission state ChromHMM upon incorporation of H3K27me3 data …….. 136

Fig. 3.10. MCA analysis of ChromHMM emission state bins in the 16-state model shows cell type-specific separation ……. 138

Fig. 3.11. MCA analysis of all ChromHMM bins in the 16-state model for each individual cell type …… 139

Fig. 3.12. MCA analysis of each individual ChromHMM emission state within $CD4^+$ T cells for the 16-emission state model …….. 141

Fig. 3.13. MCA analysis of each individual ChromHMM emission state within $CD8^+$ T cells for the 16-emission state model …….. 142

Fig. 3.14. MCA analysis of each individual ChromHMM emission state within $CD14^+$ monocytes for the 16-emission state model ……..…….. 143

Fig. 3.15. MCA analysis of each individual ChromHMM emission state within $CD19^+$ B cells for the 16-emission state model …….. 144

Fig. 3.16. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for $CD4^+$ T cells ……..….. 145

Fig. 3.17. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for $CD8^+$ T cells ……..…….. 146

Fig. 3.18. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for $CD14^+$ monocytes …….…….. 147

Fig. 3.19. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for $CD19^+$ B cells …….…….. 148

Fig. 3.20. Permutation analysis of emission states in the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in $CD4^+$ T cells ……. 151

Fig. 3.21. Permutation analysis of emission states in the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in $CD8^+$ T cells ……. 152

Fig. 3.22. Permutation analysis of emission states in the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in $CD14^+$ monocytes ...……. 153

Fig. 3.23. Permutation analysis of emission states in the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in $CD19^+$ B cells ……..……. 154

Fig. 3.24. Transition analysis of significant emission state permutations shows cell-type specific, yet highly similar, transition patterns to other ChromHMM emission states within healthy volunteers ……... 155

Fig. 3.25. Transition analysis of significant emission state permutations shows cell-type specific, yet highly similar, transition patterns to other ChromHMM emission states within AS patients …… 156

Fig. 3.26. GARFIELD enrichment plots of Cortes *et al.* (2013) AS-associated immunochip genetic variants within each of the 16 emission states including MHC-associated SNPs on chromosome 6 …… 158

Fig. 3.27. GARFIELD enrichment plots of Cortes *et al.* (2013) AS-associated immunochip genetic variants within each of the 16 emission states excluding MHC-associated SNPs on chromosome 6 …… 160

Fig. 3.28. Summarisation of MCA, permutation, and GARFIELD enrichment analyses for all 16 detected ChromHMM emission states with and without the MHC region …….. 168

CHAPTER 4: AS-ASSOCIATED GWAS LEAD SNPs SHOW *CIS*-REGULATORY LOOPING TO INFLAMMATION-RELATED GENE PROMOTERS WITH CAPTURE-C

Fig. 4.1. Experimental design for reciprocal Capture-C experiments ……... 179

Fig. 4.2. A distal active enhancer element that is repressed in $CD4^+$ T cells and $CD8^+$ T cells loops onto the *LMO1* promoter in $CD14^+$ monocytes …….…….. 181

Fig. 4.3. A distal active enhancer element loops onto the *MYC* promoter in $CD4^+$ and $CD8^+$ T cells that is Polycomb-repressed in $CD14^+$ monocytes …….. 182

Fig. 4.4. A distal bivalent promoter region loops onto the end of *INPP4B* in $CD4^+$ and $CD8^+$ T cells but not in $CD14^+$ monocytes ……. 183

Fig. 4.5. Enhancer-gene interactions between the promoter of *ERN1* and *rs196941* at the chr17q23.3 locus …… 188

Fig. 4.6. Differential epigenetic marks at the chr17q23.3 locus show the lead SNP-*ERN1* promoter interaction is more active in $CD4^+$ T cells and less accessible in $CD8^+$ T cells in AS patients …….. 189

Fig. 4.7. Enhancer-gene interactions between the promoter of *PTGER4*, *rs1992661*, *rs9283753* and *rs12186979* at the chr5p13.1 locus ……. 191

Fig. 4.8. Differential epigenetic marks at the chr5p13.1 locus show the cis-regulatory elements containing lead SNP *rs1992661* is more active in $CD4^+$ T cells and $CD14^+$ monocytes and the functional SNP *rs9283753* is more accessible in $CD8^+$ T cells ……. 193

Fig. 4.9. Enhancer-gene interactions between the promoter of *ETS2* and *rs2836883* at the

chr21q22.2 locus are specific to CD14$^+$ monocytes 195

Fig. 4.10. Differential epigenetic marks at the chr21q22.2 locus show the *ETS2*-interacting enhancers are less accessible in AS patient CD14$^+$ monocytes 197

Fig. 4.11. Complex super-enhancer-like structures form between multiple *cis*-regulatory elements and the *RUNX3* promoter at the chr1p36.11 locus 199

Fig. 4.12. Differential epigenetic marks at the chr1p36.11 shows increased enhancer activity at a distal *RUNX3*-interacting enhancer across all cell types in AS patients 201

Fig. 4.13. The lead SNP for the chr2q31.3 locus loops onto the *ITGA4* promoter in T cells and monocytes instead of the baited *UBE2E3* promoter 202

Fig. 4.14. The epigenetic marks underlying the interaction between the *ITGA4* promoter and the *rs12615545*-containing enhancer element are similar between AS patients and healthy volunteers 204

Fig. 4.15. The promoter of *ETS1* interacts with the lead SNP *rs7933433* of the chr11q24.3 locus only in the CD4$^+$ and CD8$^+$ T cells of AS patients 205

Fig. 4.16. Differential eRNA and H3K27ac marks at the *rs7933433*-containing regulatory element suggests greater enhancer activity at the chr11q24.3 locus in AS patients 207

Fig. 4.17. No reciprocal interactions are observed between the *AHR* promoter and the AS lead SNP *rs1525735* at the chr7p21.1 locus in any investigated cell type ... 209

CHAPTER 5: APPLICATION OF TILED-C IN CD4$^+$ AND CD8$^+$ T CELLS TO INTERROGATE CHROMATIN CONTACTS IN AS-ASSOCIATED GWAS LOCI

Fig. 5.1. Tiled-C analysis strategy towards identifying AS or stimulation-associated differential chromatin contacts with ICE 227

Fig. 5.2. Tiled-C matrices for the *MYC* control locus in CD4$^+$ T cells, CD8$^+$ T cells, and CD14$^+$ monocytes 232

Fig. 5.3. *Cis*-to-*trans* read ratios in the Tiled-C samples 233

Fig. 5.4. Intersection of Tiled-C and epigenetic data at the MYC locus informs the position of CD3/CD28 stimulation-induced chromosome looping 236

Fig. 5.5. Expression of *MYC* is significantly upregulated in CD4$^+$ and CD8$^+$ T cells on CD3/CD28 stimulation but not in AS patients before or after treatment with adalimumab 238

Fig. 5.6. Tiled-C interaction profiles for CD4$^+$ T cells isolated from AS patients pre and post-treatment with adalimumab are similar and lack the enhancer contacts seen in CD3/CD28-stimulated healthy volunteer cells 241

Fig. 5.7. The majority of differentially expressed genes in the 21 AS-associated Tiled-C

regions are associated with CD3/CD28-stimulated T cells instead of disease or treatment state ……. 244

Fig. 5.8. The majority of differential open chromatin regions in the 21 AS-associated Tiled-C regions are associated with CD3/CD28-stimulated T cells instead of disease or treatment state …….. 246

Fig. 5.9. Chromatin looping between stimulation-responsive genetic elements in $CD4^+$ T cells are observed in both unstimulated and CD3/CD28-stimulated healthy volunteer $CD4^+$ T cells but not with *rs199266* ……. 252

Fig. 5.10. Expression of *PTGER4* is significantly upregulated in $CD4^+$ and $CD8^+$ T cells on CD3/CD28 stimulation but not in AS patients before or after treatment with adalimumab ……. 255

Fig. 5.11. AS patients pre and post-adalimumab treatment do not show looping interactions with *rs1992661, rs928753,* or *rs12186979* …… 257

Fig. 5.12. Stimulation of $CD4^+$ T cells with CD3/CD28 T cell activation beads is associated with the formation of stronger chromatin contacts between the promoter of *BACH2*, *rs17765610*, and an internal promoter element …… 259

Fig. 5.13. Expression of *BACH2* is not differential on stimulation with CD3/CD28 activation beads, in AS or on treatment with adalimumab …… 260

Fig. 5.14. Interaction profiles at the *BACH2* locus in untreated and adalimumab-treated AS patient cells are similar to unstimulated healthy volunteers …….. 263

Fig. 5.15. Increased interaction between the shared promoter of *DEXI* and *CLEC16A* and active enhancer elements in $CD4^+$ T cells on CD3/CD28 stimulation ……. 265

Fig. 5.16. Expression of *DEXI* is differential on stimulation with CD3/CD28 activation beads but not in AS or on treatment with adalimumab …… 266

Fig. 5.17. Interaction profiles at the *CIITA-DEXI-CLEC16A-SOCS1* locus in untreated and adalimumab-treated AS patient cells are similar to unstimulated healthy volunteers ……... 269

Fig. 5.18. Linkage disequilibrium between autoimmune SNPs in the *CIITA-DEXI-CLEC16A-SOCS1* Tiled-C region …….. 279

APPENDIX

S.1. Enhancer-gene interactions between the promoter of *IL7R* and *rs991570* at the chr5p13.2 locus …… 304

S.2. Differential epigenetic marks at the chr5p13.2 locus …… 305

S.3. Enhancer-gene interactions between the promoter of *IL18RAP* and *rs4851529* at the chr2q11.2 locus ……. 306

S.4. Differential epigenetic marks at the chr2q11.2 locus ……. 307

Fig. S.5. Enhancer-gene interactions between the promoter of *BACH2*, *rs17765610*, and *rs639575* at the chr6q15 locus ……..…. 308

Fig. S.6. Differential epigenetic marks at the chr6q15 locus ……..…….. 309

Fig. S.7. Genome-wide analysis of differential genes in unstimulated healthy volunteer, CD3/CD28-stimulated healthy volunteer, untreated AS patient, and adalimumab-treated AS patient $CD4^+$ and $CD8^+$ T cells …….. 310

Fig. S.8. Genome-wide analysis of differential open chromatin regions in unstimulated healthy volunteer, CD3/CD28-stimulated healthy volunteer, untreated AS patient, and adalimumab-treated AS patient $CD4^+$ and $CD8^+$ T cells …….. 311

Fig. S.9 . Tiled-C interactions and the chromatin landscape of the *PTGER4* locus in $CD8^+$ T cells …… 313

Fig. S.10. Tiled-C interactions and the chromatin landscape of the *BACH2* locus in $CD8^+$ T cells …….. 314

Fig. S.11. Tiled-C interactions and the chromatin landscape of the *CIITA-DEXI-CLEC16A-SOCS1* locus in $CD8^+$ T cells ……... 315

Fig. S.12. Tiled-C interactions and the chromatin landscape of the *IL6R* locus in $CD4^+$ T cells ……. 316

Fig. S.13. Tiled-C interactions and the chromatin landscape of the *IL6R* locus in $CD8^+$ T cells ……. 317

Fig. S.14. Tiled-C interactions and the chromatin landscape of the *IL12RB2-IL23R* locus in $CD4^+$ T cells …… 318

Fig. S.15. Tiled-C interactions and the chromatin landscape of the *IL12RB2-IL23R* locus in $CD8^+$ T cells …… 319

Fig. S.16. Tiled-C interactions and the chromatin landscape of the *C1orf106-(INAVA)-GPR25* locus in $CD4^+$ T cells …….. 320

Fig. S.17. Tiled-C interactions and the chromatin landscape of the *C1orf106-(INAVA)-GPR25* locus in $CD8^+$ T cells …….. 321

Fig. S.18. Tiled-C interactions and the chromatin landscape of the *RUNX3* locus in $CD4^+$ T cells ……..… 322

Fig. S.19. Tiled-C interactions and the chromatin landscape of the *RUNX3* locus in $CD8^+$ T cells ……..… 323

Fig. S.20. Tiled-C interactions and the chromatin landscape of the *REL-B3GNT2* locus in $CD4^+$ T cells …… 324

Fig. S.21. Tiled-C interactions and the chromatin landscape of the *REL-B3GNT2* locus in $CD8^+$ T cells …… 325

Fig. S.22. Tiled-C interactions and the chromatin landscape of the *ICOS-CTLA4-CD28* locus in $CD4^+$ T cells …….. 326

Fig. S.23. Tiled-C interactions and the chromatin landscape of the *ICOS-CTLA4-CD28* locus in $CD8^+$ T cells 327

Fig. S.24. Tiled-C interactions and the chromatin landscape of the *CCR2-CCR5* locus in $CD4^+$ T cells 328

Fig. S.25. Tiled-C interactions and the chromatin landscape of the *CCR2-CCR5* locus in $CD8^+$ T cells 329

Fig. S.26. Tiled-C interactions and the chromatin landscape of the *EOMES-CMC1* locus in $CD4^+$ T cells 330

Fig. S.27. Tiled-C interactions and the chromatin landscape of the *EOMES-CMC1* locus in $CD8^+$ T cells 331

Fig. S.28. Tiled-C interactions and the chromatin landscape of the *CDKAL1* locus in $CD4^+$ T cells 332

Fig. S.29. Tiled-C interactions and the chromatin landscape of the *CDKAL1* locus in $CD8^+$ T cells 333

Fig. S.30. Tiled-C interactions and the chromatin landscape of the *IL22RA2-TNFAIP3* locus in $CD4^+$ T cells 334

Fig. S.31. Tiled-C interactions and the chromatin landscape of the *ILL2RA2-TNFAIP3* locus in $CD8^+$ T cells 335

Fig. S.32. Tiled-C interactions and the chromatin landscape of the *ETS1* locus in $CD4^+$ T cells 336

Fig. S.33. Tiled-C interactions and the chromatin landscape of the *ETS1* locus in $CD8^+$ T cells 337

Fig. S.34. Tiled-C interactions and the chromatin landscape of the *ATXN2-PTPN11-SH2B3* locus in $CD4^+$ T cells 338

Fig. S.35. Tiled-C interactions and the chromatin landscape of the *ATXN2-PTPN11-SH2B3* locus in $CD8^+$ T cells 339

Fig. S.36. Tiled-C interactions and the chromatin landscape of the *IKZF3-Aiolos-ORMDL3* locus in $CD4^+$ T cells 340

Fig. S.37. Tiled-C interactions and the chromatin landscape of the *IKZF3-Aiolos-ORMDL3* locus in $CD8^+$ T cells 341

Fig. S.38. Tiled-C interactions and the chromatin landscape of the *NPEPPS* locus in $CD4^+$ T cells 342

Fig. S.39. Tiled-C interactions and the chromatin landscape of the *NPEPPS* locus in $CD8^+$ T cells 343

Fig. S.40. Tiled-C interactions and the chromatin landscape of the *PTPN2* locus in $CD4^+$ T cells 344

Fig. S.41. Tiled-C interactions and the chromatin landscape of the *PTPN2* locus in $CD8^+$ T cells 345

Fig. S.42. Tiled-C interactions and the chromatin landscape of the *TYK2-ICAM1-ICAM3* locus in $CD4^+$ T cells 346

Fig. S.43. Tiled-C interactions and the chromatin landscape of the *TYK2-ICAM1-ICAM3* locus in $CD8^+$ T cells 347

Fig. S.44. Tiled-C interactions and the chromatin landscape of the *ETS2* locus in $CD4^+$ T cells 348

Fig. S.45. Tiled-C interactions and the chromatin landscape of the *ETS2* locus in $CD8^+$ T cells 349

Fig. S.46. Tiled-C interactions and the chromatin landscape of the *UBE2E3-MAPK* locus in $CD4^+$ T cells 350

Fig. S.47. Tiled-C interactions and the chromatin landscape of the *UBE2E3-MAPK* locus in $CD8^+$ T cells 351

LIST OF TABLES

CHAPTER 1: GENERAL INTRODUCTION

Table 1.1. BASDAI questionnaire given to diagnosed AS patients to assess the severity of their disease ……..… 8

Table 1.2. Timeline of the most important GWAS investigations in ankylosing spondylitis ……. 23

Table 1.3. Comprehensive list of AS-associated GWAS SNPs within European disease cohorts …….. 25

CHAPTER 2: METHODS AND MATERIALS

Table 2.1. Clinical features of AS patients included in this thesis ……. 68

Table 2.2. Sample use and genomic assays performed in recruited healthy volunteers and AS patients …… 71

Table 2.3. Primers used in the determination of Capture-C digestion efficiency ……... 78

Table 2.4. Capture-C oligonucleotide sequences and capture tiers …… 89

Table 2.5. Regions used in tiled oligonucleotide generation for Tiled-C …….. 93

CHAPTER 4: AS-ASSOCIATED GWAS LEAD SNPs SHOW *CIS*-REGULATORY LOOPING TO INFLAMMATION-RELATED GENE PROMOTERS WITH CAPTURE-C

Table 4.1. Summarisation of AS-associated GWAS loci showing *cis*-regulatory looping that intersects differential chromatin features between AS patients and healthy volunteers …… 185

Table 4.2. Evidence of enhancer-gene interactions at the nine identified AS-associated GWAS risk loci …….. 212

CHAPTER 5: APPLICATION OF TILED-C IN $CD4^+$ AND $CD8^+$ T CELLS TO INTERROGATE CHROMATIN CONTACTS IN AS-ASSOCIATED GWAS LOCI

Table 5.1 Summary of Tiled-C regions and associated figures ……..…….. 249

APPENDIX

Table S.1. SNP enrichment for ChromHMM states with an odds ratio of two or greater are linked to cis-regulatory elements in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells ……... 303

3C	Chromosome conformation capture
AHR	Aryl hydrocarbon receptor
ANOVA	Analysis of variance
AS	Ankylosing spondylitis
ASAS	Assessment of spondyloarthritis international society
ASDAS	Ankylosing spondylitis disease activity score
ATAC	Assay for transposase-accessible chromatin
axSpA	Axial spondyloarthropathy
BACH2	Broad complex-tramtrack-bric a brac and Cap'n'collar homology 2
BASDAI	Bath ankylosing spondylitis activity index
CD	Cluster of differentiation
ChIP	Chromatin immunoprecipitation
ChromHMM	Chromatin hidden Markov model
CLEC16A	C-type lectin domain family 16
CNV	Copy number variant
CRISPR	Clustered regulatory interspaced short palindromic repeats
CRISPRa/i	CRISPR activation/inhibition
CRP	C-reactive protein
CTCF	CCCTC-binding factor
DCLBD1	Discoidin, CUB, and LCCL domain containing 1
DEXI	Dexamethasone-induced protein
DMARD	Disease modifying anti-rheumatic drug
ENCODE	Encyclopaedia of DNA elements
EOR	Endoplasmic reticulum overload response
ER	Endoplasmic reticulum
ERAP	Endoplasmic reticulum aminopeptidase
ERN1	Endoplasmic reticulum to nucleus signalling 1 enzyme
eRNA	Enhancer RNA
ESR	Erythrocyte sedimentation rate
ETS	Ets proto-oncogene
EZH2	Enhancer of zeste homolog 2
eQTL	Expression quantitative trait locus
GARFIELD	GWAS analysis of regulatory/functional data with LD correction
GMDP	Guanylyl-muramyl dipeptide
GPCR	G-protein coupled receptor
GWAS	Genome-wide association study
H3K4me3	Trimethylation of lysine 4 of histone core 3
H3K27ac	Acetylation of lysine 27 of histone core 3
H3K27me3	Trimethylation of lysine 27 on histone core 3
HC	Heavy chain
Hi-C	High throughput chromosome conformation capture
HLA	Human leukocyte antigen
IBD	Inflammatory bowel disease

ICE	Iterative correction and eigenvector decomposition
IFNγ	Interferon gamma
IL	Interleukin
IL18RAP	IL-18 receptor accessory protein
IL-7R	Interleukin-7 receptor
IL-23R	Interleukin-23 receptor
INPP4B	Inositol polyphosphate-4-phosphate
IRF	Inflammatory regulatory factor
ITGA4	Integrin alpha 4
LD	Linkage disequilibrium
JAK	Janus kinase
KIR	Killer-like immunoglobulin receptor
LMO1	Rhombotin-1/LIM domain only 1
lncRNA	Long non-coding RNA
LPS	Lipopolysaccharide
MACS	Magnetic-activated cell sorting
MAPK	Mitogen-activated protein kinase
MCA	Multiple correspondence analysis
M-CSF	Macrophage colony stimulating factor
MDP	Muramyl dipeptide
MHC	Major histocompatibility complex
MRI	Magnetic resonance imagine
MS	Multiple sclerosis
mSASSS	Modified stoke ankylosing spondylitis spinal score
MYC	Myc proto-oncogene
NET	Neutrophil extracellular trap
NFκβ	Nuclear factor kappa beta
NK	Natural killer
NGS	Next generation sequencing
NO	Nitric oxide
NOS2	Nitric oxide synthase 2
nr-axSpA	Non-radiographic axial spondyloarthropathy
NSAID	Non-steroidal anti-inflammatory drug
NuRD	Nucleosome remodelling and deacetylase
PBMC	Peripheral blood mononuclear cell
PC	Principal component
PCR2	Polycomb repressive complex 2
PRR	Pattern recognition receptor
PsA	Psoriatic arthritis
PTGER4	Prostoglandin E receptor 4
RA	Rheumatoid arthritis
RANKL	Nuclear factor kappa-beta ligand
RUNX	Run-related transcription factor

SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
SpA	Spondyloarthropathy
STAT	Signal transducer and activator of transcription
T1D	Type 1 diabetes
TAD	Topologically-associated domain
Th	T-helper
TGF-β	Tumour growth factor beta
TLR	Toll-like receptor
TNF	Tumour necrosis factor
TSS	Transcription start site
Treg	T regulatory
TYK2	Non-receptor tyrosine-protein kinase
UB32E3	Ubiquitin-conjugating enzyme E2 L3
UPR	Unfolded protein response
UTR	Untranslated region
WGS	Whole genome sequencing
β2m	Beta-2-microglobulin

"Everything's science fiction until someone makes it science fact."

- Marie Lu, Warcross (2017)

1

GENERAL INTRODUCTION

Contents **Pg.**

1.1. Thesis objectives ……..…….…….……..…… 1
1.2. Overview of ankylosing spondylitis ……...……...……. 2
1.3. Genetics of ankylosing spondylitis …….. 14
1.4. Pathophysiology of ankylosing spondylitis ……..... 26
1.5. Gene regulation, genetic variants, and disease …… 43
1.6. Functional genomics and integrative analyses ……. 55
1.7. Specific aims and objectives …… 64

This chapter presents the overall aims and objectives of the thesis, and sets these in the context of the published literature on the genomics of ankylosing spondylitis with a focus on some of the major previously unaddressed genetic factors contributing to the disease.

1.1. Thesis objectives

In this thesis, I have employed functional genomic approaches to characterise ankylosing spondylitis (AS) disease-associated risk loci with a view to improving our understanding of the aetiology and pathophysiology of the disease. Genome-wide association studies (GWAS) for AS have identified numerous disease-associated genetic variants which lie outside of canonical protein-coding regions. The majority of these regions remain to be characterised to identify the causal gene and variant(s) in a disease-relevant biological and functional context. In this thesis, multiple functional genomic assays are performed to examine the local functional genomic landscape for a number of these AS risk-

association loci in immune cells. These assays include studying changes in histone modifications and gene expression in addition to inspecting regions of open chromatin to understand relationships with accessibility. Chromosome conformation capture (3C) techniques are utilised to link specific GWAS-associated variants functionally to their target genes, evaluating their molecular mechanisms of action, and ascertaining their relationship with disease state. These insights could help to identify novel druggable targets. In this chapter, an overview of the current understanding of AS pathophysiology is presented, concentrating on the genetic aetiology of the disease and potential molecular mechanisms involved. The challenges in evaluating these GWAS-associated genetic variants in such a complex disease is discussed, with further focus on the functional genomic techniques that can be applied to elucidate the chromatin landscape of specific loci. Finally, aims and objectives for each chapter will be detailed.

1.2. Overview of ankylosing spondylitis

1.2.1. Introduction to ankylosing spondylitis

AS is a member of a group of rheumatic conditions known as the spondyloarthropathies (SpA). These conditions also include, but are not limited to, psoriatic arthritis, reactive arthritis and the arthropathies associated with inflammatory bowel disease (IBD). All of these have their own specific features, but can appear indistinguishable from AS and exhibit considerable overlap in their clinical presentation (Essex *et al.*, 2024). AS is distinguished from other SpAs by its predominant axial skeletal involvement. Pathologically it is characterised by inflammation of the spinal entheses, which are specialised stress-bearing structures such as the fibrocartilaginous joints and the paraspinal attachments of ligaments and tendons to the intervertebral discs and vertebrae. Although any part of the spine can be affected, the fibrocartilaginous sacroiliac joints

connecting the pelvis and the spinal trunk are usually the first to be affected (Bray *et al.*, 2019). Inflammation at these entheses frequently leads to bone erosion. This in turn may be followed by uncontrolled osteoproliferation causing fusion of the affected joints (ankylosis). Bony outgrowths called syndesmophytes originating at the entheseal origin of the paraspinal ligaments proliferate to bridge across the intervertebral discs and fuse adjacent vertebral bodies (Tang *et al.*, 2023). In the most advanced cases of AS complete fusion of the spine may occur, resulting in the classic 'bamboo spine' and severe loss of spinal mobility (Tu *et al.*, 2022). Despite the potential severity of symptoms, the diagnosis of AS is frequently delayed. It typically takes 6 years for AS to be diagnosed even in the developed world and as few as one in three cases of AS are diagnosed within a year of onset (Deodhar *et al.*, 2016; Ogdie *et al.*, 2019).

In addition to loss of mobility and chronic pain in the axial skeleton, AS has a number of other comorbidities and higher mortality rates than the general population. Renal, cardiovascular, respiratory, and gastrointestinal complications are notably higher in those with AS (Haroon *et al.*, 2015; Exarchou *et al.*, 2016), and the excess mortality appears to be linked to the level of AS activity (Bakland *et al.*, 2011). In AS, there may be restricted lung function as a result of costovertebral joint fusion, and osteoporosis is a common complication rendering the patient susceptible to severe fractures; particularly since the risk of falls is also greater in those with compromised spinal flexibility (Ryan & Halliwell, 1999; Muñoz-Ortega *et al.*, 2014; Taurog *et al.*, 2016). AS is commonly associated with other inflammatory diseases, such as psoriasis, inflammatory bowel disease, and anterior uveitis. In our large sample from the UK, more than 40% of cases self-reported at least one episode of anterior uveitis (Brown *et al.*, 2023). Psoriasis has been reported in 17% (Lucasson *et al.*, 2022) and inflammatory bowel disease in at least 10% of AS patients; with IBD itself a risk factor in the development of the condition (de

Vlam *et al.*, 2000; Ding *et al.*, 2024). This suggests that there is substantial overlap between AS and these conditions at the level of genetic susceptibility, and that they likely share common pathophysiology through dysregulated immune or inflammatory pathways. This is a feature that has already had great effect in the identification of potential susceptibility genes through cross disease studies (Ellinghaus *et al.*, 2016).

1.2.2. Epidemiology of ankylosing spondylitis

AS is one of the most common forms of arthritis, with the highest incidence in populations of European, Inuit, and Han Chinese descent. Wider global prevalence of the condition is believed to lie between 0.1% to 1.4% (Dean *et al.*, 2014). The condition presents relatively early in life, with about 92% of AS patients presenting with the condition prior to the age of 45 (Boel *et al.*, 2022). Men were once thought to be as much as ten times more likely to develop AS than women, but more recent estimates suggest a sex ratio of closer to 2:1 – in part relating to the development of new diagnosis criteria and recognition of atypical presentations in women. AS has classically been diagnosed by the presence of radiographic changes in the sacroiliac joints according to the 1984 modified New York criteria, but these changes may take many years to develop and are considered too insensitive for modern use (Robinson *et al.*, 2021).

It is now standard practice to use magnetic resonance imaging (MRI) to identify sacroiliitis earlier in the course of the disease so that appropriate treatment can be initiated. About 12% of those with suspected AS identified by changes on MRI scans - characterised as non-radiographic axSpA (nr-axSpA) - progress to classical radiographic signs of AS within two years (Poddubnyy *et al.*, 2011). Longer follow up suggests that progression to radiographic axSpA or AS rises to 60% over ten years (Mau *et al.*, 1988). Similar figures are given for undifferentiated SpA, where roughly 10% of patients

progress to AS over two years and 23% over five to ten years (Sampaio-Barros *et al.*, 2001; Sampaio-Barros *et al.*, 2010).

When the broader diagnosis of generalised axSpA is taken into consideration, women appear to be affected almost as commonly as men (Baumberger & Khan, 2016; Taurog *et al.*, 2016). However, men and women do differ in disease manifestation. Males with axSpA tend to present with a more obviously inflammatory form of axSpA characterised by higher levels of the circulating inflammatory marker C-Reactive Protein (CRP), higher MRI scores at inflammatory sites, a higher rate of progression to the radiological stage of AS, and earlier diagnosis (Tournadre *et al.*, 2013; Boel *et al.*, 2022). Women tend to present with higher disease activity scores, a greater degree of peripheral skeletal involvement, and a reduced quality of life when compared to men (Landi *et al.*, 2016). The actual incidence of AS may therefore be higher than reported, as many cases are excluded based upon a singular interpretation of what constitutes AS.

It has been estimated that AS in the United Kingdom alone has a cost burden of roughly £19,000 per patient per year, with the majority of expenses relating to work-related issues including reduced workplace productivity and early retirement (Cooksey *et al.*, 2015). The high prevalence of AS coupled with its high socioeconomic costs and early age of onset mark it as one of the more important challenges in rheumatological healthcare.

1.2.3. Diagnosis of ankylosing spondylitis

One of the most important barriers in the study of AS is the correct diagnosis of the disease. AS shares many of its clinical features not only with other forms of SpA but also other conditions causing back pain such as degenerative disc disease (Resorlu *et al.*, 2015). Historically, the diagnosis of classical AS has relied upon the 1984 New York classification criteria, which in part defined AS on the basis of radiographic evidence of

disease from X-ray imaging of the sacroiliac joints in addition to at least one clinical criterion. In 2009 it was noted that such strict reliance on radiographic evidence would exclude patients with non-radiographic varieties of AS, which is much more common in the early stages of the disease. The ASAS (**A**ssessment of **S**pondylo**A**rthritis international **S**ociety) AS diagnostic criteria were subsequently expanded to include a second category of positive diagnosis, incorporating a positive result for human leukocyte antigen (HLA)-B27 with at least two clinical criteria (Rudwaleit *et al.*, 2009; Rudwaleit *et al.*, 2011) (**Fig 1.1.**) The role and relevance of HLA-B27 in AS predisposition is discussed further later in this chapter.

The current disease activity of AS is usually measured by BASDAI (**B**ath **A**nkylosing **S**pondylitis **D**isease **A**ctivity **I**ndex), but it can be very difficult to distinguish current disease activity from the effects of long-term damage. The BASDAI is a self-reported score based upon the subjective ranking of 0 to 10 for six questions pertaining to symptoms the patient has experienced over the preceding week. These include questions on spinal pain, peripheral joint pain and swelling, areas of localised tenderness surrounding the entheses, and the average duration and severity of morning stiffness (**Table 1.1.**).

The BASDAI is the mean of these categories with a score of 0 indicating total lack of symptoms. In practice few individuals in the general population would score zero since these symptoms are common throughout the population. A BASDAI score of 4 or more is indicative of severe AS (but could be mimicked by other conditions, including disc disease) and serves as a soft threshold for therapeutic intervention with biologic drugs such as the anti-tumour necrosis factor monoclonal antibodies (anti-TNFα) (Rudwaleit *et al.*, 2004; Keat *et al.*, 2005; Hamilton *et al.*, 2017). A decrease in the BASDAI of 50%

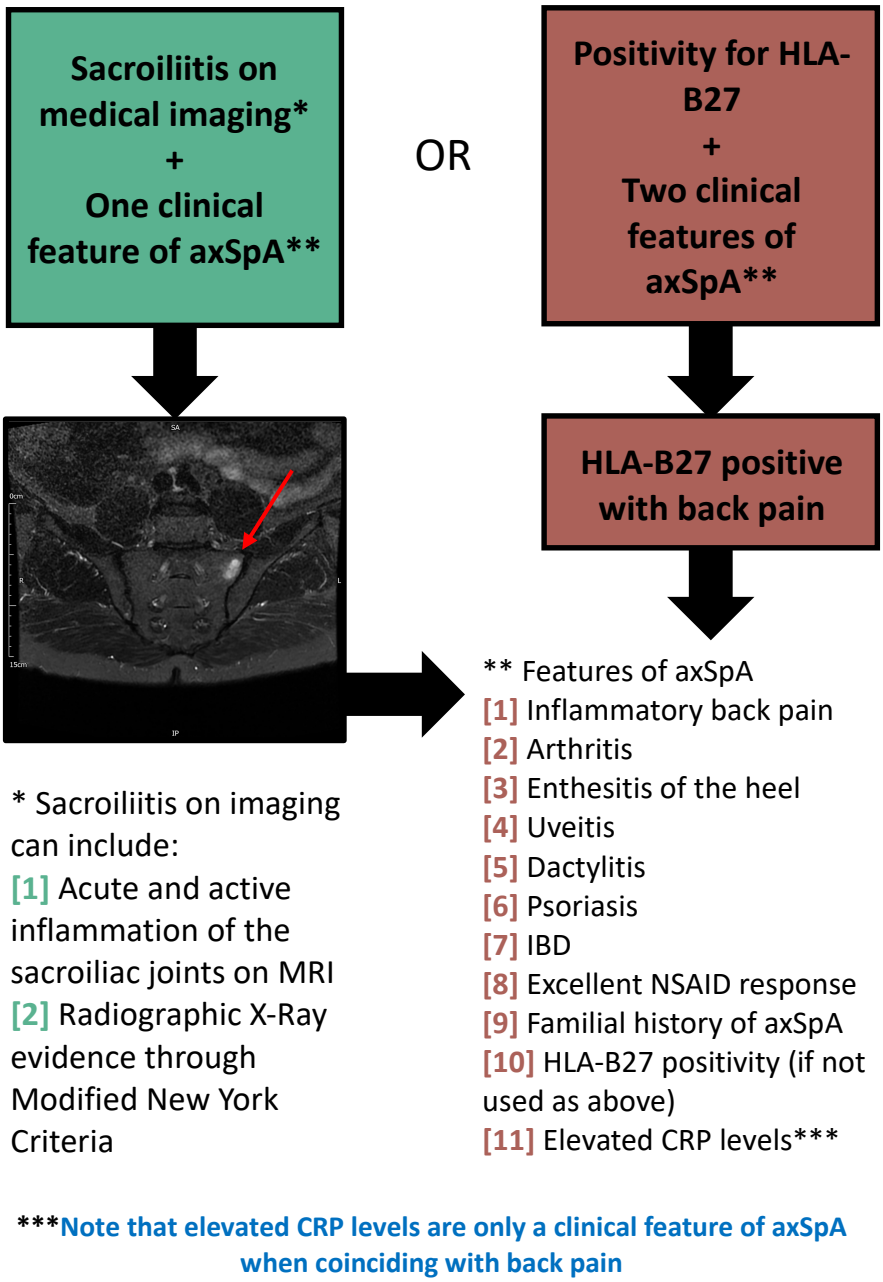

Figure 1.1. ASAS classification criteria for the diagnosis of axial spondyloarthropathy. The two ASAS categories used for the diagnosis of axSpA. Criteria are to be used in chronic cases of back pain greater than 3 months within patients under the age of 45. MRI image on the left-hand side represents an MRI image taken from an AS patient recruited for epigenetic analysis in this thesis. The red arrow points to an area shaded in white within the sacroiliac joint demonstrating the infiltration of water at the site, which indicates localised inflammation of the joint. Figure adapted from Ramiro *et al.* (2022).

Bath Ankylosing Spondylitis Activiy Index (BASDAI)
Please indicate the level for each of these questions from the proceeding week
On a scale of 0 to 10:
(1) How would you describe the overall level of fatigue/tiredness you have experienced?
(2) How would you describe the overall level of AS neck, back, or hip pain you have had?
(3) How would you describe the overall level of pain/swelling in joints other than neck, back, or hips you have had?
(4) How would you describe the level of discomfort you have had from an area tender to touch or pressure?
(5) How would you describe the level of morning stiffness you have had from the time you wake up?
(6) How long does your morning stiffness last from the time you wake up?
0 is the absence of any symptoms, and 10 is the presence of the most severe symptoms

Table 1.1. BASDAI Questionnaire given to diagnosed AS patients used to assess the severity of their disease. Each question aside from the last is scored from 0 to 10 with 0 indicating no symptoms and 10 indicating severe symptoms. The final question is scored 0 for no hours, 5 for one hour, and 10 for two hours of morning stiffness. Overall BASDAI score is calculated using the algorithm ((Q1 + Q2 + Q3 + Q4) + ((Q5 + Q6) ÷ 2))) ÷ 5. A score ≥4 is indicative of uncontrolled AS.

or to below a score of four is often used to define a successful clinical response to biologic treatment (Rudwaleit *et al.*, 2004; Keat *et al.*, 2005). Other questionnaires have been developed to assess disease activity, such as the Ankylosing Spondylitis Disease Activity Score (ASDAS), which measures responses to similar questions. However, all these questionnaires are similar in design and produce similar scores for AS activity to the BASDAI. None are truly objective and all have the same drawbacks in distinguishing current disease activity from symptoms due to longstanding damage caused by the disease. These studies also demonstrate comparable cut-off scores for therapeutic intervention (Kılıç *et al.*, 2014) and in the studies that I have reported in this thesis the BASDAI was used throughout to define disease activity.

The progressive damage of the spine seen in AS can be objectively monitored using the modified Stoke Ankylosing Spondylitis Spinal Score (mSASSS), which is a quantitative score of radiographic damage based on sclerosis, erosion and fusion around the spinal entheses in AS patients (van der Heijde *et al.*, 2019). Using mSASSS, each vertebra except the third cervical vertebra is assigned a score of zero to three. Zero indicates no abnormality of bone structure, one indicates a combination of early erosion, squaring, or sclerosis of the vertebra, two the formation of non-bridging syndesmophytes, and three total bridging of the gap between adjacent vertebrae (Creemers *et al.*, 2005). The maximum mSASSS score achievable is a total of 72 which represents complete ankylosing of the spinal column. As mSASSS is a progressive score, it is unusual to see its use in the diagnosis of AS. mSASSS is however used extensively in the assessment of patient response to various treatments, where a slower attainment of higher mSASSS scores associated with successful therapeutic intervention of the disease (van der Heijde *et al.*, 2019).

1.2.4. Contemporary and future biomarkers for ankylosing spondylitis

Whilst BASDAI is currently the standard aid for assessing AS disease activity, it is far from perfect since first, it measures the severity of symptoms that could be caused by chronic disease damage rather than current activity, and second, it is highly subjective. Ideally, objective measures of inflammation should also be included, such as blood markers of inflammation (CRP and/or erythrocyte sedimentation rate - ESR) and/or MRI of the axial skeleton. Unfortunately, the CRP is often unhelpful since only around 40% of patients with active AS mount a CRP response (Baralaikos *et al.*, 2019). In cases where a CRP response is present, CRP also poorly correlates with AS disease activity; being neither sensitive nor specific for the diagnosis of the condition (Spoorenberg *et al.*, 1999). Use of ESR also suffers from many of the same pitfalls in measuring AS activity (Reveille *et al.*, 2015).

A number of alternative biomarkers for AS are under investigation, ranging from the expression levels of several inflammation-related proteins like macrophage inhibitory factor (MIF), CD74 (HLA class II histocompatibility antigen gamma chain), anylyl aminopeptidase (ANPEP) (Baralaikos *et al.*, 2014; Li *et al.*, 2023) and composite micro-RNA (miRNA) signatures (Perez-Sanchez *et al.*, 2018). Whilst all of these studies show promise, they remain in the early stages of research and necessitate further validation as genuine biomarkers of disease before incorporation into diagnostic criteria. MRI coupled with BASDAI is undoubtedly the gold standard for identifying current disease activity in the axial skeleton, but is relatively expensive and subject to strictures in availability. It is hoped by further studying the aetiology of AS that novel biomarkers for earlier disease identification will be uncovered in addition to therapeutic targets.

1.2.5. Therapeutic strategies for treating ankylosing spondylitis

Treatment of AS typically revolves around relieving inflammatory back pain, halting or slowing the rate of vertebral fusion, reducing the severity of AS comorbidities, and maintaining flexibility of the spinal column. Non-steroidal anti-inflammatory drugs (NSAIDs) such as ibuprofen and naproxen are often employed in the first response to treatment within AS (Kroon *et al.*, 2015). These drugs work through competitive inhibition of cyclooxygenase (COX) thereby suppressing prostaglandin biosynthesis (Ju *et al.*, 2022). This pathway results in the prostanoids PGD2, PGE2, PGF2α, PGI2, and TXA2 which serve as lipid ligands for a range of G-protein coupled receptors (GPCRs) (An *et al.*, 2021). The resultant effects of these GPCRs are pleiotropic but specific, with the type of inflammation, subtype of receptor, and both the timeframe and nature of the disease affecting pathway outcome. PGE2 in particular has been studied extensively within AS, being linked to the promotion of pro-inflammatory interleukin 17 (IL-17) and interleukin 23 (IL-23) pathways within T-helper cells (An *et al.*, 2021). A meta-study of 39 independent studies demonstrated that the BASDAI of AS patients treated with NSAIDs decreased significantly by roughly 30% compared to placebo groups (Kroon *et al.*, 2015).

Whilst NSAIDs are therefore useful in reducing inflammatory back pain and the loss of mobility associated with it, other studies have found that a large proportion of patients (≥40%) still present with BASDAI scores greater than four following treatment (Baraliakos *et al.*, 2016). Many studies have also found no difference in the radiographic progression of AS-linked spinal damage between NSAID and placebo users, with changes seen to the mSASSS scores of both remaining insignificant over a two to three-year period of treatment (Zhang *et al.*, 2019). NSAIDs are therefore of subjective efficacy

in the treatment of AS but do not exert a fundamental effect on the underlying disease process to reduce the degree of spinal ankylosis.

AS patients with higher BASDAI, ASAS, or ASDAS scores are theoretically good candidates for treatment with disease-modifying anti-rheumatic drugs (DMARDs) of the sort used to treat other inflammatory forms of arthritis (Sepriano *et al.*, 2021). DMARDs are quite effective in the treatment of rheumatoid arthritis (RA), with 30% of RA patients able to achieve remission with DMARD treatment (George *et al.*, 2021). DMARDs are typically small molecule inhibitors that competitively bind to inflammation-related proteins and modulate their activity. In AS, the use of DMARDs, such as those used in RA, is somewhat controversial. In general, they do not improve symptoms in the axial skeleton but may improve peripheral arthritis. Standard slow-acting DMARDs also do not appear to affect the rate of spinal damage or axial radiographic progression in AS (Chan & Gladman, 2018). A study of 1280 AS patients treated with the DMARDs sulfasalazine or methotrexate over a period of 18-years at the same hospital found that mean adjusted mSASSS scores were not significantly different between on-interval and off-interval AS patients (Lee *et al.*, 2020).

In the UK, AS patients with high disease activity (BASDAI ≥ 4), in whom treatment with at least two different NSAIDs at their maximum dosage for over two weeks have failed, should be considered for antibody biologic therapies (Chen, 2022). These drugs act through steric hindrance of the association of various signalling receptors and their cognate ligands, thereby downregulating the associated molecular signalling cascades that promote inflammation. The most widely used biologics in AS target TNFα (i.e. anti-TNFα antibodies, such as adalimumab and golimumab, and recombinant TNF receptor/Fc fusion proteins, such as etanercept). More recently, anti-interleukin 17 (IL-17) biologics (anti-IL-17), such as secukinumab, have proved equally successful. Both of

these classes of drugs significantly reduce ASAS scores rapidly in AS patients (Cao *et al.*, 2022). Anti-TNFα and anti-IL-17 therapies are both associated with prompt reduction in sacroiliac joint inflammation, decreased axial radiographic progression, and long-term efficacy and tolerance (McVeigh & Cairns, 2006; Martín-Mola *et al.*, 2010; Blair & Dhillon, 2016; Braun *et al.*, 2019).

Despite the efficacy of biologics, there are a number of issues that limit their use. Initial and long-term efficacy and tolerance to treatment with biologics is not always maintained. Indeed, between 30 % to 40% of axSpA either do not respond to anti-TNFα therapy or become tolerant to it over the course of treatment (Antonelli *et al.*, 2015; Menegatti *et al.*, 2019). Anti-IL-17 therapy has been shown to sometimes exacerbate pre-existing IBD, and promotes the formation of new gastrointestinal complications in some AS patients, suggesting anti-IL-17 driven propagation of gut inflammatory responses in some individuals (Mahmood *et al.*, 2018). This could be particularly relevant in view of the known tendency for low-grade inflammatory changes in the terminal ileum of individuals with AS (Ciccia *et al.*, 2016).

Biologics have some immunosuppressive effects and are associated with an increased risk of infection, particularly reactivation of latent tuberculosis. Quartuccio *et al.* (2019) demonstrated that 289 of 5596 RA, AS, and psoriatic arthritis (PsA) patients treated with biologics at a single hospital from 2006 to 2017 were hospitalised with serious infection, with a significant 55% increase in infection risk also being observed within the biologic treated cohort on first time of biologic use. There is clearly a need to develop new therapies for AS when one considers the limited range of long-term treatment options for AS, the fact that up to 50% of patients in the UK treated with biologics fail to respond meaningfully in the long term (Jones *et al.*, 2020), and that there may be significant treatment side-effects in some cases. By investigating the genetic aetiology of AS in

depth, my aim in this thesis is to identify potential drug targets that can be used to increase the proportion of patients with AS achieving remission.

1.3. Genetics of ankylosing spondylitis

1.3.1. Heritability of ankylosing spondylitis

AS is highly heritable (Brown *et al.*, 2016). As in any complex polygenic disease the exact contribution from the environment and the host genome is difficult to assess. Heritability scores for diseases like AS have been typically calculated by comparing concordance rates between monozygotic and dizygotic twins that have shared a similar environmental upbringing. Using these methods, broad-sense heritability assessed from the two formal AS twin studies has yielded similar h^2 values in excess of 90% (Brown *et al.*, 1997; Pedersen *et al.,* 2008).

Results from twin-studies in AS are inevitably limited by a number of confounders (Brown & Wordsworth, 2018). First, the definition of AS by relatively strict historical criteria identifies only those with relatively severe disease. This group would be expected to be at greater genetic risk than may be typical in the AS-affected population overall. Second, despite AS being one of the more common arthritic conditions (prevalence ~ 1:200 in the UK), identifying twin pairs is labour intensive and difficult. Consequently, not only are the published AS twin-studies underpowered, but there is also perhaps an ascertainment bias towards identifying concordant rather than discordant twin pairs. This is because the former are seen as more memorable both by the affected twins and their medical attendants. As a result, statistical measures of variance, such as the confidence interval or odds-ratios within these studies are relatively high, and the conclusions may not reflect the true level of heritability. Third, twin study methodology assumes that environmental factors affect each generation equally. Underestimation of the differences

that occur between parental-child environments could lead to overestimation of the contribution of genetics within AS. Despite these reservations, the high concordance rates for AS in monozygotic twins suggest a high degree of heritability attributable mostly to the genetic makeup of an individual, of which only a minority (probably <30%) is contributed by HLA-B27 (Brown *et al.*, 1997). These findings from the largest reported twin study strongly support the goal of identifying the remaining genes contributing to AS susceptibility.

1.3.2. Genetic variation and protein expression

Genetic variation is one of the major contributors to disease susceptibility (Ku *et al.*, 2010; Claussnitzer *et al.*, 2020). Specific variation at the same given genomic position on a chromosome is known as an allele, and allelic DNA forms can arise through a number of mechanisms. The more common of these variants (>1% minor allele frequency) are referred to as single nucleotide polymorphisms (SNPs), where a single cytosine (C), adenine (A), guanine (G), or thymine (T) base is substituted for another. Other genetic variants include indels, which can cause frameshift of the DNA through small additions or deletions of base pair sequences, or structural variation, which typically relates to insertion, deletion, duplication, or inversion of DNA sequences greater than 50 nucleotides in size (Alkan *et al.*, 2011). Copy number variants (CNVs) are also included under these structural variants. CNVs are highly repetitious blocks of DNA that are sequentially copied a number of times with the number of which varying significantly between individuals.

If found in a coding region such as a gene, SNPs, indels, and structural variants can result in synonymous mutation where the redundancy of the triplet code consensus sequences still allows for the correct amino acid recruitment, a synonymous mutation, or non-

synonymous mutation, where the genetic variation results in the addition or deletion of amino acids – or substitution for entirely different amino acids – from proteins (Jilani *et al.*, 2022).

Both synonymous and non-synonymous mutations are thought to play a role in disease. The effects of deleterious non-synonymous mutations are more obvious, where the substitution of one amino acid with amino acids that have differing R-group chemical properties and characteristics – or their deletion or addition of other amino acids – can alter the tertiary structure and folding kinetics of the final protein (Sun & Yu, 2019). Although non-synonymous mutations likely explain a greater deal of the genetic aetiology behind both common and rare diseases, synonymous mutations have been determined to have a number of potential disease-relevant deleterious effects (Dhindsa *et al.*, 2022). Synonymous mutations can:

- Impact mRNA stability by altering the secondary and/or tertiary structure of the molecule. Differing base pairing properties caused by the introduction of a different base may expose or hide sequences targeted by RNA degradation machinery (Kristofich *et al.*, 2018; Gaither *et al.*, 2021). These structural changes may also affect the interaction between mRNA and other molecules such as the ribosome or miRNAs and therefore the rate of protein translation (Gaither *et al.*, 2021; Vihinen, 2022). A decrease in mRNA stability associated with synonymous structure changes can be seen at the SNP *rs1076560* in the dopamine receptor D_2 gene; where a guanine to thymine mutation results in the same amino acid sequence but a significantly less stable mRNA molecule (Duan *et al.*, 2003; Gaither *et al.*, 2021).

- Affect codon usage bias. Different synonymous codons encoding the same amino acid can be used with varying frequencies – based on factors such as the relative

availability of corresponding transfer RNAs – which can alter the speed of protein translation and ribosomal pausing (Narula *et al.*, 2019; Li *et al.*, 2021). This can impact co-translational folding and lead to misfolded or non-functional proteins; as can be seen with the ΔF508 mutation within the *CFTR* gene in cystic fibrosis (Bartoszewski *et al.*, 2010; Lan *et al.*, 2024).

- Alter the roles of the 5' and 3' untranslated regions (UTR). Mutation to the 5' UTR sequence can affect how easily translation initiates, could alter mRNA secondary structure as above, disrupt or create upstream open reading frames, modify internal ribosome entry sites, or affect the binding of ribosomes and translation initiation factors (Hinnebusch *et al.*, 2016; Gaither *et al.*, 2021; Soukarieh *et al.*, 2022; Barrington *et al.*, 2023). Mutations in the 3' UTR can affect mRNA stability, localisation, and translational efficiency; with mutations here often being found in regulatory domains for RNA-binding proteins and miRNAs and affecting post-transcriptional processes such as polyadenylation (Suhl *et al.*, 2015; Mayr *et al.*, 2017; Schuster & Hsieh, 2019; Schuster *et al.*, 2023).

- Influence splice site selection and alternative protein splice forms by mutating recognition site for splicing regulatory proteins. This can contribute towards aberrant splicing events such as exon skipping or intron inclusion, which can drastically alter the tertiary structure and function of the final peptide (Sharma *et al.*, 2019). A synonymous adenine to guanine mutation at position 414 of the von Hippel-Lindau disease tumour suppressor gene (*VHL*), for instance, leads to skipping of exon 2, which recognises hypoxia inducible factor 1 (*HIF-1*) and ordinarily contributes towards effective *HIF-1* degradation (Flores *et al.*, 2019). Constitutive *HIF-1* hyperactivation can then contribute towards the development of renal cancer, hemangioblastoma, and pheochromocytoma in variant-bearing individuals.

Whilst the effects of synonymous and non-synonymous mutations are better understood, it is lesser known what effects mutations outside of canonical protein coding regions can have towards disease. Multi-factorial traits, such as AS, demonstrate greater association with non-coding variants than their monogenic counterparts (Gaulton *et al.*, 2023). There are a number of potential mechanisms of action of such variants including effects on the expression of genes, and as such are thought to perturb the functionality of regulatory elements that modulate their expression (Duan *et al.*, 2023). Understanding how this non-coding genetic variation modulates specific genes/pathways and impacts disease is important to identify novel targets for drug development and precision medicine approaches.

1.3.3. Genome-wide association studies

GWAS have greatly facilitated the identification of genetic variants that contribute to complex traits or diseases (Visscher *et al.*, 2017; Kreitmaier *et al.*, 2023). Through direct comparison of well-powered, carefully controlled disease and age-matched cases and healthy control samples, GWAS can identify genetic variants that present with statistically higher frequencies in the disease groups, thereby implicating that particular variant or others in linkage disequilibrium with it.

A p-value of $5x10^{-8}$ is the generally accepted statistical threshold for genome-wide association; but this should also be supported by subsequent validation in a second, independent GWAS in the same ethnic group of at least nominal significance (i.e. p-value < 0.05). Variants passing this threshold can be defined confidently as disease-associated (Hammond *et al.*, 2021). Variants between with p-values between $5x10^{-4}$ and $5x10^{-8}$ are 'suggestive', and may become genome-wide significant with increasing cohort sizes and statistical power.

GWAS can be directly used to assess disease heritability in lieu of twin studies at the resolution of individual genetic loci (Voruganti & Bowness, 2020). These studies can be performed in a number of ways, including with whole genome sequencing (WGS), where the entire genomes of individuals in the disease and healthy cohorts are compared, or on microarrays. These microarrays consist of known DNA sequences (probes) attached to a sequencing chip that target specific SNPs or other genetic variants of interest, allowing for the identification of which alleles or genetic variants exist at a given gene or genetic location (the genotype) on a per individual basis in isolation from the rest of the DNA (Verlouw *et al.*, 2021).

Whilst WGS GWAS has seen some use in AS, the small number of case and control samples included in those studies renders them underpowered. There have been a number of sufficiently powered studies that have employed the genotyping microarray approach however, and two of these microarrays in particular, the Affymetrix Axiom chip, with a more general genome-wide coverage, and the Illumina Immunochip, which focuses on a subset of inflammatory or immune system-relevant genes and SNPs, estimate AS heritability to be between 30% to 70% (Burton *et al.*, 2007; Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016; Brown & Wordsworth, 2018). This was far higher than the heritability scores for related immune conditions such as IBD (20-33%) or psoriasis (18-20%) and further evidenced the strong genetic component of AS pathogenesis. However, these studies have yielded hundreds of novel individual AS-associated loci, which renders investigation of the precise causative genetic elements a significant challenge.

1.3.4. Linkage disequilibrium and statistical fine-mapping

One of the greatest challenges to the interpretation of GWAS is linkage disequilibrium (LD). LD is defined as the non-random association of alleles at different genomic loci in

a given population. The pattern of LD is reflective of the complex interplay between DNA recombination rates, selection, genetic drift, mutation rates, mating systems, and population structure that a given population has experienced (Slatkin *et al.*, 2016). LD can and does therefore vary widely between different human ethnic groups, such as Europeans and Sub-Saharan Africans (Vergara-Lope *et al.*, 2019). LD is measured as a correlation coefficient (*D*) which is defined by the equation:

$$D = p_{AB} - p_A p_B$$

Where p_A and p_B are the observed frequencies of alleles A and B at a given locus, and p_{AB} is the frequency of observing them together (Slatkin *et al.*, 2016). When p_{AB} differs from $p_A p_B$ the two loci are determined to be in LD. The correlation coefficient requires further normalisation, and can be presented as either *D'* or R^2. *D'* is calculated by dividing *D* by the theoretical maximum difference (D_{max}) between the observed and expected allele frequencies between two loci (VanLiere *et al.*, 2008). R^2 is instead calculated by taking the square of *D* between pairs of loci (VanLiere *et al.*, 2008; Slatkin *et al.*, 2016). *D'* is measured between -1 and +1, with -1 meaning that there is no association between the frequencies of two given alleles and +1 representing 'complete' LD, where the alleles are not separated by recombination. R^2 is instead measured from 0 to 1, with 0 representing no association and 1 representing 'perfect' LD, where in addition to not having been separated by recombination, the alleles occur at the same frequency in the analysed population (Slatkin *et al.*, 2016)

LD has profound effects on the interpretation of GWAS association scores, as genetic variants in LD will be inherited together and maintain the haplotype (a group of alleles that are inherited from a single parent) identity of all SNPs within an LD block. The significance of GWAS association is therefore typically presented over a block of SNPs

in high LD, and the SNP with the highest association signal in the LD block is known as the lead SNP. The lead SNP does not always have direct functional implications, however it usually will show a high *D'* or R^2 value with the true functional genetic variant.

Statistical fine-mapping is a technique that utilises imputation to provide a credible set of SNPs within the LD haplotype block that associate with the lead SNP signal (Howie *et al.*, 2012; Spain & Barrett, 2015; Schaid *et al.*, 2018). These credible sets can be used to delineate the boundaries for where the causative genetic variant will lie. Common sources used to generate these credible sets include the 1000 Genomes Project (Durbin *et al.*, 2010) or the International HapMap Project (Thorisson *et al.*, 2005). Bayesian approaches can be used on the resultant credible set to identify the likelihood that one is causal by calculating the posterior probability of each genetic variant (Mortezaei & Tavallaei, 2021). However, in the case of GWAS utilising microarray approaches, because not all genetic variants are included in their chip-based approach, multiple SNPs that share an LD block with the lead-SNP will be missed from the analysis. These regions must first be re-sequenced with next generation sequencing (NGS) technologies to identify any missed or rare genetic variants (van de Bunt *et al.*, 2015). Bayesian approaches also do not exclude instances where multiple genetic variants have a direct functional role in the disease. In these rare instances, overlap of the credible set with functional genomics data, such as that generated by the Encyclopaedia of DNA Elements (ENCODE) consortium, can facilitate the identification of perturbed genetic elements within the associated locus and further narrow down the list of putatively causal SNPs (Morova *et al.*, 2023).

1.3.5. GWAS in ankylosing spondylitis

The first association studies to be carried out in AS were not GWAS. In 1973, Brewerton *et al.* (1973) first showed that HLA-B27 was associated with AS by comparing a sample

of AS cases to healthy controls using the labour-intensive technique of microcytoxicity assays to assign HLA phenotypes, which rely on measuring cellular death in response to sera containing antibodies targeted to known HLA antigens. With their panel of 26 sera, the HLA-B27 serum was found to promote cell death in 96% of the recruited AS patients compared to only 4% of the healthy controls, thereby identifying HLA-B27 as the first genetic association to be found in AS. Since then, innumerable association studies on "candidate genes" have been undertaken in AS, and in more recent times, several GWAS have been conducted for AS (**Table 1.2**).

The first AS-directed GWAS studies were conducted in 2007 by the Welcome Trust Case Consortium and subsequently by the Australo-Anglo-American Spondylitis Consortium (Burton *et al.*, 2007). The first study used 15,333 gene targeted non-synonymous genetic variants, including 897 non-synonymous SNPs in the MHC genomic region on chromosome 6. The authors identified two AS-associated risk loci at the interleukin-23 receptor (*IL-23R*) and the endoplasmic reticulum aminopeptidase 1 (*ERAP1*). Similar studies with greater statistical power performed by Reveille *et al.* (2010) and Evans *et al.* (2011) with larger patient and healthy control populations also uncovered associations at *RUNX3*, *TNFRSF1A, IL12B*, *PTGER4*, *ANTXR2*, *CARD9*, *IL1R*, and *TBKBP1*. The largest association studies to date were conducted by Cortes *et al.* (2013) and Ellinghaus *et al.* (2016). Combined, these studies have shown that the major histocompatibility complex (MHC) class I antigen HLA-B27 demonstrates the highest observed association values within AS with a p-value of approximately10^{-900} (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016; Ko *et al.*, 2022), but that upwards of 113 other lead SNP associations can act to predispose an individual towards AS (**Table 1.3**). How a number of the best characterised AS-associated loci can functionally contribute to AS will be discussed in the next section.

Genome-Wide Association Studies conducted for Ankylosing Spondylitis

Author	Date	Cohort Numbers (AS/HV)	Coverage or Variant Type	Cohort Ethnicity	No. Identified non-MHC Genetic Variants
Burton *et al.*	2007	1000/1500	NSV	European	2
Reveille *et al.*	2010	2053/5140	GWAS	European	4
Evans *et al.*	2011	3023/8779	GWAS	European	8
Lin *et al.*	2011	1965/4301	GWAS	Han Chinese	3
Cortes *et al.*	2013	10,619/10,619	Illumina Immunochip	European, Han Chinese, Latino	24
Robinson *et al.*	2016	5040/21,333	Illumina Exomchip	European	2
Ellinghaus *et al.*	2016	8726/34,213	Illumina Immunochip	European, Han Chinese	113
Ko *et al.*	2022	206/1236	GWAS	Han Chinese	108

***Exclusive of any genetic variants within HLA-A, HLA-B, or HLA-C loci**

Table 1.2. Timeline of the most important GWAS investigations in ankylosing spondylitis. Table above shows the number of loci associated with AS at the genome-wide significance level of $5x10^{-8}$ in multiple GWAS between 2007 to 2022. The first column shows the first author of each study. The second column shows the year of study publication. The third column shows how many individuals were recruited in each study. The first number and second number show how many AS patients and healthy controls were recruited respectively in each study. The fourth column shows the type of study. The abbreviation NSV is used for GWAS that only studied non-synonymous variants. The fifth column shows the ethnic breakdown of each study. The sixth column shows the total number of non-MHC associations uncovered in each study. Table has been adapted from Li & Brown (2017).

Comprehensive List of AS-associated GWAS Risk Loci

Lead GWAS SNP	Chrom	Position	Gene	Allele (A1/A2)	Odds Ratio (OR)
rs11249215	1	25297184	*RUNX3*	A/G	1.19
rs6600247	1	25305114	*RUNX3*	T/C	0.83
rs80174646	1	67708155	*IL23R*	T/G	0.6
rs10889676	1	67722567	*IL23R*	A/C	1.16
rs4129267	1	154426264	*IL6R*	T/C	0.85
rs12131796	1	200878727	*C1orf106/GPR25*	A/G	0.83
rs2666218	2	9402988	*ASAP2*	G/A	1.12
rs10865331	2	62551472	*2p15 Intergenic*	A/G	1.3
rs4851529	2	102647300	*IL1R2*	A/G	0.88
rs871656	2	102771282	*IL1R1*	A/T	0.87
rs11306716	2	204707771	*CD28/CTLA4*	I/D	1.14
rs3749171	2	241569692	*GPR35*	T/C	1.18
rs10510607	3	28286261	*CMC1*	T/C	0.87
rs6781808	3	137441372	*SOX14*	C/T	1.14
rs12504282	4	80927001	*ANTXR2*	C/T	0.88
rs3774937	4	103434253	*NFKB1*	C/T	1.12
rs11750385	5	10521668	*Uncertain*	G/T	1.11
rs469758	5	96121715	*ERAP1*	C/T	1.34
rs2549803	5	96174929	*ERAP1*	T/C	0.87
rs2910686	5	96252589	*ERAP2*	C/T	1.21
rs6556416	5	158818745	*IL12B*	C/A	1.17
rs56167332	5	158827769	*IL12B*	A/C	1.13
rs2328530	6	20643727	*CDKAL1*	G/A	0.88
rs6908425	6	20728731	*CDKAL1*	C/T	1.2
rs67025039	6	167415404	*FGFR1OP*	I/D	1.21
rs1525735	7	17197571	*AHR*	C/T	1.17
rs1250573	10	81042475	*ZMIZ1*	A/G	0.89
rs1800682	10	90749963	*ACTA2*	G/A	0.89
rs10748781	10	101283330	*NKX2-3*	C/A	1.11
rs7115956	11	109956346	*ZC3H12C/RDX*	C/T	1.23
rs7933433	11	128194450	*ETS1*	T/G	1.12
rs1860545	12	6446777	*TNFRSF1A*	A/G	0.85
rs11616188	12	6502742	*LTBR*	A/G	1.11
rs11624293	14	88488821	*GPR65*	C/T	1.23
rs26528	16	28517709	*IL27*	C/T	1.17
rs34670647	16	30171017	*Uncertain*	D/I	1.23
rs9797244	17	26097131	*NOS2*	C/T	1.14
rs2779255	17	26137540	*NOS2*	C/T	1.11
rs12943464	17	45690351	*NPEPPS*	A/T	1.15
rs196941	17	62147192	*ERN1/ICAM2*	G/C	1.14
rs9977672	21	40463283	*21q22 Intergenic*	A/G	0.83
rs1569414	22	45727565	*FAM118A*	G/T	0.87
rs6426833	1	20171860	*Uncertain*	G/A	1.05
rs7517847	1	67681669	*IL23R*	G/T	0.93
rs183686347	1	67703442	*IL23R*	A/G	1.78
rs4845604	1	151801680	*RORC*	A/G	1.08
rs4971079	1	155130391	*Uncertain*	A/G	1.076
rs1333062	1	160846284	*LY9/CD244*	T/G	0.91
rs10800314	1	161472789	*FCGR2A*	C/A	1.07
rs61802846	1	161473873	*FCGR2A*	C/A	1.11
rs3024493	1	206943968	*IL10/IL19*	A/C	1.1
rs12075255	1	206961628	*IL10/IL19*	A/G	0.93
rs13407913	2	25097644	*Uncertain*	G/A	1.06
rs201014116	2	25500905	*DNMT3A*	C/A	1.07
rs72871627	2	163136942	*IFIH1*	G/A	1.28
rs7426056	2	204612058	*CD28/CTLA4*	A/G	1.06
rs11676348	2	219010146	*CXCR2*	T/C	1.06
rs12694846	2	231148128	*SP140*	G/A	1.08
rs4676406	2	241579108	*GPR35*	G/T	0.94
rs1001007	3	46428658	*CCR5*	A/G	0.94
rs3197999	3	49721532	*MST1*	A/G	1.09
rs1992661	5	40414989	*PTGER4*	G/A	0.93
rs9687958	5	40496423	*PTGER4*	G/T	0.95
rs71624119	5	55440730	*ANKRD55*	A/G	0.94
rs17622517	5	131803537	*C5orf56*	C/T	1.15

(Table continued on next page)

(Table continued from previous page)

rs11749391	5	150229066	*IRGM*	C/T	1.13
rs1267499	6	14715882	*6p23 Intergenic*	T/C	0.92
rs72928038	6	90976768	*BACH2*	A/G	1.08
rs582757	6	138197824	*TNFAIP3*	C/T	1.06
rs2451258	6	159506600	*TAGAP*	C/T	1.09
rs2301436	6	167437988	*FGFR1OP*	T/C	0.91
rs12718244	7	50175654	*C7orf72*	A/G	1.05
rs4917129	7	50323174	*IKZF1*	T/C	0.95
rs4728142	7	128573967	*IRF5*	A/G	1.07
rs10758669	9	4981602	*JAK2*	C/A	1.06
rs2812378	9	34710260	*CCL21*	G/A	1.09
rs726657	9	117696336	*TNFSF8*	T/C	0.92
rs4986790	9	120475302	*TLR4*	G/A	0.9
rs141992399	9	139259592	*CARD9*	G/C	0.48
rs10870077	9	139263891	*CARD9*	G/C	1.09
rs3124998	9	139389432	*NOTCH1*	T/C	1.21
rs2236379	10	6527143	*PRKCQ*	A/G	1.07
rs10761648	10	64354262	*ZNF365*	T/C	1.09
rs7915475	10	64381668	*ZNF365*	G/A	0.93
rs11190133	10	101278725	*NKX2-3*	T/C	0.88
rs10750899	11	58284951	*Uncertain*	G/A	1.23
rs11236797	11	76299649	*Uncertain*	A/C	1.105
rs11221322	11	128346793	*ETS1*	C/T	1.14
rs11221332	11	128380974	*ETS1*	T/C	1.09
rs12369214	12	107198611	*Uncertain*	A/G	0.95
rs3184504	12	111884608	*SH2B3*	T/C	0.92
rs8006884	14	35563211	*Uncertain*	C/T	1.1
rs2145623	14	35839236	*NFKBIA*	C/G	0.94
rs1569328	14	75741751	*FOS*	T/C	0.94
rs35874463	15	67457698	*SMAD3*	G/A	1.13
rs367569	16	11365500	*TNP2*	T/C	0.95
rs11574938	16	30485393	*ITGAL*	G/C	1.08
rs2066845	16	50756540	*NOD2*	C/G	0.78
rs72796367	16	50762771	*NOD2*	C/T	0.79
rs5743293	16	50763781	*NOD2*	I/D	0.82
rs9889296	17	32570547	*CCL2/7/8/11*	A/G	0.93
rs35736272	17	38032680	*ZPBP2*	C/T	1.07
rs1292035	17	57989557	*Uncertain*	A/G	1.1
rs12968719	18	12879466	*PTPN2*	A/G	1.12
rs74956615	19	10427721	*RAVER1*	A/T	0.81
rs35018800	19	10464843	*TYK2*	A/G	0.6
rs12720356	19	10469975	*TYK2*	C/A	1.09
rs587259	19	34656406	*LSM14*	T/C	0.95
rs679574	19	49206108	*FUT2*	G/C	1.07
rs4243971	20	30849517	*Uncertain*	T/G	0.92
rs6058869	20	31348750	*DNMT3B*	T/C	1.09
rs2823288	21	16820888	*21q21 Intergenic*	A/G	1.06
rs4456788	21	45616324	*ICOSLG/DNMT3L*	G/A	1.08
rs2266961	22	21928597	*UBE2L3*	G/C	1.1
rs2143178	22	39660829	*PDFGFB*	C/T	0.9

Table 1.3. Comprehensive list of AS-associated GWAS SNPs within European disease cohorts. Table above shows the complete list of all known genome-wide significant (p-value $\geq 5x10^{-8}$) AS-genetic associations from the Cortes *et al.* (2013) and Ellingaus *et al.* (2016) studies. Column one shows the lead SNP from each study, with its chromosome and base pair position in the Grhc37 human genome assembly highlighted in columns two and three. The closest gene to the lead SNP is shown in the fourth column. The possible alleles for each SNP are shown in the fifth column, with the incidence rate odds ratio (OR) the A1 major allele in AS populations shown in the sixth column. Table has been adapted from Brown & Wordsworth (2017).

Interestingly, novel diagnosis techniques are being tested using the results of AS GWAS. Many studies have begun to screen for genes that are both differentially expressed and possess differential SNP signatures within AS. For example, the gene encoding eukaryotic elongation factor 1 epsilon 1 – which is involved in the synthesis of aminoacyl t-RNA molecules – demonstrates significant upregulation in protein expression within whole-blood samples and skewing of the associated SNPs *rs7763907* and *rs7751386* within AS versus healthy controls (Fan *et al.*, 2019). Additionally, the SNPs *rs767455* and *rs1061622* of the respective *TNFRSF1A* and *TNFRSF1B* (tumour necrosis factor superfamily member 1A and 1B) genes – encoding receptors of the pro-inflammatory tumour necrosis factor alpha (TNFα) ligand – were found to diverge within AS patients that demonstrated higher initial BASDAI scores (Bugaj *et al.*, 2022). The same study also found that the G/G haplotype of *rs1061622* associated with more severe enthesitis in AS patients, whilst the A/A haplotype of the *ERAP1* – a gene important in antigen and epitope processing – SNP *rs2287987* could accurately predict comorbidity with uveitis. These studies suggest the potential superiority of genetic screening at diagnosing and predicting AS outcomes over traditional biomarkers, medical imaging, and severity scores, and encourages further study of the genetics of AS with GWAS-based approaches.

1.4. Pathophysiology of ankylosing spondylitis

1.4.1. MHC class I and class II molecules and T cells in ankylosing spondylitis

A fundamental constituent of the immune response is the presentation of peptide antigens bound to MHC molecules to mediatory T-cells (McHugh & Bowness, 2012; Halpert *et al.*, 2020). MHC molecules can be split into class I and class II. Generally, MHC class I molecules can be found on the surface of most nucleated cells, and typically present antigens to cytotoxic T-cells through the CD8 receptor for subsequent killing of

pathogen-infected or cancerous cells (Neefjes *et al.*, 2011; Bruno *et al.*, 2023). In contrast, MHC class II molecules are expressed exclusively on antigen-presenting cells of the immune system, such as dendritic cells, B cells, or macrophages, and present antigens to T-helper cells bearing the CD4 receptor to help propagate the immune response (Neefjes *et al.*, 2011; Hos *et al.*, 2022). Dysfunction of either of the signalling events of class I or class II MHC molecules is considered intrinsically linked with multiple inflammatory conditions. The enrichment for both MHC class I (such as HLA-B*27:05 and HLA-B*40:01) and class II antigens (such as HLA-DR*01:03) in AS GWAS suggests an important role for MHC antigens in AS pathogenesis (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016).

1.4.2. The MHC class I molecule HLA-B27 and ankylosing spondylitis

As previously highlighted, the MHC class I antigen HLA-B27 demonstrates the highest association with AS, with 80% to 95% of AS cases showing positivity for HLA-B27 (Schlosstein *et al.*, 1973). Moreover, HLA-B27 is relatively common among European and Han Chinese populations, which probably contributes to the relatively high prevalence of AS in these populations (Brown *et al.*, 1996; Bowness, 2015; Luo *et al.*, 2019; Wu *et al.*, 2021). Exactly how HLA-B27 predisposes an individual towards AS however is unclear, as not all HLA-B27 positive individuals will develop the disease. It is predicted that as few as 5-6% of all HLA-B27 positive individuals will be diagnosed with AS during their lifetime (Reveille *et al.*, 2013).

HLA-B27 itself is highly polymorphic, with over one hundred allotypes of the antigen. These allotypes vary in their association (Bowness, 2015). HLA-B*2702 and HLA-B*2705 are the most prevalent allotypes within European populations and show the greatest degree of AS association (Brown *et al.*, 1996; Bowness, 2015). In contrast, Han

Chinese HLA typing studies demonstrate that the allotype HLA-B*2704 is the most common, with a minority of cases presenting with HLA-B*2705. The respective rates for each allotype represents roughly 85% and 12.6% of Chinese AS patients (Luo *et al.*, 2019). Interestingly, the manifestation of multiple inflammatory comorbidities with AS could be attributed to specific HLA-B27 allotypes. Chinese AS patients positive for HLA-B*2704, for instance, may show greater incidence of uveitis compared to patients presenting with HLA-B*2705 (Li *et al.*, 2018).

A study comparing 7500 peptide antigens presented by the eight most common HLA-B27 allotypes – HLA-B*2702 to HLA-B*2709 – demonstrated that they produced similar yet unique signals in their respective antigen binding repertoires and differed quantitatively in their absolute peptide loads (Schittenhelm *et al.*, 2014). These differences could be attributed to non-synonymous genetic variation in a relatively small set of eight to ten amino acid residues in the binding cleft of HLA-B27, and may suggest that the allotypes predispose carriers to AS through several differing, yet related, molecular processes. Considering the above, HLA-B27 contributions to AS pathogenesis remain difficult to elucidate. A number of models have however been proposed.

1.4.2.1. The arthritogenic peptide hypothesis

The arthritogenic peptide hypothesis was the first model constructed to help explain HLA-B27-associated AS pathophysiology. The model proposes that HLA-B27-specific responses could be driven through auto-recognition and cross-reactivity of unique peptide-MHC complexes. These complexes could structurally resemble peptides that are targeted by autoreactive $CD8^+$ cytotoxic T-cells (Benjamin & Parham, 1990). Included amongst these could be peptides derived from pathogenic bacteria that under native contexts initiate an inflammatory immune response. Whilst the number of allotypes of

HLA-B27 that bind differential peptides could support this theory, Schittenhelm *et al.* (2014) failed to uncover any peptides in their repertoire analysis that could be considered arthritogenic.

This model was reconsidered in 2017, when the first evidence of clonally expanded TRVB9-CDR3β-TRBJ2.3 $CD8^+$ T-cells were found enriched in the synovium and eyes of patients with AS uveitis respectively (Faham *et al.*, 2017; Komech *et al.*, 2018; Hanson *et al.*, 2020). Yang *et al.* (2022) identified a motif in the BV9-CDR3β orphan T-cell receptor shared by both self and microbial antigens – such as on the bacterial UPF0324 inner membrane protein $YeiH^{bac}$ – that upon binding to BV9-CD3Rβ could stimulate inflammatory responses in $CD8^+$ T-cells. Activated $CD8^+$ T cells in AS have been shown to release chemokine attractants for other immune cells, release higher amounts of the digestive granzyme, and promote the upregulation of osteogenic genes (Martini *et al.*, 2024). Both the specific antigens and the motif were found to be conserved or similar within previously AS-linked dysbiotic pathogens such as *K. pneumoniae* (Rashid *et al.*, 2013).

These responses were only seen however when antigens containing the motif were presented on cells expressing HLA-B27*05, as opposed to the neutral or protective HLA-B27*09 allotype. Crystal structures HLA-B*2705 in contrast to HLA-B*2709 explained that differences in allotypic antigen binding and cell surface presentation could be influenced by factors such as steric hinderance. This arises due to positively charged polar residues lying near the antigen binding cleft of HLA-B*2709 as opposed to B*2705. The arthritogenic model therefore could explain the autoimmune nature of the disease and why only certain allotypes associate with AS.

1.4.2.2. The unfolded protein response hypothesis

The extracellular component of MHC class I molecules including HLA-B27 are assembled from heterotrimers consisting of an MHC class I heavy chain, a light chain beta-2 microglobulin (β2m), and a unique cell surface antigen peptide for presentation to immune cells. The heavy chain is composed of the three alpha-helical structural domains α1, α2, and α3. The domains α1 and α2 form the antigen binding groove. In contrast, α3 anchors the molecule to the transmembrane domain, and non-covalently associates with β2m whilst the peptides fold within the endoplasmic reticulum (ER) to stabilise tertiary structure. The unfolded protein response hypothesis proposes that the folding of HLA-B27 heavy chains is abnormal in regards to specific allotypes, resulting in the trafficking of empty HLA-B27 heavy chains lacking both β2m and antigen to the cell surface. Previous studies have demonstrated that HLA-B27 has a natural propensity to misfold, with the AS-associated HLA-B*2702, HLA-B*2704, and HLA-B*2705 in particular being slower to both fold and associate with β2m in the ER compared to other HLA-B27 allotypes (Mear *et al.*, 1999; Antoniou *et al.*, 2004; Merino *et al.*, 2008; Lenart *et al.*, 2022).

Lenart *et al.* (2022) found that residues 74, 77, and 116 of the F pocket – part of antigen-binding groove of HLA-B27 – could define the slower folding response of HLA-B*2705. These residues within HLA-B*2705 are filled by the negatively charged polar aspartic acid, and results in a significantly more negatively charged F pocket compared to other HLA-B27 allotypes. This could lead to the hydrophilic destabilisation of the α1 and α2 alpha-helical domains and the tertiary structure of the antigen binding groove. It had also been found early on that the shape of α1 and α2 facilitated interaction of α3 with β2m (Luthra-Guptasarma & Singh, 2004). Changes in confirmation to α1 and α2 could lead to increased dissociation of β2m from α3. The slow folding kinetics of AS-associated HLA-

B27 allotypes leads to accumulation of unfolded or misfolded peptides in the ER, and can trigger both autophagy and two stress responses in the ER-overload response (EOR) and the unfolded protein response (UPR). These responses can lead to the activation of nuclear factor kappa beta (NFκβ) pathways and the release of extracellular pro-inflammatory cytokines such as TNFα, IL-6, and IL-23 which contribute to the inflammatory mechanisms of T cells and macrophages (Smith, 2018; Raziaemanesh *et al.*, 2023). Interestingly, it was also found that misfolded HLA-B*2705 increased with *S. enterica* inoculation, which suggested infection with gut microbes could exacerbate the UPR immune response (Antoniou *et al.*, 2019; Berland *et al.*, 2023).

1.4.2.3. The homodimer hypothesis

Allen *et al.* (1999) found that HLA-B27 heavy chains can naturally form homodimeric complexes centred on cysteine 67 in the extracellular domain of the peptide. They demonstrated that the rate of homodimerization accelerates in the absence of β2m, suggesting that slower folding kinetics in relation to HLA-B27 accessory peptides can lead to a greater proportion of HLA-B27 molecules self-associating. HLA-B27 homodimers can be recognised by killer immunoglobulin-like (KIR) receptors on the surface of effector immune cells such as T cells and natural killer (NK) cells, and pathogenic allotypes such as HLA-B*2705 stimulate KIRs more strongly than protective allotypes like HLA-B*2709 (Kollnberger *et al.*, 2007; Cauli *et al.*, 2013). $CD4^+$ T cells positive for the KIR family member KIR3DL2 in particular are found enriched in HLA-B27 positive AS patients (Bowness *et al.*, 2011; Ridley *et al.*, 2016).

1.4.2.4. HLA-B27 homodimers and the UPR contribute to Th17 cell expansion

$CD4^+$ T cells can differentiate into a number of T helper subtypes. The main contributing subtype in AS are T-helper 17 (Th17) cells, which are the main producers of the pro-

inflammatory cytokine IL-17. The expression of Th17-derived IL-17 promotes the production of other pro-inflammatory cytokines and chemokines, perpetuates inflammatory signalling cascades, and contributes to the tissue damaging and bone remodelling characteristic of AS (Chisălău *et al.*, 2020). In AS, elevated levels of IL-17 have been observed in affected tissues including the spine and sacroiliac joints (DeLay *et al.*, 2009; Appel *et al.*, 2011). IL-17 is also upregulated in sacroiliac joints of diffuse idiopathic skeletal hyperostosis patients, where similar to AS the growth of bone spurs results in tendon and enthesis hardening (He *et al.*, 2023). These elevated levels correlate with disease activity and severity as measured by BASDAI, suggesting a direct link between IL-17 and inflammation in AS pathogenesis (Chen *et al.*, 2013; Quiero *et al.*, 2023). Both HLA-B27 homodimers and the UPR are believed to contribute to Th17 differentiation and the heightened expression of IL-17 in AS patients (**Fig. 1.2**).

KIR3DL2 binding to HLA-B27 activates the downstream master regulator and transcription factor RORγ (retinoic acid receptor-related orphan nuclear receptor γt), which coordinates reprogramming of gene transcription networks in naïve $CD4^+$ T cells to the Th17 phenotype (Ridley *et al.*, 2016). Similarly, IL-23 produced by the UPR in response to misfolded HLA-B27 is crucial in the expansion of Th17 cells and promotes their tissue longevity through IL-23R signalling (McGeachy & Cua, 2007; Revu *et al.*, 2018). It has also been found that a number of other, non-T cell enthesis-resident cell types bearing IL-23R could initiate enthesitis alone through genetic manipulation of the IL-23 haplotype in murine models (Sherlock *et al.*, 2012; Bridgewood *et al.*, 2020). These IL23R responses are a tissue response to inflammation that is particularly relevant to the protection of epithelial barriers, such as what occurs in in the skin and intestinal mucosa. The main contribution of HLA-B27 is therefore the constitutive activation of $CD8^+$ cytotoxic T cells with arthritogenic peptides and the production of IL-17

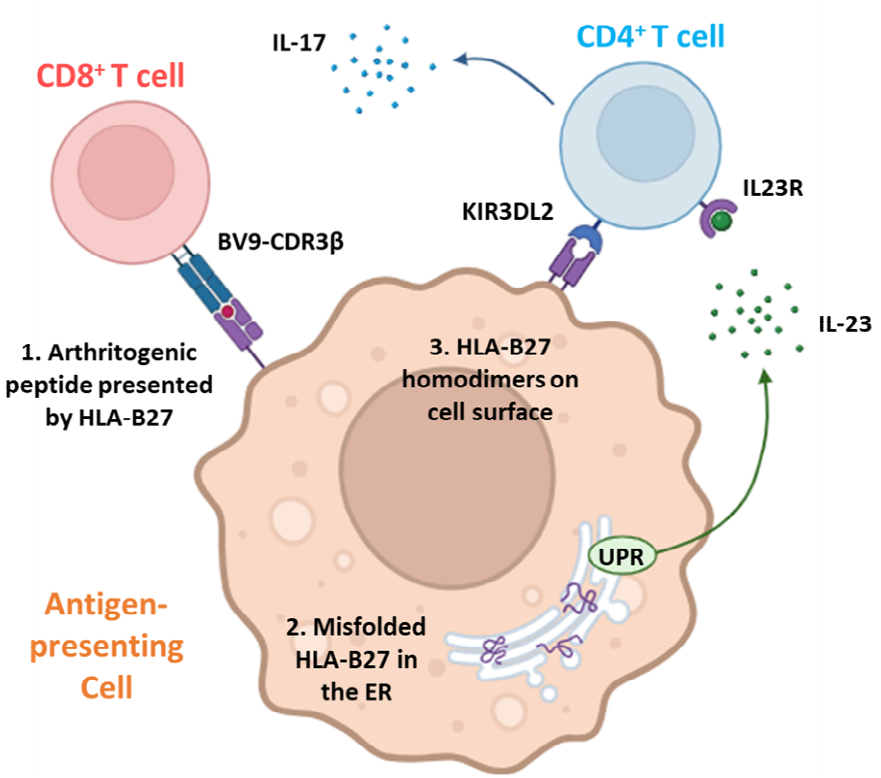

Fig. 1.2. HLA-B27 contributes to ankylosing spondylitis by upregulating the IL-23/IL-17 signalling axis in CD4$^+$ and CD8$^+$ T cells through three models. Figure above shows a summary of the three HLA-B27 disease models and how they can contribute towards signalling dysfunction in T cells. (1) HLA-B27 presents a host-derived arthritogenic peptide for which it has increased affinity for that mimics pathogenic and/or microbial peptides. This peptide is recognised by TCRs such as BV9-CDR3β in CD8$^+$ cytotoxic T cells that leads to their constitutive activation. Activated CD8$^+$ T cells contribute to osteogenic gene upregulation and release both more cell-killing associated enzymes and immune chemoattractants. (2) Misfolded HLA-B27 molecules in the ER trigger the UPR in the antigen-presenting cell, which activates pro-inflammatory cascades and results in the release of interleukins such as IL-23. (3) HLA-B27 homodimers are recognised by KIRs on CD4$^+$ T cells such as KIR3DL2 that triggers the Th17 differentiation through the transcription factor RORγt Binding of a CD4$^+$ T cell to KIR3DL2 and receipt of IL-23 signalling from the UPR results in differentiation into Th17 cells, which releases the main pro-inflammatory cytokine IL-17. IL-17 is thought to have a leading role in AS pathogenesis. Figure has been adapter from Navid *et al.* (2021) using BioRender™ (http://www.biorender.com).

producing Th17 cells from $CD4^+$ T-helper cells through misfolded proteins and KIR-homodimer signalling.

1.4.3. Non-HLA-B27 MHC contributions to ankylosing spondylitis

A number of other MHC class I or class II antigen loci have been associated with AS outside of HLA-B27. These include MHC class I HLA-A and HLA-G genes which – much like HLA-B27 – are linked to the activation of T cells and NK cells through KIR receptor families (Lynch *et al.*, 2009; Reveille *et al.*, 2014; Reveille *et al.*, 2020). Many of the GWAS genetic variants within these loci have been found to associate with AS independently of HLA-B27 positivity. Indeed, the SNP *rs2394250* within the HLA-A0201 locus, for example, is found enriched within both $HLA\text{-}B27^+$ and $HLA\text{-}B27^-$ instances of AS (Cortes *et al.*, 2013). The functional mechanisms behind how these regions could contribute towards AS pathogenesis however remain largely uncharacterised, and their future study will yield important insights into how $HLA\text{-}B27^-$ individuals can still develop AS.

1.4.4. The contribution of monocytes and macrophages to ankylosing spondylitis

Whilst the association of T cells and AS is well characterised, other immune cells are believed to contribute to AS. Monocytes are precursor cells of the myeloid lineage that can differentiate into several different cell types such as macrophages and dendritic cells. Monocytes can be isolated from peripheral blood by their expression of cluster of differentiation 14 (CD14), which binds to pathogen-derived molecules such as the bacterial lipopolysaccharide (LPS) and stimulates innate immune system pattern-recognition receptors (PRRs) such as Toll-like receptors (TLRs) (Tsukamoto *et al.*, 2018). These PRRs have a role in establishing and propagating sterile inflammation at the site of infection (Chen & Nuñez, 2010).

Monocyte-derived cells are able to ingest and phagocytose invading infectious agents aided by receptors such as PRRs, and present pathogen-derived peptides on their cell surface for recognition by other immune cell types such as T cells (Blander & Medzhitov, 2004; McCLure & Massari, 2014; Hulme *et al.*, 2021). Monocyte-derived macrophages are thought to have a role in AS pathogenesis. These macrophages differentiate from peripheral blood monocytes at the site of infection or inflammation on stimulation with immunogenic compounds such as LPS, or cytokine signals such as interleukin-10 (IL-10) (Martinez & Gordon, 2014).

Macrophages can further differentiate into osteoclasts upon receiving IL-17 signals from Th17 cells, which are involved in the resorption and remodelling of bone by releasing extracellular acidic compounds to break down the mineral components of bone (Song *et al.*, 2019; Taubmann *et al.*, 2020). These cells are often found decreased in cases of AIS, with more severe depletions in osteoclasts observed more often in male cases of disease that show higher activity (Caparbo *et al.*, 2018). These studies suggest depressed rates of bone resorption through osteoclasts may contribute to progressive ankylosis of the vertebral column.

Monocyte-derived cells can adopt a number of pro or anti-inflammatory related states that have relevance towards disease (Chen *et al.*, 2023). Macrophages, for instance, will shift from the inflammation-associated M1 phenotype to the immunosuppressive M2 state upon AICAR-adenosine signalling, which is a target for DMARD therapy (Cronstein & Sitovsky, 2017). This shift is also associated with the production of fewer pro-inflammatory cytokines and inhibition of osteoclast differentiation. Macrophages are present in most tissues of the body, although significant differences exist between tissue-resident macrophages and monocyte-derived macrophages (Wang *et al.*, 2020). The major contribution of monocyte-derived macrophages towards AS however is thought to

be the release of pro-inflammatory cytokines. Indeed, monocyte-derived macrophages isolated from AS patients have been shown to produce more M1-like phenotypes and release higher concentrations of TNFα, IL-23, and IL-1β than those derived from healthy controls (Akhtari *et al.*, 2021). Alongside IL-23, IL-1β is essential for Th17 differentiation, and increased expression may lead to expansion and faster maturation of naïve $CD4^+$ T helper cells into Th17 cells (Brescia *et al.*, 2024).

1.4.5. The contribution of B cells to ankylosing spondylitis

The primary role of B cells is the secretion of antibodies which associate with antigens on the surface of pathogens that allow for their recognition by phagocytic cells such as macrophages through the Fc receptor (Suchanek *et al.*, 2023). This results in the endocytosis of the tagged infectious agent for autophagic destruction. B cells are usually defined from within the peripheral blood by their expression of B-lymphocyte antigen cluster of differentiation 19 (CD19), a transmembrane protein that acts to recruit co-stimulatory proteins to the plasma membrane of B cells (Depoil *et al.*, 2008). B-cells have an emerging role in AS, with a number of autoreactive and non-autoreactive antibodies now implicated in the disease (Wilbrink[2] *et al.*, 2021). In order to produce antibodies, naïve B cells must first differentiate into plasmablasts, with the population of the latter cells found at significantly higher frequency within AS patients compared to healthy controls (Wilbrink[1] *et al.*, 2021).

B cells are an important target of Th17 cells, with IL-17 and interleukin-21 (IL-21) having important modulatory roles. The rate of B cell tissue-infiltration, proliferation, and differentiation to plasmablasts increases with increasing IL-17 levels (Halwani *et al.*, 2014; Pitzalis *et al.*, 2014; Subbarayal et al., 2016). The levels of IL-21, which is also important in the formation of plasmablasts, are also significantly higher within AS

patients (Rasmussen *et al.*, 2010; Xiao *et al.*, 2013). The levels of Th17 cells and plasmablasts are therefore proportionate. Indeed, methotrexate-induced reduction of Th17 cells can lead to decreases in the number of circulating plasmablasts (Nived *et al.*, 2021). Mature plasmablasts also release IL-10, which in addition to signalling for macrophage differentiation also inhibits the function of T cells as infection resolves (Matsumoto *et al.*, 2014; Zhao *et al.*, 2024). In AS, increasing levels of circulating $CD19^+$ B cells are inversely proportional to IL-10 levels, suggesting irregularities in how plasmablasts immunosuppress T cells (Chen *et al.*, 2015).

1.4.6. Other implicated cell types in ankylosing spondylitis

The cells discussed above have the strongest evidence for pathogenic roles within AS. Other cell types such as NK cells and neutrophils do however show clonal expansion in AS (Xu *et al.*, 2020; Ren *et al.*, 2022). Within humans, neutrophils are the most abundant immune cell type and are an important constituent of the innate immune response. Neutrophils ordinarily release neutrophil extracellular traps (NETS) in response to IL-8 or LPS signalling (Pieterse *et al.*, 2016; An *et al.*, 2019). These NETs are web-like structures that consist of neutrophil DNA and granule proteins that act to both aggregate microbes for efficient endocytosis by phagocytic cells, and kill pathogens with the digestive anti-microbial properties of the granules (Papayannopoulos, 2017). These NETs can also deliver pro-inflammatory cytokine loads, with neutrophil NETs in AS patients showing higher enrichment for IL-17 and IL-1β than healthy controls (Papagoras *et al.*, 2020). NETs also aid the differentiation of Th17 cells from naïve $CD4^+$ T cells, which can result in a positive inflammatory recruitment feedback loop of neutrophils to the site of infection through the Th17 produced IL-8 (Pelletier *et al.*, 2010; Wilson *et al.*, 2022).

Like neutrophils, NK cells are innate immune cells that work to kill pathogenic cells with digestive enzymes. Instead of NET granules however, NK cells achieve cell lysis by secreting granzyme and perforin upon recognition of pathogens or infected cells (Prager *et al.*, 2019). NK cells have been found within AS to activate fibroblast-like and dendritic cells to produce the pro-inflammatory cytokines interleukin-12 (IL-12) and interferon gamma (IFNγ) (Conigliaro *et al.*, 2011; Yang *et al.*, 2019). These cytokines play important roles in the skewed differentiation of $CD4^+$ T helper cells into pro-inflammatory T helper 1 (Th1) over immunosuppressive T helper 2 (Th2) cells, with the ratio of Th1 to Th2 cells demonstrating significant increase in AS (Yang *et al.*, 2007; Zhu *et al.*, 2012; Liu *et al.*, 2021). NK cell infiltration to the synovium of RA patients is also associated with higher degrees of bone damage and joint inflammation (Wu *et al.*, 2022).

1.4.7. The gut microbiome and ankylosing spondylitis

AS and IBD demonstrate concordance in about 5% to 10% of AS patients (Rudalweit & Baeten, 2006). The number of AS patients demonstrating broader gastrointestinal issues however is between 70% to 80% (Sternes *et al.*, 2022). Significant dysbiosis of the bacterial populations of the gut microbiome towards more pro-inflammatory species is seen in AS patients, with AS linked in to the skewing of *K. pneuomoniae*, *E. coli*, and *Bacteroides* within the intestine (Garcia-Montoya *et al.*, 2018; Cardoneanu *et al.*, 2021). Th17 cells expressing the gut-homing C-C chemokine receptor type 9 (CCR9) show enrichment within the intestinal wall of AS patients, suggesting widespread inflammation of the gut (Gagliani *et al.*, 2015; Ridley *et al.*, 2016).

Within AS-mimicking mouse models, treatment with indole-3 acetic acid (IAA) – which activates the aryl hydrocarbon receptor (AHR), itself an AS-associated GWAS risk locus

– decreases the population of several microbial species and promotes transdifferentiation of the enriched Th17 cells into immunosuppressive T-regulatory cells (Tregs) and alleviates AS symptoms (Shen *et al.*, 2022). This is intriguing, as HLA-B27^{+} rats raised in sterile environments will not develop symptoms of AS (Taurog *et al.*, 1994). Dysbiosis of specific pathogenic bacteria has therefore been proposed as the initial trigger for the chronic inflammation seen within AS, but how these particular species predispose the immune system to drive AS is not well understood.

1.4.8. Non-MHC AS-associated genes and ankylosing spondylitis pathogenesis

The majority of genome-wide genetic variants within AS are found outside of the MHC region on chromosome 6. Whilst the vast majority of these variants remain functionally uncharacterised, a number of broad mechanisms that predispose towards AS have been proposed, including antigen peptide processing, intracellular signalling cascades, and immune cell development. A brief description of some of the best characterised AS-associated risk loci and their functions are provided below.

1.4.8.1. *ERAP1* and *ERAP2*: MHC class I processing

AS GWAS have revealed a strong association with M1 aminopeptidases such as *ERAP1* and *ERAP2* (endoplasmic reticulum aminopeptidase 1 and 2) (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016). With a genome-wide significant p-value of approximately $1x10^{-50}$, *ERAP1* demonstrates one of the highest associations with AS (Burton *et al.*, 2007; Harvey *et al.*, 2009; Evans *et al.*, 2011; Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016). The *ERAP1* gene encodes for an ER-located aminopeptidase, and is important in the shaping the immune response by cleaving peptides from the N-terminal end to an optimal size of eight to nine residues for association with MHC class I molecules, including HLA-B27 (Chang *et al.*, 2005). The highly polymorphic nature of *ERAP1* – which currently

demonstrates twelve known SNP variants within its 99% credible set (van de Bunt *et al.*, 2015) – suggests aberrant peptide cleaving and presentation could form an important pathogenic mechanism within AS. As many as eight of these variants are non-synonymous, and directly influence the tertiary structure of ERAP1 (Harvey *et al.*, 2009). Interestingly, the association between *ERAP1* and AS was only seen within HLA-B27 positive AS patients, suggesting synergistic effects between pathogenic antigen processing and HLA-B27 (Evans *et al.*, 2011). An independent association with *ERAP2*, which lies close to *ERAP1* on chromosome 5 but occupies a separate LD-block, is also associated with AS. Unlike *ERAP1*, *ERAP2* is associated with both HLA-B27 positive and negative cases (Robinson *et al.*, 2015). Altering *ERAP1* or *ERAP2* expression has profound impact on the peptide repertoires associated with MHC class I molecules, including HLA-B27 (Chen *et al.*, 2014; López de Castro, 2018). *ERAP1* polymorphisms that are protective against AS are loss-of-function mutations that reduce the enzymatic activity and are likely involved in influencing the antigen peptide repertoire (Harvey *et al.*, 2009; Kochan *et al.*, 2011; Roberts *et al.*, 2017). Because *ERAP1* and *ERAP2* are so heavily involved in which MHC class I antigens are presented to cells of the immune system, there is scope development of small molecule inhibitors of *ERAP* functionality for use in the therapeutic intervention of AS.

1.4.8.2. *IL-23R*: Modifying the IL-17/IL-23 signalling axis

Abnormalities in the IL-23 signalling cascades responsible for Th17 cell expansion and survival are associated not only with AS, but for a number of related inflammatory conditions such as IBD, psoriasis, and psoriatic arthritis (Salas *et al.*, 2020; Chen *et al.*, 2020). Activation of IL-23R is associated with the coactivation of the downstream intracellular janus kinase (JAK) and signal transducer and activator of transcription (STAT) pathways. These JAK/STAT pathways rely on a number of the AS-associated

genes to propagate the signal cascade and alter gene expression, including *JAK2*, non-receptor tyrosine-protein kinase (*TKY2*), and *STAT3* (Deng *et al.*, 2022; McInnes *et al.*, 2022). *STAT3* in particular is another driver of Th17 cell differentiation, with studies preventing the translocation of the transcription factor upon IL-23 stimulation from the cytoplasm into the nucleus alleviating symptoms of IBD in mouse models (Zhang *et al.*, 2020). Other mouse model studies have shown that IL-23R deficiency and overexpression of IL-23 can result in SpA-like enthesitis and reductions in bone mass through osteoclast-driven mechanisms (Adamopoulos *et al.*, 2011; Sherlock *et al.*, 2021; Razawy *et al.*, 2021).

Within the *IL-23R* locus, the best characterised genetic variant is the protective non-synonymous 'A' allele of the SNP *rs11209026*. This allele inhibits IL-17 signalling cascades and Th17 cell maturation through altering IL-23R structure, and shows higher frequency in healthy controls (Di Meglio *et al.*, 2011; Deveci *et al.*, 2019). Associations independent of *rs11209026* have also been found, including the protective 'A' allele of *rs11209032*, which affects the DNA-binding of the transcription factor *TWIST1* (twist-related transcription factor 1) and is associated with Th1 cell differentiation (Roberts *et al.*, 2016; Roberts et al., 2017). Complex interplay between multiple regulatory SNPs in AS-associated genes is not limited to *IL-23R* however, with both studies by Cortes *et al.* (2013) and Ellinghaus *et al.* (2016) having revealed a number of loci with independent association signals of relevance.

1.4.8.3. *RUNX3*: Immune cell development

Run-related transcription factor 3 (RUNX3) is a transcription factor important in the differentiation of multiple different immune cell types, such as $CD4^+$ T helper cells, $CD8^+$ T cytotoxic killer cells, monocytes, NK cells, and B cells (Seo & Taniuchi, 2020). It is

particularly important in tissue-residency programs within $CD8^+$ T cells, where increased expression of *RUNX3* results in the upregulation of tumour growth factor β (TGF-β) (Li *et al.*, 2022). Ectopic expression of TGF-β and its signalling in mice has been found to increase the longevity and concentration $CD8^+$ T cells within epithelial tissues and enhance the potency of their targeted cell killing mechanisms (Fonseca *et al.*, 2022). Increased expression of *RUNX3* also downregulates other key developmental transcription factors within haematopoietic stem cell progenitor cells and inhibits maturation into cells of the myeloid lineage, favouring the formation of lymphocytic cells instead (Menezes *et al.*, 2022).

Much like *IL-23R*, *RUNX3* demonstrates multiple independent genetic associations at the SNPs *rs4648889* and *rs4265380***.** These SNPs seem to impart context-specific effects, with *rs4648889* impacting *RUNX3* expression within $CD8^+$ T cells, and *rs4265380* within $CD14^+$ monocytes (Vecellio *et al.*, 2018). Due to its importance in $CD8^+$ T cells, many studies have focused on the role of *rs4648889* within AS. The SNP has been shown to impact the DNA binding of the inflammation-associated transcription factor interferon regulatory factor 5 (IRF5), which preferentially binds to the AS-protective 'G' allele of *rs4648889* (Vecellio *et al.*, 2021). The disease-associated 'A' allele instead binds preferentially to the nucleosome remodelling and deacetylase (NuRD) complex which inhibits *RUNX3* expression. Vecellio *et al.* (2021) further demonstrated that *IRF5* knockdown resulted in the reduction of IFNγ expression, with RUNX3 having been shown to cooperate with the transcription factor T-box transcription factor TBX21 (T-bet) to modulate. Investigation of *RUNX3* in inflammatory disease is however still in the early phases of investigation, and further functional work needs to be undertaken to identify true targets for therapeutic intervention.

1.4.8.4. *TNFRSF1A*: TNFα signalling

TNFα is a ligand for a number of cell surface receptors such as TNFRSF1A or TNFRSF1B. Binding of TNFα to these receptors results in activation of the mammalian target of rapamycin (mTOR), mitogen-activated protein kinase (MAPK), and NF-κβ pathways which mediate inflammatory and apoptotic processes within a number of cells (Verzella *et al.*, 2022). TNFα signalling is also linked to osteoclastogenesis through its downstream signalling effector receptor activator of nuclear factor kappa-B ligand (RANKL), with expression of both TNFα and RANKL being necessary for osteoclast differentiation (Luo *et al.*, 2018*). TNFRSF1A* has been associated to AS through GWAS (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016).

The lead SNP *rs1860545* for *TNFRSF1A* is in LD with a number of SNPs associated with multiple sclerosis (MS) which cause the irregular splicing and removal of exon 6 from the final protein (Xing-Rong *et al.*, 2018). Exon 6 constitutes part of the transmembrane domain and solubilises the protein on deletion, resulting in significantly reduced TNFα-associated signalling (Gregory *et al.*, 2012). The AS-associated lead SNP does not however contribute to exon 6 removal in TNFRSF1A in AS, suggesting that *rs1860545* is not the causal SNP in AS or an alternative mechanism in the locus is responsible for AS predisposition (Watts *et al.*, 2019). Whilst anti-TNFα is clearly linked to AS through the propagation of immune responses, it is unclear how TNFα, its associated receptor molecules, and the SNPs within them can predispose and individual to AS.

1.5. Gene regulation, genetic variants, and disease

The previous section highlighted that AS is a complex disease involving the context-specific contribution of many genetic variants, genes and different cell types of the immune system. The impact of non-synonymous mutations in AS-associated GWAS risk

loci such as *ERAP1*, *ERAP2*, and *TNFRSF1A* can be through *in-silico* modelling of the amino-acid substitutions and the effects they have on protein structure. As highlighted by the *RUNX3* and *IL-23* GWAS loci, the functional implications behind non-coding variants are far broader and more nuanced, and not much is known about how the majority of non-coding variants identified in AS GWAS contribute to disease pathogenesis.

1.5.1. Non-coding genetic variants and eQTLs

Disease-associated non-coding genetic variants identified through GWAS almost always result either directly or indirectly in a change to the expression patterns of one or more associated genes (Lichou & Trynka, 2020). The most common method of measuring this impact is to directly assay the levels of RNA produced from the gene of interest and associate with possession of a genetic variant. This can be done on the single transcript level via quantitative real-time PCR (qRT-PCR), or on a transcriptome-wide level with RNA sequencing (RNA-seq)-based methodologies (Everaert *et al.*, 2017). Where complete genotyping and transcriptomic data are available for multiple individuals, expression quantitative trait loci (eQTLs) can be generated by estimating the average fold-change in the RNA expressed from specific genes based on the identity of the alleles in each individual, known as the allelic dose (Kerimov *et al.*, 2021). These eQTLs are therefore genetic variants that have validated functional effects on the differential expression of genes, and can be linked to disease in similar ways to GWAS variants. In immune cells, many disease-linked eQTLs are related to inflammatory processes. Indeed, eQTL mapping for the traits of IBD and vasculitis in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, neutrophils, and $CD19^+$ B cells show that eQTLs only present in active inflammation are abrogated in these conditions upon treatment with corticosteroids, the cytotoxic compound cyclophosphamide, and the $CD19^+$ B-cell depleting rituximab (Peters *et al.*, 2016).

1.5.2. *Cis*-regulatory elements in the non-coding genome

Functional genomic annotation of disease-associated genetic variants has revealed that genetic variants such as SNPs often lie within diverse regulatory elements (Fulco *et al.*, 2019; Ying *et al.*, 2023). Many of these regulatory elements are functional in specific cell types, stimulation conditions, or developmental stage. *Cis*-regulatory elements are so named as they act on gene promoters occupying the same chromosome as they are located. Hundreds of thousands of *cis*-regulatory elements exist within the human genome, and they vastly outnumber the even the highest estimates (>20,000) for the total number of coding regions (Dean *et al.*, 2021). This implies the presence of complex pleiotropic regulatory networks that govern the dynamic and context-specific expression of genes. Many types of cis-regulatory elements exist, but they can be characterised broadly as promoter, enhancer, silencers and insulator elements (Pang *et al.*, 2022).

1.5.3. The structure of *cis*-regulatory regions

Promoters are relatively short DNA sequences usually between 100 to 1000 base pairs in length adjacent to protein-coding regions. It is here that the transcription initiation complex – including RNA polymerase II and associated proteins of the pre-initiation complex – assembles and gene transcription begins (Butler *et al.*, 2002; Haberle & Stark, 2018) (**Fig. 1.3**). The spatial and temporal nature of transcription from the promoter is further modulated by the more distal enhancers, which themselves are relatively short stretches of DNA that demonstrate strong enrichment for binding sequence motifs of tissue-specific DNA-binding proteins known as transcription factors (Zboril *et al.*, 2021). Transcription factors bound to enhancer elements act to influence gene expression through the recruitment of the mediator complex, which promotes faster formation of the pre-initation complex and release of RNA polymerase II at the gene promoter. The

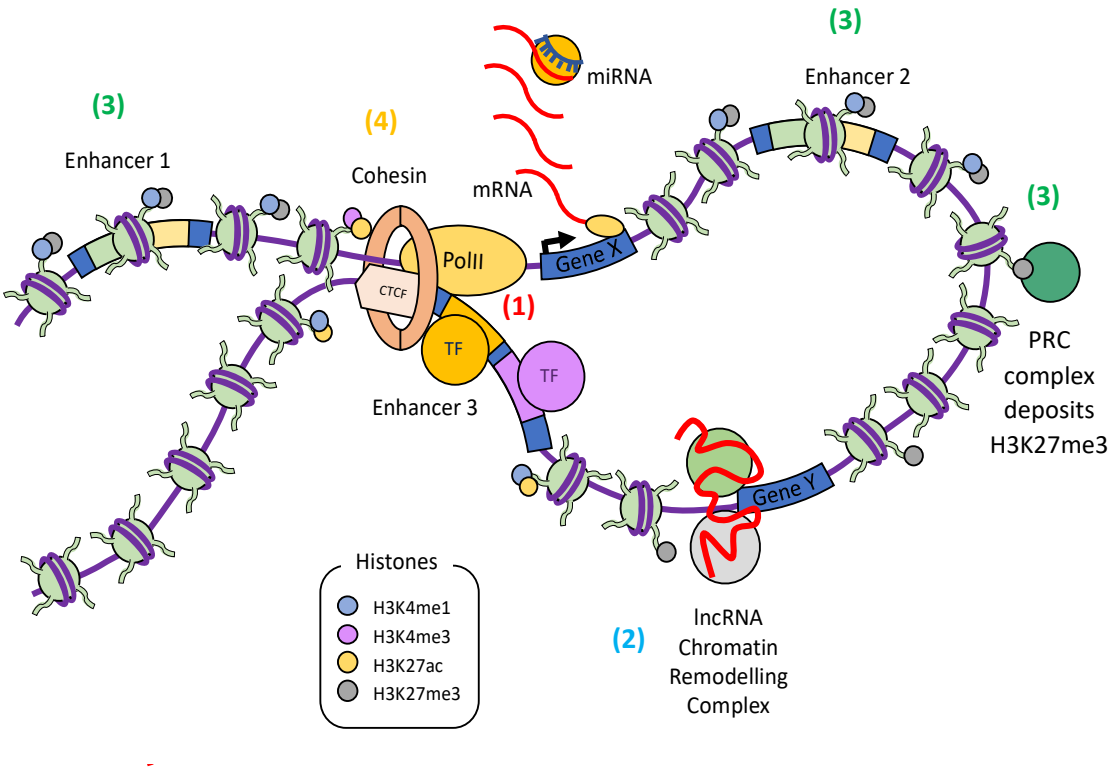

Figure 1.3. Schematic overview of regulatory interaction between enhancer and promoter elements and their histone post-translational modifications and associated proteins. Figure above shows the structure of a typical *cis*-regulatory interaction controlling gene expression. **(1)** Transcription factors (TFs) occupying enhancer elements loop through to the promoter element via the mediator complex (not shown) occupied by the enzyme RNA polymerase II (PolII) that transcribes the gene into RNA. Promoters are defined by the histone post-translation marks H3K4me3 and H3K27ac (section 1.5.3.2) and active enhancers by H3K4me1 and H3K27ac (section 1.5.3.3). **(2)** Long non-coding RNAs (lncRNA) facilitate the recruitment of chromatin remodelling complexes that can result in either the repression or activation of gene expression (section 1.5.3.4). **(3)** Context-specific yet reversible silencing of enhancer elements is achieved by exchanging H3K27ac for H3K27me3, which is catalysed by the Polycomb-repression complex (PRC) (section 1.5.3.5). **(4)** The interaction between promoter and active enhancer is stabilised by the architectural protein cohesin which is recruited by and interacts with CTCF-occupied insulators that define the boundaries of *cis*-regulatory interaction (section 1.5.3.7). Genetic variants that affect the function of many of these genetic elements can have putative pathogenic effects through the abnormal regulation of genes they control. Figure has been adapted from the thesis of Alicia Lledó Lara (Cell and tissue specific functional genomics of psoriasis and psoriatic arthritis, University of Oxford, 2019).

mediator complex demonstrates a high degree of conformational allostery, binding a wide range of coenzymatic proteins that influence how it alters gene transcription (Quievado *et al.*, 2019; Richter *et al.*, 2022).

Mediator complexes can therefore adopt different regulatory conformations depending on the identities of transcription factors bound to an enhancer region, allowing for the complex to differentially influence gene expression under distinct biological contexts including developmental time and environmental stimulation (Allen & Taatjes, 2015). Indeed, the mediator complex has been found to be the major end-point of cellular signalling cascades and the main determiner of the appropriate genomic response (van de Peppel *et al.*, 2005). In contrast to the generally fixed location of gene promoters, enhancers are often found at much more distant from the genes that they regulate and may show different states of activation or deactivation between tissues (Sanyal *et al.*, 2012). Compounded with the fact that these regions usually loop over great genomic distances to influence promoter activity, this can render identification of genes regulated by each individual enhancer challenging.

To understand why looping presents such a challenge in the identification of enhancers, ~~i~~t is necessary to understand how conformational changes within DNA occur. The total length of DNA within every cell would stretch for over 2 metres in length if left in its native conformation (Minchin & Lodge, 2019). DNA is wrapped around an octamer core of histone proteins – two each of the subunits H2A, H2B, H3 and H4 – which can then condense, coil, and supercoil into further higher order structures. This eventually terminates in the highest order structure of the chromosome. The resulting complex of DNA and its associated histones and proteins is termed chromatin. In the base chromatin state, these histone cores sterically hinder transcription factors from associating with both promoter and distal cis-regulatory elements, effectively preventing the expression of

genes condensed this way (Klemm *et al.*, 2019). The histone core can be made to associate more tightly or loosely with DNA through a number of covalent post-translational modifications to lysine residues to the tail domain of specific histone subunits.

1.5.3.1. Associated histone marks for promoter regions

Active promoter elements are usually marked by the respective acetylation and trimethylation of lysines 27 and 4 of histone 3 (H3K27ac and H3K4me3). H3K4me3 allows for the interaction of integrator complex subunit 11 (INTS11) with the pre-initiation complex, which is necessary for the latter to release RNA polymerase II and begin transcriptional elongation (Wang *et al.*, 2023). Whilst still under investigation, the H3K27ac mark has been shown to interact with the lysine acetylation-reading BRD family of bromodomain-containing chromatin remodelers to establish and maintain novel 3D chromatin contacts between promoters and cis-regulatory elements. H3K27ac hyperacetylation driven by the introduction of BRD family member BRD4-NUT fusion proteins has been shown to result in the formation of more extensive interaction domains (Alekseyenko *et al.*, 2015). Differences in H3K27ac deposition in $CD19^{+}$ B cells between healthy and diffuse large-cell lymphoma cancer patient cohorts have additionally been shown to contribute to the repositioning of multiple cis-regulatory loops that contribute towards the disease (Sungalee *et al.*, 2021). H3K27ac has also been shown to be important within promoters towards the correct positioning of the H3K4me3 mark around the TSS, with its presence necessary to initiate transcription. Introduction of an extraneous H3K27ac signal at promoters bearing the repression-associated histone 3 lysine 27 trimethylation (H3K27me3) post-translational modification has been shown to induce gene expression, but not in the reverse (Zhao *et al.,* 2021).

Disease-associated SNPs that lie in promoter regions can result in a diverse array of effects, including shifting the transcription start site (TSS) of the gene, creating or destroying upstream open-reading frames, loss of CpG methylation, effecting alternative gene splicing, or modifying the consensus binding sequence for transcription-associated transcription factors (Robert & Pelletier, 2018). Promoter SNPs have relevance in AS-associated loci, as the alleles of the SNPs rs9525641 and rs9525641 in the *RANKL* promoter – which reside near the *TNFRSF1A* association signal – altered the DNA-binding of as of yet unidentified transcription factors that was associated with changes in *RANKL* expression (Wang *et al.*, 2019). Uncovering whether AS-linked SNPs effect the expression of further genes will be an important factor in underpinning AS pathogenesis.

1.5.3.2. Associated histone marks for enhancer regions

Active enhancer elements usually retain the H3K27ac chromatin contact-associated mark, yet lose the H3K4me3 transcription-associated signal. These regions further differentiate from active promoters through the gain of the monomethylation of lysine 4 of histone 3 (H3K4me1). H3K4me1 typically associates with the barrier-to-autointegration factor (BAF) family of chromatin remodelling proteins which stabilise extruded chromatin looping interactions with the promoter, and are required for the integration of the H3K27ac mark within distal enhancer elements (Local *et al.*, 2018). The vast majority of enhancers retain H3K4me1 irrespective of activation state, and enhancers which lack the H3K27ac mark can be held in a ‘primed’ state by H3K4me1 for ready integration with gene promoters upon the correct context-specific signalling event.

Mutations within the BAF family of proteins are also noted for their association with multiple human cancers, suggesting that inconsistencies with enhancer priming and the stability of chromatin interactions can contribute towards disease (Schick *et al.*, 2019).

Active enhancers are further marked by their production of enhancer RNA (eRNA) (Sartorelli & Lauberth, 2020). The exact role of eRNAs in gene regulation has yet to be elucidated, but expression of eRNA also contributes to gene expression and may modulate protein recruitment, interaction, provide scaffolding for enhancer structure, and help initiate looping interactions (Rahman *et al.*, 2017: Arnold *et al.*, 2020; Lee *et al.*, 2023). A large proportion of disease-associated SNPs impact the DNA binding motif of transcription factors. In psoriasis for instance, risk-associated SNPs were found to disproportionally effect the binding sites of transcription factors that modulate and control immune responses, such as GATA3 (GATA-binding protein 3) and SOX5 (SRY-Box transcription factor 5), within enhancers in close proximity to the genes of inflammatory cytokines like IL-17 and IL-1β (Swindell et al., 2015).

1.5.3.3. lncRNA is associated with chromatin remodelling complexes

Long non-coding RNAs (lncRNAs) constitute a diverse class of RNA molecules characterized by their length (greater than 200 nucleotides in size) and lack of protein-coding capacity (Mattick *et al.*, 2023). Initially dismissed as transcriptional background noise, studies now reveal lncRNAs have multifaceted roles in gene regulation at the transcriptional, post-transcriptional, and epigenetic levels. Studies into the exact function of lncRNAs are ongoing, but lncRNAs have been found to modulate gene accessibility and recruit chromatin remodelling complexes and transcription factors. Expression of the lncRNA LncMyoD, for instance, represses the reprogramming of fibroblastic cells into myoblastic cells through its association with the transcription factor MyoD (myoblast determination protein 1) which increases accessibility to its consensus binding sequence (Dong *et al.*, 2020). Expression of the lncRNA ‘has_circ_0003307’ is positively correlated to inflammatory activity of AS (Li *et al.*, 2020), whilst MEG3 (maternally expressed gene 3) targets miRNA-146a and displays immunosuppressive effects by

decreasing the expression of IL-1β, IL-6, and TNFα. Variants effecting the expression of other lncRNAs in AS could have profound impact on disease susceptibility.

1.5.3.4. Associated histone marks for silencer regions

Silencers and their associated histone mark patterns are the least characterised of the cis-regulatory elements. Silencers are associated with the histone marks H3K9me3 and H3K27me3 (Pang & Snider, 2020). Whilst both marks are associated with heterochromatin, H3K27me3 is typically involved in the reversible repression of regions which are developmentally important. H3K27me3 can arise through the exchange of the acetyl group of H3K27ac for a trimethyl group catalysed by the Polycomb repressive complex 2 (PRC2) member of the Polycomb-group of repression associated chromatin remodelers (Lavarone *et al.*, 2019; Hisanaga *et al.*, 2023). Despite the important role of H3K27me3 in facultative repression of genes and formation of silencers, the exact mechanisms which govern its deposition and how it leads to repression is an area of intense study (Wiles & Selker, 2017; Liu *et al.*, 2023).

Although enhancers and promoters can be silenced through H3K27me3 substitution for H3K27ac, these silenced regions are not true silencers. True H3K27me3-rich silencing *cis*-regulatory regions loop through to the target gene promoter much like enhancers to promote the pausing of RNA polymerase II. Cai *et al.* (2021) demonstrated that inhibition of the Enhancer of Zeste Homolog 2 (EZH2) component of PRC2 – which is the component responsible for catalysation of the trimethylation of H3K27 – changes the chromatin profile of *cis*-regulatory elements. These conformational changes upon H3K27me3 loss lead to increases in the expression of multiple pro-inflammatory genes including *IGF2* (Insulin-like growth factor 2) within the K562 and HAP-1 cell lines, and suggested that H3K27me3-rich elements physically insert themselves into the three-

dimensional structure of promoter and enhancers to halt gene transcription. Silencers therefore act in opposition to enhancer elements, with release of a silencer from the mediator resulting in gene transcription.

The contribution of SNPs in H3K27me3-rich regions is not as well characterised as those in enhancers, although certain SNPs in Parkinson's disease (*rs2619364*) alter the binding sites for long-term silencing-associated transcription factors (*REST*; RE1-Silencing Transcription factor), suggesting that SNPs residing in them may affect the strength of gene repression (Sharma *et al.*, 2019).

1.5.3.5. Bivalency of histone marks at promoter regions

Interestingly, many *cis*-regulatory elements – particularly promoters – can occupy a paradoxical state between Polycomb-repression and activation through concurrent bivalent occupancy of both H3K4me3 and H3K27me3 on each nucleosome. These bivalent promoters are commonly enriched at TSS regions within stem cells, and can either be further activated or repressed through loss of one of the histone marks as the cell commits to lineage specification (Taube *et al.*, 2017; Kaukonen *et al.*, 2020; Macrae *et al.*, 2023). The exact mechanisms behind bivalent promoters are not well understood, yet bivalent promoters tend to associate tightly with regions rich in the dinucleotide repeat 5'-C-phosphate-G-3' (CpG) (Mendenhall *et al.*, 2010; Mas *et al.*, 2018).

Repetitive stretches of CpG dinucleotides termed islands can themselves be methylated by a number of enzymes from the DNA-methyltransferase family. Methylated CpGs are read by the methyl-CpG-binding domain (MBD) family of DNA-binding proteins, which facilitate the recruitment of chromatin remodelers and histone deacetylases towards transcriptional inhibition (Buchmuller *et al.*, 2020). CpG islands are typically enriched downstream of promoter elements and form an important mechanism of promoter

silencing (Klatt *et al.*, 2020). Indeed, methylated CpG islands have been found associated with H3K27me3, suggesting absence of the contact and activity-associated H3K27ac mark within these regions (Cai *et al.*, 2021). CpG islands are, however, particularly susceptible to cytosine deamination, which can result in spontaneous mutation of the underlying base pairs and loss of methylation capability (Cooper & Youssoufian., 1988; Holliday *et al.*, 1993; Sassa *et al.*, 2016). This loss can inhibit the binding of MBDs and result in abnormal activation of the repressed gene, and contribute towards disease (Buchmuller *et al.*, 2020).

In immune contexts, bivalent promoters are seen at particularly high frequencies within naïve T cells where they control the expression of transcription factors such as T-bet (*TBX21*) (Nguyen *et al.*, 2015). Dysfunction of these T cell-specific bivalent promoters has been implicated in the pathogenesis of multiple diseases. Rhombotin-1 (*LMO1*), for instance, is a transcription factor that under normal circumstances is not expressed within T cells. Once expressed, *LMO1* can contribute to tumour growth, invasion, and migration, and leads to a number of different leukemic and neuroblastomal cancers (Tremblay *et al.*, 2010; Gao *et al.*, 2022). Within T cells, *LMO1* is under the repressive influence of a bivalent promoter (Oram *et al.*, 2013). This promoter loses its H3K27me3 signal within T-acute lymphoblastic leukaemia (T-ALL), becoming associated with flanking enhancers and activating gene transcription (Oram *et al.*, 2013). The changes in the *LMO1* promoter coincide with more systemic hypomethylation of gene promoters within T-ALL, suggesting that the loss in bivalency and increased gene expression is linked to issues in methyl-CpG deposition (Touzart *et al.*, 2020). Many genes controlled by bivalent promoters implicated within cancers have also been found activated within IBD and chronic inflammation (Hahn *et al.*, 2014). To date, the exact role of these bivalent

promoters within axSpA or AS has not been studied, and could yield important insights into disease pathogenesis.

1.5.3.6. The role of insulators in maintaining looping interactions

Dependent upon cell type, not all genes encoded for in the genome will be needed for the function of that cell. Whilst *cis*-regulatory elements modulate the expression profiles of linked genes, the context-specific deactivation of genes can be attained through segregation of the chromosomes in two functionally distinct regulatory compartments in compartment A and compartment B. Compartment A is typically associated with active euchromatin and gene transcription (Zheng & Xie, 2019; Harris *et al.*, 2023), whilst compartment B is associated with gene repression and heterochromatin. These compartments can be further broken down into topologically-associated domains (TADs), which are the basic submegabasal unit of extruded looping interactions. Interactions between genetic elements are often enriched within a TAD, but interactions between regulatory elements occupying adjacent TADs are far rarer (Hanssen *et al.*, 2017; Hung *et al.*, 2024). TADs therefore act to limit the number of genes that enhancers contained within can influence. Unlike the promoter-enhancer or promoter-silencer interactions, the primary role of insulator elements is to mediate the formation of TADs and delimit their boundaries (Hung *et al.*, 2024).

Insulators can be defined experimentally by their enrichment for the architectural proteins cohesin and the zinc-finger protein CCCTC-binding factor (CTCF) without any of the histone post-translational marks common to other *cis*-regulatory elements, with cohesin being recruited to the interaction junction by CTCF which stabilises the extruded chromatin loop (Liu & Dekker, 2022). This distinction is important, as CTCF and cohesin are also important in maintaining active enhancer-promoter interactions and will show

overlap with marks such as H3K4me3 and H3K27ac outside of boundary definition (Davidson *et al.*, 2023). Disease-associated SNPs found in insulator elements can contribute to irregular TAD formation (Jablonski *et al.*, 2022). The SNP *rs12936231* is an asthma, type-1 diabetes, IBD, and primary biliary cirrhosis-associated genetic variant that alters the binding of CTCF to an insulator and weakens the boundary between two adjacent TADs (Verlaan *et al.*, 2009). The SNP is also a known eQTL that effects the expression of pro-inflammatory genes such as *ORMDL3* (Ormdl sphingolipid biosynthesis regulator 3) that can promote IL-1β signalling cascades in lung tissue (Miller *et al.*, 2014). No studies to date however have identified SNP-driven differences in TAD strength in AS.

1.6. Functional genomics and integrative analyses

Non-coding genetic variants, such as SNPs, can have broad but specific effects on *cis*-regulatory function depending on their position. A number of experimental assays that can be used to define protein-DNA interactions, open regions of chromatin, and looping *cis*-regulatory interactions can be used in tandem to identify the potential effects they have on the regulation of genes. With colleagues in the Knight group I have contributed to the generation of integrative multi-omic maps from multiple epigenetic datasets, using methods described in this section (Brown *et al.*, 2023) (**Fig. 1.4**); with some findings detailed in the results chapters to follow.

1.6.1. Experimental techniques for the detection of open chromatin regions

Certain active *cis*-regulatory regions like promoters and enhancers demonstrate open chromatin conformations to remain accessible to their associated transcription factors, co-factors, and chromatin regions they interact with. This open conformation can be exploited by the use of enzymes such as the indiscriminate DNA-digesting DNase I that

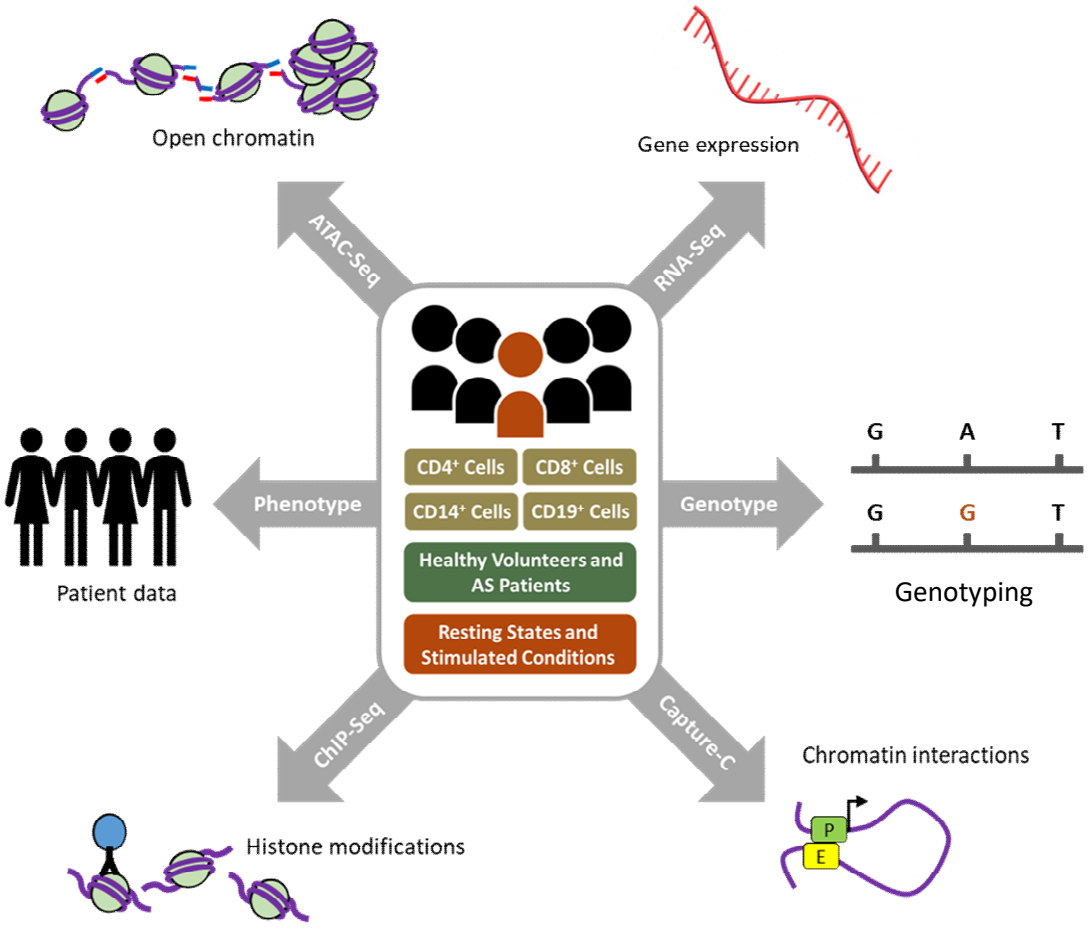

Fig. 1.4. Integrative modelling of epigenetic datasets in AS to determine the effects of disease-associated SNPs on *cis*-regulatory function and gene regulation in PBMCs used by the Knight group. For both healthy controls and AS patients, we isolate PBMC immune cells such as $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells from whole blood samples. In AS patients, we isolate PBMCs for both untreated patients, and for patients treated with biologics such as anti-TNFα and anti-IL-17. For healthy control cells, these can be further cultured with stimulants that evoke inflammatory or immune responses. AS patient data such as BASDAI, CRP levels, and MRI images are collected to identify which patients show the highest disease activity. Patients are genotyped to assess which allele is present at SNPs in AS-associated GWAS loci. Open chromatin, gene expression, histone modification, and chromatin interaction data is gathered form the purified cells to make integrated models across healthy control and AS conditions to identify differential regulatory regions that contain risk-allele SNPs across the healthy control and AS patient sample conditions.

will digest the open chromatin while being sterically hindered at heterochromatic DNA by the histone cores (Mueleman *et al.*, 2020). These 'DNase I hypersensitivity sites' can be identified as in ChIP-seq using NGS to align the non-digested fragments to a reference genome where any unmapped regions are defined as open chromatin, which are likely to include regions of DNA undergoing active transcription. DNase I hypersensitivity has been used extensively in projects such as the ENCODE consortium to characterise regions of open chromatin in over 200 distinct cell types, and it remains one of the most common methods of assessing open chromatin (Moore *et al.*, 2020).

ATAC-seq (**A**ssay for **T**ransposase-**A**ccessible **C**hromatin using **S**equencing) was developed by Buenrostro *et al.* (2013) as an alternative to DNase I hypersensitivity. ATAC-seq uses a hyperactive mutant Tn5 transposase that both cleaves at accessible chromatin regions and adds adaptor sequences for NGS in a process known as tagmentation. ATAC-seq has a number of advantages over DNase I, including being much faster than DNase I, being able to define open chromatin regions in significantly fewer cells (as few as 50,000 cells), and showing data quality equivalent to that seen in three or five orders of magnitude more cells in DNase I hypersensitivity assays (Minnoye *et al.*, 2021).

GWAS variants found outside regions of open chromatin are likely not to be functional as they remain inaccessible to transcriptional machinery and proteins (Das *et al.*, 2022). In AS, single-cell ATAC-seq has identified differential regions of open chromatin over consensus binding sequences for the transcription factors Fos, Jun, and JunB which contributes to irregular expression of genes in the TNFα signalling cascade (Xu *et al.*, 2021). Despite the increased speed and decreased starting material, one major drawback of using ATAC-seq over DNAse-seq assays is the comparative inability to produce high-quality 'footprinting' datasets – where the underlying DNA protected by bound

transcription factors can be aligned to map their specific binding site occupancy – from ATAC-seq datasets (Karabacak Calviello *et al.*, 2019). This is because the Tn5 transposase enzyme has strong intrinsic bias for particular sequences that DNAse does not have, which makes it difficult to distinguish between true transcription factor protection from the transposases favoured cutting patterns (Martins *et al.*, 2018).

Whilst DNAse-seq and ATAC-seq are therefore undoubtedly powerful in identifying which regions may adopt more open or closed conformation in disease, they cannot be used alone to differentiate between the different *cis*-regulatory element types because many *cis*-regulatory elements demonstrate open chromatin conformations.

1.6.2. Experimental techniques for the detection of histone post-translational modifications

Post-translational histone modifications can be investigated using techniques like chromatin immunoprecipitation (ChIP). Whilst many variations of ChIP exist the basic principles of the assay are the same (Schmidt *et al.*, 2009). In ChIP, formaldehyde is first used to covalently cross-link protein-DNA complexes such as that occurring at promoter-enhancer interactions. Following sonication to shear DNA, the target protein is selected using an antibody targeting an epitope on the protein surface. The cross-links are then reversed using proteinases, heat, and high salt concentrations, which denatures and destroys the protein in the fixed complex and releases the sheared, associated DNA fragment. These fragments can be analysed using qPCR if there is a specific DNA target, or in the case of genome-wide high-throughput ChIP-sequencing (ChIP-seq) with NGS (Nakato & Sakata, 2021). ChIPmentation-seq is an improvement on the basic ChIP-seq technique that incorporates the Tn5 transposase enzyme from ATAC-seq, which simultaneously fragments and adds sequencing adaptors to the DNA fragment for NGS

(Schmidl *et al.*, 2015). The difference between ATAC-seq and ChIPmentation-seq is that tagmentation with Tn5 occurs after antibody target selection. Unlike the random DNA breaks generated by sonication alone, ChIPmentation-seq can produce protein or nucleosome-level data resolution as the only DNA regions not transposed by Tn5 are those protected by cross-linked protein complexes.

In ChIP-seq and ChIPmentation-seq, the fragments generated in the NGS amplification process can be aligned to a reference genome and used to delineate precisely where in the genome the antibody-targeted protein was bound (Nakato & Sakata, 2021). ChIP-seq and ChIPmentation-seq can be used to analyse the DNA binding of many diverse proteins, including histones that have been post-translationally modified; such as H3K4me3, H3K27ac, and H3K27me3. Because *cis*-regulatory regions and their activity levels can be defined by post-translational modifications, studies that seek to characterise them often integrate multiple ChIP-seq datasets for different histone marks from the same sample which are then overlapped with each other (Carpenter *et al.*, 2022). For example, active promoter regions can be defined by an overlap of H3K4me3 and H3K27ac-associated genome-aligned sequencing reads. Co-localisation of non-coding GWAS association signals to ChIP-defined regulatory elements can help attribute function to the overlapped variant and how it may be involved in gene-regulation.

ChIP-seq or ChIPmentation-seq can be also be used to demarcate the position of proteins such as RNA polymerase II (Tan *et al.*, 2019), which would localise to actively transcribed genes and active promoters, CTCF and cohesin (Kubo *et al.*, 2021), which would mark insulators, TAD boundaries, and promoter-enhancer interactions, and transcription factors bound to their consensus binding motifs at promoters or enhancers (Chen *et al.*, 2020). In cases such as SNPs that modify transcription factor binding motifs, ChIP-based methods can be used to assess the effect of the SNP on transcription factor

binding by comparing cohorts with different SNP alleles. Transcription factor ChIP-seq has been used in AS to show that differential binding of 29 transcription factors in whole PBMCs – including T cells, monocytes, and B cells – regulated 11,823 target genes; although this dataset was not overlapped with GWAS variants (Yu *et al.*, 2020). In the same study, ATAC-seq revealed that 21 and 9 of these transcription factors bound differentially to regions of more accessible or less accessible chromatin respectively. ChIP-based methods and their overlap with open chromatin regions are therefore powerful tools in initial identification of regulatory elements and in investigating how they may differ in disease to influence gene expression.

1.6.3. Experimental techniques to assay differential gene expression

Knowing how genetic variants like SNPs from regions of differentially open chromatin or post-translational histone modification vary between healthy and patient cohorts and impact gene expression could be crucial to understanding disease pathology. NGS RNA-seq is a powerful tool for assessing the global transcriptomic profiles of genes in healthy controls and AS patients and identifying differential gene expression between the two states (Kukurba *et al.*, 2016). In RNA-seq cells are first lysed, and complementary DNA (cDNA) is produced from the freed RNA transcripts with a reverse transcriptase enzyme. Molecules of RNA are then digested with RNase, leaving only the cDNA in sample. Sequencing adaptors are then ligated directly to the cDNA for NGS, and the sequenced reads are aligned to a reference genome. This allowed for the identification of the genomic location the cDNA molecule derived from and the quantity of RNA transcripts produced from it which can be used to establish trait eQTLs. This includes the expression of non-coding RNA like lncRNA and eRNA and how this may differ between disease and healthy cohorts (Arrigoni *et al.*, 2016). RNA-seq can also be used to assess for the production of specific splice variants, where addition or subtraction of specific exons

from the same original RNA transcript can result in elongated or truncated peptide variants. These splice variants can show divergent activities or protein function upon translation and can be associated with disease (Liu *et al.*, 2022). In AS for example, PBMCs show splice variant skewing in activation-induced cytidine deaminase (AID) compared to healthy controls, with anti-TNFα treatment restoring the variant balance to ordinary levels (Kim *et al.*, 2017).

1.6.4. Experimental techniques to define chromatin interactions

Because *cis*-regulatory elements like enhancers potentially loop over great genomic distances to impart regulatory effects on target promoters, establishing physical linkage between SNP-containing regulatory elements defined by open chromatin and histone post-translational modification and the promoters of differential genes from RNA-seq data can evidence direct regulatory effects on gene expression. Recent advances in specific methods to investigate 3D genomic architecture include a number of 3C-based methods, including Hi-C (high-throughput chromosome conformation capture), Capture-C, and Tiled-C (Davidson *et al.*, 2023).

Most 3C-based assays function through similar basic principles. Like in ChIP, samples are fixed with formaldehyde to preserve protein complexes found at interacting chromatin sites (Downes *et al.*, 2022). DNA is then digested with restriction enzymes, such as DpnII or HindIII, or micrococcal nuclease (MNase). As in ChIPmentation-seq, the cross-linked complex prevents digestion of DNA at the sites of protein interactions whilst maintaining the digested ends in close proximity. DNA ligase is then used to join the two ends, resulting in a single DNA fragment consisting of two interacting restriction fragments. Traditional 3C uses either PCR with gel electrophoresis or qPCR to assess interaction in a one-vs-one fashion, where one of the primers in a pair is targeted to one of the

interacting restriction sites so that the PCR amplification will only occur if the polymerase reads through the digest junction that joins them. This is a powerful technique for assessing regions where there are suspected interactions for each other restriction fragment (i.e. previously characterised promoter-enhancer interactions) but, in contrast, 3C is limited when looking at genome-wide interactions between uncharacterised DNA regions. Other variations of basic 3C exist, such as circular 3C (4C) and carbon copy 3C (5C) (Dostie *et al.*, 2006; Werken *et al.*, 2012) but, like 3C, these methods have limitations when applied to genome-wide chromatin contacts (Davidson *et al.*, 2023).

Hi-C is an improvement on conventional 3C methodologies and functions by incorporating biotin nucleotide fill-in and blunting of the fragment end following the initial restriction enzyme digestion step, with blunt ends being subsequently ligated with a DNA ligase (Lieberman-Aiden *et al.*, 2009). The biotinylated restriction fragments can then be pulled down using streptavidin-coated beads, followed by NGS sequencing of the ligated restriction fragments and *in-silico* reconstruction of all digestion-ligation junctions across the genome to identify interacting chromatin. Unlike 3C, Hi-C can also identify interactions between DNA fragments of unknown initial function, meaning the assay does not require initial input from assays like ATAC-seq or ChIP-seq to assess chromatin interactions.

In general, 3C experiments have limitations on their resolution which is defined mostly by the average DNA fragment size produced by the selected restriction enzyme. Six base-pair cutters like HindIII produce average DNA fragments of roughly 4 kbp in size, meaning that interactions in HindIII 3C experiments can only be determined in 4 kbp intervals (Kempfer & Pombo, 2020). In contrast, the four base-pair cutter DpnII produces average restriction fragments of 256 bp and a max resolution 15-times greater than HindIII. However, increasing resolution comes with greater sequencing costs, as the

complexity of a 3C library – which is defined as the number of unique molecules containing a ligation junction in the library – increases squarely with the proportion of digestion sites (Lieberman-Aiden *et al.*, 2009; Downes *et al.*, 2022). 3C libraries generated using DpnII require far more sequencing than those generated with HindIII, as the former demonstrates a 256-fold increase in sample complexity. In Hi-C experiments, where billions of sequencing reads are already required to generate genome-wide contact maps, the associated sequencing costs put the use of Hi-C at a resolution capable of identifying the interaction between individual *cis*-regulatory elements out of the scope of most research groups (Lee *et al.*, 2022). Nevertheless, Hi-C currently remains the best method to map genome-wide chromatin contacts.

In cases where regions of interest have been previously defined, such as on the single restriction fragment (i.e. containing a disease-associated GWAS SNP) or TAD scales (i.e. an entire TAD containing a GWAS association signal), Capture-C, Tiled-C, and capture Hi-C can be used to generate one-vs-all (Capture-C) or some-vs-all (Tiled-C and capture Hi-C) interaction profiles. Capture-C and Tiled-C utilise biotinylated oligonucleotides targeted to restriction fragments of interest post-sample processing that are then pulled-down with streptavidin beads (Downes *et al.*, 2022). Capture Hi-C instead enriches for specific targets in a 3C library that used the biotinylated fill-in method. All of these techniques allow for the analysis of the targeted restriction fragment viewpoints in isolation from background restriction fragments following NGS, which can be achieved for a fraction of the sequencing cost of Hi-C. To date, there have been no studies using 3C to investigate how chromatin architecture and *cis*-regulatory interaction may be affected in AS, but such data could prove critical to identifying causal disease genetic variants.

1.7. Specific aims and objectives

The aims of this thesis are to:

1.7.1. Aim 1: To quantify and compare the changes in the epigenetic landscape that occur between AS patients and healthy volunteers (Chapter 3).

Specific chapter aims are:

[1] Generate genome-wide maps of regulatory elements by integrating open chromatin, histone post-translational modification, and RNA expression datasets from healthy control and AS patient samples.

[2] Enrich AS-associated SNPs identified through GWAS for the regulatory regions identified in the integrated model to identify which lie in active cis-regulatory elements like promoters and enhancers.

[3] Use the integrated epigenetic models to characterise how the regulatory landscape may change at AS-associated GWAS loci in AS patients and healthy volunteers.

[4] Analyse the impact that SNPs associated with MHC genes like HLA-B27 have on the enrichment of SNPs for non-MHC regulatory elements.

1.7.2. Aim 2: Characterise chromosomal interactions between AS-associated GWAS lead SNPs and target gene promoters (Chapter 4).

Specific chapter aims are:

[1] Characterise the looping interactions between lead SNPs from AS GWAS Cortes *et al.* (2013) and Ellinghaus *et al.* (2016) studies) and target gene promoters in disease-relevant immune cell types.

[2] Identify putative functional SNPs that lie in regulatory regions involved in chromatin looping at AS GWAS loci.

[3] Quantify the potential differences in looping frequency that occurs between the lead SNP and its target gene in AS patients and healthy volunteers.

1.7.3. Aim 3: Multi-omic analysis of the epigenetic environment for AS-associated GWAS loci in $CD4^+$ and $CD8^+$ T cells (Chapter 5).

Specific chapter aims are:

[1] Identify all genes in TADs that interact with AS GWAS SNPs situated in *cis*-regulatory as predicted by integrated epigenetic modelling in $CD4^+$ and $CD8^+$ T cells.

[2] Investigate how regions of open chromatin and gene expression of all genes in the TAD change in healthy volunteers and AS patients in response to immune stimulation and anti-TNFα therapeutics.

[3] Investigate how TAD restructuring or the strength of *cis*-regulatory interactions may change at AS GWAS loci to contribute towards AS pathogenesis.

"Science is no more than an investigation of a miracle we can never explain, and art is an interpretation of that miracle."

- Ray Bradbury, The Martian Chronicles (1950)

2

METHODS & MATERIALS

Contents **Pg.**

2.1. Study recruitment and ethical approval 66
2.2. Sample processing and cell culture 69
2.3. Chromatin conformation capture 74
2.4. ChIPmentation-seq 96
2.5. ATAC-seq..... 102
2.6. RT-qPCR and RNA-seq 103
2.7. Computational and statistical analyses 105

This chapter describes the recruitment of healthy volunteer and AS patient cohorts, the experimental procedures used to compare the epigenetic environments between them, and the computational methods used to analyse data throughout this thesis.

2.1. Study recruitment and ethical approval

2.1.1. AS patient and healthy volunteer sampling

Patient blood samples were collected in collaboration with Prof. Bryan Paul Wordsworth and Prof. Paul Bowness at the NHS Nuffield Orthopaedic Centre in Oxford. Healthy volunteer blood was taken in collaboration with Dr. Patrick MacLean, Dr. Hanyu Qin, Dr. Lahiru Hannundetthi, and Dr. Andrew Kwok of the Wellcome Centre for Human Genetics (WCHG) at the University of Oxford. Samples were collected under ethical approval from the National Ethics Service Committee South Central - Oxford Research

Ethics Committee B and/or C. Samples from healthy volunteers were collected under REC 06/Q1605/55. Most AS patients were collected under REC 12/SC/0063. Samples from AS patients AS301-AS305 used in Tiled-C analyses were collected pre and post-treatment with anti-TNFα biologics under REC C06/Q1606/139. Following informed written consent, up to 100mL of whole blood from AS patients and healthy volunteers were collected into 10mL BD Vacutainer™ K2 blood-tubes anticoagulated with ethylenediaminetetraacetic acid (EDTA) (BD Biosciences™, 367525) for further processing as in section 2.2.

2.1.2. Eligibility criteria and recruitment demographics

All patients included in subsequent studies were over the age of 18 and diagnosed with axial spondyloarthritis following the ASAS diagnostic criteria. Cases were reviewed by two collaborating consultant rheumatologists to confirm the presence of active disease prior to study recruitment. This was carried out according to the eligibility criteria for biologic intervention underlined by the British Society for Rheumatology (Hamilton *et al.*, 2016). Selection of appropriate AS patients was also informed through the taking of BASDAI, CRP levels, and spinal scores prior to blood draw. 87.5% of patients were positive for HLA-B27. 58% of patients were taking NSAIDs and 12.5% were taking DMARDs. No patients were taking anti-TNFα or anti-IL-17 biologics prior to their recommendation by the consultants. All patients demonstrated axial manifestations of AS. A subset of patients also presented with secondary manifestations of disease, with the most major comorbidity being uveitis. 79% of patients were male and 21% were female, reflecting the challenges of differentiating active disease between the sexes. A table showing the clinical features of AS patients prior to commencing biologic therapy has been enclosed in this chapter **(Table 2.1**). Healthy volunteers recruited in this study were required to be at least 18 years of age, with no self-reported familial history of AS

Demographics			AS Diagnostics					AS Comorbidities								Family history			Drug Treatment		
Patient	Sex	Age	HLA-B27(+)	Disease Duration	BASDAI	ASDAS-CRP	CRP (mg/L)	Psoriasis	IBD	Uveitis	Cancer	Periph. Arthritis	Vertebral fracture	Fracture	Other	AS	Other Autoimm.	Relative	Ibuprofen	Other NSAID	DMARDs
AS004	M	70-79	Yes	246	7.8	5.5	90.7	No	No	No	No	Yes	No	No	N/A	No	IBD	Parent	Current	Current	Current
AS014	F	20-29	Yes	48	5.4	1.7	0.28	No	No	Yes	No	No	No	No	N/A	Yes	No	Other	None	Ever	None
AS017	F	40-49	Yes	22	4.6	2.13	0.5	Yes	No	Yes	No	No	No	Yes	N/A	No	Psoriasis	Sibling	Current	To Start	None
AS018	M	20-29	Yes	96	9.3	-	11.2	Yes	No	No	No	No	No	No	N/A	Yes	No	Sibling	None	Current	None
AS044	M	30-39	Yes	300	7.2	3.91	9.4	No	No	No	No	No	No	No	N/A	Unsure	Unsure	Unsure	None	Current	None
AS045	M	30-39	Yes	120	5.5	4.3	30.2	No	No	No	No	Yes	No	No	N/A	No	No	Unsure	Ever	Current	None
AS046	M	30-39	Yes	36	7.6	3.9	7.2	No	No	No	No	No	Yes	Yes	N/A	Unsure	Unsure	Unsure	Ever	Current	None
AS047	M	30-39	Yes	9	6.3	3.8	8.8	No	No	No	No	Yes	No	Yes	N/A	Yes	RA	Parent	None	Current	None
AS049	F	30-39	No	16	4.2	2.2	11.6	No	Yes	No	No	No	No	No	N/A	No	No	Unsure	Ever	None	None
AS050	M	40-49	Yes	48	8.4	3.8	2.6	No	No	No	No	Yes	No	Yes	N/A	No	RA	Unsure	Ever	Ever	Current
AS051	F	50-59	No	27	7.2	3.7	10.6	No	No	Yes	No	No	No	No	N/A	Unsure	Unsure	Unsure	Ever	Ever	None
AS052	M	60-69	Yes	48	5.1	2.5	1.2	No	No	Yes	Yes	No	No	Yes	N/A	No	No	Unsure	None	None	None
AS053	F	30-39	Yes	11	5.4	3.3	15.3	No	No	No	Yes	No	No	No	N/A	Yes	Unsure	Parent	Ever	None	None
AS054	M	30-39	No	300	7.9	3.1	0.5	No	No	No	No	No	No	Yes	N/A	No	No	Unsure	Ever	Ever	None
AS058	M	40-49	Yes	18	2.7	1.5	1.9	Yes	No	No	No	Yes	No	No	N/A	Yes	Thyroid	Parent	Ever	Current	None
AS201	M	20-29	Yes	120	4.6	-	-	No	No	No	No	Yes	No	No	N/A	Yes	Unsure	Other	Ever	Current	Current
AS202	M	60-69	Yes	18	5.5	-	-	No	No	Yes	No	Yes	No	Yes	N/A	No	No	Unsure	None	Current	None
AS203	M	70-79	Yes	456	4.2	-	-	No	No	Yes	No	Yes	No	No	Thyroid	No	No	Unsure	None	None	None
AS205	M	30-39	Yes	240	5.9	-	-	No	No	No	No	No	No	No	N/A	Yes	Yes	Parent	Ever	Current	None
AS207	M	60-69	Yes	372	3.9	-	-	No	No	No	No	No	No	No	N/A	Unsure	IBD	Parent	None	Current	None
AS301	M	50-59	Yes	156	6.4	-	78.5	No	No	No	No	Yes	No	No	N/A	No	No	No	Ever	Current	None
AS303	M	50-59	Yes	190	5.6	-	2.2	No	No	Yes	No	No	No	No	N/A	No	No	No	Ever	Current	None
AS304	M	50-59	Yes	110	5.4	-	0.8	No	No	No	No	No	No	No	N/A	No	No	No	Ever	Current	None
AS305	M	20-29	Yes	36	6.6	-	23.7	No	No	Yes	No	No	No	No	N/A	No	No	Unsure	Ever	None	None

Table 2.1. Clinical features of AS patients included in all analyses in this thesis. Demographic (Sex and age), and clinical diagnostic criteria (HLA-B27 positivity, duration of AS, BASDAI scores, ASDAS-CRP scores, and CRP protein levels) for all AS patients included in the various analyses of this thesis. Whether any of these patients also demonstrated overlap with multiple common inflammatory comorbidities with AS (Psoriasis, IBD, uveitis, cancer, peripheral arthritis, vertebral fracture, or generalised fractures) has also been shown. The 'other' columns show patients with other autoimmune diseases or conditions of lesser links to AS (i.e. autoimmune thyroid disease). The family history section shows whether each patient has a family history of AS or other autoimmune conditions and the relative tab shows the closest relative to present with the disease. 'RA' = Rheumatoid arthritis; 'IBD' = Inflammatory Bowel disease. Individuals marked 'unsure' could have relatives that also present with an inflammatory disease. The drug treatment section shows a complete NSAID or DMARD therapeutic history of each patient. 'Current' indicates current use of NSAIDs or DMARDs, 'ever' indicates they have taken either drug in the past, and 'none' indicates never having taken either. Dashes represent data not collected for that particular patient.

or evidence of infection at least two weeks prior to blood collection. The ratio of male to female healthy volunteers was 53:47, and volunteers were age and sex-matched as close as possible to patients. A summary table of all AS patients (n = 24) and healthy volunteers (n = 44) included in this thesis and the assays used upon samples derived from them can be found enclosed in this chapter (**Table 2.2**). My own contributions to these datasets have been highlighted in red (processed alone) or blue (processed in collaboration) there within. AS patients that demonstrated BASDAI scores greater than four and had imaging evidence of inflammation at the sacroiliac joints or spinal entheses were recommended for anti-TNFα biologic treatment (adalimumab). Successful response to biologic treatment was defined as at least a 50% reduction of BASDAI score and/or to below a score of four. Resolution of MRI-based evidence of inflammation at the sacroiliac joints was also required. Both were retaken three months following initial blood draw and commencement of adalimumab to ascertain the success of therapy (**Fig. 2.1**).

2.2. Sample processing and cell culture

2.2.1. Peripheral blood mononuclear cell purification

Blood samples from AS patients and healthy volunteers were processed immediately following collection. Collected blood samples were combined in 50 mL Falcon™ tubes (Fisher Scientific™,10788561) and diluted 1:2 (v/v) with Hank's balanced salt solution (HBSS) (Gibco™, 14025092). PBMCs were isolated following density-gradient separation within 50mL Leucosep™ (Greiner™, 227288) tubes with 15 mL of Lymphoprep™ density gradient medium (Stemcell™ Technologies, 07851). Leucosep™ tubes were pre-prepared according to manufacturer's instructions. Samples were centrifuged at 1000xg for 10 minutes at room temperature with maximum acceleration and no braking. The PBMC layer forming at the plasma-Lymphoprep™

Individual	V2/Stim	Sample Information					RNA				ATAC				ChIPm - H3K4me3				ChIPm - H3K27ac				ChIPm - H3K27me3				ChIPm - H3K4me1				ChromHMM				Capture-C				Tiled-C			
		Condition	Sex	Age	Set	ATAC method	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+	CD4+	CD8+	CD14+	CD19+
AS004	-	AS	M	70-79	Set1	-	-	-	-	-	-	-	-	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS014	-	AS	F	20-29	Set1	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS017	-	AS	F	40-49	Set1	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS018	-	AS	M	20-29	Set1	-	-	-	-	-	X	X	-	-	-	X	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS044	-	AS	M	30-39	Set2	ATAC	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS045	-	AS	M	30-39	Set2	ATAC	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS046	-	AS	M	30-39	Set2	ATAC	X	X	X	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS047	-	AS	M	30-39	Set2	ATAC	-	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS049	X	AS	F	30-39	Set3	FATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	X	X	X	X	-	-	-	-	-	-
AS050	X	AS	M	40-49	Set3	FATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	X	X	X	X	-	-	-	-	-	-
AS051	X	AS	F	50-59	Set3	FATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	X	X	X	X	-	-	-	-	-	-
AS052	X	AS	M	60-69	Set3	FATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	X	X	X	X	-	-	-	-	-	-
AS053	X	AS	F	30-39	Set4	OATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS054	X	AS	M	30-39	Set4	OATAC	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS058	X	AS	M	40-49	Set5	OATAC	X	X	X	X	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS201	X	AS	M	20-29	Set5	OATAC	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V1	-	-	-	-	-
AS202	X	AS	M	60-69	Set6	OATAC	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-
AS203	X	AS	M	70-79	Set6	OATAC	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS205	X	AS	M	30-39	Set6	OATAC	X	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
AS207	X	AS	M	60-69	Set6	OATAC	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	X	-	-	-	-	-	-
AS301	X	AS	M	50-59	Set9	OATAC	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-
AS303	X	AS	M	50-59	Set9	OATAC	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-
AS304	X	AS	M	50-59	Set9	OATAC	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-
AS305	X	AS	M	20-29	Set9	OATAC	-	-	-	-	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	X	V2	X	-	-	-	-
OBB1169	-	HV	F	40-49	Set1	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OBB2657	-	HV	F	30-39	Set1	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OBB3073	-	HV	F	50-59	Set1	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OBB4267	-	HV	F	40-49	Set1	-	-	-	-	-	-	-	-	-	X	X	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OBB4633	-	HV	F	30-39	Set1	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL1	-	HV	M	30-39	Set2	ATAC	X	X	X	-	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL2	-	HV	M	50-59	Set2	ATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL3	-	HV	M	30-39	Set2	ATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL4	-	HV	F	40-49	Set2	ATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL5	-	HV	M	40-49	Set2	ATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTL6	-	HV	M	40-49	Set2	FATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
130_2	-	HV	M	50-59	Set3	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
81_2	-	HV	F	50-59	Set3	FATAC	X	X	X	-	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV29	-	HV	M	50-59	Set3	FATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV30	-	HV	M	60-69	Set3	FATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV31	-	HV	F	30-39	Set4	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV42	-	HV	M	40-49	Set4	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV26	-	HV	M	30-39	Set5	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV46	-	HV	F	30-39	Set5	OATAC	X	X	X	X	-	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV59	-	HV	F	20-29	Set5	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV60	-	HV	M	20-29	Set6	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV61_2	-	HV	F	40-49	Set6	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV87	-	HV	M	30-39	Set6	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV88	-	HV	M	20-29	Set6	OATAC	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV89	-	HV	F	20-29	Set6	OATAC	-	-	-	-	-	X	-	-	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV01	-	HV	F	50-59	Set7	OATAC	-	-	-	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV02	-	HV	F	50-59	Set7	OATAC	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV03	-	HV	F	30-39	Set7	OATAC	X	-	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV04	-	HV	M	50-59	Set7	OATAC	X	-	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV05	-	HV	M	40-49	Set7	OATAC	X	-	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV06	-	HV	F	40-49	Set7	OATAC	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV07	-	HV	F	40-49	Set7	OATAC	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV08	-	HV	M	60-69	Set7	OATAC	X	-	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV09	-	HV	M	30-39	Set7	OATAC	X	-	X	-	X	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
SV10	-	HV	F	50-59	Set7	OATAC	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HV117	X	HV	M	40-49	Set8	OATAC	-	X	X	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-
HV120	X	HV	F	40-49	Set8	OATAC	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-
HV121	X	HV	F	30-39	Set8	OATAC	X	X	X	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-
HV122	X	HV	M	60-69	Set8	OATAC	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-
HV123	X	HV	M	30-39	Set8	OATAC	X	X	X	-	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CN01	X	HV	M	N/A	Set9	OATAC	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-
CN02	X	HV	F	N/A	Set9	OATAC	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-
CN03	X	HV	M	N/A	Set9	OATAC	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-
CN04	X	HV	M	N/A	Set9	OATAC	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-

Table 2.2. Sample use and genomic assays performed in recruited healthy volunteers and AS patients (Continued from previous page). Table shows AS patients and healthy volunteers were used in each individual epigenomic assay. 'Sample Information' gives the randomised invidual ID (Individual), whether it has paired post-anti-TNFα treatment or post-immune stimulation data (V2/Stim), whether that individual is an AS patient or healthy volunteer (HV) (Condition), the biological sex of each recruited individual (Sex), the age bracket (Age), the recruitment cohort (Set), and the type of ATAC assay used (ATAC method). The ID of samples derived from AS patients have been shown in red, and healthy volunteers in green. Columns showing which samples and cell types ('$CD4^+$' T cells, '$CD8^+$' T cells, '$CD14^+$' monocytes, and '$CD19^+$' B cells) were used from each individual in each epigenomic assay or technique ('RNA', 'ATAC', 'ChIPm', 'ChromHMM', 'Capture-C', and 'Tiled-C') are then shown. 'X' and '-' respectively indicate whether that a sample for a given individual was used or not in an epigenomic assay. Red crosses represent samples that were processed by the thesis author, blue crosses represent samples processed and/or analysed in collaboration with other researchers, and black crosses represent samples with pre-existing data. 'V1' or 'V2' in the Tiled-C column shows if only pre or post-anti-TNFα (adalimumab) treatment data was available for that individual.

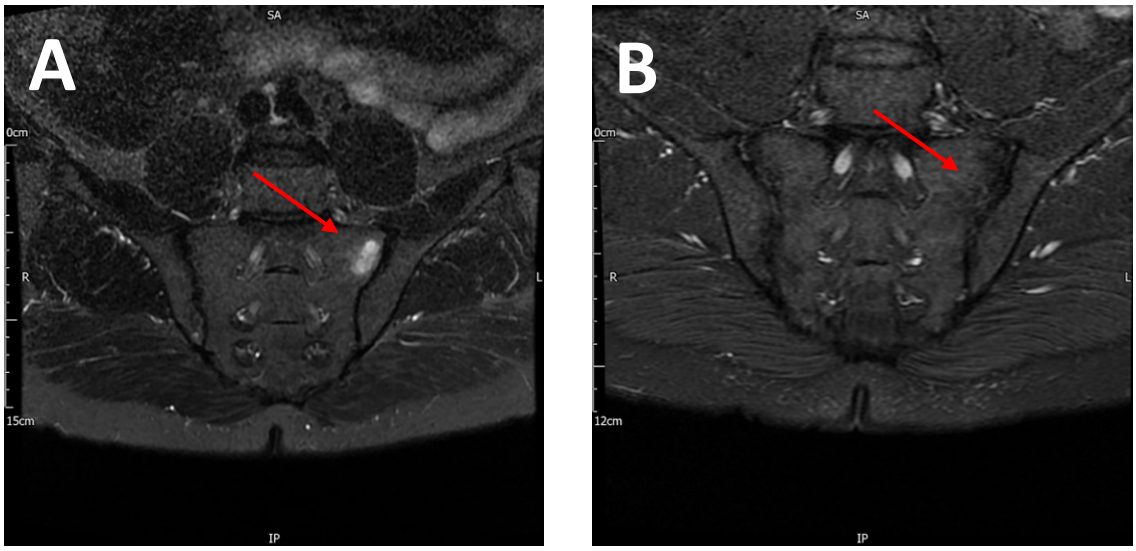

Figure 2.1. Resolution of sacroiliac inflammation within an AS patient following a three-month course of anti-TNFα therapy. (A) MRI scan of an AS patient prior to commencing anti-TNFα therapy at time of first recruitment. The red arrow points to an area shaded in white within the sacroiliac joint demonstrating water infiltration. This is indicative of localised inflammation at the sacroiliac joint. (B) MRI scan of the same AS patient following 3-month treatment with adalimumab. The absence of the white shaded area suggests resolution of sacroiliac joint inflammation and is typical of a good response to anti-TNFα therapy. All patients included in this thesis demonstrated sacroiliac inflammation on an MRI (as above) or X-ray, which was subsequently attenuated by anti-TNFα or IL17 therapy. MRI scans were acquired by Prof. Paul Wordsworth from AS patient AS303 under REC C06/Q1606/139.

interphase was removed using a Pasteur pipette into a new 50mL Falcon™ tube and washed twice with 20 mL of ice-cold HBSS. Cells were resuspended in a final volume of 20 mL of ice-cold phosphate buffered saline (PBS) (Gibco™, 10010023) supplemented with 2 mM EDTA (Sigma-Aldrich®, 102161034) and 0.5% foetal calf serum (FCS) (Gibco™, 10270106) (AutoMACS resuspension buffer). Cell counts and viabilities were determined manually following appropriate dilution with AutoMACS resuspension buffer using a haemocytometer and 1:1 (v/v) trypan-blue staining (ThermoFisher™, 15250061).

2.2.2. Magnetic-activated cell sorting

Subpopulations of target PBMCs were isolated using magnetic-activated cell sorting (MACS) LS columns (Miltenyi™, 130-042-401) following manufacturer's instructions. $CD8^+$, $CD14^+$, and $CD4^+$ cell populations from patient and healthy volunteer samples were isolated consecutively using their respective magnetic beads (Miltenyi™, 130-045-201; 130-050-201; 130-045-101). This was done to avoid contaminating the $CD4^+$ T cell fraction with CD4-expressing $CD14^+$ monocytes. Briefly, samples were incubated with magnetic beads in the dark for 15 minutes at 4°C. 15 µL of beads and 80 µL of AutoMACS resuspension buffer were used for every 10^7 isolated cells. Cell-bead mixtures were added to the columns and washed three times with 5 mL of AutoMACS resuspension buffer, with cells being eluted from the column using a final 5 mL of PBS supplemented with 1% FCS. The negative cellular flow-through was centrifuged between purifications (300g, 10 minutes, 4°C) and the supernatant was aspirated in preparation for mixing with next magnetic bead type. All cell types were centrifuged (400g, 10 minutes, 4°C) following completion of purifications, and resuspended in the appropriate cell culture media for each cell type as in section 2.2.3.

2.2.3. PBMC culture and cell preparation

2.2.3.1. $CD4^+$ and $CD8^+$ T cells

Purified $CD4^+$ or $CD8^+$ T cells were resuspended to a concentration of $2.5x10^6$ cells/mL with X-VIVO 15™ Haematopoietic Cell Culture Medium (Lonza™, LZBE02-060F) supplemented with 1% heat-inactivated human serum (Sigma-Aldrich®, H3667) and 1% penicillin-streptomycin (Gibco™, 15140122). 100 µL of this solution was added to the appropriate number of wells in a 96-well plate and topped with either an additional 100 µL of complete X-VIVO 15™ culture medium or stimulated with CD3/CD28 T cell stimulation medium (0.3 CD3/CD28 beads per cell) (Gibco™, 11161D). Stimulation beads were washed three times and resuspended in an appropriate amount of complete X-VIVO 15™ medium prior to addition as above. $CD4^+$ or $CD8^+$ T cells were then allowed to incubate for 16 hours (37°C, 5% CO_2) prior to cell harvesting.

2.2.3.2. $CD14^+$ monocytes and macrophages

For *NOS2* expression experiments monocyte-derived macrophages were generated as previously described (Ohradanova-Repic *et al.*, 2016). Briefly, purified $CD14^+$ monocytes were first resuspended to a concentration of $1x10^6$ cells with RPMI-1640 medium (Gibco™, R7388) supplemented with 2 mM L-glutamine (G8540, Sigma-Aldrich®), 1% penicillin-streptomycin, and 10% heat-inactivated human serum (H5667, Sigma-Aldrich®). 1 mL of the resuspended monocytes was then added to each well of an adherent Nunclon™ Supra Surface 6-well plates (ThermoScientific™, 140682) and allowed to adhere for 1 hour at 37°C (5% CO_2). Following cell adhesion, human macrophage colony stimulating factor (M-CSF) (81627-83-0, Sigma-Aldrich®) was added to the medium at a final concentration of 50 ng/mL. $CD14^+$ monocytes were then cultured for 7 days to produce macrophages with fresh medium change on day 4.

Following differentiation, macrophages were dissociated from adherent plates with 1 mL of Trypsin-EDTA (25200056, Gibco™) and subsequent incubation at 37°C (5% CO_2) for 5 minutes. Removal of residual adherent cells was achieved with gentle use of a cell scraper. Dissociated macrophages were then aliquoted into a new 15 mL Falcon™ tube (1495953A, ThermoScientific™) for experimental use.

For chromosome conformation capture experiments, purified $CD14^+$ monocytes were resuspended to a concentration of $2x10^6$ cells/mL with RPMI-1640 Medium supplemented with 10% FCS and 1% penicillin-streptomycin. 100 µL of this solution was added to the appropriate number of wells in a 96-well plate and topped with either an additional 100 µL of complete RPMI-1640 culture medium or stimulated with complete medium supplemented with 20 ng/mL of lipopolysaccharide (LPS) (Cell Signalling Technology, 140111S). $CD14^+$ monocytes were then allowed to incubate for 24 hours (37°C, 5% CO_2) prior to cell harvesting.

2.3. Capture-C and Tiled-C

2.3.1. Fixation of cells for Capture-C and Tiled-C

Between $2x10^6$ and $10x10^6$ MACS-purified cells (Section 2.2.3.1 to 2.2.3.3) were harvested and resuspended in 10 mL of their respective complete growth media and fixed through addition of formaldehyde (Sigma-Aldrich®, 47609) to a final concentration of 2% v/v. Cell suspensions were incubated at room temperature for 10 minutes on an end-over rotator. Samples were then quenched by adding 1 mL of ice-cold 2.5 M glycine to a final concentration of 330 mM (Sigma-Aldrich®, G1726), and incubated on ice for 15 minutes. Samples were then washed twice with 10 mL of ice-cold PBS, with the cells pelleted between wash steps with 500xg centrifugation for 10 minutes at 4ºC. Cell pellets were then resuspended in 5 mL of ice-cold Capture-C lysis buffer (10 mM Tris-HCl (pH

8); 10 mM NaCL; 0.2% Igepal CA-630 (Sigma-Aldrich®, I8896); 1 tablet cOmplete Protease Inhibitor Cocktail (04693132001, Roche®)) and incubated for 20 minutes on ice. Resultant lysates were then centrifuged as above for 20 minutes, aspirated, and resuspended in another 1 mL of ice-cold Capture-C lysis buffer, and snap-frozen in liquid nitrogen prior to storage at -80ºC.

2.3.2. Digestion of 3C libraries with DpnII

Capture-C was performed according to the methodology developed by Downes *et al.*, 2022 with modification. Cell lysates were thawed slowly at 4ºC before being pelleted at 750xg for 10 minutes at room temperature. The supernatant was subsequently removed and the cell pellet resuspended carefully in 1 mL of 1x DpnII buffer (New England Biolabs®, R0543M) with wide-bore pipette tips. Samples were centrifuged at 500xg for 10 minutes at 4ºC and resuspended into separate 200 µL aliquots of 1x DpnII buffer per a maximum of $5x10^6$ cells. An extra 50 µL was additionally added for an undigested DNA control. Single cell suspensions were created by using the grinding pestle handle of a Dounce homogeniser (Avantor™, 423-0250), with both the handle and pestle washed twice by flooding with 70% EtOH (v/v), douncing ten times, emptying, and rinsing with nuclease-free water. Handles and pestles were autoclaved between each unique sample preparation. SDS was added to each 200 µL aliquot and undigested control to a final concentration of 0.28% (v/v) (Sigma-Aldrich®, 71736), and both were then incubated in a thermomixer for 1 hour at 37ºC with constant horizontal shaking at 900 rpm.

SDS was then quenched by adding Triton-X100 (Sigma-Aldrich®, X100) to a final concentration of 1.67% and both sample and control were incubated in a thermomixer for 1 hour at 37ºC with constant horizontal shaking at 900 rpm. 500 units of DpnII (New England Biolabs®, R0543M) were then added to each reaction roughly every 8 hours

over a period of 24 hours to digest the chromatin. Samples were incubated over the 24 hours in a thermomixer at 37ºC with constant horizontal shaking at 1200rpm. Undigested controls were stored at -20ºC without DpnII addition.

2.3.3. 3C library ligation and de-crosslinking

Upon completion of digestion, samples were placed in a thermoblock at 65ºC for 30 minutes to heat inactivate the DpnII enzyme. Samples were chilled on ice for 10 minutes prior to subsequent addition of 500 µL of ice-cold nuclease-free water and 134 µL of T4 DNA ligase buffer (Life Technologies™, EL0013) to avoid ATP degradation within the buffer. 240 units of T4 DNA ligase (Life Technologies™, EL0013) was then added to each tube, and the samples were placed in a thermomixer set to 16ºC for 16 hours shaking at 1400 rpm. 30 µL from each DpnII digested library was collected prior to ligation to serve as a digested control. The digested controls were stored at -20ºC until use.

Following ligation, samples were moved to a pre-heated 65ºC thermoblock and 5 µL of proteinase K (Thermo Scientific™, EO0491) was added to remove protein complexes and formaldehyde-induced crosslinks from the re-ligated DNA. Controls were thawed and volume increased to 300 µL with nuclease-free water, with 3 µl of proteinase K added to each control tube. Both Capture-C samples and controls were incubated at 65ºC for 16 hours. Samples and controls were then cooled to 37ºC, and 10 µL or 6 µL of RNase A (Roche®, 1119915) was added into each respectively. Samples were then incubated for a further 30 minutes at 37ºC.

2.3.4. 3C library purification

Following digestion of RNA, biological replicates of samples were pooled together and an equal volume of phenol-chloroform-isoamyl alcohol (Sigma-Aldrich®, 77617) was added (v/v) to both pooled samples and digestion controls. Samples and controls were

then vortexed at max speed for 1 minute, and aliquoted into pre-prepared 5PRIME light phase-lock gel tubes (Quantabio®, 2302830) as per manufacturer's instructions. Cells were vortexed at 12,000xg for 10 minutes to separate the phases, and the aqueous phase was collected avoiding the gel interface. Samples were then mixed 1:1 (v/v) with chloroform (Sigma-Aldrich®, 650498) to remove residual phenol. Samples were again vortexed at max speed for 1 minute, and phases split using phase-lock tubes as above. To the collected aqueous phase, 100% ethanol and 3M sodium acetate (pH 5.2) (Invitrogen™, AM9740) were added to final concentrations of 70% (v/v) and 0.3M respectively. Following vigorous mixing, samples were then precipitated at -20ºC for at least 24 hours.

To collect precipitated DNA, samples were centrifuged at 14,000xg for 30 minutes at 4ºC, and the supernatant aspirated avoiding DNA pellet breakage. Undigested and digested controls were air-dried and resuspended in nuclease-free water once DNA pellet turned translucent. Capture-C sample pellets were instead washed twice with 70% ethanol, with 14,000xg for 10 minutes at 4ºC centrifugation between wash steps. Samples were then resuspended in an appropriate amount of nuclease-free water depending on pellet size, and left for 2 hours before being vortexed at max speed for 5 minutes to ensure total resuspension of DNA. 3C library yields were quantified using the broad-range dsDNA Qubit® kit (Invitrogen™, Q32853), whilst digestion and ligation profiles were obtained from genomic DNA ScreenTapes (Agilent©, 5067-5365) run on an Agilent© 4200 TapeStation (Agilent, G2964AA) as per manufacturer's instructions (**Fig. 2.2**).

A qPCR was used to determine DpnII digestion efficiency of the 3C libraries. Briefly, two sets of primers were created to span a DpnII restriction and a control site which was not cut by DpnII around the *PKDREJ* gene **(Table 2.3**). 1 µL of both the digested control and the undigested control were mixed with 10 µL of QuantiNova™ SYBR Green PCR

mix (208252, Qiagen®) and one of each primer pair to a final concentration of 0.7 µM. Each reaction was topped to a final volume of 20 µL with nuclease-free water.

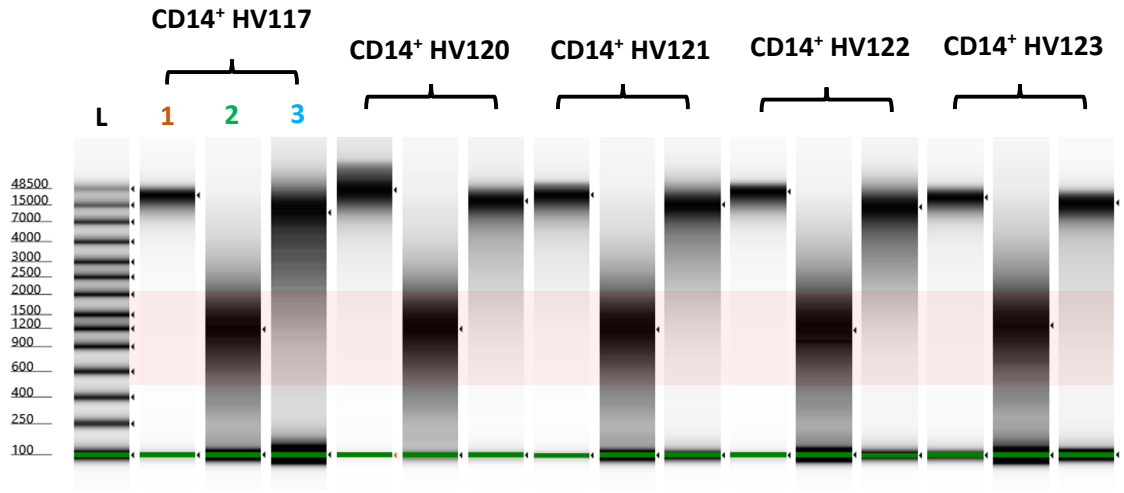

Figure 2.2. Quality control of Capture-C libraries using undigested and digested controls. This figure demonstrates the quality control steps used when generating 3C libraries. As an exemplar, results from five healthy volunteer samples from $CD14^+$ monocytes are shown. Libraries were analysed using Tapestation™. The first lane (**1**) in each sample represents the undigested, formaldehyde fixed 3C library prior to digestion with DpnII. The second lane (**2**) represents the same library following DpnII digestion as specified in this chapter. The third lane (**3**) shows the digested library after ligation. All samples are shown here after reversal of formaldehyde crosslinks. The ladder to the left (**L**) shows the approximate size of the DNA at each stage of processing. Image has been generated on an Agilent© genomic DNA ScreenTape™. The red box shows the ideal smear range for digested fragments between 2000 bp and 500 bp.

Primer Name	Primer Sequence	Amplicon Size
CapC PKDREJ Digested Forward	chr22:46653449-46653472	161bp
CapC PKDREJ Digested Reverse	chr22:46653594-46653610	
CapC PKDREJ Undigested Forward	chr22:46653596-46653612	85bp
CapC PKDREJ Undigested Reverse	chr22:46653664-46653681	

Table 2.3. Primers used in the determination of Capture-C digestion efficiency. Genomic sequence locations (GRCh37) for the *PKDREJ* primers used to determine DpnII digestion efficiency in Ceapture-C and Tiled-C samples. Where the primer pairs span a DpnII restriction site or not are specified by 'digested' and 'undigested' respectively. Amplicon size of the PCR product produced by both primer pairs has also been given.

Reactions were cycled in a Bio-Rad® C1000 thermocycler with acquisition after every cycle. Cycling consisted of an initial denaturation step of 5 minutes at 98°C, then 40 cycles of 98°C for 30 seconds, 60°C for 1 minute, and 72°C for 1 minute. CFX Manager™ software was used to determine the Cq values for each primer pair within both the undigested and digested controls. ΔCq values were calculated for undigested and digested controls by subtracting the Cq values of the control primers from the DpnII restriction site-spanning primers. ΔΔCq values were subsequently calculated by subtracting the ΔCq values for the digested control from that of the undigested control. Efficiency of DpnII digestion was then calculated by the equation: 100 x (1-($2^{-\Delta\Delta Cq}$)). Libraries that demonstrated less than 60% digestion efficiency were discarded from subsequent analyses.

2.3.5. Sonication of 3C libraries

Up to 6 µg of 3C libraries were sheared in Covaris© AFA™ Fiber 130-microTUBEs (Covaris, 520045) to an average length of less than 230 bp using an M220 Focused-Ultrasonicator (Covaris©, 500295) set to a duty cycle of 20%, an intensity of 50, cycles per burst of 200, with temperature set to 20ºC, and for 280 seconds of total sonication time. Samples were sonicated in a final volume of 120 µL. Following sonication, 1 µL of material was assessed for appropriate fragmentation size using a D1000 ScreenTape on an Agilent© 4200 TapeStation. Sonicated samples underwent clean-up and size selection using AMPure™ XP SPRI bead reagent (SPRI beads) (Beckman-Coulter®, A63881). Briefly, 1.8 times the volume of SPRI beads were added to sonicated samples (1.8x) (v/v) and mixed by pipetting a total of 10 times. Beads were allowed to bind DNA at room temperature for 5 minutes. Bead-DNA complexes were pulled down using an appropriate magnetic rack, and the supernatant was removed once clear. Samples were then cleaned twice with fresh 80% ethanol for 30 seconds apiece without disturbing beads between

washes. SPRI beads were demagnetised and air-dried following final aspiration of ethanol for 1 minute or until matte in appearance. Samples were then eluted in 86 µL of nuclease-free water, and the DNA was allowed to elute from SPRI beads for 2 minutes at room temperature before being magnetised once again. Once clear, 85 µL of the eluted sample was taken for subsequent indexing steps.

2.3.6. 3C library indexing – Tiled-C and Capture-C

The amount of sonicated DNA used for library preparation was dependent on the NEBNext® DNA library prep reagent kit (NEB®, E6040) or the NEBNext Ultra™ II DNA library prep reagent kit (NEB®, E7645) kits used for indexing. 6 µg of DNA was used for the former and up to 2 µg was used in the latter. End repair, dA tailing, and adapter ligation were carried out according to manufacturer's instructions. Samples were prepped for subsequent reactions using a 1.8x (v/v) SPRI bead clean-up step after every stage as above (section 2.3.5) and eluted in instructed volumes of nuclease-free water. The NEBNext® Multiplex Oligos for Illumina® Sets 1-4 (NEB®, E7335; E7500; E7710; E7730) and the NEBNext® Multiplex 96 Unique Dual Index Oligos for Illumina® (NEB®, E6440) primer pair kits were used for the addition of indices to Capture-C libraries and tiled-c libraries respectively. Post-adapter ligated libraries were split into two PCR reactions to help preserve sample complexity. Indices were added using the Herculase® II Fusion Polymerase kit (Agilent®, 600675) over provided NEBNext polymerases to aid retention of GC rich fragments according to manufacturer's instructions. A total of six cycles were used in both Capture-C and Tiled-C to add the indices to samples. Samples were then combined, and a 1.8x SPRI bead clean-up reaction was performed as specified above (section 2.3.5) and eluted in 123 µL of nuclease-free water. 1µL of sample was used to each ascertain yield on a broad-range dsDNA Qubit® assay and on a D1000 ScreenTape™ to check library profile (**Fig. 2.3**).

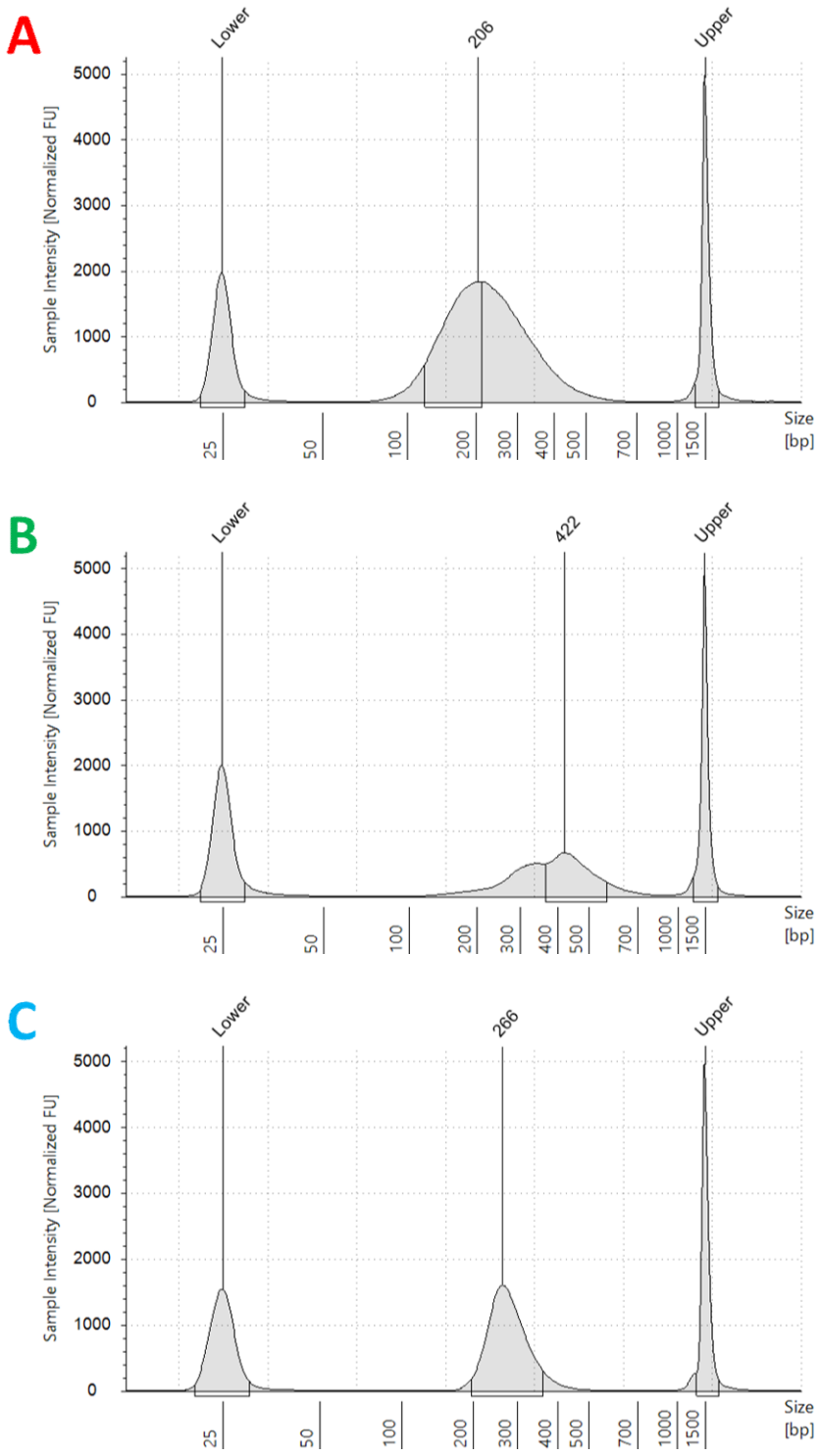

Figure 2.3. Quality control of 3C libraries through the processing stages of sonication, pre-indexing, and post-indexing. This Figure shows expected Tapestation™ profiles for high-quality sonicated libraries. (**A**), pre-indexed libraries following SPRI-bead clean-up (**B**), and post-PCR addition of sequencing indices (**C**). This example shows samples from a healthy volunteer $CD8^+$ T cell 3C library of 2 µg total input DNA. Indices were added with the NEBNext™ Ultra II indexing kit with Herculase™ II amplification. Images have been generated on an Agilent© D1000 broad-range DNA ScreenTape™.

2.3.7. Primary capture of material for Capture-C libraries

For Capture-C experiments, enrichment by oligonucleotide capture was performed using the NimbleGen® SeqCap EZ hybridisation and wash kit (Roche®, 05634261001) with second generation accessory kit (Roche®, 07145594001). Two custom biotinylated 120bp ssDNA oligonucleotides targeting DpnII restriction fragment viewpoints of interest were designed using CapSequm2 (Telenius et al., 2020) and ordered from Sigma-Aldrich®. Oligonucleotide baits for capture were designed around AS-associated regions (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016), and include 47 gene promoters, 52 AS-GWAS SNPs, and 5 control regions with known genomic interactions (**Table 2.4**).

A tiered capture approach where each bait was captured in isolation was employed at AS risk loci with multiple baited fragments to avoid contamination of adjacent bait signals. All capture oligonucleotides were designed in the GRCh37 human genome assembly. Pools of oligos were created from baits reconstituted in nuclease-free water to an exact stoichiometric ratio of 1:1. The final concentration of the pool was 2.9 nM for each individual oligo. Two oligonucleotides were designed for each viewpoint to increase restriction fragment pull-down efficiency. For viewpoints where a single oligonucleotide could not be generated, the oligonucleotide bait was added to the oligonucleotide pool at double the concentration to maintain overall pool concentration. 1-2 µg of up to six differentially indexed 3C libraries were pooled 1:1 (w/w) for subsequent capture. Prior to desiccation, 5 µg of COT DNA (Invitrogen™, 18440016) and universal blocking oligonucleotides (UBO) (Roche®, 08286396001) were added per library to the pooled samples to block repetitious DNA from hybridising to capture regions of interest. Drying was done within a SpeedVac™ SPD2030 Integrated Vacuum Concentrator (SPD2030A-230, ThermoScientific™). Hybridisation followed manufacturer's instructions with modification. Briefly, 7.5 µL of 2x hybridisation buffer and 3 µL of hybridisation

component A from the SeqCap v2 accessory kit were added per library to the desiccated pool of 3C libraries. The libraries were then reconstituted via vigorous vortexing, and the mixture heated to 95ºC for 10 minutes to denature the DNA. 4.5 µL of the 2.9 nM bait mixture per library were simultaneously heated to 47ºC in a thermocycler for 10 minutes. Denatured libraries were then added quickly to the bait mixture without removal of either from thermocycler. Samples were mixed quickly by vortexing and replaced in the thermomixer at 47ºC for a further 72 hours. A strip of parafilm® (Avantor™, 291-1214) was wrapped around the top of each PCR tube to prevent excessive evaporation of hybridisation mixture.

Oligo Name	Capture Tier	Sequence
ACTA2/FAS_Pro_1	1	GATCTCGCGCAAGAGTGACACACAGGTGTTCAAAGACGCTTCTGGGGAGTGAGGGAAGCGGTT TACGAGTGACTTGGCTGGAGCCTCAGGGGCGGGCACTGGCACGGAACACACCCTGAG
ACTA2/FAS_Pro_2	1	CGTTCCCACCGCACAGAACCCGGCGCCTATTATTGGCCAAGAAACTTGAGCAGCCTGTTTTGAA AAGTCCCTCGCTCAGAAATGCCAGCTTGCAGATGGCTAATCAAAGAGACGTGGATC
AHR_Pro_1	1	GATCTGGGCGGGGCGGGGCCGGTGAGGGGTCGGGGGTGCTCCTGCTATTCAGCCGGTGCGCGC GGCGGCGGGAGGCAGTGGCTGGGGAGTCCCGTCGACGCTCTGTTCCGAGAGCGTGCC
AHR_Pro_2	1	AGCGGCGGAAGCCGGTGCAGAAAACGTGAGTGTCCCGAGCGCGTCCTCATCGCGGGGGCTGGG CGCTCAGGCACGCGGGTGCGGGAGGCAGCCCCACCCCGCCCCCAAATCCCTGCGATC
ANKRD33B_Pro_1	1	GATCCCCCGAAGGGAGAAAACCCGCTCTGCTCCTCTGAAGCGGCGGGAGAAACCGGCCCCACCC AGGTCCGGGGGCCCGACGCCCTCGGCGGAGGCCGGCTCCAGGGGTCCGAGGTTCCC
ANKRD33B_Pro_2	1	AGGTCCCTGGCAGGGGACCATTCCTCCTCCCCTGCAGGAGACAGCCCCTCCTGCCCAGCGCCCCA TTTGGCCCCTGGGCCCCTCGCGCCGCCCTTGTCCTAACACGCTGGGGCGCAGATC
ASAP2_Pro_1	1	GATCAAAAAACATAAAAAGAAACAGATACAATTTTTAAATTGTAACTTGTTGTCTAGTGCAATG ATAAAAGCTATCCACGAATACAGCGATTTTAGAAAGTTAAATTTTACATGCAGTGC
ASAP2_Pro_2	1	GGCTGGGGGCGGGGCATGGGCGCGGCGCGCGGGGCCCGGAACGCCGAGACAATGCGGCGCGC GGCGGCCAGGAACGGCTCGGCGGGCCGGGGCGCGCGCCGCCTTCCTTCCTGCGGGATC
ASAP2_rs2666218_1	2	GATCCAGCACTGGGTTTTCTCATTTTTAATGCATCCATTCCTTAAAGCCTCTGTTCACAGTCACAA GGTGTACTTTTTAAAGGAACACAGCACACAAATGTGACCGCTAGTGGACAGCAG
ASAP2_rs2666218_2	2	GCCAGGCCCCCTTGTGCCACTGTGGCTCTCCCCAAAGCAGGGCACCAGGCTGGCTGGGGGGGTC CTGCCCAGAGCTGCCCCACGCTGCCGTGTTTGTGCTGCTGCTGTGCAGTGGTGATC
B3GNT2_Pro_1	1	GATCTCCCGCCCTTGCCCCCGCCCCGCCGAGCTGGAGCTGCTCCCGGACAAGGTAGGGCCCGGG CGCCGAGCACCCCCGCCGGCTCCCTCCCCGCTGCCCTCTCACCGCCAGGCCCCTCT
B3GNT2_Pro_2	1	AGAGGAGAGGGACACCGCCCGCTGCGGGGCTTTGGAACCTCCCAACCCATTCCGTGCCCACCTC GCCCTGGAGCCCGGGGACCTTCCCCCTCCTCGCTGCACCCCAACTTCCGGGGGATC
B3GNT2_rs10865331_1	3	GATCTCTTGATGGAAGAGAGCTCCCGCTCTGGAGCTTCTAAGTCAGCACATAGATAAGCTTGCT GTGCAGGGAGTGTCTTAAGGTCTGTGTCCTTCTGCACACTCAGGGAAATCCATTTA
B3GNT2_rs10865331_2	3	TTCTGCACACTCAGGGAAATCCATTTACCTTTTGCAGCTCCTGGAGGCCGCCACATTCCCTGGCT CATAGCCCCTTCCTTCTTCAAAGCCAGCCACGGCTGGCTGAGCCTTCCTCAGATC
B3GNT2_rs4672495_1	2	GATCTCGTTATGTATCAGTGGCATGCCCTGGCTTTCGAAAAGTTGAGGTTATGTAAGAGGTATT TTCCCTGCCTCTTATCCAGGAGGTTCTGTCTGAGTGCCAAAGTCAGCAAAGCAGTC
B3GNT2_rs4672495_2	2	GATCTCGTTATGTATCAGTGGCATGCCCTGGCTTTCGAAAAGTTGAGGTTATGTAAGAGGTATT TTCCCTGCCTCTTATCCAGGAGGTTCTGTCTGAGTGCCAAAGTCAGCAAAGCAGTC
BACH2_Pro_1	1	GATCCCAGGGTTAGCTAGCGTTTTCCCGGAAGGGCAAGAAAAGATGCTCCACTGGCCCCCACAA AGCTGACCGGACTCCCTTTTGGGTGGGGACCGGGCGAGTAAAAGCAGCCTTTTGAA
BACH2_Pro_2	1	TTCCCTAAACACCTATAACTTGGCCCTCTTGTGTACAGGCAGTATTCCATATTATTACACGTGGT GCTGATTTTTAATGAATATTATAAAAGTTTGGCTCGCTCTACTGAGAGAAGGATC
BACH2_rs17765610_1	2	GATCATAAAGTTAAAATGGAATGAAGTGGCTATAAATAGAAGAATAGTTCCAACTCACTGATGT TAATGGAGTTGCATCCCTCTGCTCTAAGCAGACCCATCTAGACTAAAATCCACTGG
BACH2_rs17765610_2	2	TTTTGAATGAAAATATTTTGAATGCCTCATTTCGAAGAACATTTTTTCATTGCACTTAAATATGTA ACTACCTGGTCTCATACTGTGCAATACTCATTCTGATACAGATGAGATGGGATC

BACH2_rs639575_1	3	GATCTTAGAGACATTTAGCCCAGTGCCTCTCAAATTTTGATGTTTATGTGAGACCCCAGGGGCAC CTTGAGAAAGTGCAGATTCAGACTCAGTAGGTGTGGGTGGCCCTGAGACTTCATT
BACH2_rs639575_2	3	ATGCTCCCTCTGTGGTGAAAATCTTTCTTAAAATAAATCCCTTGGGACAAGACATAGCATCCCAC ACATGGGACATCCAGACCTGAGAGTAGTTGAGTGAGAACATTACATTCCCTGATC
CARD9_Dwnstrm_1	2	GATCCCCGGCGGGCCGCTTCTCCCCGCAGGCCGTGTAGCGTCAGATACGGGTCTGTCCCCGGTT CCGTCCCGCCGCGCGCCTACCTTGCAGGTGTAGACGAGCTGGTAGTGCGCCGCCTC
CARD9_Dwnstrm_2	2	CGCGCGAGTGTCCGAGCGGGAATGCGGGTCACCCGTGCTGTTTATTTACGCAGCTGTGTTTTCT AACACTAATACAATGCATGCATGTATTGTGTGTTACATGGTGAAACAGAACAGATC
CARD9_Pro_1	1	GATCCCCCAGGGAGGCAGCAGAGACACAGTGACCCCGCTCCCCACCAGACCAGCCGAGGGCCC CTCCGCAGCCTTTCCCAGCCCTTGCTCCTGCCCGTTGGCATCCATGTCACCCAGATA
CARD9_Pro_2	1	CTGCCACCGTAATTCCACTGAAATGAACACATCTCCGCCCAGGGGCCGTGAAAGGGACGGGGCC GCATGGCGGAGGCTGCCTGGCTGCCTGTGGGAAGCCTCAGGCTGCTCTGGGGGATC
CARD9_rs1128905_1	2	GATCTGGGCCTGGTGAGGTGACCTGCTTCTGGAGGGGCAGGGCCGCACCCGAGAGCAGGGACA GGTGCCCGAACACAGGGTCTCCAGGACGTAGCGCCCCCCCGCATACTTGAATGTATG
CARD9_rs1128905_2	2	GGCCGGGACCAGCCCCCAGAAGCAGCAGCCCCCAGGTCAGCCCCACGACCCCCTGAGGACCCA GCCTGCTGCCGCCCACCCAGCACCCGGGGGAGGGAGGTCCTTTCCTGGAGTCTGATC
CDKAL1_Pro_1	1	GATCTCTTCGGGCCACCGTCGCTGAGCATGCGCAAATAAACGTGGCGGGACGTATGTGTCATGG CGCTCTCCATCTAAAGTCTGTGCAGCTTCCGGAGAGTGGCGGGTTGATTTTCTCAC
CDKAL1_Pro_2	1	GAGTATTTTAGGCACCTCGTTCCTTATTGTCAGGTTTAACTTCATTTGTTCTCCCACTTTATACCTT AGTGAATTTGTAGATGTGACAAGGCTTTCGCAGTTATATAGCTTTCCCAGATC
CDKAL1_rs6908425_1	2	GATCTTGCAGGTCACAGGAGGAAAAAAGCTCTTCCACTACCCTCTTAGGCTCAGTAGCTGGGGT CTATGAATCAAAATGATAAAGGACAGATTAACAGGAATAAAAACATACATATAACT
CDKAL1_rs6908425_2	2	TTTATTATGGGGCCACACAGCTTGATGGAGCTTGGTAACCCCTAGCTTAGGGCATTATTATATTG AAGCAAAGGACAGTGATAGATGATTTCCAGTAATGACACTGTCTTTACTTTGATC
CMC_Pro_1	1	GATCATTTTGAAAGCCCTGAATCCCTTGCAGGGAAGAAGCAGCTGCTCCTTGTTTTTCTAAACAT CTGTCCCATCTGTCCCAGGCCTCAGCTCACCCTTTCGTGGCTCGAGCCAGTTAAA
CMC_Pro_2	1	CTCTGGCCCCAGAAAAATTTGTATATTTGTTCTGGCGCACAGTTTTACATTTCTGTTTGGGAGTTC ATGGACCCCGAGTTCATCCTTGTAACCCAACTTAAAAACATATTTCAAAGGATC
CMC_rs11750385_1	2	GATCACCCAATTGCCGGGAAAGATGCCACAACAACACAAGAGGGAGGCCACTATCAGTTCCAG CGTCATCCCCACTAGTACCACTAGGGGGTCAGGCCACTCTGGCCACCATGGAATGCT
CMC_rs11750385_2	2	AATAGAGAGAGGTGGCCCCTAGAAAGGGTGAGAGATGCTCAGTCAAAACTCAGATGTGAGCCA GATTTAGACAATGAAGGGCCCAGTAGGCCACACTCAAGCAAGGTTGTGGCAGGGATC
CMC1_rs10510607_1	3	GATCTTTATAGAAGCTTGATTTTCATTTGGCAGCAACTTTTAATTAACTTTGTAGGGTCTCTGTTT TCCTCAAGAAACTGTGAACCCAGCGAAAAAATGCAGGTTACATGCTTATACATA
CMC1_rs10510607_2	3	ATTTGCCATAGAAAATACATTATTTTTGTCTCATCGGATTAACAACTCTCTGATGGTTGCTCTGTG AAATTTGAAAAACACAGGCTTAAGCAAAGACAATAAGAAAGAAGAATTGAGATC
CTL4_Pro_1	1	GATCATTGGTCCTGTCTACAGCTGTCACAAATTTAAGGACTCTGGTTATATTTAATCTTCACTTTT GAATTTTCTGCTTGAAAAATTTGTATTAGAAAAAAAAGTCTATCCTTTTATGGA
CTL4_Pro_2	1	GTCAATGAAATGAATTGGACTGGATGGTTAAGGATGCCCAGAAGATTGAATAAAATTGGGATT TAGGAGGACCCTTGTACTCCAGGAAATTCTCCAAGTCTCCACTTAGTTATCCAGATC
EOMES_Pro_1	1	GATCTTTGTCCCCATCCACCCACTAGCCCGCGCCTTTCTGCCCTTTTTCAATTGCCAAAGCATCGG TCAAGTTGACCACTTGGAAACTTATGGGCTGTCACTAGCTGCTTTATATATTTG
EOMES_Pro_2	1	GTCACTAGCTGCTTTATATATTTGCGGCCAGGGAGGGGCTTCCCGGCCTCACAGCCCGGAGACC CTCCTATTGGCTGGTGGAGGTGACACTAATTCAATTAGGCATTTACTAATTGGATC
EOMES_rs13093489_1	2	GATCCAACACAGAAAAGAGGCAAAGTGAAACCCAAGATGGTGATGAAGAAAGCAGAGGAAAT GCTACATGAGGCAAGATGGCAACCAGTTTATTGTGAACTGGGTCAGAAAACTCCAGGG
ERAP1_Pro_1	1	GATCTGGTAATTTCTGCATGTTCAGAATTTGACAGGAGGTATCAATGGATAGCGTGTAGCCTCC AAGACCTCCAGACATCTCGGAGTGCTCCCAGTTTACTCAAGATGAGACCTCTTTGT
ERAP1_Pro_2	1	CGATAAGTCAGAGGATTTGGCAGCATTCACTTCTAGTTTCAAAGCCCTAGCAGGGATGTAGAGG ACACAGACCTTGGAGTAAGGACTTTAAGGAGTCCTAGTTCTAGAGCCTGTGAGATC
ERN1_Pro_1	1	GATCTTCCAAGAGTGCAATTCCCACAGCTCCACTTGCTCAAGGAAGTCTCGGTGCACAGAAGGA AGCCTGCAGCCTCCCAGTACCAAGAAACCGGCAGGGCAAGTGGCAGAAGCCAGCTC
ERN1_Pro_2	1	CAGAAGCCAGCTCTCGGGTGAGAGTCCTTACTCCCTGTTGCATTCGTTCCCCAAAGTTTTCGAAT TTTTAAAAAGTTCAAACAAGGATTCGAAGCGCAGGACAGACCCTTCCCCTGGATC
ERN1_rs196941_1	2	GATCCTTCTTTGAGACCTCTGAAAGATTTTAAGGTGGTCCCTCATAGACCCTCTGCCCCCATCAA GGGAGTTCTAAAAACCATGTGTTGTCCCATTGCACTCACATTCCAAAGTCTAGAA
ERN1_rs196941_2	2	TCTAATGGAATAAATTATAATTTGACACATGGGAAGCCCACATAGTTTTTAAAGGAGTGGTATC AGCATAGACTAAAAGGAACAGAGACAGTCAAATTGAAAAAAAAGGGCTTCTAGATC
ETS1_Pro_1	1	GATCGACGGCCGCCTTCATGGTGCCAGGAGTGGGGGACGTACGGGATGGTAGCAAGTTTGCAG TTACTGTTGTTTTTCTTTTTAATGAGGATTAGTAACAGGGGGAAGGGGACGGGGGAA
ETS1_Pro_2	1	ACTTTGCGGCGAAGTGAGCGCGCTCGGGTCCCAGCCTCGCCCGCGCCGCGCCCGCTCCTCCTGCC CGGCCCTCGCCCGCTCCCTCCTCTCCGCCGGCGGCTGCCTCGTTCGCTCTCGATC
ETS1_rs7933433_1	1	GATCCTAATAATCCCTGCTCACAACACGATTACTAGCTAATCCGTCCCAGTCGAGGGTTTCTTTTC TCCTTTGAGTCACAAGGAATCCTTATTACATGAAAGATGAAAACATGAAATGCC
ETS1_rs7933433_2	1	CTGCAGCCCAAGCAAGCCCCCGTGGCAGGAGTGTGGCAGGGGCCTTCAGCTCACCCACCTCTGC TCTCTGAACATGGCACAACGAGCCTGTCAGACCACTTGCAGCCCTGCTATTTGATC

Ets2_Pro_1	1	GATCATTAGTCCTAAGCCCATTCAGAGGTTCAAGAATGGGGTCGGCTCAATTTCAGGGCCTTAT TACCCAAGCCCGGCTGCCCTTCGGTGCCACCAGCACCACTGCTCCGTCGCTGCGGA
Ets2_Pro_2	1	TCCTGAAGAGCGCGCCGCGTGGGGGACGGCCCGGTTACTTCCTCCAGAGACTGACGAGTGCGG TGTCGCTCCAGCTCAGAGCTCCCGGAGCCGCCCGGCCAGCGTCCGGCCTCCCTGATC
ETS2_rs283688395_1	2	GATCACAATCTCACCCAGAACCACAAATGAGCCTCCTGGTGTCAGTTACCTTCTTTTTCAAGAAT GACGCCAAGCTCTAATCCCTCCTCACTGAGTCTTTATGAGCTTAAAGTAACTGCA
ETS2_rs283688395_2	2	GCTTCCCAAGGCCCGGTCACTGCACGAGAGCCAAAAGCACACATCAGATGGGACAGTATGCAA TCGGAGGGGACCTGCTCTCCTGGAAAAGCCACCCTGAAAGCTTTCATTGGCCAGATC
FAM118A_Pro_1	1	GATCGCACTTGATGCTCGCCACGTCCCTGCAAAGTGCTCCCGCCCCCTTTCTGCAAATGAGGAAA CGGGACGCGCGGCTCGCCGGGCCAGGTAGGGCTAGGACGCGGACTCTAGGCCCAA
FAM118A_Pro_2	1	GCGCTCCCCAGCGCAAGCCCTCCCCCGCACATCCGACCGCGCCTGCGCTGGAGACCCCGGGGGC GTGGCCTCCCCGGTGGGCGGGACCGTGGTCACCGAGGTGGGCGGTGCCAGAGGATC
FAM118A_rs1569414_1	2	CTGAGCCTCTCTGAGATATTTGGTTAGCTGCATTGGCGATTCCCTCCTGCCTGCCTGAGTGACCC CATGCAATCTGCTTTCCCTGTGCCCACCCGACCCCGTCAGTCATGGTCACTGATC
FCGR2A_Pro_1	1	GATCTGAAAAGTAACTTAATAAAATCCTAACTGTAGACTCCCTCTCCCTTGGGGTCAAACTAGAA CCCCACTGAGGAAGGAACTCCTACAAACAAAAAAGCAGGAGTAGCCCTTCTTGAG
FCGR2A_Pro_2	1	TTCCTTTGCCACCTGCATATGTGAGGATAACATTCTTTTGCCTGTATAGAAATGAAGTGTGCATT TAATTTTCAGTGTATGGGTATTCCCCTAGGAATCTCTCTAGTCCTAGCCATGATC
FGFR1OP_Pro_1	1	GATCTTGTGCGGCGATGGGATGCGGCCGTGGGCGGTGGAGTCCAGGCGGAGTCCGCTCTCCAC ACGCGAGGGCGCCTCCTCCTCACCGCTCGCCGGCTTCAAAGCTCGCGCCTCAGGCGA
FGFR1OP_Pro_2	1	CGCAGTGGTGGCCGAGGAGGACACGGAGCTGCGGGACCTGCTGGTGCAGACGCTGGAGAAACA GCGGGGTCCTGAACCGCATCAAGGTGAGGCCGGAGGCTGGGGCCGGGCCTGGCGGATC
GPR25_Pro_1	1	GATCTGCTGTGTCCCGGCAGTGGAGTCTTTCCTAGGGAACAGGTGGATTTGGGGTGGGGCGTG CCTTAGGCCCCATACTCATAAAATCCCGGGGCTCTGTCTTCTCACCTCTAAAACAGA
GPR25_Pro_2	1	TGCCTGGCCTTCGTCAACAGCTGCGCCAACCCGCTCATCTACCTCCTGCTGGACCGCTCATTCCG AGCCCGGGCGCTGGACGGGGCCTGCGGGCGCACCGGCCGCCTGGCGCGAAGGATC
GPR25_rs41299637_1	2	GATCGAGAAGCTGTTGCAGTCCTGCCCAGGGGAGTGCCCAAGCCAATAAGGAAATAAATAGGC CATATGGGCATGGCAGGAACAAGAAAAATAGGTAAACAAAAGATAACTGCAGACATG
GPR25_rs41299637_2	2	GGCTGGGAGCCCAGAACGGGCTCCAGTCCAGAACAGCCCCTGGAAGGAAACCAGCCTGGACCA CCCCTATGAGAAGCCCAGGAAGTCTTCTGAGCCCTGGAGCGAGTCCAGGTCAGGATC
GPR35_Pro_1	1	GATCTCCGTGAGGCTCCCCTGCCTGTGTCACCCAGGGGCCTCTAGGGGTGCTCTGGAGCTGGGA GCCACTGTCCAGCACTTCCTGCCCTCTCCGGCCTCTCTTGGCTGTCTGCCAAGCAG
GPR35_Pro_2	1	CCATGGGGTCCAGGGCTGTGGGGTCCCTGGCGGTACTGTGGGCCTGCAGAGTGGGGCATGTGG GCTGAAGACCGTCTCCCCACCATGGTGGGAAGGGACAAAGGGTGGCCCTGGCAGATC
GPR35_rs4676410_1	2	GATCAGGTGCCGTGCCTTCTGCTGACCCCAGAGGCCAGACCCTGTGGCCAGAGTGAGGGATGG ATGAAATCCCAAATCTGAGCCTCCTGCCCGAGGATGAGGCCCCTCTCCAATGAACAT
GPR35_rs4676410_2	2	TTGGTGAACTCGTGTCACTCACAGCCCCATCCTGGGTCCTGGGGTGTGGTGTGCTGTGTGGGGC CCAGGCCCAGTGGGGTCACAGGTACGGGGGACTCTGGTGCCTGGGCCACAGGGATC
GPR37_Pro_1	1	GATCCAGTTGCAAGCCGAGGTCTCTAGAGGAAAAGAAAAACAAGTCCCCCTCCCCCAGACGACT CGCAGTCACAACCCCGCTCCCAGCTCGCTGGAAACACGTTGCAAACAGTCCGAAGC
GPR37_Pro_2	1	GCCTCGCCCCCGCGCTCCAGCCGCTGCGCGCGCCCGCGAAGCTCGACTCTCGGCTCGTTGCCTCC TCCCACTCCTTCGCCTCCTCCTCGGAGAGATACTGGTTTGTGGGGAGACGGGATC
GPR37_rs1525735_1	2	GATCACACCTTTATTTCACTCATCCAATCTACTGGACAACTTTGAGCTTTTCTAGATTTCTTTTTG GGGGACCGACTGTGACTAGATGTGTGTTAAAATGGACAATTGTTCTGTTTCCTG
GPR37_rs1525735_2	2	ATATAGAATTAATGGGAAGATTTTTAATCTCACATGTAGTCTAATGCTTCCTGCTTCTCTTCAAGT TCTAAGAACTAACAGGCAAGTACTTTTAGTTGTAGGTGAAGTTTAAGAGAGATC
GPR37_rs2402752_1	3	GATCTGTCTAACTCTTATTTGATATTGCGTGGCTCTTCTCTATCCCTCTCTCAAGCCTCCTTCCCCT ACCACCCAAACTACCTCTTTCCGTGTGTCAGTCAATAACAATTGCCCATAGTC
GPR37_rs2402752_2	3	TCTTCTCTATCCCTCTCTCAAGCCTCCTTCCCCTACCACCCAAACTACCTCTTTCCGTGTGTCAGTC AATAACAATTGCCCATAGTCTTTACTTTGCAACAAGTTTTTTTGGTATTGATC
GPR65_Pro_1	1	GATCATCAGAGTAATGAAATACCTACGTAACAATGCACTTGTACCCCGAACTTAAAAAAAGATA TATCATGTATAGAAACCACAGATAAGCAGATGTACGAAACCAAGCAAAAATGAGAA
GPR65_Pro_2	1	CTAAATTACTTTATAAACTATCCTGTGCTTTATAAATGTAAGGTATTAGTTTTATTATTCTCAAAG AAAGACCTACAACGGAAGTATTGAAAATCTTTACATTCCTGATATACCCTGATC
GPR65_rs11624293_1	2	GATCACCAATTATACTGACTGTGGTCTCAGAAGCTTCTGCCAGGGGGATGCCTAATGGACAGAC AGGACATCATCTCAAAGGCCAGAGGGGACCAGAAGACAGTATGAGGAGCAGCAGGT
GPR65_rs11624293_2	2	TCCTTTATTGGGTGGACAGCAGTGCTGCAGTGAACGTGCTAGGAACAGAGAATGAGCTCACTTT GAGGCATAACTGTTCTTTCAGTTCTAGCTGACCTGGAAGATTGGCTACTGTTGATC
HHAT_Pro_1	1	GATCCAGAAGCCCGAGGTTTTAATGAGAGCCACGTGATTCTGCTGAGATAAGTCTGGAGAACTC TGGCAAGCCTGTCTCTTGGCTCAGGCTTGGAGGCCTCCGAGCAGCAACATCGTCCC
HHAT_Pro_2	1	CACGTGATTCTGCTGAGATAAGTCTGGAGAACTCTGGCAAGCCTGTCTCTTGGCTCAGGCTTGG AGGCCTCCGAGCAGCAACATCGTCCCAATTATACCCCGTTGGAGCATCTTCAGATC
HHAT_rs12758027_1	2	GATCTTCCTCTGTCTCCCAACTTCCTCTCACCCTGCCTCCCTAGTTACTGCTAGTTGAGTCAGTAA TACCCAGAAAATATCTTCCTCTGCTTTAGAATGCAGAAAGTGGTAGATTAATGT
HHAT_rs12758027_2	2	TCCTCTTGTACGTGCCTGTGGCCTCCTAGAGCAGGGGCAAGACTAGAGGTAATCTTGCCTGGTG GCCTCTAAGTAGAGACACCCTACATCACTGGAGAGCATTCCAGAACAACCATGATC

ICOSLG_Pro_1	1	GATCGGCTGCCGGGCTGCGGGGAATCCCACGTGTGTGGGGGAGGGCGTGTCCGCAGACCGTGG GCACCAGCTGGCCTTGGAGAGCGCTCGGCCGGTCACCTCCCTCCTTCTCCCTTCTTC
ICOSLG_Pro_2	1	CCGCATGGTGCGGGCGGAGACCTCGGGCCGCGGAGACCTCGGGGCGCGGAGACCTCGGGGCG CGGAGAACTTGGGCCGCGGGAGACCTCGGAGAGGAGCAACCTCGGGGCGCGGGAGATC
ICOSLG_rs7282490_1	2	GATCCAACTTCCAGGATAAAATTTTAGAACCAACATTTCCATTAGCAAAAGCGGTCTTGTCTAAA ACTCTACCGACCGCGCTAAAATCCCATTCAAAACAGAGCTCTGGTTACGTAGATC
ICOSLG_rs7282490_2	2	GATCCAACTTCCAGGATAAAATTTTAGAACCAACATTTCCATTAGCAAAAGCGGTCTTGTCTAAA ACTCTACCGACCGCGCTAAAATCCCATTCAAAACAGAGCTCTGGTTACGTAGATC
IL12B_Pro_1	1	GATCCCAGCTGTCATTTTGGTTGATATCACAACTGACGACCAAGACCATCACAAATATGGGAGC AAGTCTGATTTGTAACATTATTATAATTATGAATCCAATTACTTTAAGGAATGCAC
IL12B_Pro_2	1	CTATACAATCCTTTAACTGATGGGCATTAACATTTAGGAATATACTTTTATACATCAATACACAC CACTTCAAAATAAGATAAAACAAACATTTCCACAAACTATGCATACACACAGATC
IL12B_rs6556416_1	2	CTCTTTAGGACTGGAAACCACGAAGTCAATTGAATTGAATGCACCACAACCCAGTGAGTTAAAT CTTTGTGGAAAGATTCCACAAATGCCTCTAAAGTTGCATCTATAAGCTTAATGATC
IL12B_rs6871626_1	3	GATCCAAGGGACAAGCTAGAGGCCAATAATCGGGCTGAAGGAATGTTTCGTTTGCCAAAAGAA AAAGCAGGGAAAAAAACCAGTAAGCAAGCTGTTGGCATATGTGTGATGACTGGTCTC
IL12B_rs6871626_2	3	ACATAGCTCCAACTGCCATTATGGGCTAAGTGGGTGCCTGCTGGATGGAGGCGCCTGCCCAAGT GATGAAGGACAGGTAACTTTCTCTGCCTCGGCTCACCTCTCTGGAACACCCTGATC
IL1R2_Pro_1	1	GATCAAAAGCATCTTTTTCTCTTTAATGGGGCCACAGTTGGGCAAAACGCCCATCACTTTAAAAC CACCTCTCGGCTGGAAGTACGTAATTTTTAGCGAGTCACAGAAAAATAGGGAAAC
IL1R2_Pro_2	1	GCAAAACGCCCATCACTTTAAAACCACCTCTCGGCTGGAAGTACGTAATTTTTAGCGAGTCACA GAAAAATAGGGAAACTTATGCGGCGTTTCCTTGGCCACTTCCCCATCTGGGTGATC
IL1R2_rs201423816_1	2	GATCCAATTCCTCATGCATTTTTAGGGTTAAAGTAATGGTCCACATTCAAGTACGTGACAAGACT TAGTTTTCTTCTGGACCAGTGTTAGAGCTTAGAACTGTGGAGAATAGAGGTGGTA
IL1R2_rs201423816_2	2	TCTGGAACACTTTATTTATTATATCTACTGTGTGCACACTTACTGTTCAAAATCCTGATGGCATAA CAATAAAACCAAGGTGGTTCCTGACTTTAAGGAATAGAATATGGGGAGAAGATC
IL1R2_rs4851529_1	2	GATCAATACTATAATGACCAAATTATCAAAATTGTAAGACTTAAATGTTTACAGTGTTAAGGAA ATATTATACATTGTTTTGTACAATTTTTATGGTTACAATATTGACATTTGTTTGCA
IL1R2_rs4851529_2	2	TATGGTTACAATATTGACATTTGTTTGCATCTAGACAAGGAGAAAACTTAAAAACTTGGACTTTG AAGGCAGACTTTGTGAGTTCAAGTTCAAGCACTGCCACATGGTGCCTCTGTGATC
IL23R/IL12RB2_rs11209032_1	2	GATCCAAGCTGGGCCAATCAAAGGCTTTCCCTAGGAATTGTGCAATTAGGCCTGCGAAAGAGAC AGACTCCTTCCAGTGACAGAGTGTTAGGTGGCAAGTTCAGAAGCTGTCAGTCTTGT
IL23R/IL12RB2_rs11209032_2	2	TTCCAGTGACAGAGTGTTAGGTGGCAAGTTCAGAAGCTGTCAGTCTTGTTTTTCTCCATGTGGCC AGAATGAAAGGAAGATGGCCCATAGACGCAGAATAAGAAGAATAATAAACAGATC
IL23R/IL12RB2_rs12141575_1	3	GATCTACGGTAAGGGTGGGTTCAGCTAACGACTCTTGGTGTCTTCCAGCGCTAATGATTTATAAT TAAGTTAGATTTGTAACCTTAAAATACTTTATAGCATTTACCCTGCTTGTGAGTG
IL23R/IL12RB2_rs12141575_2	3	AGAAGGCGTTTATCCATCAGGGCTCTTTTAGTACCAAAAAACAGAAAACCTGACTCCAATGAAC TTAAATGAAACTGAAAGGTCCAAGGATAGATTTTTAACTTCAAGCACACCTTGATC
IL23R/IL12RB2_rs924080_1	4	GATCCTCTTCGGATTATTGCCATAGCAGTGTGTACATTTTGTAACTTATTCTTGCACATATCATGC TTTTAGAAAAATCTTTATATCTGTTTTCCTCCCAAAATTGAGAACAGACTGAGG
IL23R/IL12RB2_rs924080_2	4	GGCACATTCAAATCCTGAGGCACTGCTACAAAACAGCAACTCACTTAGTAGCCCAGCTGGGAGC CTAGCAAATAAACCTCAACTGATTTATATATTGGTCCTACCAAAGTGAAAAAGATC
IL23R_Pro_1	1	GATCATCCAAATACTTAAAGACAGCTATATTTCCTTTTGGAAATACATGGGCCAAGTTGTTTTCT CTTGCAAGTTCTCATGATATTCCATGCAGTGGTTACTGATGAAAAGGGAATCTAT
IL23R_Pro_2	1	TTTAAATAATAATTATTGCATATTTCCATCAGTTAACTTCTAACTAGGAACATTTTTAGGGCTTCT GTTGCTTCTTATGAAGAGAAGAATTACTATTGCTAACAAATAATGAAAGGGATC
IL27/SULT1A1_rs35448675_1	2	CTCCACAGCGGGCCAGATAATAACAGATGGAGATGAGTGCCTAAAAGGAGACAGTGAGAATGA TGGGATGAAGTCACTGGGCCCTCCCCTGGGGCTACACAAGAGGAGAGGTGTCAGATC
IL27_Pro_1	1	GATCCTGGGTCATTGGCTCTGGCCTTTGAAGGCTCTGTCTCCATCCTCCAGCCTTTGGTCCCTACC TGCCCTACGTGGCATGTTACCTTCTCCACTGCCTACTGCTCTGGTTCTAGCCCC
IL27_Pro_2	1	ACGGGAAGCTGGGGTTTCGGGACATCCAAGCTCCAGACTCACTGCTAAGCTCCCACGTCCAGCC CTGTCCTCAATGTCCCATGCTTGGGTGAGCCATCCAGCCCTATTCACTGGTTGATC
IL6R_Pro_1	1	GATCCCCTTTTCTGTGGCTGCCTTGAGGCCCCGCGGGTACCTAGCTGTGTGGTCGCGGGTAGTG GCTCTTGCGCCCCCTGCTGCGCTTGCTCCCTTGGTCCGGAGCGCTCCCCGGAATGC
IL6R_Pro_2	1	TCTCCCTCTGCATTGGGGGTCCACCTCGCCTGGTCCTCCCCCCACCCCCGCGGGTGCATGTGTGG ATGACAGTGCCCTGACTATGAAGGGACTGACATCACGGGCGCTGAGTCACGGATC
IL6R_rs4129267_1	2	GATCCTACTTTGTATTCGGCACTGTGTAAGCTGCTGGCAGTGTTACAGTGAAAAAGACACATGT GGTCCCTGCCTTCAGGGAACTCAGTCTAGAGGTGGAAATGGAGAAATACTGGGAGG
IL6R_rs4129267_2	2	GGTCCCTGCCTTCAGGGAACTCAGTCTAGAGGTGGAAATGGAGAAATACTGGGAGGGGCACTT GCTCAGCTTGGAGTGGGGTCAATTCTCAAAGGAAATGACATCACCTCATCTGAGATC
IL7R_Pro_1	1	GATCACTTAGTTTCTCTGCACTCCTCACCTTATAACTGCAGTCTTGTGGCATTTTTCCAAAAGAGA GCCACTGTGAATCAAGCATAGAGTTTTATGGCCTGAATCCAGGTTAGATTATTT
IL7R_Pro_2	1	ATAAAAGCAGCTGTAGATGATACACAAATGGGTGAGGCTGTATTCCAATGAAACGTTATTTGCA AAAACAGGTGGTAGATTAAATTTGGTCCCAAGGCTTACTTGGGAAAAAAAAAGATC
IL7R_rs11742270_1	2	GATCGCAGGGGCCACATTAAAGCCTGTTTGACACACAAGCTAACTCAGTCATGCCGTTTTTAGTT CAGTCTCTTATTAAATCTTTAGTCAAATTCAGATTTCTTCTCCTCCCATTCTCCT

IL7R_rs991570_1	3	GATCTTTAAATAAATGACATTTGGGAAAATGGTTCGTTCTGCCACCTGTGTGGCGTGGGGTTAG TAAATTGTGAGCTTGAATCTCTCACATTGCAAAAGTCTTGTAAAAACATCTTGGCA
IL7R_rs991570_2	3	GTTTTGTTCCTGTACCCACATTCAACATCACCAGCTTTCAGAGACTCTGTTGGAGAGAACCACCT CCTCTTGCACACGTTTCCCTCTCTGCAGGACCATGCCAACTTCTTCAGTTTGATC
INPP4B_Ctrl_1	1	GATCCCAGGCGGGTAATTACCTCTCCCGGAGGCGGAGTGGGGGGCGGCAGCAGCAGCAGACAC TTTTAGCCTGACTTTCCTGCGTTCGCTTGCGAGCGTGTGAGCGTGTGCGCGCCCAGG
INPP4B_Ctrl_2	1	GGCCCGGCTGCCCACCCCGAAATTACCACGCTGGCTCCGTGTGCTCACACCCCGTACCCCGACCC CTTTCTGCGGCCCCTCTGCCCGCTGGGTCGCCCACCCAGACTGGGCTGTGGGATC
ITLN1_Pro_1	1	GATCCCGGAAGTCATCCAGCTGAAGGAACTGCTGTCCCTAGCAGCGTGACTCACCTGTACTCCA TCCTCTGGTGGTCGCTATGAGAAACAGCAGGAAGCTGAGTTGGTTCATTGTAATCT
ITLN1_Pro_2	1	CAATGCCCAGTGCAGGGGAGGCTTCTGGTTATCTCGGGGAGGTTATTTATACAGTGGAATTAAC AGGTTCTAAAGAAAAAACCAATCCTAAGAATCACCTGCTAGCGTTGTTCCTGGATC
LMO2_Ctrl_1	1	GATCACAAACCTTAGGAAGATTTAAAGGAATCTTATAGTTAACATGCAGCCTTAGCTGCATGTTT CTTGTGTCTTAATAACTGAACACATTTAGACTTCTTGCATTTCCGAGTACTCTAA
LMO2_Ctrl_2	1	AGTGTCAAGCTCAGCATCTTTGCCACATTTCAAACTAGGTAATTACCTTCTCTCCCAAAGGGAAT CTGCTTTATGTATTTGTGTCCTTTCTAGGTTACAAGTTTCCCAAAGGCAATGATC
MYC_Ctrl_1	1	GATCCTCTCTCGCTAATCTCCGCCCACCGGCCCTTTATAATGCGAGGGTCTGGACGGCTGAGGAC CCCCGAGCTGTGCTGCTCGCGGCCGCCACCGCCGGGCCCCGGCCGTCCCTGGCTC
MYC_Ctrl_2	1	CCGAGCTGTGCTGCTCGCGGCCGCCACCGCCGGGCCCCGGCCGTCCCTGGCTCCCCTCCTGCCTC GAGAAGGGCAGGGCTTCTCAGAGGCTTGGCGGGAAAAAGAACGGAGGGAGGGATC
NFKB1_Pro_1	1	GATCAGTATTTTATGAGCTGTGAACCAAAACCATTGCTACCACCATCACTATAATTCTATCCACA GTAATTATCATAAAGGCCTAACAATGCCTTGTAGATGAACATTCTGAGTAACTGC
NFKB1_Pro_2	1	ACAAGTCGTTTATGATAACGAAATTGTAGTTTAATCTGTGAAGAGATGTGAATGTAACTGAGAC ACGCTTAAATGGAATATACAGATGAGCTTTATTTTTATATCTGGCATGCTTGGATC
NFKB1_rs3774937_1	2	GATCATGTCTCTGCAGTGAATCCCAGTTGTCTTCACAAGCTGATTCTCGCACACCTCACCCCCTTA CTCCCTCCGGTGTTCCCCAACCATCCCCTAATCTCAAATCCTCCACACCGACTT
NFKB1_rs3774937_1	2	ATTGCTGGTTCACATAATGACTCTTTTTCTTTTAAAACATAAAGGACATATTTTTATGGATTTTTG AGGATTAATAGAATAAAAATCTAAAGTCATATTCAGATTCAAGAGAACTGGATC
NKX2-3_Pro_1	1	GATCTGTGCAGCCTCGCCCGCAGCGTCTCGGACTTCAGGACTCTATACTAGTCCTTGTCTATTTCT GGTAGAGCCGCAACATCCCACTCGCAGAGAGCAGGGGGCGTGGCCCACATGAGC
NKX2-3_Pro_2	1	CTCTCGCACCCGCAGGGGCAGGAACTTTGGTGGAGGCCAGCTGGGCGGTTTCCTCCCTTCTCTG CTTGGAGCTCAGCGTCCTTGAACCCGTGGCACTCGGTAGAGAGAGAGGAGATGATC
NKX2-3rs11190133_1	2	ACTCCGCCACCAATAGAAATCATTGATATTGTTAGAGAACATTACTGCTATGGCAGATACCTTGA AATTCTATTTACACTCATCACTGCTTTGAAATTATGGTTGTTGTTTAGACAGATC
NOS2_Pro_1	1	GATCTTGCTGTTACATCACTCTGTGTGGTGCCTCTTCACCCAACCCACCTCTTTCTGGTCCCAAGT CCAAGGTGGCAGTTCATCAGCAGGGTGGCTGCTAAGATAGAGGCACCACGGAGC
NOS2_Pro_2	1	TATTGGCACAGAGAGGCCTCTGAGATGTTGGTCTTACTAACACCTCATCCTTTAGCACTAGGTAA GCTTTGGTGGAATGGCAGGTAGGATTAAATTCCTTCCCTTATAATGGAAATGATC
NOS2_rs2531875_1	2	GATCAATTGCTGGCCTCCAAGCAGGGCCTGCCACATGCAAAGCAGAGTGATGCAGTTTCCTCCT TCACCCCAGAGATTGGACGCAGTAATAGTAGGGCCATGTCACAGTCCCTTTCTGAA
NOS2_rs2531875_2	2	GTAGGGCCATGTCACAGTCCCTTTCTGAATAGAGCACCCCCAGATGGTGGCTTTCCTACCCACAG GCCACCACGTATAATACCACTTGACCCCATCAACCTGGCTCCAAATTCCTGGATC
NPEPPS_Pro_1	1	GATCAAGCCCTTCCCGGCTCCTTCCAGTTGGAAGGCCGCCCCCAGTCCTGAGCGGCGCCGACAG GTGCAGCGCCCGCCTCCAGCCCTGCGCTCGCGCGAGCAACCCGACGCCCTCCCGCC
NPEPPS_Pro_2	1	GCGACGAGGGGAGCTTCTAGAAGGGGGGGGAGAGGGTGAGAACGAGAATGTTAATCCCAGGC AGCCTCGCTGACTTCTCCCCGCCTGCTCCAGCCCAGGGGCTGGGTTTCCAAGATGATC
NPEPPS_rs4375701_1	2	GATCTGCACTGTGGTTGATGTTTTTATTATCATCTCTTCCACCCTCCCATTTTTCAGAAAAGGACC TGAGTTCCAGAGAG
NPEPPS_rs62076549_1	3	GATCTTGCTCCACTATTTCCAAACTTTATTAGCCAATTACTTAAGTCATCTGCATTTCAGTGCAAA CCCATCTGTGGAATGGAGATAATTTCACAGATGGGTACACTGAAATGCAGAGAA
NPEPPS_rs62076549_2	3	ATGTATCTCTTTCACAGTTTGTCCTGTAGAAAACGTGAGATATATATTTGTTGCCTCCTGGACGG GTTGCCAAAAACACTTTAGATGTTATTCTTAGATTCAAAGAAGACCAATGTGATC
NPM1P17_Pro_1	1	GATCTTTGGCTGTCTTCTCATATTCAAGAGTAAGGCACTAAAAAGCTGATAAAAAGCCCTGTGT GCATGGGCCGTGCTTGTCCAAGAGCAGAACTCCCAGTTGTCCATTTCTGTGGGTCC
NPM1P17_Pro_2	1	CCCACTGCTGCCCCCTGCATCCCTAGCTTCTGCCTTTGTCCTTTAGGCTCCCTAAACCCACCTGCC CCTTTTTTCAAGGCACATCTTCACTGCTACCTCCTCAGGAAACCTTTCCTGATC
PPP2R3C_Pro_1	1	GATCCTTAAACCACATTCCACTCCGCCTCCTTTGTCGGGAGGTTTCACAGAAGCGGAAAAGGCG GCTCCCCCTTGCTCCTCCTCCCCACCTCCCCCCTACCTCTAGGAGTTTAGGGCCGG
PPP2R3C_Pro_2	1	ACTTCCGCGCAGCAGCTTCTGCATTTGCTGTTTCCTACCTGCTCCACCCCTACCAGCTCAGGCGTC AGCACCGCCAGGCCCCGTAAAGGTTAGGAAGGCCGTCCAACCACCATCTTGATC
PPP2R3C_rs8006884_1	2	GTACAGTAGGTGCACTGACAATGAAATGAACAATAAGCTTTTCTATTTGCCCTGAGGCCATAAT GCAGATGTTTTAATTTACGTTAAGACCAAAGACTCTGAATCTCAAGTAATTTGATC
PTGER4_Pro_1	1	GATCGCAGTTAGGCAGTCCTCCCAACCTCACCAAGGCTCCACCTCTCTCCAAAGCCGCAAGGCTC TGCCACCTGCGCCGGGAGAGGCTGCAATCACTGTCTCCTCCTCTTTCTTTCTCTT
PTGER4_Pro_2	1	CGCTGGGCGCACCCAAGTCTAACCCCGGGGCGCACGCGCTAGCGCAGACACCGTATTTCTTCTC CTTTCTCGGCCAACCCTAGGTAGAATCCTAAAACAACTGCCCTCTCTTCCACGATC

PTGER4_rs12186979_1	2	GATCACAATCAACATCAACAGTGCAGTATGAACCTTTGACGTATGAAAACAACATCATTTTATCT TTGTAAGCTTCCTCTCCAAAACCTCAATCTAATCATGAAAAAACAGACAAATTTA
PTGER4_rs12186979_2	2	ACTATGTGAATTGCATTTTGATTGAACTAGAATATCAATTTTACTAATCAATAATCCCCATACTG GACATCACATATTTCTCAAAAAAATTTTTTGCAATATATTCTCCATGAAATGATC
PTGER4_rs1992661_1	3	GATCACCGTCATACTTTGTCATGTCTTTGGTTCTCCCATCATCATGATTTTGCATATCCTGATAGC ATCTGTTCTTGCCA
PTGER4_rs1992661_2	3	ACAAGTATCTCCAACCTTACCCTAAGCCTCTCATATTAGGGCCTAGAAACCCACATTTTTAGCAA GCTTGTCAGGTGATC
PTGER4_rs9283753_1	4	GATCTTAATTTAATTTCACATTAAGTCACTGAATTGTAGTGTCATAAAAGCCTCTTCATGTGCTGT ATTGGGTTAATTCAGTATACTGAATTTTTCTGCAATAAAAGTATCTTGTAATAG
PTGER4_rs9283753_2	4	TAAAGACCTCAGCCATCATCAGGTACAACCTCTGAGCTGATACAGAAATCCCACTGTATTCCTTT TTCTGGAACACTTGCTCTACCAGGCTGCTCAGTGCTACCTACTCTGCACAAGATC
PTPN2_Pro_1	1	GATCCGGGGAGAGCGCTGGCGCTGCGGCGCATGCGCGCTGCGCGCCGCGCCCCGCCCCCACGC CAAGCCCTGTCGGCCGTGGGGAAAGTACCTGGAGCGTCGGACGTCAGCGCGCAGACT
PTPN2_Pro_2	1	CTGCTGACAAGATATGACTTGGTCTCTCCGACATATGTGACTCCCAACCCAACTAGTTCCTAGCT TGCTGTTTCTAAAGGACCCCAAATGTAGACTTTTGCCCCTTTTGAAGTCAAGATC
PTPN2_rs45450798_1	2	AAATAAGCAAGGTACTTTCTTGACACGTGAATTGATGACTGTTAGTTTTCTTTCCTTCCTTGTTAA CTAGTCAGCAGGAGCAGGAGTGGGAAGACAAGGGGGTACCACTTGCATCTGATC
RUNX1_Ctrl_1	1	GATCTTCAGCTAATCTTAAAACAACAATCATAAGAAGGCCCAGTCGATGACACTCAGGGTTCTA CAGCTCTCCCACATCTGTGAACTCGGGTTTGGGGATGTTGGTTAAGTTTGTGGCTG
RUNX1_Ctrl_2	1	TACTCCTCCACCGTCGCCCAGCACCCGCCGGTTAGCTGCTCCGAGGCACGAACACCCACGAGCG CCGCGTAACCGCAGCAGGTGGAGCGGGCCTTGAGGGAGGGCTCCGCGGCGCAGATC
RUNX3_Pro_1	1	GATCCCAGACCCCCTATCCTTGCTGGGAGACACAACCCAGAACTCTGCATCCCTGCAGCCACCAA AAAAGGCAAAAGCTGCCCAGAAACACTTTCTCCTCTTCCGTGGCCAAATCCACGA
RUNX3_Pro_2	1	AGGCCTTCCCAGTGTCAACCCAACCTCAGCTCACAGGATGCGAGAAGCCTGCTCGCGGCCTTGG CTCATTGGCTGGGCCGCGGTCACCTGGGCCGTGATGTCACGGCCTTTTAGAAGATC
RUNX3_rs10751775_1	2	GATCCCTCCAGGGGTGGGCTCTGAGGACGGCACCACGCGCCTAAGCAGTCTGATAGTCTCTGGG GGAGAGCCAGAGAGGTGGAGAGACCTGCCCAAGGTCCCCCAGCTTGCTGGTGGAGG
RUNX3_rs10751775_2	2	TCGGCCTGGGGAGATTAGCTGACTGAGGGGGGAGAGCGCAGGCTGGGATGCATCCTGTTTCCC CAGGAATGGGCTCGCTGCCCCCTTCTGCCCAGTGTCATCTTTGCTTAAATAATGATC
RUNX3_rs1609996_1	3	GATCCTTTCTCTTTCTTGGTTTACTCCCTTTTTTGATGGAGCATATCTTCCATTAGCTTCCTGAGAA AGCACACATGAGAGGTAGATTTATTTGGGAGCTTGTTGCATTTGATGTCTTTG
RUNX3_rs1609996_2	3	CTTTCTTGAAGTTTTGAAGATATTCTCTTTGTCTCAAGTGTTCTGAAATCCCATAGTGAGATGCCT TCATGTAATCCTGTCTTCTCTATTGTTCTGGACATTCAGTTAATCTTTTTGATC
RUNX3_rs6600247_1	4	GATCGAATGAATGCCCTGCTTTTCTGAAACATTGAAGAAGAGAGAGCTCAGATAGTTGGGCTTG GCATTTGTTTCATTTCTTTGATAGCTATGTACTGTGTACACTTACTAGGCGTCAGG
SH2B3_Pro_1	1	GATCCTCACTTCTATCTTCAGACTGTGGGACGCGAAATCGGGTGGGCACGTGGTGGGCGTGGCT TCGCAAGGCCTCGCCTTTTTCCGGACCCGCGCCAGGAGAGCCCTCCGACCCTGGCA
SH2B3_Pro_2	1	GGAAGCCCCTCTCCCCAGTTCCTGCACTGCCGAGGTCGACCGGGCCAGGCCCCCGCATGCTGCT GAATTGGGCGGTTCACCAGGGCTAGGCGTCCCAGAGACAGGTTTCTGGTATTGATC
SH2B3_rs11065898_1	2	CAGGTGGAACAGTCCGGGCGGCTCAGGAGCATCCCCTCAGTGAATCTGTGTCCAGGTGATGGG CCAAGGGTGCTCAGACCACAACTCTTGCATGTGCTTGCCAGGTAGTCAACTGTGATC
SOX4_Pro_1	1	GATCCAGTCAGATGGCTACATAGGGTGTTCTTTCCAGTAAAATTCTAAGTCTGACAACCTGGGC CTATTGCCAGCCTACATACTTTAAAAGTTTGACAAGATACATAGAAATTTACTACC
SOX4_Pro_2	1	ACCTTGGCATCCTTTCTTTGCCACATTCTTCTTAGTATTCTTCATGGAGTTCCTTTCTTTTAAAATC AATGGAATGGCAGGGTTGCCACTATTGTCCTTCTATGGTTCACATCAATGATC
TAL1_Ctrl_1	1	GATCAACAGGAGGAAAACATACCTCCTGGGTCAAAGTCCCAAGCTTTAGAGCAAAACAAGTTGT TCCAGAAAAGAGGGAGCATGGGTGGACACTTGGCTTTGAGGCTGGTCCCTCCTCCC
TAL1_Ctrl_2	1	GTTGTTCCAGAAAAGAGGGAGCATGGGTGGACACTTGGCTTTGAGGCTGGTCCCTCCTCCCTGG AGGTCTTCACACGGGTGGAGGGTCATCTCCCCAGCTGTATGGACACAGCCCTGATC
TBKBP1_Pro_1	1	GATCCTGCGGGAGGAAGTGGGGAAGTTGGGCGGGGGTGCAGCCGAGGGCGTGCCGACTGTGG AGGGGCCGACCCTCGCGCAGCTCTGGGCCTTCCTTGGGGAGGAGCTGGGGTGCCAGAG
TBKBP1_Pro_2	1	GACATGTGCTCCGCCTCCCACTTTGCCCTCATCACTGCTTACGGAGACATCAAGGAACGGCTGGG GGGCCTGGAGAGGGAGAACGCCACCCTCCGACGCCGCCTCAAAGTCTACGAGATC
TNFRSF1A_Pro_1	1	GATCTGTTTTCCAATTTGCAGAGTGGAGATAACCTGTCTCCTGGGGTTCCTGTAAGGATTTGTTC AGATGGGAGAGATATAGATAAGCGTCCGACACATGATAGGTACTCCATGTCCCTT
TNFRSF1A_Pro_2	1	GAGATATAGATAAGCGTCCGACACATGATAGGTACTCCATGTCCCTTAGTAATCCTCCCCTTTAC TCCAGAGTTCTTCACTAGCCTCCTAGTCTCCCTCTTTCTTGGGAATTTACTGATC
TYK2_Pro_1	1	GATCCCGGGGATGGGGCTTTCGCGGAGCCTACTTCCTCCTTCCTCCTTTCTCGTCCCCACTGGACC CAGACCCAGCTTTGAAGACAGGGTCGGAGTGAAGTTTGCAACTGGGAAGGCTGC
TYK2_Pro_2	1	AGTCGGCACCCAGATGCGCTGTCGCTTCTGCTATTGGCTGGCACTCGGTGGACGGAGACGGGG GAGGGGCTTGCGAGAGGCGGGAGGGAGAGCGAAGTGTGTGGGCAACGCGGCCGGATC
TYK2_rs6511701_1	2	GATCGGCCTGCAGGAGTACAGGACAGGTGCGCGCCGGGCACGCCACGTCGAGCAACAGCAGCA GGAAGAGGGGGCCCCAACATGCCACAAAGGCCAGGAGCACCACGCTGAGCGTGCGCA
TYK2_rs6511701_2	2	CGCAGTGAGTGCCACGAAGACGCCTCCCTCCCGTGCGAACCAGAGCGCGGGGGACAGTTTCAG CGTGAGCGGCCCCGACAGTAGGATGTTGGCGGCGTAGGCGGCGCCTGCCAGCAGATC

UBE2E3_Pro_1	1	GATCTTTTTTGAGGCATGAGATATTCTATCTGTTCATTCTGAGTGGAATATCCTTACTACTTATGA ACCTAGCGAACTTTAATGAACTGTTCCCCTTCTGGACTAGGAGCTCTTAGCACA
UBE2E3_Pro_2	1	ATATGGGCGATTAACAGGGTGTGTTTTCCTCTCGGAAAGGGAATTTGGTAGCTTTGACCAATGT CAGTGTGTAAATATTTAGACGTTATTAGTGACTTTTAAACATTTTTCAGCGAGATC
UBE2E3_rs12615545_1	2	GATCACTATACAGAATAGATTCATAGAGGACCTGAGGAAATATTGAAAGGCAGATGCTCACCA GTATTCATCTTTAAAAACGTATAACTTGTATTTATGAAAAAATTGACTACTGTTAGA
UBE2E3_rs12615545_2	2	AGTGTGAAGTGGGAAGGAGCAGAATGTCTGCCTCGCGTCCAGGAGCAGTGAGGCTGTCACCAG CAAGCATGGGGACCTGAGAGACCGTGTGGGTGGGATGAGGCAGATGCATTAGGGATC
UBE2L3_Pro_1	1	GATCCCCGAGGCAGGCGGCGGCTTCTCAGGGCGCTGCCGTCTGGCTTCCTGCCCTACACGGCCT CGTCGGTTCCCTGGGCCGCGCGCAGGCTCAAGGTGCTAACGGGGCTGGCCTAGGCC
UBE2L3_Pro_2	1	CGGGAGCGCAAGGCCGCGCTGGGAGCCGCTGTCGGCCCCTCAGCCCGGCCCGGGGCGTTCACG CCACTCTCCGCCCGGAGCTTGGCCGCGTCCCCCTGTCGCGGAGGCCGCGGTCCGATC
UBE2L3_rs2266961_1	2	GATCTCGTGATTGGCATCCCTGTCTCTTCTCTGTACCAGGGTAACTTTTTGACCTTTGAGACTCAG CTCAAGGGGCAATTCCTCTGCCCTTGAGGAGCTCAACAAAGGAGAAAAATGGCA
UBE2L3_rs2266961_2	2	CTGCCAGGCCAGCCTTGGAGGGTGAGCTACCTCTCTATAGAGGGGTCACTGAGGGCAGTGCTG ATGGATTTAAGGACAGGTGGCAGAGATGCCATAGCCTTTAGGTTGAGGGGCCAGATC
UBE2L3_rs2283790_1	3	GATCCAGCTCTTCAGGTCAGCCATGAGTCCTGTCTGTGGTGTCCTACAGCCAGCTAGGAACACA CCACAAGGACGTGCCATGTAGCAGCAGTCCAAGTCAGACAACCAGAAGACACCTTG
UBE2L3_rs2283790_2	3	CAAGTGGGTGAAAGACCCCAGATAGTTCGGTGCCATATGGCAAAGAGCTTCCTCTTAGTCTTAG CTGCCACAAAATGGCTTTGTGCCATGGGAGGTTGATATCAAGAGCTGTGAAGGATC
ZMIZ1_Pro_1	1	GATCACCGCCCAAGAGGCCCAGGAGACCCGCAAATCTGTGCTGCACGTCCGCACTCCCATGCGG GCCTCGTCAATCAGGAGGCGGCTCCGGCCAGGCCTGGACATCTTCGGCCCGCCCCA
ZMIZ1_Pro_2	1	CAGAGGGCCAGGCCCCAGCCTAGCAGCTCCCACGCAGCCCGCGCGCACAGCGCCCCGGGCGCTC CGCGCCGGGCCTAGGGGAATCCACGCGTCTCTGGACTGTGCTAGGTGCCGCGGATC
ZMIZ1_rs1250550_1	2	GATCCAGGAGTTGACCCCTACAGGGAGGACATTCCAGGGAGATGGGAGCTAGAGCTGTGGGGT GTGATGGGAATGATTGGTGTGCGGGGGTCTTGGGGGCAGTTTGGGTGTAGAGGGAGT
ZMIZ1_rs1250550_2	2	GGGAATGATTGGTGTGCGGGGGTCTTGGGGGCAGTTTGGGTGTAGAGGGAGTATCCAGAATG GGAGAGGAGAACTCAGGCTGGGGGAATCATTGGGGTCAGGGATGGGGTCCTGAGGATC

Table 2.4. Capture-C oligonucleotide sequences and capture tiers. All oligos included in the Capture-C experiments conducted in chapter three. The first column shows the name of each oligonucleotide in alphabetical order. The second column shows the capture 'tier', where viewpoints within the same AS-associated risk loci are baited and pulled down in independent experiments. The third column shows the full contiguous sequence from the 5' (First base on first row) to 3' (Last base on the second row). Baits with 'pro' in their name are promoter viewpoints, whilst those with a reference SNP cluster ID (i.e. ZMIZ1_rs1250550_1) are viewpoints for AS lead SNPs from Cortes *et al.* (2013) or Ellinghaus *et al.* (2016). Baits with 'Ctrl' in their name are control viewpoints with known and experimentally validated interactions. Where possible, two oligonucleotides were designed per viewpoint to better help pull-down restriction fragments of interest.

2.3.8. Washing of captured material for Capture-C

Capture reactions were washed using the NimbleGen® SeqCap EZ hybridisation and wash kit according to manufacturer's instructions with modification. All subsequent steps occurred in Eppendorf™ LoBind microcentrifuge tubes (ThermoFisher™, 0030108051) to minimise loss of sample complexity. Briefly, 100 µL of M-270 streptavidin dynabeads (Invitrogen™, 65305) per library were brought to room temperature and magnetised on an appropriate magnetic rack. The supernatant was then aspirated once clear, and the

beads were washed 3 times with 200 µL of provided 1x bead wash buffer per library. Beads were magnetised and the supernatant removed between washes. The 72-hour hybridisation reaction was then immediately added to the streptavidin beads, and the samples were mixed thoroughly by pipetting on the sides of the tube. Samples were then immediately placed into a pre-heated thermomixer set to 47ºC with agitation (600rpm) and allowed to bind to beads for 1 hour. Beads were resuspended every 10 minutes to avoid pelleting. Following incubation, 100 µL per library of pre-heated (47ºC) 1x wash buffer I were added to the streptavidin bead-DNA slurry without removing from heat and mixed by pipetting. Samples were then placed in a magnetic rack and the supernatant removed. Samples were then removed from the magnetic stand and beads resuspended in 200 µL of pre-heated (47ºC) 1x stringent wash buffer per library and incubated for 5 minutes at 47ºC. This step was repeated once more. Samples were next resuspended in 200 µL of 1x room-temperature wash buffer I per library, and the samples were mixed by vortexing for 2 minutes. Samples were then magnetised and the supernatant removed, with subsequent resuspension in 200 µL of wash buffer II per library. Samples were mixed by vortexing for 1 minute and then washed in 200 µL of wash buffer III per library as above, with vortexing for 30 seconds. Captured 3C libraries were finally resuspended in 40 µL of nuclease-free water per library and stored at -20ºC until amplification.

2.3.9. Amplification and second capture of Capture-C libraries

Libraries were amplified in 2 reactions using the KAPA™ hot-start ready mix (Roche®, K2601) kit with post-LM PCR oligos 1 and 2 (LP1 and LP2) as provided by the NimbleGen® SeqCap EZ hybridisation and accessory kits to preserve sample complexity. Samples were amplified according to manufacturer's instructions. Between 8-10 PCR cycles were used for amplification based upon how many libraries were multiplexed and number of oligonucleotide baits in the pool. Samples with two

multiplexed 3C libraries and ≤ 20 oligonucleotide capture baits, for instance, would be cycled a total of 10 times, whereas a sample of six multiplexed 3C libraries and ≥ 20 baits would have been cycled a total of 8 times. Samples were cleaned following PCR with a 1.8x SPRI bead clean-up step as above (Section 2.3.5). Samples were eluted in 52 µl, with 50 µL of material being taken forward into second capture. 1 µL of sample was analysed on a D1000 ScreenTape™ to confirm same profile as input material. Second capture of 3C material occurred as above (Section 2.3.6 to 2.3.7) treating the multiplexed 3C material as a singular library. Washed streptavidin beads were eluted in a final volume of 42 µL of nuclease-free water and stored at -20°C until further use.

2.3.10. Test qPCR and final amplification of Capture-C libraries

To avoid overamplification and loss of complexity of captured libraries, a test qPCR was performed on 2 µL of the streptavidin bead slurry using the KAPA™ reagents stated above (Section 2.3.7). As an order of magnitude (10x) more beads were used in the final library amplification, the resultant Cq value of the test reaction was rounded down by 3 to estimate appropriate cycle number for the entire library. The remaining 40 µL of streptavidin bead slurry were split into two separate PCR reactions using the rounded down Cq value as the total number of amplification cycles. PCR reactions were performed as in section 2.3.7. Libraries were cleaned-up with SPRI beads as in section 2.3.5, and eluted in 54 µL of nuclease-free water. 1µL of sample was analysed on a D1000 ScreenTape™ to confirm same DNA profile as primary capture libraries. A further 1 µL of sample was used in a universal KAPA™ library qPCR kit (Roche®, KK4824) to determine library concentration. The kit was used according to manufacturer's instructions with library dilutions of 1:10,000, 1:20,000, and 1:40,000 in triplicate. Libraries were then further diluted accordingly for sequencing with either 150 bp paired-

end NextSeq™ 500/550 300-cycle mid or high output kits (v2.5) (Illumina®, 20024905, 20024908) to a read-depth of 1 million sequencing reads per viewpoint.

2.3.11. Hybridisation and capture baits for Tiled-C

For Tiled-C experiments, tiling panels of 70 bp oligonucleotides were generated using the Oligo (v.0.2) (https://oligo.readthedocs.io/en/latest/overview.html) python programme with BLAT (v.37.1), HMMER (v.3.3.2), Tandem Repeat Finder (v.4.09.1), and RepeatMasker (v.4.1.5) from the GRCh37 human genome assembly at AS GWAS-associated risk loci (**Table 2.5**) (Oudelaar *et al*., 2020). Baits were designed according the cell-specific TADs that were reported in the promoter capture Hi-C studies from Javierre *et al.* (2016), with 100 kbp extension either end of the TAD boundaries. The Oligo pipeline was run in contiguous mode with a step size of 50 between oligonucleotides. Bait sets were deduplicated such that every DpnII restriction fragment was only captured by a single bait. Custom oligonucleotide panels were manufactured, biotinylated, and filtered for highly repetitive genomic black site regions by Twist Biosciences® through their custom oligonucleotide panel services (Twist Biosciences®, 101042). Enrichment by oligonucleotide capture was performed using the Twist standard hybridisation and wash kit (v2) with amplification mix (Twist Bioscience©, 105560). Libraries were captured according to manufacturer's instructions and protocols from Downes *et al.* (2022) with modification. Up to 400 ng of six indexed 3C libraries were multiplexed and desiccated in a vacuum concentrator as above (Section 2.3.6).

Dried DNA libraries were then resuspended in 5 µL of blocker solution and 7 µL of universal blockers (Twist Bioscience®, 100578) to block binding of highly repetitious DNA, and mixed via careful pipetting. A probe mix was simultaneously prepared by mixing 4 µL of Twist custom probe pools with 4 µL of nuclease-free water and 20 µL of

Chr[1]	Start[2]	End[3]	Locus Name[4]
1	153950000	155175000	*IL6R*
1	66950000	68300000	*IL12RB2-IL23R*
1	200300000	201425000	*C1orf106-(INAVA)-GPR25*
1	25060000	25680000	*RUNX3*
2	60600000	62850000	*REL-B3GNT2*
2	191450000	192550000	*STAT1-STAT4*
2	204075000	205075000	*ICOS-CTLA4-CD28*
3	45700000	46800000	*CCR2-CCR5*
3	27025000	28925000	*EOMES-CMC1*
5	39575000	41700000	*PTGER4*
6	19279000	22675000	*CDKAL1*
6	90001000	91600000	*BACH2*
6	136800000	138399000	*IL22RA2-TNFAIP3*
8	127800000	130850000	*MYC*
11	127725000	128950000	*ETS1*
12	55905000	58525000	*CDK2*
12	111425000	113075000	*ATXN2-PTPN11-SH2B3*
16	10500000	11600000	*CIITA-DEXI-CLEC16A-SOCS1*
17	37250000	38575000	*IKZF3-Aiolos-ORMDL3*
17	45075000	46225000	*NPEPPS*
18	12025000	13950000	*PTPN2*
19	9700000	11875000	*TYK2-ICAM1-ICAM3*
21	38950000	40875000	*ETS2*
22	20001000	23275000	*UBE2E3-MAPK1*

Table 2.5. Regions used in tiled oligonucleotide generation for Tiled-C. Regions input into Oligo (v.0.2) to generate oligonucleotides for Tiled-C experiments. The first[1] column shows the chromosome number (Chr). The second[2] and third[3] columns show the start and end coordinates of the tiled regions respectively. These coordinates are roughly equivalent to the TAD that contains an AS-associated SNP ± 100 kbp either side of the TAD of interest. Each region has been named after the suspected AS-associated SNP regulated gene in the fourth[4] column.

provided Twist hybridisation reagent. The probe mix was heated to 95ºC in a thermocycler for 2 minutes (Lid set to 105ºC) and subsequently cooled on ice for 5 minutes. Whilst probe solution was cooling, the resuspended sample of indexed 3C libraries was also heated to 95ºC in a thermomixer for 5 minutes (Lid set to 105ºC). Both libraries and probe mix were then equilibrated to room-temperature for 3 minutes, before combining and mixing carefully by pipetting. 30 µL of provided hybridisation enhancer was added to the top of the hybridisation mixture. Samples were briefly centrifuged in a microfuge to ensure the absence of bubbles and that the sample was collected at the bottom of the tube. The hybridisation reaction was then incubated in a thermocycler at 70ºC for at least 17 hours (Lid set to 85ºC). A strip of parafilm was wrapped around the top of each PCR tube to prevent excessive evaporation of hybridisation mixture.

2.3.12. Washing of captured Tiled-C libraries

Both streptavidin beads and purification beads were provided by the Twist binding and purification beads kit (Twist® Bioscience, 100578). For each capture reaction, 100 µL of streptavidin beads and 800 µL of 1x Twist bead-binding buffer were brought to room temperature. A further 275 µL of 1x Twist wash buffer 1 and 700 µL of 1x Twist wash buffer 2 per reaction were brought to 68ºC and 48ºC in a thermomixer respectively. To prepare streptavidin beads, 200 µL of 1x Twist bead-binding buffer was added and mixed thoroughly by vortexing at high speed. Samples were magnetised and supernatant was removed once clear without disturbing beads. This was repeated 2 more times for a total of 3 washes. Beads were resuspended in a final volume of 200 µL of 1x Twist bead-binding buffer, and the resuspended beads were heated to 68ºC for 10 minutes in a thermomixer. Without removing from heat, the entire hybridisation reaction was added into the bead mixture and mixed by careful pipetting. The combined samples were allowed to incubate for 10 minutes in thermomixer at 68ºC with agitation (400 rpm).

Samples were then magnetised and the supernatant removed quickly to avoid over-cooling once clear. 250 µL of 68ºC 1x Twist wash buffer 1 was then immediately added, and the beads were quickly resuspended by vortexing before being replaced on the thermomixer as above. Samples were subsequently incubated for 10 minutes at 68ºC.

The entirety of the sample was transferred into a new microcentrifuge tube – one per hybridisation reaction – to reduce the non-specific carryover of DNA bound to the tube surface. The sample was then magnetised, and the supernatant was quickly removed and replaced once clear with 200 µL of 48ºC 1x Twist wash buffer 2. Samples were mixed by careful pipetting and incubated at 48ºC in a thermomixer for 10 minutes, with subsequent magnetisation and removal of supernatant once clear. This wash step was repeated a further 2 times for a total of 3 washes, with final resuspension of captured libraries in 45 µL of room-temperature nuclease-free water. Samples were mixed via careful pipetting, and incubated on ice for at least 5 minutes. 22.5 µL of the resultant bead slurry was transferred to a PCR tube for amplification, and the rest of the sample was stored at -20ºC until further use.

2.3.13. Amplification and sequencing of Tiled-C libraries

The Equinox™ library and amp mix kit provided in the Twist standard hybridisation and wash kit was used to amplify all Tiled-C libraries. To the 22.5 µL of captured sample, 2.5 µL of ILMN primers, and 25 µL of 2x library amp mix were added and mixed via vortexing. Samples were initially heated for 2 minutes at 98ºC in a thermocycler, with subsequent amplification on a PCR cycle phase consisting of a denaturation stage of 98ºC for 15 seconds, an annealing stage of 60ºC for 30 seconds, and an extension stage of 72ºC. Lid was set to 105ºC throughout. The number of PCR cycles was determined by the overall size of probe bait sets per megabase of the genome. Reactions using the first bait

set were amplified a total of 11 times, whilst the second bait set were amplified for a total of 10. Cycles were concluded with a final extension phase of 2 minutes at 72ºC, with a final holding temperature of 4ºC. Following amplification, 90 µL of twist purification beads were added to each reaction and mixed well by vortexing. Samples were then incubated at room temperature for at least 10 minutes and washed twice with fresh nuclease-free 80% ethanol as in section 2.3.5.

Bead pellets were then allowed to dry whilst magnetised until matte in consistency and then resuspended in 33 µL of Qiagen® buffer EB from the MinElute™ reaction clean up kit. Samples were mixed via pipetting, and subsequently incubated at room temperature for at least 5 minutes to elute DNA from beads. 1 µL of sample was analysed on a high-sensitivity D1000 ScreenTape to confirm similar DNA profile to original 400 ng inputs. A further 1 µL of sample was used in a universal KAPA™ library qPCR kit to determine library concentration. The kit was used according to manufacturer's instructions with library dilutions of 1:20,000, 1:40,000, and 1:80,000 in triplicate. All Tiled-C samples were then multiplexed together and the KAPA™ assay was performed once more as stated above. The multiplexed sample was then diluted to between 2nM and 30nM and sent to Novogene© for sequencing on 150bp paired-end 200-cycle NovaSeq™ X-series 10B flow-cells (Illumina®, 20085594). Each Tiled-C library was sequenced to a minimum depth of 10 million sequencing reads per Mb of the oligonucleotide baited genome.

2.4. ChIPmentation-seq

2.4.1. Preparing cells for ChIPmentation-seq

For ChIPmentation-seq, roughly $6x10^5$ harvested MACS-purified cell populations were pelleted from growth medium and resuspended in 1mL of PBS supplemented with 1%

FCS. Cells were then fixed by adding formaldehyde to a final concentration of 1% v/v and tumbling at room-temperature for 10 minutes. 2.5M glycine was then added to a final concentration of 0.25 M to quench the formaldehyde, and cell samples were incubated at room temperature for 5 minutes. Cells were transferred into a new microcentrifuge tube and centrifuged at 1000xg for 3 minutes to pellet fixed cells. The supernatant was then removed and the samples were washed twice with unsupplemented PBS. Cells were pelleted as above between each wash step. Cell pellets were subsequently snap-frozen in a bath of dry ice and 100% ethanol and stored at -80ºC until lysis.

Fixed cell pellets were thawed and resuspended in 120 µL of SDS lysis buffer (0.25% SDS; 1 mM EDTA; 10 mM Tris-HCl (pH 8); 1x cOmplete protease inhibitor cocktail tablet) and incubated for 5 minutes tumbling at 4ºC. Samples were then aliquoted into Covaris AFA Fiber 130-microTUBEs and sonicated to 200 bp on an M220 ultrasonicator. Program settings were a duty factor of 5%, intensity of 75, cycles per burst of 200, cooled to 7ºC, and a sonication time of 480 seconds. Samples volumes were increased to 180 µL with SDS lysis buffer and centrifuged at 13,000 rpm for 10 minutes at 4ºC to pellet unwanted precipitate. The supernatant was transferred to a new microcentrifuge tube and subsequently split into six aliquots of 30 µL. Samples were snap-frozen in a dry-ice ethanol bath and stored at -80ºC. At least one of these aliquots were used as an input control.

2.4.2. Reverse cross-linking and tagmentation of ChIPmentation inputs

Inputs were thawed at room temperature and 166 µL of reverse-crosslinking buffer (10 mM Tris-HCl (pH 8); 1 mM EDTA; 300 mM NaCl; 0.4% SDS) and 4 µL of proteinase K were added to each and mixed via pipetting. Samples were then incubated for 1 hour at 55ºC followed by at least 12 hours at 65ºC in a thermomixer with agitation (600 rpm).

A MinElute™ reaction cleanup kit (Qiagen®, 28204) was then used to purify DNA for tagmentation. Input samples were eluted in 12 µL of nuclease-free water and 1 µL of sample was taken for analysis on a D1000 ScreenTape™. Quantification was done on a further 1 µL of sample with a broad-range dsDNA Qubit® assay kit. Tagmentation was carried out using the Illumina® tagment DNA TDE1 enzyme and buffer kit (Illumina®, 20034197). Input samples were initially diluted to a concentration of 1 ng/µL, and 2.5 µL of input was mixed with 5 µL of provided 2x tagment DNA buffer, 1.5 µL of nuclease-free water, and 1 µL of TDE1 transposase ((Illumina®, 15027866). Reactions were incubated for 5 minutes on a thermomixer at 55ºC without agitation, and tagmentation reaction was then stopped by adding 500 µL of Qiagen® buffer PB. Samples were purified once again with the MinElute™ reaction clean-up kit and eluted in 30 µL of EB buffer and stored at -20ºC.

2.4.3. Antibody binding and immunoprecipitation

To prepare samples for immunoprecipitation, thawed aliquots of chromatin were mixed with 1.5 times the volume (v/v) of ChIP equilibration buffer (10 mM Tris-HCl (pH 8); 1 mM EDTA; 23 mM NaCl; 1.66% Triton-X100; 1x cOmplete protease inhibitor cocktail solution). Samples were then made up to 150 µL with ChIP buffer (10 mM Tris-HCl (pH 8); 1 mM EDTA; 140 mM NaCl; 0.1% SDS; 1% Triton-X100; 1x cOmplete protease inhibitor cocktail solution) and 2 µg of H3K27me3 antibody (Diagenode®, C15410195) were added to each. Samples were transferred into PCR tubes and incubated at 4ºC for at least 16 hours on an end-over rotor. Simultaneously, 10 µL of either dynabeads™ protein A (Invitrogen™, 10001D) per sample were washed three times with 1 mL of 1x ChIP bead wash buffer (50 mM Tris-HCl (pH 8); 1 mM EDTA; 150 nM NaCl; 0.1% SDS; 1% Igepal CA-630; 1x cOmplete protease inhibitor cocktail solution). Bead suspensions were

vortexed on low speed for 1 minute prior to magnetisation and aspiration of supernatant between each wash step.

Following washing, beads were resuspended in 1 mL of 1x ChIP bead wash buffer, 1 µL of yeast tRNA (Sigma-Aldrich, R5636), and 1 µL of 100x BSA (NEB, B9000S) per library to block non-specific binding of DNA to beads. Bead suspensions were also incubated at 4ºC for at least 16 hours on an end-over rotor prior to use. After incubation, beads were washed a further 2 times with 1 mL of ChIP bead wash buffer, and resuspended in ChIP buffer to the original total volume of used beads. 10 µL of beads were then added to each antibody-bound chromatin sample and incubated at 4ºC on an end-over rotor for at least another 2 hours.

2.4.4. Washing of immunoprecipitated DNA-bead complexes

Bead-DNA complexes were washed two times with 200 µL each of 1x wash buffer 1 (20 mM Tris-HCl (pH 8); 2 mM EDTA; 150 mM NaCl; 0.1% SDS; 1% Triton-X100; 1x cOmplete protease inhibitor cocktail solution), 1x wash buffer 2 (20 mM Tris-HCl (pH 8); 2 mM EDTA; 500 mM NaCl; 0.1% SDS; 1% Triton-X100; 1x cOmplete protease inhibitor cocktail solution), 1x wash buffer 3 (10 mM Tris-HCl (pH 8); 1 mM EDTA; 250 mM lithium chloride (LiCl) (ThermoScientific™, J61895); 0.7% sodium deoxycholate (ThermoScientific™, 89904); 1% Triton-X100), and once with a final wash of 10 mM Tris-HCl (pH 8). Samples were magnetised, and beads were allowed to move through solution through moving tubes back and forth the magnetic rack a of six times to facilitate washing prior to aspiration of supernatant. Samples were cooled with ice for 10 minutes, and resuspended in 10 µL of 2x tagment DNA buffer (Illumina®, 15027866), 9 µL of nuclease-free water, and 1 µL of TDE1 transposase. All samples and reagents were kept on ice to avoid over-tagmentation of samples. Once all were resuspended, samples

were incubated at 37ºC for 10 minutes in a thermomixer without agitation. Transposase was inactivated by adding 196 µL of ice-cold ChIP elution buffer (10 mM Tris-HCl (pH 8); 5 mM EDTA; 300 mM NaCl; 0.4% SDS) and 4 µL of proteinase K to each sample. Samples were transferred each to a new microcentrifuge tube and incubated for 1 hour at 55ºC followed by 65ºC for at least 16 hours. Reactions were shaken at 1400 rpm throughout, and a strip of parafilm® was wrapped around the top of each microcentrifuge tube to prevent excessive evaporation. Following this, beads were magnetised and sample supernatants – including the input – were transferred to new microcentrifuge tube and purified with the Qiagen® MinElute™ reaction clean-up kit according to manufacturer's instructions. ChIP samples were eluted in a final volume of 23 µL Qiagen® buffer EB and stored at -20ºC.

2.4.5. Indexing and sequencing of ChIPmentation samples

Amplification of ChIPmentation samples utilised the indexes of the ATAC protocol published by Buenrostro *et al.* (2013). To determine the correct number of library amplification cycles required for each library a test qPCR using the NEBNext™ HiFi PCR master mix (NEB®, M0541) was performed with reported primer pairs. To 2 µL of purified ChIPmentation chromatin, 10 µL of 2x master mix, 0.2 µL of 100x SYBR, 6.6 µL of nuclease-free water, and 0.6 µL each of 25 µM Nextera primer Ad1_noMX and Ad2.1 were added. After initially being heated to 72º for 5 minutes and then 98ºC for 30 seconds, samples were cycled a total of 40 times with a denaturation step of 98ºC for 10 seconds, an annealing step of 63ºC for 30 seconds, and an extension step of 72ºC for 30 seconds with acquisition after every cycle. The number of amplification cycles for samples were determined from the Cq value rounded down to the nearest cycle number. 20 µL of purified ChIPmentation chromatin was then mixed with 25 µL of 2x master mix, 2 µL of nuclease-free water, and 1.5 µL each of 25 µM unique Nextera Ad1 and

Ad2 primers for individual libraries and amplified for the appropriate number of cycles as above. Amplified ChIPmentation libraries were purified using a double-sided SPRI bead clean-up step. 0.65x the volume (v/v) of room temperature SPRI beads were first added to each sample and mixed via pipetting. Samples were then incubated for 10 minutes at room temperature, magnetised for 10 minutes, and the supernatant transferred to a new low-bind microcentrifuge tube. A further 0.65x the volume of SPRI beads – for a total 1.3x volume of polyethylene glycol – was then added to each sample. Samples were again incubated for 10 minutes at room temperature, magnetised for 10 minutes, and the supernatant was discarded.

Beads were then washed twice with 1 mL of fresh 80% ethanol for 10 seconds each wash, and beads were air dried for 5 minutes at room temperature or until matte in appearance. SPRI beads were then resuspended in 15 µL of buffer RSB (Illumina®, 15027866) and mixed by pipetting. DNA was allowed to elute from the beads by incubating at room temperature for 10 minutes. Beads were magnetised for 5 minutes and the supernatant was transfered to new low-bind centrifuge tubes. Quality of eluted ChIPmentation samples was assessed using a high-sensitivity D1000 DNA TapeStation™ and quantified with a high-sensitivity dsDNA Qubit® assay. ChIPmentation samples were then diluted to a concentration of 2-4 nM and multiplexed for sequencing. 1 µL of the final multiplexed sample was used in a universal KAPA™ library qPCR kit to confirm library concentration of 2-4 nM prior to sequencing. The kit was used according to manufacturer's instructions with library dilutions of 1:10,000, 1:20,000, and 1:40,000 in triplicate. Multiplexed H3K27me3 ChIPmentation samples were then sent to Novogene© for sequencing on 150 bp paired-end 200-cycle NovaSeq™ X-series 10B flow-cells (Illumina®, 20085594). ChIPmentation libraries were sequenced to a final depth of at least 30 million uniquely mapped paired-end reads after filtering.

2.5. ATAC-seq

2.5.1. Cell preparation for ATAC-seq and transposition

10x of fresh NTM buffer (100 mM Tris-HCl (pH 7.4) (Sigma-Aldrich®, T2194), 30 mM $MgCl_2$ (Sigma-Aldrich®, 7786-30-3), 100mM NaCl) was made prior to processing ATAC-seq samples. For ATAC-Seq, $5x10^4$ of harvested MACS-purified cell populations were resuspended in 400 µL of PBS supplemented with 1% FCS. 200 µL of the supernatant was removed and the cells were pelleted via centrifugation at 500xg for 5 minutes (4ºC). Supernatant was carefully removed to avoid disturbing cell pellet and resuspended in 50 µL of ATAC lysis buffer (1x NTM buffer; 0.01% digitonin (ThermoScientific™, BN2006); 0.1% TWEEN®-20 (Sigma-Aldrich®, P1379); 0.1% Igepal CA-630). Samples were mixed by pipetting up and down a total of 3 times, and incubated on ice for 3 minutes. Lysis buffer was washed out by adding 1mL of ice-cold ATAC wash buffer (1x NTM; 0.1% TWEEN®-20) and mixed by inverting 3 times. Nuclei were pelleted at 500xg for 10 minutes (4ºC), and supernatant aspirated. Samples were then resuspended in 50 µL of ice-cold transposition mixture (2.5 µL TDE1 transposase; 25 µL 2x tagment DNA buffer; 0.5 µL 1% digitonin; 0.5 µL TWEEN®-20; 16.5 µL PBS; 5 µL nuclease-free water; 50 µL per reaction) and then incubated at 37ºC for 30 minutes on a thermomixer with agitation at 1000 rpm. 250 µL of Qiagen® buffer PB was then added to stop the transposition reaction, and samples were processed with the Qiagen® MinElute™ reaction clean-up kit according to manufacturer's instruction. DNA was eluted in 23 µL of the provided buffer EB.

2.5.2. Indexing and sequencing of ATAC-seq material

Indexing of ATAC-seq samples was performed as described by Buenrostro *et al.* (2013). A test qPCR was used as in section 2.4.5 to determine the correct number of library

amplification cycles in each ATAC-seq library. In contrast to ChIPmentation, the actual amplification used 20 µL of the ATAC-seq DNA produced in section 2.5.1, 25 µL of 2x NEBNext™ Q5™ HiFi PCR Master Mix (NEB®, M0543S), and 2.5 µL of the appropriate unique Nextera Ad1 and Ad2 primers at a concentration of 25 µM. After an initial hot-start activation step of 72°C for 5 minutes and then 98°C for 30 seconds, samples were cycled for the tested number of cycles with a denaturation step of 98°C for 10 seconds, an annealing step of 63°C for 30 seconds, and an extension phase of 72°C for 60 seconds. After cycling, 10 µL of 3M sodium acetate (pH 5.2) was added to each sample, and DNA was purified using a MinElute™ reaction cleanup kit. ATAC-seq samples were eluted in 12 µL of nuclease-free water, and 1µL each of DNA was taken for quality assessment on a D1000 high-sensitivity TapeStation™ and quantification using a universal KAPA™ library qPCR kit as in section 2.4.5 with the same conditions and dilutions. Multiplexed ATAC-seq samples were diluted to a final concentration of between 2-4 nM and sent to Novogene© for sequencing on 150bp paired-end 200-cycle NovaSeq™ X-series 10B flow-cells (Illumina®, 20085594). ATAC-seq libraries were sequenced to a final depth of at least 20 million paired-end reads after filtering.

2.6. qRT-PCR and RNA-seq

2.6.1. Purification and reverse transcription of total RNA for qPCR

Aliquots of harvested macrophages from section 2.2.3.2 were pelleted at 500xg for 1 minute at room temperature and the supernatant was aspirated. Macrophages were then processed using the Monarch® Total RNA Miniprep Kit (NEB®, T2010S) according to manufacturer's instructions for leukocyte samples under $3x10^6$ total cells and eluted in 35 µL of nuclease-free water. RNA was quantified with a qubit™ RNA High Sensitivity kit (ThermoScientific™, Q32852) and RINe degradation quality scores were calculated

using a high-sensitivity RNA ScreenTape™ (Agilent®, 5067-5579). Samples under a RINe score of 8 were discarded from analysis. 1 µg of RNA was taken forward for cDNA synthesis using the LunaScript® RT SuperMix Kit (NEB®, E3010) as per manufacturer's instructions. 3 µL of RNase A (Roche®, 1119915) was then added to each sample for 1 hour at 40°C to clear residual RNA from generated cDNA. cDNA was quantified using a high sensitivity DNA qubit™ kit.

2.6.2. qRT-PCR reaction and cycling conditions

Quantitative real-time PCR (qRT-PCR) was used to determine expression of *NOS2* in macrophages. 20 ng of cDNA was taken forward for multiplexed qRT-PCR using PrimePCR™ probe assay kits for *NOS2* (Cy5) (Nitric Oxide Synthase 2; Bio-Rad qHsaCIP0030342) and *ACTB* (FAM) (β-Actin; Bio-Rad®, qHsaCEP0036280). 1 µL of each probe-primer pair were then added to the template cDNA and 10 µL of SsoAdvanced™ Universal Probes Supermix (Bio-Rad®, 152-5280) and topped to a final reaction volume of 20 µL with nuclease-free water. A Bio-Rad® C1000 thermocycler set to acquire all fluorophore colour channels was used for amplification of samples in technical triplicate.

Following an initial denaturation step of 98°C for 2 minutes, samples were cycled a total of 40 times with a denaturation step of 98°C for 5 seconds and a joint annealing-extension step of 60°C for 30 seconds. Cq data was acquired after every cycle. Mean Cq values were calculated from the technical triplicates for both *NOS2* and *ACTB*. Normalised *NOS2* ΔCq values were then calculated for each well by taking the Cq value for the *ACTB* in-well control from the Cq of the *NOS2* target gene. ΔΔCq values were calculated for each time-point by subtracting the normalised *NOS2* ΔCq values for the 0-hour control from the normalised *NOS2* ΔCq for each time-point. The fold-change over the 0-hour

control was then calculated using the equation $2^{-\Delta\Delta Cq}$ for each time-point. Standard deviation of the ΔΔCq values was calculated by $\sqrt{(S_{NOS2})^2 + (S_{ACTB}^2)}$ where S_{NOS2} and S_{ACTB} are the standard deviations of the *NOS2* and *ACTB* Cq values respectively. Statistical significance of time-point comparisons was assessed by one-way analysis of variance (ANOVA) with Tukey's range test post-hoc analysis. The null hypothesis for this experiment was no change in *NOS2* expression between time-points.

2.6.3. Processing of cells for RNA-seq and sample sequencing

For RNA-seq experiments, $2x10^6$ cell of each cell type collected in sections 2.2.1 to 2.2.2 were pelleted at 500xg for 10 minutes at 4°C and resuspended in 300 µL of buffer RLT (Qiagen®, 79216) supplemented with 1% β-mercaptoethanol (Sigma-Aldrich®, 60-24-4). RNA was then extracted using the AllPrep DNA/mRNA/miRNA Universal kit (Qiagen®, 80224) according to manufacturer's instructions using QIAshredder spin columns (Qiagen®, 79656). Quality of eluted total RNA was assessed on a high sensitivity RNA TapeStation and quantified with a high-sensitivity RNA qubit. Samples below a RINe score quality threshold of 8 were removed from subsequent analysis. 2 µg of isolated RNA for each sample were then sent to GENEWIZ®/Azenta® Life Sciences for ribodepletion, library indexing, and standard RNA-seq. RNA-seq libraries were sequenced by Azenta® to a depth of 25 million uniquely mapped paired-end reads after filtering.

2.7. Computational and statistical analyses

2.7.1. Analysis of Capture-C data

Analysis for NG Capture-C was led by Gabriele Migliorini and performed as outlined by Brown *et al.* (2023). FastQC (v.0.11.4) and TrimGalore (v.0.4.4) were used to ascertain the quality of sequenced reads and remove Illumina® adapters from the reads

respectively. Two sets of merged and unmerged sequenced reads were generated for subsequent analysis following the FLASH96 overlap and merge criteria. A modified version of FLASH (v.1.2.11) was used to digest flashed and unflashed sequenced reads after merging at predicted DpnII digest sites. Bowtie2 (v.2.3.5.1) (Langmead *et al.*, 2009) was then used to count and align reported *in-silico* digested reads to the human genome assembly GRCh37. Read pairs containing sub-reads shorter than 20bp after flashing were discarded, and remaining reads were passed into SAMTools (v.1.1.2) for enumeration and deduplication of PCR-introduced read copies.

Pairs with low mapping quality on either end of the read – corresponding to a phred quality score of 30 or less – were also excluded at this stage. Our in-house analysis pipeline ICeCAP was then used to identify genuine Capture-C interactions by enumerating and removing experimental artefacts – such as re- and self-ligations events – from the deduplicated reads (https://zenodo.org/records/7760067). CaptureCompare (v.1.0.0) (Telenius *et al.*, 2020) was used to compare the mean number of interactions for each bonafide promoter or SNP viewpoint between healthy volunteer and AS patient Capture-C samples. One standard deviation of the mean was also plotted to facilitate identification of differential regions of interaction. Peaky (v.1.0.0) (Eijsbouts *et al.*, 2019) was used to assess the significance of interaction frequencies within each AS-associated GWAS locus, with median values derived from the top 10% of peaky scores for the given region of interest.

Samples were plotted alongside ChromHMM and other epigenetic datasets in R (v.4.3.1) with Gviz (v.46.1) (Hahne & Ivanek, 2016) and GenomicInteractions (v.1.36.0) (Harmston *et al.*, 2015). CaptureCompare tracks are plotted as the mean counts from three AS patients and three healthy volunteer samples in bins of 400bp with one standard

deviation of the mean shading. PeakY interaction tracks are plotted as the -$\log_{10}$ of the FDR value below each CaptureCompare track with a sliding window of 400bp.

2.7.2. Analysis of Tiled-C data

Tiled-C interaction matrices and heatmaps were generated using CapCruncher (v.0.3.13.dev1) (GitHub-sims-lab/CapCruncher) as per author's instructions. FastQC (v.0.11.4) was used to perform quality control on fastq files. Fastq files were split into separate parts of 1 million reads per part to reduce pipeline memory requirements and allow parallelisation of pipeline processes. PCR duplicates were removed from the fastq files using the CapCruncher package to reduce the memory and CPU requirements of the pipeline. TrimGalore (v.0.4.4) was then used to trim Illumina® adaptor sequences from the ends of fastq reads. A modified version of FLASH (v.1.2.11) was used to digest flashed and unflashed sequenced reads after merging at predicted DpnII digest sites. Bowtie2 was then used to count and align reported *in-silico* digested reads to the human genome assembly GRCh37.

The CapCruncher package was then used to annotate the bowtie2 aligned digested reads, filter for experimental or assay artefacts such as self or re-ligation events, and remove aligned PCR duplicates not identified in the previous removal step. Interaction matrices were generated from the CapCruncher pipeline with ICE (Iterative correction and eigenvector decomposition) (v.0.5.3) normalisations. As ICE generates a relative contact probability for each individual Tiled-C experiment and does not produce an associated p-value, the normalisation allows for comparison of contact matrices across samples with differing total read depths and without multiple testing correction (Imakaev *et al.*, 2012). The hdf5 files produced by CapCruncher were plotted in Python (v.3.10.8) using a mixture of CapCruncher plotting functionality and CoolBox (v.0.3.8) (Xu *et al.*, 2021).

2.7.3. Analysis of ChIPmentation-seq data

Analysis of ChIPmentation-seq was performed as previously described (Brown *et al.*, 2023) with modification and the assistance of Dr. Andrew Brown and Dr. Gabriele Migliorini. Sequenced reads for ChIPmentation samples and their controls were first aligned to the GCRh37 human genome assembly using Bowtie2 SAMtools (v.1.12) (Li *et al.*, 2009) and the Picard toolkit (v.3.1.1) from the Broad Institute were then used to filter PCR-duplicated reads, low quality reads (MAPQ score < 30), and mitochondrial reads from each sample. Aligned reads were then enumerated with HTSeq (v.2.0.2) (Anders et al., 2015). Bedgraph files were generated from the read pileups using the bamtobed function of BEDtools (v.2.31.0), and bedgraphs were converted to bigwig files with bedGraphToBigWig (v.2.8) (Kent *et al.*, 2010) for plotting in R (v.4.3.1) with Gviz (v1.46.1) and GenomicInteractions (v.1.36.0). Peak calling of H3K27me3 samples and controls was performed using MACS2 v2.1.2 (Zhang *et al.*, 2008).

Due to the broader and weaker nature of the H3K27me3 signal in comparison to histone marks such as H3K27ac and H3K4me, peak calling was performed in broad-peak mode with the less stringent parameters than reported (--nomodel, shift -100, extsize 200, q 0.1, keep-dup all, call-summits) and no p-value cut-off. All peaks were extended by 500bp upstream and downstream of the peak summit to create normalised uniform peak lists in all samples. All peaks were then merged, compared between control and H3K27me3 samples, and peaks called in less than 20% of all H3K27me3 ChIPmentation samples were removed from the analysis.

As H3K27me3 is typically weakly enriched at TSS, instead of the TSS enrichment quality control reported in Brown *et al.* (2023) the fraction of reads in peaks (FRiP) method was instead used. FRiP enumerates the number of reads assigned to peaks and their relative

strength for each ChIPmentation sample to calculate the signal-to-noise ratio for each peak (Schmidl *et al.*, 2015). Successful ChIPmentation demonstrated a FRiP score of at least 0.01. Differential analysis and limma batch correction of H3K27me3 peaks between conditions were performed with DESeq2 (v.1.43.2) (Love *et al.*, 2014) with Limma (v.3.58.1) (Ritchie *et al.*, 2015) for batch effect correction. Both the Wald test and the Benjamini-Hochberg method were used to correct for multiple testing. The X and Y chromosomes were removed from the analysis by DESeq2 to limit batch effect from sample sex. H3K27me3 peaks were considered significant if they demonstrated a false discovery rate (FDR) value of less than 0.05. Tracks showing the upregulation or downregulation of called H3K27me3 peaks were created from bed files generated from the DESeq2 output with BEDtools (v2.31.0) for plotting alongside the other epigenetic datasets in R (v.4.3.1) with Gviz (v.1.46.1) and GenomicInteractions (v.1.36.0). ChIPmentation-seq was visualised with sliding windows of 400bp.

2.7.4. Analysis of ATAC-seq data

ATAC-seq data was processed exactly as in section 2.6.3 until the calling of peaks with MACS2 (v.2.1.2) which used different parameters in narrow-peak mode (--nomodel, shift -100, extsize 200, p 0.1, keep-dup all, call-summits). A p-value cut-off of $1x10^{-8}$ was also applied to called peaks, and peaks were removed upon overlap with ENCODE highly-repetitive blacklisted regions (https://www.encodeproject.org/annotations). Peak heights were extended by 250bp in either direction to create a list of 500bp non-overlapping homogenous peaks. In cases where more than one summit was identified in a single detected peak, the median peak height was used. All called peaks in all samples were then collapsed into a masterlist and peaks present in less than 20% of the samples were removed from the analysis. As promoters are well known regions of open chromatin, the quality of ATAC-seq samples was determined by using TSS enrichment (±1000bp), with

samples showing a TSS enrichment score of 5 or less being removed from the analysis. TSS locations were derived from Ensembl for the GRCh37 human genome assembly (Hunt *et al.*, 2018).

Differential analysis and limma batch correction of ATAC peaks between conditions were carried out as for ChIPmentation-seq with DESeq2 and limma correction. Both the Wald test and the Benjamini-Hochberg method were used to correct for multiple testing. ATAC-seq peaks were considered differential if they demonstrated a false discovery rate (FDR) value of less than 0.05. DESeq2 ATAC-seq analysis presented in chapter five was restricted to comparison of peaks only within the bounds of tiled regions. Tracks showing the upregulation or downregulation of called H3K27me3 peaks were created from bed files generated from the DESeq2 output with BEDtools (v2.31.0) for plotting alongside the other epigenetic datasets in R (v.4.3.1) with Gviz (v.1.46.1) and GenomicInteractions (v.1.36.0). ATAC-seq was visualised with sliding windows of 200bp. Bedgraph files were generated from the read pileups using the bamtobed function of BEDtools (v.2.31.0), and bedgraphs were converted to bigwig files with bedGraphToBigWig (v.2.8).

2.7.5. Analysis of RNA-seq data

RNA-seq analysis was carried out as published (Brown *et al.*, 2023). RNA-seq reads were mapped to the GRCh37 human genome assembly using STAR (v.2.7.11) (Dobin *et al.*, 2013) and reads were counted with featureCount (v.2.0.6) (Liao *et al.*, 2014). Duplicated RNA reads were marked and removed using the Picard toolkit (v.3.1.1) and SAMtools (v.1.12). To identify eRNA peaks from amongst the RNA-seq data, uniquely mapped RNA reads within ATAC-seq peaks ±3000kbp from a gene coding region were enumerated with featureCounts (v.2.0.6). Differential analysis and limma batch

correction of both RNA and eRNA peaks between conditions were performed with DESeq2 and limma correction. Both the Wald test and the Benjamini-Hochberg method were used to correct for multiple testing. RNA and eRNA peaks were considered significantly differential if they demonstrated a false discovery rate (FDR) value of less than 0.05 and a fold change of greater than 1.5. Bedgraph files were generated from the read pileups using the bamtobed function of BEDtools (v.2.31.0), and bedgraphs were converted to bigwig files with bedGraphToBigWig (v.2.8). For visualisation, RNA-seq was plotted as a total RNA track in bins of 1000bp using the $\log_2$ count as a single representative bigwig from the healthy volunteer HV30 and the as patient AS053. Tracks showing the upregulation or downregulation of called peaks specifically for eRNA were created from bed files generated from the DESeq2 output with BEDtools (v2.31.0) for plotting alongside the other epigenetic datasets in R (v.4.3.1) with Gviz (v.1.46.1) and GenomicInteractions (v.1.36.0).

2.7.6 ChromHMM

ChromHMM (v.1.25) (Ernst & Kellis, 2017) was performed as previously described (Brown *et al.*, 2023) with the addition of the H3K27me3 histone modification. A subset of four AS patients and four healthy volunteer samples that received complete single-batch processing and binarised sequencing data for ATAC, RNA, H3K4me3, H3K27ac, and H3K27me3 were used for ChromHMM. A generic track for H3K4me1 from an independent healthy volunteer was also incorporated into the ChromHMM. Unlike the differential ATAC-seq, RNA-seq, and ChIPmentation-seq data which used MACS2, peaks for each dataset were calculated from the ChromHMM package directly. A coding track was generated as a binarised file showing the presence of absence of known gene sequences within each 200bp bin. In order to determine the appropriate model size for ChromHMM, model sizes from 1 emission state to 30 emission states were generated on

the binarised data tracks. The CompareModels functionality of ChromHMM was then used to compare the correlation value for each state in the 30-state model to that in the lower state models to ascertain how well that particular state is captured from the 1 to 30 emission state models. The median correlation value was then calculated for each model size. A 16-emission state model was chosen as the most appropriate as it represented a local median correlation value peak and where the median correlation value began to plateau, suggesting diminishing returns in model informativeness past 16-states. Manual curation was used to assign putative functions to each of the ChromHMM emission states according to literature, combinations of epigenetic marks, and their genomic distribution.

Multiple correlation analyses (MCA) were performed either on the global distribution of ChromHMM emission states or on each emission state individually using FactoMineR (v.2.10) (Lê *et al.*, 2008). To reduce computational time, ChromHMM bins were only kept in global analyses if they demonstrated more than 1 emission state and if the quiescent U14 emission state was present in less than 2/3rds of all samples. In individual emission state analyses, all bins where one-third of the individuals demonstrated a non-quiescent emission state were filtered out from the analysis. Regions that showed the same emission state within the 8 individuals were collapsed. Fisher exact tests were used to ascertain whether significant differences in each ChromHMM bin existed between healthy volunteers and AS patients on both a cell type-specific and emission-state specific basis. Due to binarisation of ChromHMM and sample size limitations (n=4 AS patients, n=4 healthy volunteers), the minimum p-value achievable by Fisher exact tests is 0.05, which on multiple testing correction results in a q-value greater than 0.05, which represents an emission state present in 4 of the healthy volunteers and 0 AS patients, or vice versa.

To identify ChromHMM emission states with significant differences in their distributed regions between AS patients and healthy volunteers than expected by chance alone, I permuted the disease status the maximum possible number of times (70 total combinations) and compared the observed frequency of Fisher's p-value significant regions with the permuted background distribution. The minimum achievable p-value for this analysis was 0.015, corresponding to 1 out of 70 possible combinations. Comparative emission state transition analysis was also performed upon Fisher's significantly permuted ChromHMM bins by comparing the state in the four healthy volunteers to the four AS patients. Transitions present in four healthy volunteers and no AS patients and in four AS patients and no healthy volunteers were plotted using Circlize (v.0.4.11) (Gu *et al.*, 2014).

2.7.7. Enrichment of genetic variants in ChromHMM emission states

Enrichment of AS-associated variants in regulatory regions was determined using GARFIELD (GWAS analysis of regulatory and functional information enrichment with LD correction) for each of the emission states defined by ChromHMM. To achieve this, GARFIELD requires five input files. The first and second files are the genomic locations of the trait-associated GWAS variants and their associated p-values. A third file detailing the position of each ChromHMM bin that contains a GWAS variant in each cell type is then required to link these variants to their putative ChromHMM emission states. From two LD tag files, GARFIELD then generates a list of independent genetic variants ($r^2 \geq 0.01$) for a given reference population, and calculates whether any of the listed variants or those in direct LD ($r^2 \geq 0.8$) can be found within the DNA feature of interest (Iotchkova *et al.*, 2019; Han *et al.*, 2020). Logistic regression is then used to produce both an odds ratio and enrichment p-values for the grouped overlapping variants on a per-feature basis. Enrichment of AS variants from European sub-analysis was assessed at four GWAS

significance thresholds: $p < 1$, 0.1, $5 \times 10{-}7$ and $5 \times 10{-}8$. The UK10K variant set, pre-processed in GARFIELD, was used as the reference population for these analyses; with correction for multiple testing performed using the Bonferroni method set to a confidence interval of 95% (Q-value = 0.05). GARFIELD enrichment results were plotted using custom scripts and Circlize (v.0.4.11).

2.7.8. Script and data availability

Count data for RNA-seq, ATAC-seq, ChIPm and eRNA (raw and normalised), Capture-C mean pile-up tracks and PeakY scores, and ChromHMM data can be found deposited online on the Knight group Zenodo page (https://zenodo.org/records/7576256). Accession numbers are listed in the key resources table of Brown *et al.* (2023). Tiled-C hdf5 files will be deposited online in the Knight lab Zenodo at a later date. This thesis additionally analyses existing, publicly available data (GWAS data and PCHiC data) with their accession numbers also being be found in the key resources table of the publication. Modified GARFIELD (https://github.com/CIDavidson/ChromHMM_GARFIELD) will be deposited at a later date on GitHub.

"Deep in the human unconscious is a pervasive need for a logical universe that makes sense. But the real universe is always one step beyond logic."

- Frank Herbert, Dune (1965)

3

INTEGRATIVE MODELLING OF THE EPIGENETIC LANDSCAPE IN AS-ASSOCIATED GENETIC LOCI SHOWS ENRICHMENT OF GWAS SNPs FOR *CIS*-REGULATORY ELEMENTS

Contents **Pg.**

3.1. Chapter introduction …….. 115

3.2. Functional characterisation of the *NOS2* locus ……. 123

3.3. Enrichment of AS-associated SNPs in regulatory regions ….. …….. 130

3.4. Chapter discussion ……. 161

This chapter describes the integration of genomic, transcriptomic, and epigenetic datasets – alongside variant enrichment analyses – to further our understanding of the regulatory landscape surrounding AS-associated GWAS genetic risk loci in T cells, B cells, and monocytes.

3.1. Chapter introduction

3.1.1. Identification of disease-specific gene expression networks

As highlighted in the general introduction, GWAS – and related studies such as the AS immunochips employed by Cortes *et al.* (2013) and Ellinghaus *et al.* (2016) – produce

trait-association signals surrounding polymorphic variants upon comparison of healthy and patient cohorts. The variant for a given locus with the highest statistical strength of association in the analysis is often termed the lead SNP. Despite such high signals, calling the lead SNP as a disease-driving variant directly in a condition is problematic, as identification of causal variants through GWAS is complicated by factors such as LD and gene density. As LD is defined as the non-random association between two variants in a given population, the causal effects of one variant are often masked by the presence of another when they are in LD (Abell *et al.*, 2022). Greater gene density additionally increases the number of potentially linked disease-associated candidates at a given locus. Statistical fine-mapping of the lead SNP usually narrows the signal to a credible set of potentially causative SNPs based on LD (section 1.3.4). Without further study however, all credible set variants putatively contribute towards the observed phenotype and regulation of genes in the locus. It is therefore necessary to carry out multiple different post-GWAS approaches to shorten the list of associated variants and further scrutinise those which could be truly causal.

Disease-associated genetic variants may be linked to the expression patterns of one or more associated genes. Observed SNP-driven changes in target gene expression are known as eQTLs. Establishing eQTLs for every gene-to-genetic variant pairing is however both time-consuming and requires hundreds of healthy volunteer and patient samples to have enough statistical power to detect allelic changes in expression due to one SNP. This is particularly difficult in diseases such as AS, which demonstrate hundreds of risk-association loci with relatively small disease contribution effect sizes. In these cases, multiple different techniques – including pathway analysis – can be used to prioritise which variants to investigate further. Pathway analyses operate by first annotating each gene proximal to SNPs with a direct function from gene ontology

databases. The resultant gene list is then statistically tested for significant overrepresentation of any one term (such as gene involvement in inflammatory or antigen presentation pathways), implicating that ontology in disease pathogenesis (Reimand *et al.*, 2019). Pathway analyses therefore allow for the initial identification of functionally related GWAS risk-loci that could share the same biological pathway, narrowing loci for functional follow-up to those with direct context-specific relevance (White *et al.*, 2020).

3.1.2. Nitric oxide synthase 2 as an ankylosing spondylitis-associated locus

To date, pathway analysis of AS has only been conducted in a handful of studies (Ding *et al.*, 2018; Yu *et al.*, 2020; Brown *et al.*, 2023; Sun *et al.*, 2023). From the project outset, I wanted to identify functional AS GWAS SNPs and chose the nitric oxide synthase 2 (*NOS2*) locus due to its high GWAS association value (p-value = $1.2x10^{-10}$) and its direct relevance towards inflammation (Cortes *et al.*, 2013). The limited pathway analyses also show that *NOS2* is an important player in gut inflammation in AS patients, with the locus constituting part of a highly druggable pro-inflammatory pathway alongside other AS-associated genes (such as *JAK2*) that can be targeted with repurposed therapeutic compounds such as methazolamide (Ellinghaus *et al.*, 2016; Sun *et al.*, 2023). *NOS2* encodes for an inducible form of the enzyme nitric oxide synthase, and expression of the peptide leads to the production of large quantities of nitric oxide (NO) in response to a wide array of stimuli. These include immune-relevant proinflammatory compounds (including LPS) and several AS-relevant inflammatory cytokines (such as TNFa and IFNγ) (Salim *et al.*, 2016). The function of *NOS2* is remarkably pleiotropic, with some of its key effects in the immune system relating to the direct bactericidal effects of NO and its ability to act as a pro-inflammatory mediator through the nitrosylation of peptide residues; influencing the stability of the tertiary protein structure of multiple intracellular signalling molecules (Shinozaki *et al.*, 2014; Xue *et al.*, 2018; Zhao *et al.*, 2021).

NOS2 is expressed natively in a number of murine cell types, including monocytes, macrophages, dendritic cells, and both T helper and cytotoxic T cells (Xue *et al.*, 2018; Sonar *et al.*, 2019). Interestingly, the locus shows genetic association not only with AS (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016), but also with other arthritic and inflammatory conditions including other spondyloarthropathies such as psoriatic arthritis and juvenile idiopathic arthritis (Thompson *et al.*, 2012; Stuart *et al.*, 2015). The locus additionally shows genome-wide significance with common AS comorbidities such as IBD and acute anterior uveitis (Jostins *et al.*, 2012; Dhillon *et al.*, 2014; Huang *et al.*, 2020). Notably, monocytes and their subsets in human psoriatic arthritis patients show robust increases in *NOS2* expression upon LPS and IFNγ stimulation when compared to those of healthy volunteers (Zhong *et al*., 2018). With a strong AS-association and clear mechanistic links towards both immune-related and inflammatory conditions, *NOS2* presents as one of the most promising AS risk loci for functional characterisation. Despite such strong links to inflammatory diseases however, studies showing how *NOS2* expression may differ in disease are seldom linked to any particular genetic variant. The epigenetic environment surrounding the *NOS2* locus in human primary immune cells additionally remains sorely understudied, making identification of any putative pathogenic variant towards targeted gene analysis difficult.

3.1.3. Using ChromHMM to identify cell-type specific genomic annotations

The epigenetic environment of the *NOS2* locus is to be analysed in this chapter using Chromatin hidden Markov models (ChromHMM), which are a method employed to annotate the genome by combining epigenetic (Ernst & Kellis, 2017) datasets such as ChIP-seq (which assays for the enrichment of various biologically divergent histone post-translational modification marks) and ATAC-seq or DNAse-seq (which probes for regions of open chromatin). The model identifies length-defined stretches of chromatin

signatures (bins) along the DNA that demarcate major reoccurring patterns (emission states) and their genomic location from the combinatorial epigenetic data. As mentioned in the general introduction, manual curation of ChromHMM emission states would, for instance, be able to discern the H3K4me3 and H3K27ac enriched promoters from the H3K27ac and H3K4me1 presenting enhancer-like elements. The default length of a ChromHMM bin is 200bp, which constitutes the average length of a singular nucleosome and associated spacer regions (Luger *et al.*, 1997; Ernst & Kellis, 2017). The biological function of these 200bp regions can be inferred from the prevalence and relative strength of each epigenetic mark within each bin, thereby mapping the biological function of DNA and its associated proteins at a single-nucleosome resolution. This model is then to be intersected with both fine-mapped SNPs at specific loci of interest (i.e. *NOS2*) and later other genome-wide non-MHC genetic variants. This can then help identify whether any putatively causal AS-associated genetic variants that lie outside of protein-coding regions reside in predicted *cis*-regulatory regions such as promoters or enhancers.

We in the Knight lab have previously used ChromHMM to define genomic annotations in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes from both healthy volunteers and AS patients (Brown *et al.*, 2023). This model incorporates ATAC-seq and RNA-seq data alongside ChIPmentation-seq data for the promoter-specific and transcription activity-associated mark H3K4me3, the general contact and activity-associated mark H3K27ac, and the poised enhancer-associated mark H3K4me1. State annotation was facilitated by exploring TSS enrichment and genomic annotations (such as CpG islands and exons) for each state. We identified 14 emission states, with state 1 representing the gene TSS, states 2 to 4 representing flanking promoter-like states, 5 to 7 representing states associated with transcription, 8 to 11 representing various strengths of *cis*-regulatory or enhancer-like elements, and 12 to 14 representing quiescent or intronic DNA (**Fig. 3.1**).

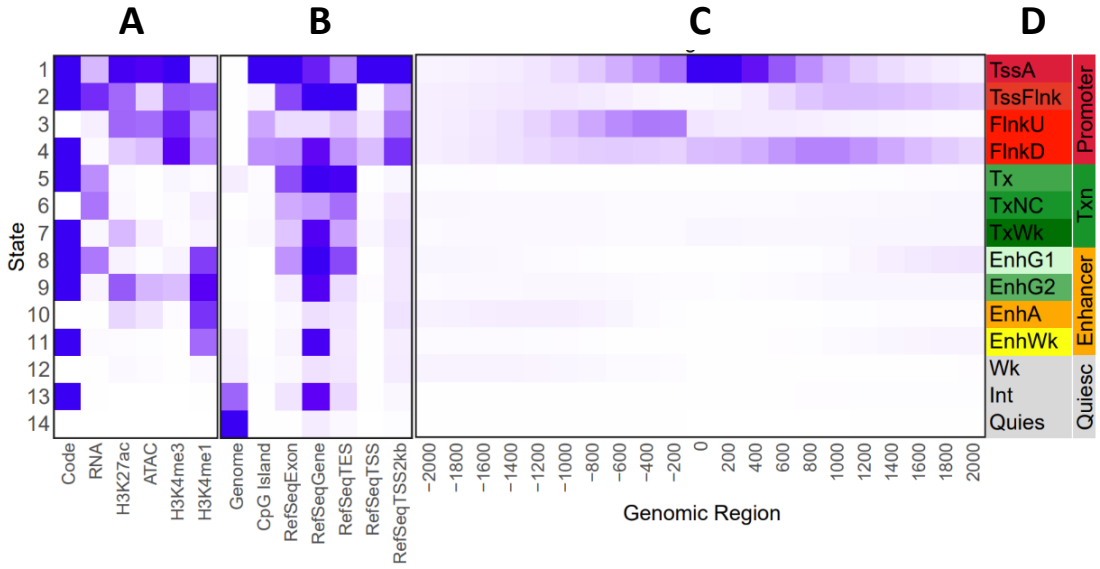

Figure 3.1. ChromHMM emission states from the Knight group. A 14-state ChromHMM model was generated from AS-specific and healthy volunteer epigenetic datasets. (**A**) Abundance of epigenetic inputs in each ChromHMM emission state. (**B**) Overlap of ChromHMM state with genomic features. These include the percentage that each state makes up of the genome (Genome%), the relative enrichment for CpG islands within each state (stronger CpG values are indicative of a greater ability to be modulated through methylation of cytosine) (CpGIsland), whether these states reside in exons (RefSeqExon) or in the wider gene (RefSeqGene), and whether they are enriched in transcriptional start sites (TSS) (RefSeqTSS), transcriptional end sites (TES) (RefSeqTES), or are found within 2 kbp of the TSS (RefSeqTSS2kb). (**C**) The TSS neighbourhood shows the positional enrichment of each state relative to the TSS. The strength of each input is shown by the intensity of the blue boxes. (**D**) Biological identity of each emission state is shown on the far right of the figure. Tx indicates transcriptional states. Enh indicates enhancer-like states. Figure adapted from Brown *et al.* (2023).

3.1.4. Enrichment of disease-associated genetic variants at regulatory genomic regions

Genetic variants which change the underlying epigenetic structure and activation state of regulatory elements often have the most robust effect on gene expression (Wang *et al.*, 2023). Indeed, Kasowski *et al.* (2013) found that differential gene expression was most often observed when greater than 60% of linked enhancers varied in their activation state. These variable enhancer regions additionally demonstrated greater enrichment of SNPs relative to their non-variable counterparts. These molecular changes could be reflected in differential ChromHMM states arising between the variant-bearing and control cohorts**.** A number of post-ChromHMM analyses can be used to (A) deduce the extent of epigenetic differences between two groups and (B) facilitate identification of emission states that show enrichment for specific AS-associated variants:

[A] In order to ascertain the extent of epigenetic differences between AS patients and healthy volunteers, in this chapter I employ quantitative methods (i.e. comparing the total number of bins sharing chromatin state identities), MCA (to detect how variable the underlying ChromHMM models are between patients and controls), and Fisher's permutation analysis (to detect chromatin state switches at individual bins).

[B] The R package GARFIELD can identify whether disease-associated polymorphisms within a given ChromHMM emission state are present at frequencies higher or lower than would be expected through chance alone (Iotchkova *et al.*, 2019; Vu & Ernst, 2022). GARFIELD enriches SNPs from GWAS summary data in functional epigenetic annotation data whilst accounting for LD structure and gene density at multiple genome-wide significance thresholds. GARFIELD has previously been shown to successfully enrich immune-relevant regions of open chromatin assayed through ATAC-seq for

multiple sclerosis-associated GWAS variants on a celltype-specific basis (Qin *et al.*, 2023). In this chapter, I use GARFIELD to functionally annotate Cortes *et al.* (2013) SNPs for my ChromHMM emission states

Some of the data in this chapter was generated and analysed by other members of the Knight group and has been published (Brown *et al.*, 2023). A table of my contributions can be found in the methods and materials chapter (**Table 2.2**). In this chapter, I have added to this work by considering repressive histone post-translational modifications in the epigenomic model and added the functional analysis of AS-associated SNPs within $CD19^+$ B cell populations.

3.1.5. Chapter aims

The overall aim of this chapter is to identify epigenetic differences that arise between healthy individuals and AS patients and to characterise genomic regions for further functional follow-up. The specific chapter aims are:

[1] To analyse the *NOS2* AS-associated GWAS risk locus with in-house epigenetic data and fine-mapped credible SNP lists to identify regions of differential chromatin accessibility, histone post-translational modification identity, and gene expression that could implicate causal genetic variants.

[2] To integrate the epigenetic data into a singular model with ChromHMM and segment the genome into biologically meaningful genomic segmentations.

[3] To identify differential and variant enriched genomic segmentations that could be important in disease.

[4] To analyse the impact of MHC on SNP enrichment patterns in regulatory regions.

[5] Investigate transition bias for each emission state in AS patients and healthy controls.

3.2. Functional characterisation of the *NOS2* locus

3.2.1. ChromHMM analysis of the *NOS2* locus in T cells and monocytes

I first aimed to understand the regulatory genomic landscape on a locus-specific basis in order to identify non-coding regulatory regions that overlap with AS-relevant genetic variants. I chose the *NOS2* locus due to its significant genetic association and relevant functional characteristics. I analysed the locus using our previously developed AS ChromHMM (Brown *et al.*, 2023). The lead SNP for *NOS2* (*rs2531875*; p-value = $1.2x10^{-10}$) does not overlap any ChromHMM predicted regulatory elements nor any differential peaks for any of the individual epigenetic assays, including H3K4me3, H3K27ac, ATAC, or RNA, suggesting it is not the causal SNP in either $CD4^+$ T cells, $CD8^+$ T cells, or $CD14^+$ monocytes (**Fig. 3.2**). However, cell-type specific ChromHMM regulatory regions are present at the NOS2 locus near some of the AS-associated credible set variants. For example, *rs7210647* is located between two monocyte-specific enhancers 14 kbp upstream of *NOS2* (**Fig. 3.2: Red shaded box**).

Total RNA-seq tracks can determine whether active transcription is observed in coding regions and can also identify regulatory regions through the detection of eRNAs. Whilst the role of eRNAs in enhancer biology remains to be fully elucidated, they appear to be involved in loop stabilisation through promoting cohesin binding and histone acetyltransferase catalytic function (Sartorelli & Lauberth, 2020). Using total RNA-seq data, non-coding transcription can be seen underlying the three monocyte specific enhancers. Differential promoter use is an important driver of functional variation in the human genome (Garieri *et al.*, 2017). *NOS2* has two alternative promoters that are used differentially between $CD14^+$ monocytes compared to $CD4^+$ and $CD8^+$ T cells (**Fig 3.2: Blue shaded box**).

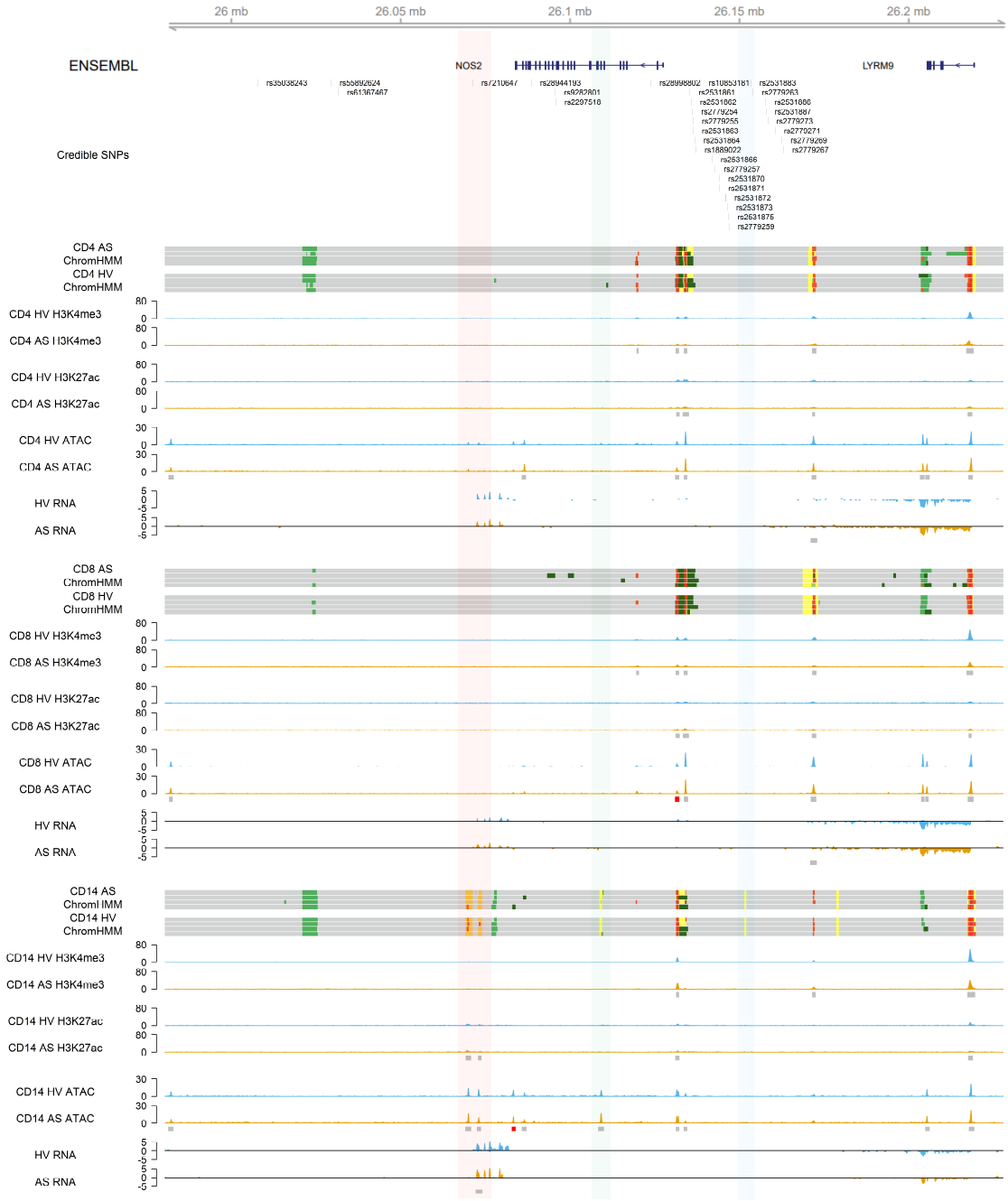

Figure 3.2. The epigenetic landscape surrounding the *NOS2* AS-associated GWAS risk locus in CD4$^+$ T cells, CD8$^+$ T cells, and CD14$^+$ monocytes. This plot shows the overlap of AS-associated SNPs with context-specific epigenetic features in CD4$^+$ T cells, CD8$^+$ T cells, and CD14$^+$ monocytes. From the top down: 'ENSEBML' shows the genomic position of selected coding regions in the *NOS2* locus from the Ensembl consortium. 'Lead SNP' shows *rs2531875*, the lead AS-associated SNP from Cortes *et al.* (2013). 'Credible SNPs' shows fine-mapped SNPs produced for the *NOS2* locus. Three panels are then shown underneath for CD4$^+$ T cells (top), CD8$^+$ T cells (middle), and CD14$^+$ monocytes (bottom). Each panel shows: ChromHMM tracks from the 14-emission state model (**Fig. 3.1**) (Brown *et al.*, 2023) in both AS patients and healthy volunteers and differential epigenomic assay tracks for H3K27ac, H3K4me3, ATAC, and

RNA that underlie the ChromHMM (Orange = AS patients, blue = healthy volunteers). Grey blocks underneath each track set show MACS2-called peaks. Red and blue blocks represent differential MACS2-called peaks between AS patients and healthy volunteers, with blue indicating higher peak heights in healthy volunteers and red higher peak heights in AS patients. RNA tracks at the bottom show transcription from the positive (above the grey track line) or negative (below the grey track line) DNA strand. MACS2 called peaks and their associated differentials under the RNA tracks are specific for eRNA. The red shaded box shows the location of the $CD14^+$ monocyte context-specific enhancers and the SNP *rs7210647* that lies in-between them. The blue shaded box shows the location of a $CD14^+$ monocyte-specific promoter not seen in the T cell populations. The green shaded box shows the position of a weak intronic enhancer specific to $CD14^+$ monocytes. Sample numbers for each assay given in Table 2.2.

Furthermore, another promoter-like region is present in $CD4^+$ and $CD8^+$ in a *NOS2* intron region and there are two $CD14^+$ monocyte specific enhancer-like regions in the *NOS2* gene itself (surrounding exon six and roughly 20 kbp upstream of the transcription start-site) that bear the hallmarks of weak or poised enhancers (**Fig. 3.2: Green shaded box**). Two distinct regions of LD were found at the *NOS2* locus using LDMatrix (**Fig 3.3**), which could explain the two independent GWAS association signals detected by Cortes *et al.* (2013). One of these LD blocks centres upon the promoter region, whilst the other lies upstream of the gene and overlaps the $CD14^+$ monocyte-specific ChromHMM enhancers. Taken together, these findings suggest that regulation of the *NOS2* locus in $CD14^+$ monocytes could be influenced by AS-associated SNPs in two distinct regions.

3.2.2 NOS2 is epigenetically silenced in monocyte-derived macrophages

NOS2 is widely expressed in the murine immune system, particularly in myeloid cells such as monocytes and macrophages (Gross *et al.*, 2014; Sonar *et al.*, 2019). In contrast, *NOS2* is not usually expressed in healthy human immune cells. This is presumably due to the strong CpG methylation in close proximity to the *NOS2* TSS that suggests long-term repression of the gene (Gross *et al.*, 2014). As such, there are few models to express *NOS2* in human monocytes and macrophages. One study by Landes *et al.* (2015)

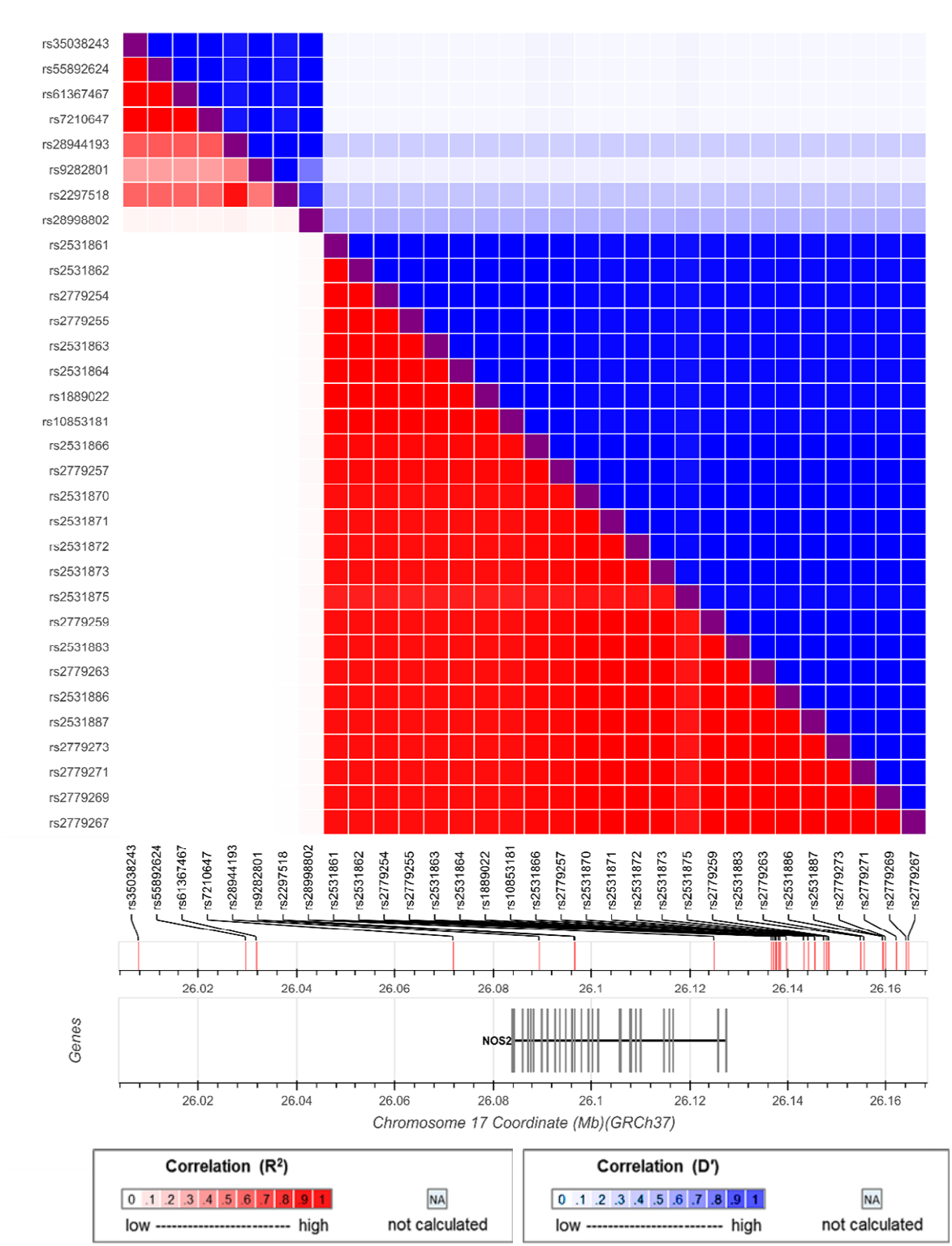

Figure 3.3. Linkage disequilibrium at the *NOS2* locus. Heatmap showing LD between SNPs in the NOS2 credible set in a EUR(GBR) population. Two distinct blocks of LD can be seen separated by the SNP *rs2899880*. Data analysed using LDmatrix from LDlink 5.0 (www.ldlink.nci.nih.gov).

expressed *NOS2* by stimulating monocyte-derived macrophages with muramyl dipeptide (MDP) and guanylyl muramyl dipeptide (GMDP). To establish a consistent *NOS2* expression model within human primary cells, I aimed to replicate the results of the Landes *et al.* (2015) study to induce expression of *NOS2* with MDP and GMDP using qRT-PCR as a functional read-out. I differentiated MACS-isolated $CD14^+$ monocytes into macrophages and stimulated with either MDP or GMDP for 1 hour, 2 hours, 6 hours, 12 hours, 24 hours, or 48 hours. A control population of monocyte-derived macrophages that received no stimulation (0-hour) was used for comparison. In contract to Landes *et al.* (2015), I observed no differential *NOS2* induction in monocyte-derived macrophages at any of the defined time-points (**Fig 3.4**).

I next sought to characterise the regulatory elements that lead to silencing of the *NOS2* gene in the immune cell types of interest. CpG islands are typically enriched by histones bearing the H3K27me3 post-translational modification mark, which associated with the silencing PRC2 Polycomb-complex. H3K27me3-mediated repression is often seen at developmentally important genes that could be expressed upon maturation and differentiation of the cell type, in addition to alternative stimulation conditions (McCarthy *et al.*, 2021; Kfourey-Beaumont *et al.*, 2022). To ascertain whether *NOS2* could be repressed through the action of H3K27me3 – and whether this repression may differ between AS patient and healthy cohorts – I performed ChIPmentation-seq of the H3K27me3 histone mark upon chromatin samples isolated from four AS patients with active disease (BASDAI score ≥ 4.0 and MRI evidence of inflammation) and four healthy volunteers. I observed a peak of H3K27me3 enrichment in close proximity to the *NOS2* promoter (**Fig. 3.5: Red shaded box**). This peak was invariant across $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes, and no differential peaks were seen between the healthy volunteers and AS patients in any of the cell types. This suggests that *NOS2* is strongly

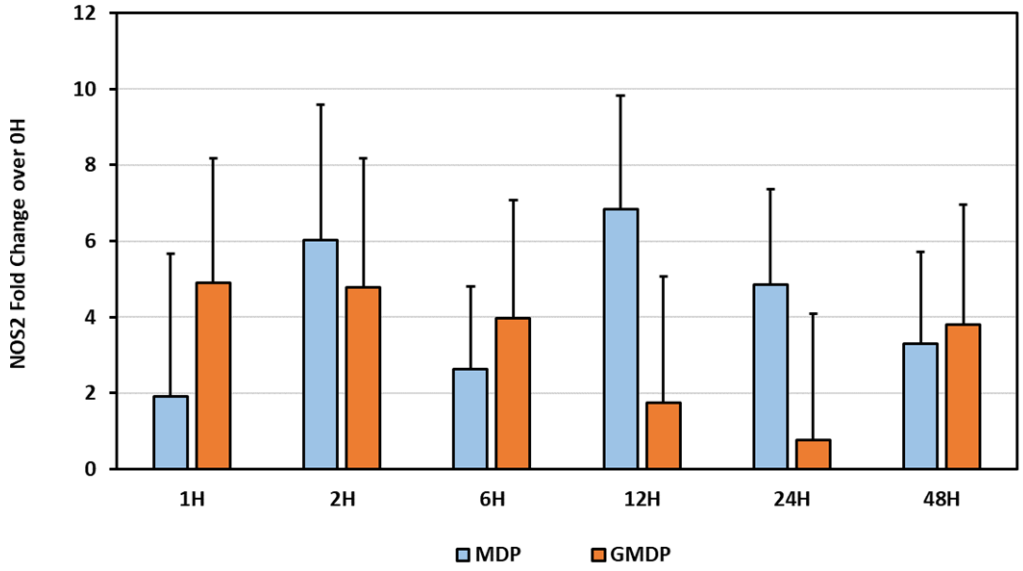

Figure 3.4. No induction of *NOS2* is seen upon stimulation with either MDP or GMDP in CD14^{+} monocyte-derived macrophages. Bar chart showing RT-qPCR results for *NOS2* in monocyte-derived macrophages 1-hour (1H), 2-hours (2H), 6-hours (6H), 12-hours (12H), 24-hours (24H), and 48-hours (48H) post-stimulation with either MDP or GMDP (n = 3 for each timepoint). Results were normalised to control Cq values for β-actin, and are represented as the fold change over the unstimulated control (0H). Neither MDP (blue) nor GMDP (orange) were able to stimulate significant differential expression of *NOS2* at any time point; as measured by one-way ANOVA with Tukey-Kramer post-hoc test. Error bars show one standard deviation of the mean.

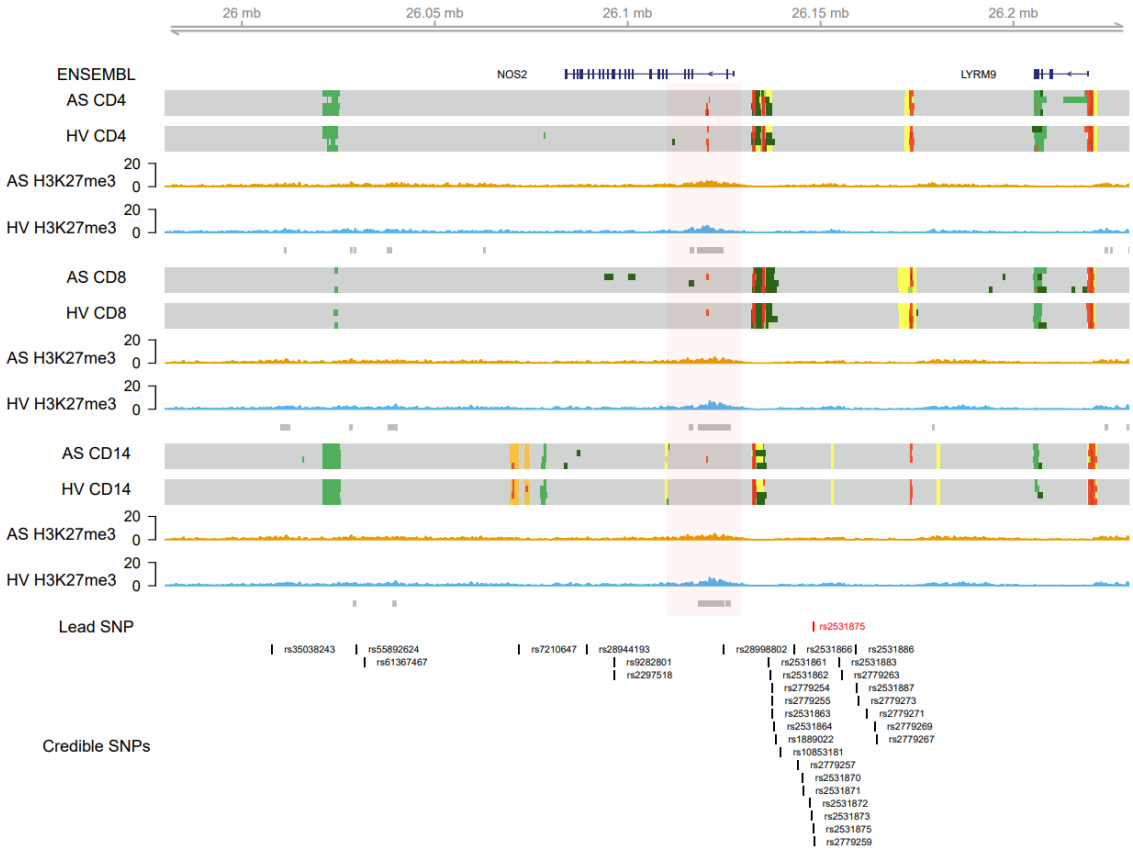

Figure 3.5. The expression of NOS2 is repressed within CD4$^+$ and CD8$^+$ T cells and CD14$^+$ monocytes through an intragenic H3K27me3-rich silencing element. ChromHMM from Brown *et al.* (2023) (Fig. 3.1). From the top down: ChromHMM tracks for AS patients and healthy volunteers for CD4$^+$ T cells, CD8$^+$ T cells, and CD14$^+$ monocytes. Tracks showing the cell-specific enrichment for the H3K27me3 post-translational histone modification mark (AS H3K27me3 and HV H3K27me3) are shown underneath their respective cell type (n = 4 AS patients; n = 4 healthy volunteers). Orange tracks show peaks for AS H3K27me3-rich features whilst blue shows the healthy equivalent. Grey blocks underneath each track show the MACS2-called peaks between AS patients and healthy volunteers. No MACS2-called H3K27me3 peaks in the NOS2 locus were differential in the DESeq2 analysis. 'Lead SNP' shows the NOS2 AS-associated lead SNP from Cortes *et al.* (2013). 'Credible SNPs' shows the fine-mapped credible set produced by Dr. Anna Sanniti have been added underneath to intersect putative causal SNPs with the H3K27me3 silencing signal.

epigenetically silenced across the three cell types by the Polycomb group of proteins in both the AS patients and healthy volunteers; but it does not exclude a context-specific role for the locus in other cell types.

Looking more broadly across the genome, I identified 49,825 MACS2-called H3K27me3 peaks in $CD4^+$ T cells, 75,604 peaks in $CD8^+$ T cells, and 45,559 peaks in $CD14^+$ monocytes. As there is a growing body of evidence to suggest B cell involvement in AS pathogenesis (Wilbrink et al., 2021), I additionally isolated $CD19^+$ B cells from the four AS patients and healthy volunteers and performed H3K27me3 ChIPmentation-seq to characterise B cell specific silencing elements. 19,779 H3K27me3 peaks were identified in B cells. DESeq2 was used to determine which of the H3K27me3 peaks were differential between AS patients and healthy volunteers (**Fig. 3.6**). Interestingly, the only H3K27me3 peak to pass both the DESeq2 $\log_2$ fold change threshold of 0.585 and the p-value threshold of $1x10^{-5}$ was found internal to the gene *DCLBD1* (discoidin, CUB and LCCL domain containing 1) in $CD14^+$ monocytes ($\log_2$ fold change = -0.97; adjusted p-value = 4.06×10^{-7}). *DCLBD1* is involved in the integrin signalling pathway and is found to be depleted within invasive breast cancers (Cardin *et al.*, 2021), yet its potential role in AS pathogenesis is poorly understood owing to lack of study into gene function. These results therefore suggest that H3K27me3 signals might not vary significantly between AS patients and healthy volunteers.

3.3. Enrichment of AS-associated SNPs in regulatory regions

3.3.1. An updated epigenomic map in AS incorporating Polycomb repression-associated post-translational histone modifications

I sought to maximise the informativeness of the H3K27me3 ChIPmentation-seq data by incorporating it into a new ChromHMM model to define the chromatin states in immune

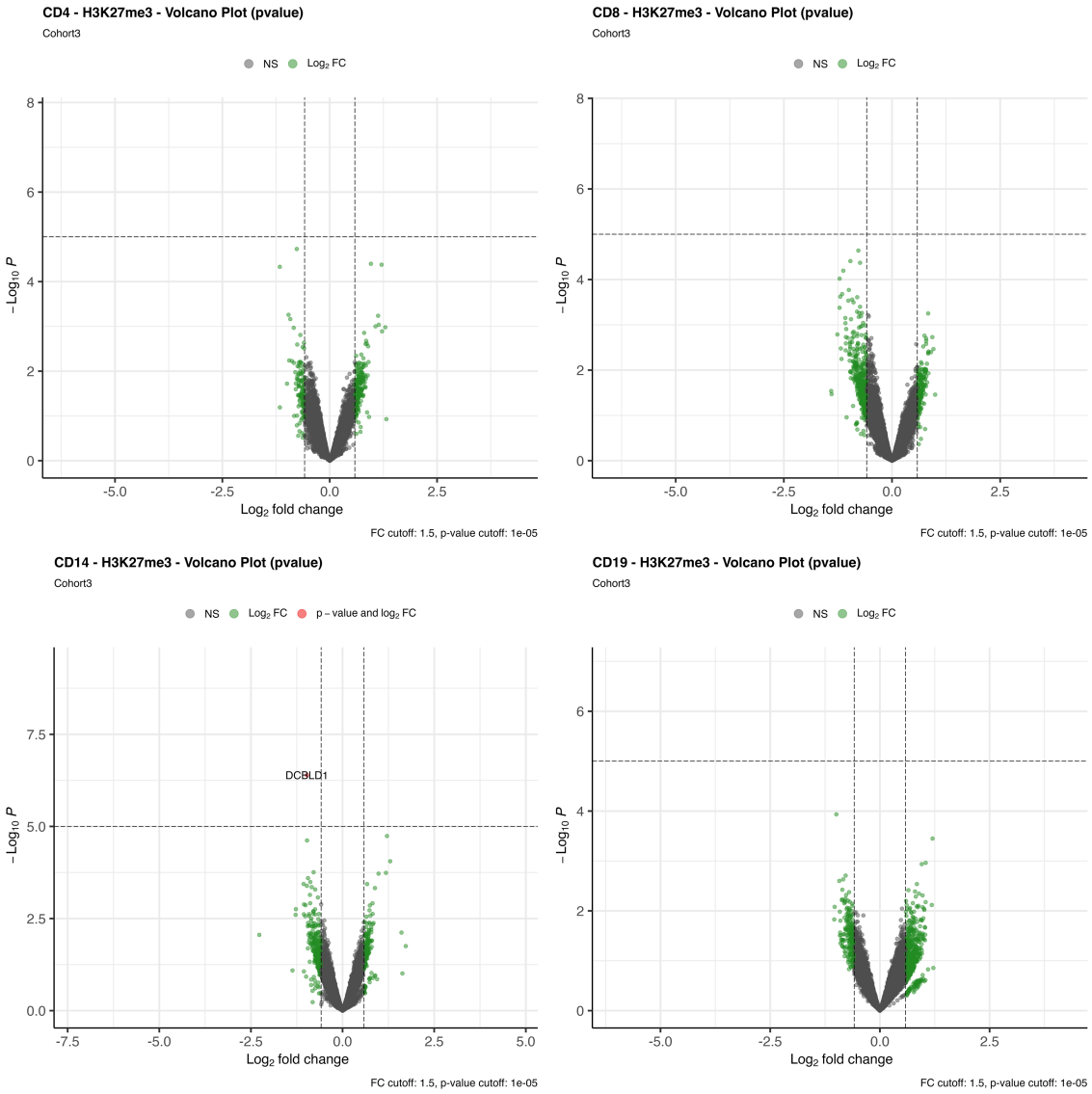

Figure 3.6. Differential genes linked to H3K27me3 peaks within all investigated cell types. Figure above is comprised of four volcano plots showing differential genes linked to MACS2-called H3K27me3 post-translational histone modification peaks in $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). Genes were linked to MACS2-called H3K27me3 peaks by the closest RefSeq transcriptional start-site within 50kbp of the called peak. Log_2 old change (FC) of the gene has been shown on the x-axis whilst the log_{10} of the associated p-value has been shown on the y-axis of each plot. Genes were assumed to be differential if they passed a log_2 FC threshold of 1.5 upregulated or downregulated and a p-value threshold of $1x10^{-5}$. These thresholds have been plotted on the x and y axes, respectively. In all cell types, only a single H3K27me3 peak in the *DCBLD*1 gene in $CD14^+$ monocytes was found to pass both filters. H3K27me3 peaks were kept if present in at least 20% of analysed individuals and if greater than 10 reads were attributed to each peak. Results have been produced and plotted using the MACS2 and the DESeq2 R packages.

cells from AS patients and healthy volunteers. I integrated the H3K27me3 ChIPmentation-seq data with H3K4me1, H3K4me3, H3K27ac, RNA, and ATAC datasets that were previously used to generate the 14-emission state ChromHMM in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes (**Fig. 3.1**) (Brown *et al.*, 2023). The same datasets were also integrated for $CD19^+$ B cells, with the above epigenetic assays pre-processed by Dr. Andrew Brown in addition to my H3K27me3 data (**Table 2.2**).

The exact number of states output by ChromHMM can vary by experiment, with factors such as cell type, epigenetic marks used, and the number of output states specified in the ChromHMM resulting in profound model differences. Accounting for this, I first defined the optimal size for my ChromHMM output by creating models that contained between 1 and 30 emission states (**Fig. 3.7**). Correlation values for each of these emission states were calculated from the emission probabilities for all states in the 30-state model, and compared with that in each of the lower-state models. The calculated correlation values correspond to how well that particular emission state can be captured by other states within the lower-state models, and where each state could begin to stop splitting into biologically meaningful ChromHMM emissions. I observed that the median correlation value began to plateau upon the model reaching 16 emission states; suggesting diminishing increases in model informativeness passed this point and resolution of all baseline, biologically divergent key chromatin states achievable from the data integration.

I generated the 16-state ChromHMM model and compared this with the 14-state model we had previously produced to facilitate identification of the biology behind each emission state (**Fig 3.8**). I found that state compositions between the two models were near identical for the first 14 states, with two novel states in U15 and U16 arising from the addition of H3K27me3. U15 was enriched for H3K27me3 signals and showed

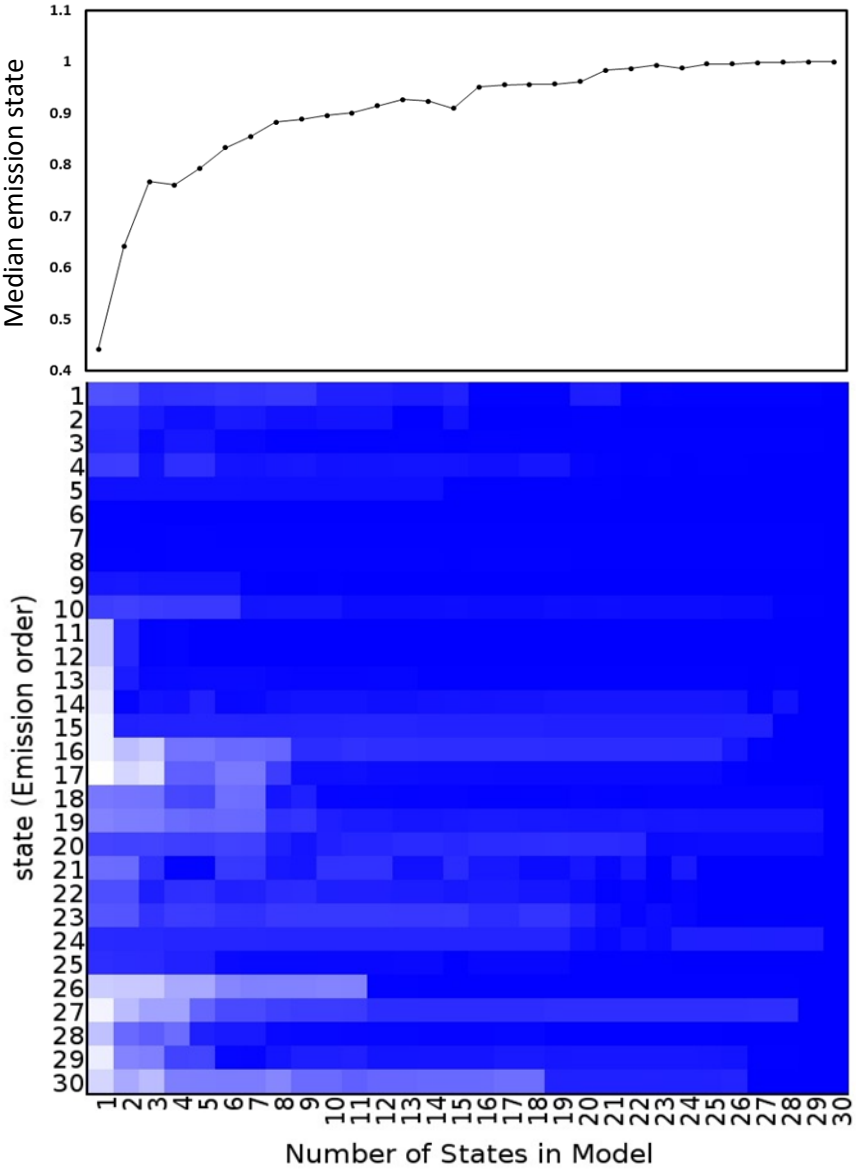

Figure 3.7. Correlation of emission parameters reveals the 16-emission state ChromHMM to be the most appropriate model size upon incorporation of H3K27me3 ChIP-seq data. The line plot (top panel) demonstrates the median correlation values for all emission states in each ChromHMM size iteration from 1 to 30. The heatmap plot (bottom panel) shows how well a specific emission state correlates to other emission states within ChromHMM iterations of different emission sizes. Each row corresponds to a state from a maximal 30-state ChromHMM and each column a model with a different number of states. The intensity of the blue colour in each individual cell is indicative of the maximum emission parameter correlation of any state in the model of the column with the state of the row from the 30-state model. As diminishing increases in median correlation values are seen past the 16-state model, it was chosen as the most appropriate size for subsequent ChromHMM generation. All 30 detected emission states are additionally well captured in the 16-state model, suggesting states begin to split passed 16 emission states based on variation within the emission states themselves; rather than capturing key biologically divergent, baseline chromatin states.

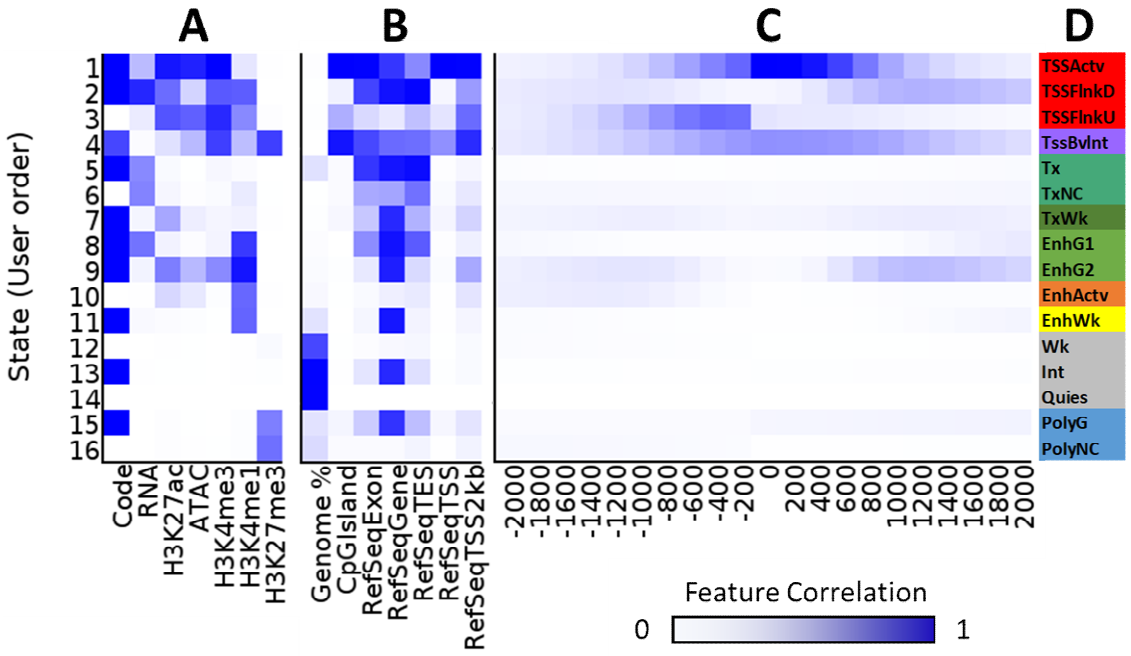

Figure 3.8. Identifying the biology behind each emission state within the 16-state ChromHMM. Figure above shows the enrichment of genomic features for each of the expanded 16 ChromHMM emission states generated upon inclusion of H3K27me3 ChIP-seq data. (A) Heatmap showing the correlation values for each of the feeder epigenomic assays. (B) Heatmap showing the correlation values and overlap for several common genomic features within each emission state, such as the percentage total genomic coverage of each emission state, CpG island presence, whether states enrich for exons or the wider gene region as a whole, and enrichment at the RefSeq transcriptional start (TSS) and end (TES) sites. (C) Heatmap showing the breakdown of the RefSeq TSS enrichment scores from (B) in 200bp increments. (D) Suspected biological function of each of the 16 emission states based upon the combination of all preceding heatmaps. TSSActv is indicative of an active TSS element; TSSFlnkD and TSSFlnkU of downstream and upstream promoter elements; TSSBvlnt of a bivalent promoter TSS element bearing an H3K27me3 signal; Tx and TxNC of active transcription from a genic and non-coding sources; TxWk of overall weaker levels of transcription; EnhG1 and EnhG2 of enhancers overlapping coding elements of the genome; EnhActv and EnhWk of active and poised enhancers; Wk of an overall weak signal for all feeder epigenomic assays; Int for non-transcribed intronic regions; Quies for DNA of either no or unknown functionality; and PolyG and PolyNC for Polycomb-repressed genes and non-coding regions respectively. Intensity of the colour blue within each cell is indicative of a higher correlation value between each genomic feature or annotation and each respective emission state.

overlap with coding regions of the genome, suggesting this state represented long-term Polycomb repressed genes. U16 instead enriched solely for H3K27me3, suggesting the state represented regions of Polycomb repression that are not associated with genes. U4, having previously been identified as a downstream flanking promoter element, now demonstrated a robust H3K27me3 signal. This state is likely representative of a bivalent promoter, bearing both the activity-associated H3K4me3 and the repression-associated H3K27me3 histone marks (Mas *et al.*, 2018). Consistent with this, U4 shows a high degree of TSS enrichment, although much weaker than U1 TSSActv. These bivalent promoters typically exist in a poised state that can result in activation or repression of the associated gene through loss of one of the histone marks as a cell commits and terminally differentiates, and also form an important feature in mounting responses to the extracellular environment and stimuli (Kitazawa *et al.*, 2021).

The novel U15 and U16 states mostly arise *de novo* from regions that were previously characterised as quiescent, intronic, or weak by the 14-emission state ChromHMM; suggesting its superiority over the 14-emission state model for the identification of silencing elements at AS-associated GWAS risk loci. Comparison of both models at the *NOS2* locus, for instance, reveals a large block of H3K27me3-rich region in close proximity to the promoter that could be responsible for silencing the gene that the 14-state model predicts as quiescent (**Fig. 3.9: Red dashed box**). This block also contains the credible set SNP *rs28998802*, which may affect the silencing of *NOS2* in the investigated cell types. Additionally, an internal U1 TssActv promoter element within the *NOS2* gene seen primarily in the T cell populations detected by the 14-state ChromHMM is now identified as a bivalent-promoter-like U4 TssBvlnt ChromHMM emission state in the same region. Overall, the novel 16-emission state model is efficient at identifying



Figure 3.9. The 16-emission state ChromHMM reveals a number of repression-associated regulatory regions that were hidden in the 14-emission state ChromHMM upon incorporation of H3K27me3 data. Figure above shows tracks for the 14-emission state ChromHMM (S14) in Brown *et al.* (2023) compared to the 16-emission state model produced in this chapter (S16) for $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells in AS patients (AS) (n = 4) and healthy volunteers (HV) (n = 4). 'ENSEMBL' track shows the location of selected genes from the Ensembl consortium in the locus. 'Credible SNPs' shows the credible set of *NOS2* SNPs produced by Dr. Anna Sanniti based on the association signal from Cortes *et al.* (2013). ChromHMM colours are consistent with figure 3.8. A large block of Polycomb-repressed U15 and U16 emission state bins (regions of light blue shown in red-dashed box) is seen in close proximity to the *NOS2* promoter. This region also contains a U4 bivalent promoter-like emission state in T cells, although this is highly variable. The SNP *rs28998802* found in the *NOS2* credible set can be found within this repressed block in all cell types.

new silencing genomic elements which can contain AS-associated genetic variants that did not show interesting intersection with regulatory regions in the 14-emission state model.

3.3.2. Global and state-specific differences in chromatin conformation between AS patients and healthy volunteers

I next sought to ascertain whether any of the states in the new ChromHMM model demonstrated differential enrichment between AS patients and healthy volunteers. A skew in the number of any of the 16 states to the side of disease when compared to a healthy population could be indicative of global changes in chromosome composition in the disease state that could be linked to AS pathogenesis. To determine whether the updated ChromHMM could capture the variation between individual states, I first performed an MCA to establish whether the identities of each cell type are maintained in the ChromHMM output (**Fig. 3.10**). I observed that the ChromHMM data clustered according to cellular identities, with the $CD4^+$ and $CD8^+$ T cell populations clustering closer together than the more distantly related $CD14^+$ monocytes and $CD19^+$ B cells. Separation between AS patients and healthy volunteers was however not observed at this stage of the analysis.

To investigate whether any differences existed in ChromHMM emission state identities between AS patients and healthy volunteers in each cell type, I performed MCA upon all detected ChromHMM bins (**Fig. 3.11**). The data was first filtered to include only the most polymorphic bins that showed at least two of the ChromHMM emission states across all cell types and individuals. Bins were also excluded from the analysis if greater than two-thirds of the individuals identified for the U14 quiescent state. I found that $CD4^+$ T cells, $CD8^+$ T cells, and $CD19^+$ B cells show separation of AS patients and healthy volunteers

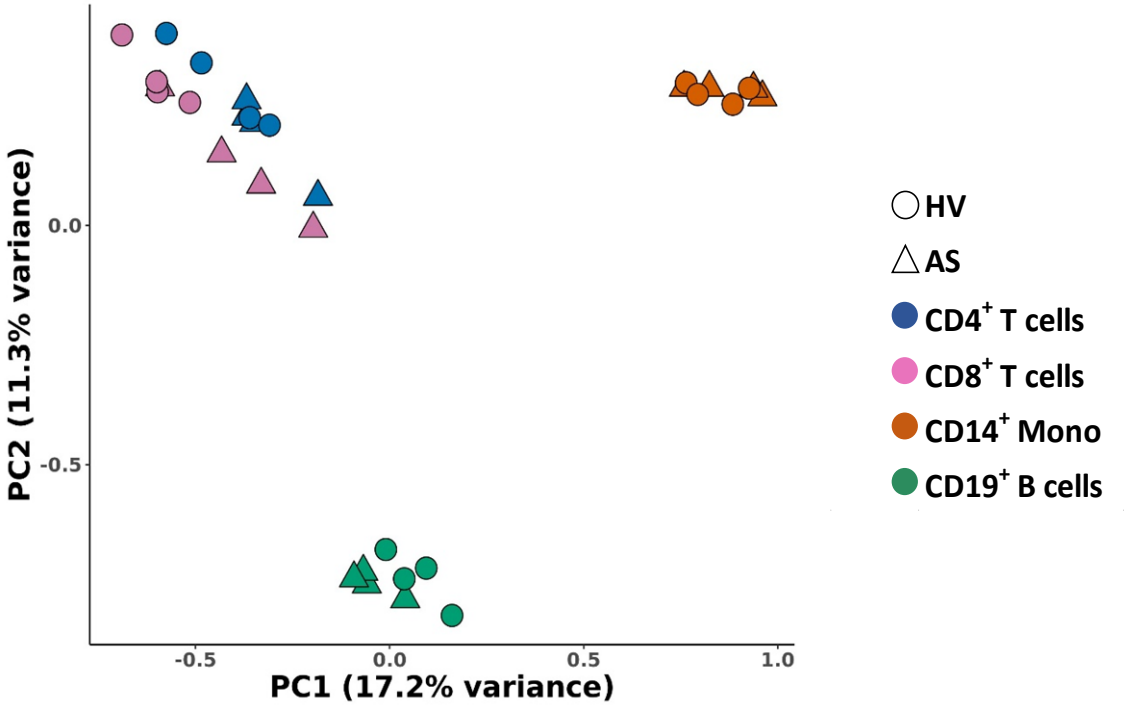

Figure 3.10. MCA analysis of ChromHMM emission state bins in the 16-state model shows cell type-specific separation. Figure above shows the separation a subsampled population of 10,000 emission state bins from $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells derived from the generated 16-state ChromHMM. Whilst no separation is seen by either AS patient (circles) or healthy volunteers (triangles), a strong separation of the immune cell types is clearly observed.

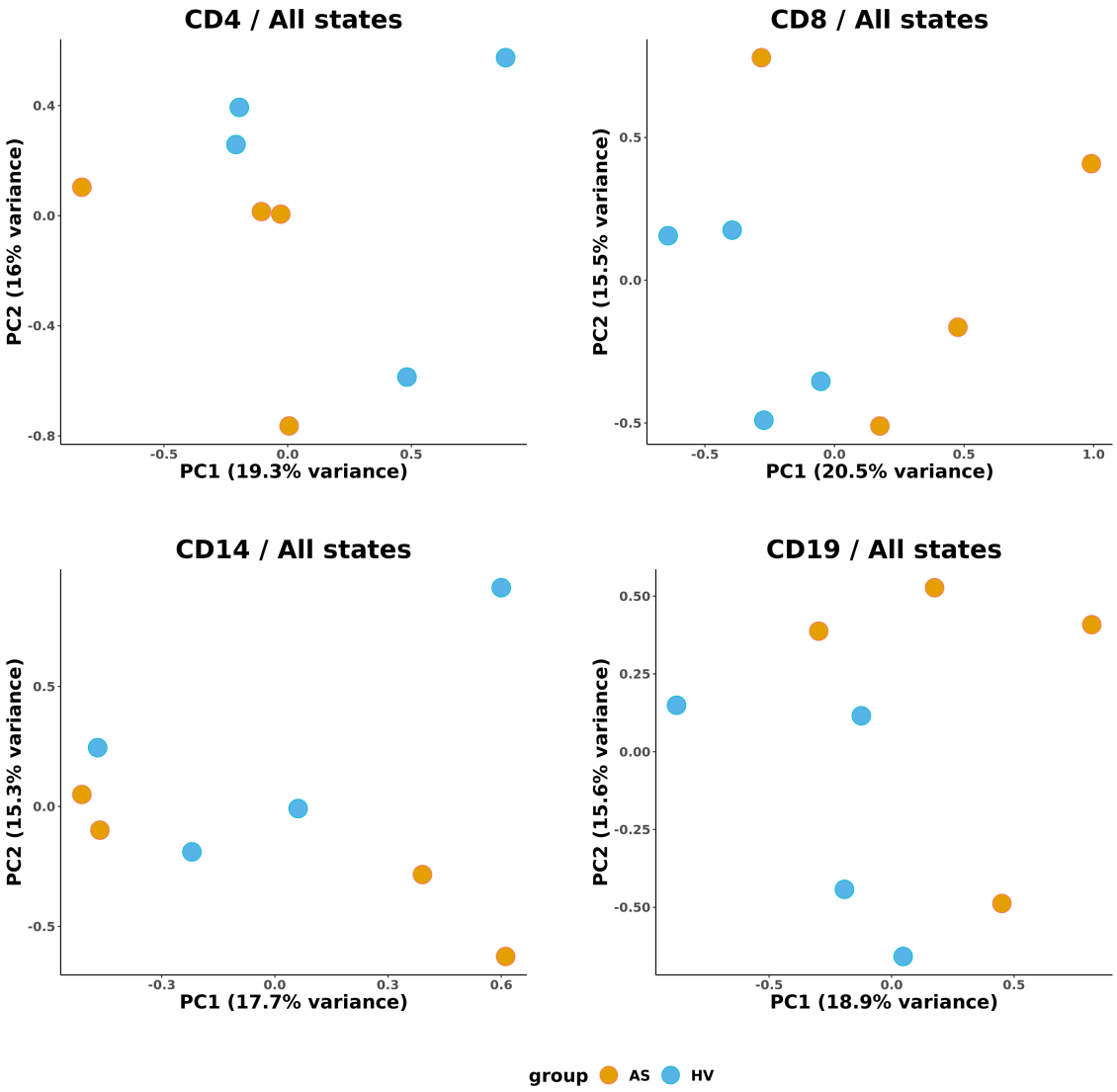

Figure 3.11. MCA analysis of all ChromHMM bins in the 16-state model for each individual cell type. Figure above shows individual dot-plots for the separation of AS patient and healthy volunteer samples in $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). T cells and B cells demonstrate separation on dimensions one and two, whilst monocytes do not show separation of AS patients and healthy volunteers.

on MCA. In contrast to Brown *et al.* (2023), $CD14^+$ monocytes however did not show this separation. This may suggest significant variation in chromatin conformation between AS patients and healthy volunteers in T cells and B cells.

To identify which ChromHMM emission states could be driving global changes in chromatin composition, I performed MCA of the 16 identified ChromHMM emission states separately in $CD4^+$ T cells (**Fig. 3.12**), $CD8^+$ T cells (**Fig. 3.13**), $CD14^+$ monocytes (**Fig. 3.14**), and $CD19^+$ B cells (**Fig. 3.15**). Here, ChromHMM bins were filtered out only if greater than one-third of an individual bin across all cell types and individuals identified for the U14 quiescent state. Separation of AS patients versus healthy volunteers was seen in a limited number of cases, such as U1 TSSActv and U4 TSSBvlnt in B cells. $CD4^+$ T cells, $CD8^+$ T cells, and $CD19^+$ B cells also showed separation in the U14 quiescent state and Polycomb-repressed U15 PolyG and U16 PolyNC states. Only the U11 EnhWk emission state was observed to separate in $CD14^+$ monocytes.

3.3.3. $CD14^+$ monocytes and $CD19^+$ B cells show quantitative differences in ChromHMM emission states in AS patients compared to healthy volunteers

I next sought to determine whether chromatin context could differ between AS patients and healthy volunteers by quantifying the proportion of each ChromHMM emission state in each cell type. The number of bins in each state for all cell types was highly variable. This was most apparent in the T-cell populations, where no state showed statistically significant differences between AS and healthy volunteers as measured by the Wilcoxon signed-rank test (**Fig. 3.16; Fig. 3.17).** The total numbers of U1 TSSActv (adjusted p-value = 0.03; percent change = -9.8%) and U3 TSSFlnkU (adjusted p-value = 0.03; percent change = -16.7%) active promoter emission state bins were significantly decreased in the $CD14^+$ monocytes of AS patients compared to healthy volunteers (**Fig.**

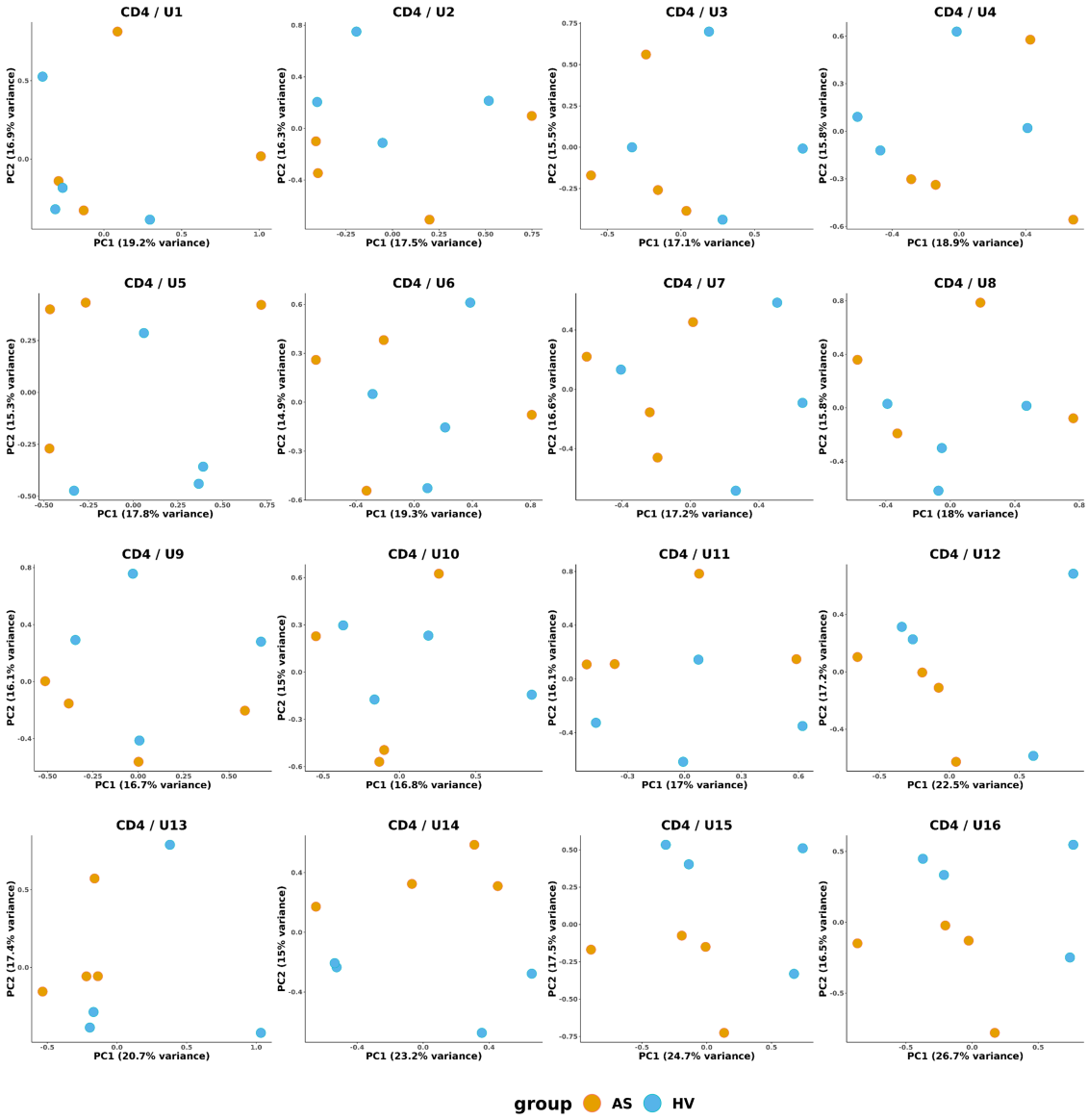

Figure 3.12. MCA analysis of each individual ChromHMM emission state within CD4$^+$ T cells for the 16-emission state model. Dot plots above show the genomic distribution of the 16 detected ChromHMM emission states upon incorporation of H3K27me3 into the model specifically for CD4$^+$ T cells. Each plot compares the four AS patients (orange dots) to the four healthy volunteers (blue dots). Separation of AS patients and healthy volunteer samples is indicative of considerable variation in the genomic distribution of each ChromHMM bin between the two conditions. Separation is seen in either dimension one or dimension two for the states of U14 Quies, U15 PolyG, and U16 PolyNC.

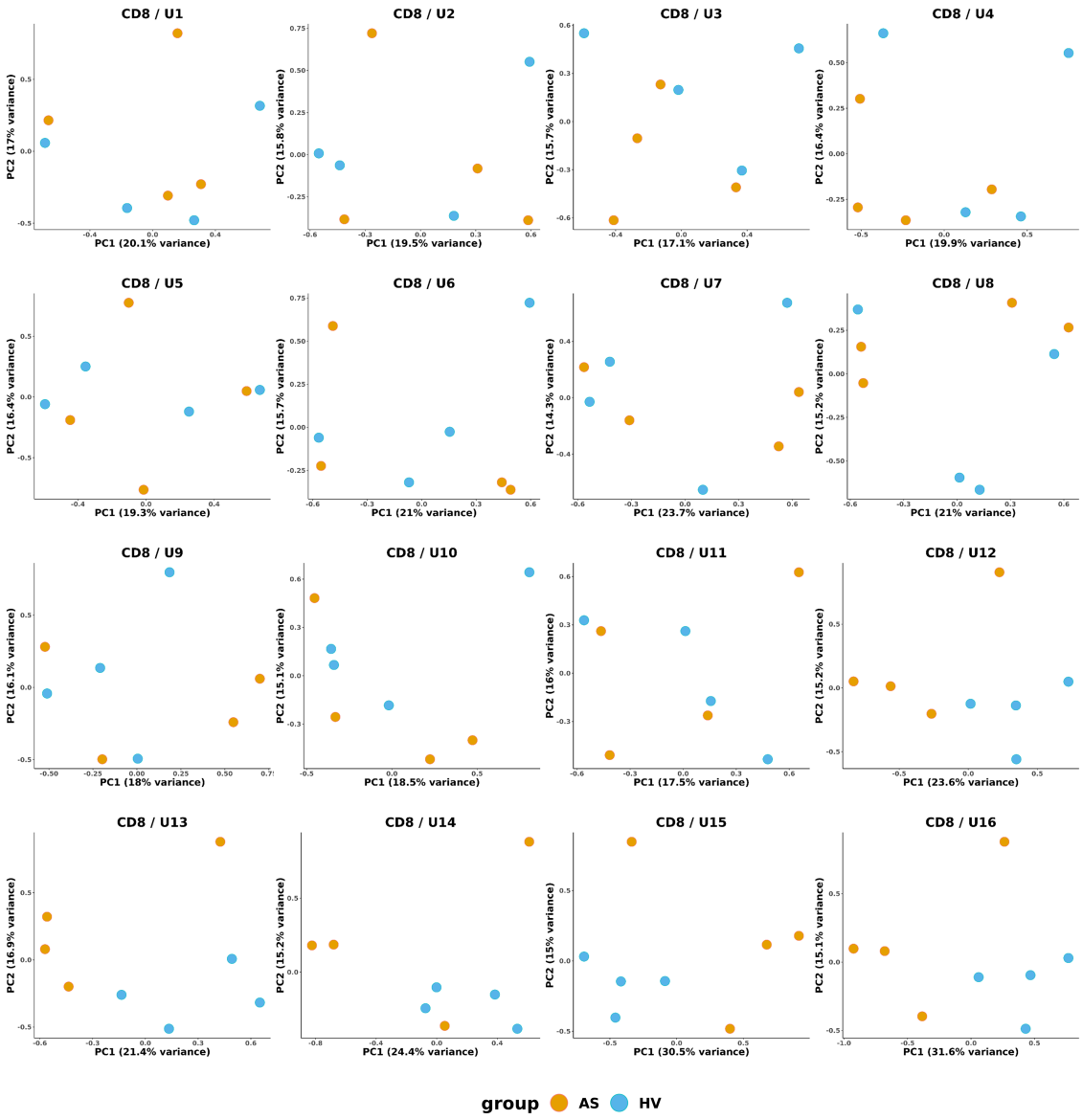

Figure 3.13. MCA analysis of each individual ChromHMM emission state within CD8$^+$ T cells for the 16-emission state model. Dot plots above show the genomic distribution of the 16 detected ChromHMM emission states upon incorporation of H3K27me3 into the model specifically for CD8$^+$ T cells. Each plot compares the four AS patients to the four healthy volunteers. Separation of AS patients and healthy volunteer samples is indicative of considerable variation in the genomic distribution of each ChromHMM bin between the two conditions. Separation is seen in either dimension one or dimension two for the states of U12 Int, U13 Wk, U15 PolyG, and U16 PolyNC.

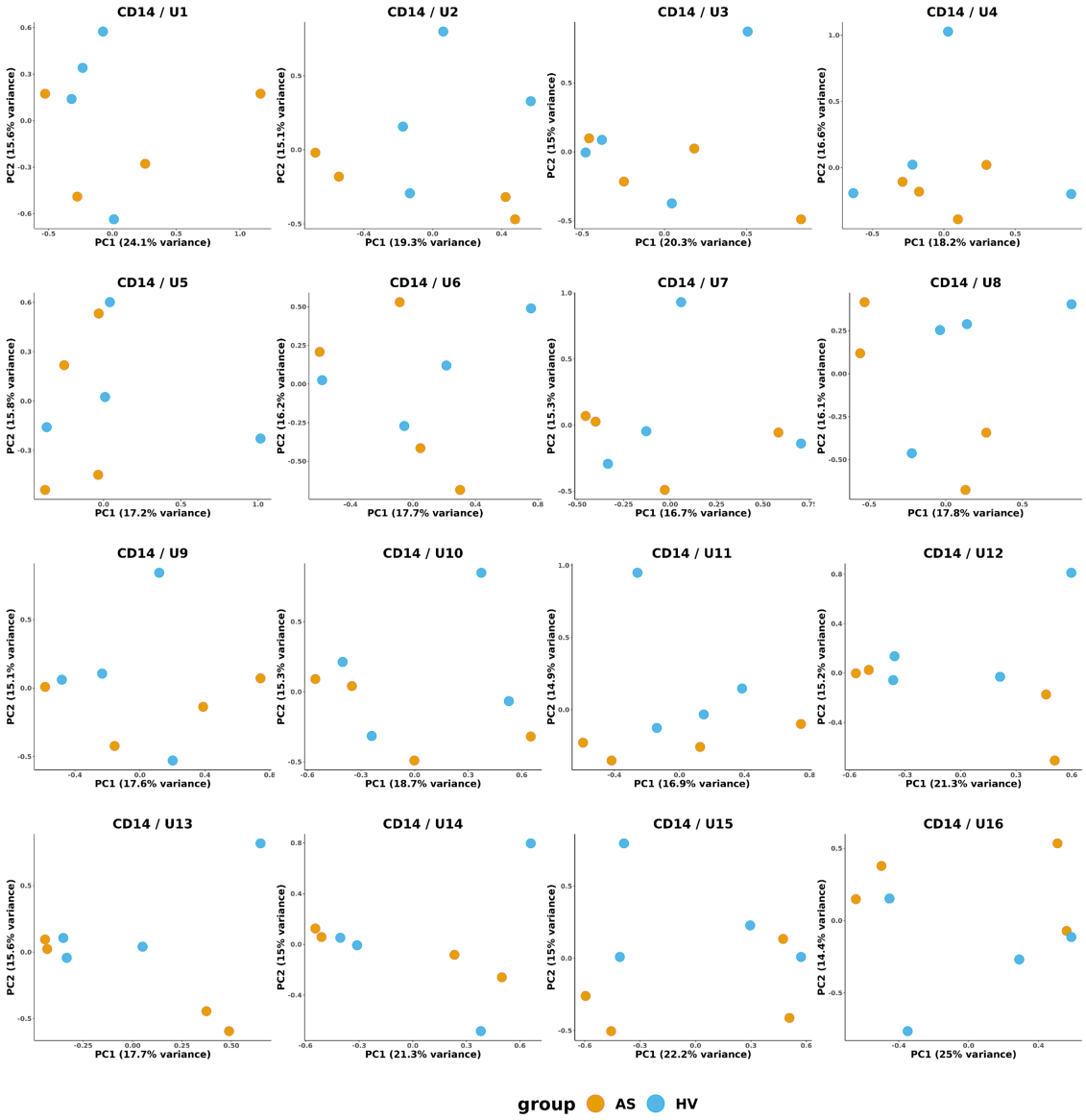

Figure 3.14. MCA analysis of each individual ChromHMM emission state within CD14$^+$ monocytes for the 16-emission state model. Dot plots above show the genomic distribution of the 16 detected ChromHMM emission states upon incorporation of H3K27me3 into the model specifically for CD14$^+$ monocytes. Each plot compares the four AS patients to the four healthy volunteers. Separation of AS patients and healthy volunteer samples is indicative of considerable variation in the genomic distribution of each ChromHMM bin between the two conditions. Separation is seen in either dimension one or dimension two for the singular state of U11 EnhWk.

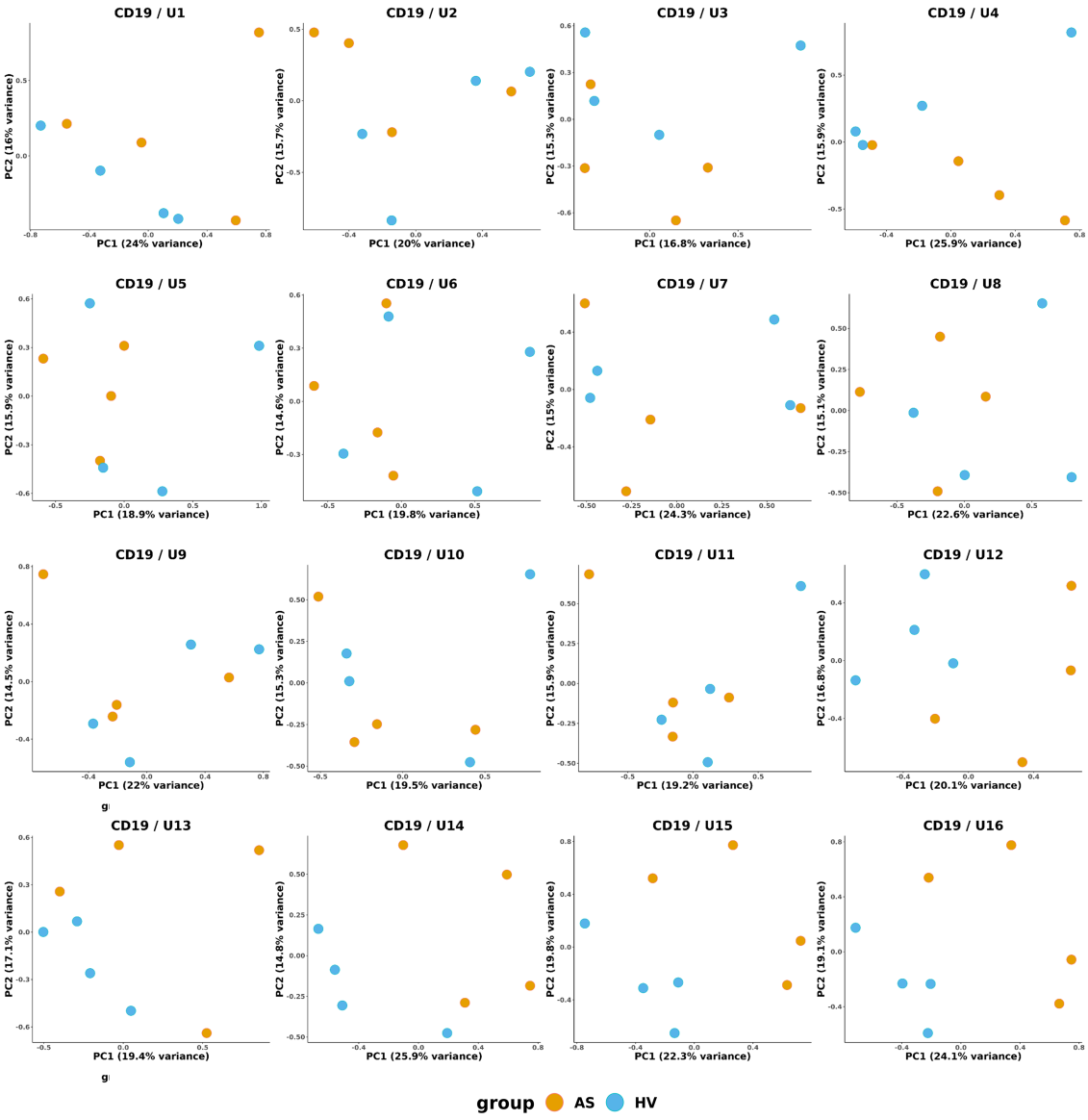

Figure 3.15. MCA analysis of each individual ChromHMM emission state within CD19$^+$ B cells for the 16-emission state model. Dot plots above show the genomic distribution of the 16 detected ChromHMM emission states upon incorporation of H3K27me3 into the model specifically for CD19$^+$ B cells. Each plot compares the four AS patients to the four healthy volunteers. Separation of AS patients and healthy volunteer samples is indicative of considerable variation in the genomic distribution of each ChromHMM bin between the two conditions. Separation is seen in either dimension one or dimension two for the singular state of U1 TssActv, U12 Int, U13 Wk, U14 Quies, U15 PolyNC, and U16 PolyG.

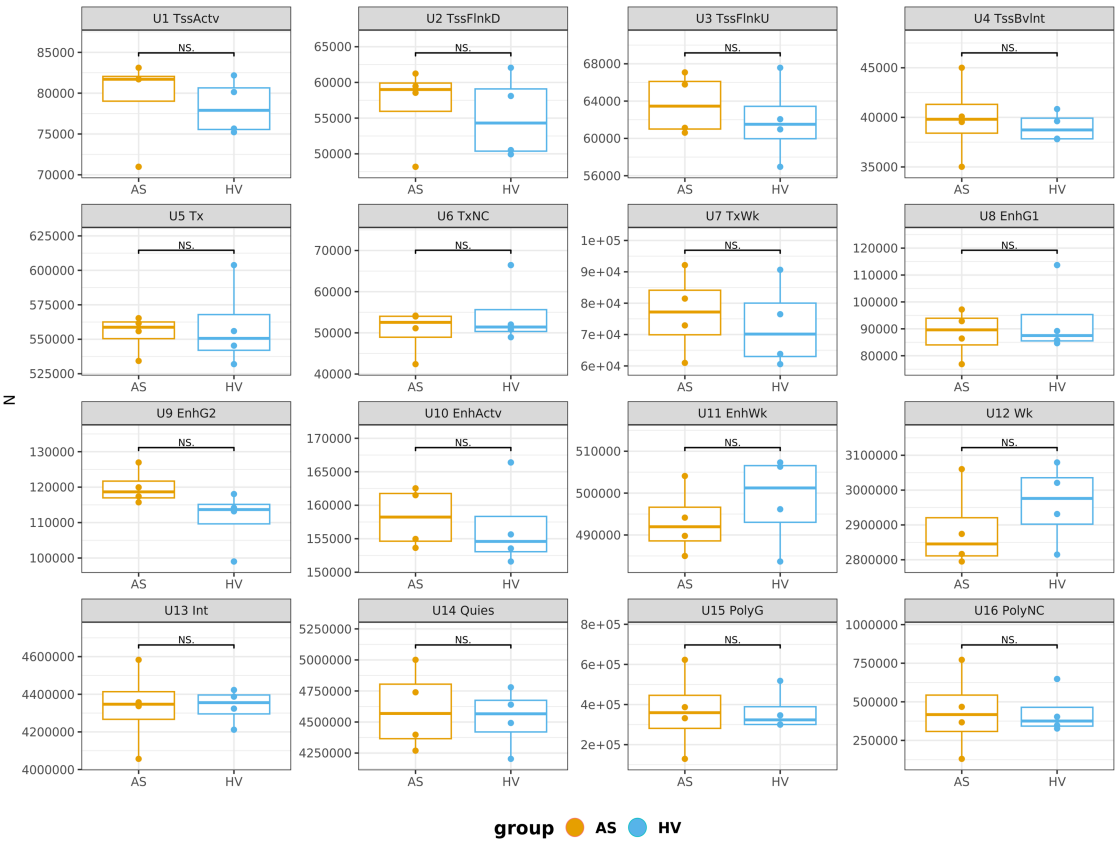

Figure 3.16. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for CD4$^+$ T cells. Shown above are box plots comparing the total counts for each of the 16 emission states between AS patients (n = 4) and healthy volunteers (n = 4) in CD4$^+$ T cells. The number of 200bp bins detected for each state (*N*) has been shown on the y-axis. Significance has been calculated using the Wilcoxson signed-rank test with Bonferroni correction (NS = not significant). No emission states within CD4$^+$ T cells showed significant skewing between AS patients and healthy volunteers.

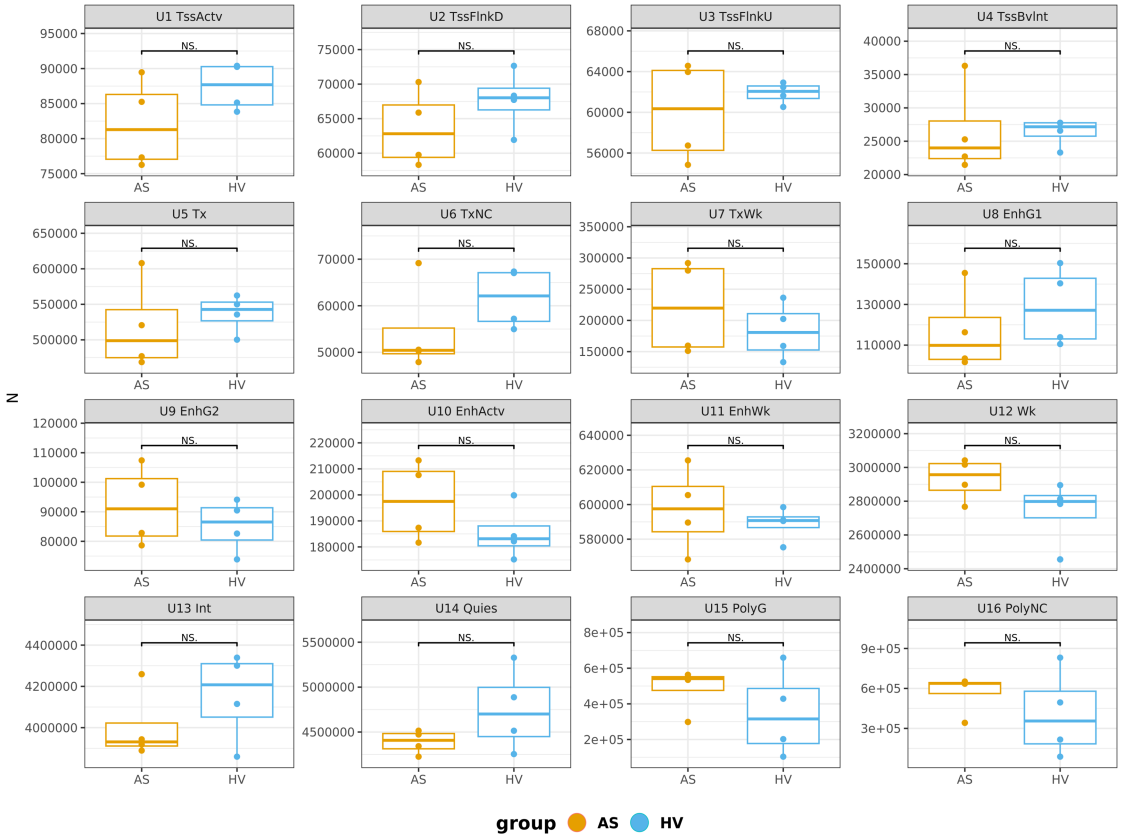

Figure 3.17. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for CD8$^+$ T cells. Shown above are box plots comparing the total counts for each of the 16 emission states between AS patients (n = 4) and healthy volunteers (n = 4) in CD8$^+$ T cells. The number of 200bp bins detected for each state (*N*) has been shown on the y-axis. Significance has been calculated using the Wilcoxson signed-rank test with Bonferroni correction (NS = not significant). No emission states within CD8$^+$ T cells showed significant skewing between AS patients and healthy volunteers.

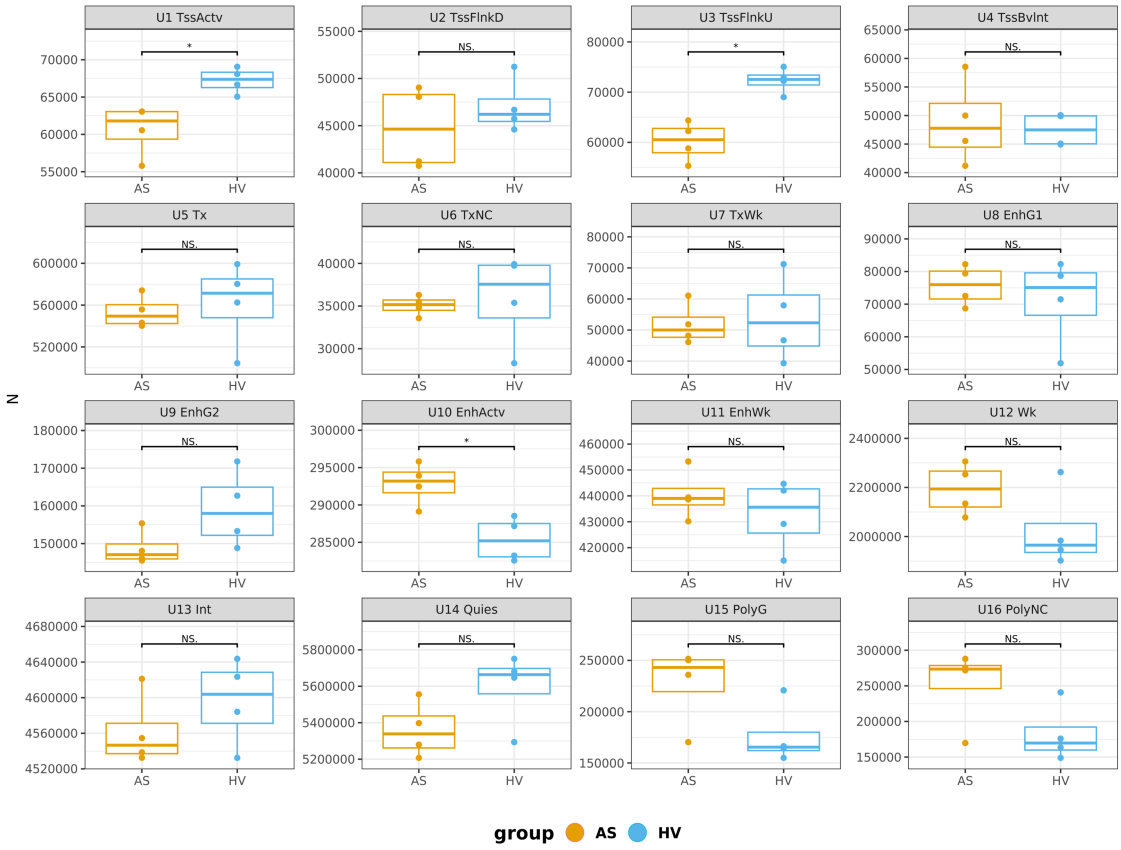

Figure 3.18. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for CD14$^+$ monocytes. Shown own above are box plots comparing the total counts for each of the 16 emission states between AS patients (n = 4) and healthy volunteers (n = 4) in CD14$^+$ monocytes. The number of 200bp bins detected for each state (*N*) has been shown on the y-axis. Significance has been calculated using the Wilcoxson signed-rank test with Bonferroni correction (NS = not significant; * ≤ adjusted p-value of 0.05). Significant differences in total detected bins could be seen for the U1 TSSActv, U3 TSSFlnkU, and U10 EnhActv emission states.

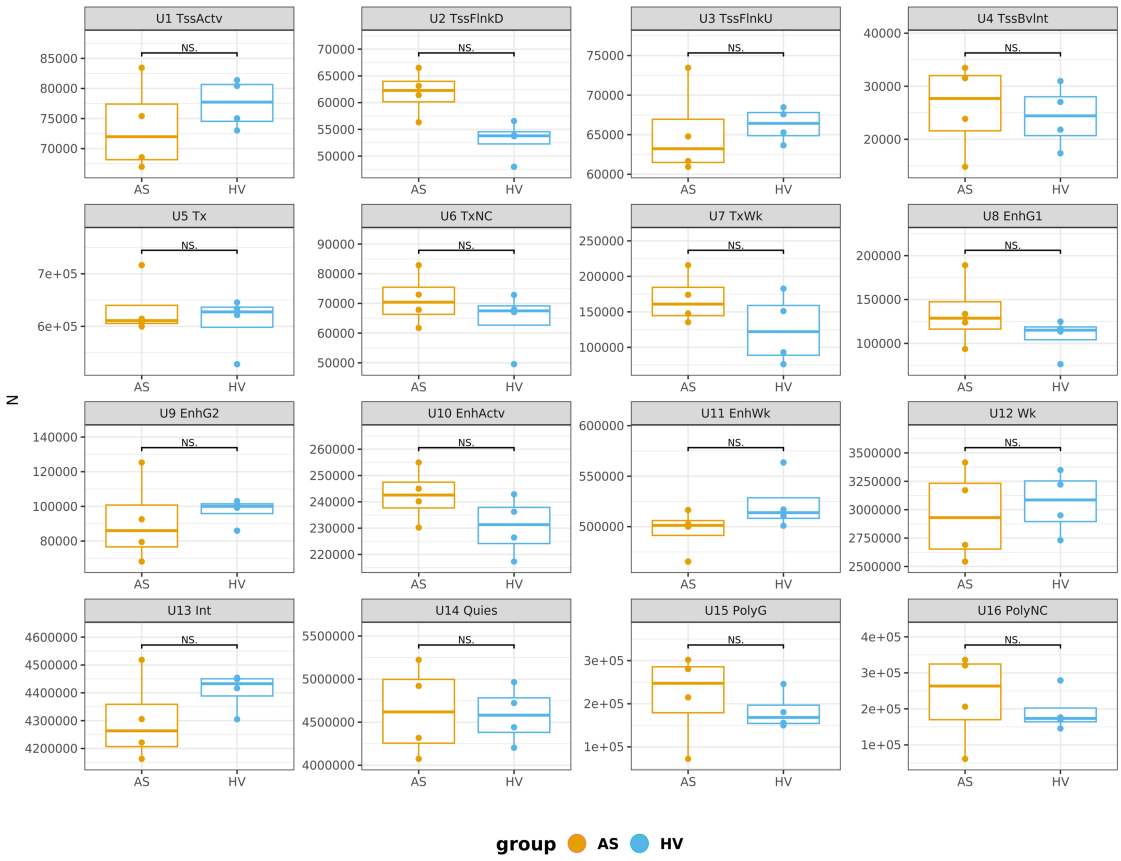

Figure 3.19. Quantification of genomic segmentations per ChromHMM emission state within the 16-state model for CD19$^+$ B cells. Shown above are box plots comparing the total counts for each of the 16 emission states between AS patients (n = 4) and healthy volunteers (n = 4) in CD19$^+$ B cells. The number of 200bp bins detected for each state (*N*) has been shown on the y-axis. Significance has been calculated using the Wilcoxson signed-rank test with Bonferroni correction (NS = not significant). No emission states within CD19$^+$ B cells showed significant skewing between AS patients and healthy volunteers.

3.18). The U10 EnhActv emission state also showed significant increases in the total number of bins of AS patient $CD14^+$ monocytes (adjusted p-value = 0.03; percent change = 2.5%). Like in the T cells, no statistically significant differences were observed in $CD19^+$ B cells (**Fig. 3.19**). These results suggest that whilst widespread reconfiguration and differential epigenetic usage of chromatin is not seen within all AS-implicated cell types, it could form an important part of the $CD14^+$ monocyte-driven aspects of disease pathogenesis.

3.3.4. Permutation analysis reveals significant differences in the identify of quiescent and Polycomb-repressed emission states within AS compared to healthy volunteers

I next sought to perform permutation analysis to validate the significance of the difference in the number of emission state alterations in AS patients and in the ChromHMM outputs of all four PBMC cell types. Due to the binary nature of the emission states and limited sample size for both conditions (n = 4 for AS patients; n = 4 for healthy volunteers) however, the smallest p-value achievable through a Fisher's exact test was 0.05, which after multiple testing correction resulted in an adjusted p-value greater than 0.05. This corresponded to a state present in either four AS patients and no healthy volunteers (AS4vsHV0), or no AS patients and four healthy volunteers (AS0vsHV4). Towards the identification of ChromHMM emission states with a higher distribution of differential regions in AS than expected by chance alone, I permuted all possible combinations of the four AS patients and healthy volunteers. This resulted in a total of 70 combinations between all healthy volunteers and AS patients. The experimentally observed frequency of the Fisher's p-value significant regions was compared to the background permuted distribution on a per-state basis. Using this methodology, the lowest Fisher p-value possible for a state is 0.015; corresponding to 1/70 possible permutations and a complete switch from one chromatin state to another.

Whilst there were no emission states in $CD4^+$ T cells (**Fig. 3.20**) showing significant differences in the experimentally observed frequency of Fisher's significant bins versus the permuted background, emission states U12, U13, U15, and U16 were however shown to be significantly permuted in $CD8^+$ T cells (**Fig. 3.21**). There were no significantly permuted states in $CD14^+$ monocytes (**Fig. 3.22**). Emission states U12, U14, U15, and U16 additionally saw significant permutation in $CD19^+$ B cells (**Fig. 3.23**). These states correspond to either weak, intronic, quiescent, or Polycomb-repressed regions of DNA, and could suggest shifts to more active emission states in either the disease or healthy cohort. Interestingly, no *cis*-regulatory-like or transcription-like emission states showed significant permutation, suggesting that the majority of promoter-like, enhancer-like, and transcription-like elements and their ChromHMM-defined level of activity may be relatively stable between both AS patients and healthy volunteers.

3.3.5. Analysis of transition profiles of significantly permuted ChromHMM emission states reveals bias in emission state transformations

Having shown that there are disease-specific enrichments of particular chromatin states in each cell type, I next wanted to find out how these states were changing between the AS patients and healthy volunteers. To track state transitions, I analysed the state identities of each bin where the state was present in all four AS cases and absent from all healthy volunteers (4ASvs0HV) and where the state was present in all four healthy volunteers and absent from all AS patients (0ASvs4HV). I observed that all cell types shared similar patterns in state transitions (**Fig. 3.24; Fig. 3.25**). In T cells and B cells, U15 PolyG and U16 PolyNC regions in AS patients primarily switch with U13 Int and U12 Quies respectively in healthy volunteers. This was not seen in $CD8^+$ T cells; potentially explaining why the state did not show permutation significance. Some U15 PolyG and U16 PolyNC regions transition to the U4 TSSBvlnt state. This pattern is

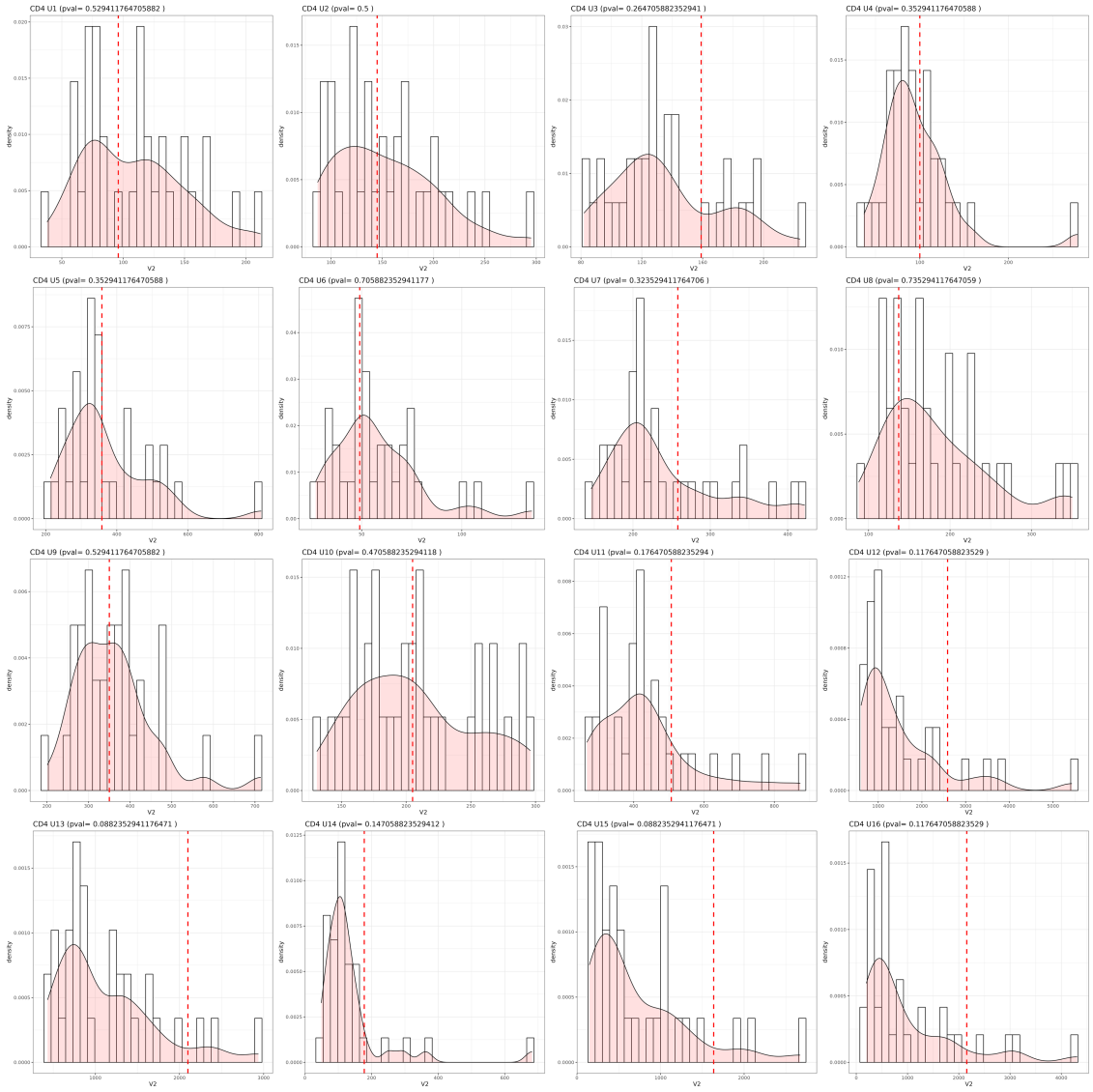

Figure 3.20. Permutation analysis of emission states in the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in CD4$^+$ T cells. Plots above show the observed frequency of significantly permuted regions for each ChromHMM emission state (Fisher's p-value = 0.05) seen in CD4$^+$ T cells for four AS patients and 0 healthy volunteers, compared to a background distribution created from a total of 70 possible permutations of the disease and healthy states. Here, the background distribution has been shown as the pink area under the curve. The x-axis of each plot represents the number of permuted states (V2) whilst the y-axis represents the density of each possible permutation (density). The observed frequency of significantly permuted regions in each emission state is shown as a red dashed-line on each plot. The permutation-associated p-value for each state has been indicated in the title of each plot. A p-value of 0 is indicative instead of a p-value of 0.015, the lowest possible achievable p-value in the permutation analysis. No states within CD4$^+$ T cells showed significant permutation.

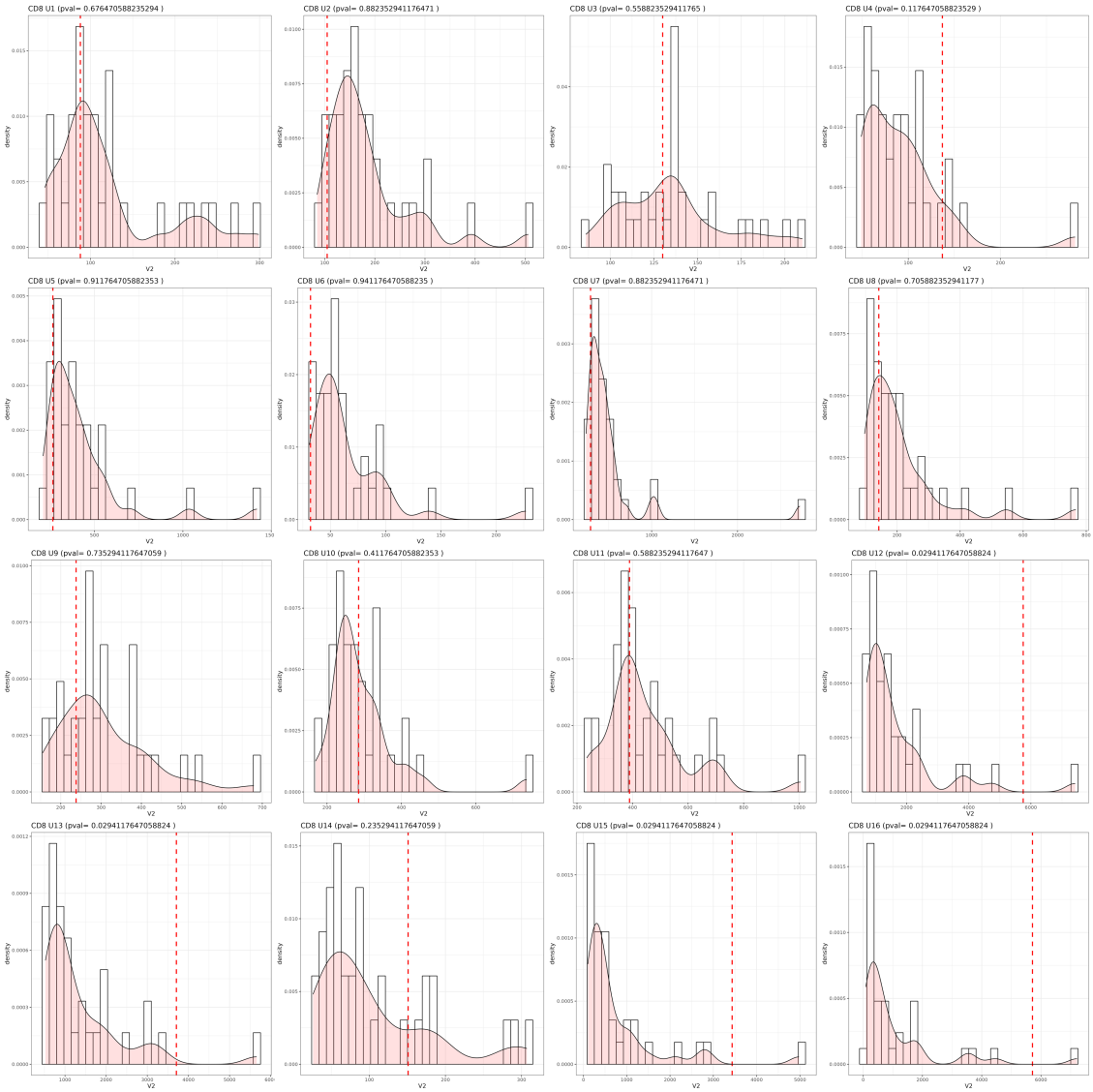

Figure 3.21. Permutation analysis of emission states in the 16-state ChromHMM shows significant differences for quiescent, intronic, weak, and Polycomb-repressed chromatin states in CD8$^+$ T cells . Plots above show the observed frequency of significantly permuted regions for each ChromHMM emission state (Fisher's p-value = 0.05) seen in CD8$^+$ T cells for four AS patients and 0 healthy volunteers, compared to a background distribution created from a total of 70 possible permutations of the disease and healthy states. Here, the background distribution has been shown as the pink area under the curve. The x-axis of each plot represents the number of permuted states (V2) whilst the y-axis represents the density of each possible permutation (density). The observed frequency of significantly permuted regions in each emission state is shown as a red dashed-line on each plot. The p-value for each state has been indicated in the title of each plot. A p-value of 0 is indicative instead of a p-value of 0.015, the lowest possible achievable p-value in the permutation analysis. The states U12, U13, U15, and U16 were significantly permuted in CD8$^+$ T cells.

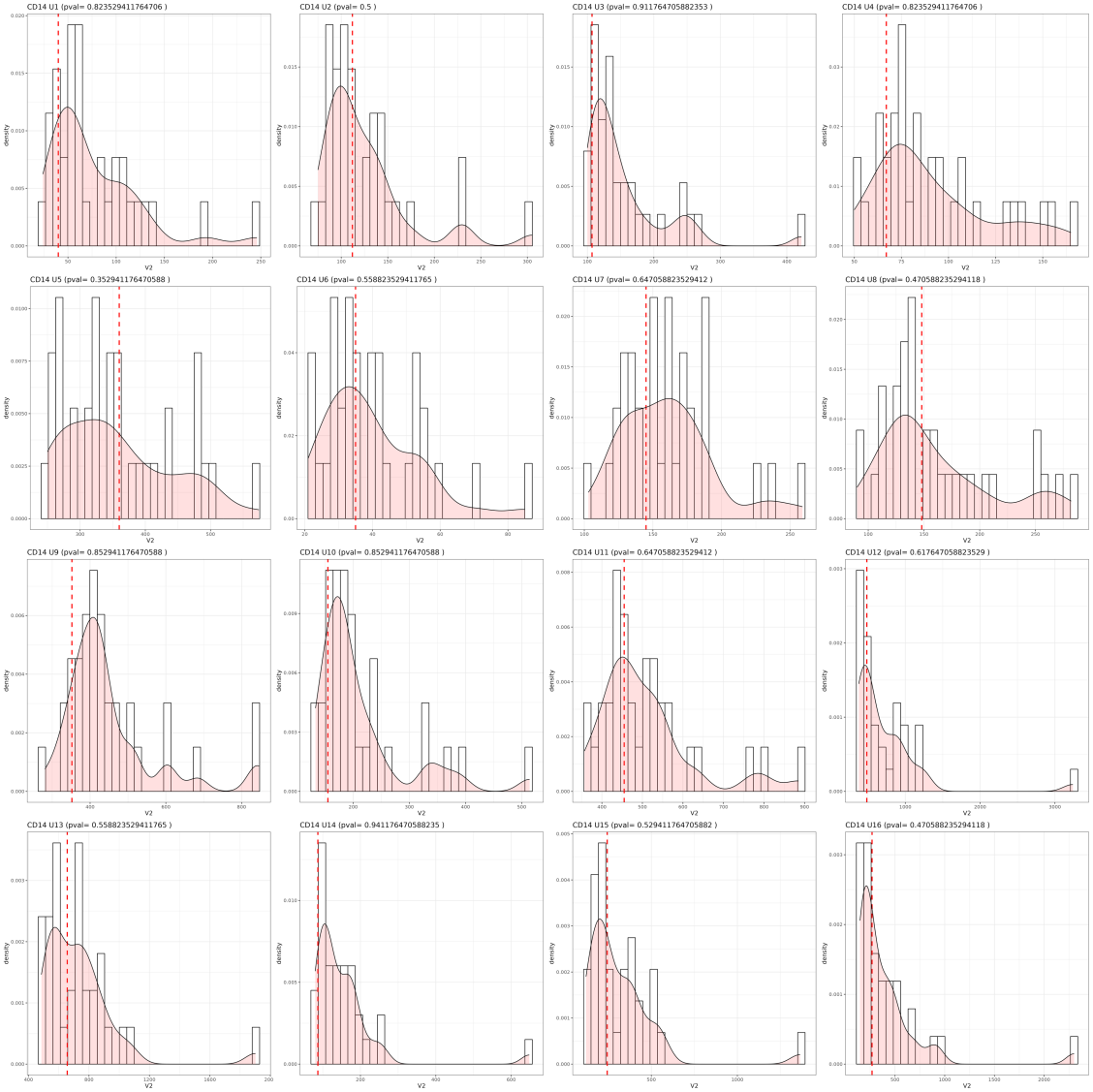

Figure 3.22. Permutation analysis of emission states within the 16-state ChromHMM shows no significant differences in the frequency of significantly permuted regions in CD14$^+$ monocytes . Plots above show the observed frequency of significantly permuted regions for each ChromHMM emission state (Fisher's p-value = 0.05) seen in CD14$^+$ monocytes for four AS patients and no healthy volunteers, compared to a background distribution created from a total of 70 possible permutations of the disease and healthy states. Here, the background distribution has been shown as the pink area under the curve. The x-axis of each plot represents the number of permuted states (V2) whilst the y-axis represents the density of each possible permutation (density). The observed frequency of significantly permuted regions in each emission state is shown as a red dashed-line on each plot. The permutation-associated p-value for each state has been indicated in the title of each plot. A p-value of 0 is indicative instead of a p-value of 0.015, the lowest possible achievable p-value in the permutation analysis. No states within CD14$^+$ monocytes demonstrated permutation significance.

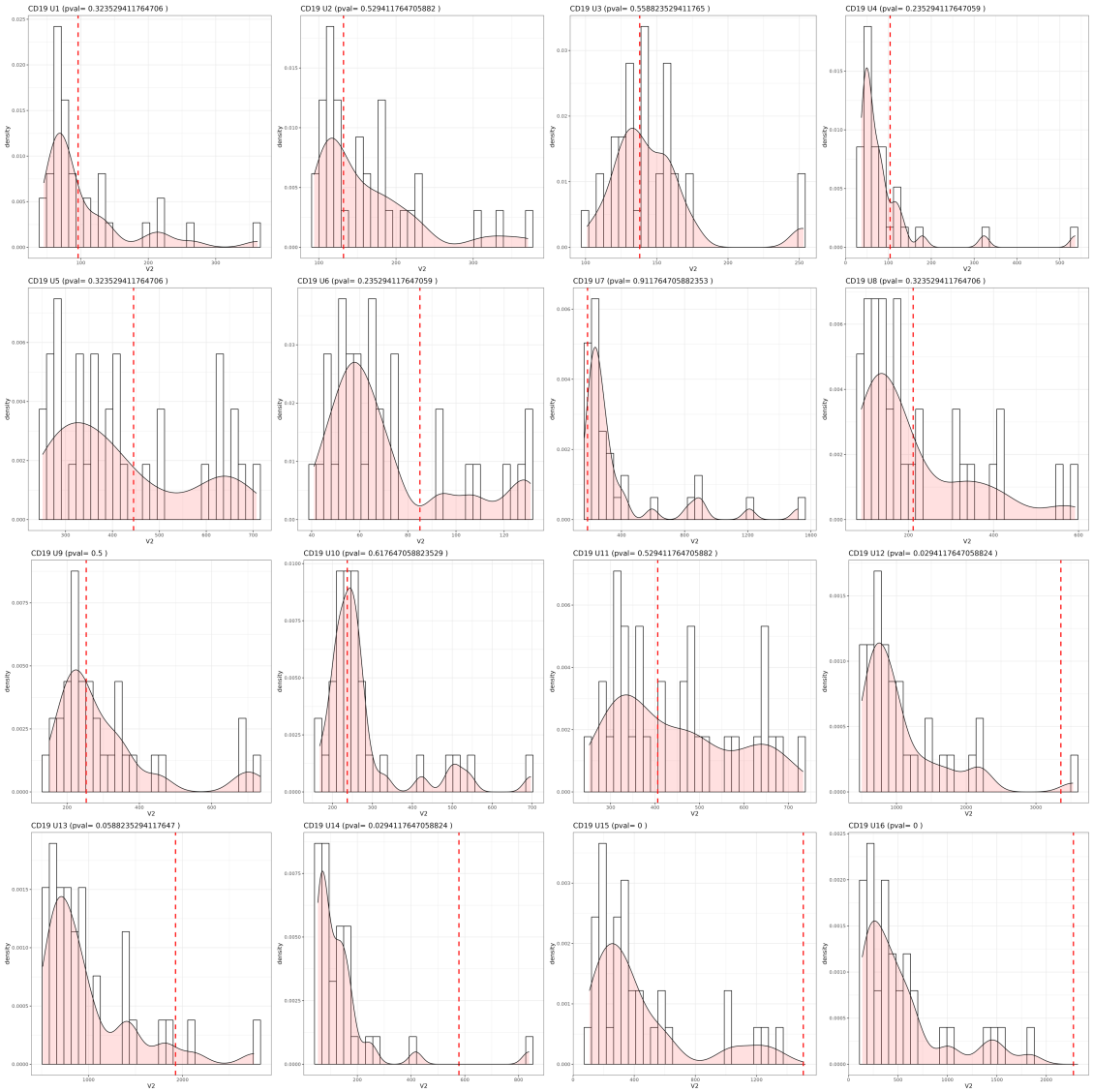

Figure 3.23. Permutation analysis of emission states within the 16-state ChromHMM shows significant differences for quiescent, intronic, weak, and Polycomb-repressed chromatin states in CD19$^+$ B cells . Plots above show the observed frequency of significantly permuted regions for each ChromHMM emission state (Fisher's p-value = 0.05) seen in CD19$^+$ B cells for four AS patients and no healthy volunteers, compared to a background distribution created from a total of 70 possible permutations of the disease and healthy states. Here, the background distribution has been shown as the pink area under the curve. The x-axis of each plot represents the number of permuted states (V2) whilst the y-axis represents the density of each possible permutation (density). The observed frequency of significantly permuted regions in each emission state is shown as a red dashed-line on each plot. The permutation-associated p-value for each state has been indicated in the title of each plot. A p-value of 0 is indicative instead of a p-value of 0.015, the lowest possible achievable p-value in the permutation analysis. The states U12, U14, U15, and U16 showed permutation significance in CD19$^+$ B cells.

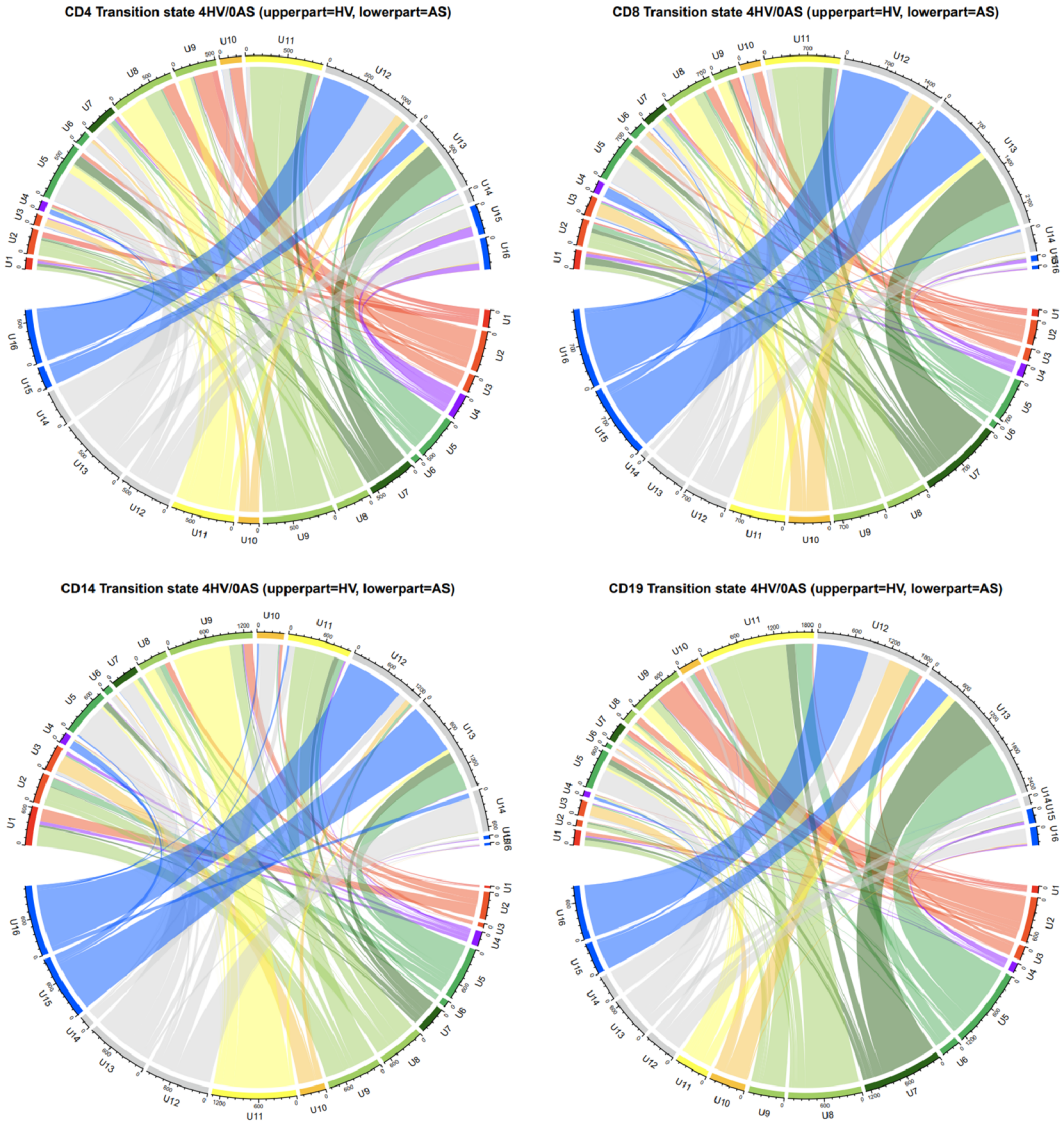

Figure 3.24. Transition analysis of significant emission state permutations shows cell-type specific, yet highly similar, transition patterns to other ChromHMM emission states within healthy volunteers. Circos plots above show the directional transition of equivalent emission states in none of the AS patients and four of the healthy volunteers for $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). Each emission state has been indicated on the circumference by number. Circle circumference is representative of the number of differential ChromHMM fragments, with one tick indicative of 200 ChromHMM bins. Emission states with similar functional roles have been group based on colour consistent with figure 3.8. Red is equal to promoter-like states, purple to bivalent promoter-like states, orange to active enhancer-like states, yellow to weak enhancer-like states, light green to genic enhancers, dark green to strongly transcribed regions, forest green to weakly transcribed regions, blue to Polycomb-repressed states, and grey to quiescent or weak states. The top half of the circle shows the emission state in the healthy volunteers, whilst the bottom half shows the equivalent emission state in AS patients. All ChromHMM bins included in the analysis have a Fisher's p-value of 0.05.

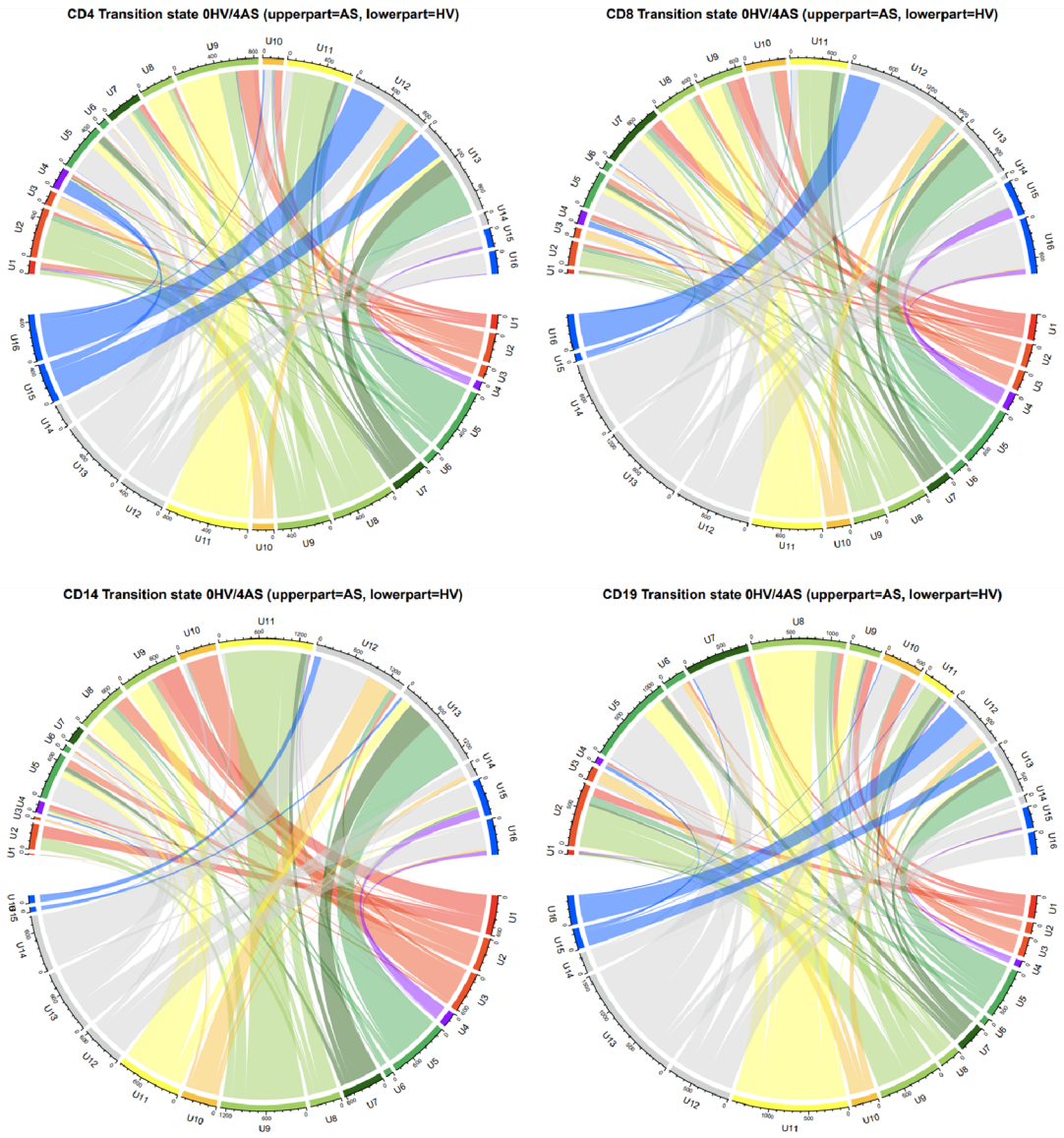

Figure 3.25. Transition analysis of significant emission state permutations shows cell-type specific, yet highly similar, transition patterns to other ChromHMM emission states within AS patients. Circos plots above show the directional transition of equivalent emission states in the four AS patients and no healthy volunteers for $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). Each emission state has been indicated on the circumference by number. Circle circumference is representative of the number of differential ChromHMM fragments, with one tick indicative of 200 ChromHMM bins. Emission states with similar functional roles have been group based on colour consistent with figure 3.8. Red is equal to promoter-like states, purple to bivalent promoter-like states, orange to active enhancer-like states, yellow to weak enhancer-like states, light green to genic enhancers, dark green to strongly transcribed regions, forest green to weakly transcribed regions, blue to Polycomb-repressed states, and grey to quiescent or weak states. The top half of the circle shows the emission state in the AS patients, whilst the bottom half shows the equivalent emission state in healthy volunteers. All ChromHMM bins included in the analysis have a Fisher's p-value of 0.05.

maintained both in 4ASvHV and 0ASvs4HV. U4 TSSBvlnt regions were observed to transition to the active promoter states U1 TSSBvlnt, U2 TSSFlnkD,and U3 TSSFlnkU, and the enhancer states U10 EnhActv and U11 EnhWk, suggesting a small proportion of U4 TSSBvlnt regions become active gene promoters or enhancers. These transitions likely indicate the silencing and activation of unique gene sets in AS and healthy volunteers through the loss or gain of the H3K27me3 post-translational histone-modification. The identity of these sets could potentially contribute towards AS pathogenesis. This is further evidenced by U12 Wk and U13 Int states transitioning to the transcription-associated U5 Tx, U6 Tx NC, and U7 Tx Wk states, suggesting activation of gene transcription from a previously quiescent state.

3.3.6. Enrichment of AS-associated GWAS variants in cis-regulatory and Polycomb-repressed regions

Having shown that there is variation between the chromatin state profiles of healthy volunteers and AS patients, I wanted to ascertain if the AS-associated SNPs from Cortes *et al.* (2013) – which were validated in the more powerful cross-disease Ellinghaus *et al.* (2016) study – are enriched in regulatory regions to delineate chromatin states responsible for the genetic association. Enrichment analysis was carried out using GARFIELD. As the MHC link to AS is well established through HLA-B27, GARFIELD was performed with and without the MHC region on chromosome six to highlight the contribution of each emission state to both MHC and less understood AS-associated risk loci.

The AS-associated genetic variants show similar enrichment patterns in all cell types when MHC SNPs were included (**Fig 3.26**). The most significant levels of enrichment were seen in the state U6 TxNC across all cell types. Significant enrichment was also seen in the promoter states U1 TSSActv, U2 TSSFlnkD, and U3 TSSFlnkU, and the

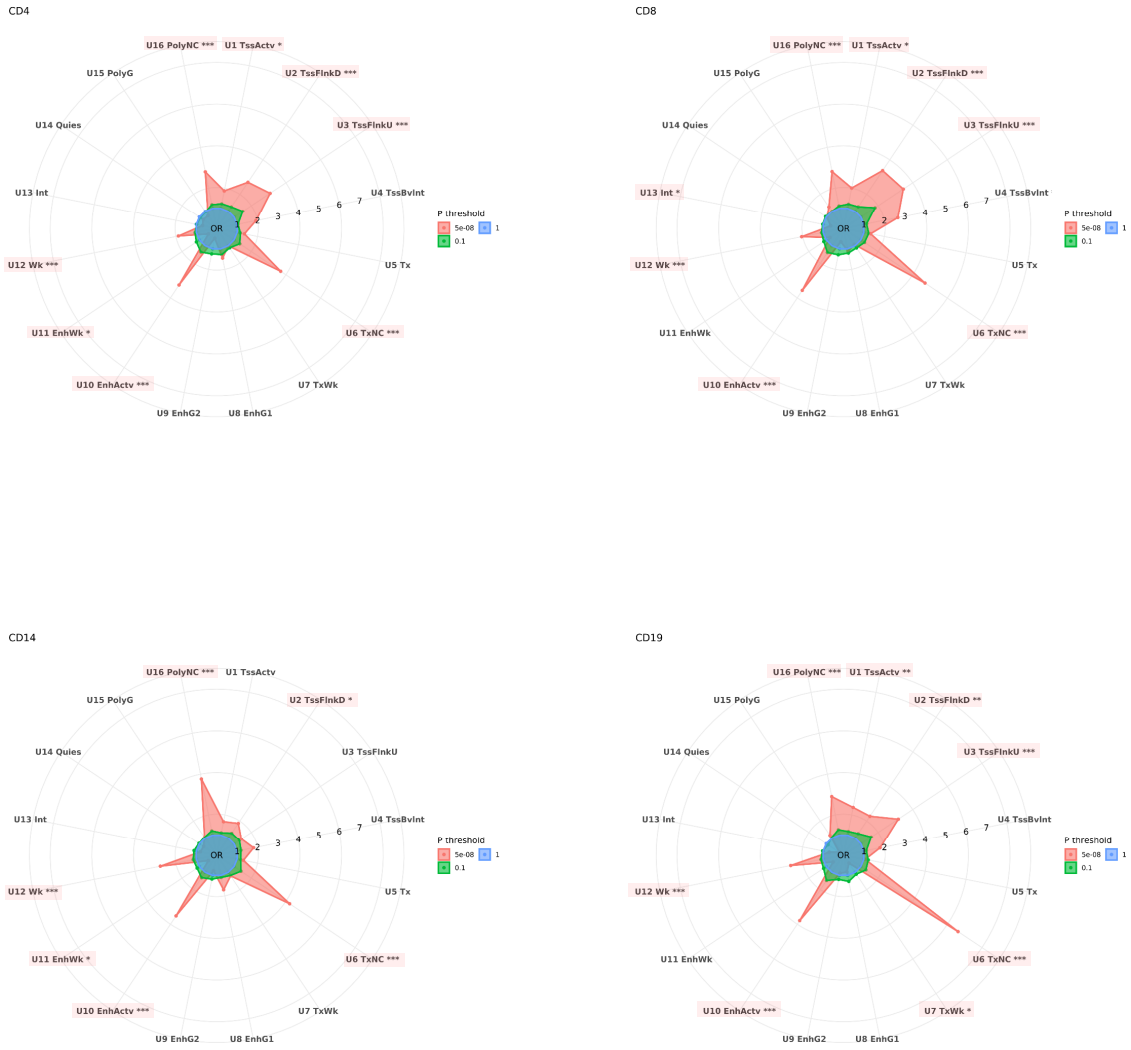

Figure 3.26. GARFIELD enrichment plots of Cortes et al. (2013) AS-associated immunochip genetic variants within each of the 16 emission states including MHC-associated SNPs on chromosome 6. Radial plots above showthe enrichment patterns for each emission state for AS-associated SNPs detected through immunochip by Cortes *et al.* (2013) for $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). All SNPs were considered for enrichment in this analysis, including the MHC-associated region between the genomic coordinates of 25,000,000 and 35,000,000 on chromosome 6. Three significance thresholds of genetic variant associations were tested for enrichment. Blue shows the odds ratio (OR) enrichment pattern for variants with a p-value threshold of 1, green the OR enrichment pattern for variants with a p-value threshold of 0.1, and pink shows the OR enrichment pattern for variants with a p-value threshold of $5x10^{-8}$. The latter corresponds to the minimum GWAS variant score required for genome-wide significance. ChromHMM emission states that are significantly enriched for by AS-associated GWAS SNPs have been highlighted in red shaded boxes. The p-value of enrichment in each ChromHMM emission state for AS-associated GWAS variants is shown starred at three levels of significance (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

enhancer state U11 EnhWk. This supports the hypothesis that AS genetic variants modulate the regulatory volunteer of linked genes. Enrichment was also seen in the U12 Wk and U16 PolyNC. In contrast, upon removal of the MHC-associated SNPs from the analysis, the genetic variant enrichment patterns were far more-cell type specific (**Fig. 3.27**). None of the cell types demonstrated variant enrichment for the U6 TxNC or the U16 PolyNC chromatin states.

Interestingly, the emission states U15 PolyG and U16 PolyNC were found to be negatively associated with the Cortes *et al.* (2013) SNPs. Indeed, both states had an average enrichment odds ratio of 0.12 and 0.42, respectively. This heavily suggested that causative SNPs in non-MHC risk loci are underrepresented in regions of Polycomb-repression higher than chance alone, and also suggests that the previously seen enrichment of U16 PolyNC is driven by MHC-associated SNPs. Non-MHC AS SNPs were enriched in at least one of the four promoter states in three of the four investigated cell types. Non-MHC SNPs were also enriched in U10 EnhActv across all cell types aside for $CD19^{+}$ B cells, suggested that variants which directly influence *cis*-regulatory function remain important within non-MHC loci.

AS-associated SNPs also showed enrichment for the chromatin states U12 Wk and U13 Int. Interestingly, both of these states comprise the majority of transitions away or to U15 PolyG or U16 PolyNC in both four healthy volunteers and four AS patients. In B cells, from the previous eight enriched chromatin states, the only states significantly enriched for AS-associated SNPs upon removal of MHC-associated genetic variants were TSSFlnkD and U12 Wk. This suggests that the majority of functionally enriched SNPs in B cells are related to MHC. Overall, these data suggest that the majority of non-coding SNPs involved in AS are present in quiescent, promoters, enhancers, or Polycomb-repressed regions of the genome.

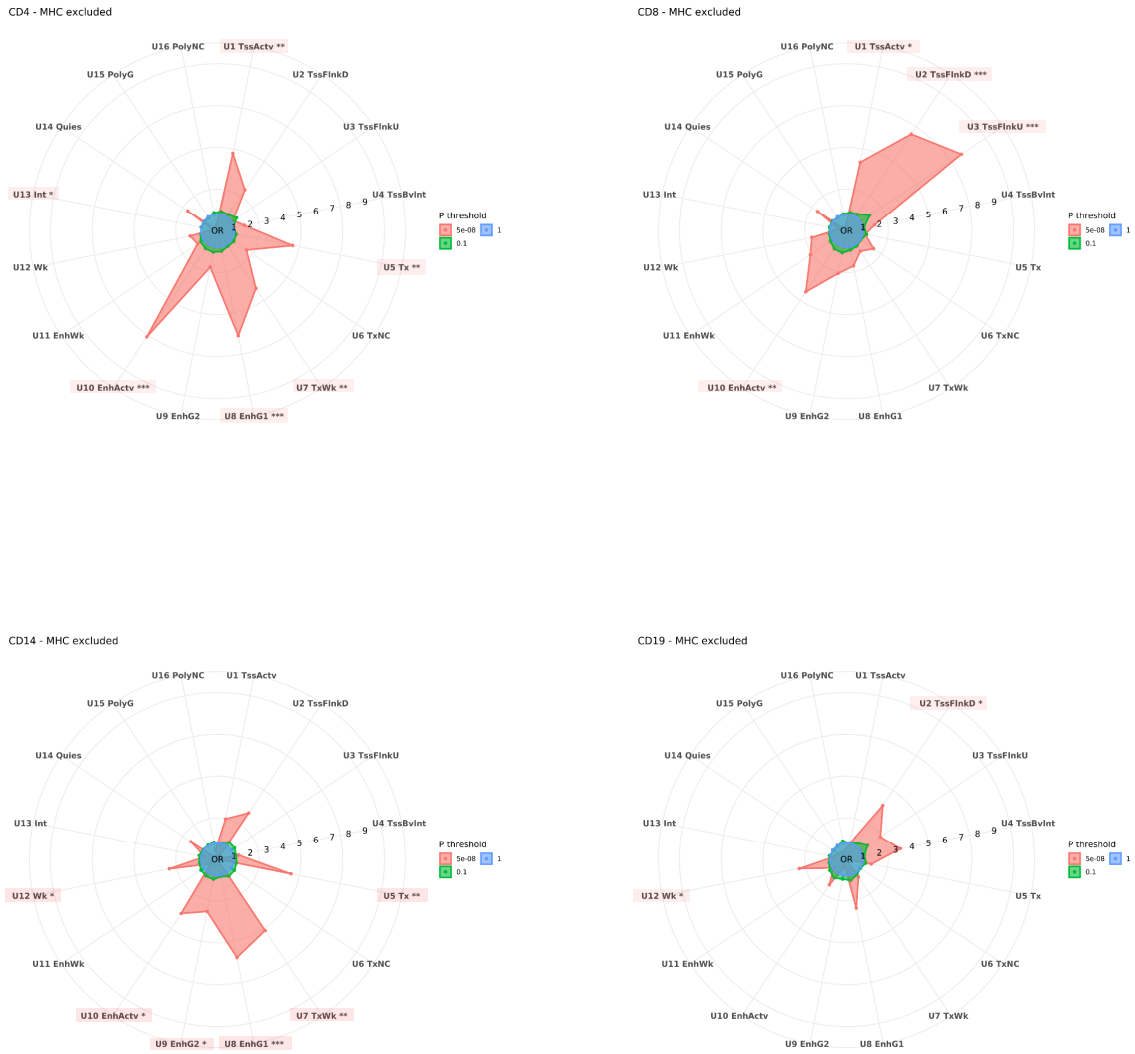

Figure 3.27. GARFIELD enrichment plots of Cortes et al. (2013) AS-associated immunochip genetic variants within each of the 16 emission states excluding MHC-associated SNPs on chromosome 6. Radial plots above show the enrichment patterns for each emission state for AS-associated SNPs detected through immunochip by Cortes *et al.* (2013) for $CD4^+$ T cells (top left), $CD8^+$ T cells (top right), $CD14^+$ monocytes (bottom left), and $CD19^+$ B cells (bottom right). MHC-associated SNPs were excluded from the analysis by removing all SNPs on chromosome six between the genomic coordinates of 25,000,000 and 35,000,000. Three significance thresholds of genetic variant associations were tested for enrichment. Blue shows the odds ratio (OR) enrichment pattern for variants with a p-value threshold of 1, green the OR enrichment pattern for variants with a p-value threshold of 0.1, and pink shows the OR enrichment pattern for variants with a p-value threshold of $5x10^{-8}$. The latter corresponds to the minimum GWAS variant score required for genome-wide significance. ChromHMM emission states that are significantly enriched for by AS-associated GWAS SNPs have been highlighted in red shaded boxes. The p-value of enrichment in each ChromHMM emission state for AS-associated GWAS variants is shown starred at three levels of significance (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

One of the GARFIELD capabilities is the generation of a list of AS-associated GWAS SNPs that drive the enrichment of chromatin states. This analysis was limited to chromatin states that showed an odds ratio of 2 or greater in the MHC analysis, corresponding to at least a 100% increase in the odds that AS-associated variants are enriched in a given chromatin state (**Table S.1.**). All AS-associated ChromHMM emission states that showed an OR of greater than two in the global GARFIELD analysis were associated with promoter, enhancer, or silencer function. U2 TSSFlnkU and U3 TSSFlnkD were seen in the T cells and B cells, suggesting AS-variants influence promoter function in these cells. U6 TxNC, U10 EnhActv, and U16 PolyNC were seen across the PBMC cell types, suggesting important mechanisms for non-coding transcription, active enhancers, and non-coding Polycomb-repression in all isolated cell populations.

3.4. Chapter discussion

In this chapter, I have described the construction of high-quality epigenetic maps of AS-associated genetic loci by combining independent multi-omic epigenetic datasets into a singular ChromHMM model, and shown its relevance for functional prioritisation by enriching GWAS variants for the 16 identified chromatin states. A number of model limitations must however be considered.

3.4.1. The *NOS2* association signal is silenced by H3K27me3 in T cells, B cells, and monocytes and likely does not contribute towards AS pathogenesis in these cells.

As explained in the general introduction, it has been proposed that GWAS genetic variants that lie outside of known protein-coding regions can contribute towards disease by disrupting complex *cis*-regulatory regions and altering the expression profile of disease-associated genes. It is estimated that upwards of 98% of the genome does not

code for proteins and that over 90% of GWAS genetic variants can be found in these regions (Maurano *et al.* 2012). This is also true of AS, with many of these regions remaining functionally uncharacterised. This highlights our need to develop tools to interpret how GWAS variants can perturb the underlying regulatory landscapes and cause disease. In addition, the association between AS and MHC is relatively well known through HLA-B27, yet the contribution of non-MHC loci towards AS is poorly understood. The approach in this chapter was to define AS associated regulatory regions by combining ChIP-Seq, RNA-Seq, and ATAC-seq datasets from patients and volunteers to generate integrative models fit for mapping the regulatory environment outwith the MHC class-I association.

The first aim of this chapter was to identify putatively causal SNPs within the *NOS2* AS-associated GWAS risk locus. *NOS2* was chosen due to its high-association with AS, its unknown aetiological contribution towards the disease, and its direct association with the inflammatory response and immune system in murine immune cells. I first investigated the GWAS signal from Cortes *et al.* (2013) for the *NOS2* locus with our own early ChromHMM (Brown *et al.*, 2023) for regulatory features of interest. Having found evidence of context-specific enhancers in $CD14^{+}$ monocytes, I then attempted to produce a gene expression model utilising MDP and GMDP in monocyte-derived macrophages based upon the work of Landes *et al.* (2015). Observing no difference in the expression of the gene, these experiments suggested the silencing of *NOS2* in $CD14^{+}$ monocyte-derived cells. A later ChIPmentation-seq experiment suggested this could arise due to long-term Polycomb protein group repression through a large chromatin block enriched for the post-translational histone modification H3K27me3 in close proximity to the gene promoter region. This region was invariant across the T cells, monocytes, and B cells and was not differential between the AS and healthy states, suggesting strong repression of

NOS2 and lack of AS contribution in these cell types. The first aim of the chapter was therefore unsuccessful, as *NOS2* expression could not be recapitulated to investigate the context-specific ChromHMM elements predicted in $CD14^+$ monocytes.

It should be noted that in the *NOS2* analysis, I analysed the epigenome of monocytes yet conducted the expression experiments in monocyte-derived macrophages. This was due to limited literature on which human primary cells, if any, produce *NOS2* in the immune system. Whilst these macrophages derived from $CD14^+$ monocytes, there may be significant genomic variation at the *NOS2* locus in the two cell types that was not captured by ChromHMM. Indeed, it has been shown that monocyte-derived macrophages gain and lose about 6000 and 3000 enhancers respectively upon differentiation from monocytes (Hoeksema & de Winther, 2014). Unlike monocytes, macrophages exist on a cell state spectrum between M1 (pro-inflammatory macrophages) or M2 (immunosuppressive macrophages). In mice, it is M1 macrophages that produce *NOS2* (Palmieri *et al.*, 2020). Use of common macrophage differentiation compounds such as M-CSF on cultured human $CD14^+$ monocytes instead produce M2 monocytes (Wu *et al.*, 2022). Taken together, these *NOS2* experiments highlight the need to either carry out further ChromHMM profiling in cell types derived from the naïve cells analysed in this chapter – for example, Th17 from naïve $CD4^+$ T cells and macrophages from $CD14^+$ monocytes – or to limit the functional analysis to genes already expressed in analysed cell types.

3.4.2. Genome-wide H3K27me3 signals were not found to be differential in AS patients and healthy volunteers and suggested consistent Polycomb-associated gene repression between the two states

When H3K27me3 signals were investigated genome-wide and in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells, only a single H3K27me3 peak internal to

DCBLD1 in $CD14^+$ monocytes which has no known AS functionality was detected across the entire analysis. Based on this, I concluded that there was no difference in the peak heights of H3K27me3-rich regions between AS patients and healthy volunteers and that the vast majority of genes repressed by H3K27me3 – including *NOS2* – likely remain repressed in AS in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells.

This is surprising however, as recent study of Polycomb-associated repression signals in whole blood PBMCs suggested the expression levels of H3K27me3 and *EZH2* (Enhancer of zeste homolog 2) – a methyltransferase responsible for deposition of trimethyl groups on lysine 27 on histone 3 – were significantly lower in AS patients as measured by ELISA (Chen *et al.*, 2024). This decrease also correlated with an increase in the Th17 master-regulator transcription factor RORγt, suggesting an inverse link between DNA silencing and inflammation. As this was a global analysis on all cell types of the peripheral blood, it could be that H3K27me3 plays a more pathogenic roles in other blood cells outside of T cells, monocytes, or B cells.

It should be noted however that H3K27me3 signals are broader and far more widely distributed genome-wide than marks like H3K4me3 and H3K27ac (Zhao *et al.*, 2020). This makes H3K27me3 peaks difficult to define due to the diffuse signal. Whilst this was accounted for in the MACS2 pipeline using an adjusted q-value threshold of 0.1 *in lieu* of a p-value threshold of $1x10^{-8}$ for the H3K4me3 and H3K27ac marks (Brown *et al.*, 2023), it was not in the DESeq2 differential pipeline. The conditions of the latter pipeline may have been too strict to confidently detect differential H3K27me3 signals using the reported thresholds, and this merits further investigation to reach a proper conclusion on differential H3K27me3 expression.

3.4.3. Incorporation of H3K27me3 into ChromHMM identifies regions of Polycomb repression that can help identify inactive genes in T cells, B cells, and monocytes

Whilst these H3K27me3-rich regions may not be differential, elements containing H3K27me3 in addition to other histone marks may still show significant differences in AS patients compared to healthy volunteers. We have shown that the H3K4me3 histone mark that delineates promoters and transcription is considerably differential between AS patients and healthy volunteers in previous publication (Brown *et al.*, 2023). Regulatory elements that contain H3K4me3 can also demonstrate bivalency for H3K27me3, and these bivalent promoters are important in the expression of genes involved in cell development and differentiation (Más *et al.*, 2018). The second aim of this chapter was therefore to generate an integrated epigenetic model that will allow for the identification of biological DNA states that would be able to differentiate between repressed, bivalent, and active regulatory elements. Towards this goal, I incorporated the genome-wide H3K27me3 ChIPmentation-seq data into ChromHMM alongside H3K4me1, H3K4me3, H3K27ac, ATAC, and RNA data to segment the genome into biologically meaningful emission states. This was done not only for $CD14^+$ monocytes, but also for the $CD4^+$ T cells and $CD8^+$ T cells included in the 14-state model and *de-novo* for $CD19^+$ B cells.

16 emission states were identified that largely matched with our previous 14-state ChromHMM (Brown *et al.*, 2023), including Polycomb-repressed coding and non-coding elements (U15 PolyG & U16 PolyNC) and a bivalent promoter element (U4 TSSBvlnt). *NOS2* was used as the primary example to show successful state modelling. Whilst the promoter of *NOS2* was defined as U1 TSSActv, U2 TSSFlnkD and U3 TSSFlnkU in both the 14-state and 16-state models, a large block of Polycomb-associated U15 PolyG was found in the *NOS2* gene in all cell types. Dreger *et al.* (2016) – who additionally found that *NOS2* in endothelial cells is repressed by H3K27me3 at this site – has shown that

silencing RNA (siRNA) knockdown of *EZH2* lead to an upregulation of *NOS2* expression. This suggests that expression of the gene may also be repressed in the analysed immune cells by the same Polycomb-repressed chromatin block. Interestingly, in the new ChromHMM, the only genetic variant from the Cortes *et al.* (2013) *NOS2* credible set that directly overlaps with a functional element is found in the newly identified Polycomb-repressed chromatin block. The variant is *rs28998802*, which is strongly associated with psoriasis, IBD, and a poorer anti-TFNα therapy outcome (Vedak *et al.*, 2016). Stuart *et al.* (2015) also identified it as the likely causal SNP in both psoriasis vulgaris and cutaneous-only psoriasis, with the SNP modifying the consensus binding sequence of up to seven transcription factors.

Stuart *et al.* (2015) also identified that *NOS2* expression was differential in the lesional tissue – not peripheral blood – of psoriasis patients. Unless *rs28998802* is involved in the modulation of the repressive H3K27me3 mark directly, it is unlikely that $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, or $CD19^+$ B cells are the appropriate cell type for *NOS2* investigation. Whilst ultimately unsuccessful in my first chapter aim of identifying and functionally validating AS-associated genetic variants in the *NOS2* locus due to issues with H3K27me3-associated repression, my second aim to integrate this H3K27me3 data into a ChromHMM that is fully capable of identifying novel regions of repression and the SNPs within them was successful. The 16-emission state ChromHMM may therefore be used over the 14-emission state ChromHMM to additionally:

(1) Identify AS-associated GWAS risk loci that are extensively Polycomb-repressed in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells that remain epigenetically repressed between AS patients with active disease and healthy volunteers.
(2) Identify bivalent *cis*-regulatory regions that hint towards developmental, inflammatory, or immune importance, and how they may differ between AS patients and

healthy volunteers.

(3) Identify AS-associated genetic variants that lie in regulatory elements that were previously uncharacterised or of quiescent states in the 14-state model in AS patients and healthy volunteers.

3.4.4. AS-associated GWAS SNPs are enriched for *cis*-regulatory ChromHMM chromatin states that may differ between AS patients and healthy volunteers.

The third aim of this chapter was to identify which of these chromatin states are implicated in AS pathogenesis. Differential use of chromatin states in AS patients and healthy volunteers was assessed through permutation analysis and MCA, and enrichment of AS-associated GWAS variants was performed through GARFIELD enrichment analyses. The cumulative summary of these analyses can be found below (**Fig. 3.28**). MCA and permutation analyses revealed that the distribution of active enhancer and promoter chromatin states such as U1 TSSActv and U10 EnhActv was likely not significant between AS patients and healthy volunteers. This was to be expected, as the genomic positions of cis-regulatory elements are largely fixed within a given cell type regardless of disease or stimulation condition (Pollex *et al.*, 2024). Instead, context-specific regulation is believed to derive from dynamic interaction and activation of *cis*-regulatory elements in regards to environmental cues (Pollex *et al.*, 2024; Yang & Hansen, 2024).

Ultimately, there will be a number of genes in any given cell type under active enhancer influence, such a housekeeping genes, that will outweigh those for inflammatory or immune-linked genes. Indeed, Nasser *et al.* (2021) were able to identify 6,316,021 enhancer-gene connections for a total of 269,539 genes across 131 samples derived from immune cell lines – including T cell-like, B cell-like, and monocyte-like immune cell

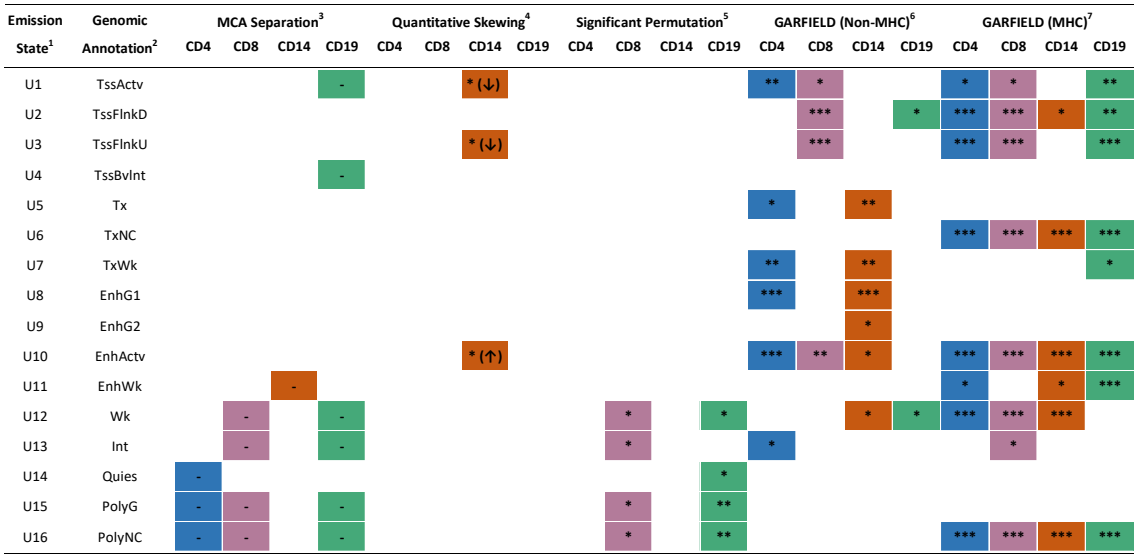

Figure 3.28. Summarisation of MCA, permutation, and GARFIELD enrichment analyses for all 16 detected ChromHMM emission states with and without the MHC region. Figure above shows the enrichment for each emission state within the 16-emission state ChromHMM produced over the course of chapter three. The first column shows the numerical identity of the 16 output ChromHMM emission states as in figure 3.8[1] . The associated predicted biological role is shown in the second column[2]. Whether each emission state separates AS patient and healthy volunteer samples on either dimension one or two of the MCA plot is shown in the third column[3]. No associated p-value has been given for this analysis. Quantitative skewing of emission states between AS patients and healthy volunteers is shown in the fourth column[4]. States with significantly greater frequency (↑) or lesser frequency (↓) in AS patients upon comparison with the healthy volunteers are indicated by respective arrows to show directionality of skewing. The fifth column shows those emission states that were significantly permuted within four AS patients and no healthy volunteers[5]. GARFIELD enrichment without the MHC region is shown in the sixth column[6], and the enrichment with the MHC region included is shown in the seventh column[7]. Cells have been coloured based on the colours of figure 3.9. Stars within each box are indicative of the significance level of the adjusted p-value for each emission state within the quantitative, permutation, or GARFIELD analyses (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

lines – and found that only 577 of those genes were linked to enhancers containing GWAS variants for 17 inflammatory conditions. Global permutation of *cis*-regulatory chromatin states therefore cannot be expected when comparing disease to healthy volunteer cohorts. Casella *et al.* (2022) have however shown that permutation analysis of individual genomic features can still help identify causal SNPs in enhancer elements at risk loci of complex traits if overlapped with further functional information.

Despite the lack of permutation or MCA separation, I found a significant increase in number of U10 EnhActv bins and significant decreases in the number of U1 TSSActv and U3 TSSFlnkD bins in $CD14^+$ monocytes. This was interesting, as AS-associated GWAS SNPs were found significantly enriched for U10 EnhActv in $CD14^+$ monocytes and U2 TSSFlnkU in $CD19^+$ B cells within both the MHC and non-MHC GARFIELD analyses. It could be that multiple of the AS-associated GWAS SNPs enrich for the new active promoter and enhancer elements that are unique to AS patients. This *de-novo* activation of *cis*-regulatory elements and chromatin remodelling, for example, is a common mechanism of hyperinflammatory macrophage priming on stimulation with IL-4 and LPS. (Kaikkonen *et al.*, 2013; Glass & Natoli, 2015; Czimmerer *et al.*, 2022). Many of these macrophage priming *cis*-regulatory regions intersect GWAS SNPs for multiple immune disorders (Alasoo *et al.*, 2018). T cells and B cells did not demonstrate quantitative skewing, MCA separation, or significant permutation, but GWAS SNPs in these cells were still enriched for promoter and enhancer chromatin states in both GARFIELD analyses.

Information on LD structure used to prune GARFIELD genetic variants was also generated from the UK10K cohort. As mentioned in the general introduction, LD can vary highly between different ethnicities, and these results may currently only be applicable to identifying genetic variants lying in *cis*-regulatory regions in populations of

European descent. These findings should be compared and contrasted to similar samples isolated from other ethnicities with high incidence of AS (i.e. East Asians) in order to identify those *cis*-regulatory element variants that are shared between the two populations. Identification of SNPs is also limited to the GWAS summary statistics data available for GARFIELD enrichment analysis.

Ideally, GARFIELD should have been run on both available GWAS summary statistics to generate two lists of enriched SNPs that could be compared, with high-confidence SNPs that appeared in both lists in high LD with the reported lead SNPs representing the best SNPs for further functional follow-up. Unfortunately, the Ellinghaus *et al.* (2016) summary statistics data has not been made readily available for use by the authors as of writing this thesis, and the GARFIELD was limited to association data for 24 of the 113 total loci reported across the studies. Due to the success of the enrichment analyses employed in this chapter however, incorporating these summary statistics when made available, and the summary statistics of future AS GWAS, will improve the identification of AS-associated SNPs that enrich for *cis*-regulatory ChromHMM chromatin states. Based on third chapter aim, despite these limitations, I conclude that I was successful in identifying that AS-associated GWAS SNPs at both MHC and non-MHC AS risk loci could be involved in dysfunction of *cis*-regulatory elements as predicted by ChromHMM, with these SNPs putatively contributing to altered gene expression in the disease.

3.4.5. AS-associated SNPs enrich for cis-regulatory elements but only MHC-associated AS SNPs enrich for Polycomb-repressed chromatin states

The fourth aim of this chapter was to analyse the impact that the MHC SNPs could have on the SNP enrichment pattern in ChromHMM-predicted regulatory regions. It was shown that AS-associated GWAS SNPs are enriched for *cis*-regulatory chromatin states

in both MHC and non-MHC analyses. AS-associated SNPs for MHC regions – but not non-MHC SNPs – were found to be enriched for Polycomb-repressed U16 PolyNC in all cells. This tied to the fifth chapter aim, which was to identify transition biases for the significantly permuted states. U16 PolyNC was significantly permuted and showed biased switching to U13 Int and U4 TSSBvlnt. As AS-associated SNPs enrich for U16 PolyNC at MHC loci only, this could suggest loss of Polycomb-repression and activation of MHC genes. The observation of MHC U16 PolyNC SNP enrichment and switching away from the chromatin state is striking, as it has been shown previously that decreased occupancy of H3K27me3 signals near the promoter regions of MHC class-I loci as a consequence of abrogating *EZH2* expression results in the increased expression of MHC class-I peptides; including HLA-B27 (Lehmann *et al.*, 2021). My findings therefore suggest that SNPs found in Polycomb-repressed regions of the genome could contribute towards the MHC-driven aspects of AS pathogenesis.

Furthermore, the contribution of MHC-associated variants towards global enrichment patterns was varied in the different cell types. In contrast to $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes, the enrichment of most states in B cells almost entirely relied on the MHC AS-associated SNPs, with a drop from eight chromatin states enriched in AS-associated GWAS SNPs to only U2 TSSFlnkU and U12 Wk in the non-MHC analysis. Although a clear role for B cells in AS has still to be elucidated, my findings suggest that AS-associated SNPs at MHC loci may have a prominent role in B cells in their contribution to disease pathogenesis. In the other cell types, the most major difference between MHC and non-MHC GARFIELD analyses is the loss of U16 PolyNC enrichment, suggesting the modulation of gene silencing is not a prominent feature of AS-associated GWAS variants at non-MHC loci.

3.4.6. Concluding remarks

In GWAS, the majority of association signals lack a functional characterisation due to residing outside of known protein-coding regions. In this chapter, I have shown that although there are limitations, integrative modelling of the epigenetic landscape using ChromHMM, permutation analyses, and GARFIELD enrichment analysis surrounding AS-associated GWAS loci can be used to facilitate the identification of novel putatively causal genetic variants. By looking at individual cell types, I established a putative key role for non-coding genetic variants in *cis*-regulatory elements at non-MHC loci in all cell types, with particular emphasis on $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes. This chapter has also highlighted the importance of investigating the functional effects of SNPs in appropriate cellular models and disease contexts in AS and other immune or inflammatory conditions.

"You must know that a person's ability to discern the truth is directly proportional to his knowledge."

- Cixin Liu, The Three-Body Problem (2008)

4

AS-ASSOCIATED GWAS LEAD SNPs SHOW *CIS*-REGULATORY LOOPING TO INFLAMMATION-RELATED GENE PROMOTERS WITH CAPTURE-C

Contents **Pg.**

4.1. Chapter introduction 173

4.2. Experimental design and control loci 178

4.3. SNP-promoter interactions at AS-associated loci 184

4.4. Complex and context-specific interaction profiles 196

4.5. Chapter discussion 208

In this chapter, reciprocal Capture-C is used to investigate chromatin looping interactions between AS-associated lead SNPs and gene promoters in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes.

4.1. Chapter introduction

4.1.1. Chromosome conformation capture can link SNPs in regulatory regions with cognate genes

Two major limitations arise from the ChromHMM analysis in chapter three. First, many SNPs within AS risk loci can overlap independent ChromHMM-predicted regulatory elements that may contribute to expression of genes in the locus (Benton *et al.*, 2019).

These regulatory elements can also demonstrate different activity levels in different cell types despite sharing the same ChromHMM emission state (Vu & Ernst, 2022). Second, multiple genetic variants in LD can overlap the same predicted regulatory element. In the study of complex traits, genetic variants exhibiting LD can contribute through influencing multiple functionally related enhancers, although this is not the case for every locus (Corradin *et al.*, 2014; Nasser *et al.*, 2021). It now appears that large blocks of long-range linkage disequilibrium are more frequent than previously expected; the linkage of two genetic variants in *cis*-regulatory-like predicted ChromHMM emission states is actually extremely common and complicates the process of identifying causal variants precisely (Park, 2019). However, *cis*-regulatory elements contact cognate genes in 3D space through context-specific mediator complexes to control gene expression, and these contacts can be revealed using 3C-based methodologies. ChromHMM and chromatin conformation capture studies therefore complement one another, with the overlap of ChromHMM allowing functional characterisation of the DNA underlying a significant chromatin interaction, and the intersection of capture data showing which of the ChromHMM bins physically interact with gene promoters and contribute to regulation.

4.1.2. Capture-C has previously been used successfully in inflammatory conditions to show that differential chromatin interactions may be involved in disease pathogenesis

The control of enhancer-promoter contacts is important in the context-specific expression of genes, and alterations in when, where, and how strongly enhancers contact gene promoters can contribute to disease pathogenesis. Indeed, Oh *et al.* (2021) were able to demonstrate through a combination of targeted silencing screens and Hi-C, that the timed release of an enhancer from its respective promoter – and its subsequent looping to alternative promoters within the same locus – formed an important pathogenic

mechanism within the respective cancer loci of *CLPTM1L–TERT*, *ZCCHC7–PAX5* and *PVT1–MYC*, and also the Parkinson's disease-associated *NUCKS1–RAB7L1* locus. Similar studies have also implicated differential chromatin contacts in the pathogenesis of several inflammatory arthropathies, – including psoriatic arthritis, rheumatoid arthritis and juvenile idiopathic arthritis) – and other inflammatory/autoimmune conditions, such as, IBD, type 1 diabetes (T1D), multiple sclerosis (MS), and systemic lupus erythematosus (SLE) (Gateva *et al.*, 2009; Martin *et al.*, 2015; Meddens *et al.*, 2016; Carrini *et al.*, 2018; Ge *et al.*, 2021, Zhao *et al.*, 2023). These examples demonstrate how alteration of physical contact between a promoter and enhancer element can alter gene expression patterns in disease. To date however, no studies have identified differential chromatin looping events in AS.

4.1.3. Selecting the appropriate the appropriate chromosome conformation capture assay for investigation of AS-associated lead SNP interactions.

In this chapter, a reciprocal Capture-C approach is used to characterise the looping interactions between 54 AS-associated lead SNPs (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016) and 44 of their most proximal gene promoters (**Table 2.4**). These findings are also intersected with ChromHMM to identify SNP-containing *cis*-regulatory elements that interact with promoters that could influence abnormal gene expression.

Conceptually, Hi-C studies provide the most comprehensive all-vs-all genome-wide image of chromatin interactions. However, these studies are limited by experimental resolution. Chromatin conformation capture studies such as Hi-C have historically used either six-base pair cutting restriction enzymes, such as HindIII, or four-base pair cutters, such as DpnII, to fragment the genome for capture. DpnII digestion results in an average fragment size of 500bp whereas HindIII produces fragments closer to 4kbp (Downes *et*

al., 2022). As the bins of the 16-emission state ChromHMM generated in chapter three are 200bp in size, DpnII would be a logical choice for fragmentation of the genome. This would effectively map looping interactions to the resolution of two and a half ChromHMM bins at AS-associated GWAS risk loci. However, deeper resolution necessitates deeper sequencing read depth, and the cost of producing DpnII Hi-C contact maps at the depth required to investigate differential interactions in AS is considerable (Greenwald *et al.*, 2019; Yardımcı *et al.*, 2019).

The Cortes *et al.* (2013) and Ellinghaus *et al.* (2016) GWAS clearly identified regions of AS association for further investigation. After fixation of the protein complexes found at *cis*-regulatory elements and digestion with DpnII, restriction fragments containing lead SNPs were targeted with biotinylated oligonucleotides employed in Capture-C approaches to enrich for specific capture viewpoints of interest. This drastically decreases sequencing costs whilst providing interaction profiles for the regions of DNA with the highest disease association. Because genome-wide eQTL maps show that the average distance between a causal variant and target gene is far below 100 kbp and typically affect the expression of genes in close proximity (Veyrieras *et al.*, 2008; Kumosaka *et al.*, 2019), the nearest gene TSS to each lead SNP was also captured and interactions between the two were compared.

In this chapter, Capture-C is performed in healthy volunteer and AS patient $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ T cells. $CD19^+$ B cells were omitted from subsequent analyses due to difficulty in isolating adequate cell numbers for Capture-C. As these experiments were completed prior to the revision of the ChromHMM and related analyses, the SNPs that enriched for active enhancer (U10 EnhActv) chromatin states were not baited for capture. Capture-C data presented in this chapter are also intersected with ChromHMM to identify which lead SNP interactions lie in any of the predicted *cis*-regulatory

chromatin states. To assess the statistical significance of Capture-C interactions in each of the conditions, the R package Peaky is used (Eisjbouts *et al.*, 2019; Brown *et al.*, 2023). The most important interactions show both a high mean interaction count and a high peaky score (-$\log_{10}$ FDR).

4.1.4. Chapter aims

It was hypothesised that disease-specific chromatin looping events could form between gene promoters and the lead SNPs of AS GWAS. The existence of differential interactions could help underpin how causative genetic variants or *cis*-regulatory elements contribute to thepathogenesis of AS. The overall aim of this chapter is to investigate how chromatin interactions between lead SNPs and/or regulatory regions may differ in AS. These studies should also help to prioritise AS-associated risk loci for further study. The specific aims of this chapter are:

[1] To characterise chromatin looping from candidate gene promoters at AS-associated GWAS loci.

[2] To identify putative functional SNPs that lie in regulatory regions involved in chromatin loops at AS GWAS loci.

[3] To characterise quantitative differences in looping frequency between AS patients and healthy volunteers.

4.2. Experimental design and analysis of known interactions at control loci for method validation

4.2.1. Experimental Design

Because the targeting of two distal fragments in the same reaction can introduce significant signal bias for interactions between the promoter and lead-SNP viewpoints when compared to adjacent, untargeted restriction fragments (Downes *et al.*, 2022), Capture-C experiments were carried out in independent capture bait sets or 'tiers' (**Fig. 4.1**). The first capture bait set consisted of the baits for restriction fragments that contained known TSS regions that were closest to the lead SNPs. The second capture bait set consisted of the baits for restriction fragments containing the lead SNP within the same locus in either the Cortes *et al.* (2013) or Ellinghaus *et al.* (2016) studies. The third and fourth capture bait sets were used in loci that either demonstrated multiple LD blocks with independent GWAS association signals, or those where multiple SNPs in the same LD block have proven influence over the expression of genes within the locus. This baiting strategy allowed for reciprocal validation of promoter-lead SNP interactions whilst accounting for lead SNPs that may interact with alternative gene promoters, or promoters that interact with multiple different regulatory elements outside of the GWAS association signal.

4.2.2. Interactions at control loci provides confidence in Capture-C data quality

In order to provide confidence in the quality of the Capture-C data for comparison of interactions between healthy volunteers and AS patients, three control loci – *INPP4B*, *MYC*, and *LMO1* – were captured in addition to the AS-associated regions. Previous literature had identified verified interactions between regulatory elements at these regions which should be recapitulated on analysis of the capture data.

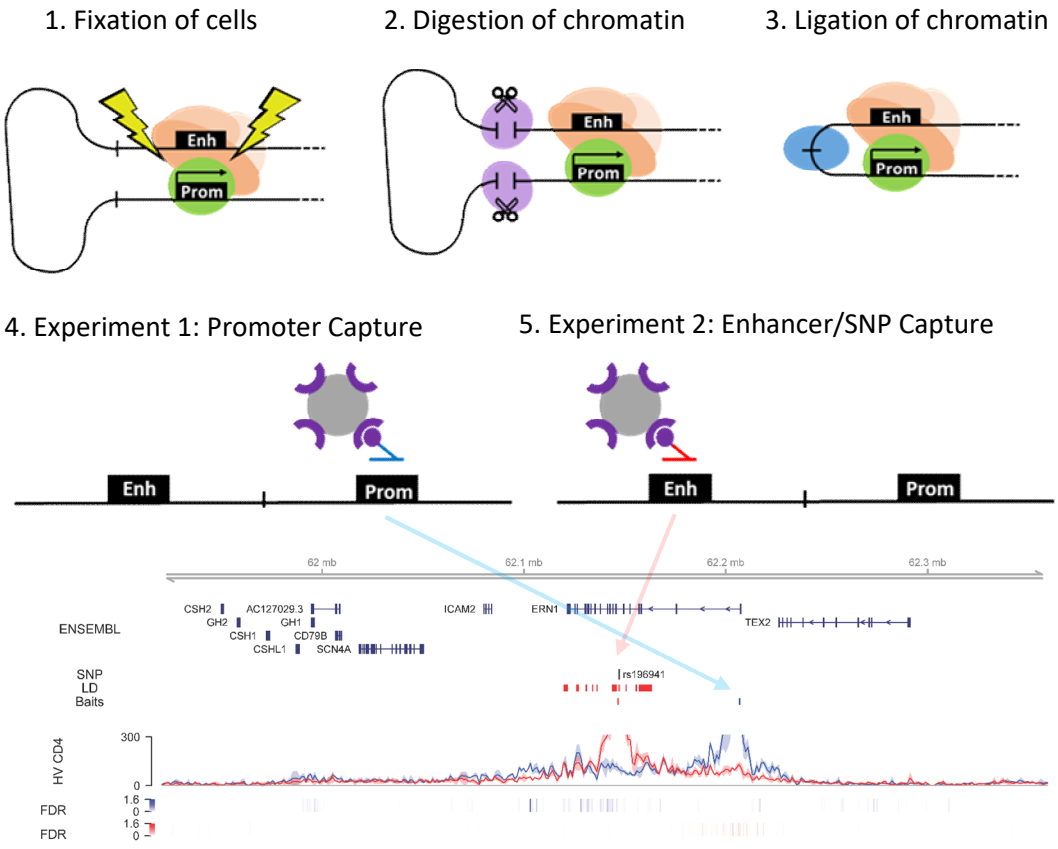

Figure 4.1. Experimental design for reciprocal Capture-C experiments. Experimental workflow largely followed protocols from Downes *et al.* (2022). (1) Cells were first fixed with formaldehyde to preserve the peptide interactions between *cis*-regulatory elements. (2) Chromatin was then digested with a restriction enzyme (DpnII) to fragment the genome whilst preserving the DNA underneath the fixed protein complexes. (3) DNA was then ligated, and the crosslinks reversed to produce DNA fragments consisting of genetic elements in close spatial proximity. (4 & 5) DNA was then separated into different samples for the tiered oligonucleotide capture – here either the promoter (Prom) oligo (blue) or Enhancer (Enh)/lead SNP oligo (red) viewpoints – where biotinylated oligonucleotides targeting regions of interest were pulled-down using streptavidin beads. (6) Pulled-down DNA fragments were then multiplexed for sequencing, and ligation junctions were recreated *in-silico* prior to mapping to the genome with ICeCAP. Interaction significance was then calculated from Peaky and the mean read counts was plotted using CaptureCompare. The reciprocal interaction profile for the *ERN1* promoter and the lead-SNP *rs19691* is shown. 'SNP' shows the location of the lead SNP. 'LD' shows regions in high linkage disequilibrium ($R^2 \geq 0.8$) with the lead SNP. 'Baits' shows the position of the oligonucleotide capture baits (Promoter bait in blue, lead SNP bait in red). 'HV CD4' shows the mean interaction counts in AS patient $CD4^+$ T cells. 'FDR' shows the $-\log_{10}$ FDR Peaky interaction score for both the promoter restriction fragment (blue) and lead-SNP restriction fragment (red). Both the promoter and lead SNP restriction fragments show interactions nearby each other and with another upstream element.

The locus chr11p13 contains the *LMO2* (LIM domain only 2) gene. An interaction exists between a viewpoint just downstream of the *LMO2* promoter and a boundary domain 120kbp upstream in Jurkat cells (Hnisz *et al.*, 2016; Dekker *et al.*, 2017). In the Capture-C data, this interaction was observed in healthy volunteer (Maximum Peaky -$\log_{10}$ FDR = 7.0) and AS patient $CD14^+$ monocytes (Maximum Peaky -$\log_{10}$ FDR = 6.4) (**Fig. 4.2: Red shaded box**). The distal element was characterised as U10 EnhActv by ChromHMM specifically in $CD14^+$ monocytes, suggesting active enhancer activity. No interactions were seen for *LMO2* in either $CD4^+$ or $CD8^+$ T cells at this distal element. This was expected since this region is characterised as polycomb-repressed (U15 PolyG and U16 PolyNC) by ChromHMM. This suggested silencing of interacting *cis*-regulatory elements seen in the $CD14^+$ monocytes.

Similarly, known cell-type specific interactions were also detected at the chr8q24.21 locus which contains *MYC* (Myc proto-oncogene) (Hnisz *et al.*, 2013) (**Fig. 4.3**) and chr4q31.21 which contains *INPP4B* (Inositol polyphosphate-4-phosphate) (Javierre *et al.*, 2016) (**Fig. 4.4**). These loci showed high mean interaction counts at the correct genomic positions for the validated interactions but no associated Peaky FDR score due to the distance over which these interactions occur. To save time and computational resourced, Peaky analysis in this chapter was limited to 500 kbp with both *MYC* and *INPP4B* interactions occurring over this limit. Despite this, the Capture-C data appears to be of high quality and is able to recapitulate the profiles of previously validated interactions. Intersection with ChromHMM shows that interactions occur at active enhancer elements (i.e. U1 TSSActv, U10 EnhActv) and is inhibited by Polycomb-repression (i.e. U15 PolyG, U16 PolyNC).

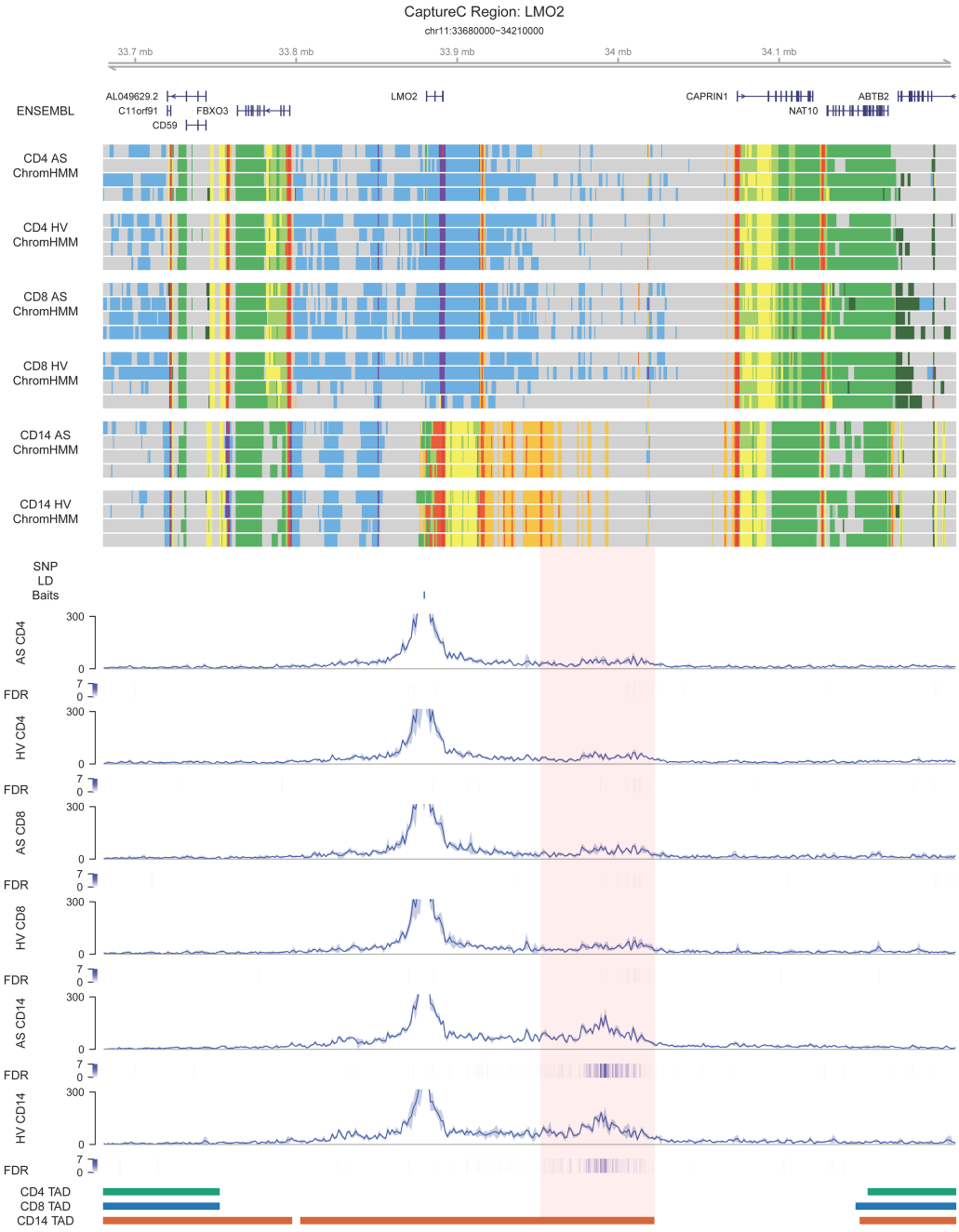

Figure 4.2. A distal active enhancer element that is repressed in CD4$^+$ T cells and CD8$^+$ T cells loops onto the *LMO1* promoter in CD14$^+$ monocytes. Interaction profile a the chr11p13 control locus. From top to bottom: Scale bar showing the genomic coordinates of genetic elements in the locus. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows the ChromHMM from chapter three in CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region(s) in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

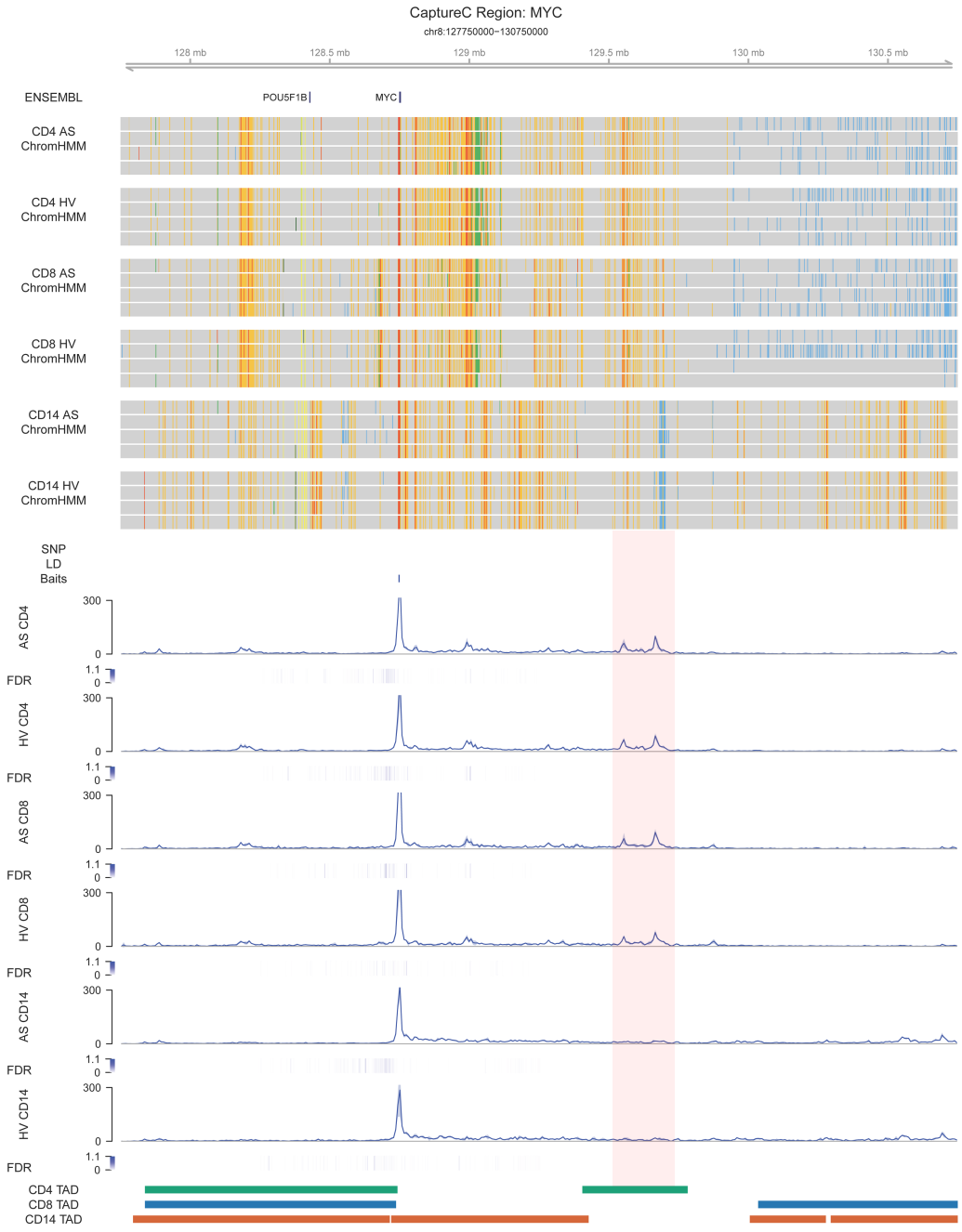

Figure 4.3. A distal active enhancer element loops onto the *MYC* promoter in CD4$^+$ and CD8$^+$ T cells that is Polycomb-repressed in CD14$^+$ monocytes. Interaction profile a the chr8q24.21 control locus. From top to bottom: Scale bar showing the genomic coordinates of genetic elements in the locus. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

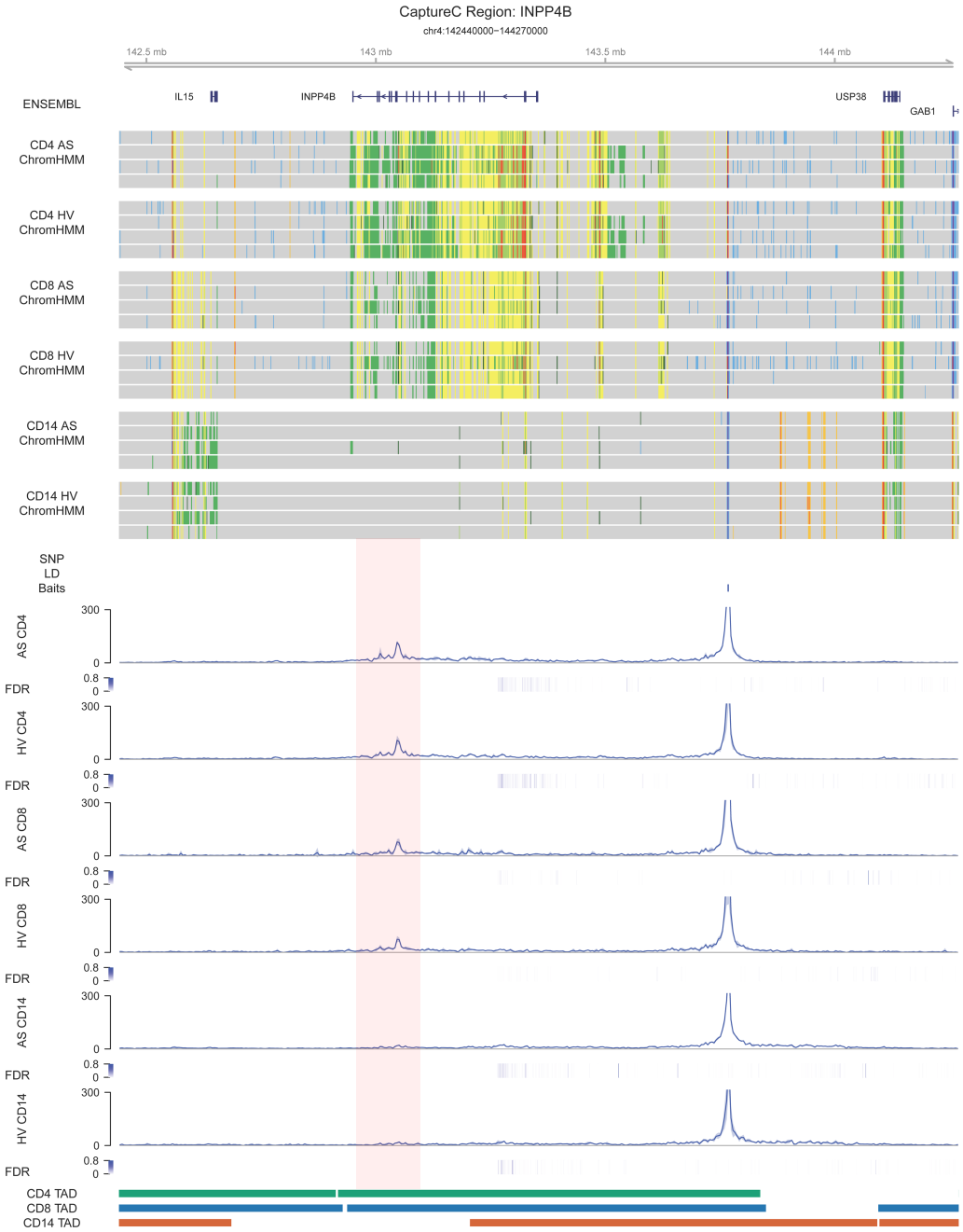

Figure 4.4. A distal bivalent promoter region loops onto the end of *INPP4B* in CD4$^+$ and CD8$^+$ T cells but not in CD14$^+$ monocytes. Interaction profile a the chr4q31.21 control locus. From top to bottom: Scale bar showing the genomic coordinates of genetic elements in the locus. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

4.3. SNP-promoter interactions at AS-associated loci

4.3.1. Summary of interactions and intersection of Capture-C with epigenetic data

Of the 47 gene promoters and 52 AS-associated signals investigated with reciprocal Capture-C, nine of the analysed regions contained *cis*-regulatory interactions between gene promoters and SNPs in ChromHMM-defined regulatory elements (**Table 4.1**). These loci either demonstrated reciprocal looping between lead SNP and targeted promoter viewpoints, looping of the AS lead SNP to untargeted gene promoters, or looping of targeted gene promoters onto other regulatory elements outside of the AS lead SNP or their associated 99% credible set intervals (van de Bunt *et al.*, 2015). Interactions were defined in this chapter by both a high mean interaction count and high Peaky interaction scores. The Peaky algorithm finds sets of regions at which peaks of varying strengths combinedly explain the normalised Capture-C signal (Eisjbouts *et al.*, 2019).

Peaky scores are however subjective and must be analysed on a per-locus basis with intersecting epigenetic information to determine if the interaction occurs between genetic elements that putatively contribute to AS pathogenesis. In the subsection, three loci which show strong reciprocal context-specific interactions with *cis*-regulatory chromatin states and AS-GWAS lead SNPs – *ERN1, PTGER4*, and *ETS2* – will be discussed in detail. In order to ascertain how chromatin interactions at AS-associated GWAS loci could contribute to disease pathogenesis, it should be noted that the Capture-C data is intersected with a number of different epigenetic assays to inform how the biology between interacting genetic elements could change in AS:

[1] The 16-emission state ChromHMM is used to characterise the overarching biology behind each of the interactions. Specifically, this is used to assess whether interactions occur between active *cis*-regulatory elements like promoters and enhancers, and whether

Locus[1]	Gene[2]	Capture-C[3] CD4 CD8 CD14	H3K4me3[4] CD4 CD8 CD14	H3K27ac[5] CD4 CD8 CD14	H3K27me3[6] CD4 CD8 CD14	ATAC[7] CD4 CD8 CD14	eRNA[8] CD4 CD8 CD14	Capture-C Figures[9]	Differential Figures[10]
chr1p36.11	*RUNX3*	■ ■ ■						Fig. 4.9	Fig. 4.10
chr2q11.2	*IL18RAP*	■ ■		■		■ ■ ■		Fig. S.3	Fig. S.4
chr2131.3	*ITGA4*	■ ■ ■	■ ■	■		■ ■ ■	■	Fig. 4.11	Fig. 4.12
chr5p13.2	*IL7R*	■ ■	■ ■ ■	■ ■				Fig. S.1	Fig. S.2
chr5p13.1	*PTGER4*	■ ■ ■	■ ■ ■	■		■ ■ ■	■	Fig. 4.7	Fig. 4.8
chr6q15	*BACH2*	■ ■						Fig. S.5	Fig. S.6
chr11q24.3	*ETS1*	■ ■	■ ■ ■	■ ■ ■		■ ■ ■	■ ■	Fig. 4.13	Fig. 4.14
chr17q23.3	*ERN1*	■ ■ ■	■	■ ■ ■		■ ■		Fig. 4.5	Fig. 4.6
chr21q22.2	*ETS2*	■				■		Fig. 4.8	Fig. 4.9

Table 4.1. Summarisation of AS-associated GWAS loci showing *cis*-regulatory looping that intersects differential chromatin features between AS patients and healthy volunteers. 'Locus[1]' shows the name of each of the AS-associated loci from Cortes *et al.* (2013) and Ellinghaus *et al.* (2016). 'Gene[2]' shows the name of the gene found to associate with lead and/or functional SNP(s) in each locus. 'Capture-C[3]' shows whether the interaction was observed in $CD4^+$ or $CD8^+$ T cells or $CD14^+$ monocytes. 'H3K4me3[4]', 'H3K27ac[5]', 'H3K27me3[6]', 'ATAC[7]', and 'eRNA[8]' show which of the epigenetic marks were significantly differential between AS patients and healthy volunteers at each locus. 'Capture-C Figures[9]' shows the figure number for each locus showing the Capture-C interaction profiles. 'Differential Figures[10]' shows the figure number for each locus showing the intersection of interactions with the various differential epigenetic marks. Genetic features in $CD4^+$ T cells are shown in blue, $CD8^+$ T cells in pink, and $CD14^+$ monocytes in orange.

these interactions could affect the expression of associated genes.

[2] Differential ATAC-seq data is used to assess where regions of chromatin adopt more open or closed conformations in the disease state, implying where accessibility to transcription factors or other associated DNA binding proteins may be attenuated or strengthened in AS.

[3] Differential ChIPmentation-seq data is used to inform the position and relative strength of multiple histone post-translational modifications:

- H3K4me3 is associated with RNA polymerase II priming and is used to assess the transcriptional readiness of gene promoters in each locus.
- H3K27ac is associated with the stabilisation of looping interactions and *cis*-regulatory activity and is used to assess how readily genetic elements interact in the AS and healthy states.
- H3K27me3 is associated with Polycomb-repression and is used to determine whether genes and regulatory elements are more or less repressed in AS patients compared to healthy volunteers.

[4] Total RNA-seq data was used to determine differential eRNA tracks to derive the quantity of eRNA produced from enhancers in each locus. eRNA is also linked to loop stabilisation and is used as another metric to measure enhancer strength in the disease and healthy states.

4.3.2. The chr17q23.3 locus shows clear reciprocal looping interactions between the promoter of *ERN1* and *rs196941* in CD4$^+$, CD8$^+$, and CD14$^+$ cells

The chr17q23.3 AS-associated risk locus contains the *ERN1* (endoplasmic reticulum to nucleus signalling 1 enzyme) gene that is critical in the sensing of ER stress and the

unfolded protein response. Also known as IRE1α (inositol-requiring 1 alpha), IRE1α/ERN1 activates the transcription factor sXBP1 (spliced X-box–binding protein 1) under ER stress and induces the expression of stress alleviating genes (Asada *et al.*, 2015). The downstream IRE1α/sXBP1-signalling target TNAP (tissue-nonspecific alkaline phosphatase) shows increased expression in response to misfolded AS-associated HLA-B*2705 molecules in the ER, which has been linked to the mineralisation of extraneous spine-introduced mesenchymal stem cells and the formation of vertebrae-spanning syndesmophytes in mouse models (Liu *et al.*, 2019). Changes in ERN1/IRE1α regulation and expression may therefore contribute to AS pathogenesis through promotion of spinal ossification in patients.

The lead SNP for the locus is *rs196941*, which demonstrates genome-wide significance (p-value = 8.2 x 10^{-10}) (Ellinghaus *et al.*, 2016). Reciprocal interactions that show high Peaky scores (Maximum Peaky -$\log_{10}$ FDR = 1.3) are observed between *rs196941* and the *ERN1* promoter region baits in all cell types (**Fig. 4.5: Red shaded box**). *rs196941* in an intron of *ERN1*, with ChromHMM predicting this region to constitute active promoter chromatin states (U1 TSSActv and U2 TSSFlnkD) in the T cell populations, suggesting an internal cis-regulatory element. In monocytes, this region is instead predicted as the weak enhancer-like U11 EnhWk. Overall, the Capture-C data supports a model where an intragenic promoter element containing a disease-associated SNP loops onto the *ERN1* promoter in T cells.

In $CD4^+$ T cells, this region is defined by both significantly stronger H3K27ac deposition and significantly less accessible chromatin in AS patients (**Fig. 4.6: Red shaded box**). In $CD8^+$ T cells, this region is instead defined by significantly weaker ATAC-seq peaks in AS patients. This suggests that the interacting promoter element could be more active in $CD4^+$ T cells and less active in $CD8^+$ T cells, with the increased deposition of H3K27ac

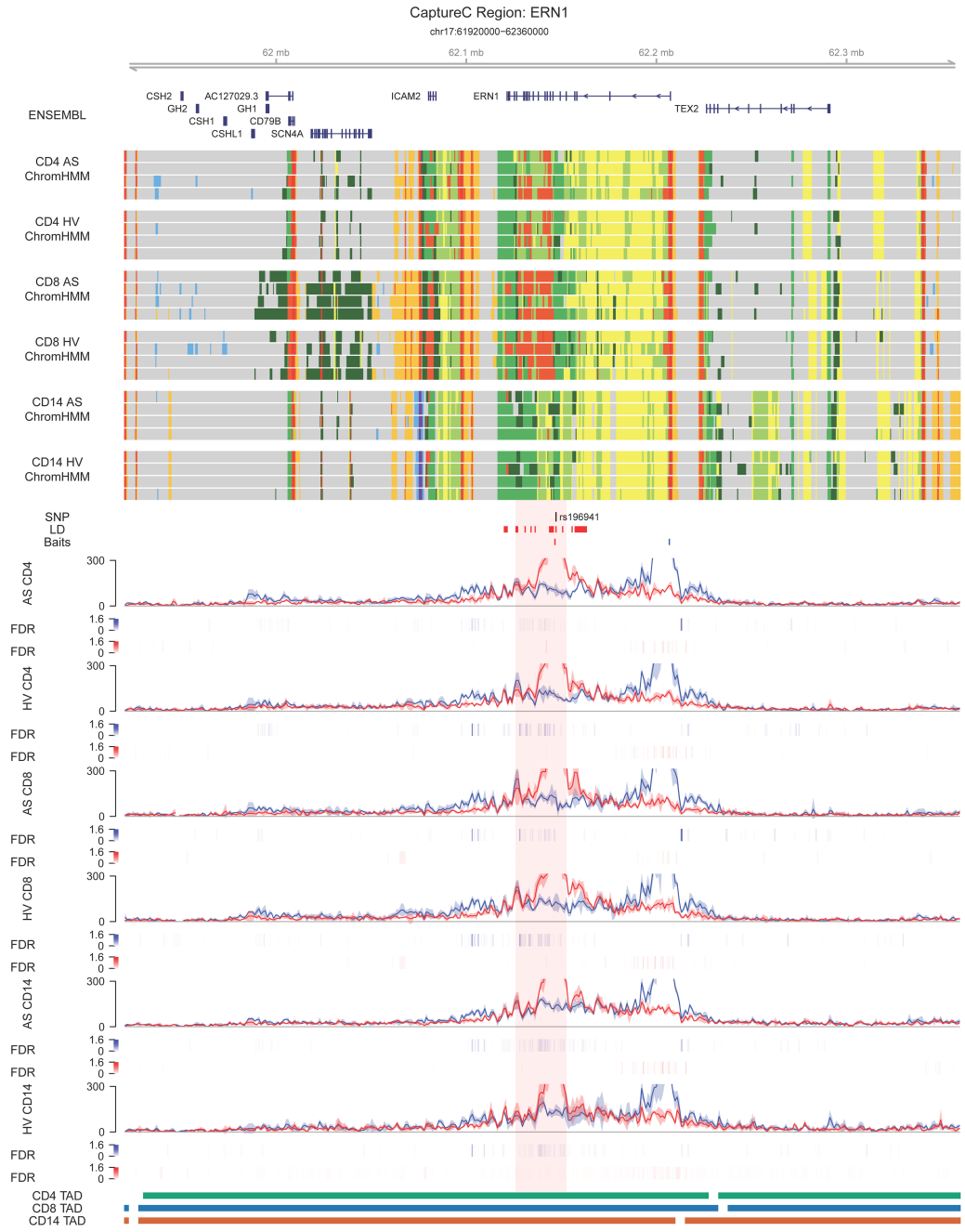

Figure 4.5. Enhancer-gene interactions between the promoter of *ERN1* and *rs196941* at the chr17q23.3 locus. From top to bottom: Scale bar showing the genomic coordinates of genetic elements in the locus. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

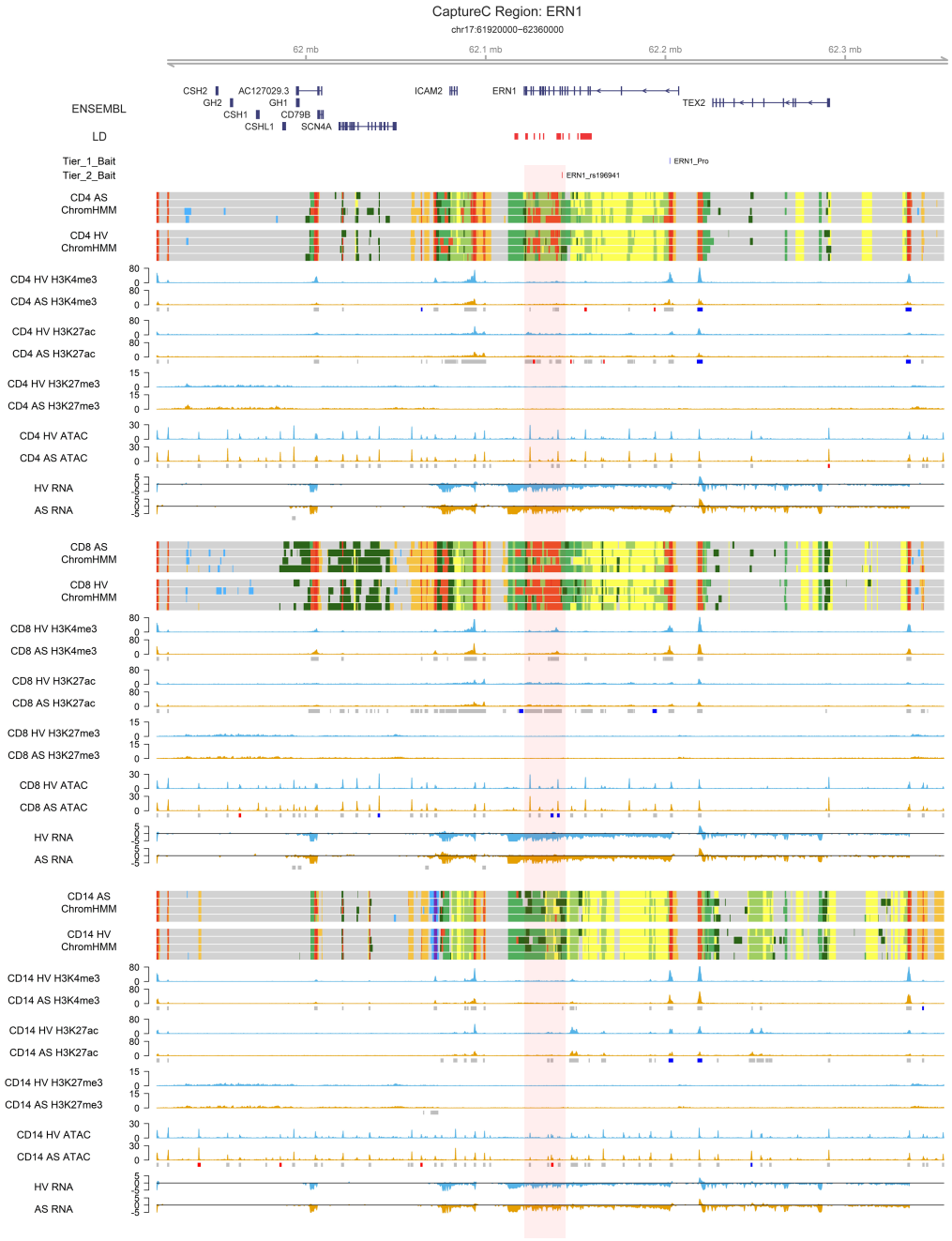

Figure 4.6. Differential epigenetic marks at the chr17q23.3 locus show the lead SNP-*ERN1* promoter interaction is more active in CD4$^+$ T cells and less accessible in CD8$^+$ T cells in AS patients. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red shaded box shows the location of the *ERN1* interacting lead SNP-containing *cis*-regulatory regions.

promoting DNA looping and the decreased chromatin accessibility preventing association of DNA-binding proteins.

4.3.3. Cell-type specific interactions are found between two independent AS-associated SNPs and the promoter of *PTGER4*

The *PTGER4* gene within the within the chr5p13.1 AS-associated risk locus encodes for prostaglandin E receptor 4. *PTGER4* is a GPCR which recognises the prostanoid ligand PGE_2. *PTGER4* signalling in naïve $CD4^+$ T helper cells has been shown to enhance IL-23 signalling cascades and promote the pro-inflammatory expansion of Th1 and Th17 T helper cell subsets (Yao *et al.*, 2009; Lee *et al.*, 2019). *PTGER4* is also important in T cell exhaustion – which is defined as a reduced or dysfunctional response to chronic antigen stimulation – with *PTGER4* expression increased particularly within exhausted $CD8^+$T cells (Wu *et al.*, 2023). In $CD14^+$ monocyte-derived macrophages, PGE_2 signalling is typically associated with suppression of TNFα with increased expression observed upon LPS stimulation (Tang *et al.*, 2017). With increased *PTGER4* expression in Th17 cells correlating with higher BASDAI scores in AS patients (Klasen *et al.*, 2019), the broad immune and inflammatory-related effects of *PTGER4* in several immune cell types makes the GPCR an attractive therapeutic target within AS.

Two suggestive independent AS-association signals are described at the *PTGER4* locus centred around the SNPs *rs12186979* (p-value = 4.3×10^{-6}) (Cortes *et al.*, 2013) and *rs1992661* (p-value = 3.2×10^{-5}) (Ellinghaus *et al.*, 2016). Another SNP in *rs9283753* has also been shown to lie within an enhancer-like element and alter *PTGER4* expression in lymphoblastoid cell lines (Tewhey *et al.*, 2016). To better characterise the interaction of each of these genetic variants with the *PTGER4* promoter, all three of the SNPs were baited in four independent capture bait sets (**Fig. 4.7**). I observed interactions between

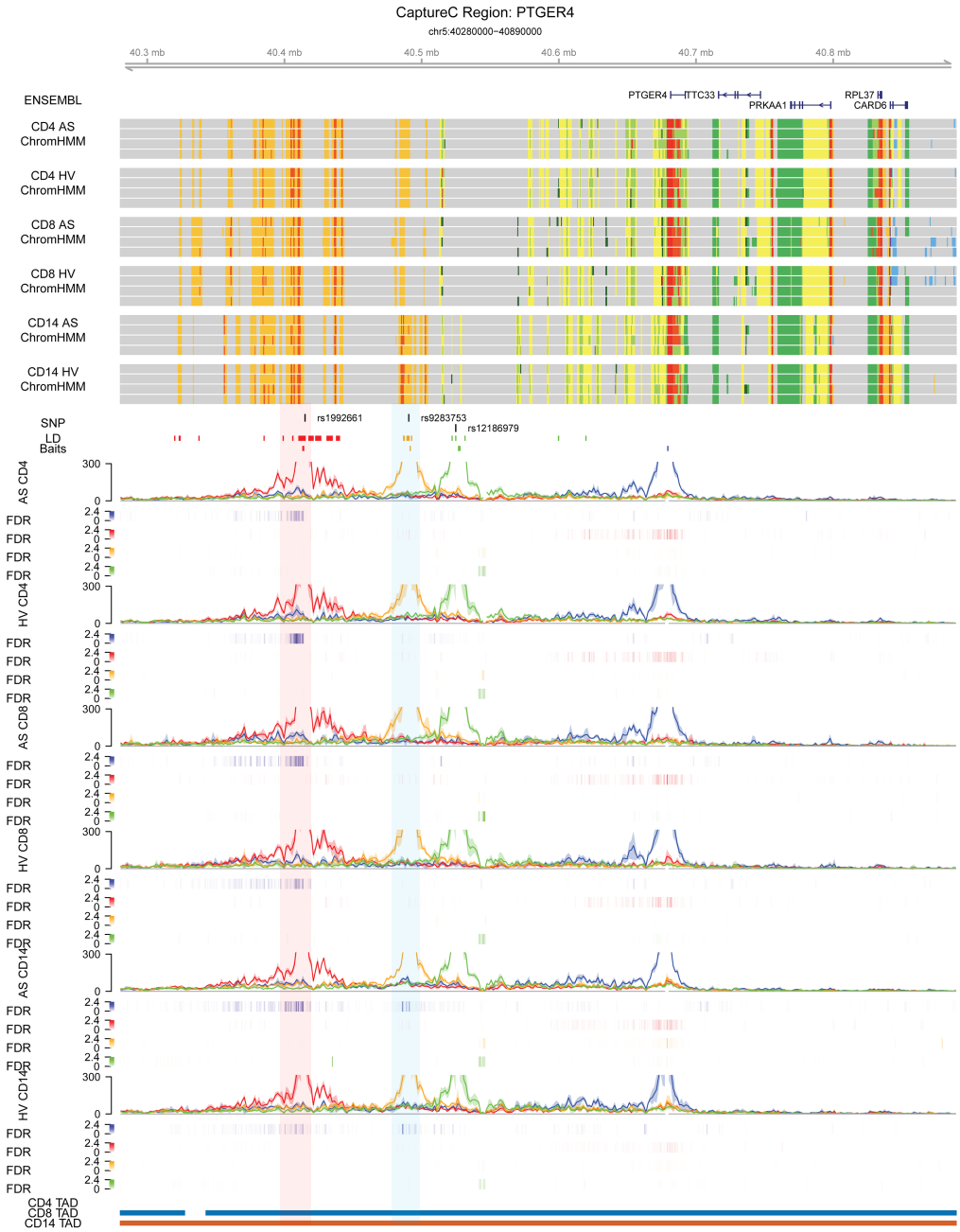

Figure 4.7. Enhancer-gene interactions between the promoter of *PTGER4*, *rs1992661*, *rs9283753* and *rs12186979* at the chr5p13.1 locus. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

the *PTGER4* promoter and *rs1992661* in all analysed cell types (Maximum Peaky -$\log_{10}$ FDR = 2.4) (**Fig. 4.7: Red shaded box**). *rs1992661* overlaps an enhancer (U10 EnhActv) flanked by promoter (U2 TSSFlnkD and U3 TSSFlnkU) chromatin states. I also observed a significant Peaky interaction (Maximum Peaky -$\log_{10}$ FDR = 2.3) between the *PTGER4* promoter and the functional SNP *rs9283753* specific to $CD14^+$ monocytes (**Fig. 4.7: Blue shaded box**). This interaction overlapped an active cis-regulatory region showing enhancer (U10 EnhActv), promoter (U2 TSSFlnkD and U3 TSSFlnkU) chromatin states. This interaction was absent in T cells despite ChromHMM predicting the SNP to intersect an active enhancer enhancer (U10 EnhActv) elements; suggesting the *rs928375* enhancer may specifically alter *PTGER4* expression in $CD14^+$ monocytes. No Peaky interactions were seen between the *PTGER4* promoter and *rs12186979* in $CD4^+$ T cells, $CD8^+$ T cells, or $CD14^+$ monocytes. This suggests the SNP does not contribute to the regulation of *PTGER4* expression in these cell types.

A number of differential epigenetic marks intersect the interaction *cis*-regulatory elements at the *PTGER4* locus and suggest their pathogenicity in AS. The predicted *cis*-regulatory element around *rs1992661* shows significantly stronger H3K27ac peaks in AS patient $CD4^+$ T cells and $CD14^+$ monocytes compared to healthy volunteers, suggesting enhancer activity and contact between genetic elements in the *rs1992661* LD block and the *PTGER4* promoter is stronger in these cells in the disease state (**Fig. 4.8: Red shaded box**). Coinciding with increased H3K27ac, *rs1992661* also overlapped a differential eRNA signal that was stronger in AS patient $CD4^+$ and $CD8^+$ T cells, further suggesting increased enhancer activity at this region in these cells. These results support a model where the active enhancer overlapping the AS-lead SNP *rs1992661* loops to the promoter of *PTGER4*. The presence of differential H3K27ac and eRNA marks underlying this interaction in AS patient may hint towards an increase in enhancer activity in AS patients,

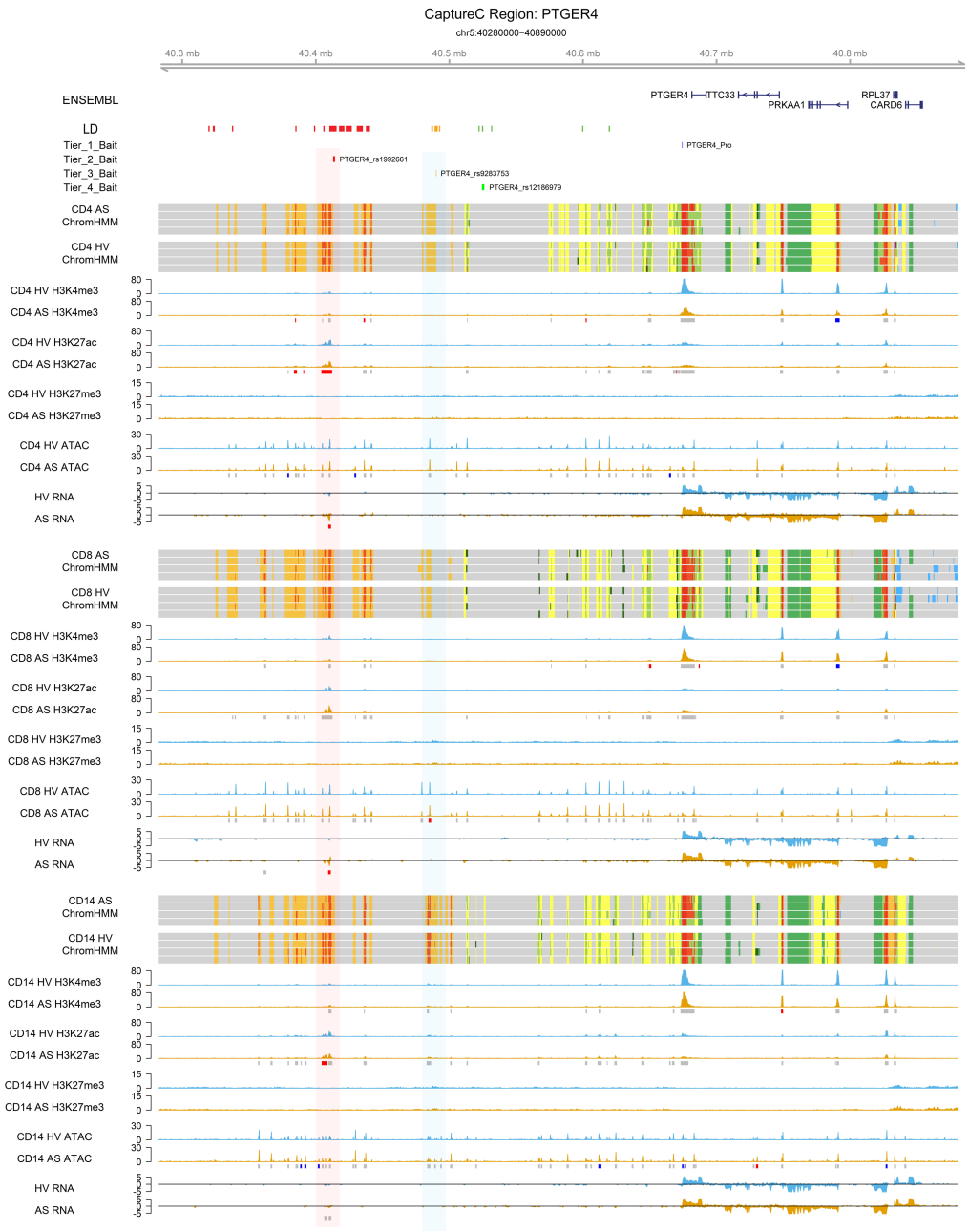

Figure 4.8. Differential epigenetic marks at the chr5p13.1 locus show the cis-regulatory elements containing lead SNP *rs1992661* is more active in CD4$^+$ T cells and CD14$^+$ monocytes and the functional SNP *rs9283753* is more accessible in CD8$^+$ T cells. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks for a single representative track from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' tracks show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red and blue shaded box shows the location of the *PTGER4* interacting lead or functional SNP-containing *cis*-regulatory regions.

and may contribute to altered *PTGER4* expression and disease pathogenesis.

4.3.4. The AS-associated SNP *rs2836883* shows context-specific looping onto the *ETS2* promoter in CD14$^+$ monocytes but not in CD4$^+$ or CD8$^+$ T cells

The *ETS2* gene within the chr21q22.2 AS-associated risk locus encodes for the Ets proto-oncogene 2 transcription factor. *ETS2* has been linked to a number of pleiotropic roles in the immune system and is involved in the suppression of pro-inflammatory cytokines such as TNFα, IL-6, and IFN-β (Ma *et al.*, 2019). *ETS2* is also found enriched at transcription factor binding motifs surrounding multiple chemokine genes, and is an important mediator of immune cell recruitment to inflammatory sites (Pitaressi *et al.*, 2016). *ETS2* also sees use as a biomarker for pro-inflammatory M1-polarised macrophages in patients presenting with membranous nephropathy; highlighting its importance in the monocyte-driven aspects of inflammatory conditions (Wan *et al.*, 2022). *ETS2* has also been shown to antagonistically compete with the related, pro-inflammatory transcription factor *ETS1* (Ets proto-oncogene 1) for the same sequence binding motifs to activate distinct pro and anti-inflammatory downstream signalling pathways (Ren *et al.*, 2023).

In the *ETS2* locus the lead SNP is *rs2836883* (p-value = 5.70 x 10^{-40}). I observed a strong Peaky interaction (Maximum -$\log_{10}$ FDR = 0.9) between *rs28636883* and the ETS2 promoter in CD14$^+$ monocytes that was characterised by active enhancer (U10 EnhActv) and flanking promoter (U2 TSSFlnkD and U3 TSSFlnkU) chromatin states (**Fig. 4.9: Red shaded box**). This interaction was not seen in either the CD4$^+$ or CD8$^+$ T cells. An additional interaction just at an active enhancer (U10 EnhActv) element shows strong Peaky interaction scores in AS patient and healthy volunteer CD14$^+$ monocytes (**Fig. 4.9: Blue shaded box**). This interaction was however much weaker in healthy volunteers

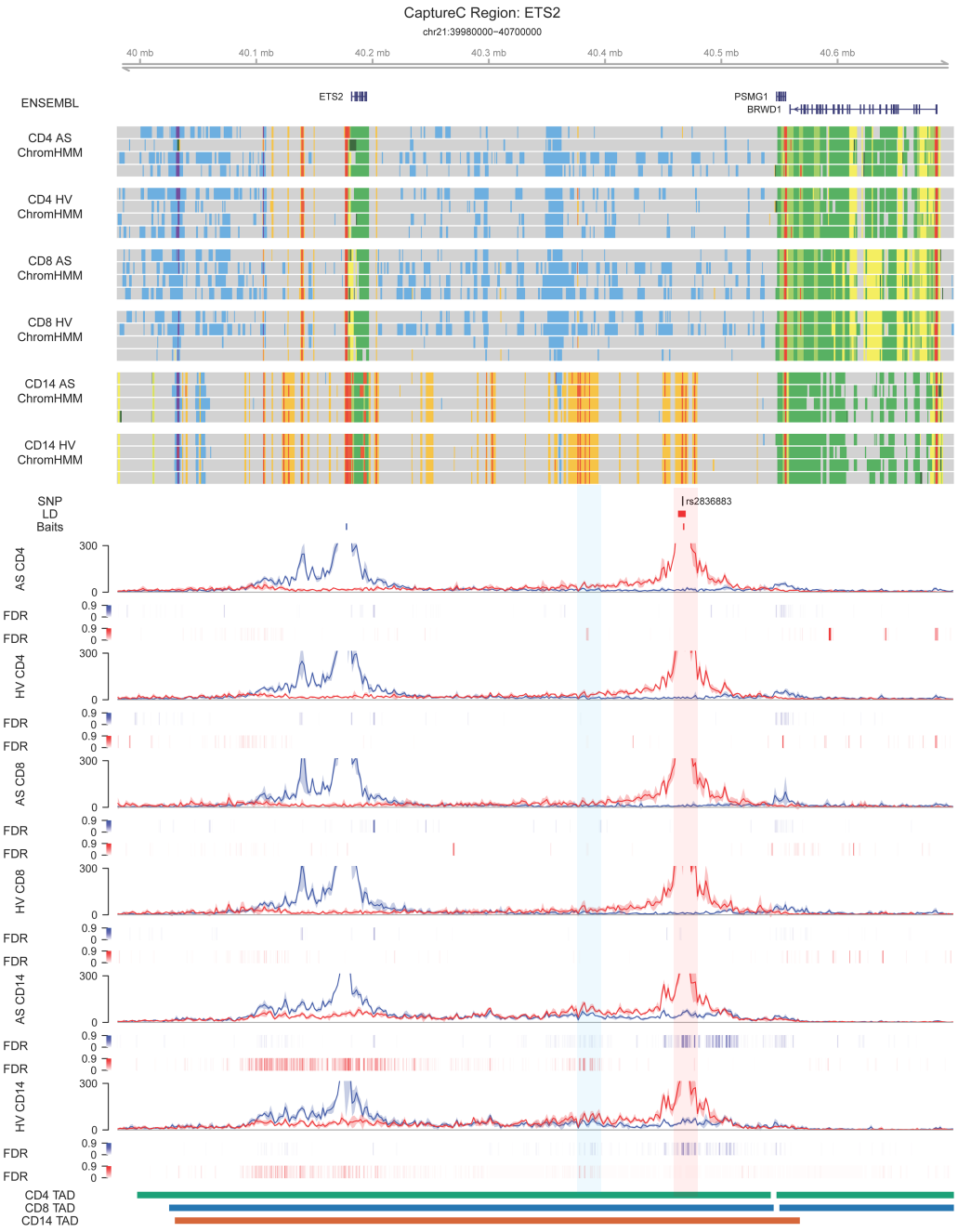

Figure 4.9. Enhancer-gene interactions between the promoter of *ETS2* and *rs2836883* at the chr21q22.2 locus are specific to CD14$^+$ monocytes . From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

(Maximum -$\log_{10}$ FDR = 0.9) than in AS patients (Maximum -$\log_{10}$ FDR = 0.7), suggesting enhancer-promoter contacts may be attenuated at the locus in AS.

At chr21q22.2, the epigenetic landscape of T cells is characterised by extensive Polycomb-associated repression (U16 PolyNC) up and downstream of *ETS2*, suggesting T cell-specific silencing of *cis*-regulatory elements in these cells (**Fig. 4.10**). Consistent with lower Peaky score in $CD14^+$ monocytes, two ATAC peaks each in the two promoter-interacting enhancers were weaker in AS patients compared to healthy volunteers and suggest these enhancers are less accessible to DNA-binding proteins in the disease state (**Fig. 4.10: Red and blue shaded boxes**). Overall, these results support a model where two active enhancer elements that adopt more closed chromatin conformation in AS – one containing the Ellinghaus *et al.* (2016) AS-associated lead SNP – loop onto the *ETS2* promoter in $CD14^+$ cells with reduced frequency in AS.

4.4. Complex and context-specific interaction profiles

I analysed Capture-C interactions at 44 loci. As shown above, 5 AS-risk associated loci (*ERN1, PTGER4, ETS2, IL7R, BACH2*) showed clear interaction with at least one lead SNP and a gene promoter. Of the four other loci, interactions were either highly complex and involved multiple cis-regulatory elements in close proximity to each other (*RUNX3),* looping of lead SNP-containing lead SNPs to the promoters of untargeted genes (*ITGA4*, *IL18RAP*), or looping of the targeted gene promoter to cis-regulatory elements outside of the lead SNP association signal (*ETS1*). The interaction profiles for the remaining 34 loci, such as *AHR*, did not show interactions for either the promoter bait or the AS-associated SNP bait. An example of each of these scenarios will be discussed in detail below.

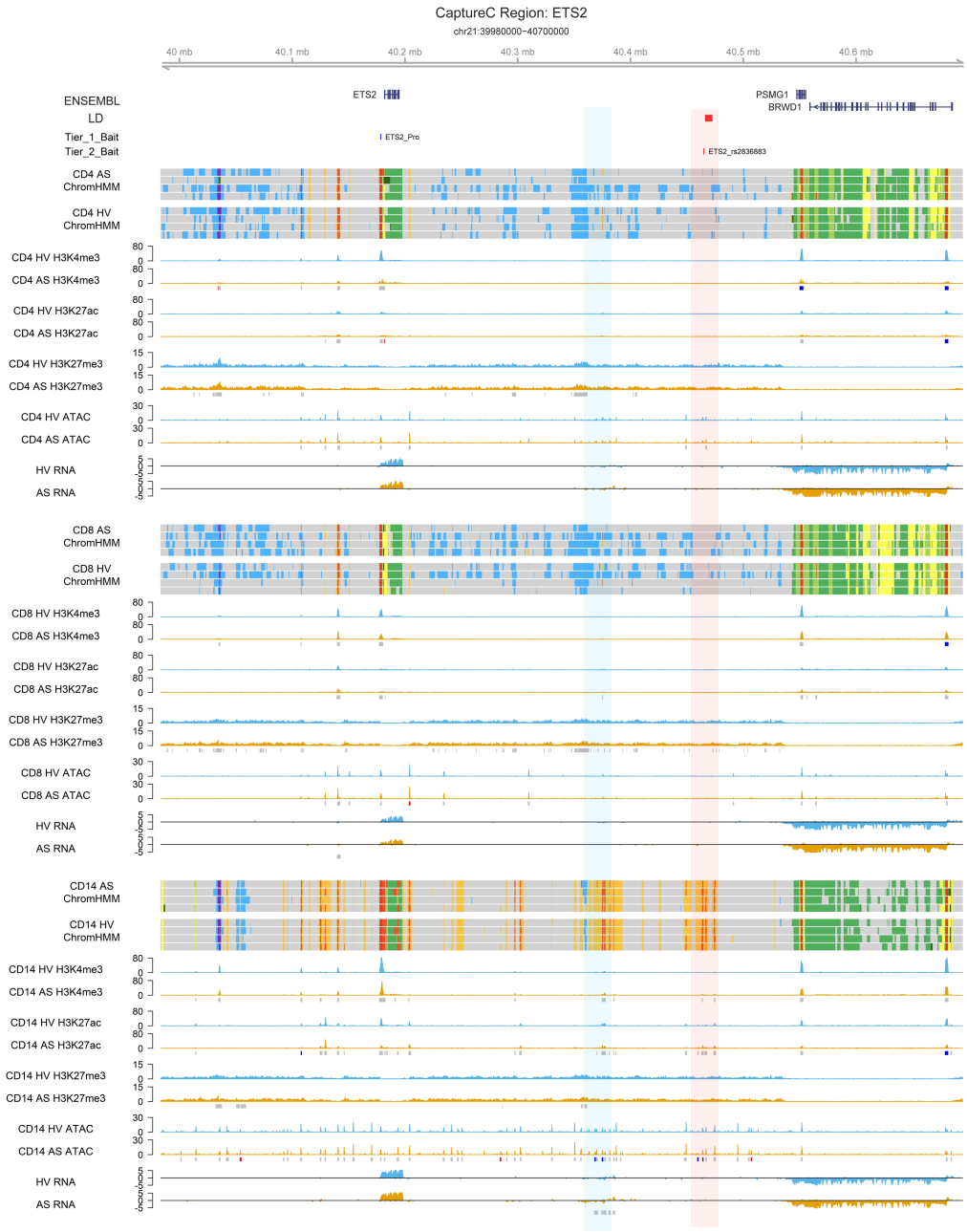

Figure 4.10. Differential epigenetic marks at the chr21q22.2 locus show the *ETS2*-interacting enhancers are less accessible in AS patient CD14$^+$ monocytes. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red and blue shaded box shows the location of the *ETS2* interacting lead SNP-containing (red) or other (blue) *cis*-regulatory regions.

4.4.1. The interactions in the chr1p36.11 locus shows super-enhancer-like structure between the promoter of *RUNX3* and distal regulatory elements in T cells and monocytes

The epigenetic environment surrounding *RUNX3* at the chr1p36.11 locus is among the most well characterised of all AS-associated GWAS risk loci (Cohen *et al.*, 2021) (Lead SNP for the locus is *rs6600247*, p-value = 2.6 x 10^{-15}; Cortes *et al.*, 2013). Previous work using a mixture of luciferase-based reporter assays, electrophoretic mobility shift assays (EMSA), ChIP, and qRT-PCR has described two functional SNPs with disease-specific allelic effects. *rs4265360* alter *RUNX3* expression in $CD14^+$ monocytes on LPS stimulation and *rs464889* regulated RUNX3 in $CD8^+$ T cells (Vecellio *et al.*, 2018; Vecellio *et al.*, 2021). The lead SNP may regulate *RUNX3* expression, with the AS-risk associated C allele perturbing a c-Myc transcription factor binding consensus sequence in $CD14^+$ monocytes that results in significantly lower binding of c-Myc; as measured by western blot-EMSA (Cohen *et al.*, 2021).

I found that upon performing Capture-C with baits at *rs6600247* and the RUNX3 promoter that the close spatial proximity of the lead SNP to the promoter obscured the significance of any interactions within a 1 Mbp window around the *RUNX3* promoter and made confirmation of reciprocal association difficult (**Fig 4.11**). A sharp interaction peak between the RUNX3 promoter and *rs6600247* was observed, however due to the noise of the Capture-C data these interactions are not called by Peaky (**Fig. 4.11: Red shaded box**). This was true for both the promoter and lead SNP viewpoint in all investigated cell types. Some differences between AS patients and healthy volunteers were seen at more distal regulatory elements (**Fig. 4.11: Blue shaded box**). ChromHMM shows that the lead SNP lies in an active enhancer (U10 EnhActv), and a strong interaction (Maximum Peaky -$\log_{10}$ FDR = 2.2) was observed between the lead AS SNP and a *cis*-regulatory

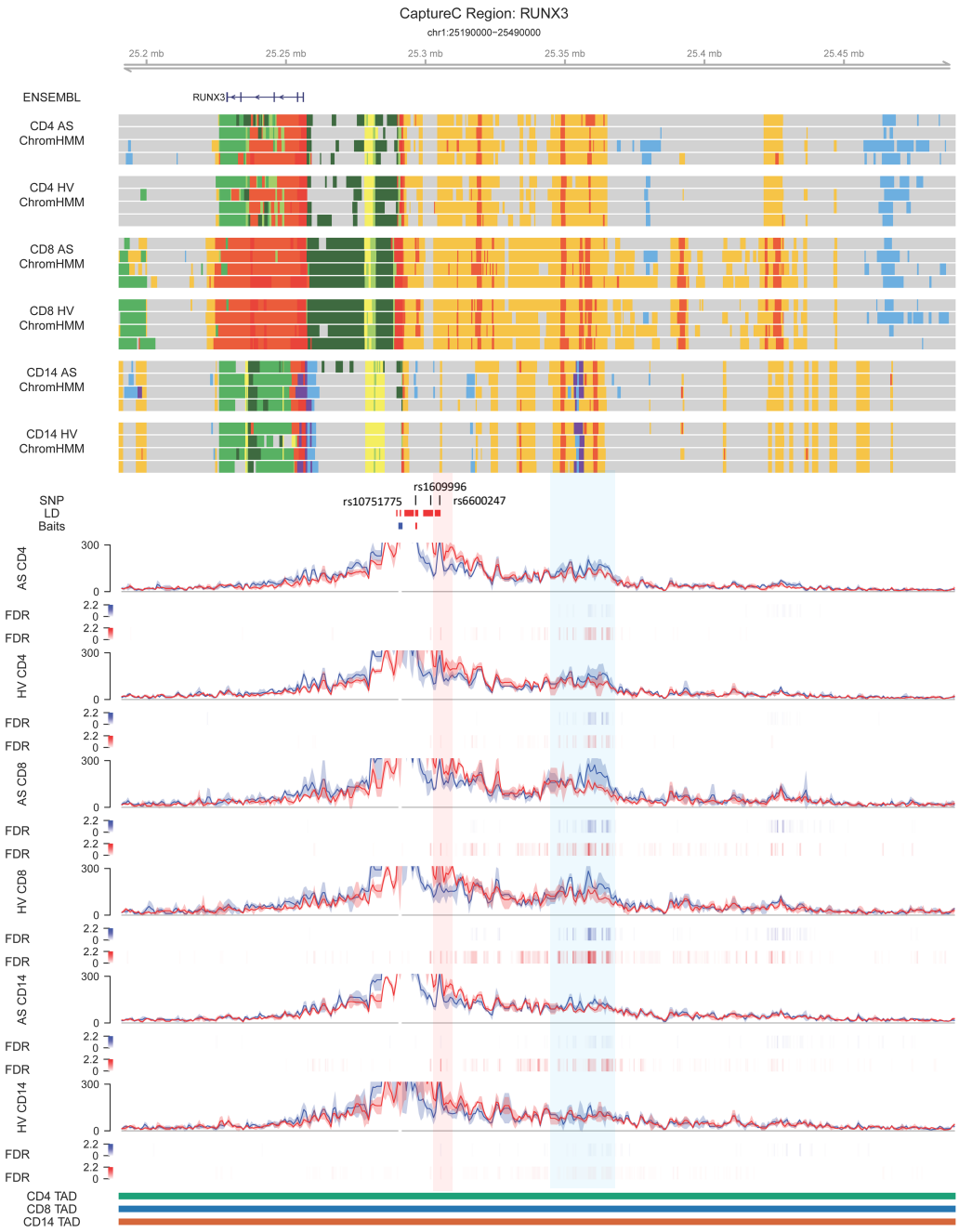

Figure 4.11. Complex super-enhancer-like structures form between multiple cis-regulatory elements and the *RUNX3* promoter at the chr1p36.11 locus. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

element (characterised by the U2 TSSFlnkU, U3 TSSFlnkD, and U10 EnhActv chromatin states) outside of the LD block in all cell types. The same region also interacts weakly with the *RUNX3* promoter (Maximum Peaky -$\log_{10}$ FDR = 1.3). In T cells, these interactions are present in both AS patients and healthy volunteers although appear much weaker in the healthy volunteers. This is seen most strongly in the $CD14^+$ monocytes of patients for the lead SNP viewpoint (Maximum Peaky -$\log_{10}$ FDR = 2.2) compared to those of healthy volunteers (Maximum Peaky -$\log_{10}$ FDR = 0.7) where the regulatory element is instead characterised as a bivalent promoter (U4 TSSBvlnt).

The distal interacting enhancer contains eRNA peaks that are much stronger in the $CD4^+$ T cells and $CD14^+$ monocytes of AS patients, suggesting greater activity from these regulatory elements in the disease state (**Fig. 4.12: Blue shaded box**). This was not seen in the $CD8^+$ T cells. These results support a multi-factorial regulation model where multiple *cis*-regulatory elements with putative differential activities in AS loop onto the *RUNX3* promoter in $CD4^+$ and $CD8^+$ T cells to control its expression.

4.4.2. The lead SNP *rs12615545* in the chr2q31.3 locus loops onto the promoter of *ITGA4* instead of the most proximal gene *UBE2E3* in $CD4^+$ and $CD8^+$ T cells and $CD14^+$ monocytes

At chr2q31.3, the lead SNP is *rs12615545* (p-value = 1.0 x 10^{-9}) (Cortes *et al.*, 2013). Due to its proximity to *rs12615545*, the promoter for *UBE2E3* (Ubiquitin-conjugating enzyme E2 L3) was baited. ChromHMM shows that the lead SNP lies in an active enhancer (U10 EnhActv) element, with the Capture-C data revealing that *rs12615545* loops onto the promoter of *ITGA4* (Integrin alpha 4) (Maximum Peaky -$\log_{10}$ FDR = 0.8) in all investigated cell types (**Fig. 4.13: Red shaded box**). No interactions were observed between the promoter viewpoint of *UBE2E3* and the *rs12615545*, suggesting *ITGA4*, not

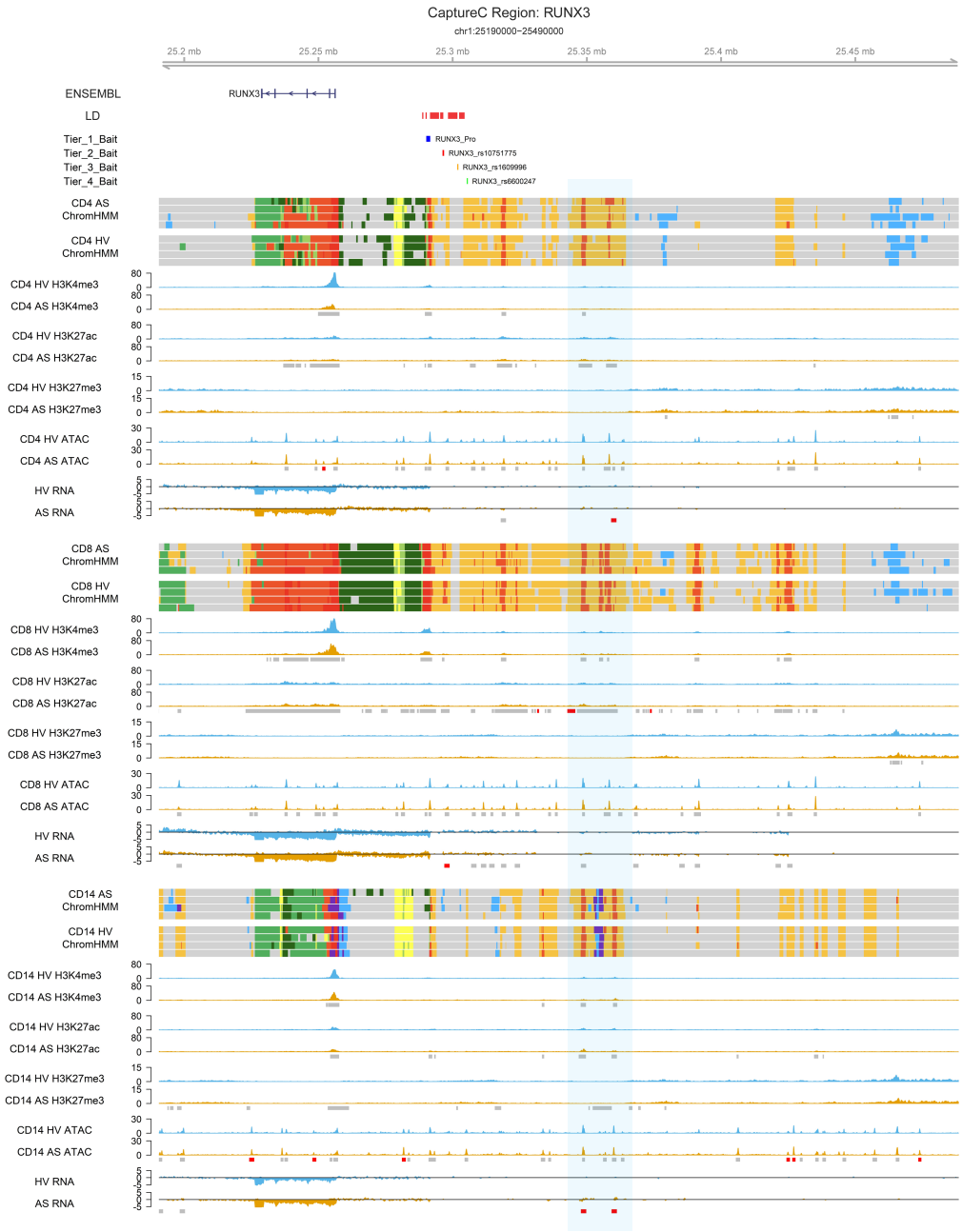

Figure 4.12. Differential epigenetic marks at the chr1p36.11 shows increased enhancer activity at a distal *RUNX3*-interacting enhancer across all cell types in AS patients. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks for a single representative track from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes in AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The blue shaded box shows the location of the distal *RUNX3* promoter-interacting *cis*-regulatory element.

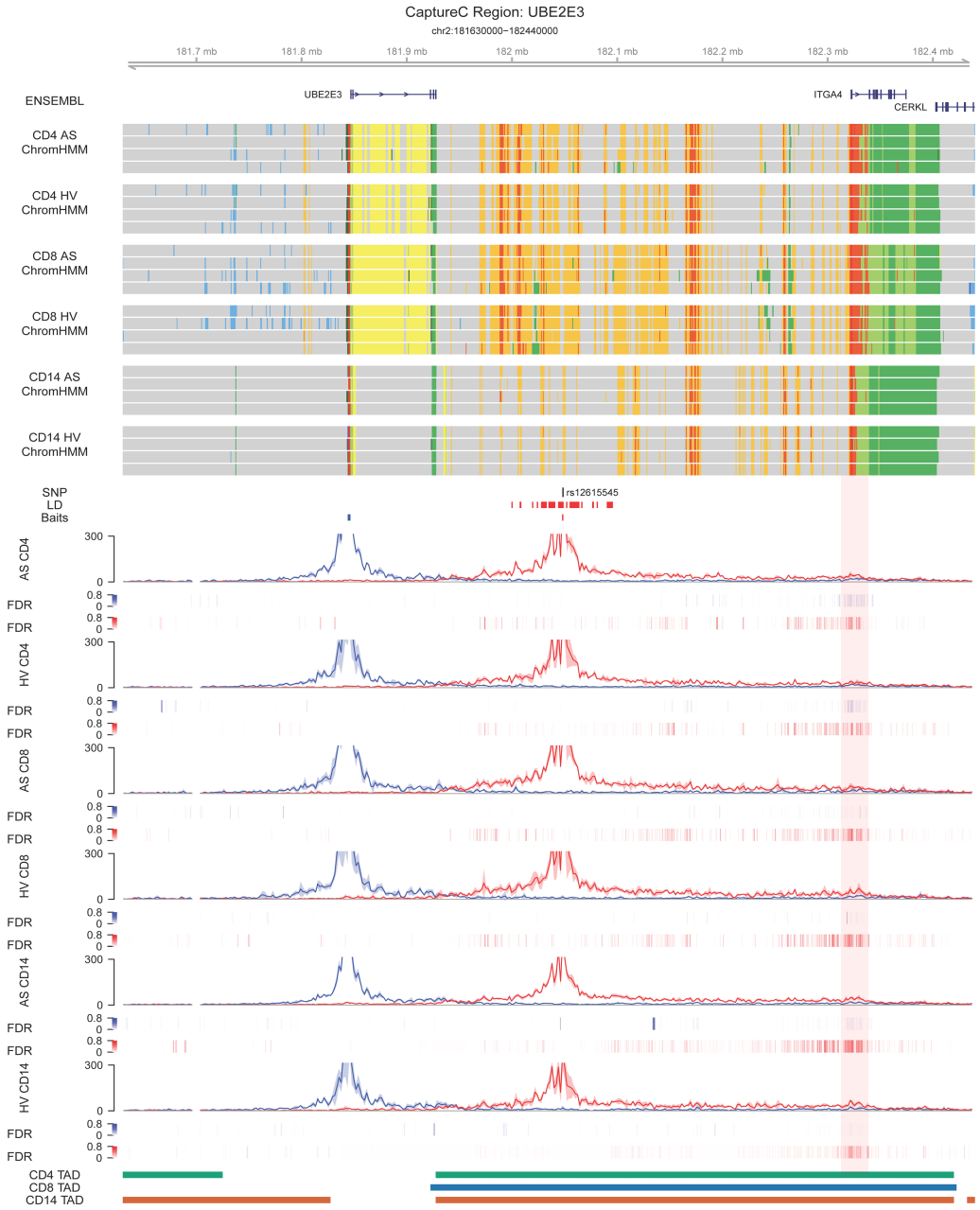

Figure 4.13. The lead SNP for the chr2q31.3 locus loops onto the *ITGA4* promoter in T cells and monocytes instead of the baited *UBE2E3* promoter. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

UBE2E3, is the gene contributing towards AS in these cell types. *ITGA4* encodes integrin alpha 4 which is involved in the trafficking of immune cells to sites of infection or inflammation, with knockout of *ITGA4* in murine MS models attenuating the migration of particular of pro-inflammatory Th1 cells into the spine and central nervous system (Glatigny *et al.*, 2015; Kendirli *et al.*, 2023). No differential marks were observed intersecting the lead SNP regulatory element in any cell type however, suggesting the activity of the underlying enhancer is similar in the disease and healthy states and that the lead SNP may not be the functional variant for the locus (**Fig. 4.14: Red shaded box**).

4.4.3. The promoter of *ETS1* loops onto an active enhancer element outside of the lead-SNP containing LD block in $CD4^+$ and $CD8^+$ T cells but not in monocytes

The *ETS1* gene at the chr11q24.3 AS-associated risk locus encodes for the ETS proto-oncogene 1 transcription factor which has broad regulatory roles in immune cells. In T cells, ETS1 expression is particularly associated with the production of immunosuppressive Tregs from naïve T cells post-antigen challenge; with knockdown of *ETS1* resulting in polarisation of naïve $CD4^+$ T cells towards Th2 and Th17 T helper cell subsets (Moisan *et al.*, 2007; Mouly *et al.*, 2010; Luchtel *et al.*, 2022). The ETS1 transcription factor is also a competitive binding-site antagonist for the ETV6 (translocation-Ets-leukemia virus) transcription factor in monocytes, which normally suppresses inflammatory cytokines and controls their differentiation to macrophages and dendritic cells within mouse models (Villar *et al.*, 2023).

A number of predicted active enhancers (U10 EnhActv) are present downstream of *ETS1,* one of which contains the AS-associated lead SNP *rs7933433* (p-value 1.3×10^{-10}) (Ellinghaus *et al.*, 2016) in $CD4^+$ and $CD8^+$ T cells, and $CD14^+$ monocytes (**Fig. 4.15**). Capture-C data shows a looping interaction from an enhancer downstream of *rs7933433*

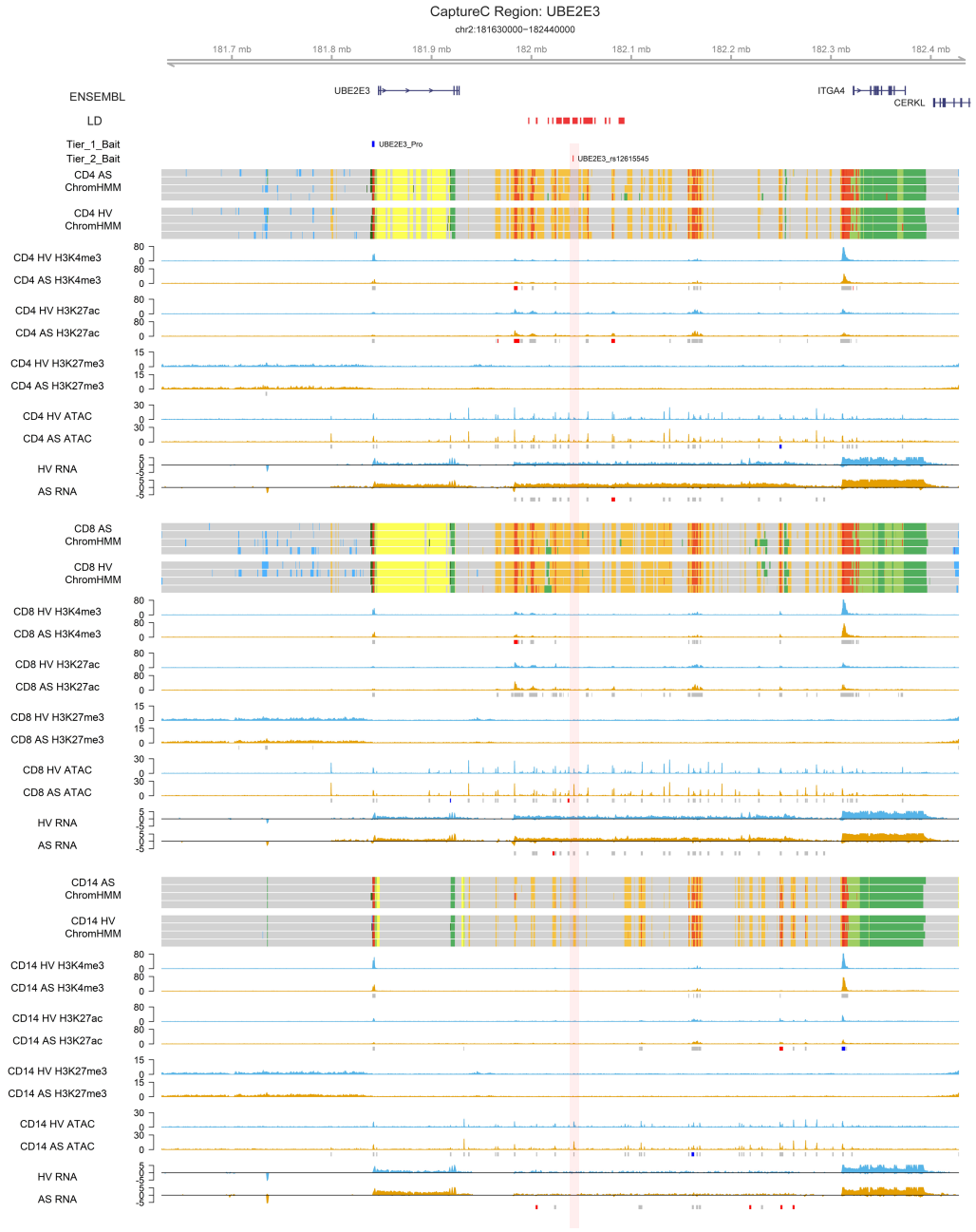

Figure 4.14. The epigenetic marks underlying the interaction between the *ITGA4* promoter and the *rs12615545*-containing enhancer element are similar between AS patients and healthy volunteers. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red shaded box shows the location of the distal *ITGA4* promoter-interacting *cis*-regulatory element.

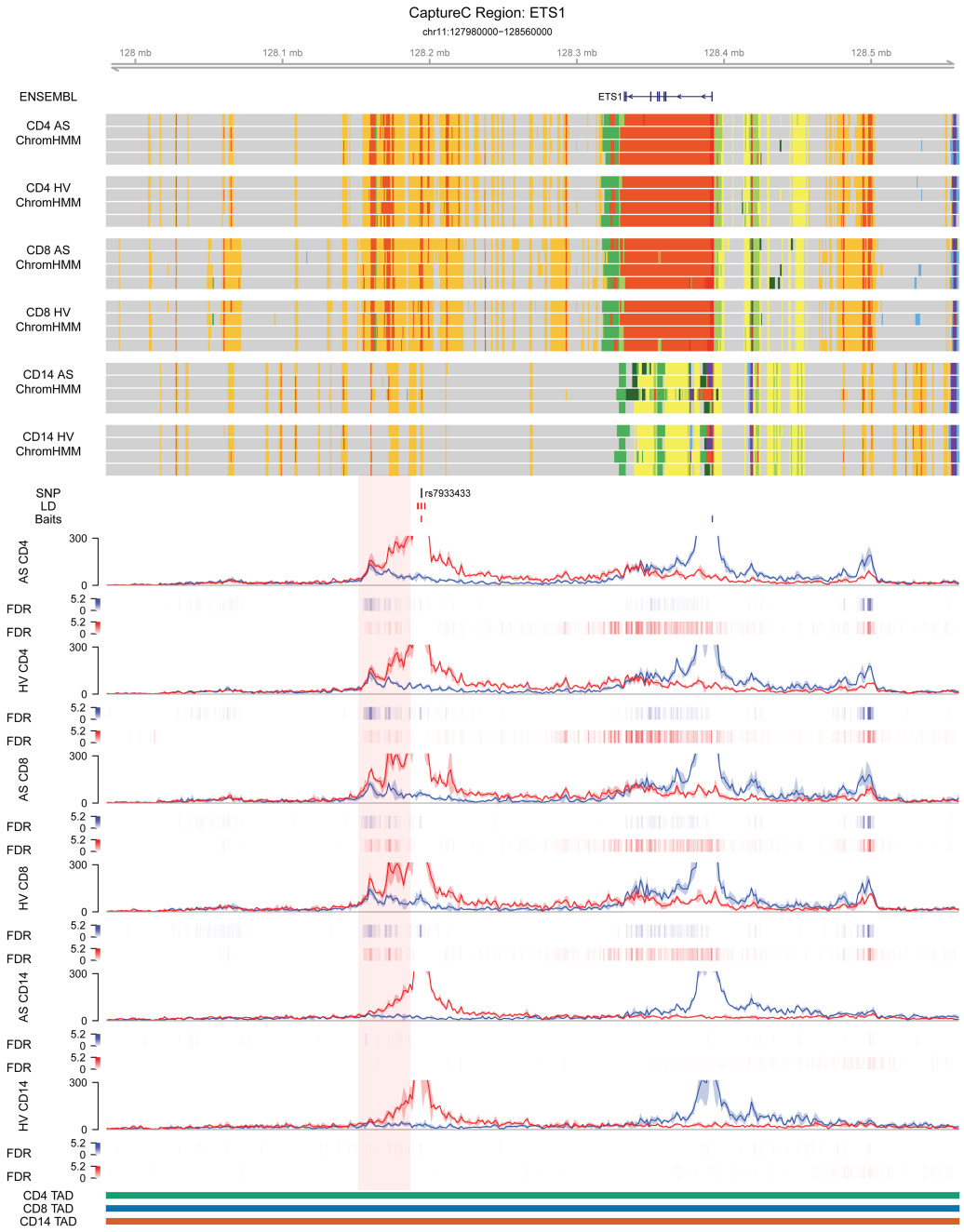

Figure 4.15. The promoter of *ETS1* interacts with the lead SNP *rs7933433* of the chr11q24.3 locus only in the CD4$^+$ and CD8$^+$ T cells of AS patients. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from CD4$^+$ T cells ('CD4'), CD8$^+$ T cells ('CD8'), and CD14$^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

in $CD4^+$ and $CD8^+$ T cells outside of the LD block containing *rs7933433* (Maximum Peaky -$\log_{10}$ FDR = 4.7) across all cell types in AS patients and healthy volunteers (**Fig. 4.15: Red shaded box**). Reciprocal interaction for this regions was seen in the interaction tracks of both the ETS1 promoter and *rs7933433*. In healthy volunteers, the promoter viewpoint showed very weak interaction with the lead SNP (Maximum Peaky -$\log_{10}$ FDR = 2.9). In AS patients, the interaction between the *ETS1* promoter and *rs7933433* was much stronger (Maximum Peaky -$\log_{10}$ FDR = 4.1).

None of these elements were found to interact in $CD14^+$ monocytes, suggesting these enhancer-promoter interactions were unique to T cells. In both $CD4^+$ and $CD8^+$ T cells, the lead SNP-containing regulatory element was associated with stronger eRNA and ATAC peaks in AS patients, suggesting increased enhancer activity and chromatin accessibility in the disease (**Fig. 4.16: Red shaded box**). These findings support a model where the lead SNP enhancer adopts a more open chromatin conformation, produces more loop stabilising-associated eRNA, and inserts into pre-existing contacts between the *ETS1* promoter and a distal active enhancer element in AS to potentially influence gene expression.

4.4.4. The 7p21.1 locus shows a lack of interaction between the lead SNP and any gene promoters in the region

The 7p21.1 locus is a gene desert that contains the gene for the aryl hydrocarbon receptor (*AHR*). *AHR* is a ligand-activated transcription factor that can help regulate differentiation responses in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes (Dean e*t al.*, 2023). The genome-wide significant lead SNP for the locus is *rs1525735* (p-value = 4.6×10^{-10}) (Ellinghaus *et al.*, 2016). ChromHMM shows that the lead SNP lies in a weak enhancer element (U11Wk) in all investigated cell types, and Capture-C reveals a lack of

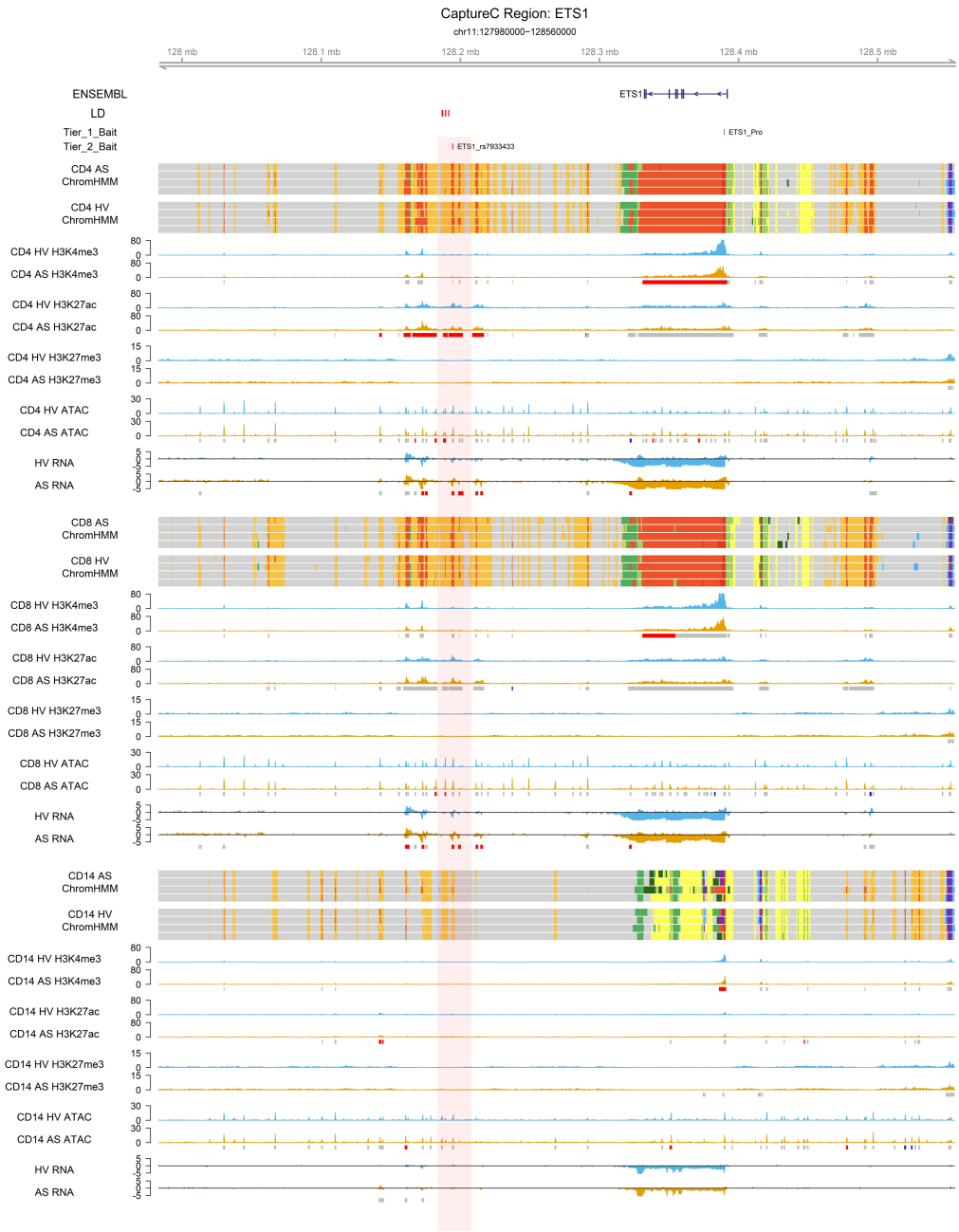

Figure 4.16. Differential eRNA and H3K27ac marks at the *rs7933433*-containing regulatory element suggests greater enhancer activity at the chr11q24.3 locus in AS patients. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the epigenetic datasets in AS patients (AS054) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The blue shaded box shows the location of the distal *ITGA4* promoter-interacting *cis*-regulatory element.

reciprocal interaction between the *rs1525735* baited restriction fragment and the *AHR* promoter bait in any of the analysed cell types (**Fig. 4.17**). Together, these results suggest that *rs1525735* is not associated with *AHR* in $CD4^+$ T cells, $CD8^+$ T cells, or $CD14^+$ monocytes, and may not be the functional causative AS-associated SNP for the locus in these cell types. Of the 44 analysed AS-associated GWAS loci, 35 of them showed similar lack of reciprocal interaction to chr7p21.1.

4.5. Chapter discussion

3C-based methodologies are invaluable tools in the field of functional genomics. Over the last 20 years, 3C techniques have been widely used to identify and estimate interaction frequency at multiple genomic loci (Dekker *et al.*, 2002; Lieberman-Aiden *et al.*, 2009; Hansen *et al.*, 2018). 3C approaches also allow for the elucidation of how GWAS genetic variants and different genomic elements can physically interact with gene promoters to directly influence gene expression, which may differ between AS patients and healthy volunteers. In this chapter, I applied reciprocal Capture-C of gene promoters and AS-associated lead SNPs to primary $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes of AS patients and healthy volunteers. Through the combination of Capture-C techniques with ChromHMM I was able to map cell-specific interaction profiles for 53 independent AS-associated genetic variants and 44 gene promoters in AS patients and healthy volunteers.

4.5.1. A subset of the AS-associated GWAS risk loci show strong looping interactions between ChromHMM-predicted *cis*-regulatory regions

The first and second aims of this chapter were jointly to characterise how *cis*-regulatory elements at AS-associated GWAS risk loci interact with target genes to potentially predispose and individual towards developing AS. To ensure that the Capture-C data was

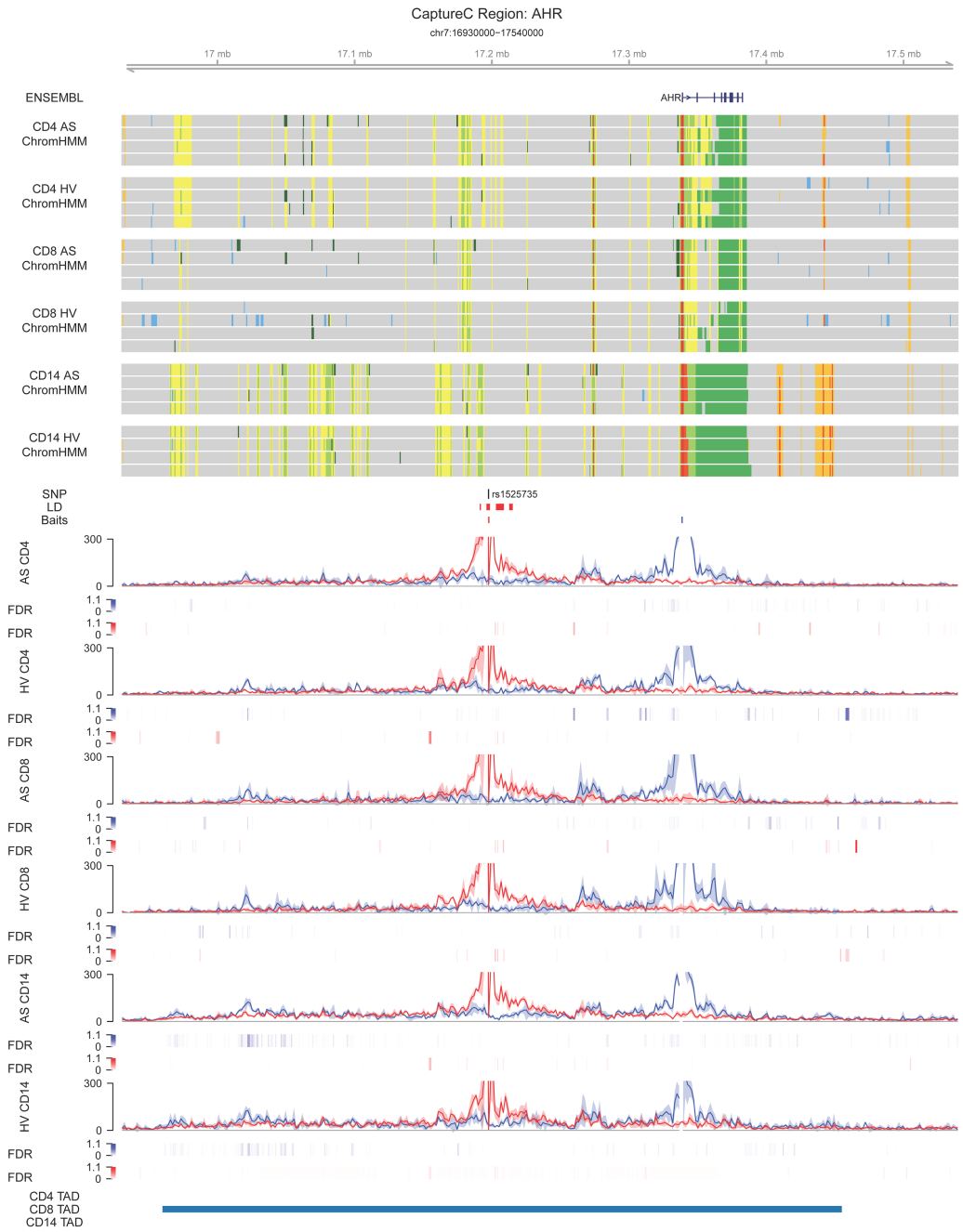

Figure 4.17. No reciprocal interactions are observed between the *AHR* promoter and the AS lead SNP *rs1525735* at the chr7p21.1 locus in any of the investigated cell types. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

of high enough quality to investigate these aims, I first analysed how previously validated interactions at the *LMO2*, *MYC*, and *INPP4B* loci formed in the capture dataset. The Capture-C data managed to recapitulate these interactions at all three loci (*LMO2* in $CD14^+$ monocytes and *MYC* and *INPP4B* in $CD4^+$ and $CD8^+$ T cells). Intersection with ChromHMM also showed that these interactions occurred between active enhancer and promoter chromatin states. Where interactions were not observed, these regions were instead characterised by Polycomb-repressed or quiescent chromatin states. I therefore concluded that not only is the Capture-C data of high quality, but also can be integrated alongside ChromHMM to explain the biological reasoning behind why two of the DpnII restriction fragments are interacting.

It should be noted that for the *MYC* and *INPP4B* loci that the interactions lay outside of the maximum Peaky interaction score window specified in this investigation (≤ 500 kbp) (Brown *et al.*, 2023). This threshold was chosen because the majority of enhancer-promoter interactions occur below this threshold, with recent studies by Chen *et al.* (2024) finding that the median distance these contacts form over in ten embryonic murine tissues to be approximately 410 kbp using Hi-C. Despite the lack of Peaky score, *INPP4B* and *MYC* showed clear, high mean interaction counts, suggesting these interactions were successfully captured in the experiment. These interactions were known to occur at these loci however, and in the AS-associated loci, where interaction profiles are lesser known, analysis of interactions with the promoter and lead SNP viewpoints were kept within this 500 kbp window.

Following successful characterisation of the control loci, of the 44 AS-associated regions analysed, 9 of them showed disease-relevant interactions with high Peaky interaction scores between active ChromHMM characterised enhancers (U10 EnhActv) and promoters (U1 TSSActv, U2 TSSFlnkD, and U3 TSSFlnkU); 5 of these loci showed

reciprocal interaction between a lead SNP-containing enhancer and the closest gene promoter (*ERN1*, *PTEGR4*, *ETS2*, *BACH2*, *IL7R*), 2 showed interaction from a lead SNP-containing enhancer to an untargeted promoter element (*IL18RAP*, *ITGA4*), 1 showed interaction with an enhancer element outside of the lead SNP LD block (*ETS1*), and 1 showed complex interaction between multiple cis-regulatory elements in close proximity to each other (*RUNX3*). The genes identified here have potential relevance in AS pathogenesis, with direct involvement in promoting immune system or inflammatory pathways (**Table 4.2**). I was therefore partially successful in my first two chapter aims of identifying which of the lead SNPs lied in *cis*-regulatory chromatin states and of identifying which genes they interacted with in their respective loci.

4.5.2. The majority of loci showed no interactions for either the promoter viewpoint or the lead SNP viewpoint

Despite the success in identifying these nine loci that contain *cis*-regulatory elements intersecting AS-lead SNPs, I observed a lack of Capture-C interaction signal at 35 of the AS-associated loci in any cell type within AS patients and healthy volunteers. This may be due context-specificity (i.e. where there may be interactions in other cell types or stimulatory states) or baiting for the wrong gene and/or SNP. A good example of this is the *AHR* locus. Despite the role of *AHR* in immune signalling in the T cell and monocyte populations and maintenance of IL-17 signalling (Sondermann *et al.*, 2023), no interactions were found between genetic variant and gene promoter. It should be noted that high expression levels of *AHR* are only seen in primary $CD4^+$ T cells upon immunosuppressive TGFβ1 (transforming growth factor beta 1) signalling (de Lima *et al.*, 2018). Whilst TGFβ1 can derive from a multitude of cellular sources, expression of TGFβ1 is significantly increased in monocyte-derived macrophages on stimulation with LPS (Sun *et al.*, 2018) and in T cells upon TCR stimulation (Saadey *et al.*, 2023).

Locus[1]	Gene[2]	Full Gene Name[3]	Gene Function[4]	AS Lead SNP[5]	Gene-SNP Distance[6]	Other Evidence[7]	Putative Mechanism[8]
chr1p36.11	*RUNX3*	Runt-related transcription factor 3	Transcription factor important in T cell differentiation	*rs6600247* [a,b]	13612	Functional evidence for disrupted transcription factor binding[1]	SNPs alter TF binding, localised chromatin looping of enhancers
chr2q11.2	*IL18RAP*	Interleukin-18 receptor accessory protein	Constituent peptide of receptor for pro-inflammatory IL-18	*rs4851529* [a,b]	387952	-	Long-range enhancer interaction
chr21q31.3	*ITGA4*	Integrin alpha 4	Involved in motility and migration of immune cells	*rs12615545* [a]	273165	LncRNA in $CD4^+$ and $CD8^+$T cells[2]	Long-range enhancer interaction
chr5p13.2	*IL7R*	Interleukin-7 receptor	Constituent peptide of receptor for pro-inflammatory IL-7	*rs11742270* [a]	1738	Regulatory SNP and splice-variant[3] and differential eRNA	Two signals: Regulatory enhancer SNP and IL-7R splice variant
chr5p13.1	*PTGER4*	Prostaglandin E receptor 4	Prostoglandin GPCR involved in IL-23 and TNFα pathways	*rs12186979* [a], *rs1992661b* [b]	155171	Functional SNP[4], and differential eRNA	Long-range enhancer interaction
chr6q15	*BACH2*	BTB domain and CNC homolog 2	Transcription factor important in T cell regulation	*rs17765610* [a]	0	-	Long-range enhancer interaction from within the gene
chr11q24.3	*ETS1*	Ets prot-oncogene 1	Transcription factor that regulates cytokine and chemokine production	*rs7933433* [b]	134204	Differential eRNA	Long-range enhancer interaction
chr17q23.3	*ERN1*	Endoplasmic reticulum to nucleus signalling 1	Unfolded protein response	*rs196941* [b]	0	-	Intronic enhancer interaction
chr21q22.2	*ETS2*	Ets proto-oncogene 2	Transcription factor involved in T and B cell regulation	*rs2836883* [a], *rs9977672* [b]	269866	Differential eRNA	Long-range enhancer interaction

Table 4.2. Evidence of enhancer-gene interactions at the nine identified AS-associated GWAS risk loci. 'Locus[1]' shows each of the AS-associated loci that showed interaction between lead SNPs, *cis*-regulatory elements, and/or targeted genes in the Capture-C analysis. 'Gene[2]' shows the name of the gene in the locus with evidence of looping interactions with other genetic elements. 'Full Gene Name[3]' shows the full name of each interacting gene. 'Gene Function[4]' gives a short description of the biology behind each of the genes and how they may apply to AS. 'AS Lead SNP[5]' shows the Cortes *et al.* (2013) and/or Ellinghaus *et al.* (2016) lead SNP for each locus. 'Gene-SNP Distance[6]' shows the distance between the interaction genetic element and the promoter of each identified gene. 'Other Evidence[7]' shows brief findings from this thesis and other studies to suggest further functional evidence for AS contributions. In this column: [1] = Cohen *et al.*, (2021); [2] = Castellanos-Rubio *et al.* (2016); [3] = Al-Mossawi *et al.* (2019). [4] = Tewhey *et al.* (2016). 'Putative Mechanism[8]' speculates the putative biological role of each interaction based on the intersection of Capture-C and ChromHMM data in this chapter.

It could therefore be that interactions between AS-associated SNPs and target promoters may only be observed in the right immune signalling context (such as *AHR*/*TGFβ1*). Since AS is characterised by immune dysregulation, identification of chromosome looping in AS patient pre and post-therapeutic intervention could identify regulatory regions that respond to immune or inflammatory signals. This is important as enhancers that control inflammation-associated genes are amongst the first genomic elements to respond to inflammatory stimuli. Heward *et al.* (2015) for instance, showed that stimulation with LPS significantly and robustly increased eRNA levels in the monocyte-like THP-1 cell line. This was mediated by the NFκβ and MAPK pathways, with certain NKκβ genes such as the NFκβ subunit p105 (*NFKB1*) and the NFκβ-inhibitor alpha (*NFKBIA*) binding protein showing AS GWAS association (Ellingaus *et al.*, 2016); suggesting these pathways may be up or downregulated in AS patients and healthy controls.

Indeed, many components of the NFκβ pathway have been identified as potential therapeutic targets of methazolamide treatment in AS patients (Sun *et al.*, 2023). These immune stimulus-induced eRNA are involved in the formation and stabilisation of looping interactions at inflammation-related enhancer elements, suggesting the inflammatory or immune stimulus is required prior to organisation of genomic elements at the target promoter and gene expression (Wan *et al.*, 2022). Some immune cell types, including the $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes presented in this chapter, may therefore demonstrate significant differences in their interaction profile upon immune stimulus, and this should be considered in future chromosome conformation capture experiments.

4.5.3. Other cell types and certain T cell and monocyte subtypes may show different interaction profiles between AS-associated lead SNPs and their target genes

Additionally, although I have shown in this chapter how these loci may operate in primary naïve $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes, the complex contributions of other cell types in AS – such as neutrophils, NK cells, and subtypes of each of those cell types analysed here (like Th17 cells) – implicate potential functional roles for SNPs in those cell types. Like T cells and monocytes, this functional role may too be attenuated or amplified by pro-inflammatory signalling. Additionally, because the main sites of inflammation in AS are at the sacroiliac joints and the spinal entheses, the cell types isolated from PBMCs in this chapter may not properly reflect the changes seen in cells isolated from AS patient synovial or spinal bone biopsies.

There is some evidence to suggest that the epigenetic landscape of tissue-resident cells can significantly differ to that of those residing in the peripheral blood following tissue infiltration. Comparison of tissue-resident $CD8^+$ T cells to those in the peripheral blood using bulk RNA-seq and ATAC-seq for instance, shows that the tissue-resident cells across a diverse range of tissues (such as the kidney, liver, salivary glands, adipose tissue, and spleen) produce global transcriptomic and open chromatin signatures more similar to each other than to circulating $CD8^+$ T cells (Crowl *et al.*, 2022). This suggests that there may be significant differences in which SNPs and regulatory elements interact with a given promoter element in the cells contributing to AS pathogenesis at the site of inflammation when compared to the cells analysed in this chapter. It could therefore be the case that the right gene promoters and lead SNPs were baited in this approach, but have been investigated in cell types where neither enhancer or gene are active.

Because this analysis was also performed in bulk purified $CD4^+$, $CD8^+$, or $CD14^+$ cell populations, these samples reflect a broad array of pro and anti-inflammatory subtypes in the peripheral blood that have antagonistic functions, such as immunosuppressive Tregs and pro-inflammatory Th17 cells. The Capture-C and Peaky data presented in this chapter is the mean interaction profile for all of these distinct subtypes, and will mostly represent the interactions seen in the most populous of the subtypes. Rarer subtypes with known disease-contributions may have their interactions diluted in these samples. In AS patients, where Th17 cells are thought to have a prominent role in AS pathogenesis, Th17 cells constitute less than 2.5% of all peripheral blood $CD4^+$ T cells (Shen *et al.*, 2009). If there are pathogenic interactions that occur in Th17 cells that are not seen in the majority of the bulk $CD4^+$ purified T cell population, then they would not be properly captured in the approach used in this chapter. In addition to including more stimulation conditions, future 3C-based experiments should incorporate a more diverse number of cell types and their subtypes reflective of the complexity of their contributions to AS pathogenesis.

4.5.4. The interaction of *cis*-regulatory elements with the *RUNX3* promoter were complex and resembled super-enhancer-like structures

Despite multiple promising initial insights into how AS-associated lead-SNPs interacted with several disease-linked genes, this chapter has also highlighted the complexity of the 3D genomic landscape surrounding a number of the loci. This is particularly pronounced at regions such as *RUNX3*, where the close spatial proximity of both the lead-SNP and SNPs within the same LD block are obscured due to localised Brownian DNA motion, which derives from the randomised movement of molecules suspended in fluid that may bring together two segments of DNA in close proximity by chance alone. The closer these segments are together, the higher the chance that these segments will be ligated in 3C-based approaches like Capture-C.

Whilst analytic packages like Peaky attempt to account for Brownian motion in their algorithms by modelling distance decay as a function, the usual result is that interaction peaks closer to the promoter viewpoint are attenuated whilst those more distal are amplified (Eisjbouts *et al.*, 2019). This will result in potentially important interactions peaks, like what was seen between the *RUNX3* promoter and the lead SNP *rs6600273*, being called as insignificant by Peaky. Investigation of *RUNX3* and other loci that demonstrate lead SNPs in close proximity to capture viewpoints using the approach described in this chapter is therefore difficult, and will require investigation with other techniques such as Micro Capture-C (MCC) (Downes *et al.*, 2022). Instead of a 6 or 4 bp-cutter, MCC uses a micrococcal nuclease to fragment the genome. Micrococcal nuclease does not recognise any one specific cut site in the genome and cuts indiscriminately, resulting in the digestion of all DNA not protected by formaldehyde-fixed protein complexes. The interaction profile produced by MCC is subnucleosomal, which allows potential experimental resolution on the base-pair scale, and would better allow for the resolution of interactions between the *RUNX3* promoter and its functional and/or lead SNPs.

4.5.5. Some regions may show differential interaction between AS patients and healthy volunteers

The third aim of this chapter was to investigate whether any of the observed interactions were differential between AS patients and healthy volunteers. Some of the loci appeared to show differential interactions between the two states. At chr11q24.3, the Peaky interaction scores between the *ETS1* promoter and the lead SNP *rs7933433* were far higher in AS patient $CD4^+$ and $CD8^+$ T cells compared to healthy volunteers. At chr21q22.2, the reciprocal interaction between the lead SNP *rs2836883* and the *ETS2* promoter appeared much weaker in healthy volunteer $CD14^+$ monocytes compared to AS

patient cells. The Peaky values suggest that the active enhancer (U10 EnhActv) elements both of these SNPs lie in contact the enhancer with increased frequency in AS patients, and also suggests greater control of gene expression in their associated cell types. The lead SNP-containing regulatory elements were also intersected with differential chromatin marks to better inform putative pathogenic contributions. Four of the loci (*ERN1*, *PTGER4*, *ETS2, IL7R*) intersected differential H3K27ac marks between AS patients and healthy volunteers, three intersected differential ATAC peaks (*PTGER4*, *ETS1*, *ETS2*), and two intersected differential eRNA (*ETS1*, *PTGER4*).

It was however difficult to deduce whether differential interaction existed for each of the loci with Peaky alone however, as mean interaction counts tended to show high standard deviation. This suggested a large degree of variation exists in loci interaction profiles between in the three healthy volunteers and three AS patients. This also has an effect on the Peaky scores, making them appear weaker when standard deviation is higher (Eisjbouts *et al.*, 2019). As such, conclusions that are drawn from the Peaky score may not truly reflect differences in AS biology and the reciprocal approach described in this chapter cannot be used alone to deduce which enhancers may contribute to AS pathogenesis. Additionally, as we recruited only three individuals for condition, the effects of genotype on interaction frequency cannot be underestimated (McGovern *et al.*, 2016). In juvenile idiopathic arthritis for example, the CC genotype of the SNP *rs9979383* is associated with significant increases in interaction frequency with the *RUNX1* promoter over the AA genotype (Taylor *et al.*, 2017). Because we did not recruit based on genotype, as it is infeasible to recruit based on the identify of every SNP in a credible for 53 AS-associated lead SNPs, and because these sample cohorts are so small, the variation seen in the mean interaction counts may arise from SNPs that genuinely effect interaction frequency that differ in at least one of the patients or healthy volunteers,

but not all. Now that interaction profiles have been generally defined at these loci for $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes, future Capture-C work should focus on increasing the number of individuals in each cohort to increase the power of Peaky analysis and give greater confidence that these interactions are truly statistically differential between AS patients and healthy volunteers.

4.5.6. Tiled-C can be used to investigate the interactions between differential epigenetic marks at each locus not captured in the Capture-C experiments

It should be noted that discussion of intersection with differential epigenetic marks was limited to only those genetic elements that contained AS lead SNPs. At every locus at least one of differential open chromatin, eRNA, H3K4me3, or H3K27ac marks were seen outside of the lead SNP-containing *cis*-regulatory element. This suggests that a proportion of the epigenetic differences at AS-associated loci reside outside of the lead SNP signal. Whilst these marks could be investigated individually with Capture-C or MCC, Tiled-C offers a better alternative approach to characterising which genes these differential marks impact the expression of. Tiled-C is a some-vs-all approach that involves baiting every restriction fragment between two given genomic coordinates with biotinylated oligonucleotides, with capture of all fragments performed in a single reaction (Downes *et al.*, 2022). This allows for construction of contact matrices, where the interaction pattern for each individual viewpoint can be analysed in consideration of every other restriction fragment. Because of the sequencing and oligonucleotide generation costs involved in Tiled-C, the use of the technique is prohibitive in diseases with a large number of disease-associated loci as in AS which demonstrated greater than 100 known disease-associated signals. This chapter however narrowed down 44 of the AS-associated loci to nine candidates that showed potential differential interactions between AS patients and healthy volunteers in $CD4^+$ and $CD8^+$ T cells and $CD14^+$

monocytes. In the next chapter, Tiled-C will be performed on these loci to better characterise how regulatory elements within them could contribute to AS pathogenesis.

4.5.7. Concluding remarks

The use of ChromHMM to identify putative causative elements is limited due to its lack of information on interactions between predicted *cis*-regulatory elements. In this chapter, I have described the generation of Capture-C data for overlap with ChromHMM to identify interactions that are differential themselves or overlap differential epigenomic features. Nine AS-associated loci showed any interaction between a lead SNP-containing regulatory element and/or the target gene promoter of the 44 total baited loci. The majority of loci did not demonstrate interaction between the lead SNP and any genomic element, suggesting potential context-dependency outside of T cells and monocytes. Intersection of stronger Peaky interaction scores and differential open chromatin, eRNA, and histone post-translational modification peaks show that these interactions occur between regulatory elements that appear more active in AS patients compared to healthy volunteers. The majority of differential peaks did not show interaction with either the lead SNP viewpoint or targeted promoter viewpoint. These nine loci will be taken forward for complete profiling of TAD-based interactions with Tiled-C in chapter five.

"Man has gone out to explore other worlds and other civilizations without having explored his own labyrinth of dark passages and secret chambers, and without finding what lies behind doorways that he himself has sealed."

- Stanisław Lem, Solaris (1961)

5

APPLICATION OF TILED-C IN CD4$^+$ AND CD8$^+$ T CELLS TO INTERROGATE CHROMATIN CONTACTS IN AS-ASSOCIATED GWAS LOCI

Contents **Pg.**

5.1. Chapter introduction 220

5.2. Tiled-C at control loci identifies known interactions 230

5.3. Genome-wide epigenetic changes in untreated, treated, and stimulated cells 243

5.4. Example Tiled-C regions 249

5.5. Chapter discussion 267

In this chapter, chromatin interaction data from Tiled-C is intersected with epigenetic data for CD14+ monocytes, CD4$^+$ T helper cells and CD8$^+$ cytotoxic T cells from healthy volunteers before and after stimulation, and from AS patients with active disease pre and post-anti-TNFα therapy, to characterise the disease-relevant responses of regulatory elements associated with immune or inflammatory stimuli at AS GWAS risk loci.

5.1. Chapter introduction

5.1.1. Tiled-C can be used to comprehensively map chromosome looping in TADs

In chapter four, I used Capture-C to identify looping interactions between *cis*-regulatory elements containing AS-associated lead SNPs and cognate genes. However, other notable

interactions likely exist, such as those that initiate from disease-specific regulatory regions. Since most looping interactions and SNP-driven gene expression effects are often limited to TADs (Way *et al.*, 2017; Bolt *et al.*, 2022), alternative approaches like Tiled-C may be used to analyse all looping interactions within the TADs containing the AS GWAS association signals. Tiled-C enables detection of all interactions across specific TADs at higher resolution and lower cost than other methods such as Hi-C (Oudelaar *et al.*, 2020).

The Tiled-C methodology is similar to Capture-C, except where Capture-C baits a single restriction fragment of interest, Tiled-C baits every restriction fragment between two genomic coordinates. Tiled-C has been used to map enhancer-promoter contacts in mouse embryonic stem cells during their differentiation to haematopoietic progenitor cells in the TAD containing the gene *RUNX1* (Owens *et al.*, 2022). This approach has been ground-breaking in demonstrating how dynamic sub-TAD chromatin boundaries might play a role in first establishing TAD structure, and then coordinating gene expression. In this chapter, I have used Tiled-C to identify looping interactions relevant to AS pathogenesis in $CD4^+$ and $CD8^+$ T cells.

5.1.2. Experimental design of Tiled-C experiments

5.1.2.1. Prioritisation of genetic loci

Over 100 GWAS loci are associated with AS (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016), making investigation of every TAD containing a risk-associated genetic variant with Tiled-C prohibitively expensive in both sequencing costs and oligonucleotide bait generation. I therefore prioritised 21 AS-associated loci that showed the best evidence for involvement in AS pathogenesis in the cell types studied throughout this thesis (**Table**

2.5). This included the nine regions that showed reciprocal or complex interactions by Capture-C in chapter four.

5.1.2.2. Choice of cell types and stimulation conditions

I chose to perform Tiled-C in the same bulk MACS-purified cell types as in previous chapters, namely $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes. In chapter four, I observed a lack of Capture-C interaction at 35 of the 44 baited AS-associated loci, such as *AHR*, and proposed that interactions may only be remodelled upon immune stimulation.

Indeed, use of capture Hi-C in LPS-stimulated THP-1 cell line-derived macrophages showed that 10,683 out of 97,423 significant promoter-distal enhancer interactions were either significantly stronger (64% of contacts) or weaker (36% of contacts) following LPS administration for 2 hours (Pradhananga *et al.*, 2020). Furthermore, stimulation of T cells with anti-CD3 and anti-CD28 antibodies induces rapid epigenetic changes (such as increased chromatin accessibility and H3K27ac deposition) that overlap with GWAS variants from a number of different inflammatory conditions (Soskic *et al.*, 2019). These changes also correlated with the increased expression of genes involved in the early stages of T cell activation.

In this chapter, the $CD4^+$ and $CD8^+$ T cells of healthy volunteers were stimulated with anti-CD3/CD28 beads, and healthy $CD14^+$ monocytes were stimulated with LPS. These stimuli result in immune activation phenotypes that are more reflective of cellular states that are observed in the peripheral blood of AS patients (Gupta *et al.*, 2016; Martini *et al.*, 2024), and may yield novel insights into how AS-associated loci and TADs organise to regulate the expression of genes within these cells in response to inflammatory milieu.

5.1.2.3. Choice of patient samples for Tiled-C experiments

In addition to the cells of healthy volunteers, I also performed Tiled-C on the $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes of AS patients, allowing for identification of differential looping events that may explain disease-associated differential expression of genes. As mentioned in the general introduction (section 1.2.5), anti-TNFα biologics are highly successful in AS treatment, but their exact mechanisms of action in AS is not well understood. In order to better understand how biologic therapy affects the epigenetic landscape of AS-associated GWAS risk loci, AS patients were recruited and $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes were purified from PBMCs both before (AS patients with active disease prior to biologic therapy) and following at least three months of treatment with an anti-TNFα therapeutic (adalimumab) (anti-TNF treated AS patients). Comparison of patients before and after biologic therapy could provide insight into which AS-associated GWAS loci are modulated by anti-TNFα treatment. Adalimumab was chosen for two main reasons:

(A) Adalimumab is a fully humanised antibody therapy that is not known to exacerbate the inflammatory status of other pre-existing comorbidities (Mounach & El Maghrouai, 2014). The use of other non-humanised or partially humanised biologics, such as etanercept, have been found to bring on or worsen inflammatory gastrointestinal issues in AS co-presenting with IBD (Tolu *et al.*, 2018; Mihai *et al.*, 2023).

(B) Adalimumab remains the most common form of anti-TNFα biologic used to treat AS in the UK, with the British Society for Rheumatology biologics register suggesting that adalimumab is prescribed in 60% of AS cases (Derakhshan *et al.*, 2021). Generating data in AS patients treated with adalimumab would therefore represent the potential epigenetic changes that occur in the majority of anti-TNFα treated AS patients.

AS patients were carefully selected to ensure the correct diagnosis of active AS, especially since a limited number of AS patients (n = 3 for each condition and cell type) were analysed due to sequencing cost budgetary constraints. Patients were diagnosed with AS following MRI imaging and confirmation of sacroiliitis, and AS cases were reviewed by two consultant rheumatologists. At the time of first sample collection, every patient had a BASDAI scores above 4 – the typical threshold for the diagnosis of uncontrolled AS. The presence of common AS comorbidities was kept to a minimum, with the only confirmed comorbidities being uveitis (n = 4), peripheral arthritis (n = 3), and non-vertebral fractures (n = 1) (**Table. 2.1: AS201, AS202, AS205, and AS301-AS305**).

All patients recruited for Tiled-C experiments showed good response to adalimumab treatment, having complete resolution of sacroiliitis as observed on MRI and reduction of BASDAI scores to less than four three months after first administration of therapeutics.

5.1.3. Defining areas of differential chromatin interaction in the Tiled-C data

Whilst Tiled-C can provide comprehensive overview of chromatin structure at AS-associated loci, identification of statistically significant interactions using this approach is difficult. The Peaky analysis used in chapter four, for instance, cannot be used to define significant interactions across TADs in Tiled-C data. This is because Peaky primarily relies on bait-to-prey interactions (derived from interactions between oligo-baited and non-baited restriction fragments) to model distance decay, accordingly adjust interaction p-values, and identify statistically significant interactions (Eisjbouts *et al.*, 2019). Since in Tiled-C every restriction fragment across the AS-associated TADs is baited, the vast majority of interactions in Tiled-C analysis will instead derive from bait-to-bait interactions. These bait-to-bait interactions are inherently biased as both ends of the

ligation product are targeted by individual baits. Due to the differing bait sequence identities, this double enrichment leads to greater variability in fragment capture efficiencies compared to bait-to-prey interactions where only one end of the fragment is targeted (Ben Zouari *et al.*, 2019; Freiere-Pritchett *et al.*, 2021; Aljogol *et al.*, 2022). Bait-to-bait fragments therefore have different experimental 'visibilities' and higher degrees of statistical background noise that must be removed prior to statistical analysis (Imakaev *et al.*, 2012; Freiere-Pritchett *et al.*, 2021).

In Tiled-C data however, this would lead to the removal of most interactions and too few bait-to-prey fragments being passed for distance modelling which would additionally lead to less accurate interaction significance calls. This issue also applies to other common statistical packages used to call significant interactions in Capture-C and CHi-C – such as CHiCAGO (Capture Hi-C Analysis of Genomic Organisation) (Cairns *et al.*, 2016) and CHiCANE (Capture Hi-C Analysis Engine) (Holgersen *et al.*, 2021) – which also highly recommend the removal of bait-to-bait interactions from capture data for accurate interaction modelling.

The heavy bait-to-bait bias in Tiled-C datasets therefore limits which statistical tools, if any, can be used to investigate differential interactions. In this chapter, I analysed the Tiled-C data (which has been sequenced to a depth of $10x10^6$ fragments per Mbp of the genome tiled to produce a resolution of 2 kbp) using ICE (Iterative Correction and Eigenvector decomposition) (Imakaev *et al.*, 2012). This was done using the CapCruncher pipeline (Downes *et al.*, 2022) with specific details found in chapter two (section 2.7.2). ICE allows for the normalisation of Tiled-C data by modelling interactions to the assumptions that the overall bias between two bins results from the sum of many individual biases (i.e. technical biases such as bait-to-bait interactions, but also biological biases such as the distinct interaction frequencies between different

chromatin states) and that all bins in the analysis should interact a similar number of times and have equal experimental visibility (Imakaev *et al.*, 2012; Servant *et al.*, 2018). This produces a relative contact probability score that can be further normalised to the total number of *cis*-reads (those sequencing reads that align to the same chromosome the Tiled-C region and represent the majority of the valid short-range promoter-enhancer interactions of interest) to account for difference in sequencing depth in each of the three biological replicates prior to merging (**Fig. 5.1**).

The Tiled-C data is presented in this chapter as a comparison of these ICE-normalised interaction matrices between the unstimulated and stimulated healthy volunteer cells and the cells from untreated and adalimumab-treated patients. Specifically, I sought to identity looping interactions that were (a) gained or lost on either stimulation or treatment or (b) showed different relative contact probability scores in at least one of the conditions. These matrices are also overlapped with ChromHMM (Chapter four) to help identify if any changes in the interaction profile can be associated with *cis*-regulatory function.

5.1.4. Intersection of Tiled-C data with other genomic modalities can facilitate interpretation of disease-associated interactions

As comparison between ICE-normalised interaction matrices is not a true quantitative statistical measure of differential contact, I sought to intersect the Tiled-C data with other epigenetic datasets to further evidence to support differential interactions. Here, I chose to perform RNA-seq and ATAC-seq analysis in paired samples and cell types from the same individuals as the Tiled-C experiments. Both assays facilitate the interpretation of the Tiled-C data by defining how gene expression levels and chromatin accessibility changes across the four conditions.

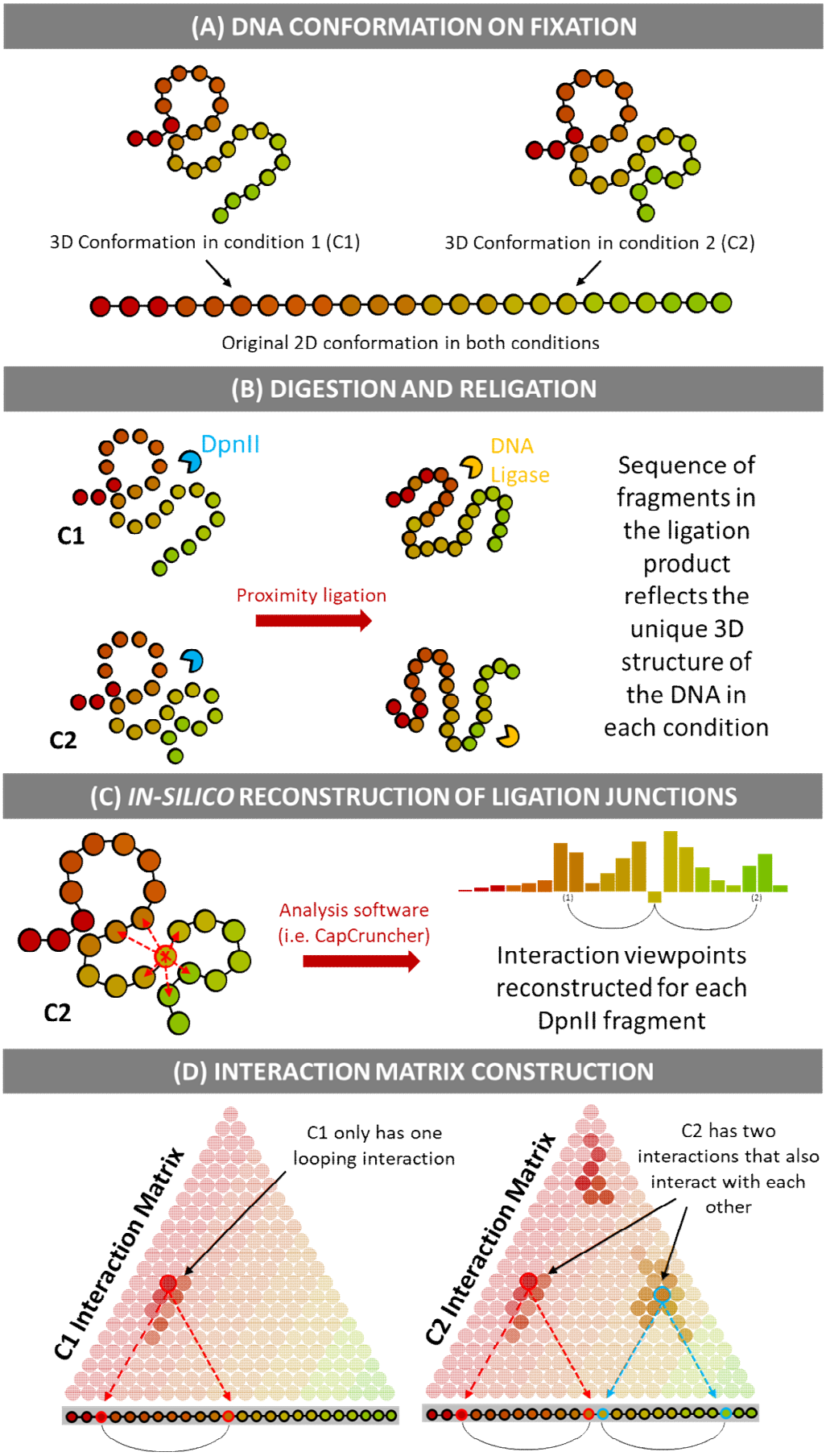
(A) DNA CONFORMATION ON FIXATION
3D Conformation in condition 1 (C1)
3D Conformation in condition 2 (C2)
Original 2D conformation in both conditions
(B) DIGESTION AND RELIGATION
DpnII
DNA Ligase
C1
Proximity ligation
C2
Sequence of fragments in the ligation product reflects the unique 3D structure of the DNA in each condition
(C) IN-SILICO RECONSTRUCTION OF LIGATION JUNCTIONS
C2
Analysis software (i.e. CapCruncher)
Interaction viewpoints reconstructed for each DpnII fragment
(D) INTERACTION MATRIX CONSTRUCTION
C1 Interaction Matrix
C1 only has one looping interaction
C2 Interaction Matrix
C2 has two interactions that also interact with each other

Figure 5.1. Tiled-C analysis strategy towards identifying AS or stimulation-associated differential chromatin contacts with ICE (Continued from previous page). (A) Differing contexts and conditions can create unique chromatin loops between genomic regions (i.e. context-specific promoter-enhancer contacts). Underlying DNA sequence in both conditions on fixation with formaldehyde however remains identical. (B) Digestion of DNA with restriction enzymes (i.e. DpnII) and religation creates unique DNA structures (concatemers) based on the higher chance of ligating DNA fragments in close proximity. These concatemers are more reflective of the chromatin architecture in both conditions. (C) Once the Tiled-C samples have been fully processed (section 2.3), analytic pipelines (CapCruncher) first determine the contact frequencies between each bin genome-wide (one versus all). This is done for each bin in a Tiled-C region, giving complete interaction information at predefined resolution for the entire locus (some vs all). (D) Following ICE and total *cis*-read normalisation to remove bias from each viewpoint, these contact frequencies can be represented in triangular 2D heatmaps known as contact matrices. In these matrices, the intensity of the colour at the intersection of two bins is representative of the relative contact probability. Comparison of the relative contact probabilities in these matrices between conditions can help identify condition-specific looping events and identifying whether these condition-responsive interactions intersect with genetic variants (i.e. Credible or lead AS GWAS SNPs) could also yield further insight into where gene regulation might be altered in disease. Figure adapted from Davies *et al.* (2017).

Moreover, the Tiled-C matrices are also intersected with the H3K3me3 (promoter activity), H3K27ac (*cis*-regulatory looping activity), and H3K27me3 (Polycomb-associated silencing) histone mark data from unstimulated healthy volunteers and untreated AS patients shown in chapter four. This creates a comprehensive resource of the epigenetic landscape in healthy immune cells across immune, inflammatory, and/or activatory stimulation conditions, and in AS patients before and after treatment with anti-TNFα therapeutics, that will help identify SNPs that might be associated with differential chromatin interactions found in these different cell types and states.

5.1.5. Chapter aims

The overall aim of this chapter was to investigate how chromatin looping interactions at AS-associated loci could contribute to AS pathogenesis. This will be achieved using a combination of Tiled-C, ChromHMM, RNA-seq, and ATAC-seq in healthy volunteer

cells (unstimulated/stimulated) and AS patients (before and after biologic therapy). The specific chapter aims are:

[1] To generate a comprehensive epigenetic map of chromosome looping interactions, gene expression, and accessible chromatin in immune cells isolated from healthy volunteers and AS patients.

[2] To identify looping interactions that overlap AS-associated SNPs and identify the gene promoters, regulatory regions, or other evidence supporting a potential role for the SNP in modulating a specific gene.

[3] Identify differential looping interactions on comparison of:

(i) Unstimulated and stimulated immune cells.

(ii) AS patients and healthy volunteers

(iii) AS patients before and after treatment with anti-TNFα therapeutics.

5.2. Tiled-C at control loci identifies known interactions

5.2.1. Assessment of data quality at the volunteer locus *MYC*

Prior to performing Tiled-C on the 21 AS-associated GWAS regions, I wanted to ensure that the experiment would produce data of high enough quality to identify putative disease-relevant interactions. Towards this, I first performed a pilot Tiled-C experiment focusing on the *MYC* locus as the epigenetic landscape and enhancers that surround *MYC* are amongst the best characterised across the genome. Studies have used CRISPR (Clustered regularly interspaced short palindromic repeats) interference (CRISPRi) in lymphocytic K562 cell lines to demarcate the exact position of enhancers involved in *MYC* regulation (Fulco *et al.*, 2019).

CRISPRi sterically hinders the binding of transcription factors to a targeted DNA region by guiding a catalytically dead Cas9 protein to the genomic site with the use of complementary single-stranded guide RNA (sgRNA). High CRISPRi scores are indicative of DNA regions with disproportionate effect on gene expression (i.e. enhancer and promoter regions). By tiling 1.2. Mbp of the *MYC* locus with sgRNAs, Fulco *et al.* (2019) were able to detect a total of seven enhancer regions and two repressive elements with high CRISPRi scores characterised by an ENCODE Consortium ChromHMM. Interactions between the *MYC* promoter and the two enhancers have also been reported by Hnisz *et al.* (2016) and I was able to recapitulate these in Chapter four using Capture-C (**Fig. 4.3**).

Here, I sought to identify whether these known promoter-enhancer interactions were present in my Tiled-C data and to determine the overall quality of the data across the four different conditions in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes.

5.2.2. Tiled-C data quality varied across cell types

Having obtained interaction matrices for the *MYC* locus in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes for the four conditions, I noticed that after ICE normalisation the contact probabilities in the unstimulated and LPS-stimulated healthy volunteer $CD14^+$ monocyte data at the *MYC* locus were much lower than those observed in the other cell types and conditions (**Fig. 5.2**). An overall four to five-fold reduction in the ICE and total cis-read normalised relative contact probabilities could be observed when comparing the LPS-stimulated healthy volunteer $CD14^+$ monocytes to any other cell type and/or condition. To determine if this was caused by lower sample quality, I compared the number of *cis*-reads to the number of *trans*-reads (those sequencing reads that map to other chromosomes that arise from a mixture of genuine interchromosomal interactions and random internuclear ligation events) produced for all Tiled-C samples. The quality of a Tiled-C library is partly defined by the ratio of these *cis*-to-*trans* sequencing reads (Akgol Oksuz *et al.*, 2021), with a high quality Tiled-C library typically requiring a *cis*-read proportion in excess of 60% of all sequenced reads for downstream analysis (Downes *et al.*, 2022). The healthy volunteer $CD14^+$ monocyte samples had a greater number (~10%) of *trans*-reads compared to other Tiled-C samples (**Fig. 5.3**).

The average number of *cis*-reads across the three unstimulated and stimulated healthy volunteer monocyte biological replicates was also below the typical 60% quality threshold. This suggests that while the Tiled-C data from all T cell samples and AS patient $CD14^+$ monocytes is of similar quality, the data from the healthy $CD14^+$ monocytes is of substantially lower quality and is unsuitable at this moment for further use. This chapter will therefore focus on identifying disease-relevant looping interactions found only in the $CD4^+$ and $CD8^+$ T cell Tiled-C data.

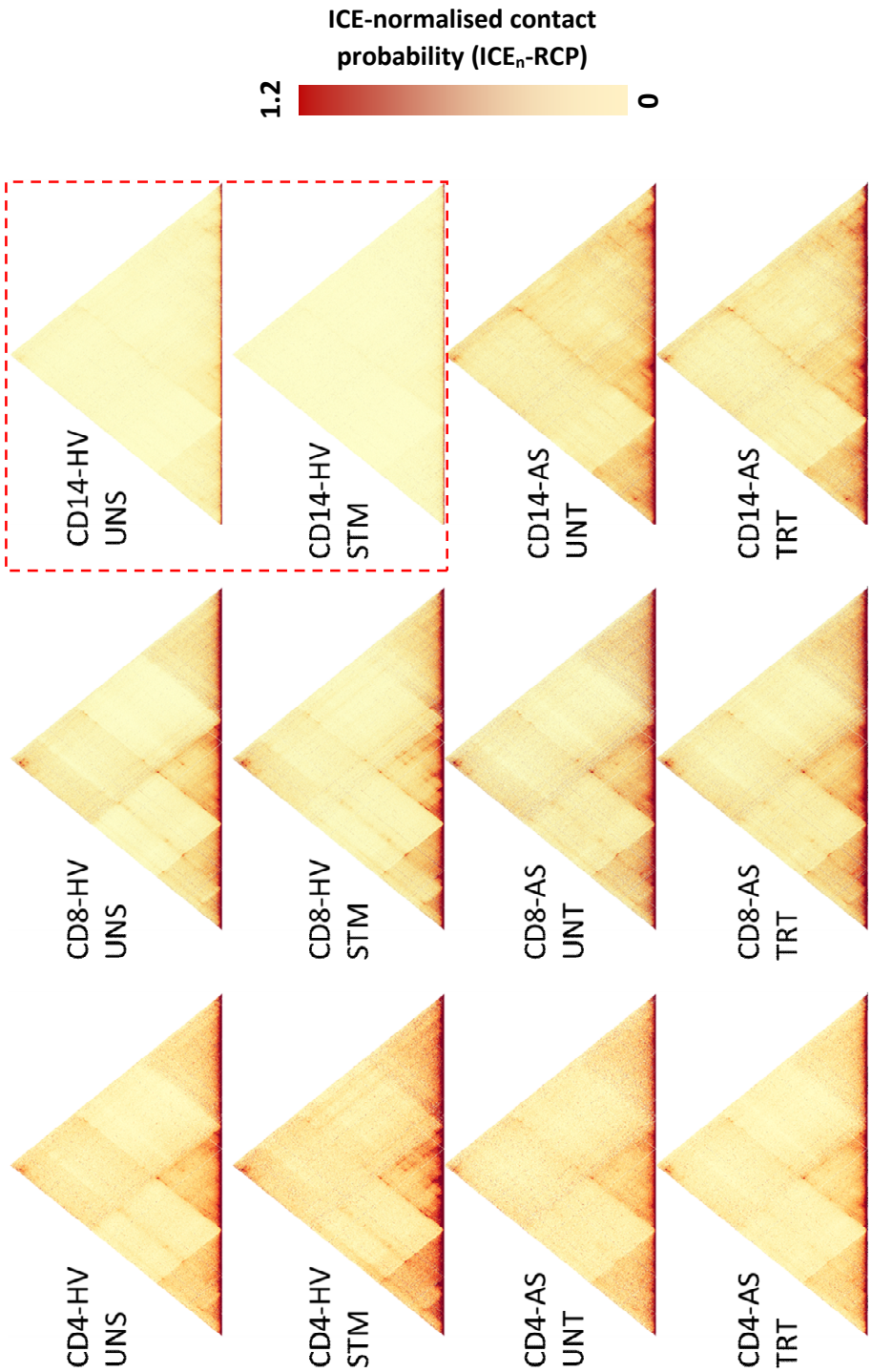

Figure 5.2. Tiled-C matrices for the *MYC* control locus in CD4$^+$ T cells, CD8$^+$ T cells, and CD14$^+$ monocytes. Each plot shows the mean relative ICE contract probabilities normalised to the proportion of *cis*-reads in each sample for CD4$^+$ T cells (CD4), CD8$^+$ T cells (CD8), CD14$^+$ monocytes (CD14) across the *MYC* locus (chr8: 127800000-130850000). 'UNS' and 'STM' plots show unstimulated healthy volunteer (HV) samples, whilst 'UNT' and 'TRT' plots show untreated and anti-TNFα (adalimumab) treated AS patient (AS) samples. Dashed red box shows mean matrices for the lower quality healthy volunteer unstimulated and stimulated CD14$^+$ monocyte samples.

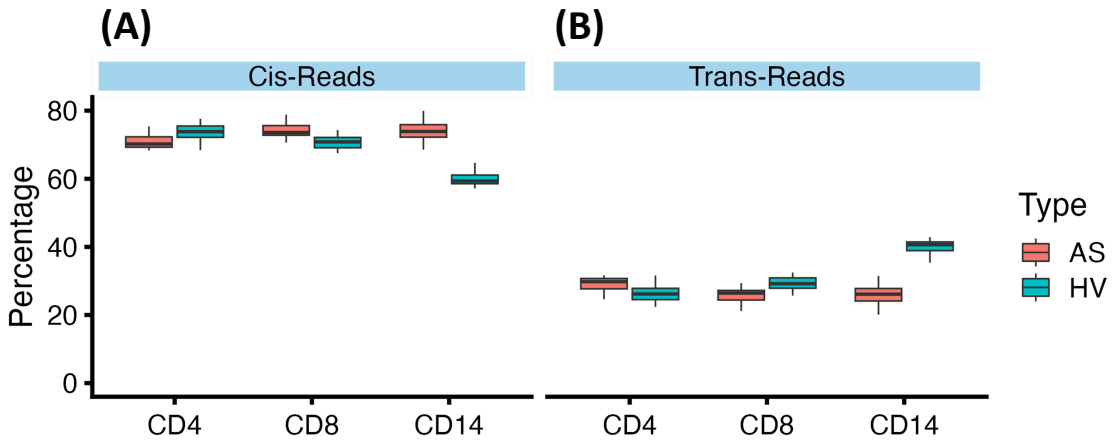

Fig. 5.3. *Cis*-to-*trans* read ratios in the Tiled-C samples. Box plots show the percentage of *cis* (A) and *trans* (B) reads in Tiled-C samples following NGS and alignment to human genome GRCh37 for $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes. Conditions for each of the cell types were combined (n = 6), and reads are shown as the average *cis* or *trans* reads reported across all tiled regions. 'AS' = AS patients; 'HV' = healthy volunteers.

5.2.3. The regulatory landscape of *MYC* is altered by stimulation of CD4$^+$ and CD8$^+$ T cells with CD3/CD28 T cell activation beads

It has previously been shown that the expression of *MYC* and chromosome looping interactions in the locus are increased on T cell activation with anti-CD3/CD28 beads (Glinos *et al.*, 2020; Marchingo *et al.*, 2020). In order to ascertain whether these responses could be observed in the Tiled-C data, I intersected the validated interactions from Hnisz *et al.* (2016) and the high CRISPRi enhancer features from Fulco *et al.* (2019) with the Tiled-C interaction matrices I generated for unstimulated and CD3/CD28-stimulated healthy volunteer CD4$^+$ T cells (**Fig. 5.4**).

In the unstimulated healthy volunteer CD4$^+$ T cell interaction matrix, the contacts between the *MYC* promoter (**Fig 5.4: i**) and the two validated enhancer interactions from Hnisz *et al.* (2016) (**Fig. 5.4: ii**) are clearly recapitulated. These interactions are also seen with similar contact frequency in the CD3/CD28-stimulated CD4$^+$ T cell interaction matrix; suggesting the Tiled-C data is of high quality and can identify known enhancer-promoter interactions. In addition, five new contacts are formed with the *MYC* promoter on stimulation of CD4$^+$ T cells (**Fig. 5.4: iii**). The Tiled-C interaction matrices were intersected with ChromHMM chromatin states (generated in chapter three) and ATAC-seq data from CD4$^+$ and CD8$^+$ T cells across the four conditions. RNA-seq was also performed in order to determine whether the formation of contacts on stimulation could be associated to differential gene expression in these cells.

Notably, ChromHMM analysis shows that both the validated enhancer and stimulation-sensitive contacts intersect with active *cis*-regulatory element-associated chromatin states (i.e. U1 TSSActv, U2 TSSFlnkU, U3 TSSFlnkD, and U10 EnhActv) in both unstimulated healthy volunteers and untreated AS patients.

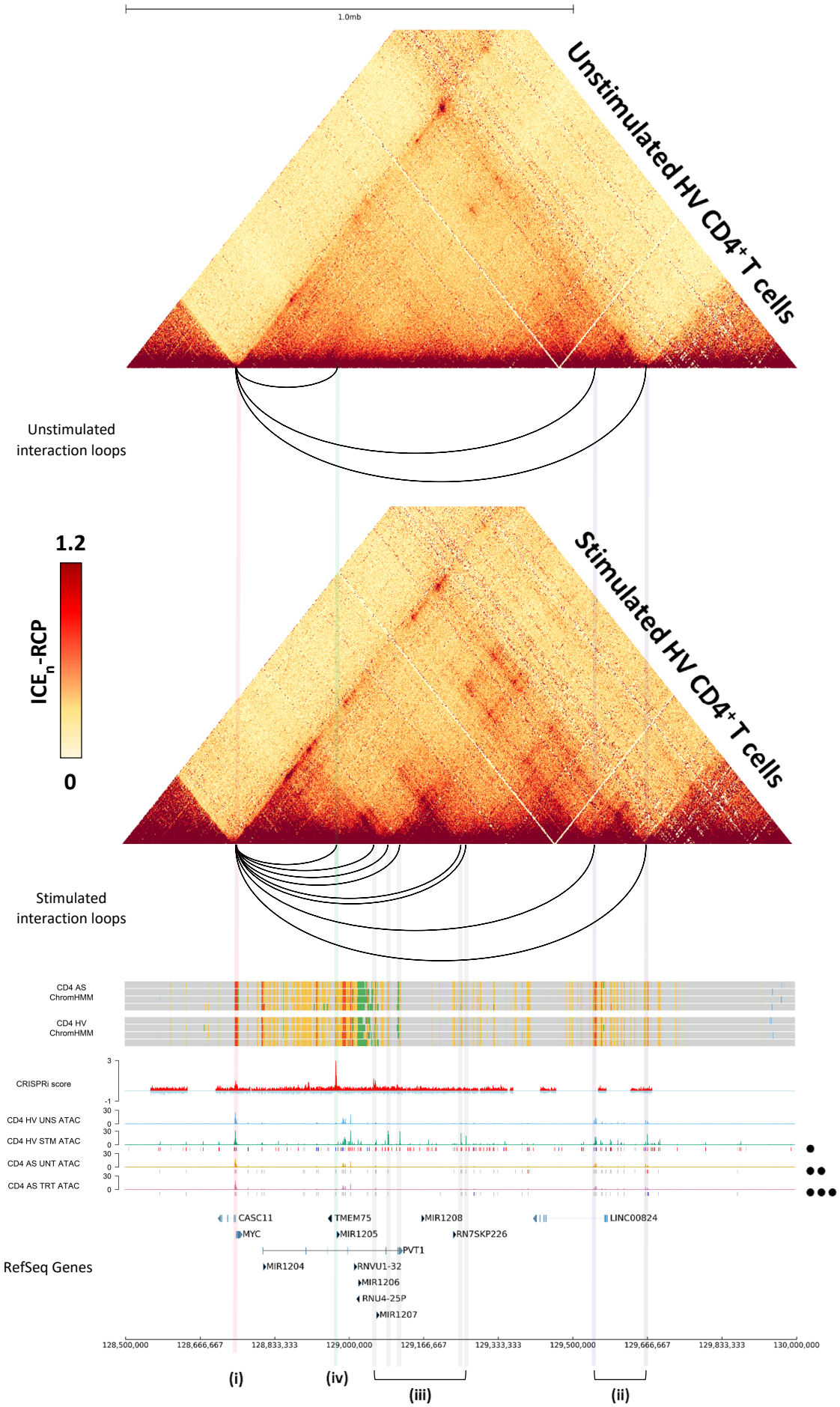
1.0mb
Unstimulated HV CD4+ T cells
Unstimulated interaction loops
1.2
ICE$_n$-RCP
0
Stimulated HV CD4+ T cells
Stimulated interaction loops
CD4 AS ChromHMM
CD4 HV ChromHMM
CRISPRi score
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
RefSeq Genes
CASC11
MYC
TMEM75
MIR1205
MIR1208
RN7SKP226
LINC00824
PVT1
MIR1204
RNVU1-32
MIR1206
RNU4-25P
MIR1207
128,500,000
128,666,667
128,833,333
129,000,000
129,166,667
129,333,333
129,500,000
129,666,667
129,833,333
130,000,000
(i)
(iv)
(iii)
(ii)

Figure 5.4. Intersection of Tiled-C and epigenetic data at the MYC locus informs the position of CD3/CD28 stimulation-induced chromosome looping (Continued from previous page). From top to bottom: Mean Tiled-C interaction matrices (2 kbp resolution) for unstimulated (n=3) and CD3/CD28-stimulated (n=3) bulk $CD4^+$ T cells from healthy volunteers (HV). Matrices represent the ICE and total *cis*-read normalised relative contact probability score (ICE_n-RCP) between each 2 kbp bin. 'Unstimulated Interaction Loops' and 'Stimulated Interaction Loops' show manually annotated interaction loops between the MYC promoter and other *cis*-regulatory features of interest in unstimulated and stimulated $CD4^+$ T cells respectively. 'CD4 AS ChromHMM' shows the 16-state ChromHMM for the four AS patients from chapter three, and 'CD4 HV ChromHMM' shows the equivalent in healthy volunteers. 'CRISPRi score' shows the interference scores from Fulco *et al.* (2019). 'CD4 HV UNS ATAC', 'CD4 HV STM ATAC', 'CD4 AS UNT ATAC', and 'CD4 AS TRT ATAC' show mean ATAC-seq peaks from unstimulated healthy volunteers (n = 31), stimulated healthy volunteers (n = 5), untreated AS patients (n =20), and adalimumab-treated AS patients (n = 20) respectively. Boxes underneath the ATAC tracks represent significantly differential peaks (Red = upregulated peaks in second condition; Grey = no change between conditions; Blue = downregulated peaks in second condition). ● = unstimulated versus stimulated cells from healthy volunteers; ●● = unstimulated cells from healthy volunteers versus untreated AS patients; ●●● = untreated versus treated AS patients. 'RefSeq Genes' shows the position of known genes in the locus. Scale bar at the bottom of figure shows genomic position of genetic features. Vertical lines depict regions of epigenetic interest: (i) *MYC* promoter (red); (ii) Validated interactions from Hnisz *et al.* (2016) in lymphocyte cell lines (purple); (iii) CD3/CD28-stimulated sensitive enhancers (Grey); (iv) Area of highest CRISPRi score (Green). Locus shown between chr8:128,500,000-130,000,000.

$CD4^+$ T cell stimulation was associated with significant, locus-wide increases in chromatin accessibility (**Fig. 5.4: ●**), with a total of 62 differential open chromatin peaks appearing on CD3/CD28 stimulation. Activation-associated ATAC-seq peaks were also observed to intersect with the two validated enhancer interactions and six ChromHMM-defined active enhancers. This suggests that the contacts formed with the *MYC* promoter in $CD4^+$ T cells on CD3/CD28 stimulation have putative regulatory functions and are only open to accessory proteins (i.e. transcription factors) in the stimulated state.

Three of the identified interactions aligned with regions of high CRISPRi score; including the *MYC* promoter, the leftmost enhancer of the stimulation-associated enhancers, and the region of highest CRISPRi score (**Fig. 5.4: iv**). Unlike the other two regions however, the later region was not predicted by ChromHMM with the current input datasets to have defined *cis*-regulatory function, nor did it show an ATAC-seq peak in either the unstimulated or CD3/CD28-stimulated $CD4^+$ T cells. This suggests an alternative regulatory role for this specific looping interaction outside of conventional promoter-enhancer contacts.

The formation of these looping *cis*-regulatory contacts in the stimulated healthy volunteer $CD4^+$ T cells correlates with the upregulation of *MYC* gene expression (Log_2 fold change = 2.01; adjusted p-value = 1.00×10^{-24}) (**Fig. 5.5: A**). Taken together, these findings suggest that the interactions that appear on CD3/CD28 stimulation evidenced by Tiled-C occur between validated enhancers and the *MYC* promoter, and that these contacts may contribute to the upregulation of *MYC* expression in activated $CD4^+$ T cells. These trends can also be seen in $CD8^+$ T cells (**Fig. S.6**), which show similar upregulation of *MYC* on CD3/CD28 stimulation (Log_2fold change = 2.17, adjusted p-value = 2.77×10^{-10}) (**Fig. 5.5: B**).

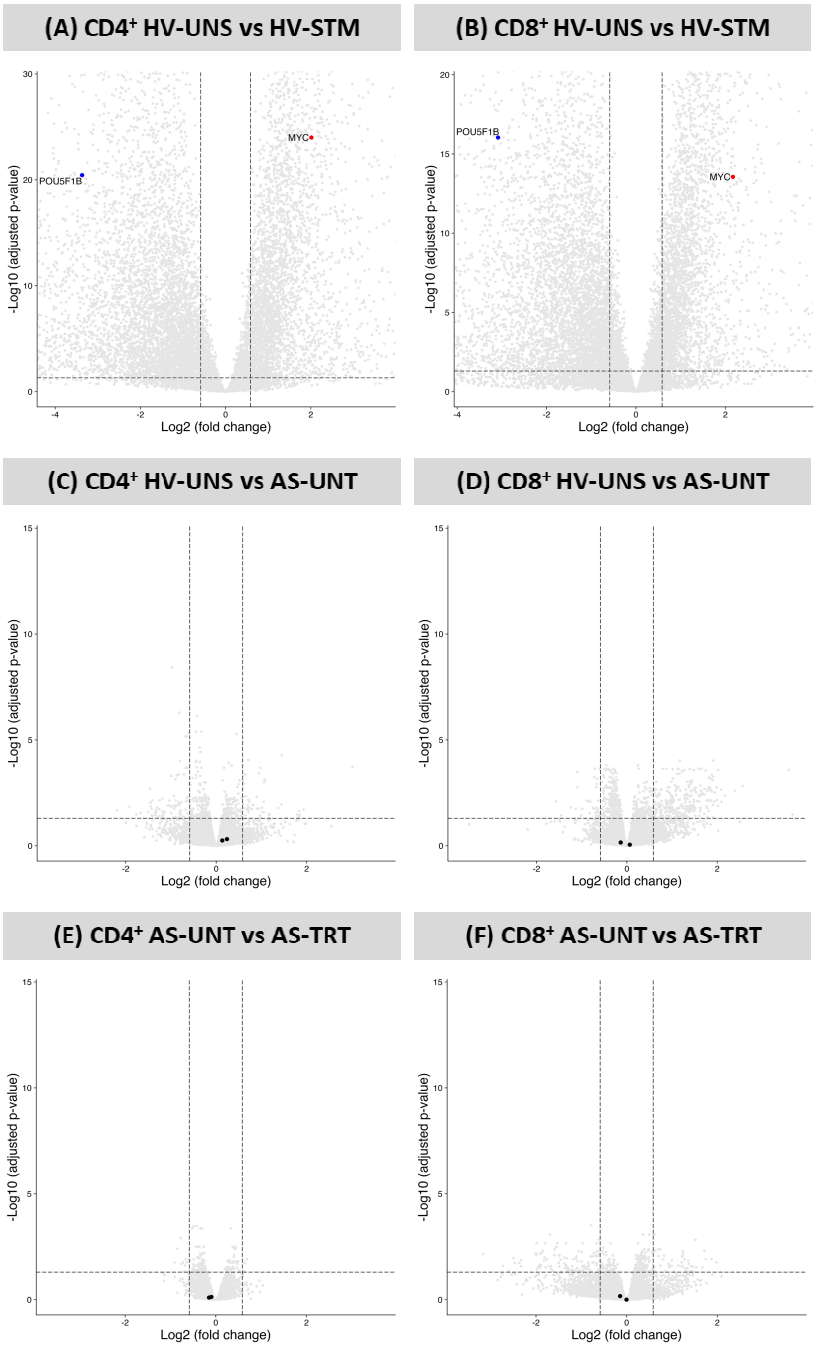

Figure 5.5. Expression of *MYC* is significantly upregulated in CD4$^+$ and CD8$^+$ T cells on CD3/CD28 stimulation but not in AS patients before or after treatment with adalimumab. Volcano plots showing differential gene expression in the *MYC* Tiled-C locus (chr8: 127,800,000 to 130,850,000). Upregulated (red) and downregulated (blue) genes in stimulated vs unstimulated cells are indicated. Adjusted p-value (0.05) and log2 fold change (0.585) significance thresholds are shown. (A) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer CD4$^+$ T cells. (B) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer CD8$^+$ T cells. (C) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) CD4$^+$ T cells. (D) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) CD8$^+$ T cells. (E) Untreated (n = 10) versus adalimumab-treated (n = 10) AS patient CD4$^+$ T cells. (F) Untreated (n = 11) versus adalimumab-treated (n = 11) AS patient CD8$^+$ T cells.

5.2.4. AS patients pre and post-treatment with adalimumab lack the CD3/CD28-associated enhancer contacts seen in stimulated healthy volunteer cells

Having observed that the formation of new enhancer interactions is associated with the upregulation of *MYC* on CD3/CD28-stimulation, I next sought to determine whether the interactions found in stimulated T cells were also detected in AS patients pre and post-treatment with adalimumab. Whilst *MYC* has not been identified as an AS susceptibility locus by any GWAS study to date, certain AS-associated genetic variants – such as *rs6600247* at the *RUNX3* locus – are thought to disrupt *MYC* binding consensus motifs and potentially effect gene expression (Cohen *et al.*, 2021). *MYC* is also upregulated in regard to increased pro-inflammatory cytokine signalling pathways (such as TNFα and TGF-β) which are also seen elevated in AS (Liu *et al.*, 2015). Therefore, although I have used *MYC* as a control locus to assess Tiled-C quality, understanding whether the same epigenetic mechanisms that contribute to the upregulation of *MYC* in stimulated healthy volunteer T cells are maintained in the AS patient cells pre and post-treatment could yield important insights into AS pathogenesis.

In both untreated AS patient and adalimumab-treated AS patient $CD4^+$ T cells, the contacts between the *MYC* promoter (**Fig. 5.6: i**), both the Hnisz enhancers (**Fig. 5.6: ii**), and the region of highest CRISPRi score (**Fig. 5.6: iii**) are observed at similar relative contact probability scores as in unstimulated healthy volunteer CD4+ T cells. This suggests that the AS patient Tiled-C data is of similar overall quality to the healthy volunteer data, and can be used to identify differential interactions that may occur between unstimulated healthy volunteers and untreated AS patients, or untreated and treated AS patients. The stimulation-sensitive contacts observed in the stimulated healthy volunteer $CD4^+$ T cells are however not seen in either the untreated or treated AS patient contact matrices.

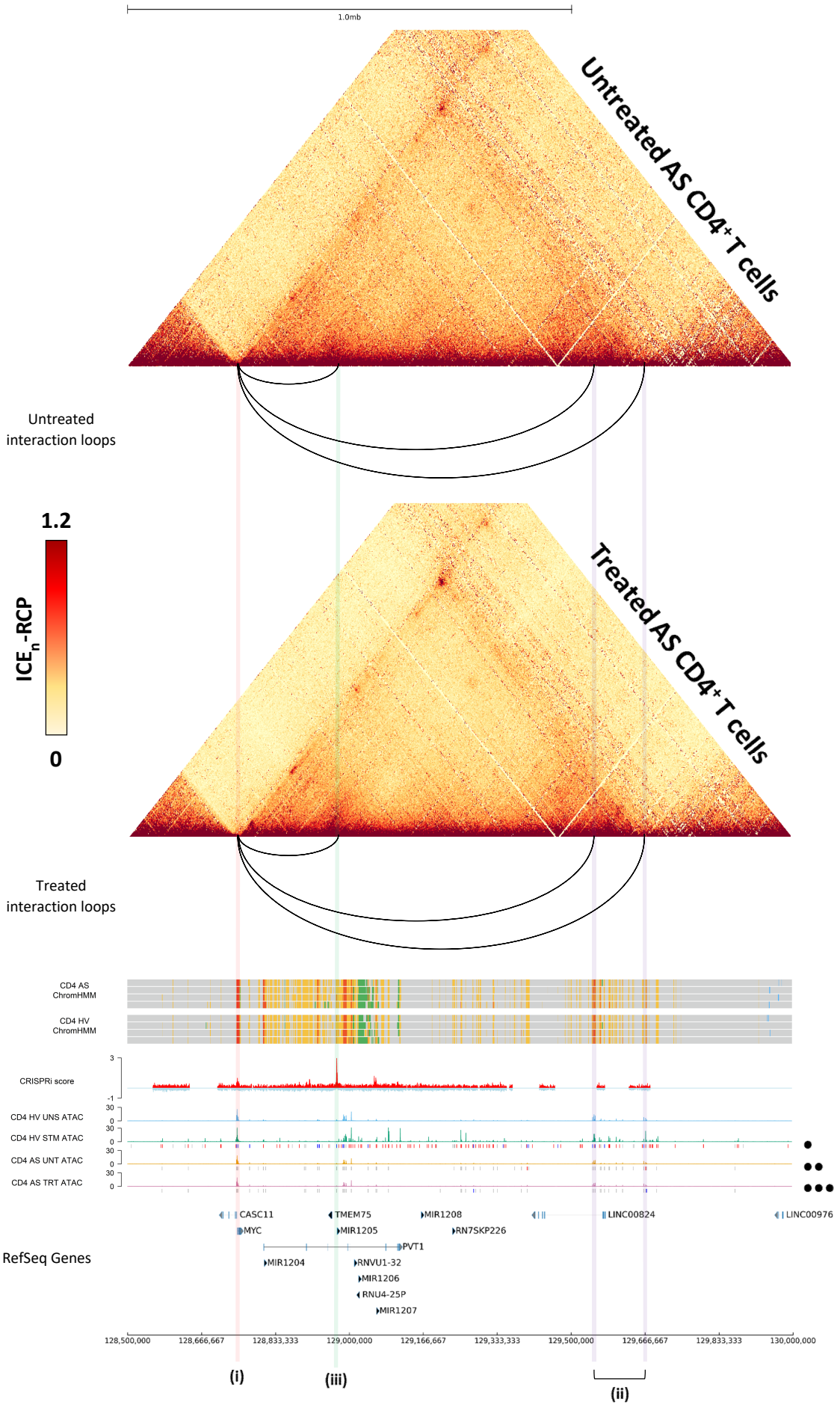
1.0mb
Untreated AS CD4+ T cells
Untreated interaction loops
1.2
ICE_n-RCP
0
Treated AS CD4+ T cells
Treated interaction loops
CD4 AS ChromHMM
CD4 HV ChromHMM
CRISPRi score
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
RefSeq Genes
CASC11
TMEM75
MIR1208
LINC00824
LINC00976
MYC
MIR1205
RN7SKP226
PVT1
MIR1204
RNVU1-32
MIR1206
RNU4-25P
MIR1207
128,500,000
128,666,667
128,833,333
129,000,000
129,166,667
129,333,333
129,500,000
129,666,667
129,833,333
130,000,000
(i)
(iii)
(ii)

Figure 5.6. Tiled-C interaction profiles for CD4$^+$ T cells isolated from AS patients pre and post-treatment with adalimumab are similar and lack the enhancer contacts seen in CD3/CD28-stimulated healthy volunteer cells (Continued from previous page). From top to bottom: Mean Tiled-C interaction matrices (2 kbp resolution) for untreated (n=3) and adalimumab-treated (n=3) bulk CD4$^+$ T cells from AS patients (AS). Matrices represent the ICE and total *cis*-read normalised relative contact probability score (ICE_n-RCP) between each 2 kbp bin. 'Untreated Interaction Loops' and 'Treated Interaction Loops' show manually annotated interaction loops between the *MYC* promoter and other *cis*-regulatory features of interest in untreated and treated CD4$^+$ T cells respectively. 'CD4 AS ChromHMM' shows the 16-state ChromHMM for the four AS patients from chapter three, and 'CD4 HV ChromHMM' shows the equivalent in healthy volunteers. 'CRISPRi score' shows the interference scores from Fulco *et al.* (2019). 'CD4 HV UNS ATAC', 'CD4 HV STM ATAC', 'CD4 AS UNT ATAC', and 'CD4 AS TRT ATAC' show mean ATAC-seq peaks from unstimulated healthy volunteers (n = 31), stimulated healthy volunteers (n = 5), untreated AS patients (n =20), and adalimumab-treated AS patients (n = 20) respectively. Boxes underneath the ATAC tracks represent significantly differential peaks (Red = upregulated peaks in second condition; Grey = no change between conditions; Blue = downregulated peaks in second condition). ● = unstimulated versus stimulated healthy volunteers; ●● = unstimulated healthy volunteers versus untreated AS patients; ●●● = untreated versus treated AS patients. 'RefSeq Genes' shows the position of known genes in the locus. Scale bar at the bottom of figure shows genomic position of genetic features. Vertical lines depict regions of epigenetic interest: (i) *MYC* promoter (red); (ii) Validated interactions from Hnisz *et al.* (2016) in lymphocyte cell lines (purple); (iii) CD3/CD28-stimulated sensitive enhancers (Grey); (iv) Area of highest CRISPRi score (Green). Locus shown between chr8:128,500,000-130,000,000.

The lack of these contacts coincides with no significant differential expression in the RNA-seq data when comparing unstimulated healthy volunteers to AS patients with active disease (**Fig. 5.5: C and D**) or AS patients with active disease (untreated) versus after anti-TNF treatment (adalimumab-treated/treated) (**Fig. 5.5: E and F**). Furthermore, ATAC-seq comparison between unstimulated healthy volunteers and untreated AS patients (**Fig. 5.6**: ●●) or untreated and treated AS patients (**Fig. 5.6:** ●●●) suggests that chromatin accessibility at this region is not as significantly differential in AS before and after biologic therapy for 3 months as it is on CD3/CD28 stimulation. Indeed, in contrast to the 62 differential open chromatin regions identified between unstimulated and stimulated $CD4^+$ T cells, only two peaks with significantly higher accessibility in the unstimulated healthy volunteer versus untreated AS patient comparison, and two peaks with significantly lower chromatin accessibility in the untreated versus adalimumab-treated AS patient samples, were identified at the *MYC* locus. Similar results are seen in untreated and treated $CD8^+$ T cells (**Fig. S.6**), which also lack the stimulation-associated contacts, do not show differential *MYC* expression, and present with similar open chromatin profiles as unstimulated healthy volunteer $CD8^+$ T cells. These results collectively suggest that the epigenetics of the *MYC* locus in untreated and treated AS patient peripheral blood $CD4^+$ and $CD8^+$ T cells are similar to unstimulated healthy volunteer cells.

5.3. Genome-wide epigenetic changes in untreated, treated, and stimulated cells

5.3.1. The gene expression and chromatin accessibility profiles of AS patient $CD4^+$ and $CD8^+$ T cells pre and post-adalimumab treatment more closely resemble unstimulated healthy volunteer cells over those stimulated with CD3/CD28 beads

Having seen no differences in the interaction, open chromatin, and expression profiles of unstimulated healthy volunteer and untreated or treated AS patient $CD4^+$ and $CD8^+$ T cells at the *MYC* locus, I next sought to identify whether this similarity was a feature of all of the captured loci or whether stimulation-shared features could be identified at any of the 21 AS-associated Tiled-C regions. Towards this, I investigated how the ATAC-seq and RNA-seq signatures across the captured loci differ amongst the four different conditions with DESeq2 for a total of six comparisons:

- Unstimulated healthy volunteers versus stimulated healthy volunteers.
- Unstimulated healthy volunteers versus untreated AS patients.
- Unstimulated healthy volunteers versus treated AS patients.
- Stimulated healthy volunteers versus untreated AS patients.
- Stimulated healthy volunteers versus treated AS patients.
- Untreated AS patients versus treated AS patients.

The analysis showed that the major driver of differential gene expression (**Fig. 5.7**) in the Tiled-C regions is CD3/CD28 stimulation. In $CD4^+$ T cells, 526 of the 531 differential genes involved comparisons that included stimulated cells (**Fig. 5.7: A**). Results were similar in $CD8^+$ T cells, where 494 of the 512 differential genes also arose from comparison with stimulated cells (**Fig. 5.7: B**). All identified differential genes (aside from *PAPOLG* in the untreated AS patient versus treated AS patient $CD8^+$ T cell

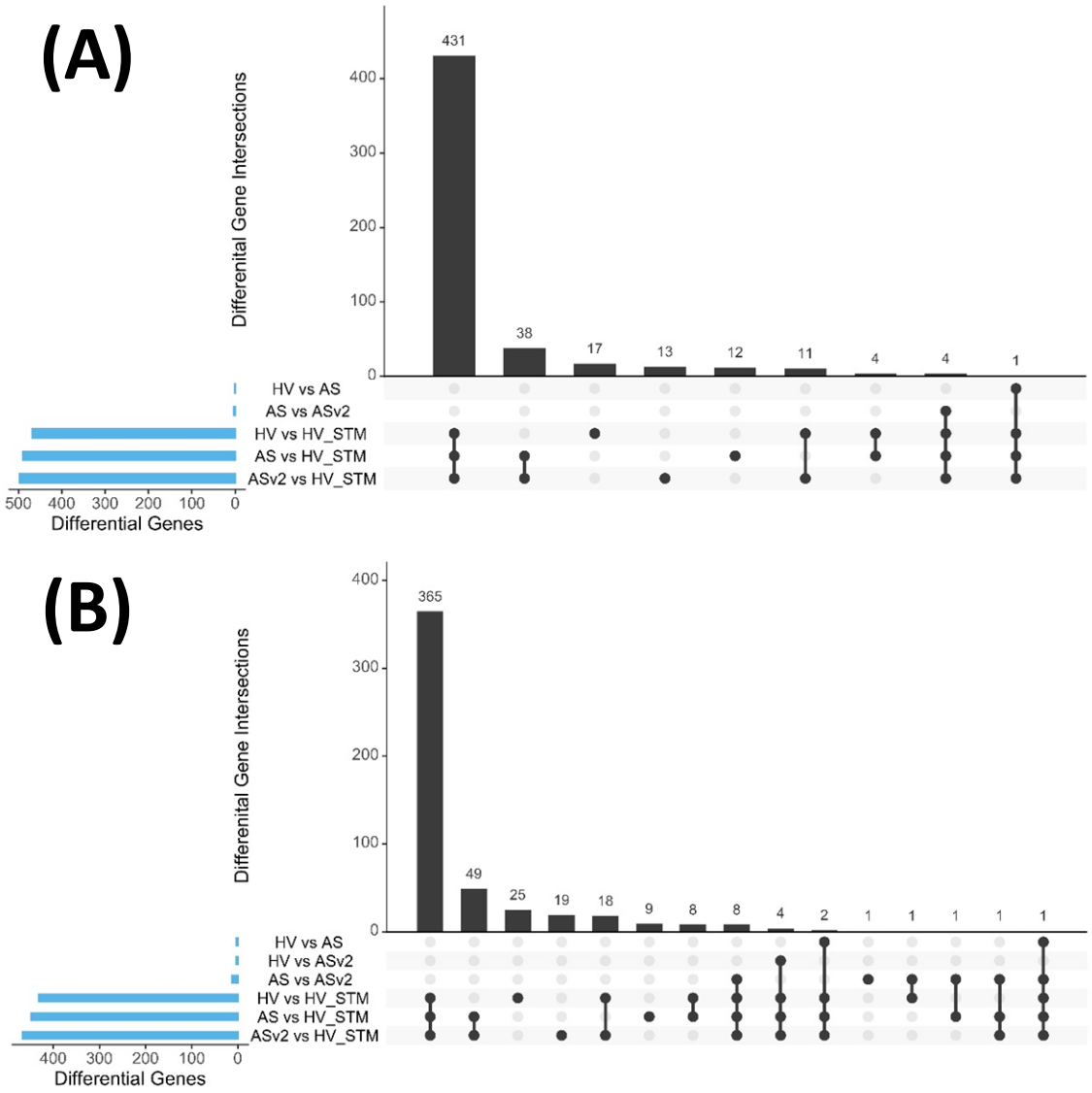

Figure. 5.7. The majority of differentially expressed genes in the 21 AS-associated Tiled-C regions are associated with CD3/CD28-stimulated T cells instead of disease or treatment state. UpSet plots showing the number of differentially expressed genes for each condition comparison in (A) $CD4^+$ T cells and (B) $CD8^+$ T cells. HV = Unstimulated healthy volunteer cells; HV_STM = CD3/CD28-stimulated healthy volunteer cells; AS = untreated AS patient cells; ASv2 = adalimumab-treated AS patient cells Filled black dots and connecting bars show which differentially expressed genes are shared between each of the comparisons. Blue bars show the total number of differentially expressed genes for each comparison. Note that the HV vs ASv2 comparison in $CD4^+$ T cell produced no significantly differential expressed genes (adjusted p-value ≤ 0.05; $\log_2$ fold change $\geq \pm 0.585$).

comparison) could also be found in the comparisons with CD3/CD28-stimulated $CD4^+$ and $CD8^+$ T cell samples; suggesting these genes may also play a role in modulating T cell activation across the four conditions. Five of these differential genes were found in $CD4^+$ T cells; including *SCRN2* (*NPEPPS* locus), *CTLA4* (*ICOS-CTLA4-CD28* locus), and *CDC37* (*TYK2-ICAM1-ICAM3*). In $CD8^+$ T cells, a total of 23 differential genes were identified; including *TNFAIP3* (*IL22RA2-TNFAIP3* locus), *EOMES* (*EOMES-CMC1* locus), *TAGAP* (*TAGAP* locus), and *ITGB3* (*NPEPPS* locus).

The trends observed in the RNA-seq data are even more pronounced in the ATAC-seq data (**Fig. 5.8**). It has previously been suggested that proximal (≤ 50 kbp from a gene TSS) *cis*-regulatory elements are important in defining the transcriptional activity of neighbouring genes (Wilczynski *et al.*, 2012). Differential peaks found within 50 kbp either end of a TSS were therefore assumed to have strong influence on gene expression, and the ATAC-seq analysis reported here was limited to these proximal open chromatin peaks across the 21 Tiled-C regions. Under these parameters – despite the overall identification of 2,397 and 2,386 differential open chromatin regions in $CD4^+$ and $CD8^+$ T cells respectively – I found that no differential ATAC-seq peaks in $CD4^+$ T cells (**Fig. 5.8: A**) and only three differential peaks in $CD8^+$ T cells (**Fig. 5.8: B**) did not associate with CD3/CD28 stimulation.

Genome-wide analysis of the shared differential peaks across the four $CD4^+$ (**Fig. S.7: A**; **Fig. S.8: A**) and $CD8^+$ (**Fig. S.7: B**; **Fig. S.8: B**) T cell conditions imply that these trends are maintained across the genome. Taken together, these results suggest that AS patient $CD4^+$ and $CD8^+$ T cells from peripheral blood pre and post-treatment with adalimumab are more similar in terms of chromatin accessibility and gene expression profile to unstimulated healthy volunteer cells than to CD3/CD28-activated T cells.

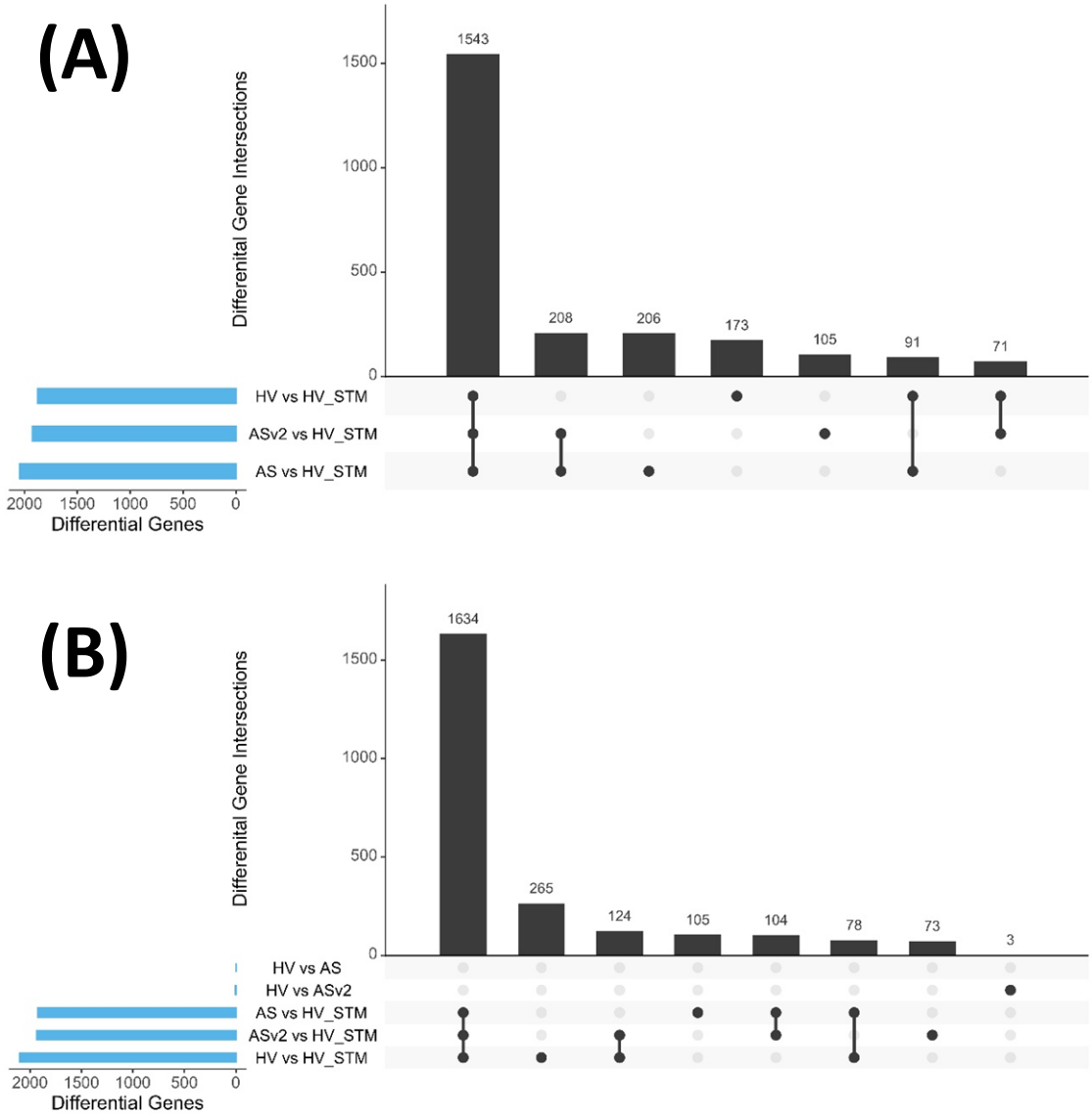

Figure. 5.8. The majority of differential open chromatin regions in the 21 AS-associated Tiled-C regions are associated with CD3/CD28-stimulated T cells instead of disease or treatment state. UpSet plots showing the number of differential open chromatin regions ± 50 kbp of a TSS for each condition comparison in (A) $CD4^+$ T cells and (B) $CD8^+$ T cells. HV = Unstimulated healthy volunteer cells; HV_STM = CD3/CD28-stimulated healthy volunteer cells; AS = untreated AS patient cells; ASv2 = adalimumab-treated AS patient cells Filled black dots and connecting bars show which differentially expressed genes are shared between each of the comparisons. Blue bars show the total number of differential open chromatin regions for each comparison. Note that the HV vs AS, HV vs ASv2, and AS vs ASv2 comparisons in $CD4^+$ T cells and the AS vs ASv2 comparison in $CD8^+$ T cells produced no differential open chromatin regions (adjusted p-value ≤ 0.05; $\log_2$ fold change $\geq \pm 0.585$).

5.3.2. AS-associated SNPs are found in regulatory elements that respond to T cell activation and stimulation with CD3/CD28 beads

Identification of SNPs that affect genes involved in T cell activation could be relevant in AS, as such genes can polarise T cells to distinct pro-inflammatory or immunosuppressive phenotypes based on how expression is influenced. Since several genes involved in T cell activation are located in AS GWAS-associated regions (such as *IL7R*, *SH2B3*, *BACH2*, *ICOSL*, *ZMIZ1*, *EOMES,* and *TBX21*), genes which show altered interactions between regulatory elements on CD3/CD28 stimulation may prove important towards AS pathogenesis (Rosine *et al.*, 2023). Having generated high-quality Tiled-C data and comprehensive interaction matrices and epigenetic maps for immune or inflammation-related loci in $CD4^+$ and $CD8^+$ T cells across the four conditions, I next addressed whether Tiled-C could help inform the functional basis of observed genetic associations in AS from GWAS. Specifically, I sought to delineate which gene(s) had the best evidence of physical interaction with disease-associated variants to help generate hypotheses as to the basis of GWAS association, identify whether chromatin contacts between promoters and genetic variants are responsive to stimulation with CD3/CD28 T cell activation beads, and if chromatin architecture in the stimulated $CD4^+$ and $CD8^+$ T cells were recapitulated in untreated and/or treated AS patients.

Across my Tiled-C regions, I found that the promoter of 13 genes (*IL-23R*, *SERPB1*, *IL12RB2*, *ICOS*, *CTLA4*, *PTGER4*, *BACH2*, *DEXI*, *CLEC16A*, *RMI2*, *KPBN1*, *TBKBP1*, and *TBX21*) across six tiled loci showed either stronger chromatin interactions (Greater ICE_n-RCP scores in the CD3/CD28-stimulated cells compared to the unstimulated cells) and/or the formation of contacts absent in the unstimulated cells with AS GWAS association signals (either lead GWAS SNPs or the 99% credible set) on CD3/CD28 stimulation (**Table 5.1**).

Chr[1]	Start[2]	End[3]	Locus Name[4]	Lead SNP[5]	Stimulated Interactions[6]	Tiled-C Figures[7]	
						$CD4^+$ T cells	$CD8^+$ T cells
1	153950000	155175000	*IL6R*	*rs4129267*	-	Fig. S.12	Fig. S.13
1	66950000	68300000	*IL12RB2-IL23R*	*rs11209026*	*IL-23R, SERPB1, IL12RB2* ($CD4^+$)	Fig. S.14	Fig. S.15
1	200300000	201425000	*C1orf106-(INAVA)-GPR25*	*rs41299637*	-	Fig. S.16	Fig. S.17
1	25060000	25680000	*RUNX3*	*rs6600247*	-	Fig. S.18	Fig. S.19
2	60600000	62850000	*REL-B3GNT2*	*rs35018800, rs12720356*	-	Fig. S.20	Fig. S.21
2	204075000	205075000	*ICOS-CTLA4-CD28*	*rs5837881*	*ICOS, CTLA4* ($CD4^+$)	Fig. S.22	Fig. S.23
3	45700000	46800000	*CCR2-CCR5*	*rs1001007*	-	Fig. S.24	Fig. S.25
3	27025000	28925000	*EOMES-CMC1*	*rs13093489*	-	Fig. S.26	Fig. S.27
5	39575000	41700000	*PTGER4*	*rs6908425*	*PTGER4* ($CD4^+$)	Fig. 5.9 (HV) and Fig. 5.11 (AS)	Fig. S.9
6	19279000	22675000	*CDKAL1*	*rs2328530, rs714830*	-	Fig. S.28	Fig. S.29
6	90001000	91600000	*BACH2*	*rs17765610*	*BACH2* (Both)	Fig. 5.12 (HV) and Fig. 5.14 (AS)	Fig. S.10
6	136800000	138399000	*IL22RA2-TNFAIP3*	*rs582757*	-	Fig. S.30	Fig. S.31
11	127725000	128950000	*ETS1*	*rs7933433*	-	Fig. S.32	Fig. S.33
12	111425000	113075000	*ATXN2-PTPN11-SH2B3*	*rs110658989*	-	Fig. S.34	Fig. S.35
16	10500000	11600000	*CIITA-DEXI-CLEC16A-SOCS1*	*rs367569*	*DEXI, CLEC16A, RMI2* (Both)	Fig. 5.15 (HV) and Fig. 5.17 (AS)	Fig. S.11
17	37250000	38575000	*IKZF3-Aiolos-ORMDL3*	*rs35736272*	-	Fig. S.36	Fig. S.37
17	45075000	46225000	*NPEPPS*	*rs9901869*	*KPNB1, TBKBP1, TBX21* ($CD4^+$)	Fig. S.38	Fig. S.39
18	12025000	13950000	*PTPN2*	*rs45450798*	-	Fig. S.40	Fig. S.41
19	9700000	11875000	*TYK2-ICAM1-ICAM3*	*rs35164067*	-	Fig. S.42	Fig. S.43
21	38950000	40875000	*ETS2*	*rs2836883*	-	Fig. S.44	Fig. S.45
22	20001000	23275000	*UBE2E3-MAPK1*	*rs2266961*	-	Fig. S.46	Fig. S.47

Table 5.1. Summary of Tiled-C regions and associated figures (Continued from previous page). Column one[1] shows the chromosome number of each Tiled-C locus. Column two[2] and three[3] show the start and end coordinates of the tiled regions. The fourth column[4] shows the assigned name of each locus. The fifth column[5] shows the lead SNP for each locus from Cortes *et al.* (2013) and/or Ellinghaus *et al.* (2016). The sixth column[6] shows the names of genes that show stimulation-associated looping interactions between their promoter region and either the AS-associated lead SNP or, where available, the 99% credible set from van de Bunt *et al.* (2015). The seventh column gives the figure number and location of each Tiled-C plot in $CD4^+$ and $CD8^+$ T cells.

Only four genes (*BACH2*, *DEXI*, *CLEC16A*, and *RMI2*) in two AS-associated loci were similarly identified in $CD8^+$ T cells. None of the Tiled-C loci showed substantial differences in their interaction profiles between the $CD4^+$ and $CD8^+$ T cells of untreated AS patients and adalimumab-treated AS patients. Three genes that show the strongest evidence of stimulation-associated *cis*-regulatory looping events (*PTGER4*, *BACH2*, and *DEXI*) with AS GWAS signals are illustrated as examples below.

5.4. Example Tiled-C regions

5.4.1. Upregulation of *PTGER4* in $CD4^+$ T cells on CD3/CD28 activation requires the coordination of multiple stimulation-specific enhancers

As previously mentioned in chapter four, the *PTGER4* locus has three SNPs with potential putative functional effect in T cells (section 4.3.2); two independent AS-association values at the SNPs *rs12186979* and *rs1992661*, and the functional *rs9283753* which effects *PTGER4* expression in lymphoblastoid cell lines. Contact was observed between the *PTGER4* promoter and *rs1992661* in the Capture-C data across both healthy volunteers and AS patients, but it was unclear if the interaction frequency varied between the two groups, and whether it influenced expression of *PTGER4*. The area of genetic association from the immunochip studies (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016) is clearly defined within a TAD downstream of the *PTGER4* promoter (chr8:40,280,000–

40,700,000). Tiled-C was used to assess the chromatin interactions at the *PTGER4* locus in greater detail and across stimulation conditions and following AS patient treatment. In $CD4^+$ T cells there are clear differences in the interaction profile of the *PTGER4* locus before and after CD3/CD28 stimulation (**Fig. 5.9**). Consistent with previous findings in chapter four, the interaction between the *PTGER4* promoter (**Fig. 5.9: i**) and *rs1992661* (**Fig. 5.9: ii**) can be seen; with stronger relative contact probabilities in the stimulated cells.

There is additionally an overall increase in interaction frequency in the area surrounding the SNP *rs9283753* (U10 EnhActv) (**Fig. 5.9: iii**) and *rs12186979* (U11 EnhWk) (**Fig. 5.9: iv**). The Tiled-C data also evidences an interaction between the *PTGER4* promoter and a region approximately 700 kbp upstream of the promoter (**Fig. 5.9: v**). This region is defined as quiescent by ChromHMM, suggesting the interaction between the two elements is not related to enhancer function. The ATAC-seq data (**Fig. 5.9: ●**) shows that the increase in interaction frequency is associated with significant increases in the overall accessibility of enhancer elements in the locus. *rs1992661* is accessible in both conditions, consistent with the unchanged chromosome looping interaction between this SNP and the promoter. In unstimulated cells, the chromatin at *rs9283753* and *rs12186979* remains inaccessible. On CD3/CD28 stimulation however, a significant increase in the accessibility of *rs9283753*, but not *rs12186979*, can be seen in the $CD4^+$ T cells accompanying the overall increase in interaction observed in the CD3/CD28-stimulated $CD4^+$ T cells. The most proximal downstream enhancer to *rs1992661* (**Fig. 5.9: vi**) – which intersects with a region of high relative contact probability with the *PTGER4* promoter – shows considerable and significant increases in chromatin accessibility. The interaction profiles of both unstimulated and CD3/CD28-stimulated $CD8^+$ T cells is highly similar to that seen in CD3/CD28-stimulated $CD4^+$ T cells (**Fig. S.9**); suggesting

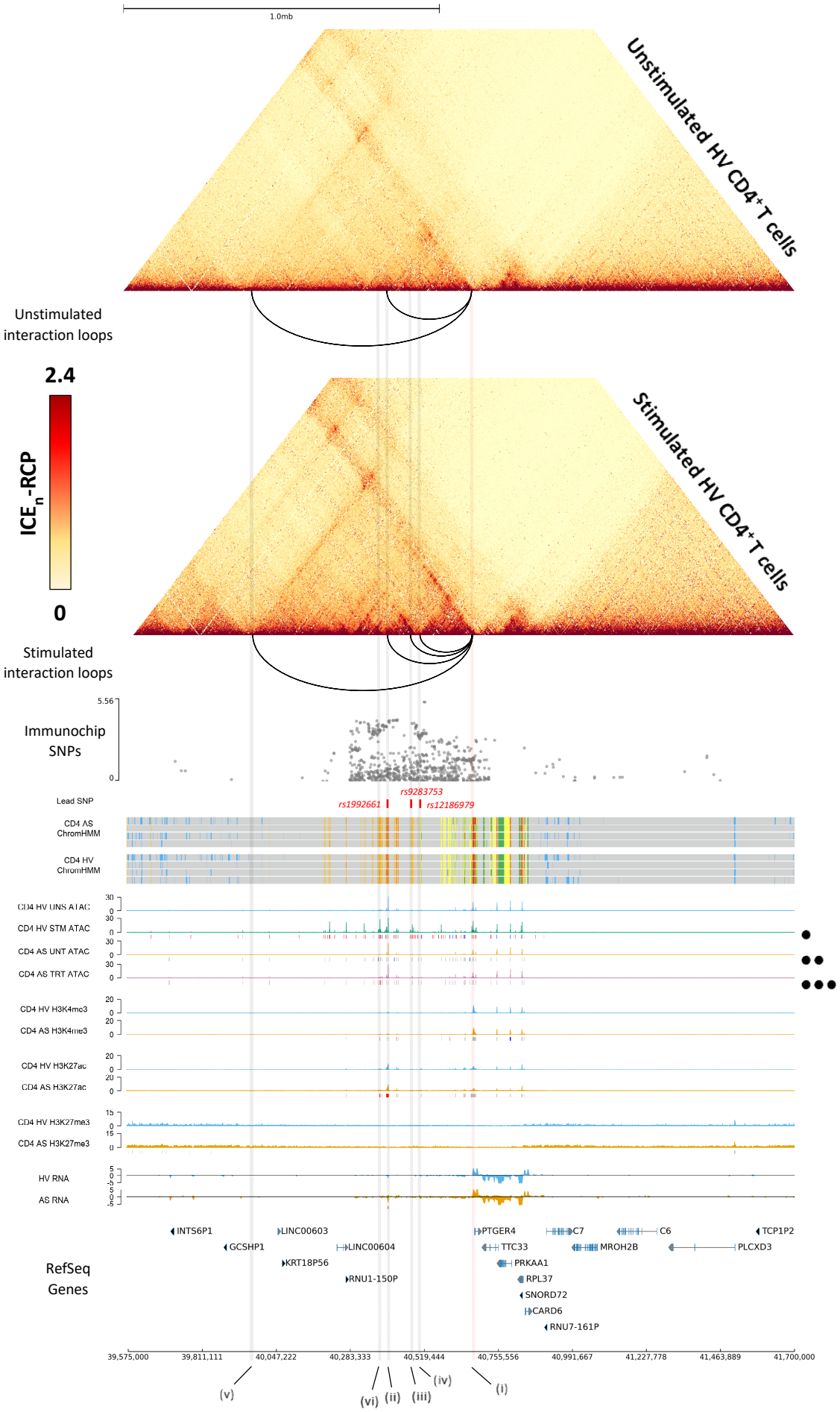
1.0mb
Unstimulated HV CD4+ T cells
Unstimulated interaction loops
2.4
ICE_n-RCP
0
Stimulated HV CD4+ T cells
Stimulated interaction loops
Immunochip SNPs
5.56
0
Lead SNP
rs9283753
rs1992661
rs12186979
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
RefSeq Genes
INTS6P1
LINC00603
PTGER4
C7
C6
TCP1P2
GCSHP1
LINC00604
TTC33
MROH2B
PLCXD3
KRT18P56
PRKAA1
RNU1-150P
RPL37
SNORD72
CARD6
RNU7-161P
39,575,000
39,811,111
40,047,222
40,283,333
40,519,444
40,755,556
40,991,667
41,227,778
41,463,889
41,700,000
(v)
(vi)
(ii)
(iii)
(iv)
(i)

Figure 5.9. Chromatin looping between stimulation-responsive genetic elements in CD4$^+$ T cells are observed in both unstimulated and CD3/CD28-stimulated healthy volunteer CD4$^+$ T cells but not with *rs199266* (Continued from previous page). From top to bottom: Mean Tiled-C interaction matrices (2 kbp resolution) for unstimulated (n = 3) and CD3/CD28-stimulated (n = 3) bulk CD4$^+$ T cells cultured from healthy volunteers. Matrices represent the ICE and total cis-read normalised relative contact probability (ICE_n-RCP). 'Unstimulated interaction loops' and 'stimulated interaction loops' show manually curated arcs showing suspected interactions with the *PTEGR4* promoter in unstimulated and stimulated healthy volunteers respectively. 'Immunochip SNPs' shows the $\log_{10}$ p-value score for all SNPs included in the Cortes *et al.* (2013) GWAS. SNPs that pass the genome-wide significant p-value threshold of 5×10^{-8} are coloured red. 'Lead SNP' shows the position and name of significant or suggestive lead SNPs identified by either Cortes *et al.* (2013) or Ellinghaus *et al.* (2016). 'CD4 AS ChromHMM' shows the 16-state ChromHMM from chapter three in AS patients, and 'CD4 HV ChromHMM' shows the equivalent in healthy volunteers. 'ATAC' tracks show mean ATAC-seq peaks from unstimulated healthy volunteers (n = 31; blue), stimulated healthy volunteers (n = 5; green), untreated AS patients (n = 21; orange), and adalimumab-treated AS patients (n = 20; pink). Boxes underneath the ATAC-seq tracks represent significantly differential (adjusted p-value ≤ 0.05; $\log_2$ fold change ≥ 0.585) (Red = upregulated peaks in second condition; Grey = no change between conditions; Blue = downregulated peaks in second condition). ● = unstimulated healthy volunteers versus stimulated healthy volunteers; ●● = unstimulated healthy volunteers versus untreated AS patients; ●●● = untreated AS patients versus treated AS patients. 'H3K4me3' tracks show mean H3K4me3 ChIPm-seq peaks from unstimulated healthy volunteers (n = 7; blue) and untreated AS patients (n = 7; orange). 'H3K27ac' tracks show mean H3K27ac ChIPm-seq peaks from unstimulated healthy volunteers (n = 7; blue) and untreated AS patients (n = 7; orange). 'H3K27me3' tracks show mean H3K27me3 ChIPm-seq peaks from unstimulated healthy volunteers (n = 4; blue) and untreated AS patients (n = 4; orange). 'RNA' tracks show mean RNA-seq peaks from unstimulated healthy volunteers (n = 24; blue) and untreated patients (n = 15; orange). Like ATAC-seq, boxes underneath each track represent differential epigenetic peaks, but only for peaks in untreated AS patients relative to unstimulated healthy volunteers (adjusted p-value ≤ 0.05; $\log_2$ fold change ≥ 0.585). 'RefSeq Genes': Position of genes from RefSeq. Scale bar at the bottom of figure shows genomic position of genetic features. Vertical lines depict regions of epigenetic interest: (i) *PTGER4* promoter; (ii) *rs1992661;* (iii) *rs928753;* (iv) *rs12186979;* (v) CD3/CD28-stimulation sensitive enhancers; (vi) Differential H3K27ac enhancer region. Locus shown between chr5:39,575,000-41,700,000.

these contacts are formed in $CD8^+$ T cells regardless of stimulation state.

Notably, the increased interaction frequency of these enhancer-promoter contacts and enhancer accessibility on stimulation was associated with a significant upregulation of *PTGER4* in $CD4^+$ T cells (adjusted p-value = 2.40×10^{-47}; $\log_2$ fold change = 1.79) (**Fig. 5.10**: **A**). *PTGER4* was also seen upregulated in CD3/CD28-stimulated $CD8^+$ T cells (adjusted p-value = 3.20×10^{-7}; $\log_2$ fold change = 0.99) (**Fig. 5.10: B**). *CARD6* in $CD4^+$ T cells (adjusted p-value = 5.96×10^{-14}; $\log_2$ fold change = -2.03;) and *TTC33* in both $CD4^+$ (adjusted p-value = 1.04×10^{4}; $\log_2$ fold change = 1.00) and $CD8^+$ T cells (adjusted p-value = 4.99×10^{-5}; $\log_2$ fold change = 0.98) were also differential; although this did not appear to coincide with increased interaction between the promoters of either gene and any of the previously identified regions in the contact matrices of unstimulated or CD3/CD28-stimulated $CD4^+$ or $CD8^+$ T cells.

I next sought to identify whether the contacts that coincided with increased *PTGER4* expression could also be seen in the $CD4^+$ T cell untreated AS patient and adalimumab-treated AS patient data **(Fig. 5.11**). The data shows that there is no *PTGER4* promoter (**Fig. 5.11: i**) and *rs1992661* (**Fig. 5.11: ii**) contact in either the untreated AS patients or the adalimumab-treated patients. Furthermore, the contact that is established on CD3/CD28 stimulation of $CD4^+$ T cells with the area of DNA encompassing *rs9283753* (**Fig. 5.11: iii**) and *rs12186979* (**Fig. 5.11: iv**) is absent in both untreated AS patients and adalimumab-treated AS patients.

Contact between the *PTGER4* promoter and the 700 kbp upstream element (**Fig. 5.11: v**) is apparent in both matrices, but is much stronger and more comparable to the contacts frequencies seen in the healthy volunteer data in adalimumab-treated AS patients, suggesting this interaction may be attenuated in untreated AS patients. The accessibility

of chromatin in the locus in the untreated AS patients is similar to unstimulated healthy volunteers (**Fig. 5.11: ●●**) and to adalimumab-treated AS patients (**Fig. 5.11: ●●●**), with most stimulation-associated interacting enhancers seen in the unstimulated healthy volunteer data remaining inaccessible before and after treatment. Consistent with the hypothesis that these contacts may be required for *PTGER4* upregulation, *PTGER4* expression is not significantly different in either $CD4^+$ or $CD8^+$ T cells when comparing either unstimulated healthy volunteers to untreated AS patients (**Fig. 5.10: C & D**) or untreated AS patients to adalimumab-treated AS patients (**Fig. 5.10: E & F**). Overall, these results suggest that the stronger contacts seen between *rs12186979*, *rs9283753*, and *rs1992661* and the *PTGER4* promoter on CD3/CD28 stimulation may contribute to *PTGER4* regulation in activated $CD4^+$ and $CD8^+$ T cells.

5.4.2. *rs17765610* forms CD3/CD28 stimulation-associated contacts with the BACH2 promoter that are absent in AS patient cells from pre and post-adalimumab treatment for three months

In chapter four, Capture-C was used to analyse the interaction profile at the chr6q15 locus between the promoter of *BACH2* (Broad complex-tramtrack-bric a brac and Cap'n'collar Homology 2) and two independent AS-association signals centred on *rs17765610* (genome-wide significant; p-value = 5.3×10^{-8}) and *rs639575* (suggestive significance; p-value = 8.6×10^{-5}) (Cortes *et al.*, 2013) (**Fig. S.5**). Additionally, the 99% credible set of AS disease association with *rs17765610* is limited to a 30 kb region that overlaps introns five and six and exons five, six, and seven of *BACH2* (van de Bundt *et al.*, 2015). In $CD4^+$ and $CD8^+$ T cells, strong interactions were observed over the entire *rs17765610* 99% credible set – which sits primarily in a weak enhancer element (U11 EnhWk) and – the *BACH2* promoter, but not with *rs639575* (also defined as U11 EnhWk).

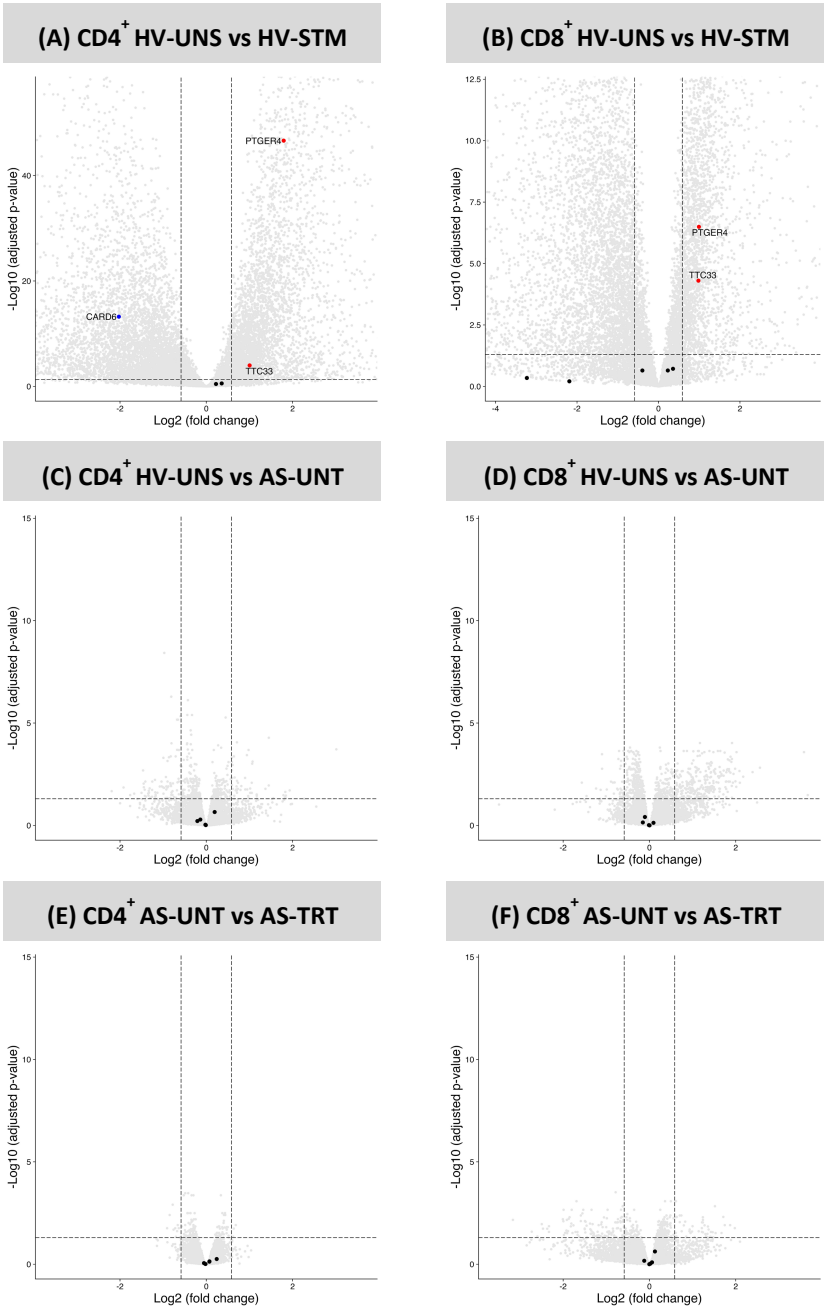

Figure 5.10. Expression of *PTGER4* is significantly upregulated in CD4$^+$ and CD8$^+$ T cells on CD3/CD28 stimulation but not in AS patients before or after treatment with adalimumab. Volcano plots showing differential gene expression in the *PTGER4* Tiled-C locus. Upregulated (red) and downregulated (blue) genes in stimulated vs unstimulated cells are indicated. Adjusted p-value (0.05) and log2 fold change (0.585) significance thresholds are shown. (A) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer CD4$^+$ T cells. (B) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer CD8$^+$ T cells. (C) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) CD4$^+$ T cells. (D) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) CD8$^+$ T cells. (E) Untreated (n = 10) versus adalimumab-treated (n = 10) AS patient CD4$^+$ T cells. (F) Untreated (n = 11) versus adalimumab-treated (n = 11) AS patient CD8$^+$ T cells.

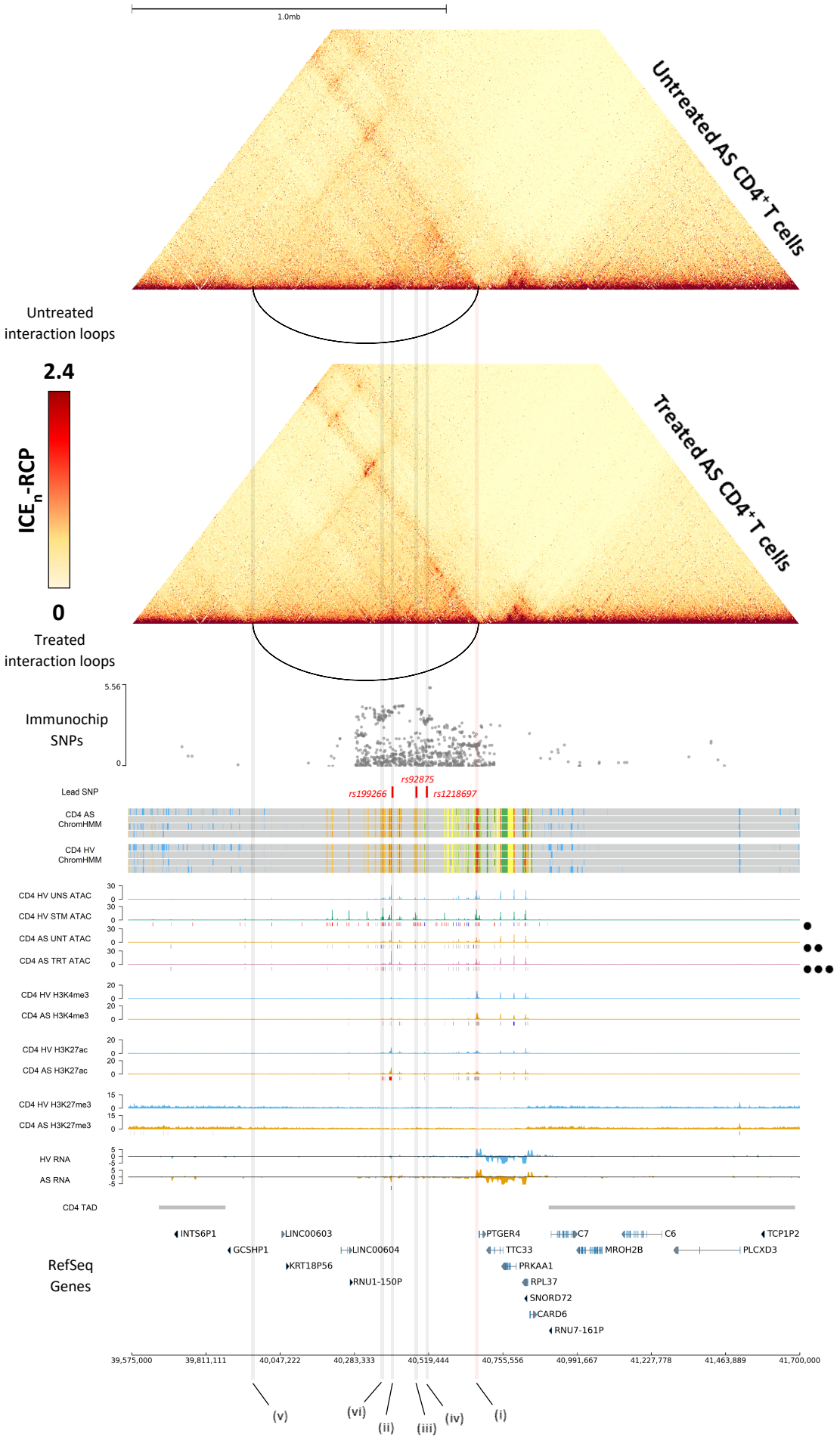
1.0mb
Untreated AS CD4+ T cells
Untreated interaction loops
2.4
ICE_n-RCP
0
Treated AS CD4+ T cells
Treated interaction loops
Immunochip SNPs
5.56
0
Lead SNP
rs92875
rs199266
rs1218697
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
CD4 TAD
RefSeq Genes
INTS6P1
LINC00603
PTGER4
C7
C6
TCP1P2
GCSHP1
LINC00604
TTC33
MROH2B
PLCXD3
KRT18P56
PRKAA1
RNU1-150P
RPL37
SNORD72
CARD6
RNU7-161P
39,575,000
39,811,111
40,047,222
40,283,333
40,519,444
40,755,556
40,991,667
41,227,778
41,463,889
41,700,000
(v)
(vi)
(ii)
(iii)
(iv)
(i)

Figure 5.11. AS patients pre and post-adalimumab treatment do not show looping interactions with *rs1992661, rs928753,* or *rs12186979* (Continued from previous page). Mean Tiled-C interaction matrices (2 kbp resolution) for untreated (n = 3) and adalimumab-treated (n = 3) bulk $CD4^+$ T cells isolated from AS patients. Matrices represent the ICE and total cis-read normalised relative contact probability (ICE_n-RCP). 'Untreated interaction loops' and 'treated interaction loops' show manually curated arcs showing suspected interactions with the *PTEGR4* promoter in unstimulated and stimulated healthy volunteers respectively. See figure 5.11 on pages 249-250 for full detail of accessory tracks. Vertical lines depict regions of epigenetic interest: (i) *PTGER4* promoter; (ii) *rs1992661;* (iii) *rs928753;* (iv) *rs12186979;* (v) CD3/CD28-stimulation sensitive enhancers; (vi) Differential H3K27ac enhancer region. Locus shown between chr5:39,575,000-41,700,000.

Analysis of the Tiled-C data of healthy volunteer $CD4^+$ T cells shows that the interaction between the *BACH2* promoter (**Fig. 5.12: i**) and *rs17765610* (**Fig. 5.12: ii**) is maintained in both the unstimulated and CD3/CD28-stimulated cells; although higher interaction frequency scores are observed in the stimulated cells. In the CD3/CD28-stimulated $CD4^+$ T cells however, the *BACH2* promoter and *rs17765610* interacts strongly with an internal downstream promoter element (U2 TSSFlnkD) (**Fig. 5.12: iii**). This interaction is not observed in the unstimulated $CD4^+$ T cells. Both the *BACH2* promoter and *rs17765610* additionally interact with a distal upstream block of active enhancers (**Fig. 5.12: iv**) with far greater relative contact probability scores in the CD3/CD28-stimulated $CD4^+$ T cells compared to the unstimulated cells. ATAC-seq data shows the internal promoter element and the upstream block of enhancers were associated with significant increases in chromatin accessibility in the CD3/CD28-stimulated cells (**Fig. 5.12: ●**).

Interaction between *rs639575* (**Fig. 5.12: v**) and the *BACH2* promoter could not be determined in $CD4^+$ T cells due to their relatively close proximity; with any potential interaction being occluded by the Brownian motion of local DNA. Overall similar interaction results were observed in the $CD8^+$ T cell data for unstimulated and CD3/CD28-stimulated $CD8^+$ T cells (**Fig. S.10**).

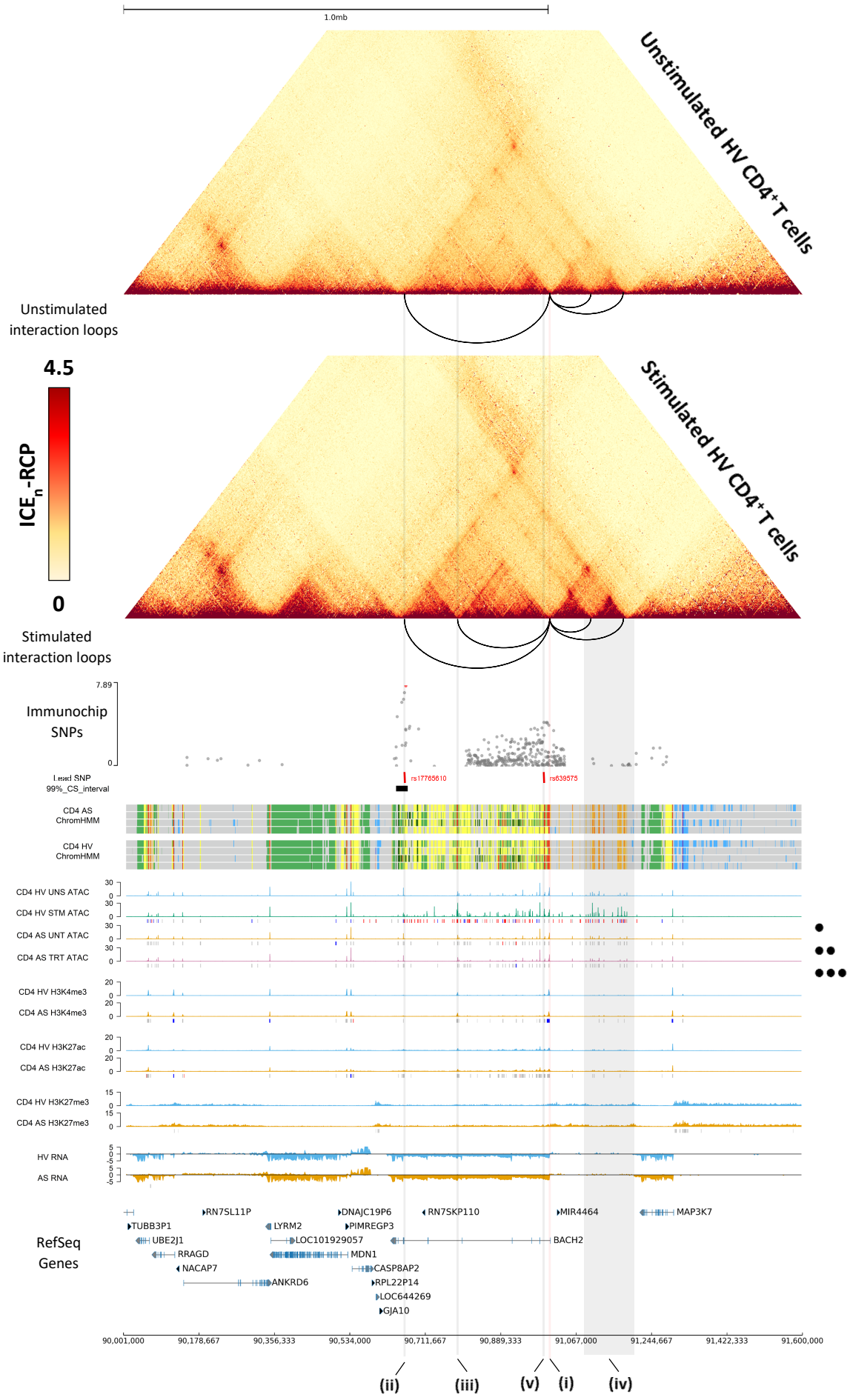
1.0mb
Unstimulated HV CD4+ T cells
Unstimulated interaction loops
4.5
ICE_n-RCP
0
Stimulated HV CD4+ T cells
Stimulated interaction loops
7.89
Immunochip SNPs
0
Lead SNP 99%_CS_interval
rs17765610
rs639575
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
RefSeq Genes
RN7SL11P
DNAJC19P6
RN7SKP110
MIR4464
MAP3K7
TUBB3P1
LYRM2
PIMREGP3
UBE2J1
LOC101929057
BACH2
RRAGD
MDN1
NACAP7
CASP8AP2
ANKRD6
RPL22P14
LOC644269
GJA10
90,001,000
90,178,667
90,356,333
90,534,000
90,711,667
90,889,333
91,067,000
91,244,667
91,422,333
91,600,000
(ii)
(iii)
(v)
(i)
(iv)

Figure 5.12. Stimulation of CD4$^+$ T cells with CD3/CD28 T cell activation beads is associated with the formation of stronger chromatin contacts between the promoter of *BACH2*, *rs17765610*, and an internal promoter element (Continued from previous page). Mean Tiled-C interaction matrices (2 kbp resolution) for unstimulated (n = 3) and CD3/CD28-stimulated (n = 3) bulk CD4$^+$ T cells cultured from healthy volunteers. 'Unstimulated interaction loops' and 'stimulated interaction loops' show manually curated arcs showing suspected interactions with the *BACH2* promoter in unstimulated and stimulated healthy volunteers respectively. See figure 5.11 on pages 249-250 for full detail of accessory tracks. Vertical lines depict regions of epigenetic interest: (i) *BACH2* promoter; (ii) *rs17765610;* (iii) downstream internal promoter element*;* (iv) distal block of stimulation-associated enhancers*;* (v) *rs639575*. Locus shown between chr6:90,001,000-91,600,000.

Despite the clear formation of these enhancer contacts with the *BACH2* promoter across both bulk T cell populations however, the expression of *BACH2* was not differential between unstimulated healthy volunteer and CD3/CD28-stimulated CD4$^+$ (**Fig. 5.13: A**) or CD8$^+$ (**Fig. 5.13: B**) T cells. Five other genes were also found to be differential across the *BACH2* locus in CD4$^+$ and CD8$^+$ T cells on CD3/CD28 stimulation: *RRGAD* and *MDN1* were significantly upregulated and *CASP8AP2*, *ANKRD6*, and *GABRR2* were significantly downregulated in both cell types. *GJA10* was additionally significantly downregulated only in the CD4$^+$ T cells. Although there was a general increase in relative contact frequency across genes when comparing unstimulated CD4$^+$ and CD8$^+$ T cells to CD3/CD28-stimulated cells, no contact was observed to intersect with either *rs639575* or *rs17765610*.

I next sought to identify whether these interactions could instead influence the expression of *BACH2* in AS patients. Notably, neither the untreated AS patients nor the adalimumab-treated patients showed convincing interaction between the *BACH2* promoter (**Fig. 5.14: i**), *rs17765610* (**Fig. 5.14: ii**), or the downstream promoter element (**Fig. 5.14: iii**). *rs17765610* itself does appear to interact with the upstream block of enhancer elements (**Fig. 5.14: iv**), although the overall relative contact probabilities are far weaker than can

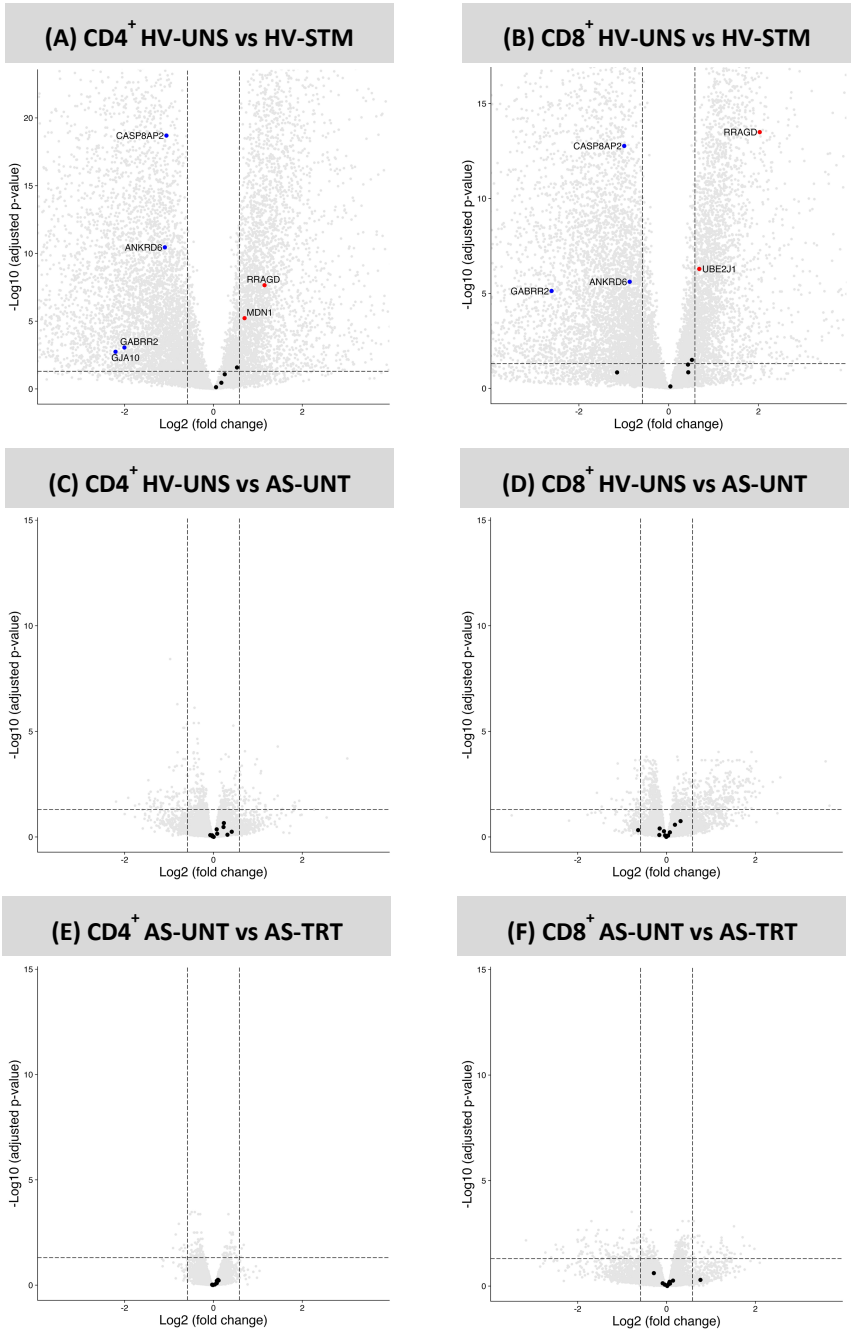

Figure 5.13. Expression of *BACH2* is not differential on stimulation with CD3/CD28 activation beads, in AS or on treatment with adalimumab. Volcano plots showing differential gene expression in the *BACH2* Tiled-C locus. Upregulated (red) and downregulated (blue) genes in stimulated vs unstimulated cells are indicated. Adjusted p-value (0.05) and log2 fold change (0.585) significance thresholds are shown. (A) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer $CD4^+$ T cells. (B) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer $CD8^+$ T cells. (C) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) $CD4^+$ T cells. (D) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) $CD8^+$ T cells. (E) Untreated (n = 10) versus adalimumab-treated (n = 10) AS patient $CD4^+$ T cells. (F) Untreated (n = 11) versus adalimumab-treated (n = 11) AS patient $CD8^+$ T cells.

be seen in the unstimulated healthy volunteers and CD3/CD28-stimulated healthy volunteers. Consistent with these findings, no significant difference in chromatin accessibility can be seen either between unstimulated healthy volunteers and untreated AS patients (**Fig. 5.14: ●●**) or untreated AS patients and adalimumab-treated AS patients (**Fig. 5.14: ●●●**). Intersecting these features with the ChIPmentation-seq and RNA-seq data from both untreated AS patients and unstimulated healthy volunteers show that the *BACH2* promoter shows significantly lower H3K4me3 deposition in untreated AS patients compared to healthy volunteers. No other region however showed differential H3K4me3, H3K27ac, H3K27me3, or eRNA marks, suggesting a high degree of epigenetic similarity between untreated AS patients and unstimulated healthy volunteer $CD4^+$ T cells. These results are similarly observed in the untreated and adalimumab-treated $CD8^+$ T cell data (**Fig. S.10**).

Like between unstimulated healthy volunteers and CD3/CD28-stimulated healthy volunteers, *BACH2* expression (and expression of the other differential genes seen on CD3/CD28 stimulation) is not differentially expressed when comparing unstimulated healthy volunteers and untreated AS patients (**Fig. 5.13: C & D**) or untreated AS patients and adalimumab-treated AS patients (**Fig. 5. 13: E and F**).Taken together, these findings suggest that the region encompassing the 99% credible set and the SNP *rs17765610* interact more strongly with the *BACH2* promoter on T cell activation. This interaction may however be absent in the $CD4^+$ and $CD8^+$ T cells of AS patients.

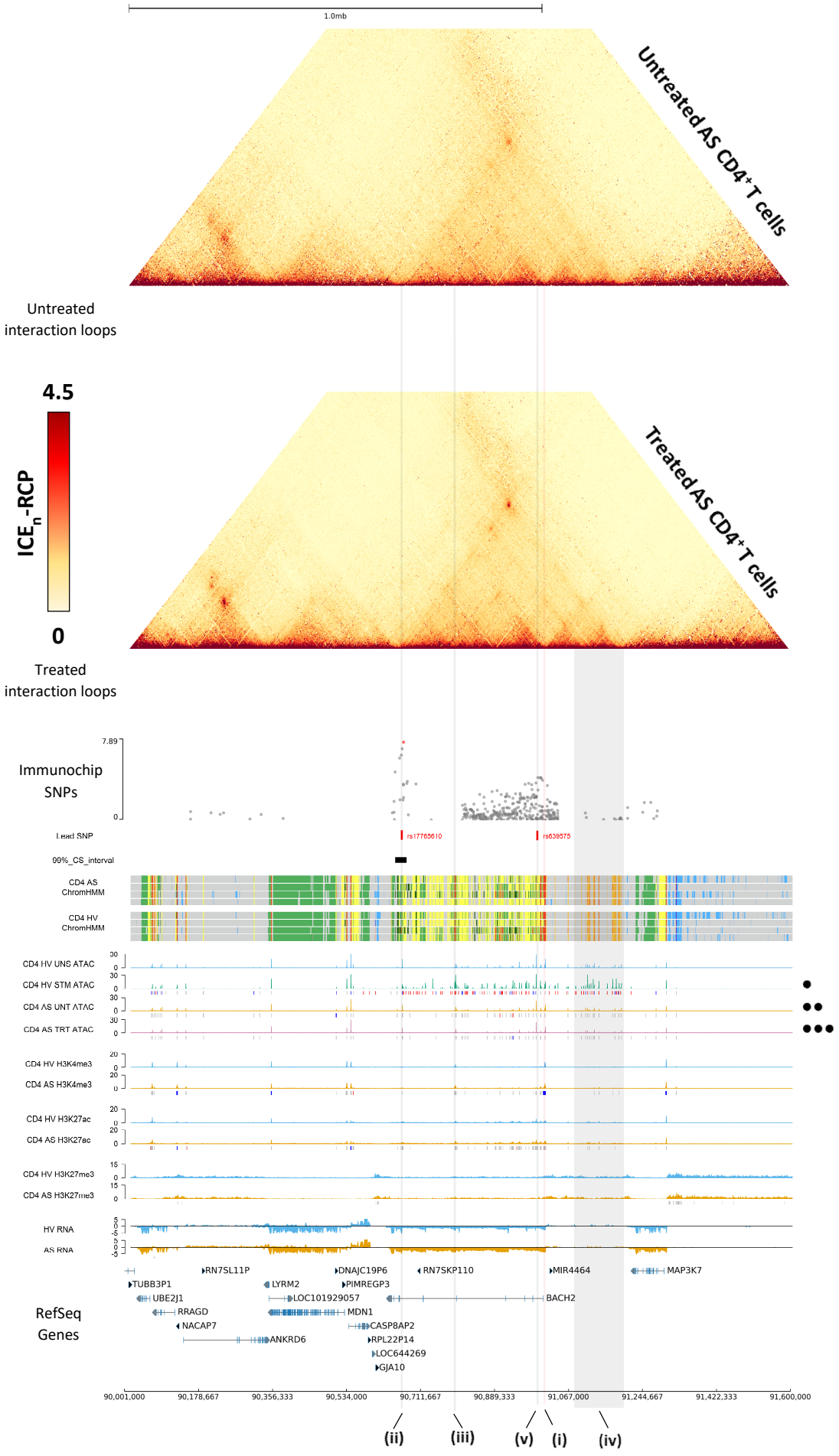
1.0mb
Untreated AS CD4+ T cells
Untreated interaction loops
4.5
ICE_n-RCP
0
Treated AS CD4+ T cells
Treated interaction loops
Immunochip SNPs
7.89
0
Lead SNP
rs17765610
rs639575
99%_CS_interval
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
RefSeq Genes
RN7SL11P
DNAJC19P6
RN7SKP110
MIR4464
MAP3K7
TUBB3P1
LYRM2
PIMREGP3
UBE2J1
LOC101929057
BACH2
RRAGD
MDN1
NACAP7
CASP8AP2
ANKRD6
RPL22P14
LOC644269
GJA10
90,001,000
90,178,667
90,356,333
90,534,000
90,711,667
90,889,333
91,067,000
91,244,667
91,422,333
91,600,000
(ii)
(iii)
(v)
(i)
(iv)

Figure 5.14. Interaction profiles at the *BACH2* locus in untreated and adalimumab-treated AS patient cells are similar to unstimulated healthy volunteers (Continued from previous page). Mean Tiled-C interaction matrices (2 kbp resolution) for untreated (n = 3) and adalimumab-treated (n = 3) bulk $CD4^+$ T cells isolated from AS patients. 'Untreated interaction loops' and 'treated interaction loops' show manually curated arcs showing suspected interactions with the *BACH2* promoter in unstimulated and stimulated healthy volunteers respectively. See figure 5.11 on pages 249-250 for full detail of accessory tracks. Vertical lines depict regions of epigenetic interest: (i) *BACH2* promoter; (ii) *rs17756510;* (iii) downstream internal promoter element*;* (iv) distal block of stimulation-associated enhancers*;* (v) *rs639575*. Locus shown between chr6:90,001,000-91,600,000.

5.4.3. An active enhancer element in close proximity to *rs367569* interacts more strongly with the promoter of *DEXI* on CD3/CD28 stimulation

Whilst *PTGER4* and *BACH2* do show interaction between the lead SNP(s) and target gene promoters, some of the loci show increased interaction outside of the lead SNP signal on CD3/CD28 stimulation. For example, the *CIITA-DEXI-CLEC16A-SOCS1* locus was tiled between the coordinates of 10,500,000-11,600,000 of chromosome 16. Although Cortes *et al.* (2013) did not identify a genome-wide significant association value in the locus, there is a genome-wide significant AS-association identified by Ellinghaus *et al.* (2016) centred on the lead SNP *rs367569* (adjusted p-value = 5.03×10^{-14}). Analysis of the Tiled-C data in healthy volunteer $CD4^+$ T cells shows that there are no clear interactions between *rs367569* (**Fig. 5.15: i**) and any gene promoter region across the locus in either unstimulated or CD3/CD28-stimulated cells. ChromHMM shows that *rs367569* lies in a region of quiescent DNA (U14 Quies) in both healthy volunteers and AS patients, suggesting the SNP may not be functional in $CD4^+$ T cells. Instead, an interaction can be seen between the shared promoter of *DEXI* (dexamethasone-induced protein) and *CLEC16A* (C-type lectin domain family 16) (**Fig. 5.15: ii**) and a block of enhancers approximately 10 kbp downstream of *rs367569* (**Fig. 5.15: iii**).

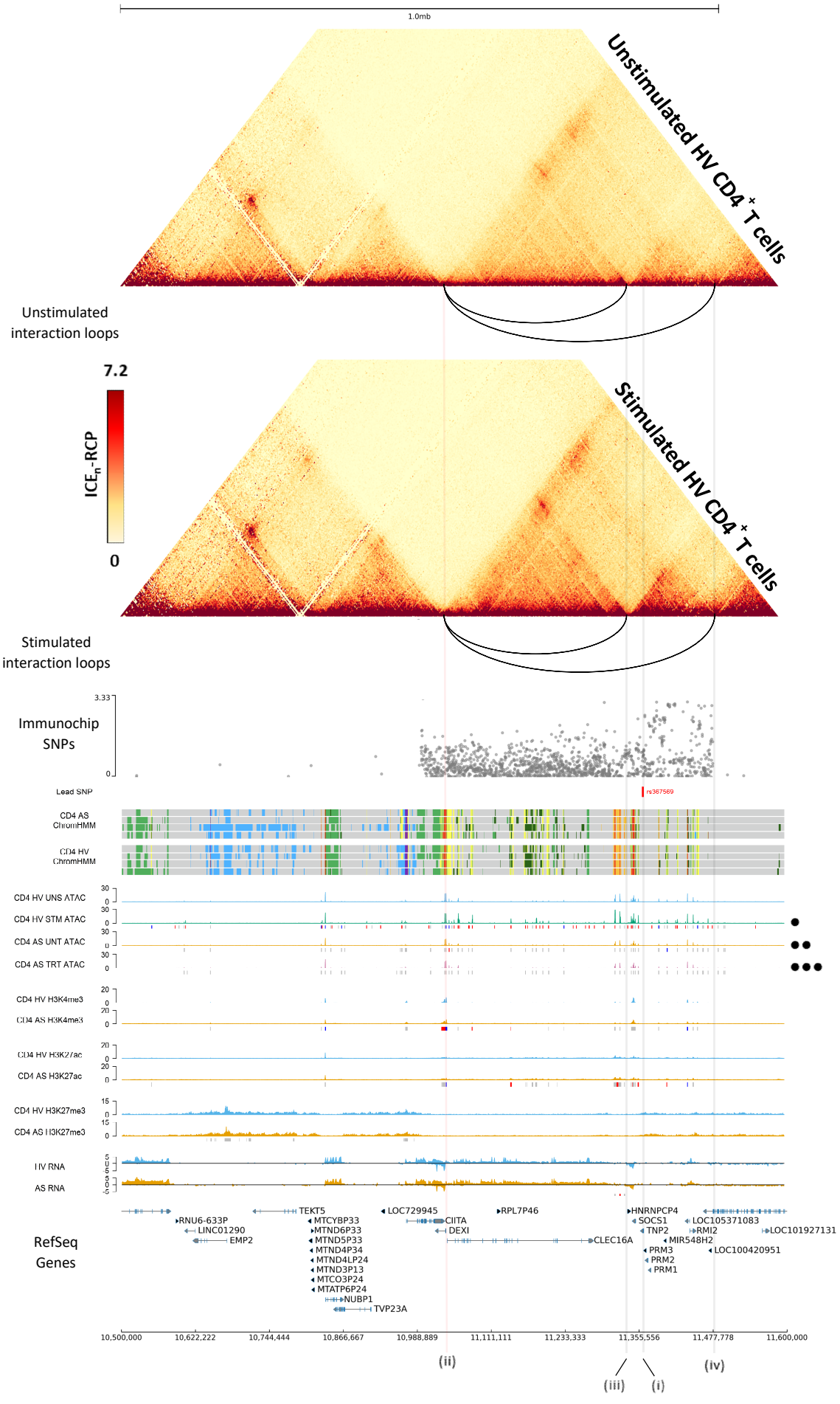
1.0mb
Unstimulated HV CD4+ T cells
Unstimulated interaction loops
7.2
ICE_n-RCP
0
Stimulated HV CD4+ T cells
Stimulated interaction loops
Immunochip SNPs
3.33
0
Lead SNP
rs367569
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
RefSeq Genes
RNU6-633P
LINC01290
EMP2
TEKT5
MTCYBP33
MTND6P33
MTND5P33
MTND4P34
MTND4LP24
MTND3P13
MTCO3P24
MTATP6P24
NUBP1
TVP23A
LOC729945
CIITA
DEXI
RPL7P46
CLEC16A
HNRNPCP4
SOCS1
TNP2
PRM3
PRM2
PRM1
LOC105371083
RMI2
MIR548H2
LOC100420951
LOC101927131
10,500,000
10,622,222
10,744,444
10,866,667
10,988,889
11,111,111
11,233,333
11,355,556
11,477,778
11,600,000
(ii)
(iii)
(i)
(iv)

Figure 5.15. Increased interaction between the shared promoter of *DEXI* and *CLEC16A* and active enhancer elements in CD4$^+$ T cells on CD3/CD28 stimulation (Continued from previous page). Mean Tiled-C interaction matrices (2 kbp resolution) for untreated (n = 3) and adalimumab-treated (n = 3) bulk CD4$^+$ T cells isolated from AS patients. 'Unstimulated interaction loops' and 'stimulated interaction loops' show manually curated arcs showing suspected interactions with the *DEXI-CLEC16A* promoter in unstimulated and stimulated healthy volunteers respectively. See figure 5.11 on pages 249-250 for full detail of accessory tracks. Vertical lines depict regions of epigenetic interest: (i) *rs367569*; (ii) *DEXI-CLEC16A* promoter; (iii) *DEXI-CLEC16A* interacting enhancer elements; (iv) quiescent DNA interaction. Locus shown between chr12: 111,425,000 to 113,075,000.

Contact is also established between the promoter and a region of DNA approximately 120 kbp upstream of *rs367569* (**Fig. 5.15: iv**) although this region too constitutes quiescent DNA. Whilst these contacts are established in both unstimulated and CD3/CD28-stimulated healthy volunteer CD4$^+$ T cells, the relative contact probability values between the two regions are much stronger in the stimulated cells. Similar differences in interaction frequency can also be observed in the healthy volunteer CD8$^+$ T cells (**Fig. S.11**). As in *PTGER4* and *BACH2*, the increased interaction frequency between these elements coincides with locus-wide increases in chromatin accessibility (**Fig. 5.15: ●**).

Unlike the previous two examples, *DEXI* (but not *CLEC16A*) is significantly downregulated on CD3/CD28 stimulation in both CD4$^+$ (adjusted p-value = 4.39 x 10^{-2}; $\log_2$ fold change = -0.62) and CD8$^+$ (adjusted p-value = 3.47 x 10^{-2}; $\log_2$ fold change = -0.69) T cells (**Fig. 5.16: A & B**). The genes *CIITA*, *AC099489.1*, *ATF7IP2*, and *TVP23A* were also downregulated and *SOCS1* upregulated in both cell types. *RMI2* was upregulated specifically in CD4$^+$ T cells, whilst *NUBP1* was upregulated and *EMP2* was upregulated in CD8$^+$ T cells. Changes in the interaction profile between the promoters of these genes and other *cis*-regulatory elements however could not be determined by the Tiled-C data.

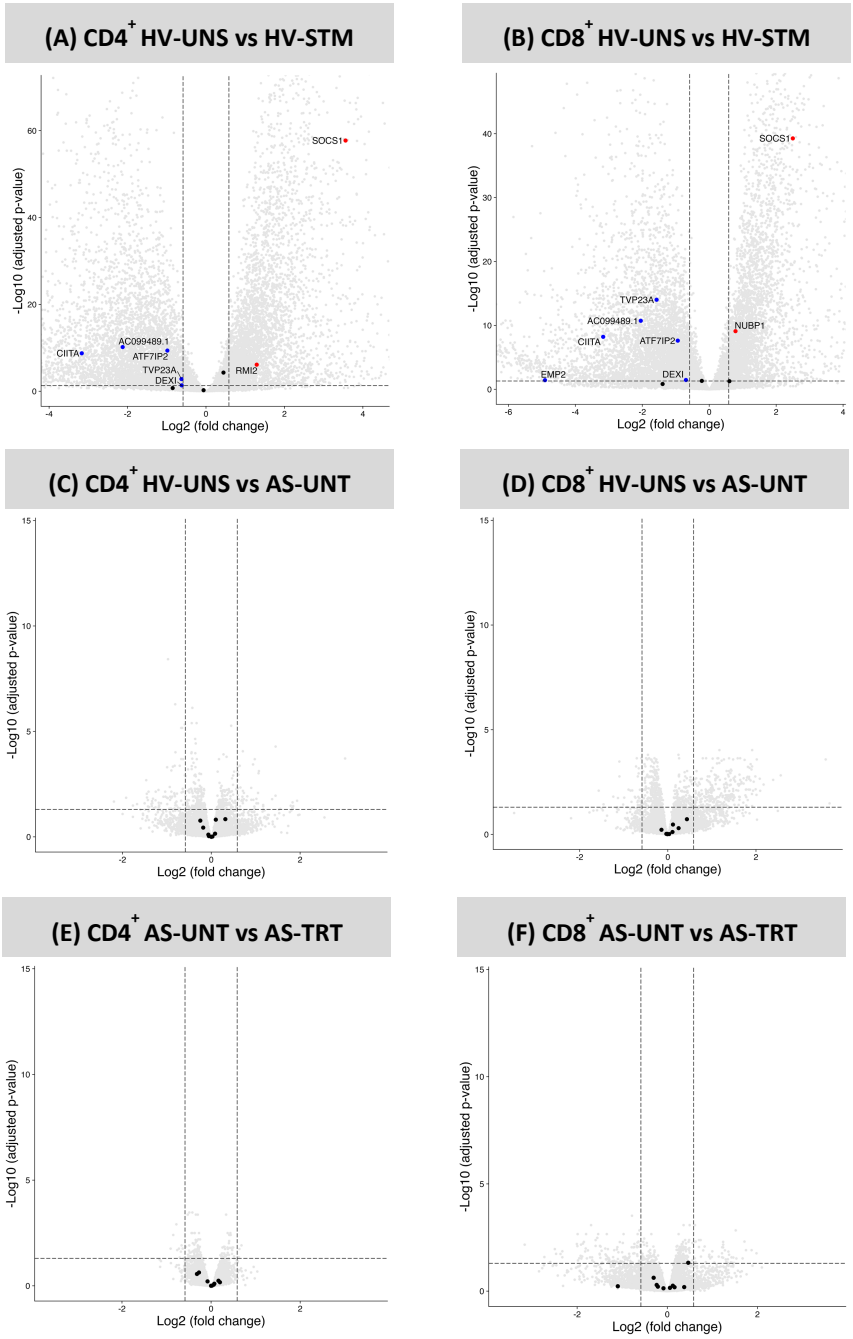

Figure 5.16. Expression of *DEXI* is differential on stimulation with CD3/CD28 activation beads but not in AS or on treatment with adalimumab. Volcano plots showing differential gene expression in the *CIITA-DEXI-CLEC16A-SOCS1* Tiled-C locus. Upregulated (red) and downregulated (blue) genes in stimulated vs unstimulated cells are indicated. Adjusted p-value (0.05) and log2 fold change (0.585) significance thresholds are shown. (A) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer $CD4^+$ T cells. (B) Unstimulated (n = 18) versus CD3/CD28-stimulated (n = 5) healthy volunteer $CD8^+$ T cells. (C) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) $CD4^+$ T cells. (D) Unstimulated healthy volunteer (n = 13) versus untreated AS patient (n = 12) $CD8^+$ T cells. (E) Untreated (n = 10) versus adalimumab-treated (n = 10) AS patient $CD4^+$ T cells. (F) Untreated (n = 11) versus adalimumab-treated (n = 11) AS patient $CD8^+$ T cells.

In AS patients, the interaction profiles of both untreated and adalimumab-treated AS patient $CD4^+$ T cells are highly similar to both each other and those seen in unstimulated healthy volunteers (**Fig. 5.17**). Like in the *PTGER4* and *BACH2* loci, chromatin accessibility is not significantly different between unstimulated healthy volunteer and untreated AS patients (**Fig. 5.17: ●●**) or untreated and adalimumab-treated AS patients (**Fig. 5.17: ●●●**). There is however significant downregulation of the H3K4me3 and H3K27ac signal at the *DEXI-CLEC16A* promoter (**Fig. 5.17: ii**) and upregulation of H3K27ac at the downstream block of enhancers (**Fig. 5.17: iii**) between unstimulated healthy volunteers and untreated AS patients. Despite these changes, none of the genes found in the locus are differential when comparing unstimulated healthy volunteers to untreated AS patients (**Fig. 5.16: C & D**) or untreated AS patients to adalimumab-treated AS patients (**Fig. 5.16: E & F**) across $CD4^+$ and $CD8^+$ T cells. These findings are similarly found in the $CD8^+$ T cell AS patient Tiled-C data (**Fig. S.11**).

5.5. Chapter discussion

In this chapter, I have described the generation of high-quality chromatin interaction maps of 24 AS-associated GWAS loci using Tiled-C in primary peripheral blood $CD4^+$ and $CD8^+$ T cells for both healthy volunteers (cultured cells, both unstimulated and stimulated with CD3/CD28 T cell activation beads) and AS patients (sampled from patients before and 3-months following first administration of anti-TNFα therapeutics). I have further shown their application to identify how looping interactions between ChromHMM-characterised regulatory elements that also show differential epigenetic marks could contribute to the altered expression of their associated genes, and identified three genes in particular (*PTEGR4*, *BACH2*, and *DEXI*) that show increased interaction with AS association signals and/or regulatory elements on T cell activation.

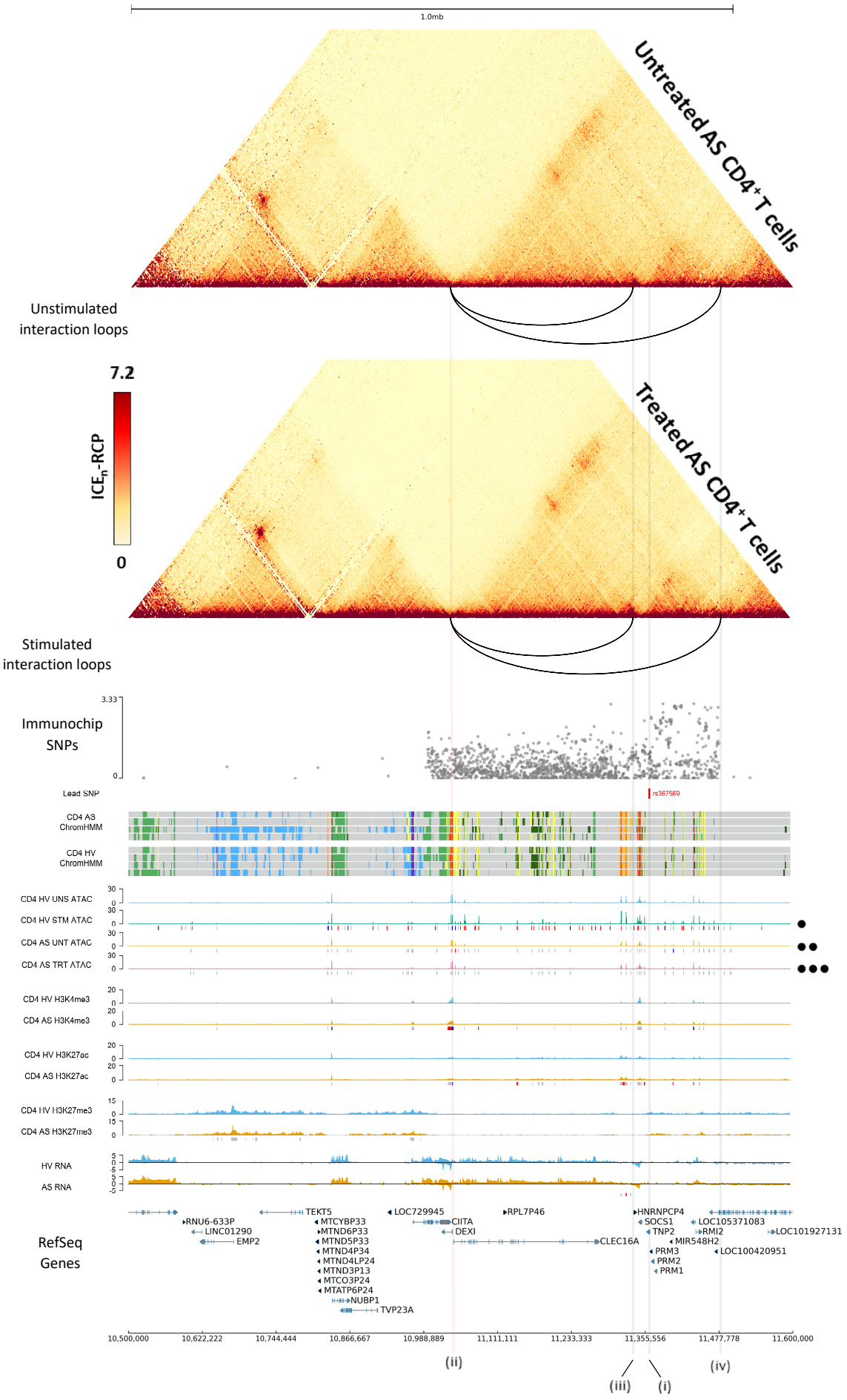
1.0mb
Untreated AS CD4+ T cells
Unstimulated interaction loops
7.2
ICE_n-RCP
0
Treated AS CD4+ T cells
Stimulated interaction loops
Immunochip SNPs
3.33
0
Lead SNP
rs367569
CD4 AS ChromHMM
CD4 HV ChromHMM
CD4 HV UNS ATAC
CD4 HV STM ATAC
CD4 AS UNT ATAC
CD4 AS TRT ATAC
CD4 HV H3K4me3
CD4 AS H3K4me3
CD4 HV H3K27ac
CD4 AS H3K27ac
CD4 HV H3K27me3
CD4 AS H3K27me3
HV RNA
AS RNA
RefSeq Genes
RNU6-633P
LINC01290
EMP2
TEKT5
MTCYBP33
MTND6P33
MTND5P33
MTND4P34
MTND4LP24
MTND3P13
MTCO3P24
MTATP6P24
NUBP1
TVP23A
LOC729945
CIITA
DEXI
RPL7P46
CLEC16A
HNRNPCP4
SOCS1
TNP2
MIR548H2
PRM3
PRM2
PRM1
LOC105371083
RMI2
LOC100420951
LOC101927131
10,500,000
10,622,222
10,744,444
10,866,667
10,988,889
11,111,111
11,233,333
11,355,556
11,477,778
11,600,000
(ii)
(iii)
(i)
(iv)

Figure 5.17. Interaction profiles at the *CIITA-DEXI-CLEC16A-SOCS1* locus in untreated and adalimumab-treated AS patient cells are similar to unstimulated healthy volunteers (Continued from previous page). Mean Tiled-C interaction matrices (2 kbp resolution) for untreated (n = 3) and adalimumab-treated (n = 3) bulk $CD4^+$ T cells isolated from AS patients. 'Untreated interaction loops' and 'treated interaction loops' show manually curated arcs showing suspected interactions with the *DEXI-CLEC16A* promoter in unstimulated and stimulated healthy volunteers respectively. See figure 5.11 on pages 249-250 for full detail of accessory tracks. Vertical lines depict regions of epigenetic interest: (i) *rs367569*; (ii) *DEXI-CLEC16A* promoter; (iii) *DEXI-CLEC16A* interacting enhancer elements; (iv) quiescent DNA interaction. Locus shown between chr12: 111,425,000 to 113,075,000.

5.5.1. Generation of comprehensive looping interaction maps at AS-associated GWAS risk loci

Whilst identification of enhancers carrying disease-associated variants alone is useful in identifying which of the variants are putatively causal, establishing which of these enhancers contact specific gene promoters in different cellular contexts across AS-associated loci could yield important insight into which *cis*-regulatory elements might be the primary contributor towards AS pathogenesis. The first aim of this chapter was to develop comprehensive maps of chromatin looping interactions across $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes that could be used alongside chromatin accessibility, gene expression, and histone modification data to first link SNPs and enhancer elements to genes and then to identify enhancers that could be further investigated with functional assays (i.e. CRISPR-based gene editing approaches or subcloning and enzyme reporter assays).

5.5.1.1. Data quality

Before the main analysis, I first performed quality control on the Tiled-C data to ascertain whether it could be used to identify enhancer-promoter interactions. All of the $CD4^+$ and $CD8^+$ T cell Tiled-C libraries showed excellent overall quality control metrics with an

overall *cis*-to-*trans* ratio exceeding 70%. Despite similar *cis*-to-*trans* ratios being seen in the $CD14^+$ monocyte AS patient samples, the ratios achieved in the healthy volunteer samples were far lower. This suggests possible issues specifically with the healthy volunteer $CD14^+$ monocyte culture conditions that the AS patient samples were not subjected to, and the analysis of interactions in $CD14^+$ monocytes was excluded from this chapter.

Certain subsets of $CD14^+$ monocytes – such as $CD14^+CD16^{low}$ monocytes which constitute the majority of the MACS-purified $CD14^+$ cell fraction (Brown *et al.*, 2023) – have been shown to be prone to apoptosis both spontaneously and in response to increase oxidative stress in culture (Zhao *et al.*, 2010; Lage *et al.*, 2022). Monocyte apoptosis is associated with the internal fragmentation of nucleosomal DNA (Fahy *et al.*, 1999; Atkin-Smith *et al.*, 2020). This could help explain the increased *trans* reads in the healthy volunteer samples pre and post-stimulation with CD3/CD28 beads, as the increased number of apoptotic DNA fragments will also be digested in the nucleus and randomly interact with DNA fragments produced from fixed protein complexes. Although the quality of the T cell data is high based on *cis*-to-*trans* read ratios, we have also shown that $CD14^+$ monocytes show greater epigenetic differences between unstimulated healthy volunteers and untreated AS patients (Brown *et al.*, 2023). The AS patient Tiled-C data from $CD14^+$ monocytes could therefore yield important insights into how these differences influence target genes and into disease pathogenesis. In future, Tiled-C libraries from uncultured, unstimulated healthy volunteer $CD14^+$ monocytes should be prepared similarly to the AS patient libraries in order to avoid possible apoptosis in culture.

Cis-to-*trans* ratios are a good quality metric, but do not provide insight into how well interactions were captured. The *MYC* locus was therefore included as a control region to

investigate whether I had captured known interaction profiles. Aside from the extensive study of the regulatory landscape surrounding *MYC* with CRISPRi (Fulco *et al.*, 2019), *MYC* is known to upregulate on TCR activation where it aids in promoting metabolic reprogramming of activated T cells (i.e. promoting glycolysis and glutaminoolysis pathways) (Wang *et al.*, 2011; Preston *et al.*, 2015; Ma *et al.*, 2024). *MYC* was therefore chosen as an excellent control locus to (a) ascertain the quality of the Tiled-C data and whether it could identify validated chromatin interactions and (b) investigate how the regulatory elements may react to stimulation with CD3/CD28 T cell activation beads.

On investigation of the Tiled-C data, I was able to detect five CD3/CD28 responsive regulatory elements that intersected with active ChromHMM enhancer states and increased chromatin accessibility in stimulated $CD4^+$ and $CD8^+$ T cells. It should be noted however that Fulco *et al.* (2019) identified seven total enhancer elements in their CRISPRi experiments, and four of these peaks aligned poorly with the interaction data. This could be due to the considerable epigenetic differences between the myeloid, antigen-presenting K562 cell lines used by the authors and the primary, lympoid naïve $CD4^+$ and $CD8^+$ T cells used in this chapter (Ito *et al.*, 2024). Reassuringly though, three of these regulatory elements intersected with the regions of highest CRISPRi score. This suggests the Tiled-C data was able to detect context-specific looping events with experimentally validated contributions to *MYC* regulation, and further highlighted the overall high quality of the $CD4^+$ and $CD8^+$ T cell interaction data.

5.5.1.2. Comparison of conditions

The Tiled-C work carried out in this chapter in primary $CD4^+$ and $CD8^+$ T cells is novel. Very little has been done in previous studies to physically link regions of AS association to target genes and even less has been done to understand how the structure of chromatin

at these loci responds to immune stimulus or treatment; which correspond to aims two and three of this chapter. Of the 24 captured AS-associated regions, six of them showed interaction between at least one gene promoter and an AS-associated lead SNP signal. Two of these loci – *BACH2* and *PTGER4* – showed convincing increases in the strength of these signals on $CD4^+$ and $CD8^+$ T cell activation with CD3/CD28 beads. The *CIITA-DEXI-CLEC16A-SOCS1* locus also showed increased interaction frequencies in the CD3/CD28-stimulated cells, but did not interact directly with the AS lead SNP signal; although interactions with the *DEXI-CLEC16A* promoter were found in close proximity to the lead SNP. Interestingly, the differences between untreated AS patient, adalimumab-treated AS patient, and unstimulated healthy volunteer cells were far subtler. Indeed, the majority of loci showed near identical interaction profiles between the three conditions. This is not entirely surprising, as we have previously shown that the epigenetic differences between healthy volunteer and AS patient $CD4^+$ and $CD8^+$ T cells are relatively small (Brown *et al.*, 2023).

The epigenetic data of healthy volunteer $CD4^+$ and $CD8^+$ T cells stimulated with CD3/CD28 were the most divergent of the four conditions. Strikingly, the ATAC-seq data showed that chromatin accessibility almost unanimously increases significantly across the AS-associated loci. Most of these ATAC-seq peaks cannot be observed in the unstimulated healthy volunteer cells, untreated AS patient cells, or adalimumab-treated AS patient cells. A plurality of the stimulation-associated ATAC-seq peaks also intersected regions predicted to have *cis*-regulatory function by ChromHMM; suggesting that activation of $CD4^+$ and $CD8^+$ T cells with CD3/CD28 beads results in the opening of stimulation-specific regulatory elements. This is further evidenced by the RNA-seq data, which showed that hundreds of genes in $CD4^+$ (526) and $CD8^+$ (494) T cells were differentially expressed on CD3/CD28 activation. These findings corroborate those of

Zhang *et al.* (2023), who also used ATAC-seq, RNA-seq, and high resolution Hi-C to investigate how chromatin architecture differs in murine Th1, Th2 and Th17 activated T cell subsets – which are seen expanded in AS cases (Liu *et al.*, 2021) – compared to naïve circulating $CD4^+$ T cells The authors found that induction of naïve $CD4^+$ T cells with CD3/CD28 beads and further differentiation to Th1 cell or Th17 cells with their associated cytokines lead to wide-scale TAD rearrangement and both gains and losses in enhancer promoter contacts unique to each cell subset. This was most pronounced at genes associated with functions specific to activated T cells such as *Rorc* (which encodes the Th17 lineage specifying transcription factor RORγt), where the promoter was shown to interact with a group of enhancer elements 80 kbp downstream of the promoter that was absent in naïve $CD4^+$ T cells. Comparison of Hi-C profiles shown that $CD8^+$ T cell activation is also associated with extensive chromatin remodelling when comparing naïve cells to effector and memory subsets (Russ *et al.*, 2023). The findings in this chapter have highlighted that the effects of some AS-associated SNPs may only be investigated in activated $CD4^+$ and $CD8^+$ T cells, and that the reconfiguration of chromatin at *BACH2*, *PTGER4*, and *DEXI* may better reflect the epigenetic landscape present in disease-associated cell types such as Th1 and Th17 cells.

5.5.2. Considerations for AS pathogenesis

5.5.2.1. The SNPs *rs1992661* and *rs9283753* both interact with the PTGER4 promoter on CD3/CD28 stimulation

As mentioned in chapter four (section 4.3.2), *PTGER4* is a GPCR involved in pro-inflammatory signalling cascades in $CD4^+$ and $CD8^+$ T cells that recognises the PGE2 ligand (section 1.2.5). *PTGER4* has also been found critical in the differentiation of Th17 cells, with deletion of the gene preventing the accumulation of IL-17^+ T cells in the skin

of psoriatic patients (Lee *et al.*, 2019). Genetic variants that interfere with the regulation of *PTGER4* expression in activated T cells may therefore bias $CD4^+$ T cells towards specific pro-inflammatory subtypes. The Capture-C data in chapter four showed a looping interaction between the *PTGER4* promoter and the SNP *rs1992661*, but not *rs9283753* or *rs12186979.* Whilst the contact between the *PTGER4* promoter and *rs1992661* could also be seen in the unstimulated healthy volunteer Tiled-C data, it also showed that both *rs9283753* and *rs12186979* additionally interact with the *PTGER4* promoter on CD3/CD28 stimulation. All of which coincided with increased *PTGER4* expression in $CD4^+$ and $CD8^+$ T cells.

The effects of *rs1992661* and *rs12186979* on *PTGER4* expression have not been functionally investigated, but *rs9283753* is an eQTL for *PTGER4* expression in blood (https://genetics.opentargets.org/Variant/5_40490507_C_T/association) (Tewhey *et al.*, 2016). Cap analysis of gene expression by the FANTOM5 consortium has also revealed that the transcription level – a marker of enhancer activity – of the enhancer containing *rs9283753* correlates with the increased *PTGER4* TSS activity in activated T cells (Noguchi *et al.*, 2017). The SNP intersects consensus binding motifs for the transcription factors CEBP and CREB, and homozygosity for the AS risk-associated T allele increases *PTGER4* expression in lymphoblastoid cell lines (Tewhey *et al.*, 2016). The data presented in this chapter did not find strong evidence for *rs12186979* functionality, as the chromatin at the intersecting active enhancer element remained inaccessible – as evidenced by ATAC-seq – in both the unstimulated and stimulated healthy volunteer cells. *rs1992661* was found to lie in an active enhancer region of significantly higher looping interaction-associated H3K27ac and eRNA signals in untreated AS patients compared to healthy volunteers; suggesting the enhancer activity of this region may be higher in AS patients compared to healthy volunteers. Taken together, these findings

highlight putative functional roles for the regulatory elements underlying *rs9283753* and *rs1992661* towards *PTGER4* regulation in AS.

PTGER4 has recently been identified as an important therapeutic target in osteoarthritis. Targeting the GPCR with the antagonist HL-43 has been shown to inhibit downstream *PTGER4*-realted signalling cascades (i.e. the phosphoinositol and MAPK pathways) in osteoclasts (Jiang *et al.*, 2022) of osteoarthritis patients has resulted in reduction of angiogenesis and nerve growth at the site of inflammation; helping alleviate pain in osteoarthritis patients in a fashion different to NSAIDs. Pain is the second-most prevalent quality of life concern amongst AS patients, with greater than 80% of patients identifying it as an area for therapeutic improvement (Ward, 1999; Bagcivan *et al.*, 2015). Increased *PTGER4* expression on the surface of Th17 cells in AS patients is also correlated to increased BASDAI scores and patient pain (Klasen *et al.*, 2019). Although more studies on *PTGER4* in AS are required to prioritise the gene as a bonafide therapeutic target – particularly around how *rs9283753* and *rs1992661* effect *PTGER4* expression in primary T cells – compounds like HL-43 may provide alternate analgesic avenues for AS treatment.

5.5.2.2. *BACH2* expression may be modulated by *rs17756510* in activated T cells

BACH2 encodes an eponymous transcription factor with putative pathogenic functional roles in AS. *BACH2* has been found to be a key regulator of $CD4^+$ T cell differentiation. Increased *BACH2* expression has been shown to inhibit the transition of naïve cells to effector Th1, Th2, and Th17 cells whilst promoting Treg-associated transcriptional responses and controlling immune system activation (Roychodhuri *et al.*, 2013; Yang *et al.*, 2019; Sidwell *et al.*, 2020). In $CD8^+$ T cells, *BACH2* expression critically inhibits transcriptional programs associated with T cell exhaustion and protects long-term

effector functionality (Yao *et al.*, 2021). *BACH2* is also a key inhibitor of T cell activation in $CD8^+$ T cells, with deletion of *BACH2* resulting in a shift to more activated phenotypes and DNA configurations in naïve primary $CD8^+$ T cells (Russ *et al.*, 2023). Contact between the lead SNP *rs17756510* and the BACH2 promoter was much stronger on stimulation of both $CD4^+$ and $CD8^+$ T cells with anti-CD3/CD28 beads. The effects of the risk-associated G allele compared to the C allele of *rs17756510* on the expression of any gene in the locus has yet to be established. Upregulation or downregulation of *BACH2* driven by *rs17756510* could however influence the activation efficacy of T cells and the speed they adopt pro-inflammatory phenotypes. Indeed – although no such studies have been conducted in AS patients or other SpAs – expression levels of *BACH2* within the Th1, Th2, and Th17 cells of SLE patients are inversely proportional with SLE disease activity index scores and pro-inflammatory cytokines (Long *et al.*, 2023).

It should be noted however that *BACH2* expression was not differential between any of the conditions despite the increased contact between *rs17756510* and the gene promoter. This suggests that CD3/CD28 stimulation alone may not be sufficient to upregulate or downregulate *BACH2* expression. In $CD4^+$ T cells for instance, *BACH2* has been shown to upregulate in response to stimulation together with CD3/CD28, IL-6, and IL-21 (Long *et al.*, 2023); suggesting that *BACH2* may require input from multiple independent signalling cascades to affect gene expression. As IL-6 and IL-21 also promote the development and expansion of Th17 cells (Nurieva *et al.*, 2007; Shi *et al.*, 2019; Harbour *et al.*, 2020), how the CD3/CD28-stimulated contact between *rs17756510* and the *BACH2* promoter is affected by IL-6 and IL-21 should be a priority for future functional investigation.

BACH2 is not only associated with AS, but a plethora of other inflammatory conditions including IBD, SLE, T1D, RA, and MS. This makes *BACH2* an attractive therapeutic

target for inflammatory or autoimmune conditions as a whole (Zwick et al., 2024). Investigation into how *BACH2* contributes to these conditions however is still in its infancy, and no compounds capable of directly modulating the activity of *BACH2* have yet been identified. Combined with its pleiotropic and often contradictory role across multiple distinct lineages of immune cells, this makes *BACH2* currently difficult to design therapeutics for. *BACH2* does however appear to be highly dependent on *HDAC3* (histone deacetylase) for transcription factor binding. Treatment with the *HDAC3* inhibitor RGFP966 has been shown to significantly downregulate *BACH2* expression in activated B cells (Inaba *et al.*, 2023). The authors also found that decreased *BACH2* occupancy at enhancers in the IL-10 locus lead to increased expression of immunosuppressive IL-10 from RGFP966-treated B cells; which attenuated the TNFα response of peripheral $CD4^+$ T cells when co-cultured. It is therefore clear that *BACH2* is a multifaceted transcription factor with influence in AS extending passed $CD4^+$ and $CD8^+$ T cells, and future studies should incorporate more immune cell types in order to fully understand the role of *BACH2* in disease.

5.5.2.3. Expression of *DEXI* may be controlled in activated T cells by differential AS-associated *cis*-regulatory elements

CD3/CD28 stimulation induced greater interaction frequency between the shared *DEXI-CLEC16A* promoter and active enhancer elements in close proximity to the lead SNP *rs367569*. ChromHMM predicted *rs367569* to lie within a region of quiescent DNA; suggesting the SNP may not be a causal variant for the locus in T cells. The *DEXI-CLEC16A* promoter and its interacting enhancers however showed differential *cis*-regulatory activity between unstimulated healthy volunteer and untreated AS patient $CD4^+$ T cells: with the *DEXI-CLEC16A* promoter showing simultaneous downregulation and upregulation of H3K4me3 signals and the enhancers upregulation of eRNA and

H3K27ac signals. These contacts appeared to specifically downregulate *DEXI* expression and not *CLEC16A*, as the expression of the latter was not significantly different between unstimulated healthy volunteers and CD3/CD28-stimulated T cells. *DEXI* is of unknown function in T cells. In pancreatic cells however, abrogating *DEXI* expression has been shown to reduce activation of inflammation-associated transcription factors (i.e. STAT1) and the expression of proinflammatory cytokines and chemokines (Dos Santos *et al.*, 2019). Differential expression of *DEXI* in AS patients may therefore alter the local inflammatory milieu and the recruitment of other immune cells to the site of inflammation.

DEXI is also a candidate gene in a number of other autoimmune conditions such as T1D (Tomlinson *et al.*, 2014), MS (Zuvich *et al.*, 2011), and SLE (Zhang *et al.*, 2011). The NHGRI-EBI GWAS catalogue (Buniello *et al.*, 2019) shows that 35 SNPs have been associated with the broad definition of 'autoimmune conditions' in the *CIITA-DEXI-CLEC16A-SOCS1* Tiled-C region (https://www.ebi.ac.uk/gwas/efotraits/EFO_0005140). None of these SNPs however show high LD ($R^2 \geq 0.8$) with *rs367569*. As in any of the AS-associated regions, there are likely to be multiple mechanisms in the locus that predispose disease pathogenesis. Indeed, whilst the looping interactions described here in primary peripheral blood $CD4^+$ and $CD8^+$ T cells are novel, looping interactions have been described between the SNPs *rs12708716* and *rs8062322* (not in the NHGRI-EBI GWAS catalogue) in intron 19 of *CLEC16A* in $CD14^+$ monocytes (Dos Santos *et al.*, 2019). This interaction however is noticeably absent in both $CD4^+$ and $CD8^+$ T cells before and after CD3/CD28 stimulation, and the T cell interaction absent in the $CD14^+$ monocyte looping data. Both of these SNPs are found to be associated with T1D and MS but not AS, and are in poor LD with the AS-associated lead SNP *rs367569* (**Fig. 5.15**).

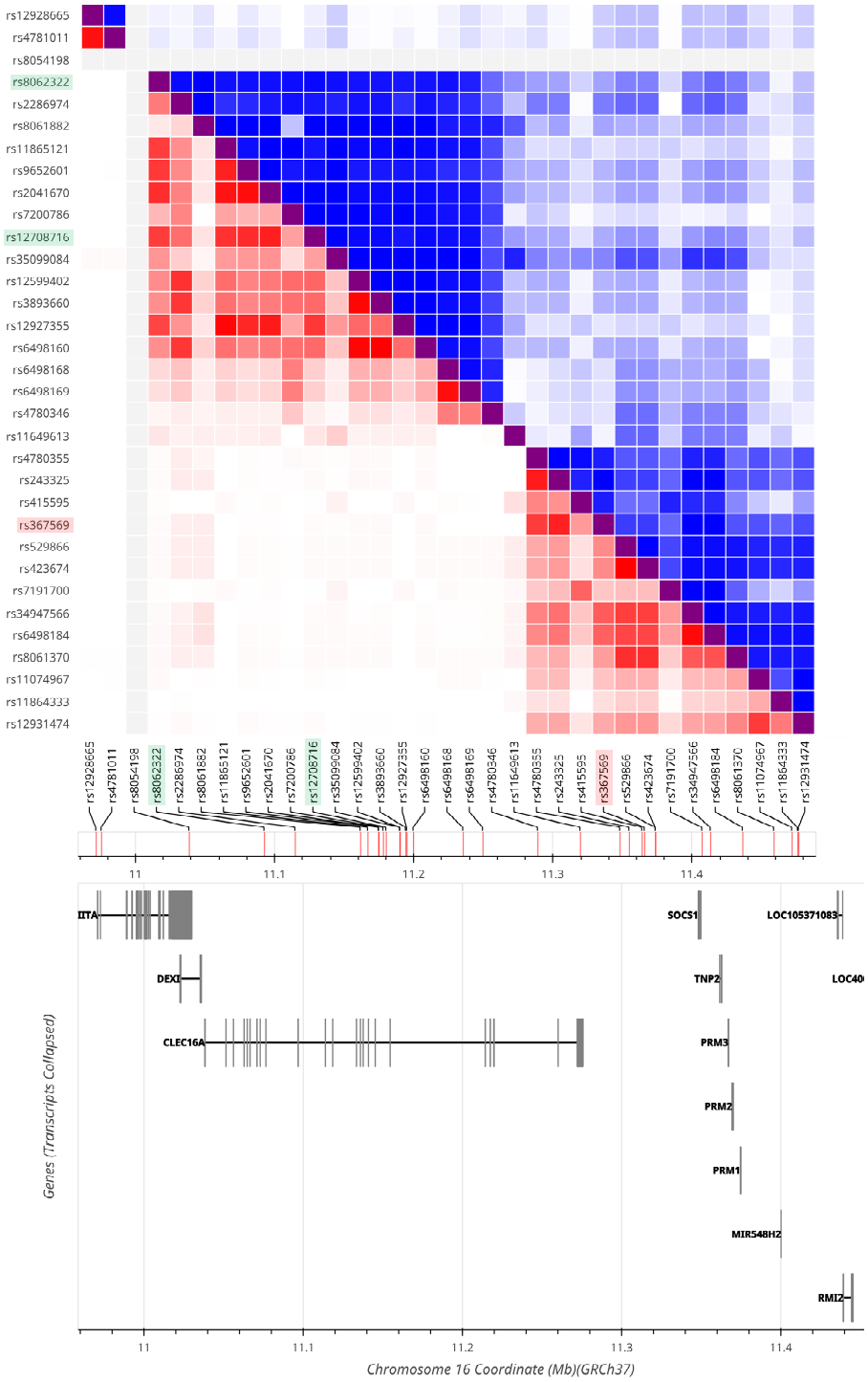

Fig 5.18. Linkage disequilibrium between autoimmune SNPs in the *CIITA-DEXI-CLEC16A-SOCS1* Tiled-C region. Heatmap showing LD between all SNPs currently associated with an autoimmune condition from the NHGRI-EBI GWAS catalogue between the coordinates of 10,500,000 to 11,600,000 on chromosome 16 in a EUR(GBR) population. Two distinct blocks of LD can be seen separated by the SNP centred on the SNP *rs11649613*. The two SNPs with looping interactions in CD14^{+} monocytes (*rs12708716* and *rs8062322*) are highlighted in green, and the AS-associated lead SNP (*rs367569*) is highlighted in red. Data analysed using LDmatrix from LDlink 5.0 (www.ldlink.nci.nih.gov).

Although the Tiled-C data evidences strengthened promoter-enhancer contacts between the *DEXI-CLEC16A* promoter and enhancers with differential epigenetic marks in AS, further fine-mapping of the GWAS signal centred on *rs367569* will be needed to help identify putatively causal SNPs that may lie in the interacting regions. Functional characterisation of many genes in the locus, such as *DEXI* and *CLEC16A*, will also need to performed to ascertain the therapeutic value of the locus in AS.

5.5.3. Study limitations and future work

5.5.3.1. Cellular epigenetics may differ between AS flares and patients in remission and at the site of inflammation

The differences between AS patient $CD4^+$ and $CD8^+$ T cells isolated pre and post-adalimumab therapy were minimal compared to the same cells from unstimulated healthy volunteer cultures. The findings presented from AS patients presented in this chapter may be limited by isolating the cells under variable inflammatory contexts. AS is a condition that is in reality characterised by disease flares. Flares typically involve temporary, but significant, worsening of AS symptoms and clinical manifestations over base disease activity (Stone *et al.*, 2008). As many as 75% of AS patients will experience a flare, which last on average for four days with mean recurrence every 35 days (Barnett *et al.*, 2021). Current consensus suggests that AS flares can occur in response to a variety of stimuli, including microbial gut stress or in response to the use of specific biologics or their altered therapeutic dosage (Sharif et al., 2020; Kwon *et al.*, 2021; Yeo *et al.*, 2024). Due to the lack of formal definitions, tests, or biomarkers that could help identify AS in flare, the number of studies investigating the molecular and cellular causes of AS flares are limited and they remain poorly understood. Flares in related conditions (such as IBD) are associated with significantly increased inflammatory T cell signatures

(among many other cell types such as $CD14^+$ monocytes and $CD19^+$ B cells) the peripheral blood and at the site of inflammation (Rubin *et al.*, 2019). Cells taken during an AS flare may therefore have better reflected the activated epigenetic phenotypes observed in the CD3/CD28-stimulated $CD4^+$ and $CD8^+$ T cells.

Sampling from AS patients in flare was considered during this study, however neither BASDAI scores nor MRI imaging are currently effective ways to diagnose an AS flare without further refinement; with high overall specificity but low sensitivity (Zheng *et al.*, 2021; Constantino *et al.*, 2022). BASDAI scores require an average back pain score over the preceding week – at which point most flares would have resolved – and MRI suffer from lack of immediate availability (most patients would receive appointments too late into a flare) and are relatively expensive. Whilst every patient recruited in this chapter used for functional epigenetic analysis has definitive and active AS following the most current diagnostic criteria, these patients could represent a mixed population of individuals who are experiencing active flares and those in remission and who reflect variable inflammatory statuses. It is important to note that the levels of the broad-range inflammatory blood marker CRP varied in my recruited AS patients (Range = 0.8 to 78.5; Mean = 26.3; SEM = 18.2) (**Table 2.1**); suggesting that some patients had far higher circulating inflammatory statuses than others. Whilst CRP does correlate poorly with AS activity, CRP should be considered in the recruitment criteria in future studies in order to establish a baseline inflammatory level for better comparison between the recruited individuals, and capture AS cases when they may be in their most inflammatory state.

Furthermore, as mentioned in chapter four (section 4.5.3) – and aside from the significant differences between cell subtypes, developmental stages, and stimulation and/or activation – the epigenetics of tissue-resident immune cells (including $CD4^+$ and $CD8^+$ T cells) are significantly different to circulating immune cells. Correspondingly, Tiled-C

performed on T cells isolated from the site of inflammation in AS (i.e. sacroiliac joint biopsies or spinal synovial fluid) may show different interaction profiles across the baited regions. To date, no study has however investigated how epigenetics may differ between peripheral circulating and tissue-resident immune cells in AS. Future work sampling cells from tissue biopsies and using single-cell-based technologies will help elucidate whether research should be focused on PBMCs (as is currently applied) or on the tissue-resident cells (Simone *et al.*, 2021). It should also be noted that *ex-vivo* $CD4^+$ and $CD8^+$ T cells from AS patients pre and post-treatment were compared to *in-vitro* cultured $CD4^+$ and $CD8^+$ T cells stimulated with CD3/CD28. The results of this chapter show the importance of context towards gene regulation, and in future AS patient cell samples for ChIP-seq, ATAC-seq, RNA-seq, and Tiled-C should also be generated from CD3/CD28-stimulated cells to investigate whether the ability for these looping interactions to form is altered in activated patient T cells.

The Tiled-C data (n = 3) and CD3/CD28-stimulated ATAC-seq (n = 5) and RNA-seq (n = 5) data in this chapter was also limited by modest sample sizes. Interactions, chromatin accessibility, and gene expression are known to be affected by individual SNPs, and their presence may cause highly variable data across specific individuals. In future, these sample sizes should be increased to avoid possible outliers influencing the interaction and epigenetic data and ensure that the increase in interactions between AS SNPs in stimulated healthy volunteers and lack of interaction change in AS is genuine.

5.5.3.2. Machine-learning for the identification of differential interactions in AS

The Tiled-C analysis was limited by its lack of true statistical comparison between conditions to identify differential interactions. As mentioned in the chapter introduction (section 5.1.3), whilst many statistical packages were considered none could handle the

high proportion of bait-to-bait interactions produced by the captured Tiled-C loci. Alternative Tiled-C analysis methods could be adopted in future, such as 'Twins' (Al-jibury *et al.*, 2023). Twins uses machine deep-learning models and Siamese neural networks (a neural network with two or greater distinct subnetworks that share the exact same parameters and weightings; i.e. unstimulated and stimulated interaction matrices) to visually inspect normalised contact matrices for each biological replicate and call statistically significant differences in interaction strength between conditions. This could be used to detect gained and lost differential features across conditions in my Tiled-C data. Although the comparison between relative contact probability scores have identified interactions that appear different between gene promoters and SNPs like *rs17765610* and *rs928753* in both unstimulated and stimulated $CD4^+$ and $CD8^+$ T cells, approaches like Twins should be considered as an extra layer in the analysis to give further confidence that these SNPs are found in differential interaction domains. Furthermore, Twins has been shown to be highly sensitive to the changes in chromatin conformation caused by CTCF degradation or distal enhancer deletion. This suggests the pipeline may be sensitive enough to analyse the subtle changes in chromatin conformation that may occur between unstimulated healthy volunteers and AS patients pre and post-treatment with adalimumab.

5.5.4. Concluding remarks

Underpinning how the epigenetic landscape differs between cell types and between AS patients and healthy volunteers is crucial towards our better understanding of AS pathology and identification of novel therapeutic targets. Although there are limitations, integrating Tiled-C data with gene expression, chromatin accessibility, and ChromHMM datasets uncovered that AS-associated SNPs in bulk $CD4^+$ T cells and three in bulk $CD8^+$ T cells formed stronger chromatin with ten genes such as *BACH2*, *PTGER4*, and *TNFAIP3*. These contacts did not form in unstimulated T cells, highlighting the potential

importance of activated T cells in contributing to AS pathogenesis. Importantly, the work performed in this chapter has produced comprehensive maps of the *cis*-regulatory architecture at 21 AS-associated loci that can be used to target regions for future functional investigation (i.e. CRISPR/Cas9-based enhancer knockdown studies) to further uncover the molecular mechanisms driving pathogenesis and provide further insight into the T cell-driven aspects of AS.

"Stuff your eyes with wonder, live as if you'd drop dead in 10 seconds. See the world. It's more fantastic than any dream made or paid for in factories."

- Ray Bradbury, Fahrenheit 451 (1953)

6

GENERAL DISCUSSION

Contents **Pg.**

6.1. Thesis summary ……. 285

6.2. Modelling the epigenetic landscape in AS ……. 287

6.3. T cell activation and AS-associated genetic variants ……..... 291

6.4. Limitations and future work ……..… 295

6.5. Concluding remarks …….. 301

In this chapter, the results of this thesis are summarised in the context of published literature and the overall thesis goals. The overall limitations and considerations for future work stemming from this thesis are also discussed.

6.1. Summary of findings

The overall aim of this thesis was to understand the functional basis of AS disease-associated risk loci in order to improve understanding of the aetiology and pathophysiology of this disease. By applying a functional genomic approach to understanding gene regulation, the work presented here provides novel evidence linking specific AS-associated GWAS genetic variants with *cis*-regulatory regions to the genes they interact with in disease-relevant cell types and cellular contexts. AS has a strong polygenic component involving genetic variants in greater than 100 loci (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016). Prior work has shown that many of these AS-associated

genetic variants intersected cis-regulation-associated histone marks in immune cells; suggesting they may lie in active regulatory elements (Li *et al.*, 2017). It was not known, however, how the regulome varied between healthy individuals and AS patients with active disease and the nature of chromosomal interactions at AS GWAS risk loci.

To address this, I built on a recent effort to systematically map the chromatin landscape genome-wide in AS (section 3.3.1) using an integrated ChromHMM approach (Ernst & Kellis, 2017) incorporating both transcriptomic and epigenetic profiling assays (Brown *et al.*, 2023). In addition to the original H3K4me3 and H3K27ac ChIPmentation-seq, ATAC-seq, and RNA-seq datasets used therein, I specifically performed and integrated H3K27me3 ChIPmentaion-seq data into the 14-state ChromHMM to identify Polycomb-repressed chromatin states. Differential analysis was then performed on the new ChromHMM datasets to identify where the epigenome of AS patients and healthy volunteers differed. This analysis revealed AS-specific changes in the chromatin landscape and showed that AS-associated SNPs are enriched in *cis*-regulatory chromatin states such as active promoters and enhancers. Some MHC SNPs were also found to be enriched in Polycomb-repressed chromatin states.

I next wanted to identify which genes might be regulated by these SNPs in regulatory regions. Both Capture-C (section 4.2.1) and Tiled-C (section 5.3.1) analyses were used to identify looping interactions between AS-associated SNPs intersecting ChromHMM-defined active enhancer regions and the promoters of proximal genes. The Tiled-C importantly identified looping interactions between AS-associated SNPs in enhancers and promoters that were formed or strengthened in $CD4^+$ and $CD8^+$ T cells on stimulation with anti-CD3/CD28 activation beads; a finding consistent with the proposed importance of activated effector T cell subtypes such as Th1, Th2, and Th17 cells in AS pathogenesis (Dong, 2021; Su *et al.*, 2022).

Taken together, these findings provide comprehensive maps of the AS-associated chromatin landscape in peripheral blood immune cells and suggests hypotheses for how AS-associated SNPs could influence gene regulation and cellular function in AS. By investigating both naïve and stimulated bulk T cell populations in this thesis, I have demonstrated a key role for activated T cells in AS pathogenesis. This work also emphasizes the importance of determining the functional effects of genetic variants in appropriate cell types and disease contexts not only in AS, but also in related immune-mediated conditions such as the other SpAs.

6.2. Modelling the epigenetic landscape of AS

6.2.1. *NOS2* was not a tractable locus for further study

AS has a complex cellular aetiology involving a number of immune (i.e. T cells and monocytes) and non-immune cells (i.e. osteoclasts) (section 1.4). Because of this, identifying which AS-associated variants are functional in which tissues and cell types is a challenging endeavour. This thesis has focused on functional genomic analysis of healthy volunteer and AS patient $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells given that: $CD4^+$ and $CD8^+$ T cells are thought to be an important early target in models of HLA-$B27^+$ predisposition to AS (section 1.4.2); $CD14^+$ monocytes can differentiate into macrophages and bone-reabsorbing osteoclasts on signalling with LPS, IL-10, or IL-17 (section 1.4.4) with the proportion of osteoclasts inversely proportional to AS disease activity (Caparbo *et al.*, 2018). Further studies evidence that macrophages in AS patients produce more pro-inflammatory cytokines such as TNFα, IL-23, and IL-1β (Akhtari *et al.*, 2021); $CD19^+$ B cells have an emerging role in AS. They differentiate into antibody-producing plasmablasts. A number of autoreactive antibodies have been uncovered in AS cases such as those raised against sclerostin and collagen

(Wilbrink[2] *et al.*, 2021). Plasmablasts are also expanded in AS cases and correlate to altered levels of immunosuppressive IL-10 in the disease (Chen *et al.*, 2015; Matsumoto *et al.*, 2014; Zhao *et al.*, 2024).

An initial goal of chapter three was to investigate how the epigenetic landscape differed between healthy volunteers and AS patients at the *NOS2* locus (section 3.2.1). The locus was selected both for its function in the immune system and its strong AS GWAS association (Cortes *et al.*, 2013; Ellinghaus *et al.*, 2016). As the locus had not been fully studied in human systems, *NOS2* initially appeared to be a good candidate for functional investigation. Subsequent analysis of chromatin state at *NOS2* confirmed the gene is most probably not transcribed in naïve $CD4^+$ T cells, $CD8^+$ T cells or $CD14^+$ monocytes. Previous literature suggested stimulation of $CD14^+$ monocyte-derived macrophages with MDP and GMDP could induce *NOS2* expression (Landes *et al.*, 2015), although attempts to replicate this experiment did not show significant *NOS2* induction. H3K27me3 ChIPmentation-seq revealed a strong Polycomb-complex associated repression signal internal to *NOS2* that suggested long-term gene repression in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ T cells. The work highlighted the importance of investigating the effects of AS-associated variants and genes in the appropriate cell types and contexts where they may be active.

6.2.2. AS-associated genetic variants enriched primarily for *cis*-regulatory elements and Polycomb-repressed DNA

I found a large number of sites of Polycomb repression-associated post-translational histone modification H3K27me3 peaks – between 19,779 ($CD19^+$ T cells) and 75,604 peaks ($CD14^+$ monocytes) – using the selected MACS2 parameters (section 2.7.3). However only a single peak across the analysis in $CD14^+$ monocytes was differential

between AS patients with active disease and healthy volunteers (section 3.2.2). This result was surprising, as Cribbs *et al.* (2020) had shown that inhibition of H3K27me3 histone demethylases – which remove H3K27me3 in response to cellular signalling – increases H3K27me3 occupancy at the promoters of key metabolic transcription factors in T cells and increases sensitivity to self-antigens that bias $CD4^+$ T cell development towards AS-associated Th17 phenotypes.

I built a novel ChromHMM using our previously reported datasets (Brown *et al.*, 2023) which additionally incorporated the H3K27me3 ChIPmentation-seq data to identify new hidden repressive chromatin states (section 3.3.1). I identified 2 additional chromatin states that corresponded to both coding and non-coding Polycomb-repressed elements. I then investigated whether AS-associated variants were enriched in any of the chromatin states (section 3.3.6). I observed two important findings:

[1] AS-associated SNPs were enriched primarily for active cis-regulatory chromatin states; including promoters and enhancers in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells.

[2] AS-associated SNPs specifically in the MHC region (defined as chr6: 25,000,000 – 35,000,000) showed enrichment for Polycomb-repressed and non-coding transcription chromatin states in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells.

The first finding we previously reported (Brown *et al.*, 2023) and suggested that functional analysis of AS-associated variants should be first limited to enhancers and promoters. Interestingly, the MCA (section 3.3.2), quantitative (section 3.3.3), and permutation (section 3.3.4) analyses revealed that – whilst no significant differences existed in the proportion of promoters and enhancers between AS patients and healthy volunteers in $CD4^+$ T cells, $CD8^+$ T cells, and $CD19^+$ T cells – $CD14^+$ monocytes showed

decreased numbers of promoters and increased numbers of enhancers in AS patients. These changes are likely reflective of epigenetic changes as cells differentiate into the more inflammatory phenotypes often seen enriched in AS cases. Indeed, Tserel *et al.* (2010) have shown that deposition of H3K4me3 at promoters decreased as $CD14^+$ monocytes differentiated into their derived macrophages and dendritic cells. Minderjahn *et al.* (2022) conversely were able to show significant reconfiguration of H3K27ac signals across the genome during this differentiation; with a general gain in the number of active super-enhancers in the macrophages and dendritic cells. The promoter and enhancer changes detected in this analysis could therefore represent early epigenetic changes reflective of monocytes preparing to differentiate. These findings highlight a putative important role for $CD14^+$ monocytes and their derived cell types in AS, as some of these novel active enhancer regions may overlap SNPs that are not seen in healthy volunteers.

Although the second finding was not investigated further in this thesis, H3K27me3-mediated silencing of MHC genes is a critical modulator of their expression. In one study, Burr *et al.* (2019) performed CRISPR-Cas9 knockout screens genome-wide to investigate how MHC genes are regulated. The authors found that ablation of *PRC2* was associated with significant upregulation of HLA-B allotypes in erythroleukemia cells. In MS, at least three SNPs in the MHC region close to the genes *H2-Ab1* and *CYP24A1* overlap strong context-specific H3K27me3 signals that upon INFγ treatment lose the repressive mark, and gain increased H3K4me3, H3K27ac, and CTCF signals; suggesting they are silenced *cis*-regulatory elements (Meijer *et al.*, 2022). These SNPs also only showed interaction with their respective gene promoters – as measured by HiChIP – on IFNγ treatment. The AS-associated SNPs that overlap with my Polycomb-repressed chromatin states in the MHC region could represent silenced regulatory elements awaiting specific activation cues to modulate MHC gene expression in specific cell states. Consistent with this idea,

the Polycomb-repressed chromatin states were the only regions to show significant permutation and MCA separation in $CD8^+$ T cells and $CD19^+$ B cells. This suggested shifts away from Polycomb-repression in the disease state to other, more active biological roles.

6.3. T cell activation and AS-associated genetic variants

6.3.1. Capture-C was used to link AS-associated lead SNPs to their cognate genes

Although AS-associated SNPs enriched for regulatory elements, ChromHMM is unable to physically link these genetic variants to their target genes. In chapter four, a reciprocal Capture-C approach was used to determine whether lead SNPs from Cortes *et al.* (2013) and Ellinghaus *et al.* (2016) interacted with the promoter of their closest gene. DpnII was used to fragment the genome – allowing for a resolution of roughly 2 kbp – and the fragments containing the lead SNP and promoters were both baited with complimentary oligonucleotides. The data was analysed both with normalised read pileups and with Peaky to determine interaction significance. A total of nine genes showed significant interaction with an AS-associated lead SNP across $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes (section 4.3.1). Seven of these were between the lead SNP and their nearest gene. As an example, at the chr21q22 locus there was a clear reciprocal interaction between the promoter of *ETS2* and the lead SNP *rs2836883* specific to $CD14^+$ monocytes. This SNP is also associated with other inflammatory and AS-related conditions such as IBD (Xie *et al.*, 2021).

A recent study has shown that *ETS2* is a central mediator of the inflammatory response in $CD14^+$ monocyte-derived macrophages (Stankey *et al.*, 2024). Through CRISPR-Cas9-based gene editing, the authors were able to measure the effects of an IBD, AS, primary sclerosing cholangitis, and Takayasu's arthritis-associated SNP (*rs2836882*)

intersecting an enhancer in the same LD block as the AS-lead SNP *rs2836883.* They found that the risk-associated G-allele was associated with significantly higher *ETS2* expression and enhancer activity when compared to the normal A-allele as measured by massive parallel reporter assays. It was also determined that the G-allele significantly increased the binding of the transcription factor PU.1 to the enhancer. Other studies have found that PU.1 primes and prevents the silencing of enhancers surrounding pro-inflammatory genes in mature macrophages (Tagore *et al.*, 2015). Stankey *et al.* (2024) also determined that CRISPR-Cas9-mediated knockout of *ETS2* significantly decreased the expression of a set of genes enriched for macrophage activation and pro-inflammatory production ontologies in pathway analysis. Importantly, this study highlights that the Capture-C work in chapter four can identify interactions with functional genetic variants that have important roles in mediating inflammation. Similar studies should be conducted for the interacting regions in the other identified loci.

6.3.2. Tiled-C was used to map the complete interaction profiles of AS-associated TADs in multiple contexts

Some interaction profiles were complex. At *RUNX3* for instance, the lead SNPs and functional variants (Vecellio *et al.*, 2018; Cohen *et al.*, 2021) are found in close proximity to the promoter viewpoint. This increased the statistical noise through the Brownian motion of DNA and made determining genuine interactions difficult with Peaky. In addition, Capture-C was not able to identify looping interactions between the majority of differential epigenetic features found in AS patients (i.e. open chromatin, H3K4me3, H3K27ac, or eRNA) and a target gene. In chapter five, I improved upon the Capture-C analysis by performing Tiled-C on 21 of the AS-associated loci in $CD4^+$ T cells, $CD8^+$ T cells, and $CD14^+$ monocytes including the nine previously identified. In addition to the healthy volunteer and AS patient cells, I expanded this analysis to also include stimulated

cells (CD3/CD28 T cell activation beads for $CD4^+$ and $CD8^+$ T cells; LPS for $CD14^+$ monocytes) and cells from responding AS patients with active disease following 3-months treatment with anti-TNFα therapy (adalimumab). I also performed ATAC-seq and RNA-seq on matched samples to determine if changes in the Tiled-C data corresponded to changes in chromatin accessibility and gene expression.

The Tiled-C libraries in $CD4^+$ and $CD8^+$ T cells showed excellent quality with the majority (≥ 70%) of all sequencing reads deriving in *cis*. Whilst AS patient samples were similar to the T cells, the healthy volunteer $CD14^+$ monocyte data was far poorer and had a much higher proportion of *trans* reads. Although untested, previous studies had shown higher degrees of spontaneous monocyte cell death in culture (Zhao *et al.*, 2010; Lage *et al.*, 2022). Because no baseline, healthy comparisons were available for AS patient cells, the $CD14^+$ monocytes were omitted from this analysis.

One of the key findings of chapter five was the putative importance of T cell activation towards AS. In $CD4^+$ and $CD8^+$ T cells, the control locus *MYC* showed a CD3/CD28 stimulation-specific response. New interactions formed between the *MYC* promoter and as many as five active ChromHMM-predicted enhancers. These regions showed good concordance with previously validated interactions (Hnisz *et al.*, 2016) and CRISPRi studies (Fulco *et al.*, 2019). Moreover, these interactions were not seen in the AS patients with active disease either before or after treatment with adalimumab. The expression of *MYC* was also differential only in the CD3/CD28-stimulated $CD4^+$ and $CD8^+$ T cells and suggested these contacts may contribute to the upregulation of *MYC* on activation. This was further evidenced through the ATACs-seq data, which showed that the chromatin at these regions were only accessible on CD3/CD28 stimulation.

6.3.3. T cell activation was associated with increased chromatin interactions at multiple inflammation-related loci in $CD4^+$ and $CD8^+$ T cells

The trends seen at the *MYC* locus extended to a number of the AS-associated loci. In chapter five, I observed six regions in $CD4^+$ T cells and four regions in $CD8^+$ T cells were associated with increased chromatin contact between an AS-associated SNP and a gene promoter on CD3/CD28 stimulation. The *PTGER4* (section 5.4.1), *BACH2* (section 5.4.2), and *DEXI* (section 5.4.3) loci were used as examples to illustrate the importance of activated phenotypes in analysing AS-associated risk loci in T cells. All of these loci demonstrated looping interactions with enhancers in or near AS-association signals that were either only seen in, or strengthened by, CD3/CD28 stimulation.

At the 5p13.1 locus, for instance, it was shown that the lead SNP *rs1992661* interacted with the *PTGER4* promoter across all conditions. On CD3/CD28 stimulation however, both *rs199266* and the *PTGER4* promoter were also seen to interact with a region of open chromatin containing the SNP *rs9283753*. Tewhey *et al.* (2018) have shown that *rs9283753* has direct effect on *PTGER4* expression in lymphoblastoid cell lines, with the risk-associated T allele significantly increasing *PTEGR4* expression. It is likely that both *rs9283753* and *rs1992661* may control *PTGER4* expression, with the potential effects of *rs9283753* necessitating study in activated T cells.

These findings recapitulate those found by *Farh et al*. (2015), who performed comparative analysis of GWAS-associated SNPs and epigenetic marks to underpin common pathogenic cell types shared between immune-mediated conditions. The authors developed a propriety algorithm trained on 39 immune conditions – including AS and other SpAs – and generated genome-wide RNA, H3K4me3, and H3K27ac datasets across 56 distinct cell types. The authors noted that the H3K27ac signal at *cis*-regulatory

elements of naïve $CD4^+$ T cells stimulated with CD3/CD28 activation beads correlated well with those found in CD3/CD28-stimulated Th1 and Th2 cell populations. Th1 cells are pro-inflammatory and release cytokines such as IFNγ and TNFα. Th2 cells are instead immunosuppressive and release cytokines such as IL-4 and IL-10. Both populations and their cytokines are significantly altered in AS, with Th1 cells showing significant enrichment and Th2 cells showing significant depletion (Wen *et al.*, 2017). Farh *et al.* (2015) also showed that the stimulated naïve T cell, Th1, and Th2 cluster enriched significantly for immune-associated GWAS SNPs; suggesting CD3/CD28 stimulation may be an important context for the investigation of AS SNPs.

Interestingly, the authors found that Th17 clustered separately from the CD3/CD28-stimulated cells; with Th17 cells instead clustering with IL-17 negative, phorbol 12-myristate 13-acetate and ionomycin stimulated $CD4^+$ T cells. This suggested that CD3/CD28 stimulation may not capture the H3K27ac-rich *cis*-regulatory landscape of Th17 cells. As previously mentioned, naïve T cells, Th1 cells, and Th17 cells have been shown to have distinct chromatin conformations (Zhang *et al.*, 2023). Although the CD3/CD28 stimulation of bulk $CD4^+$ T cells did show increased contact with AS-associated SNPs at a handful of loci, these studies highlight the need to adopt more targeted isolations of T cell subtypes to identify where GWAS genetic variants may be active.

6.4. Limitations and future work

6.4.1. Sample numbers and extended recruitment

Throughout this thesis analysis has been limited by relatively low sample numbers. This was due to a mix of both logistical and funding limitations, as the assays used to generate comprehensive maps of the epigenome in healthy volunteer and AS are labour intensive

and expensive. In chapter three, the ChromHMM was limited to four healthy volunteers and four AS patients. One of the strengths of the ChromHMM derives from the fact that all of the assays were performed on the same individual at a single blood draw. The ChromHMM therefore represents a matched snapshot of the chromatin landscape of each individual at recruitment. Downstream analysis was limited by the small sample size. In order to determine whether there were statistically significant differences in chromatin state between AS cases and healthy controls we performed a Fisher's exact test for each chromHMM bin, which is significant where a chromatin state is present in four healthy volunteers and no AS patients or *vice-versa*). This significance is however not maintained following multiple testing correction. This would mean a relatively high rate of false-positives and an inability to conclude that chromatin states between the AS patient and healthy volunteer comparisons were different. The permutation analysis circumvented this by comparing the number of Fisher's exact test significant regions to a maximally permuted background of 70 possible combinations. This identified chromatin states where the number of four versus zero bins were higher than by expected by chance alone. Increasing the number of ChromHMM samples would allow the corrected Fisher's exact p-value to be used.

In chapters four and five, the Capture-C and Tiled-C experiments were each performed on three individuals. Although the absolute minimum for statistical analysis with packages like Peaky, the low sample numbers are likely to introduce variation in interaction frequency due to patient clinical heterogeneity, their individual responsiveness to anti-TNFα treatment, and their haplotype. Additionally, whilst all of the cell type samples from Capture-C were matched from the same individual, this was not the case in the Tiled-C data owing to poor sample quality from certain libraries.

6.4.2. Additional cell types and tissue sampling

Although the $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells analysed in this thesis have evidence for roles in AS pathogenesis, a number of other cell types are suspected to be involved. Osteoclasts and osteoblasts control bone resorption and deposition and likely have an important role in modulating the spinal damage seen in AS patients (Liu *et al.*, 2021). AS-associated GWAS variants are also enriched for H3K4me1 and H3K27ac signals across the epigenome of gut epithelial cells (Li *et al.*, 2017). Previous studies have shown substantial epigenetic differences between tissue-resident immune cells and PBMCs. One study by Qaiyum *et al.* (2021) used single cell RNA-seq to identify unique pro-inflammatory mature $CD103^+CD49a^+CD8^+$ T cell phenotypes in the synovial fluid of AS patients that were not found in peripheral blood of AS patients, RA patients, or healthy volunteers. As peripheral blood cells may not reflect the epigenetics at the inflamed tissue, recruitment of new patients and tissue samples should therefore consider collection of matched sacroiliac biopsies and synovial fluid to identify any putative pathogenic differences between PBMCS and tissue-resident cells. Additionally, in chapter five, only patients treated with anti-TNFα (adalimumab) therapeutics were considered. Other biologic treatments that effect different signal cascades, such as anti-IL17, are also effective in AS treatment. Future work should additionally investigate the functional effects that these additional therapeutics may have on the regulatory landscape of AS patient cells.

6.4.3. Incorporating more epigenetic datasets and stimulation conditions into ChromHMM

The ChromHMM analysis was able to detect chromatin states for promoters, enhancers, repressed elements, and active transcription. One key *cis*-regulatory element that was not

investigated was insulator elements. As mentioned in the discussion, insulators have an important role in defining interaction distance and maintaining TAD structure (section 1.5.3.6). SNPs in insulator regions have been shown to have potential effects on gene regulation. Ostrovsky *et al.* (2021) for instance, have shown that *rs4426765*, which is found in a CTCF binding site, could alter the expression of heparanase. *rs4426765* is a risk-allele for graft-versus-host disease, where a transplant donors cells raise an immune response against the tissue of the recipient causing systemic inflammation (Ferrara *et al.*, 2024). Heparanase is heavily involved in general T cell tissue infiltration, and the C-allele of *rs4426765* correlates with increased T cell activation levels and expansion. Insulator elements could be identified by performing similar ChIPmentation-seq for CTCF, which could then be intersected with AS-associated SNPs to investigate whether similar cases to *rs4426765* may occur in AS.

The ChromHMM datasets described in this thesis were also only performed in unstimulated PBMCs. Consistent with the extensive reorganisation of the chromatin environment in AS-associated Th1, Th2, and Th17 cells (Zhang *et al.*, 2023), the ATAC-seq data in chapter five suggested widespread and significant increases in chromatin accessibility across all 21 AS-associated loci in CD3/CD28-stimulated $CD4^+$ and $CD8^+$ T cells that were not observed in unstimulated healthy volunteers, untreated AS patients with active disease, or in adalimumab-treated AS patients with active disease. Tiled-C also suggested the formation or strengthening of interactions in a handful of loci in T cells on CD3/CD28 stimulation. Aside from these findings, previous studies have highlighted the extensive differences between naïve T cells and their activated effector and/or memory subtypes (Jonkman *et al.*, 2022; Rose *et al.*, 2023). H3K4me3, H3K27ac, and H3K27me3 ChIPmentation-seq should be performed on my matched ATAC/RNA-seq individuals in order to build a ChromHMM for CD3/CD28 stimulated T cells in both

AS patients and healthy volunteers. This could then be compared to naïve healthy volunteer T cell data as the untreated AS patients with active disease were in chapter three.

6.4.4. Functional follow-up of identified variants

Although I have clearly linked *cis*-regulatory SNPs to gene promoters and correlated the strength of these contacts to changes in gene expression, this thesis does not perform direct functional analysis on any of the defined SNPs and the expression of their linked genes. As shown by both Fulco *et al.* (2019) at the *MYC* locus and Stankey *et al.* (2024) at the *ETS2* locus, the use of CRISPR-based gene editing and CRISPRi/a can help interrogate enhancer function (Li *et al.*, 2020). CRISPR-based methods have recently shown a great deal of success in primary human T cells (Schmidt *et al.*, 2022; McCutcheon *et al.*, 2023; Ito *et al.*, 2024). Adopting a similar strategy to Stankey *et al.* (2024), CRISPR-could be used to inhibit, active, or delete enhancers containing AS-associated lead SNPs to determine their effect on gene expression. Tiled-C showed a clear interaction between the SNP *rs17765610* and the *BACH2* promoter in both $CD4^+$ and $CD8^+$ T cells with increased interaction frequency on CD3/CD28 stimulation. The effects of the risk-associated A-allele of *rs17765610* over the normal T-allele on *BACH2* expression has not been established. CRISPR-based gene editing could be used to generate two populations of primary T cells – one with the T-allele and one with the A-allele – and qRT-PCR could be used to assay differential *BACH2* expression. In addition, other assays such as EMSA and ChIP could be used to interrogate whether specific allelic variants affect the binding of transcription factors. We have previously used this to great effect at the *RUNX3* locus, where the risk-associated C-allele of the SNP *rs6600247* significantly reduced the binding of c-MYC to its consensus sequence (Cohen *et al.*, 2022).

One of the major limitations of the Tiled-C data was the lack of statistical tests to determine differential interactions between the four conditions. This was due to the abundance of bait-to-bait interactions in the Tiled-C data that rendered the distance models of popular differential interaction tools like Peaky and ChICAGO highly inaccurate (section 5.1.3). Novel machine-learning algorithms, such as Twins, now offer alternative ways to detect differential interactions (Al-jibury *et al.*, 2023). Tiled-C data analysed with these algorithms could be combined with CRISPR-based gene editing to interrogate how the allelic variants of AS-associated SNPs may affect the looping interactions between enhancers and promoters.

6.4.5. Pathway to new therapies

The translation of GWAS findings to therapeutic targets remains a challenge. The development of most drugs will fail due to poor efficacy in prioritising targets, although it has been shown that therapeutic targets with genetic support have a far higher chance of being valid therapeutically (Plenge *et al.*, 2013; Hay *et al.*, 2014; Nelson *et al.*, 2015). In chapter five for instance, I discussed the potential for targeting *PTGER4* with HL-43 (section 5.5.4.1). Ideally AS therapeutics should target genes with the greatest potential therapeutic benefit. To identify novel therapeutic targets in our previous study (Brown *et al.*, 2023) we had incorporated the findings into the Pi Priority Index (Fang *et al.*, 2019) (http://pi.well.ox.ac.uk). By integrating several different epigenetic datasets, the Pi pipeline prioritises and ranks each gene based on genomic evidence, annotation evidence, target tractability, target druggability, and estimations of effects of drugs on gene expression from multiple publicly available datasets.

The most up-to-date version of Pi ranks *PTGER4* sixth out of 14,776 genes for overall therapeutic target priority behind *LCE3D*, *ETS2*, *IRF1*, *ICAM1*, and *CREM*. This suggests

PTGER4 is amongst the most druggable of targets in AS (http://pi.well.ox.ac.uk:3010/pidb/discovery/AS). Aside from small molecules like HL-43, Pi also highlights the extracellular component of *PTGER4* as a high-confidence potential target for antibody therapy using pathway analyses. It should be noted that whilst this version of Pi incorporates data from the 14-state ChromHMM and the Capture-C data from chapter four, the H3K27me3 and Tiled-C data has yet to be added to the database. Adding the H3K27me3 data may further aid prioritisation by identifying genes that are silenced by the Polycomb complex in T cells and monocytes, whilst the Tiled-C would provide chromatin conformation capture information at a much deeper resolution than is currently available for the analysed genes. The Pi rankings could therefore benefit in future from both increasing the depth of the functional analysis presented in this thesis by adding new epigenetic datasets, and by extending the Tiled-C analysis to other genes and cell types to refine the database and identify novel therapeutic targets.

6.5. Concluding remarks

AS is a complex disease with a poorly understood genetic aetiology. The combination of GWAS and functional genomics approaches are proving invaluable in piecing together how disease-associated genetic variants effect gene regulation and contribute towards pathogenesis. In this thesis, I present comprehensive epigenetic maps of multiple AS-associated GWAS loci incorporating ATAC-seq, ChIPmentation-seq, RNA-seq, and interaction data to identify putatively functional genetic variants. These findings show that AS-associated variants are enriched for *cis*-regulatory elements (such as promoters and enhancers), non-coding transcription, and Polycomb-repressed regions of DNA. The chromatin conformation capture data linked genetic variants in these regulatory regions to the expression of multiple inflammation-associated genes including *PTGER4*, *BACH2*,

and *DEXI*. The results of this thesis also show the importance of context-specificity in investigating AS-associated GWAS loci and suggest important roles for enhancer-intersecting AS-associated genetic variants in CD3/CD28-activated T cells.

"Was there anything more exciting in life than seeking answers?"

- Isaac Asimov, Prelude to Foundation (1988)

A

APPENDIX

Cell Type[1]	ChromHMM State[2]	Odds Ratio (OR)[3]	Enrichment Significance[4]
$CD4^+$ T cells	U2 TSSFlnkU	2.70	1.95E-04
	U3 TSSFlnkD	3.09	1.01E-04
	U6 TxNC	3.69	1.71E-04
	U10 EnhActv	3.26	1.69E-06
	U16 PolyNC	2.80	4.16E-07
$CD8^+$ T cells	U2 TSSFlnkU	3.38	1.98E-07
	U3 TSSFlnkD	3.45	1.82E-05
	U6 TxNC	4.71	7.34E-07
	U10 EnhActv	3.57	1.14E-08
	U16 PolyNC	2.82	4.16E-07
$CD14^+$ monocytes	U6 TxNC	4.22	2.13E-04
	U10 EnhActv	3.52	7.28E-08
	U16 PolyNC	3.76	1.70E-09
$CD19^+$ B cells	U2 TSSFlnkU	2.28	1.79E-03
	U3 TSSFlnkU	3.17	1.14E-04
	U6 TxNC	6.61	9.32E-12
	U10 EnhActv	3.79	1.63E-09
	U16 PolyNC	2.90	2.09E-06

Table S.1. SNP enrichment for ChromHMM states with an odds ratio of two or greater are linked to *cis*-regulatory elements in $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and $CD19^+$ B cells. Enrichment information of AS-associated SNPs in ChromHMM emission states with an OR of two or greater, corresponding to a threshold for a 100% increase in the chance of biased SNP enrichment for each state. The first column shows the cell type that each ChromHMM emission state is enriched in[1]. The second column shows the name of each ChromHMM emission state[2]. The third column shows AS-associated GWAS SNP enrichment OR for each state[3]. The fourth column shows the associated enrichment p-value for each state[4]. These values are derived from the GARFIELD analysis containing both MHC and non-MHC loci.

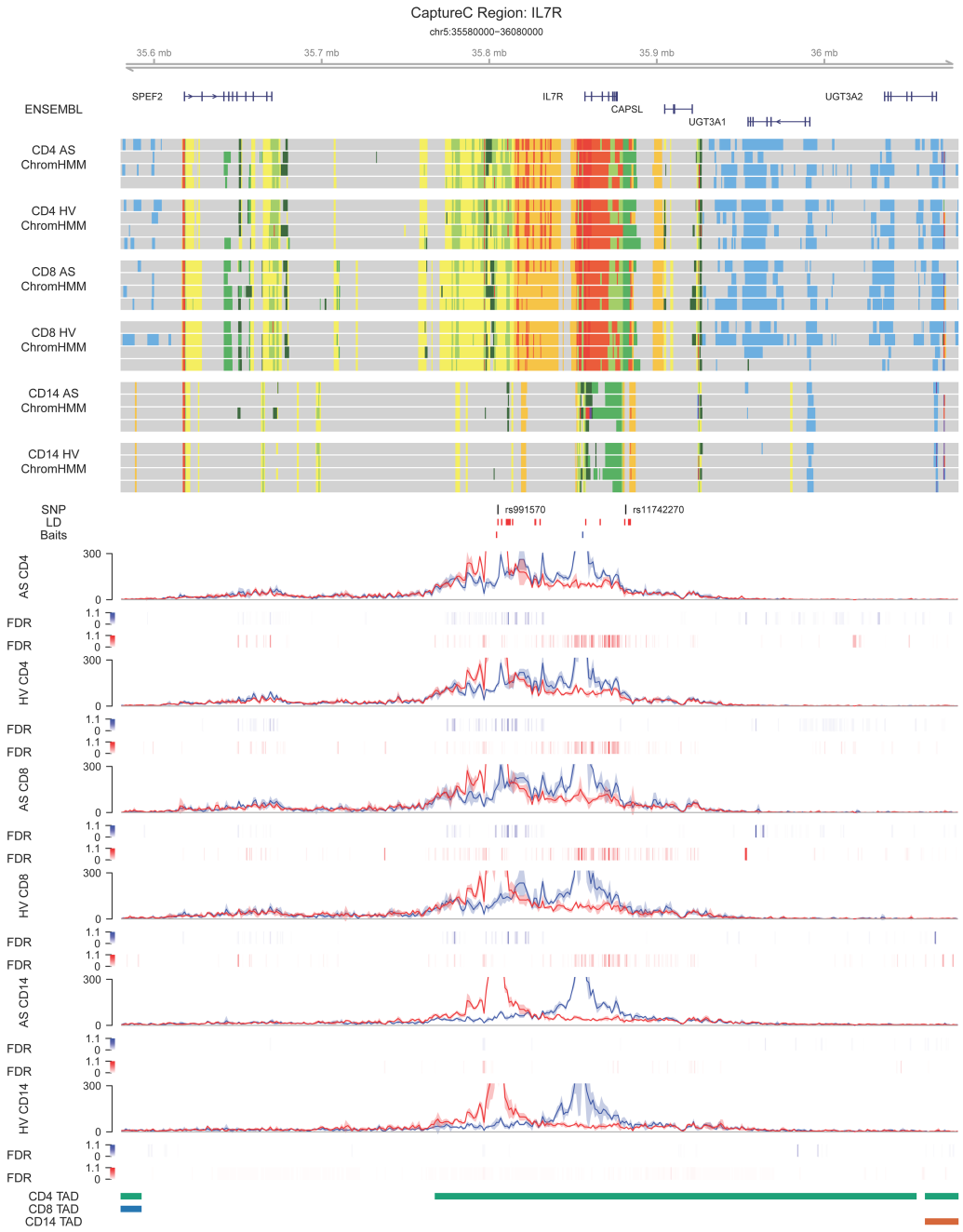

Figure S.1. Enhancer-gene interactions between the promoter of *IL7R* and *rs991570* at the chr5p13.2 locus. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).



Figure S.2. Differential epigenetic marks at the chr5p13.2 locus. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the different epigenetic datasets in AS patients (AS053) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red and blue shaded box shows the location of the *PTGER4* interacting lead or functional SNP-containing *cis*-regulatory regions.

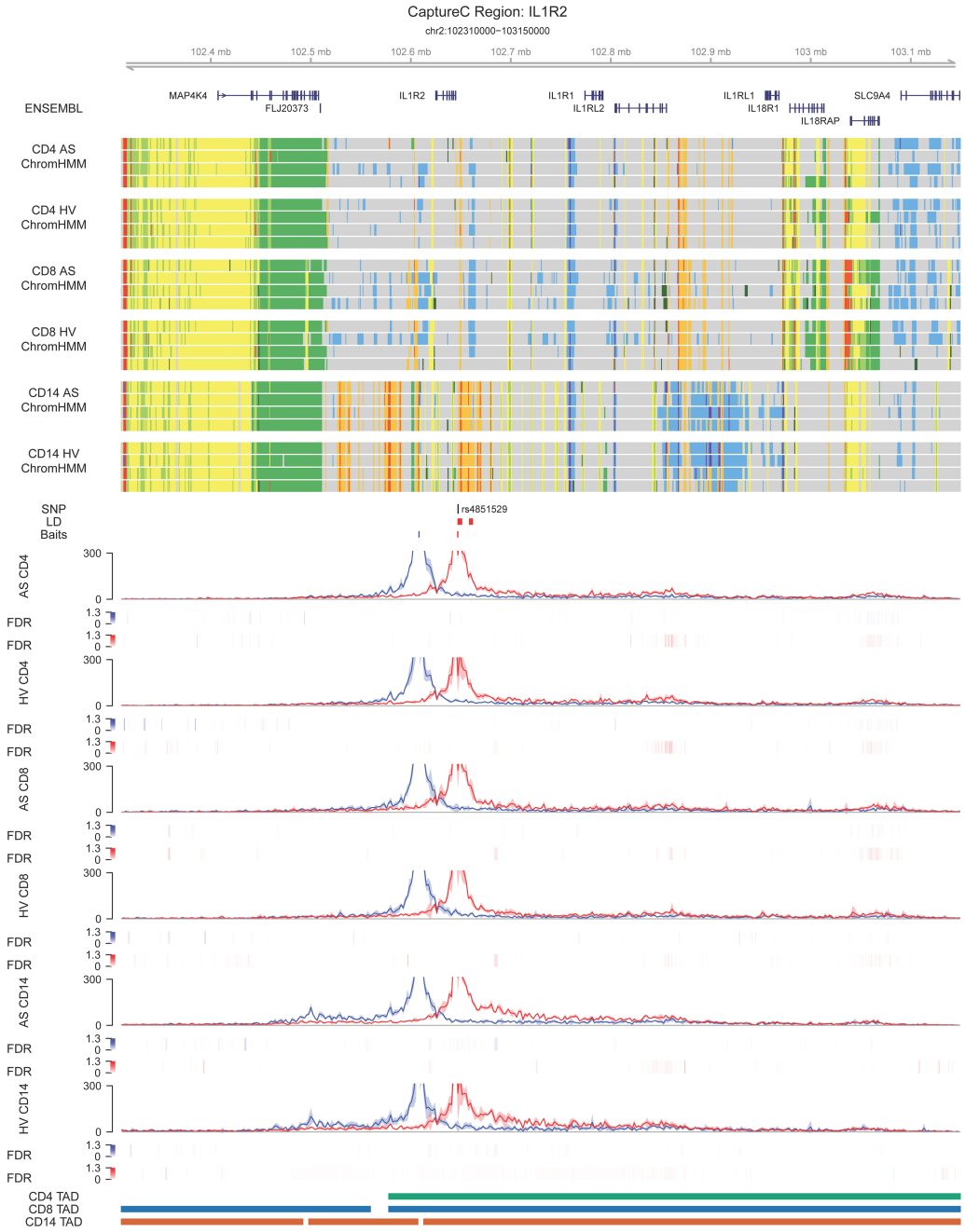

Figure S.3. Enhancer-gene interactions between the promoter of *IL18RAP* and *rs4851529* at the chr2q11.2 locus. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

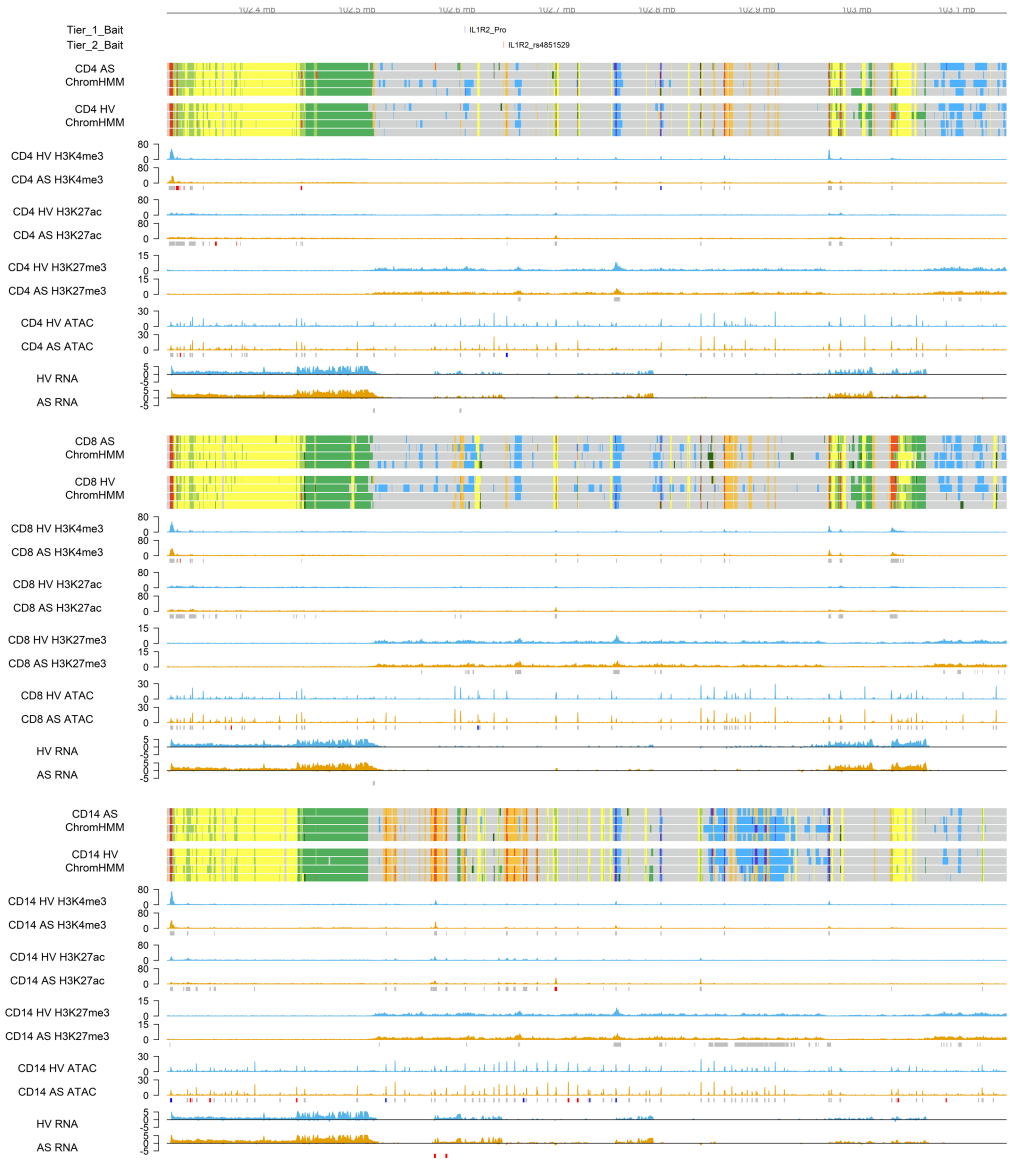

Figure S.4. Differential epigenetic marks at the chr2q11.2 locus. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for different epigenetic datasets in AS patients (AS053) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red and blue shaded box shows the location of the *PTGER4* interacting lead or functional SNP-containing *cis*-regulatory regions.

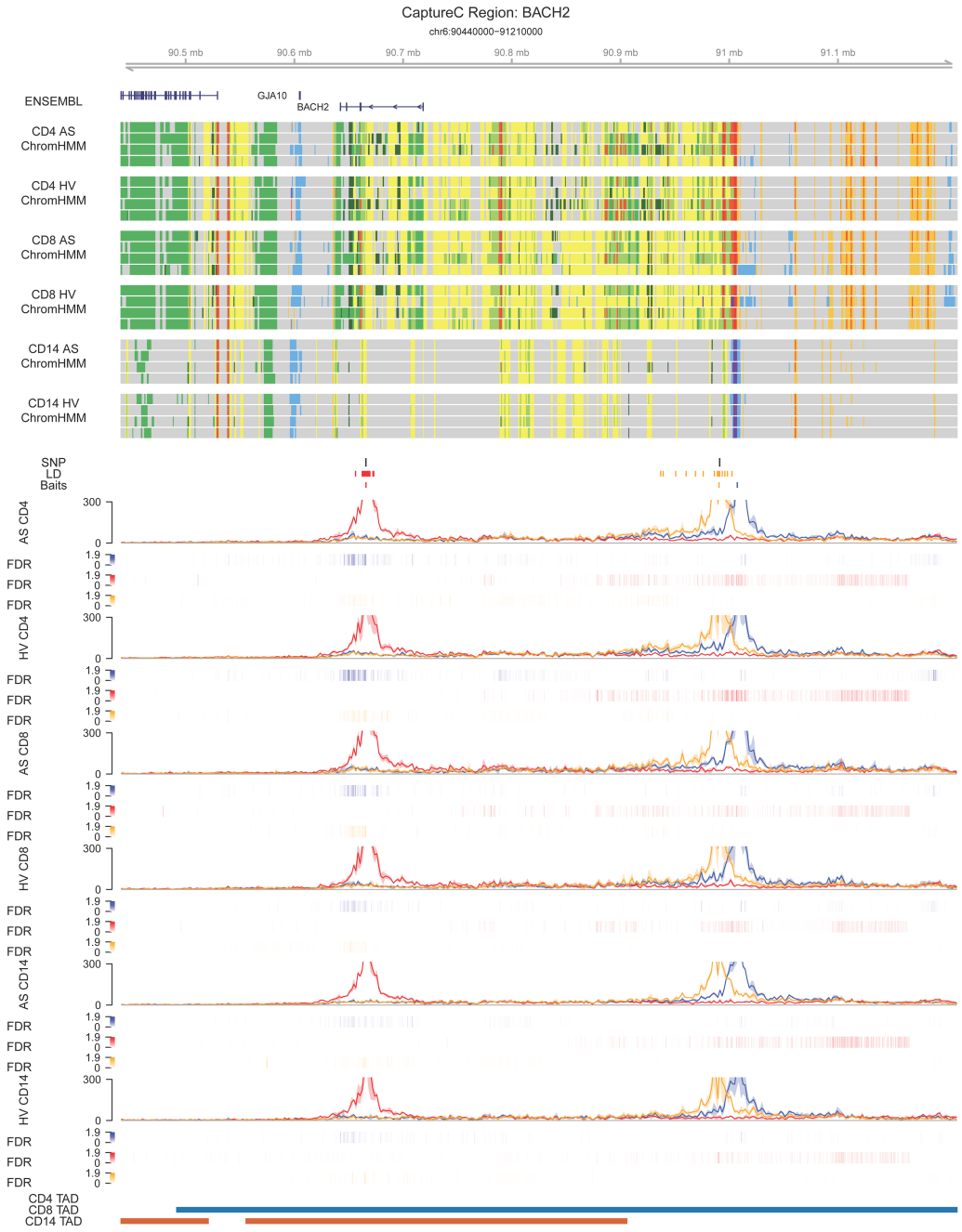

Figure S.5. Enhancer-gene interactions between the promoter of *BACH2*, *rs17765610*, and *rs639575* at the chr6q15 locus. From top to bottom: Scale bar shows the genomic position of genetic features and interactions. 'ENSEMBL' shows the location Ensembl genes in the locus. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'SNP' shows location of the functional or AS-associated lead SNP(s). 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Baits' shows the location of the baited DpnII fragment(s). Tracks underneath show the mean Capture-C interaction counts for each cell type from AS patients ('AS') and healthy volunteers 'HV'). 'FDR' tracks show the Peaky interaction scores in 400 bp bins. 'TAD' tracks show the TAD(s) in each locus for each cell type from Javierre *et al.* (2016).

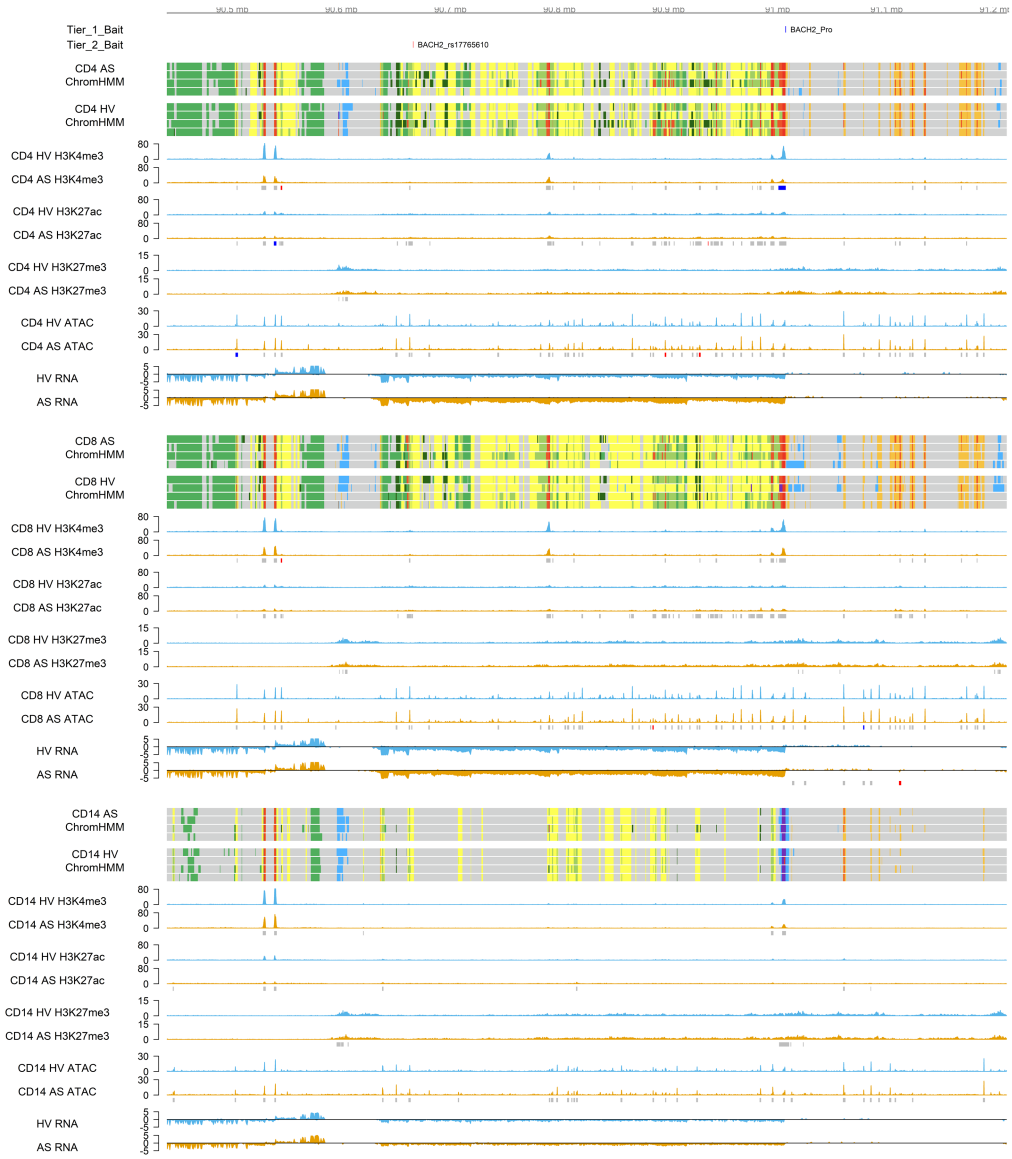

Figure S.6. Differential epigenetic marks at the chr6q15 locus. From top to bottom: Scale bar shows the genomic position of genetic features below. 'ENSEMBL' shows the location Ensembl genes in the locus. 'LD' shows the region in high LD ($R^2 \geq 0.8$) with the lead SNP(s). 'Tier_1/2_bait' shows the location of baited DpnII restriction fragments in capture bait sets 1 or 2. 'ChromHMM' shows tracks from $CD4^+$ T cells ('CD4'), $CD8^+$ T cells ('CD8'), and $CD14^+$ monocytes from AS patients ('AS') and healthy volunteers ('HV'). Colours are consistent with Figure 3.8. 'H3K4me3', 'H3K27ac', 'H3K27me3', 'ATAC', and 'RNA' show a single representative track for the different epigenetic datasets in AS patients (AS053) ('AS': Orange) and healthy volunteers (HV31) ('HV': Blue). Boxes underneath each mark show the strength of MACS2 called peaks in AS patients compared to healthy volunteers (Red = significantly stronger peaks in AS; Grey = no change in peak strength; Blue = significantly weaker peaks in AS). The red and blue shaded box shows the location of the *PTGER4* interacting lead or functional SNP-containing *cis*-regulatory regions.

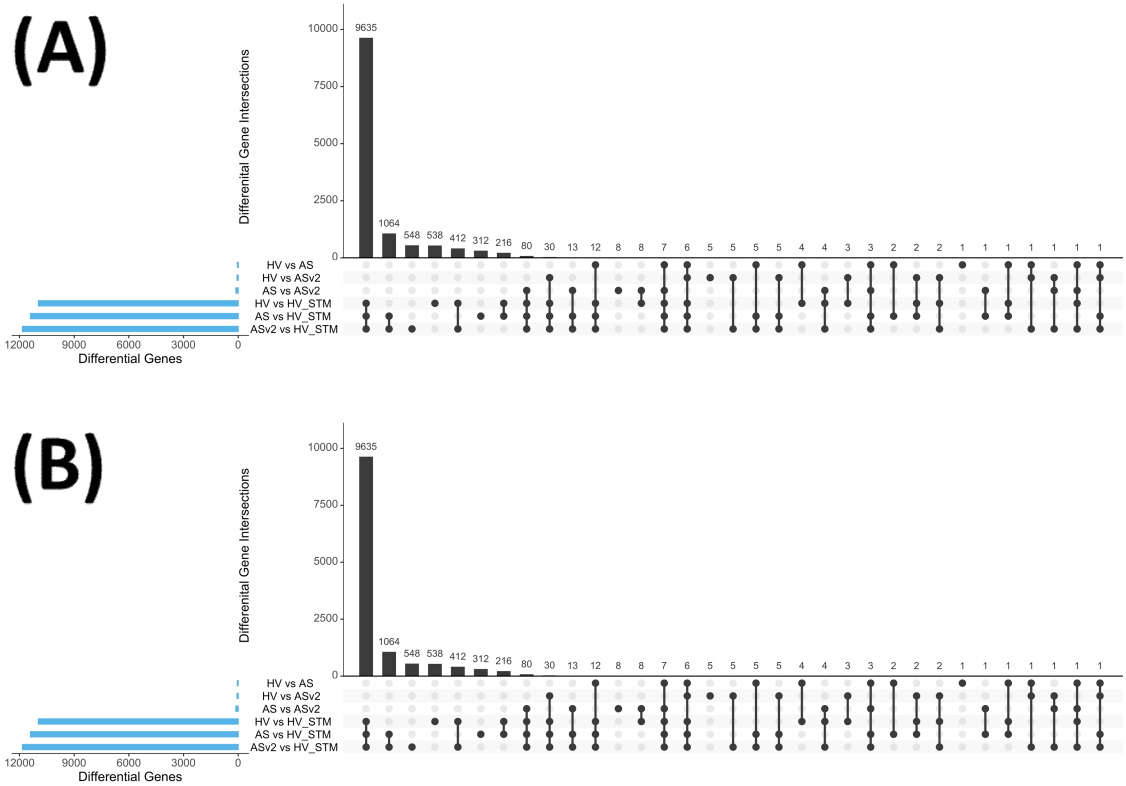

Figure. S.7. Genome-wide analysis of differential genes in unstimulated healthy volunteer, CD3/CD28-stimulated healthy volunteer, untreated AS patient, and adalimumab-treated AS patient CD4$^+$ and CD8$^+$ T cells. UpSet plots showing the number of differentially expressed genes for each condition comparison in (A) CD4$^+$ T cells and (B) CD8$^+$ T cells. HV = Unstimulated healthy volunteer cells; HV_STM = CD3/CD28-stimulated healthy volunteer cells; AS = untreated AS patient cells; ASv2 = adalimumab-treated AS patient cells Filled black dots and connecting bars show which differentially expressed genes are shared between each of the comparisons. Blue bars show the total number of differentially expressed genes for each comparison.

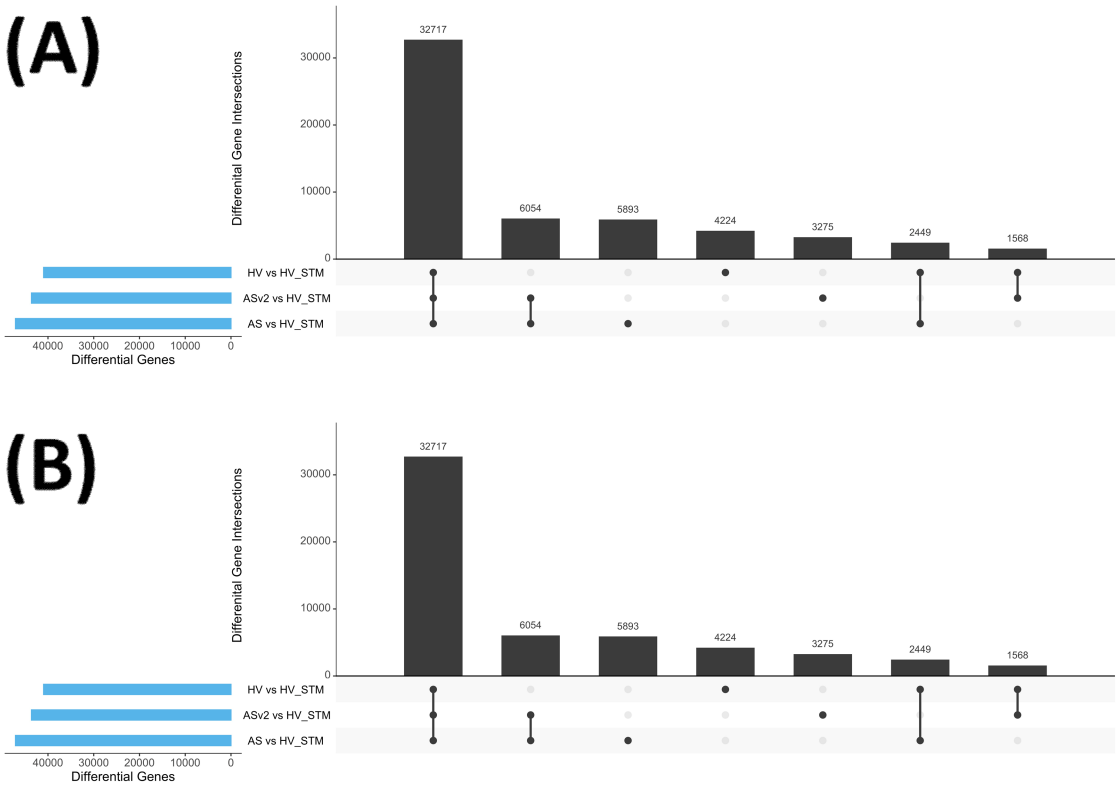

Figure. S.8. Genome-wide analysis of differential open chromatin regions in unstimulated healthy volunteer, CD3/CD28-stimulated healthy volunteer, untreated AS patient, and adalimumab-treated AS patient CD4$^+$ and CD8$^+$ T cells. UpSet plots showing the number of differential ATAC-seq peaks for each condition comparison in (A) CD4$^+$ T cells and (B) CD8$^+$ T cells. HV = Unstimulated healthy volunteer cells; HV_STM = CD3/CD28-stimulated healthy volunteer cells; AS = untreated AS patient cells; ASv2 = adalimumab-treated AS patient cells Filled black dots and connecting bars show which differentially expressed genes are shared between each of the comparisons. Blue bars show the total number of differentially expressed genes for each comparison.

Description of figures S.9 to S.47. Mean Tiled-C interaction matrices (2 kbp resolution) for $CD4^+$ or $CD8^+$ T cells in unstimulated cultured healthy volunteer cells (n = 3), CD3/CD28-stimulated cultured healthy volunteer cells (n = 3), cells from untreated AS patients with active disease (n = 3), and cells from adalimumab-treated AS patients with active disease (n = 3). Matrices represent the ICE and total cis-read normalised relative contact probability (ICE_n-RCP). 'Immunochip SNPs' shows the $\log_{10}$ p-value for all SNPs included in the Cortes *et al.* (2013) GWAS. SNPs that pass the genome-wide significant p-value threshold of 5×10^{-8} are coloured red. 'Lead SNP' shows the position and name of significant or suggestive lead SNPs identified by either Cortes *et al.* (2013) or Ellinghaus *et al.* (2016). '99%_CS_interval' shows the extent of the credible set from van de Bunt *et al.* (2015) for applicable loci. 'CD4/CD8 AS ChromHMM' shows the 16-state ChromHMM from chapter three in AS patients and 'CD4/CD8 HV ChromHMM' shows the equivalent in healthy volunteers. 'ATAC' tracks show mean ATAC-seq peaks from unstimulated healthy volunteers (n = 31 for $CD4^+$ T cells; n = 24 for $CD8^+$ T cells (blue), stimulated healthy volunteers (n = 5 for $CD4^+$ T cells; n = 5 for $CD8^+$ T cells (green), untreated AS patients (n = 21 for $CD4^+$ T cells; n = 20 for $CD8^+$ T cells) (orange) and adalimumab-treated AS patients (n = 20 for $CD4^+$ T cells; n = 19 for $CD8^+$ T cells) (pink). Boxes underneath the ATAC-seq tracks represent significantly differential (adjusted p-value ≤ 0.05; $\log_2$ fold change ≥ 0.585) (Red = upregulated peaks in second condition; Grey = no change between conditions; Blue = downregulated peaks in second condition). ● = unstimulated healthy volunteers versus stimulated healthy volunteers; ●● = unstimulated healthy volunteers versus untreated AS patients; ●●● = untreated AS patients versus treated AS patients. 'H3K4me3' tracks show mean H3K4me3 ChIPm-seq peaks from unstimulated healthy volunteers (n = 7 for $CD4^+$ T cells; n = 7 for $CD8^+$ T cells) (blue) and untreated AS patients (n = 7 for $CD4^+$ T cells; n = 6 for $CD8^+$ T cells) (orange). 'H3K27ac' tracks show mean H3K27ac ChIPm-seq peaks from unstimulated healthy volunteers (n = 7 for $CD4^+$ T cells; n = 7 for $CD8^+$ T cells) (blue) and untreated AS patients (n = 7 for $CD4^+$ T cells; n = 6 for $CD8^+$ T cells) (orange). 'H3K27me3' tracks show mean H3K27me3 ChIPm-seq peaks from unstimulated healthy volunteers (n = 4 for $CD4^+$ T cells; n = 4 for $CD8^+$ T cells) (blue) and untreated AS patients (n = 4 for $CD4^+$ T cells; n = 4 for $CD8^+$ T cells (orange). 'RNA' tracks show mean RNA-seq peaks from unstimulated healthy volunteers (n = 24 for $CD4^+$ T cells; n = 24 for $CD8^+$ T cells) (blue) and untreated patients (n = 15 for $CD4^+$ T cells; n = 16 for $CD8^+$ T cells (orange). Like ATAC-seq, boxes underneath each of the epigenetic tracks represent differential epigenetic peaks, but only for peaks in untreated AS patients relative to unstimulated healthy volunteers (adjusted p-value ≤ 0.05; $\log_2$ fold change ≥ 0.585). 'RefSeq Genes': Position of genes from RefSeq. Scale bar at the bottom of figure shows genomic position of genetic features.

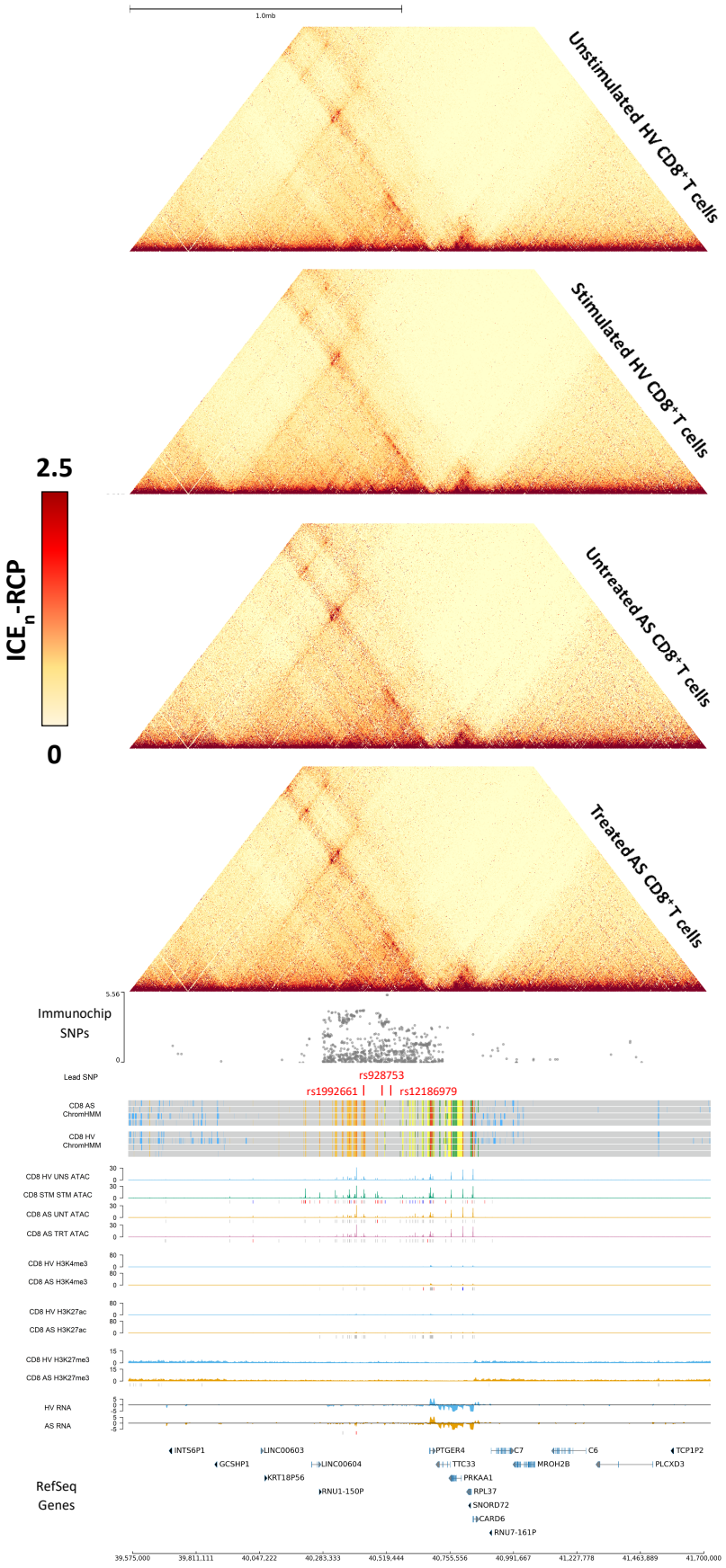

Fig S.9. Tiled-C interactions and the chromatin landscape of the *PTGER4* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr5:39,575,000-41,700,000.

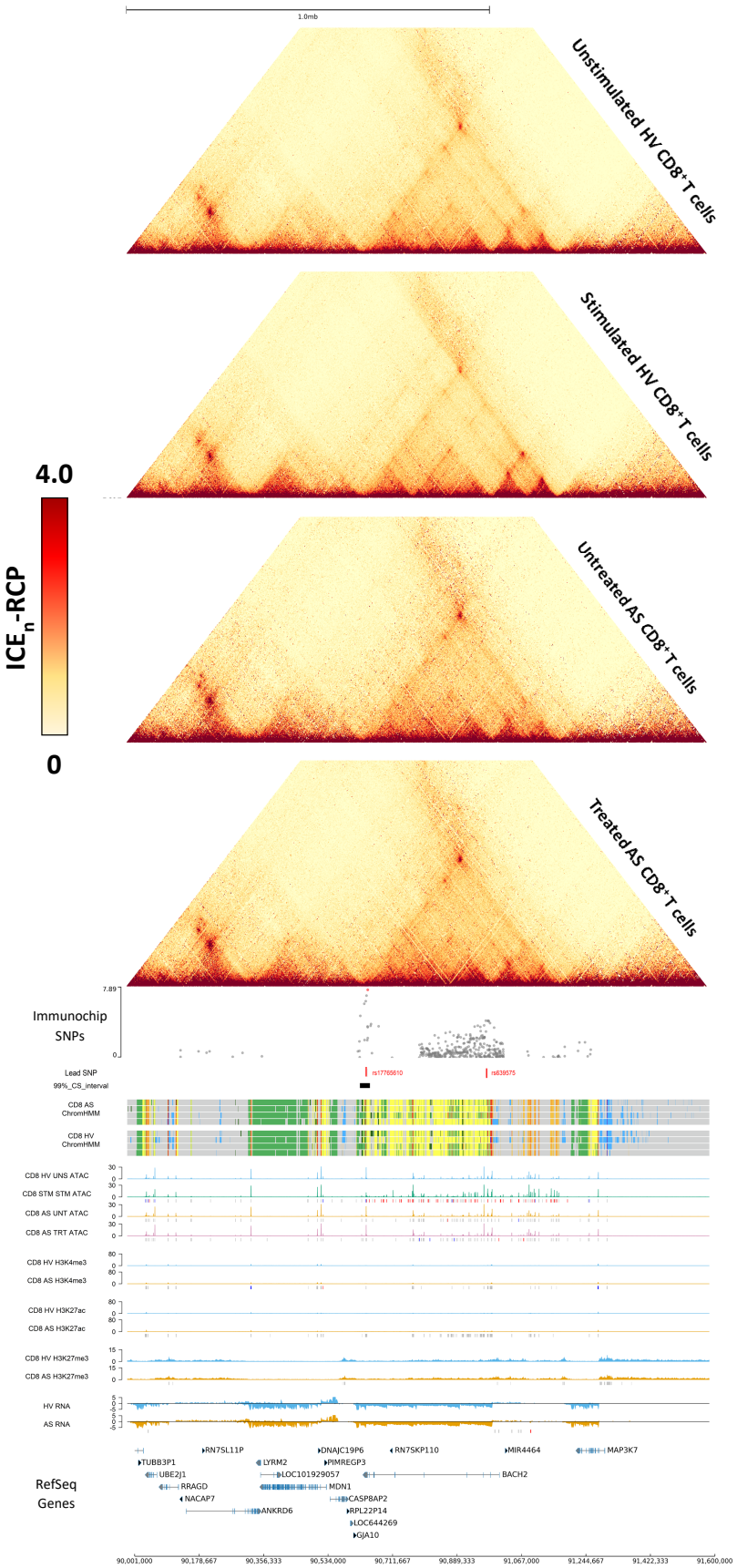

Fig S.10. Tiled-C interactions and the chromatin landscape of the *BACH2* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr6:90,001,000-91,600,000.

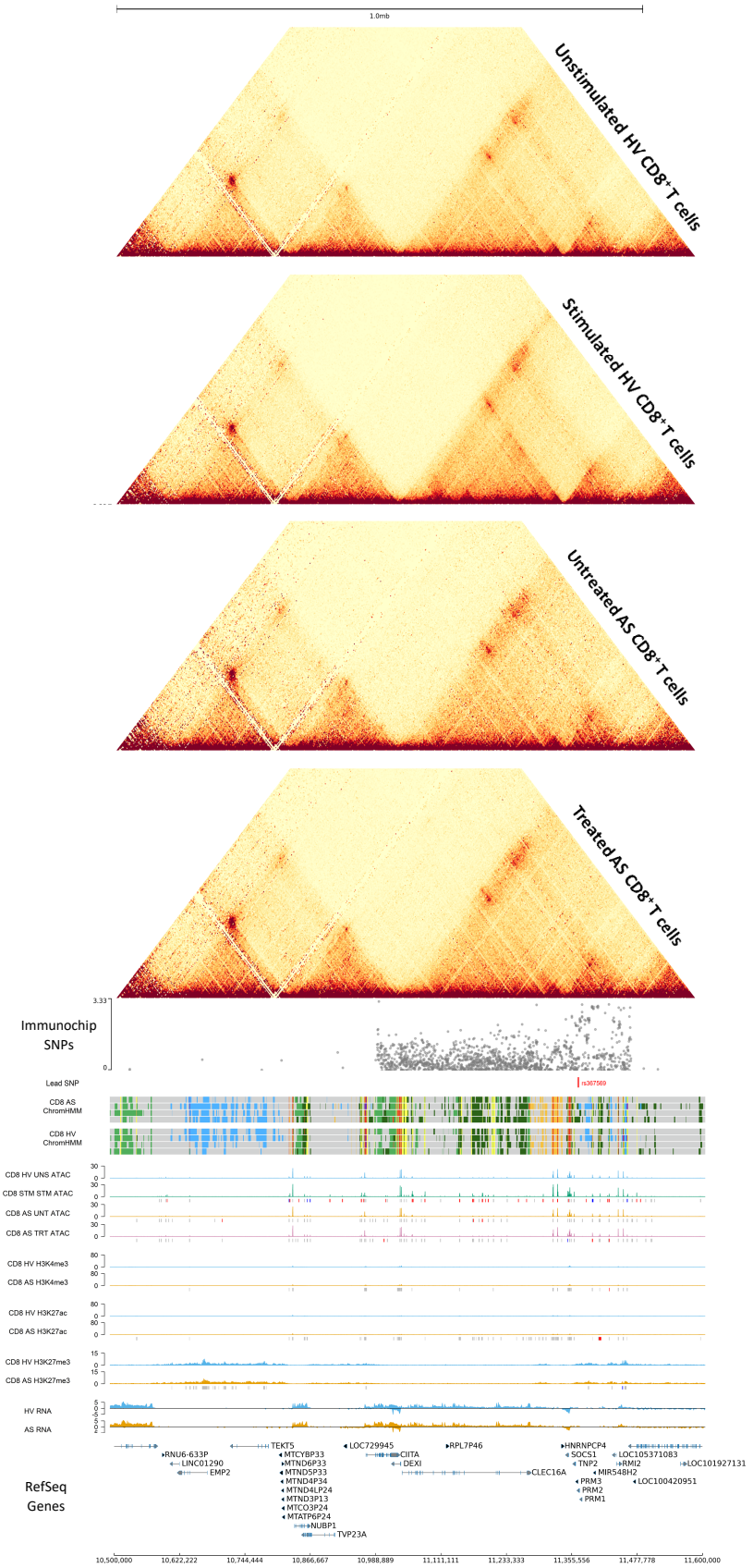

Fig S.11. Tiled-C interactions and the chromatin landscape of the *CIITA-DEXI-CLEC16A-SOCS1* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr16:10,500,000-11,600,000.

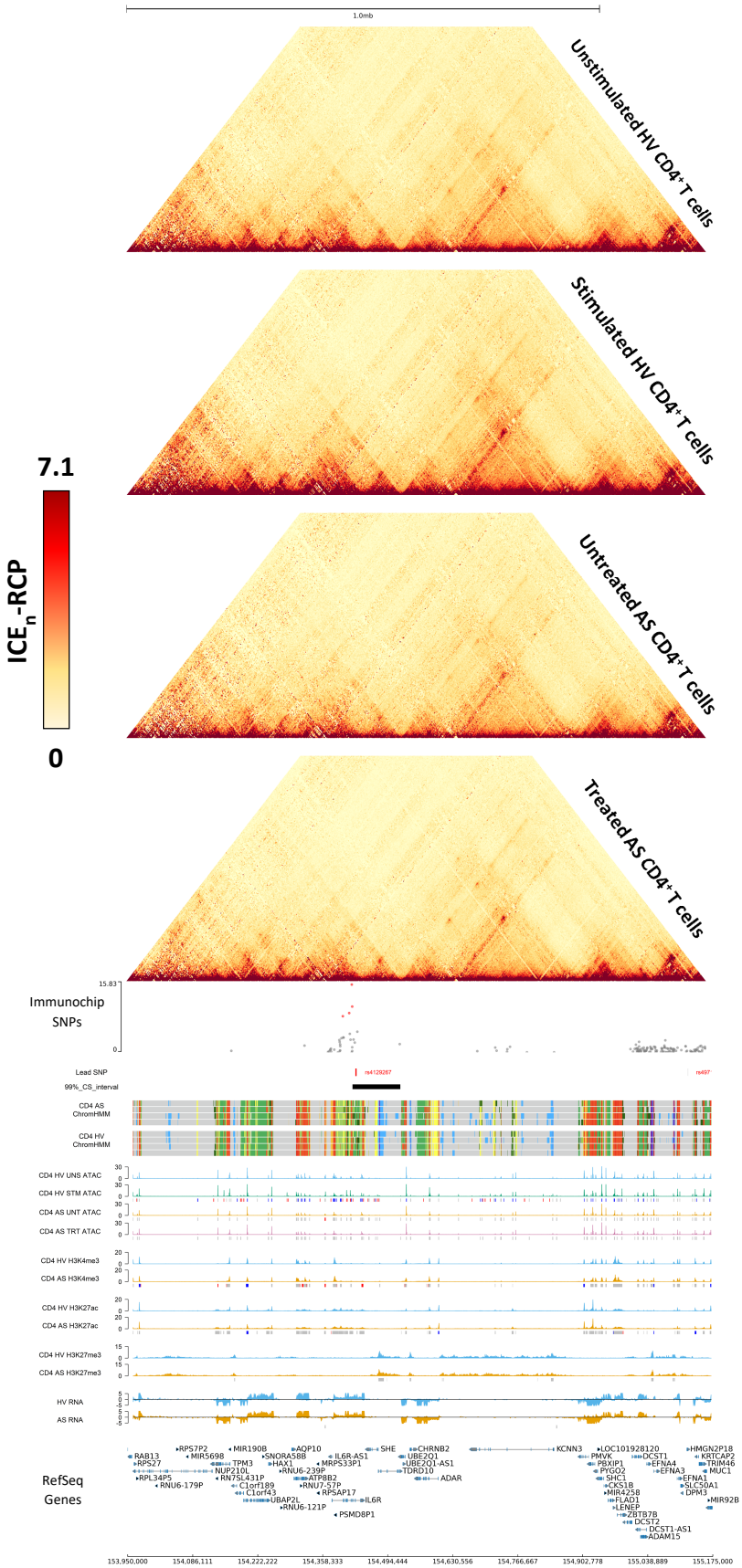

Fig S.12. Tiled-C interactions and the chromatin landscape of the *IL6R* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:154,950,000-155,175,000

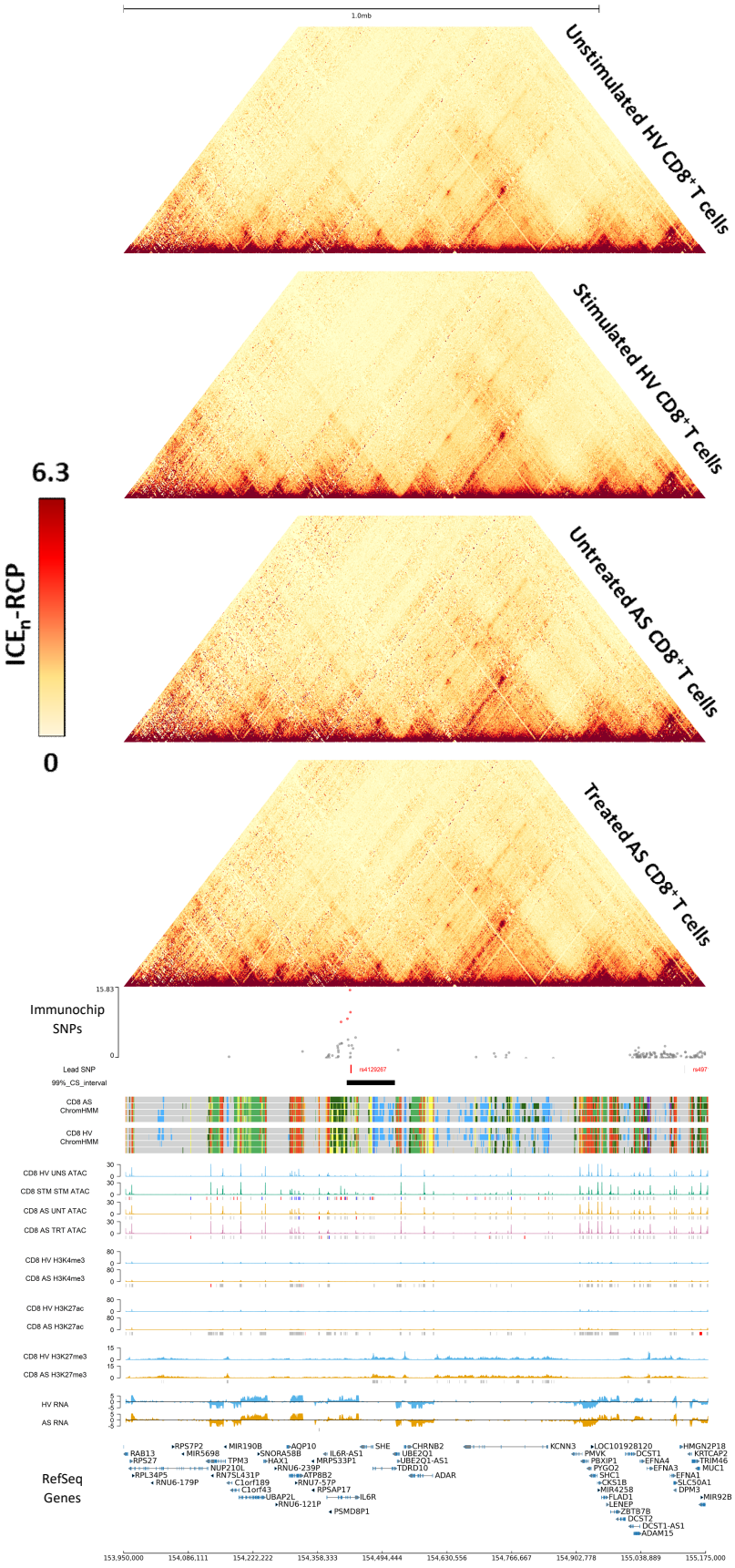

Fig S.13. Tiled-C interactions and the chromatin landscape of the *IL6R* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:154,950,000-155,175,000

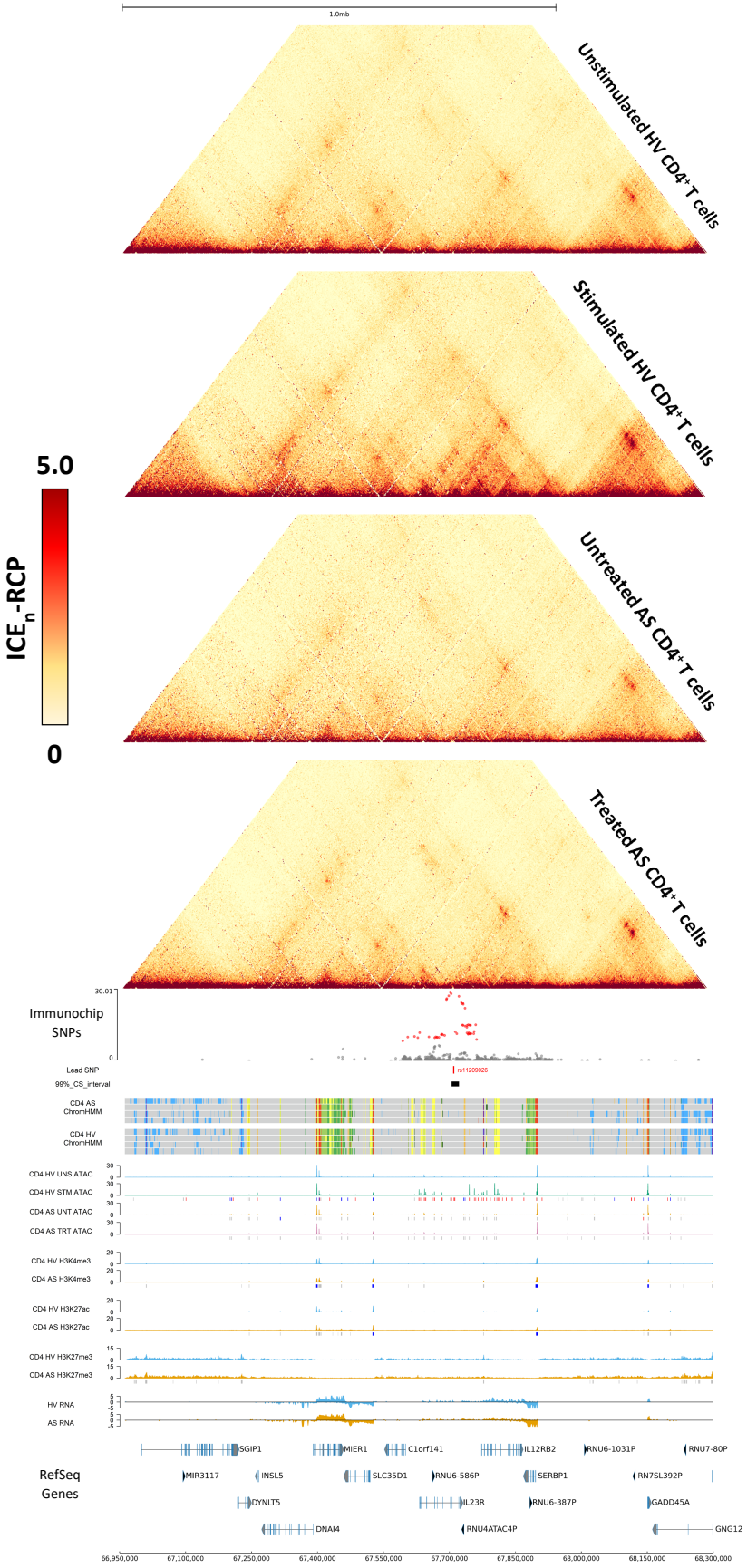

Fig S.14. Tiled-C interactions and the chromatin landscape of the *IL12RB2-IL23R* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:66,950,000-68,300,000.

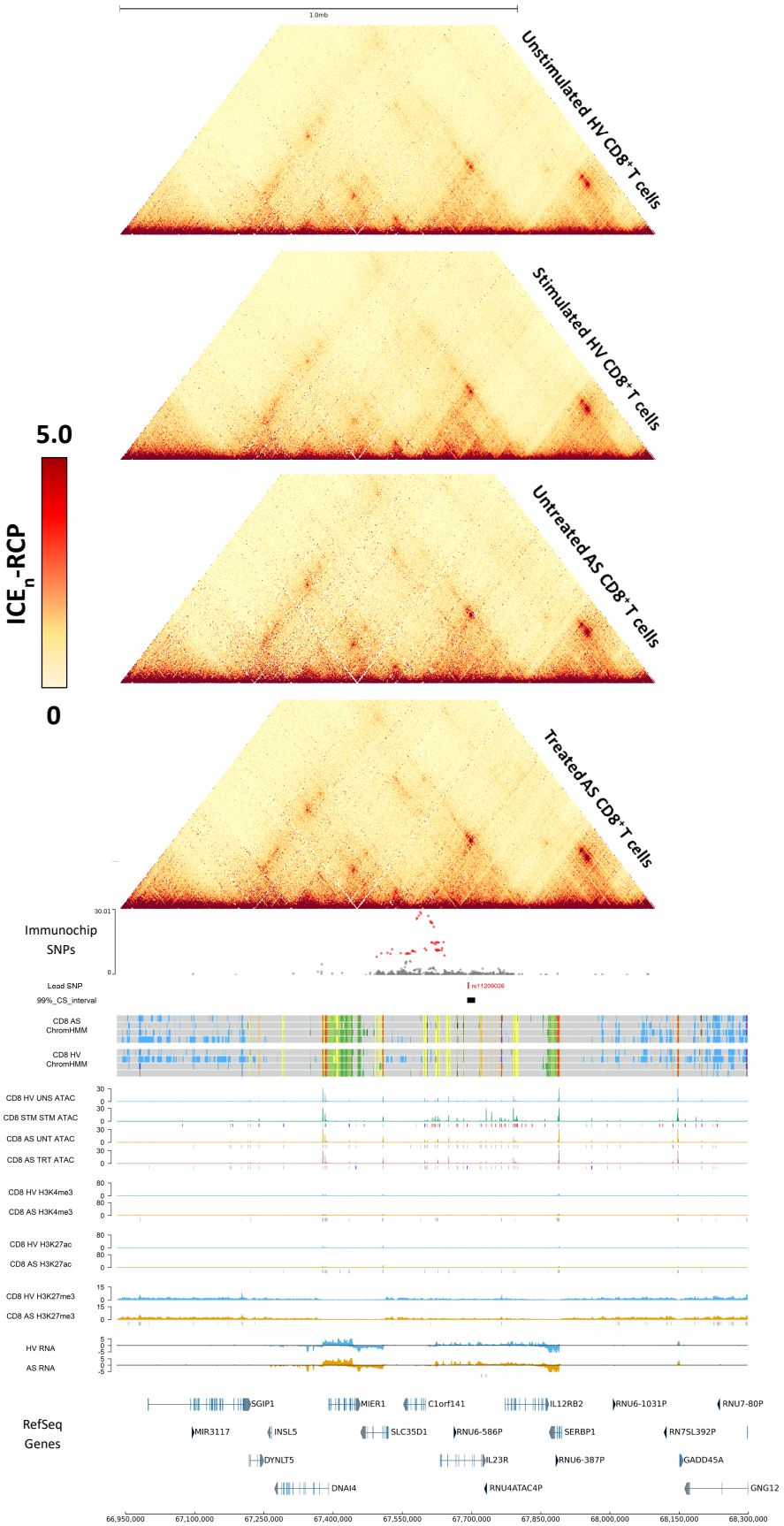

Fig S.15. Tiled-C interactions and the chromatin landscape of the *IL12RB2-IL23R* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:66,950,000-68,300,000.

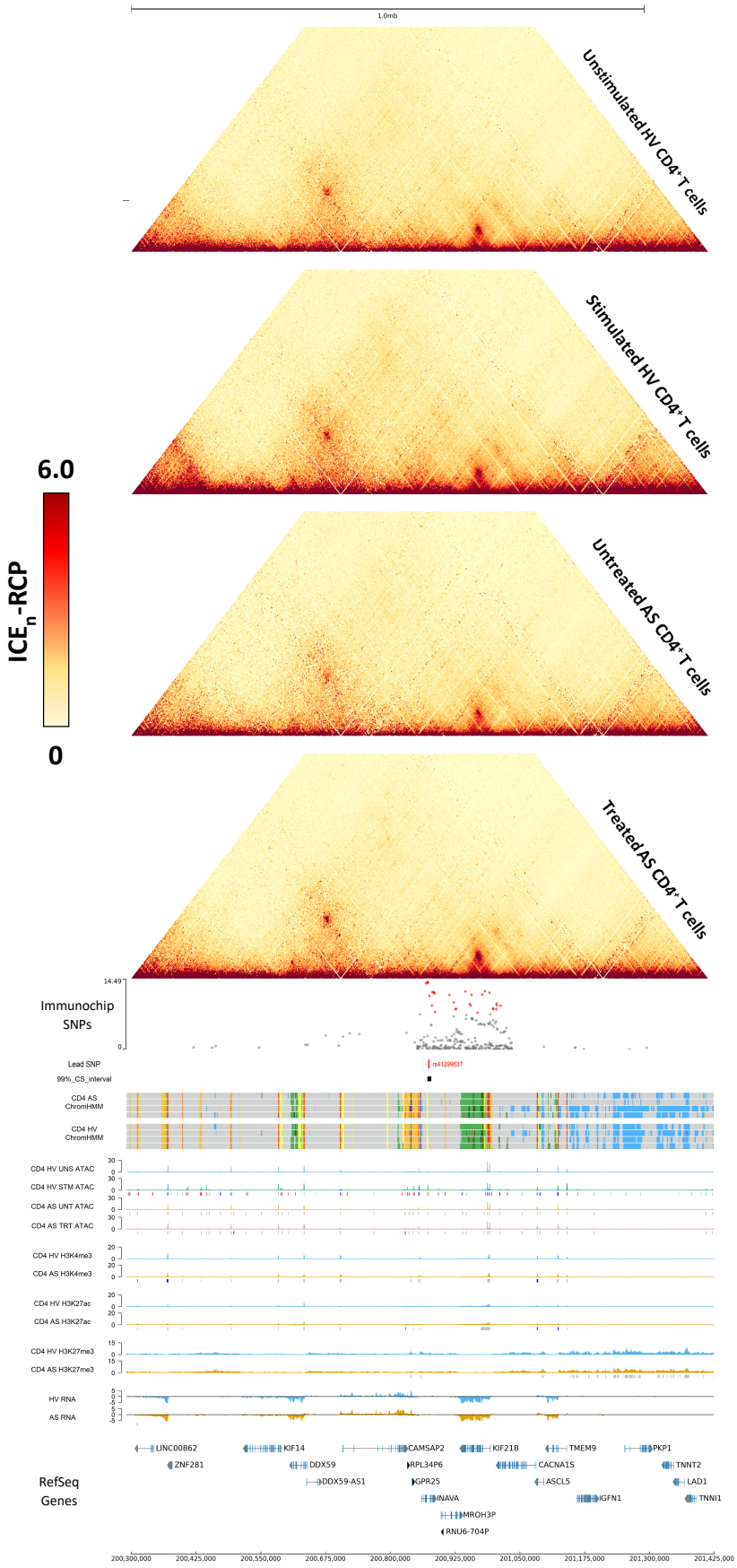

Fig S.16. Tiled-C interactions and the chromatin landscape of the *C1orf106-(INAVA)-GPR25* locus in CD4$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:200,300,000-201,425,000.

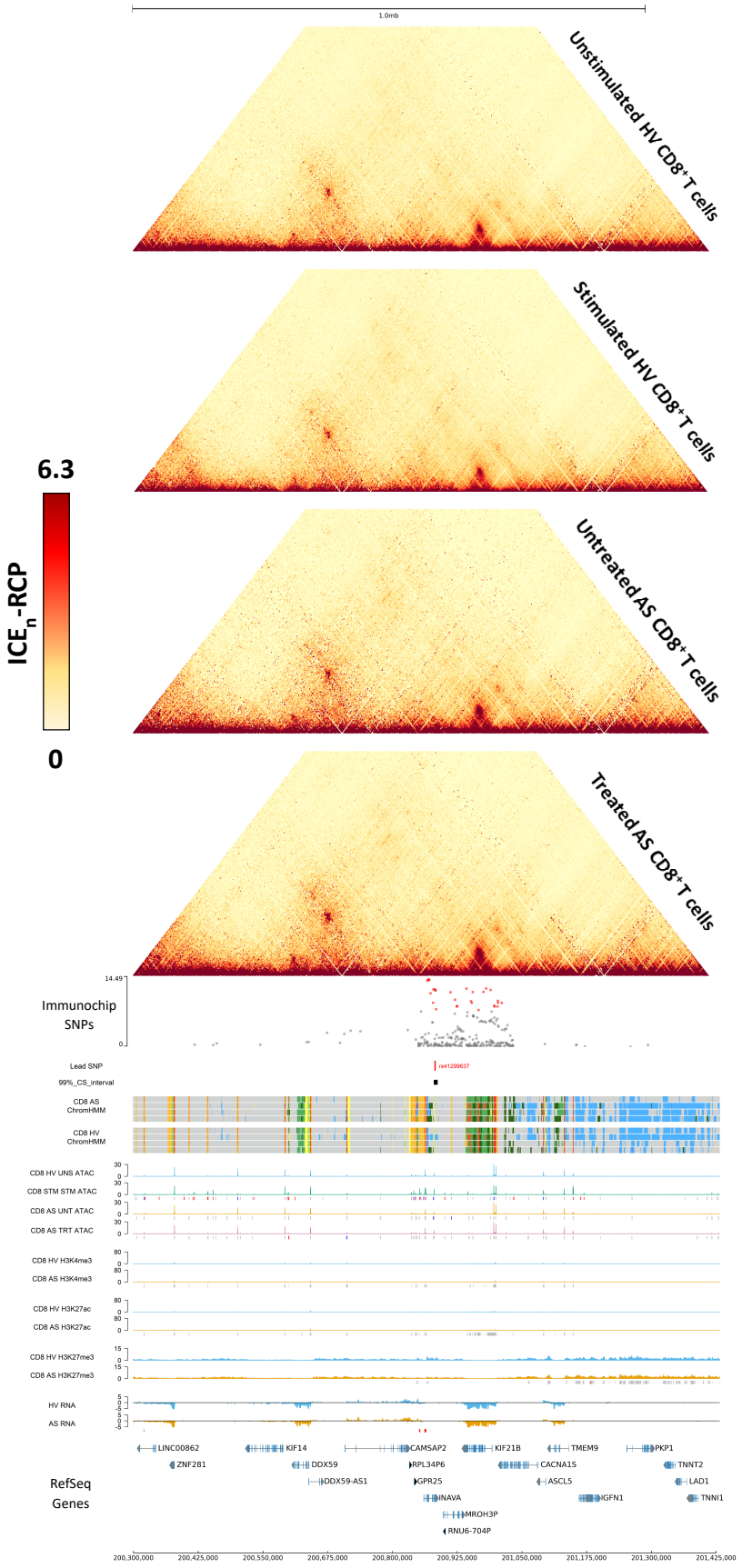

Fig S.17. Tiled-C interactions and the chromatin landscape of the *C1orf106-(INAVA)-GPR25* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:200,300,000-201,425,000.

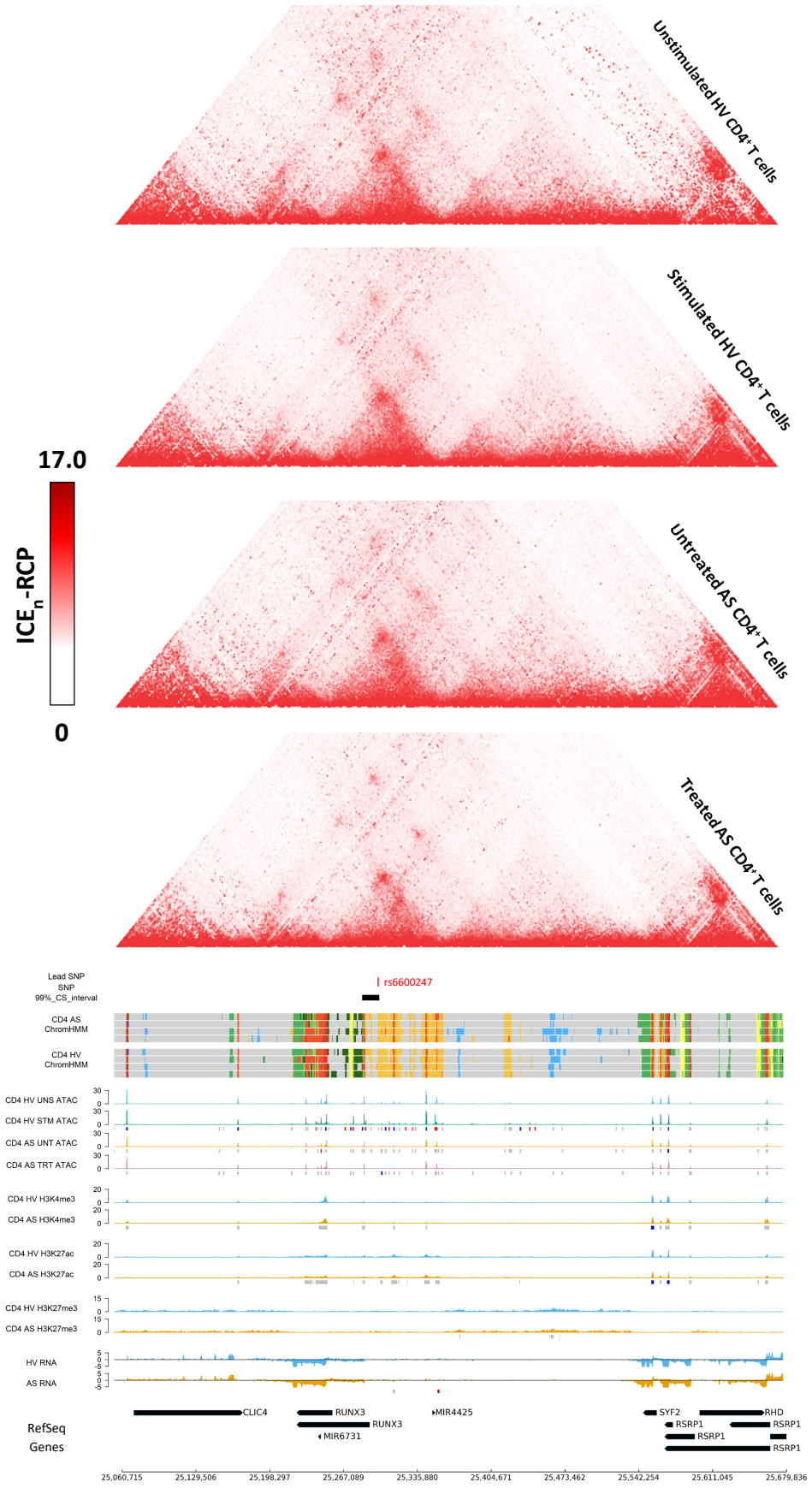

Fig S.18. Tiled-C interactions and the chromatin landscape of the *RUNX3* locus in CD4$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:25,060,000-25,680,000.

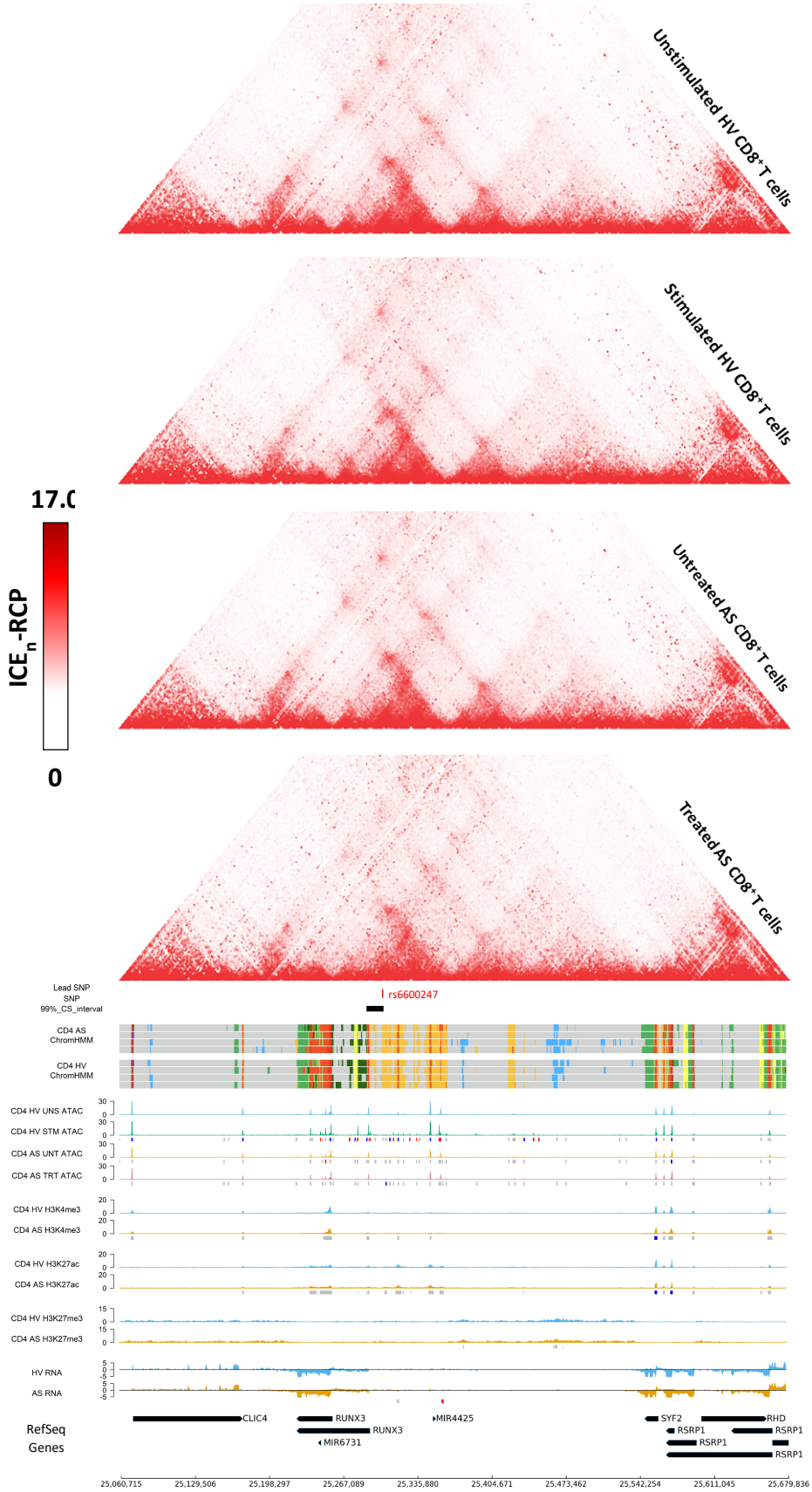

Fig S.19. Tiled-C interactions and the chromatin landscape of the *RUNX3* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr1:25,060,000-25,680,000.

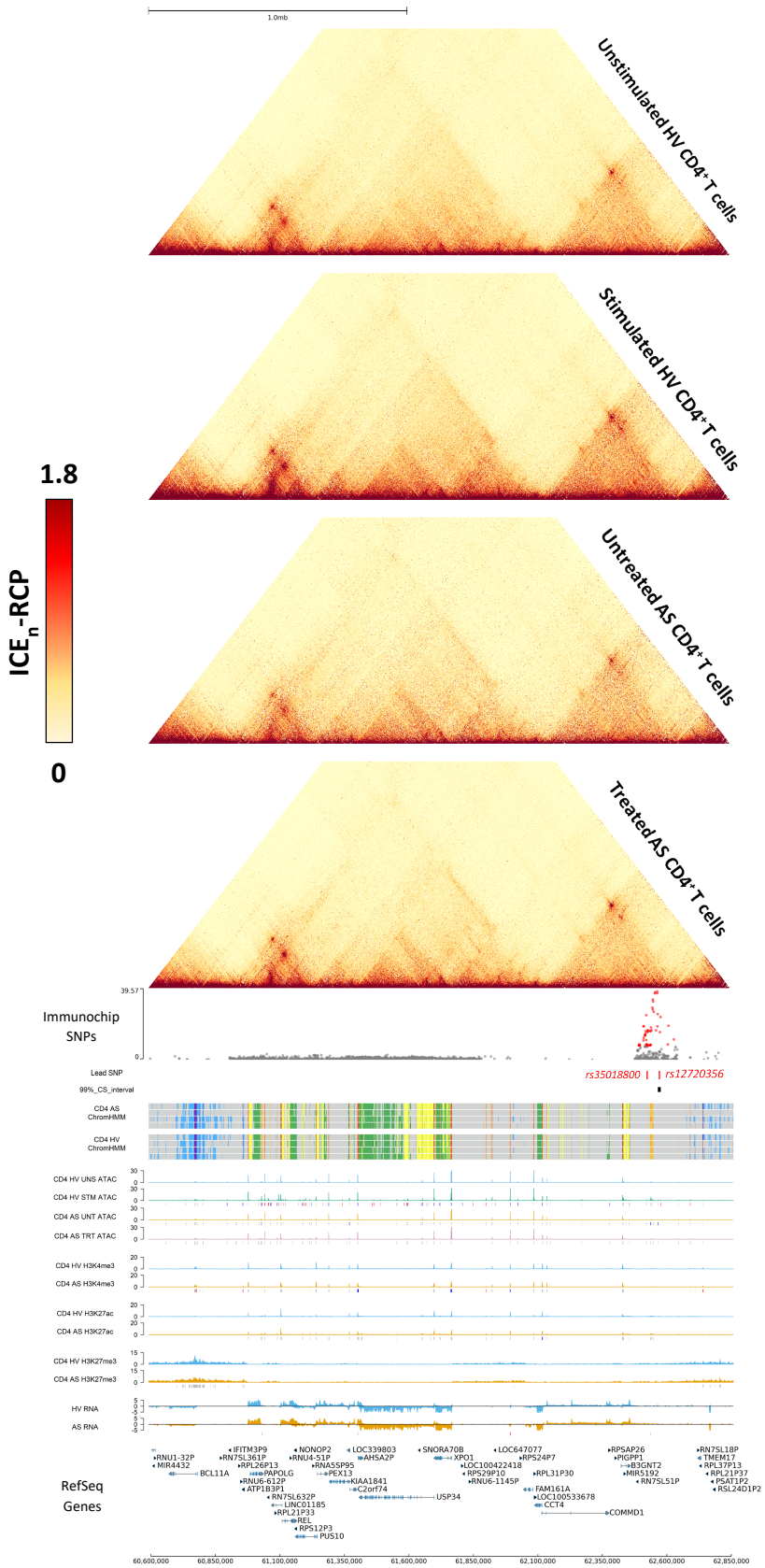

Fig S.20. Tiled-C interactions and the chromatin landscape of the *REL-B3GNT2* locus in CD4$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr2:60,600,000-62,850,000.

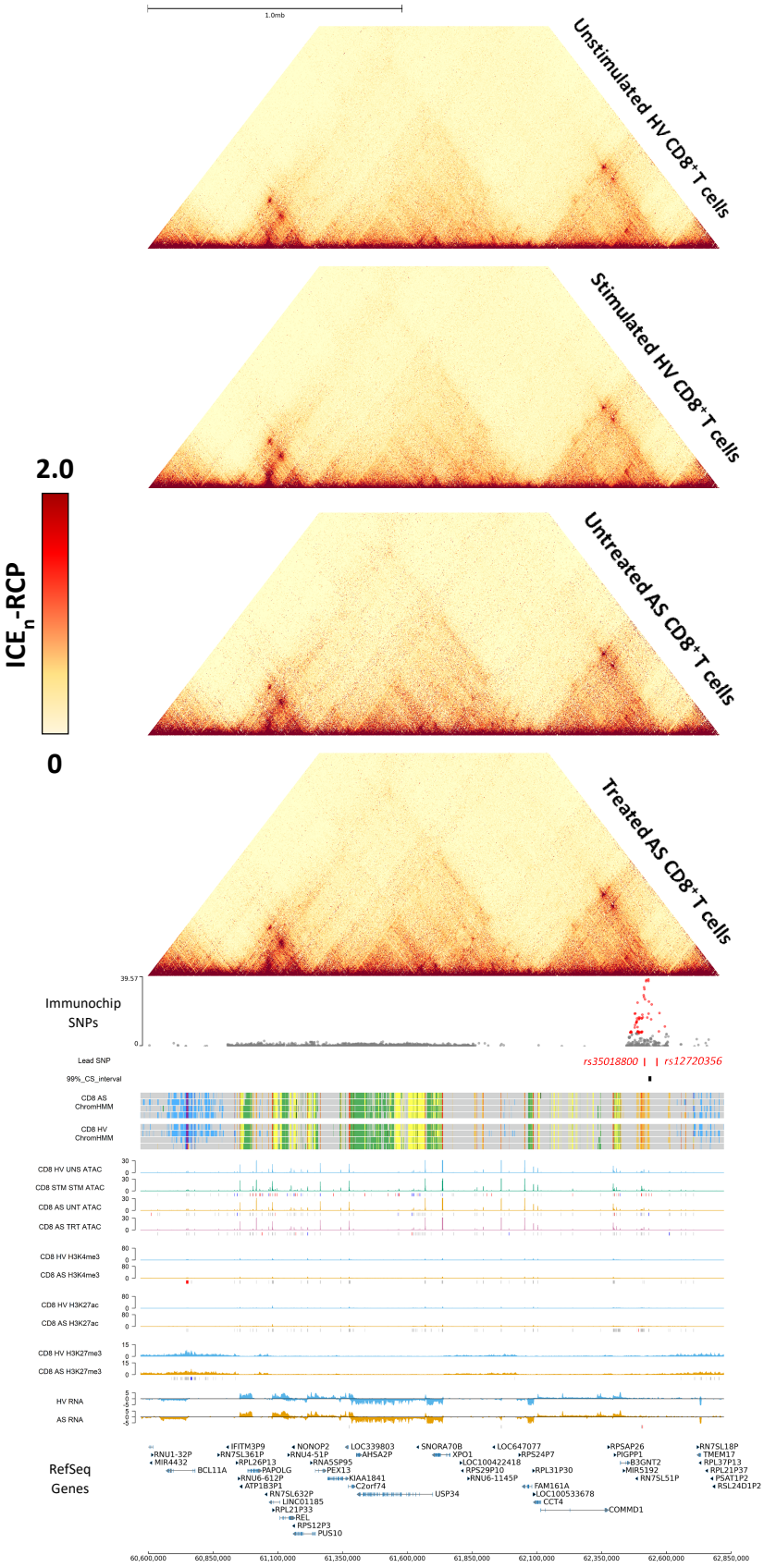

Fig S.21. Tiled-C interactions and the chromatin landscape of the *REL-B3GNT2* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr2:60,600,000-62,850,000.

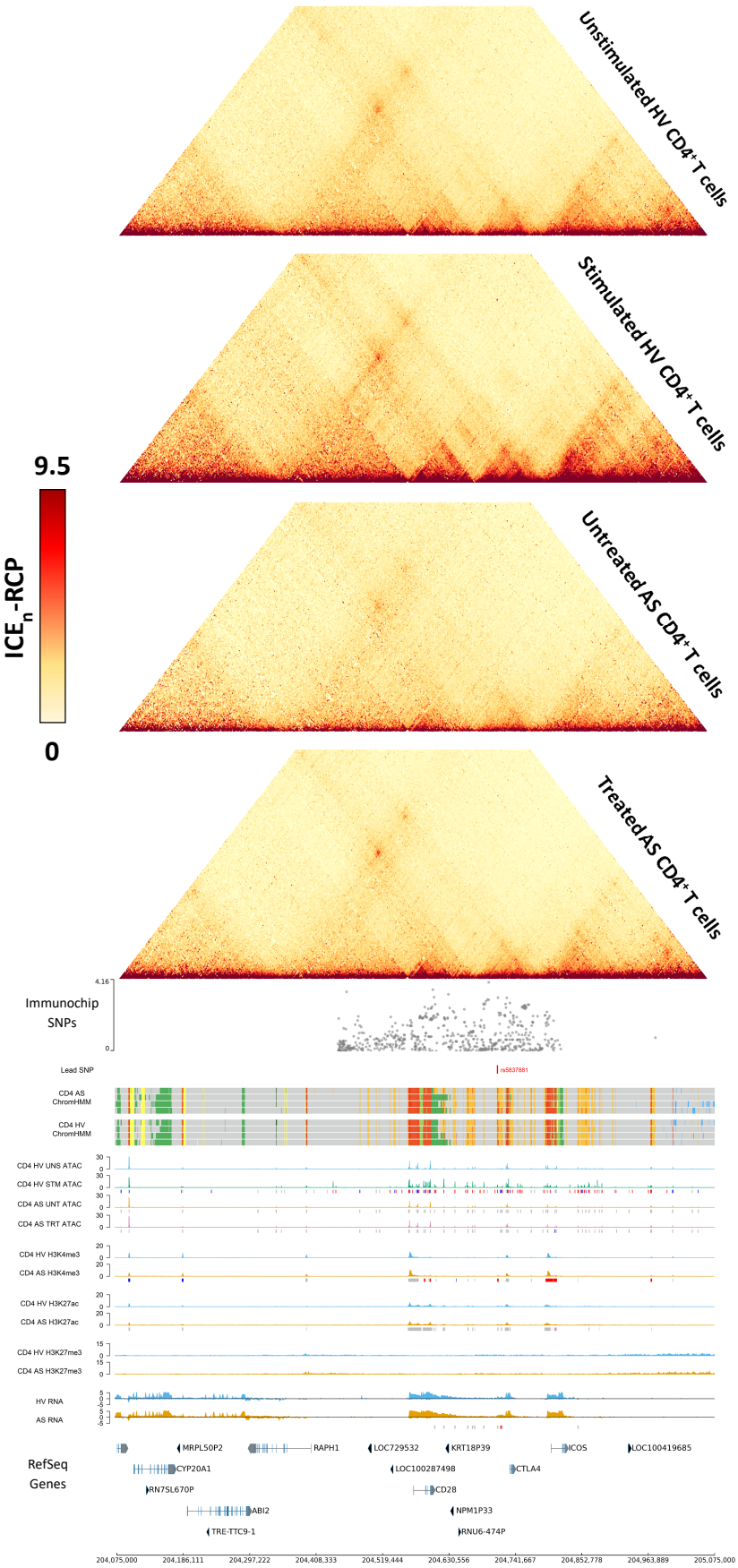

Fig S.22. Tiled-C interactions and the chromatin landscape of the *ICOS-CTLA4-CD28* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr2:204,075,000-20,5075,000.

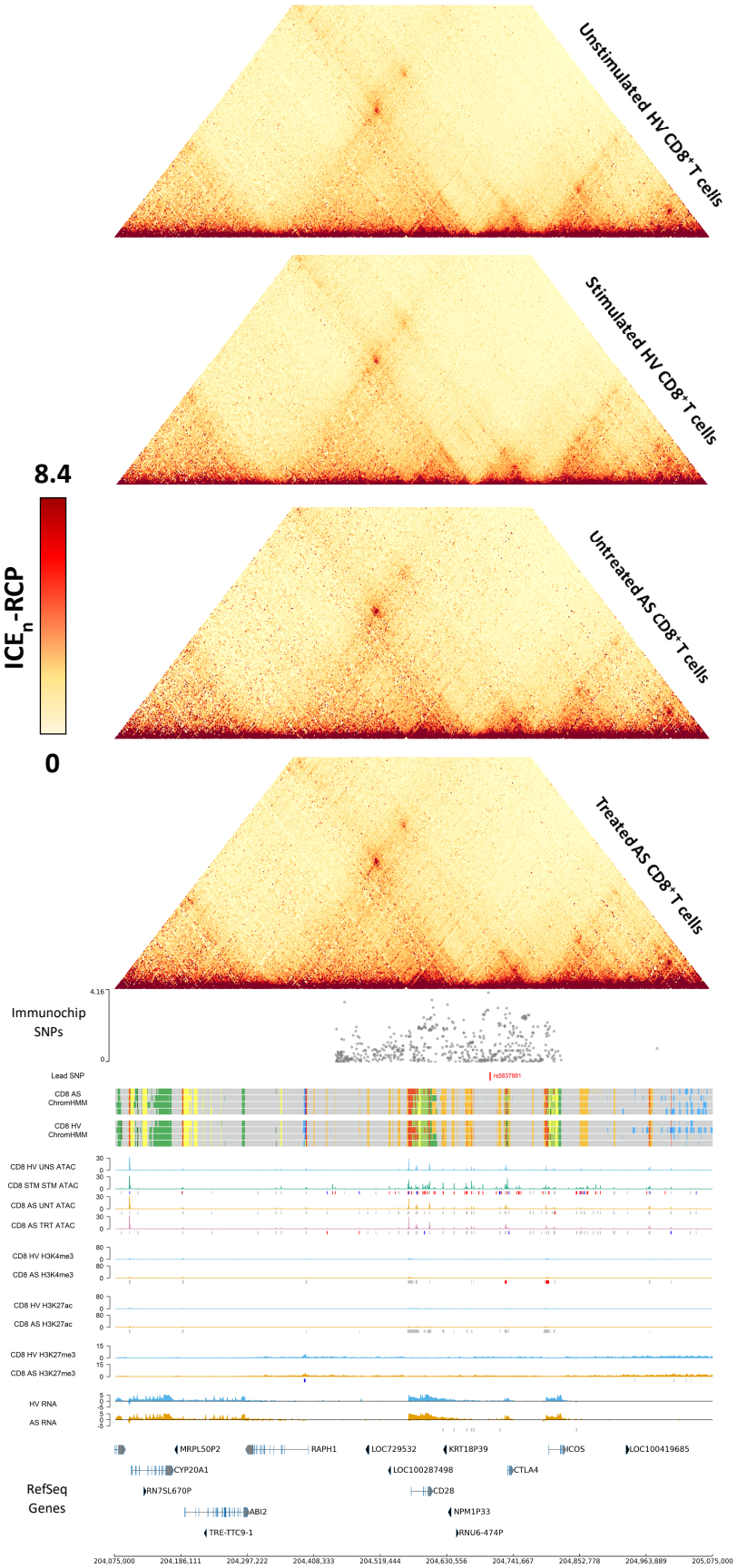

Fig S.23. Tiled-C interactions and the chromatin landscape of the *ICOS-CTLA4-CD28* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr2:204,075,000-20,5075,000.

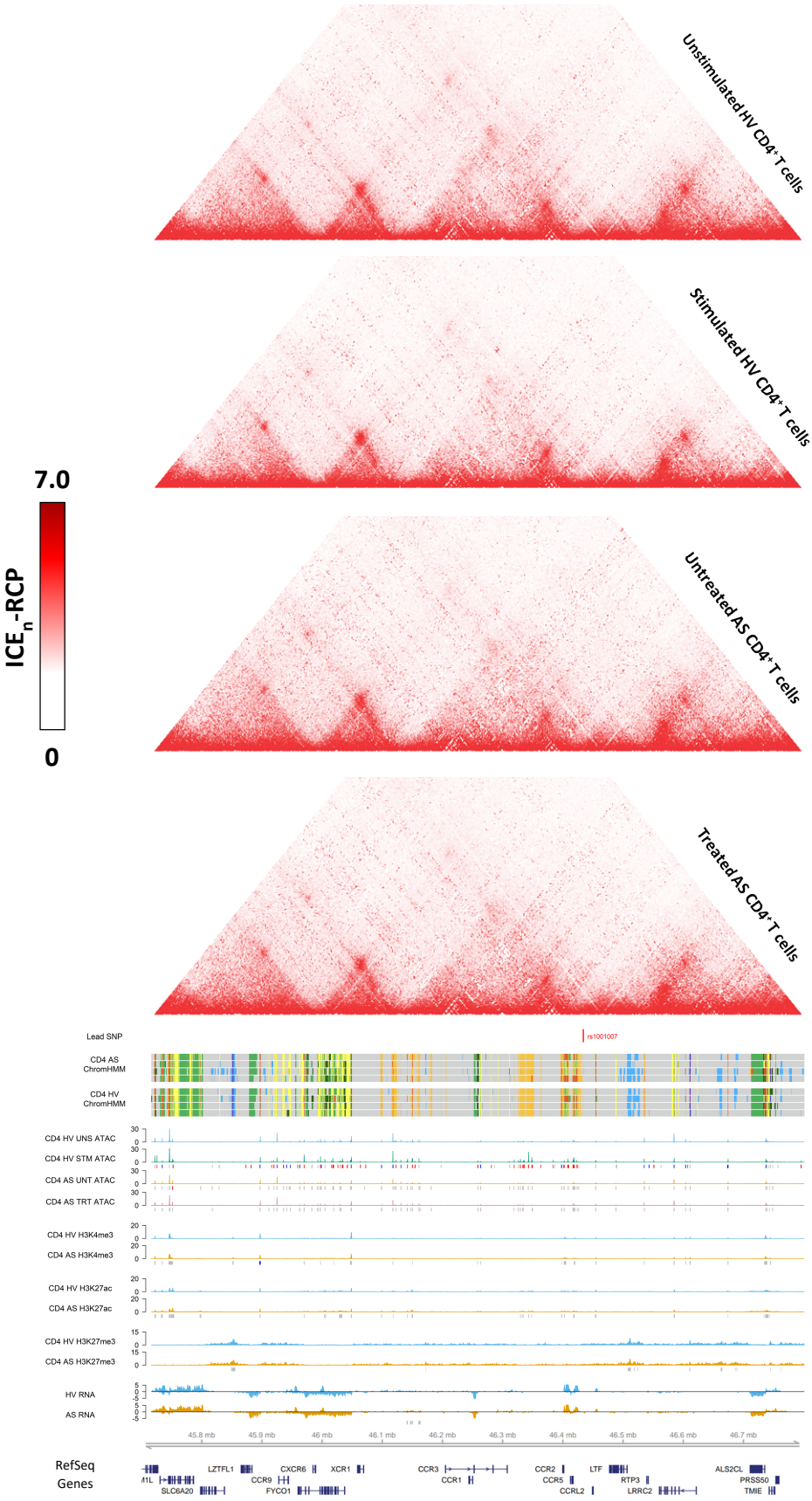

Fig S.24. Tiled-C interactions and the chromatin landscape of the *CCR2-CCR5* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr3:45,700,000-46,800,000.

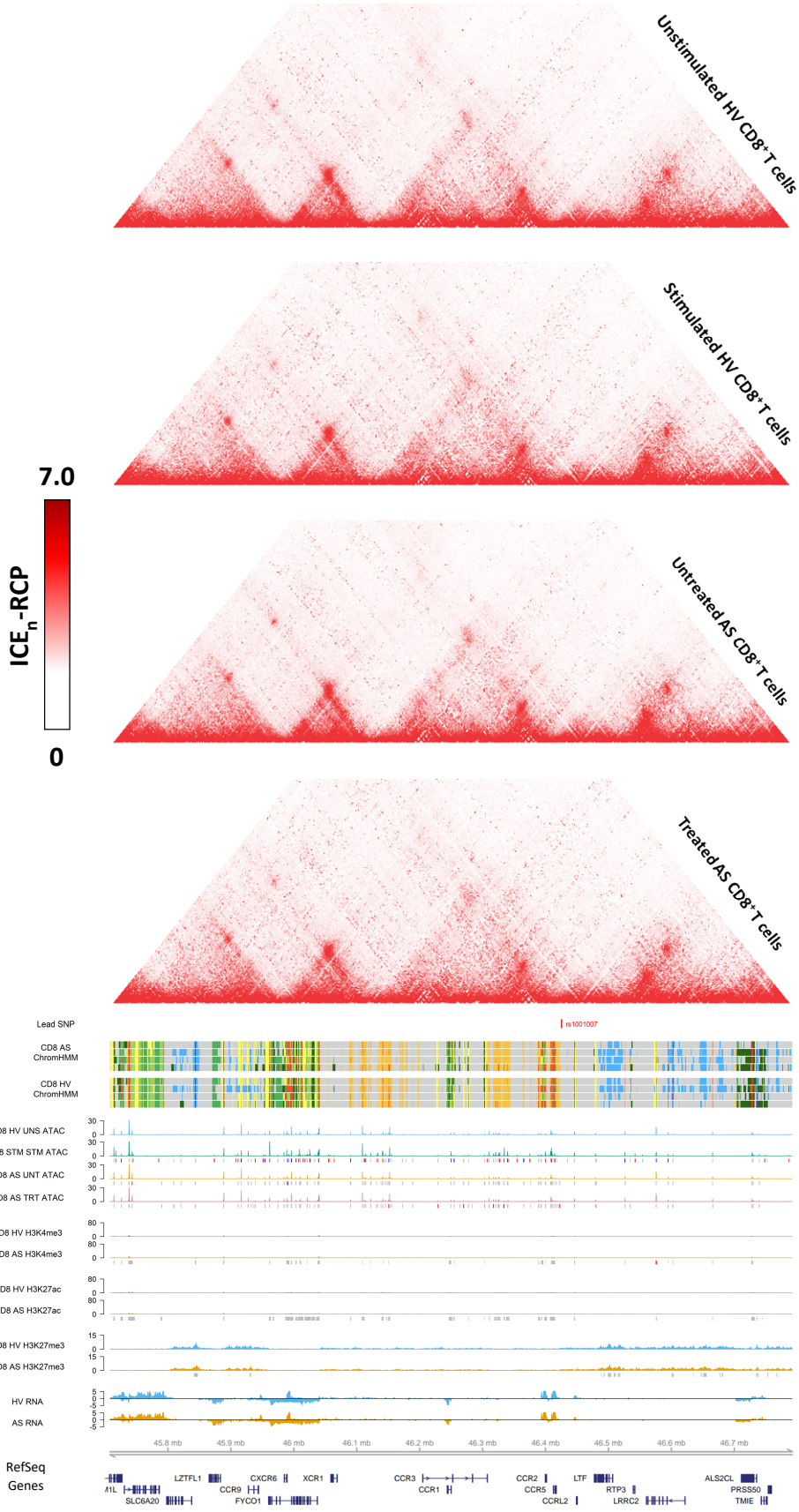

Fig S.25. Tiled-C interactions and the chromatin landscape of the *CCR2-CCR5* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr3:45,700,000-46,800,000.

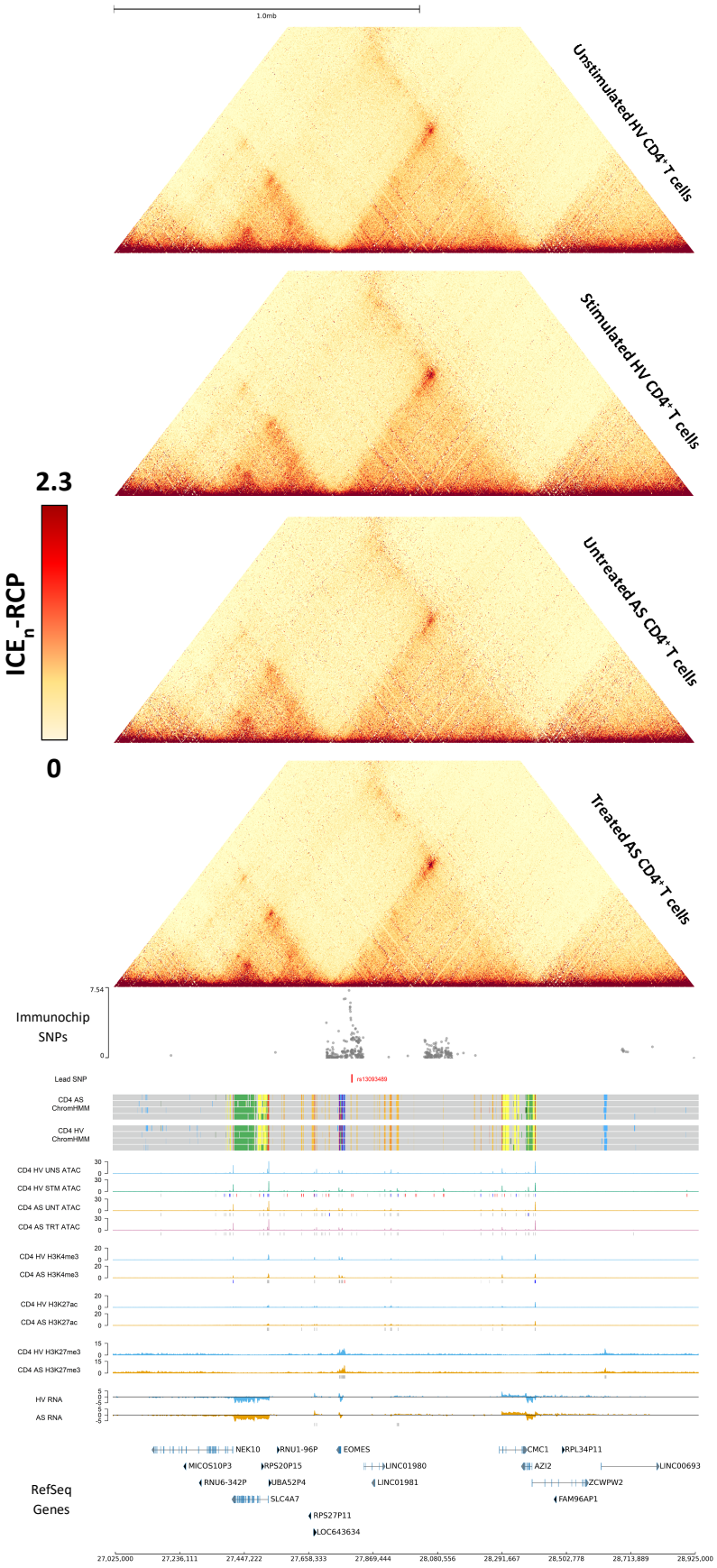

Fig S.26. Tiled-C interactions and the chromatin landscape of the *EOMES-CMC1* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr3:27,025,000-28,925,000.

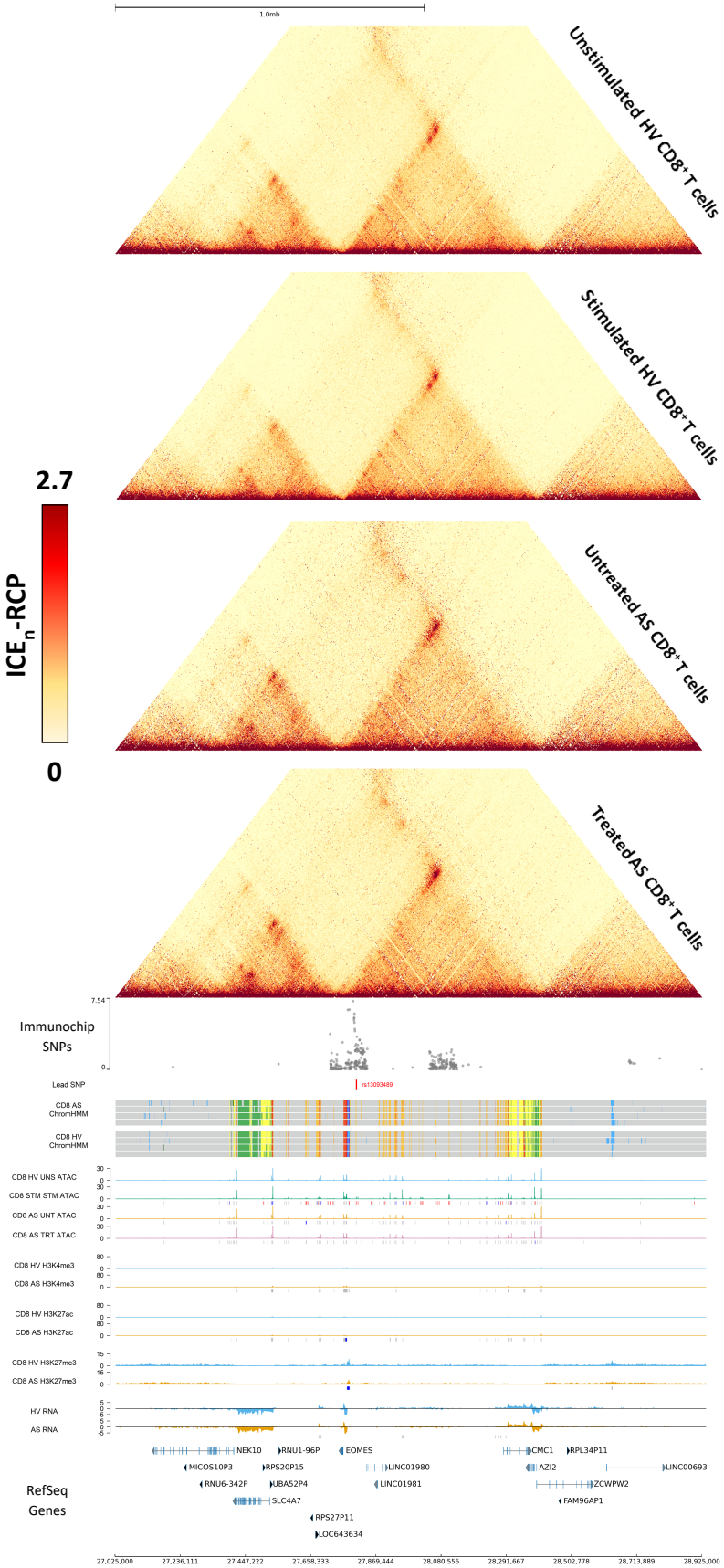

Fig S.27. Tiled-C interactions and the chromatin landscape of the *EOMES-CMC1* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr3:27,025,000-28,925,000.

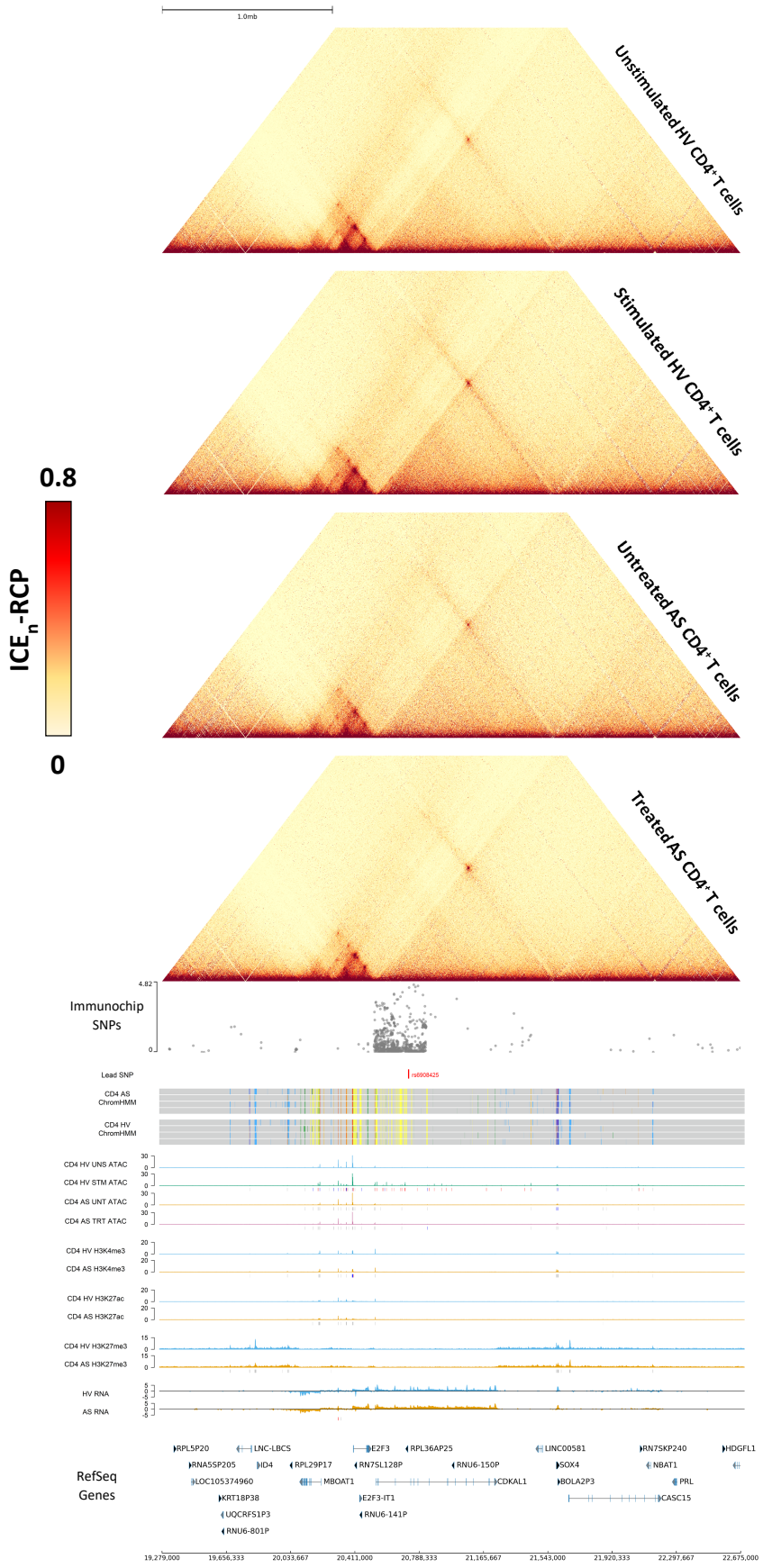

Fig S.28. Tiled-C interactions and the chromatin landscape of the *CDKAL1* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr6:19,279,000-22,675,000.

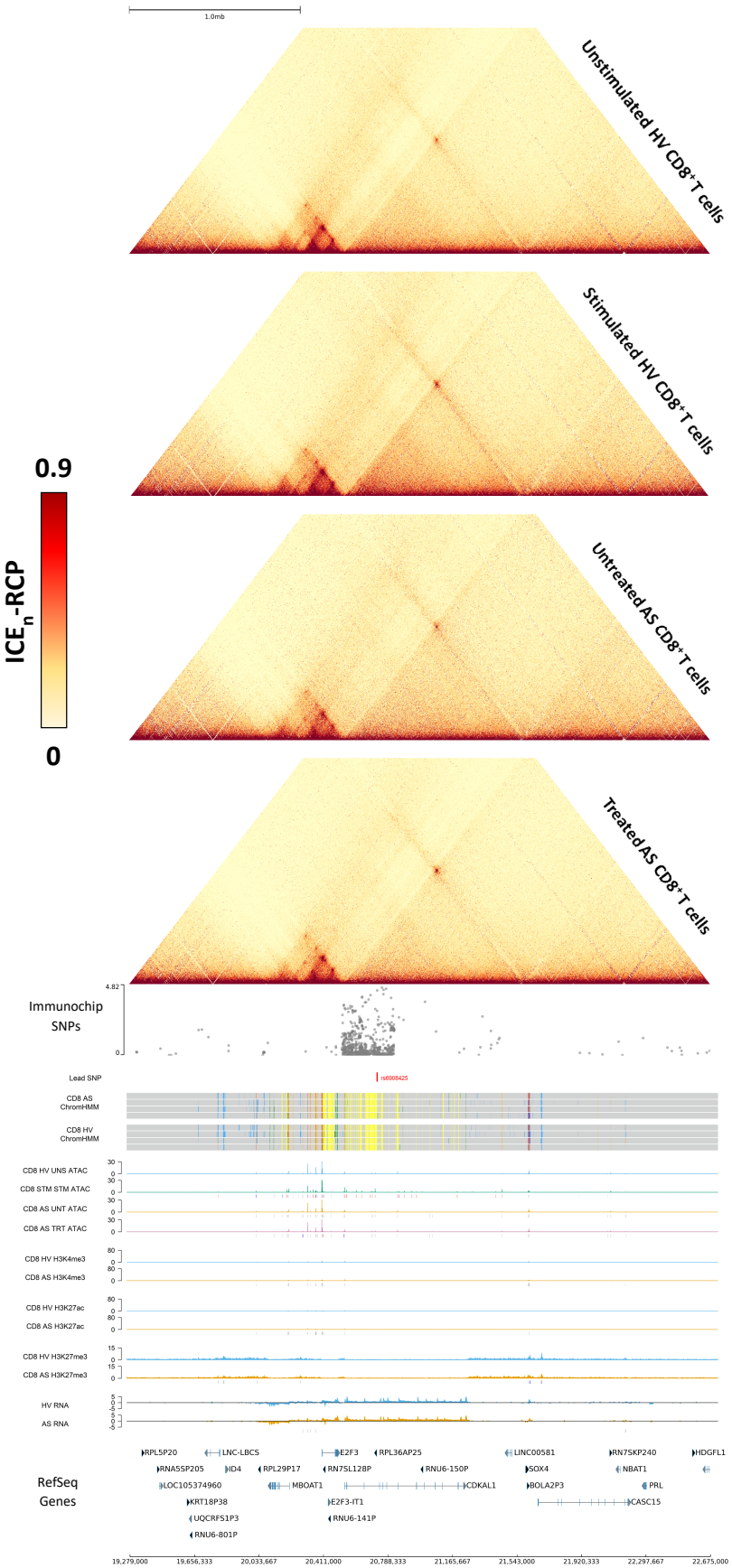

Fig S.29. Tiled-C interactions and the chromatin landscape of the *CDKAL1* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr6:19,279,000-22,675,000.

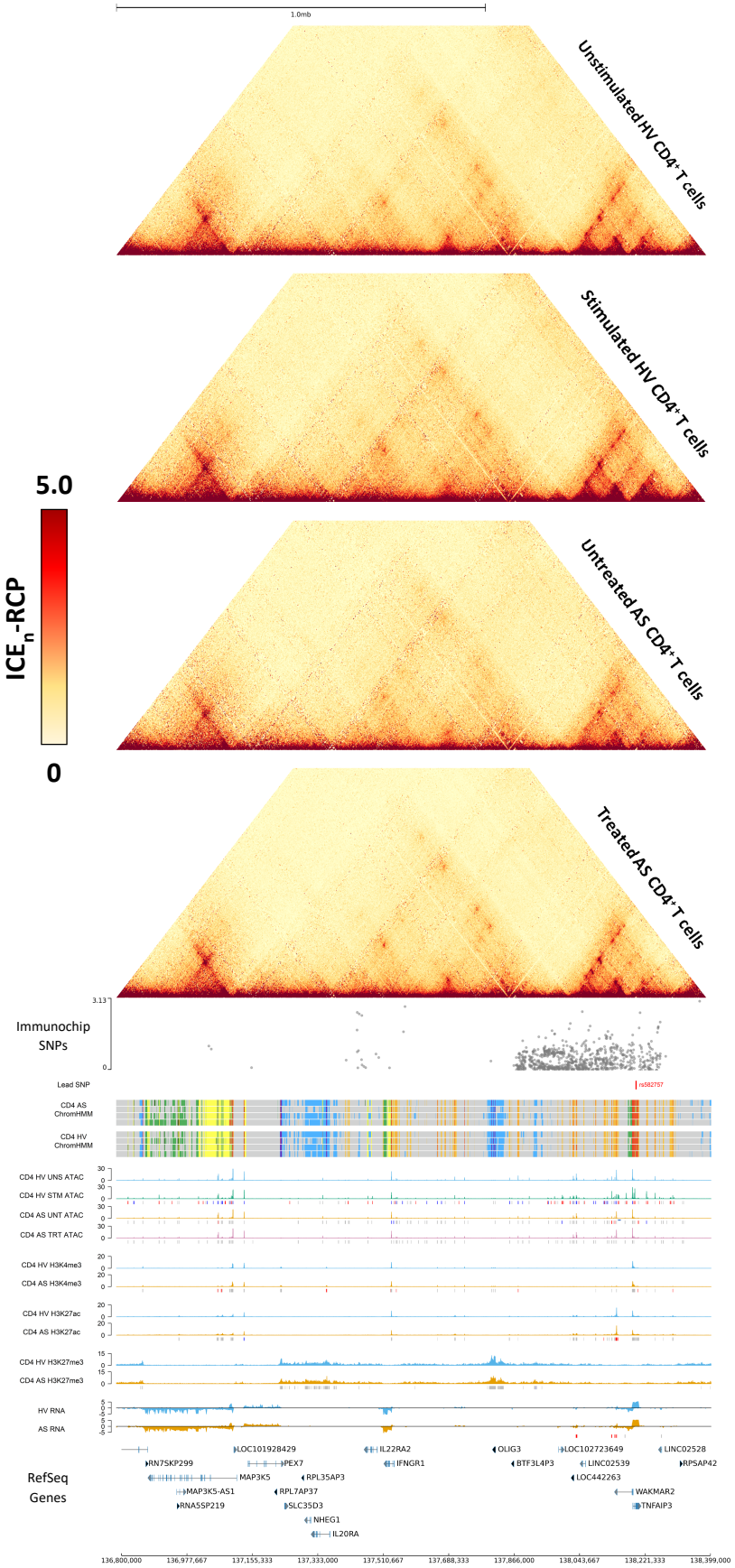

Fig S.30. Tiled-C interactions and the chromatin landscape of the *IL22RA2-TNFAIP3* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr6:136,800,000-138,399,000.

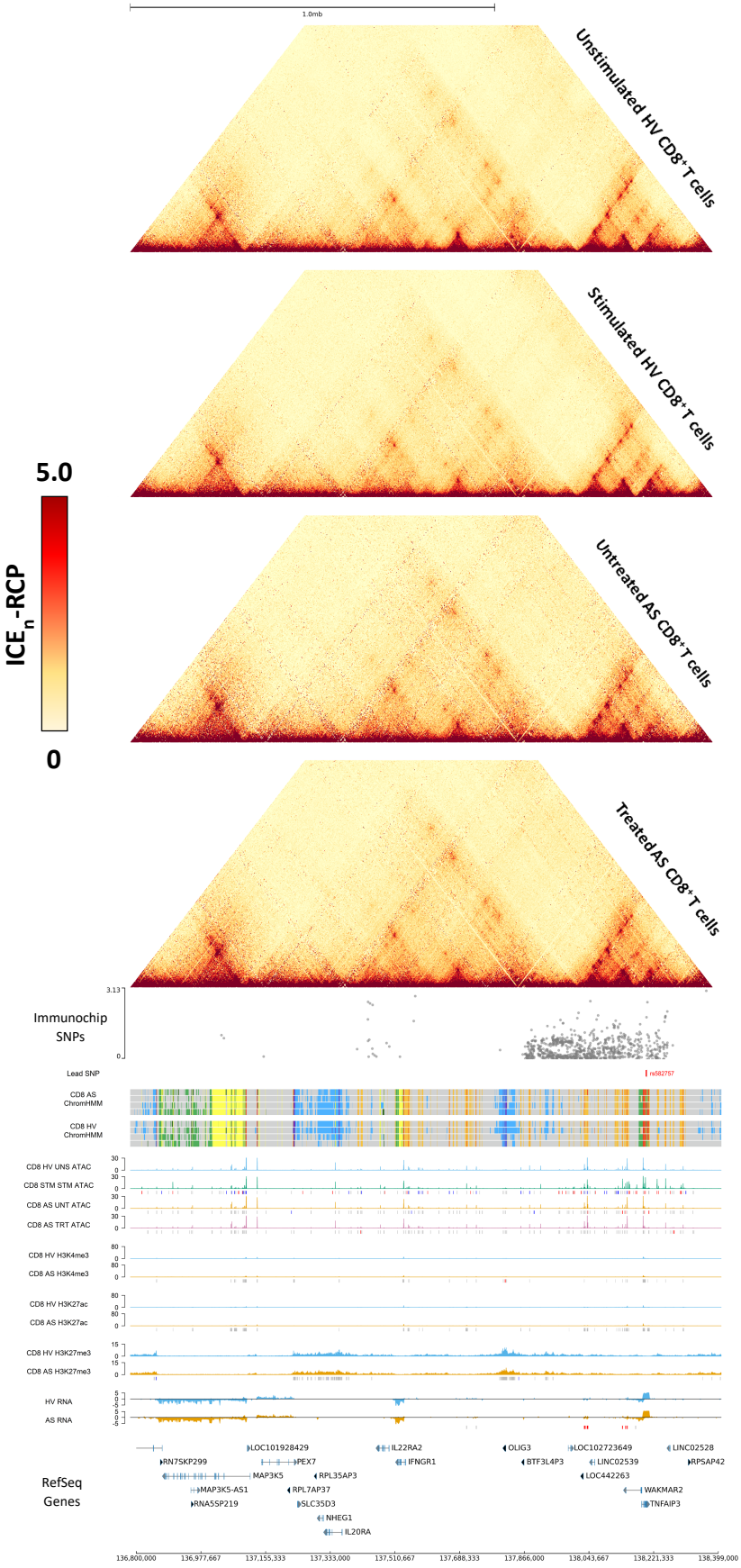

Fig S.31. Tiled-C interactions and the chromatin landscape of the *ILL2RA2-TNFAIP3* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr6:136,800,000-138,399,000.

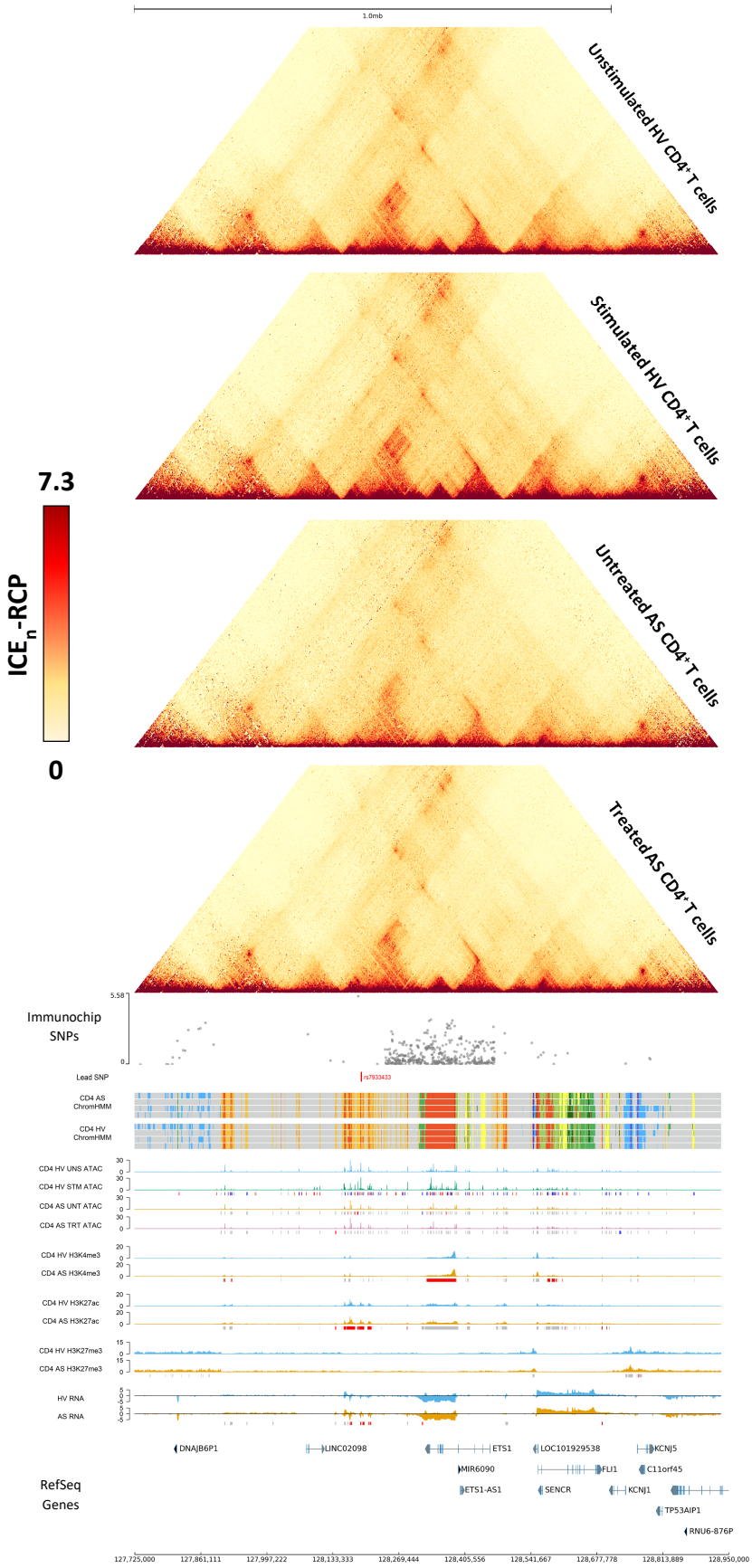

Fig S.32. Tiled-C interactions and the chromatin landscape of the *ETS1* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr11:127,725,000-128,950,000.

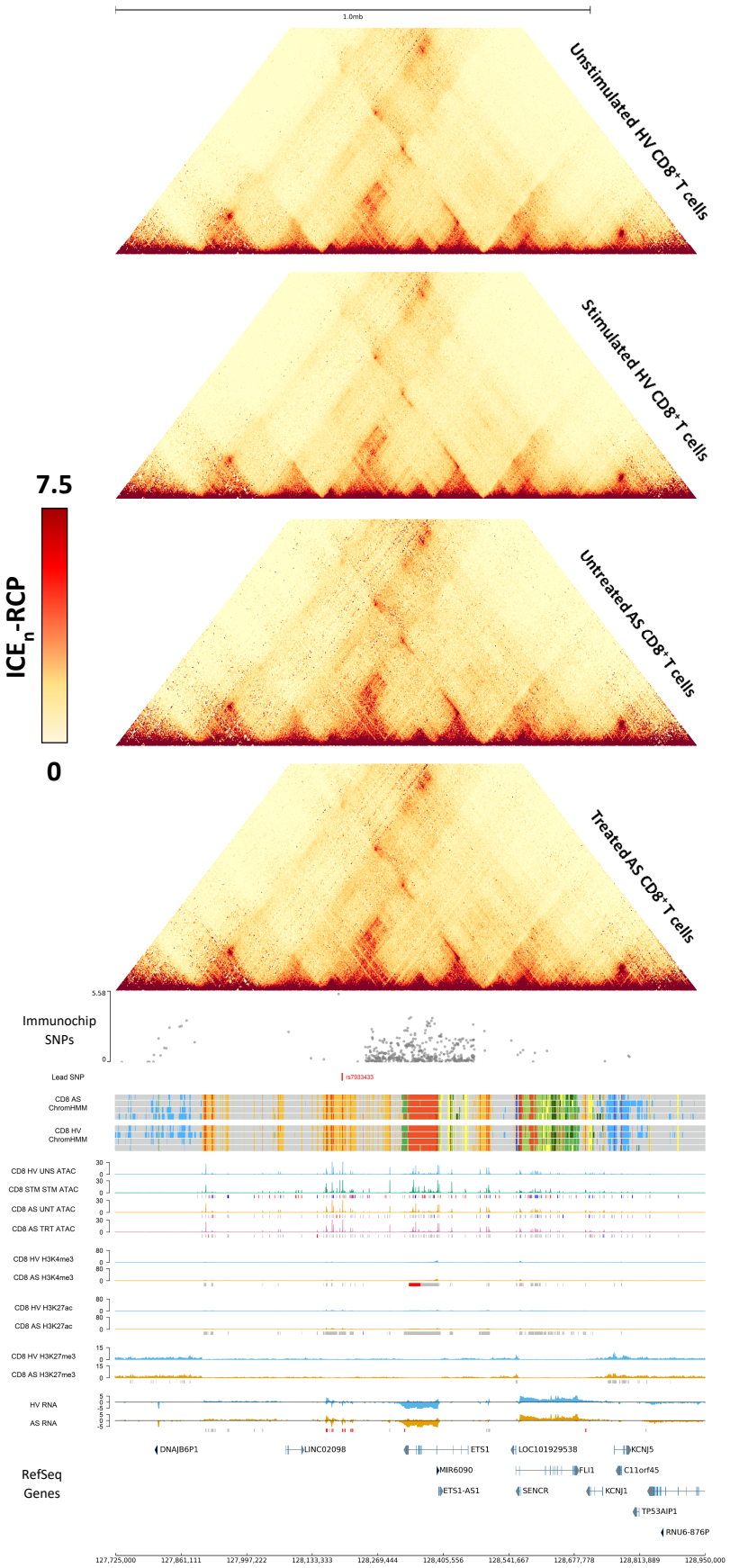

Fig S.33. Tiled-C interactions and the chromatin landscape of the *ETS1* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr11:127,725,000-128,950,000.

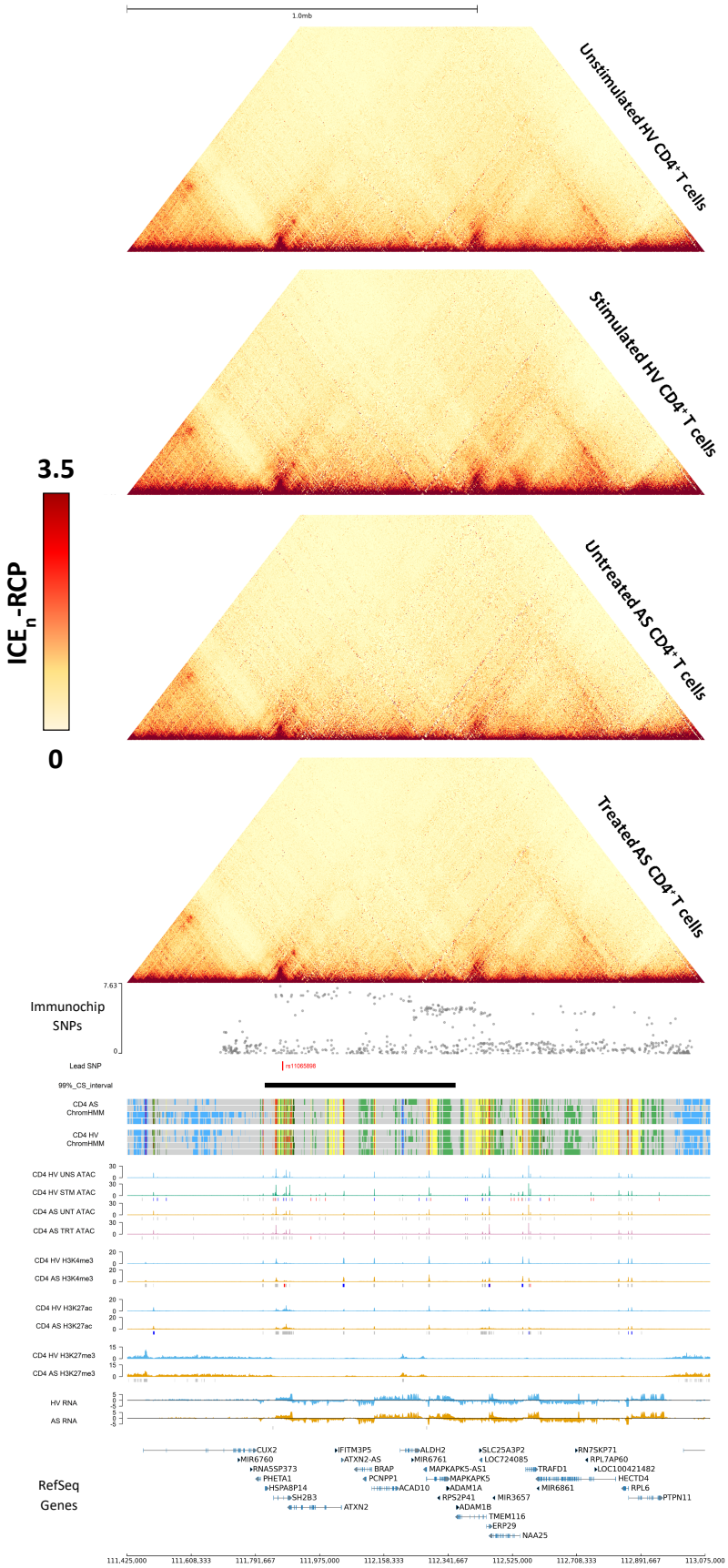

Fig S.34. Tiled-C interactions and the chromatin landscape of the *ATXN2-PTPN11-SH2B3* locus in CD4$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr12:111,425,000-113,075,000.

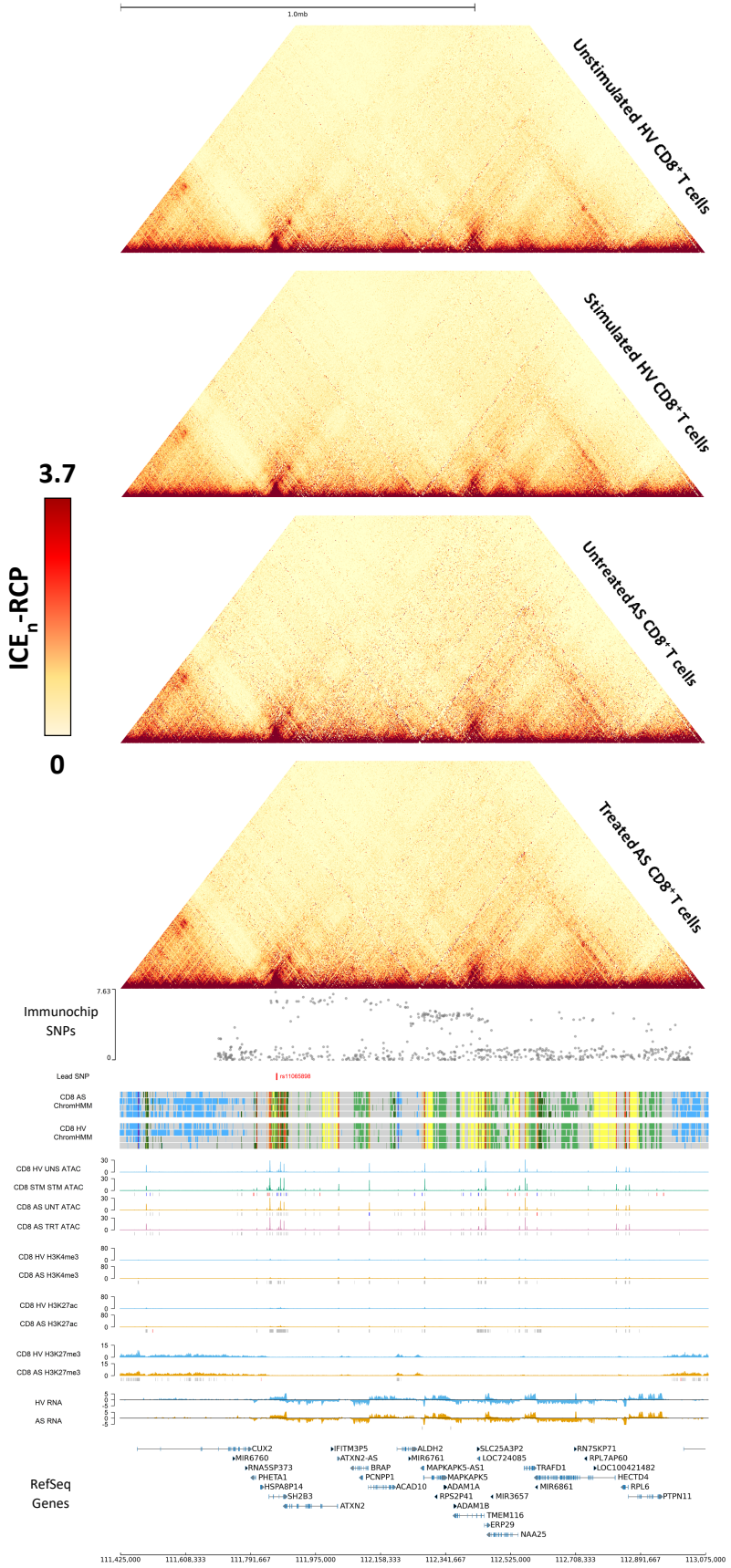

Fig S.35. Tiled-C interactions and the chromatin landscape of the *ATXN2-PTPN11-SH2B3* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr12:111,425,000-113,075,000.

\\

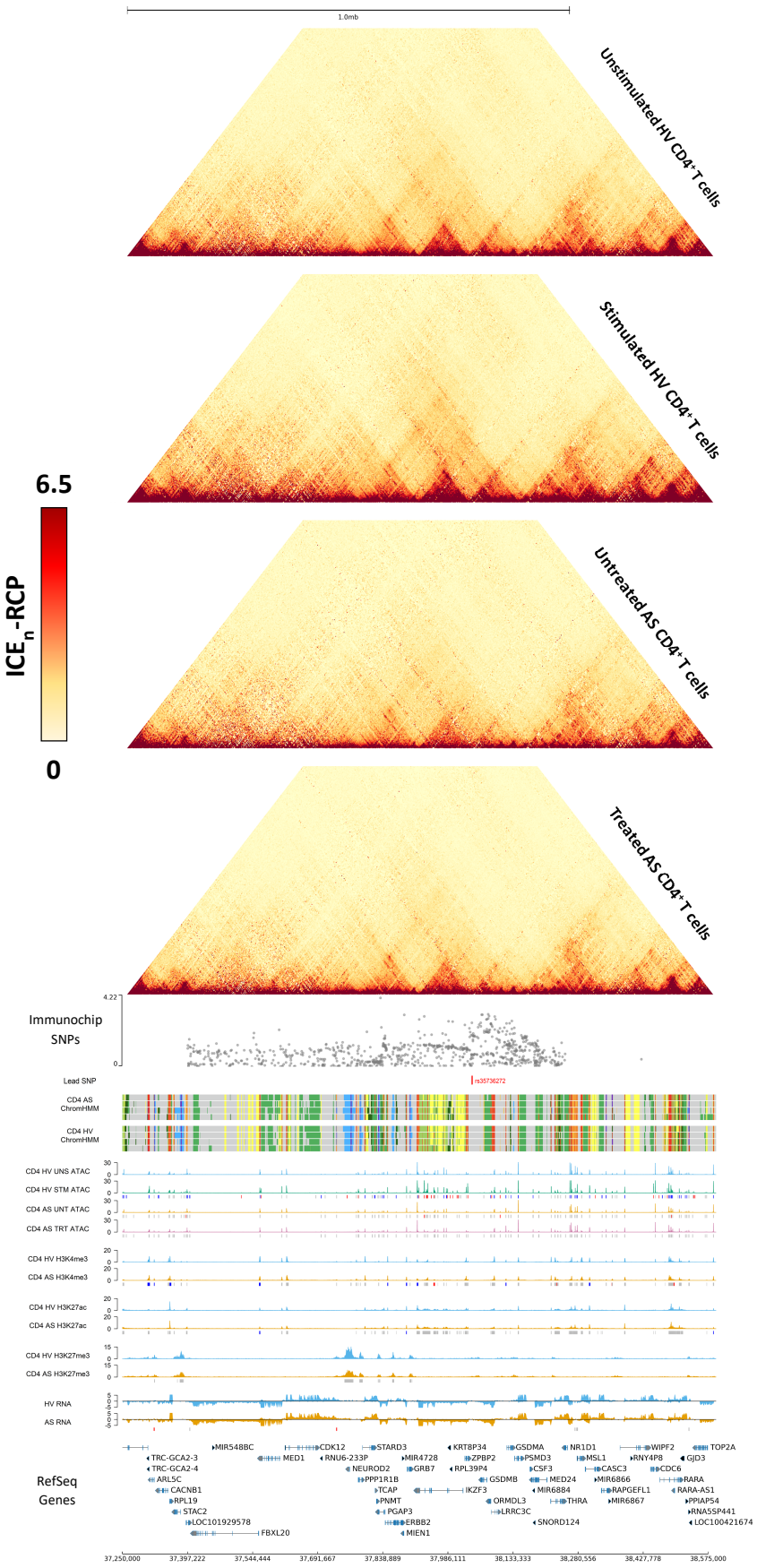

Fig S.36. Tiled-C interactions and the chromatin landscape of the *IKZF3-Aiolos-ORMDL3* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr17: 37,250,000-38,575,000.

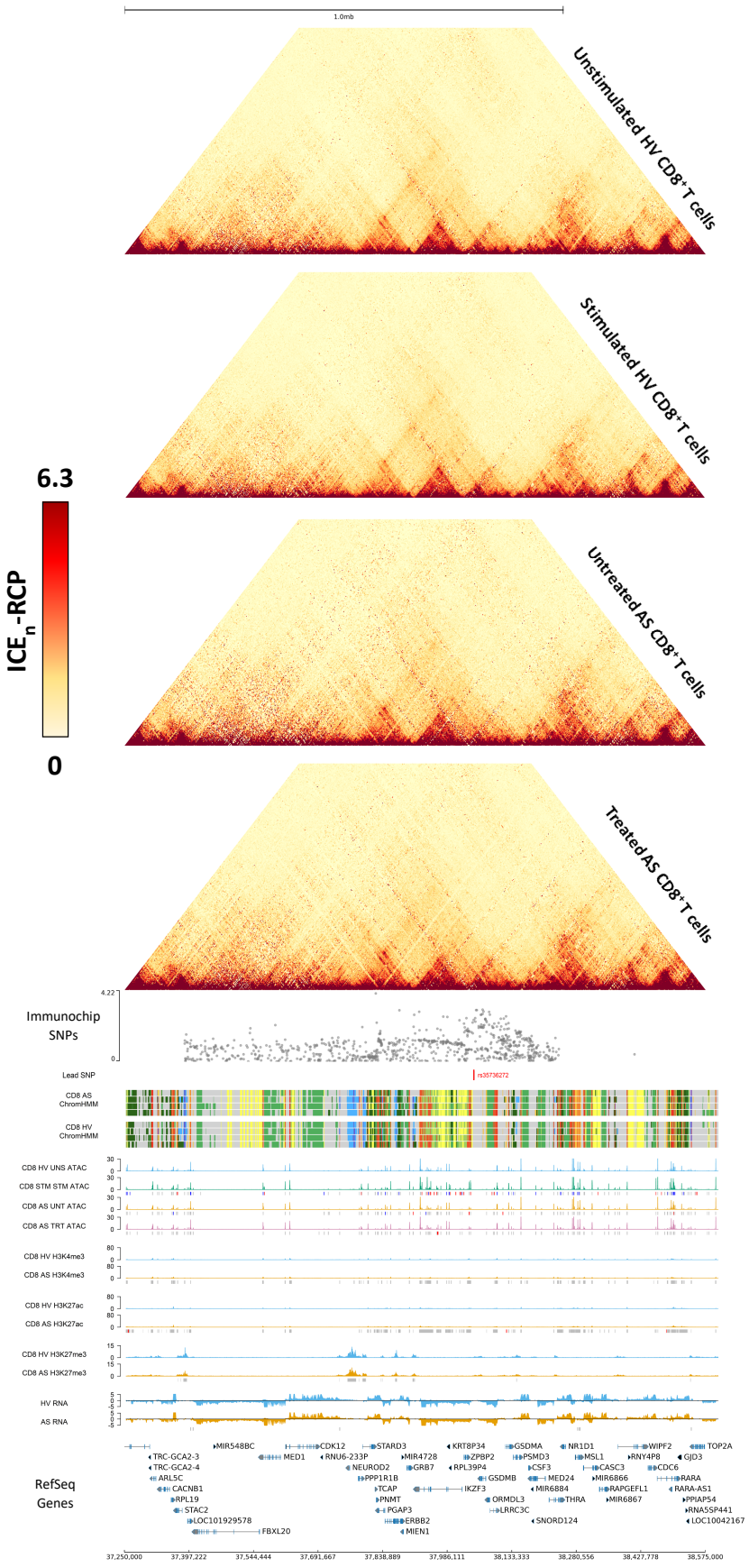

Fig S.37. Tiled-C interactions and the chromatin landscape of the *IKZF3-Aiolos-ORMDL3* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr17: 37,250,000-38,575,000.

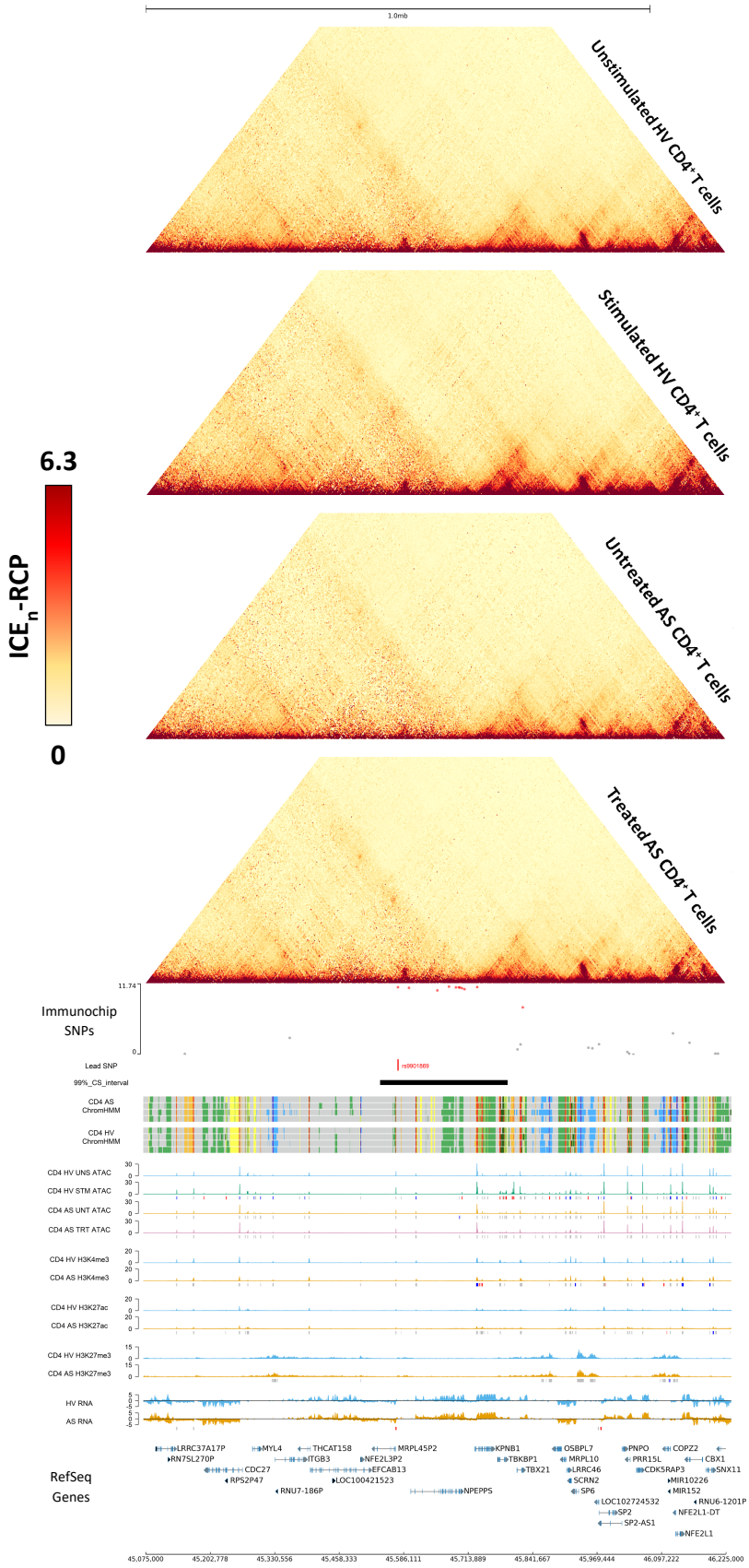

Fig S.38. Tiled-C interactions and the chromatin landscape of the *NPEPPS* locus in CD4$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr17:45,075,000-46,225,000.

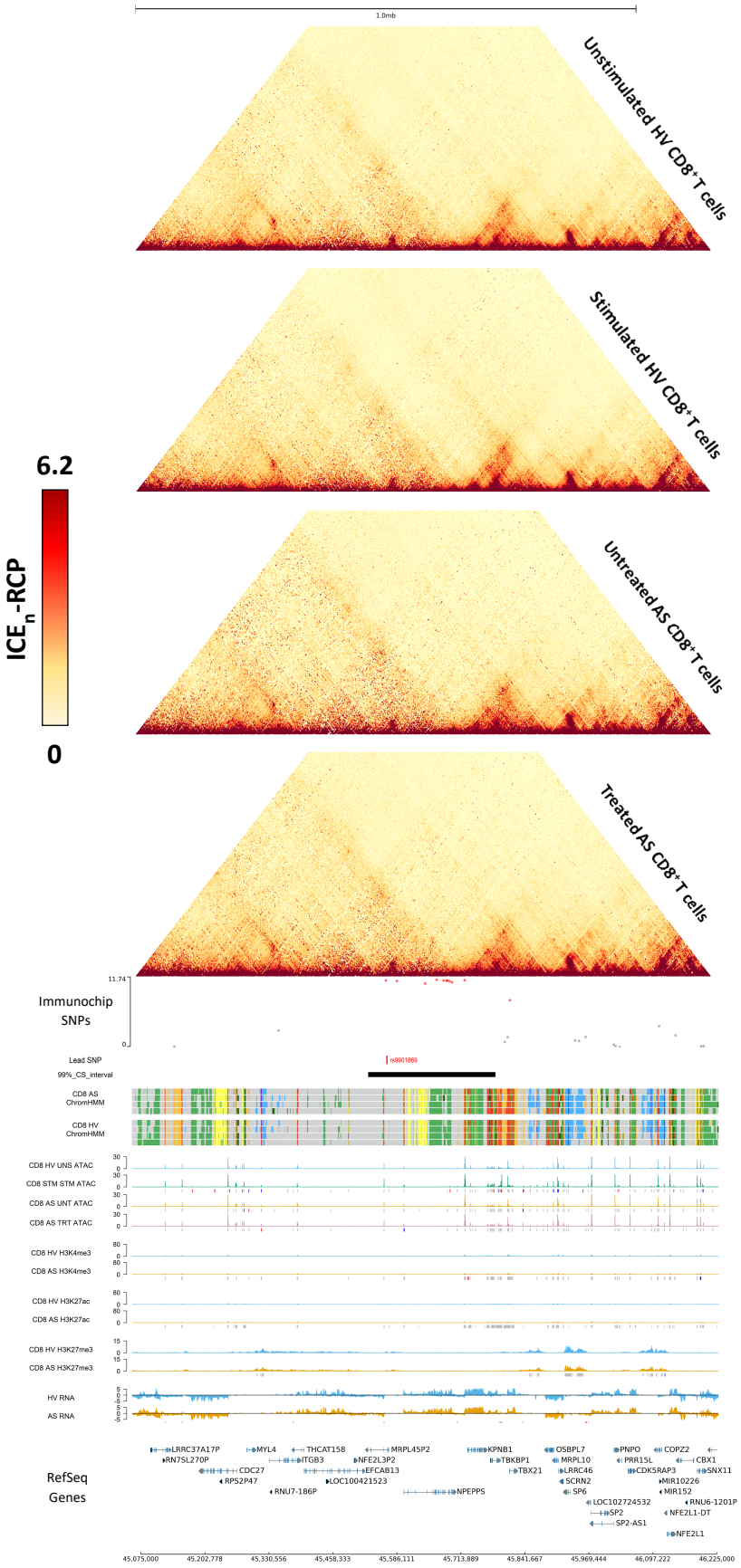

Fig S.39. Tiled-C interactions and the chromatin landscape of the *NPEPPS* locus in CD8$^+$ T cells. See description of page 308 for complete figure legend. Locus shown between chr17:45,075,000-46,225,000.

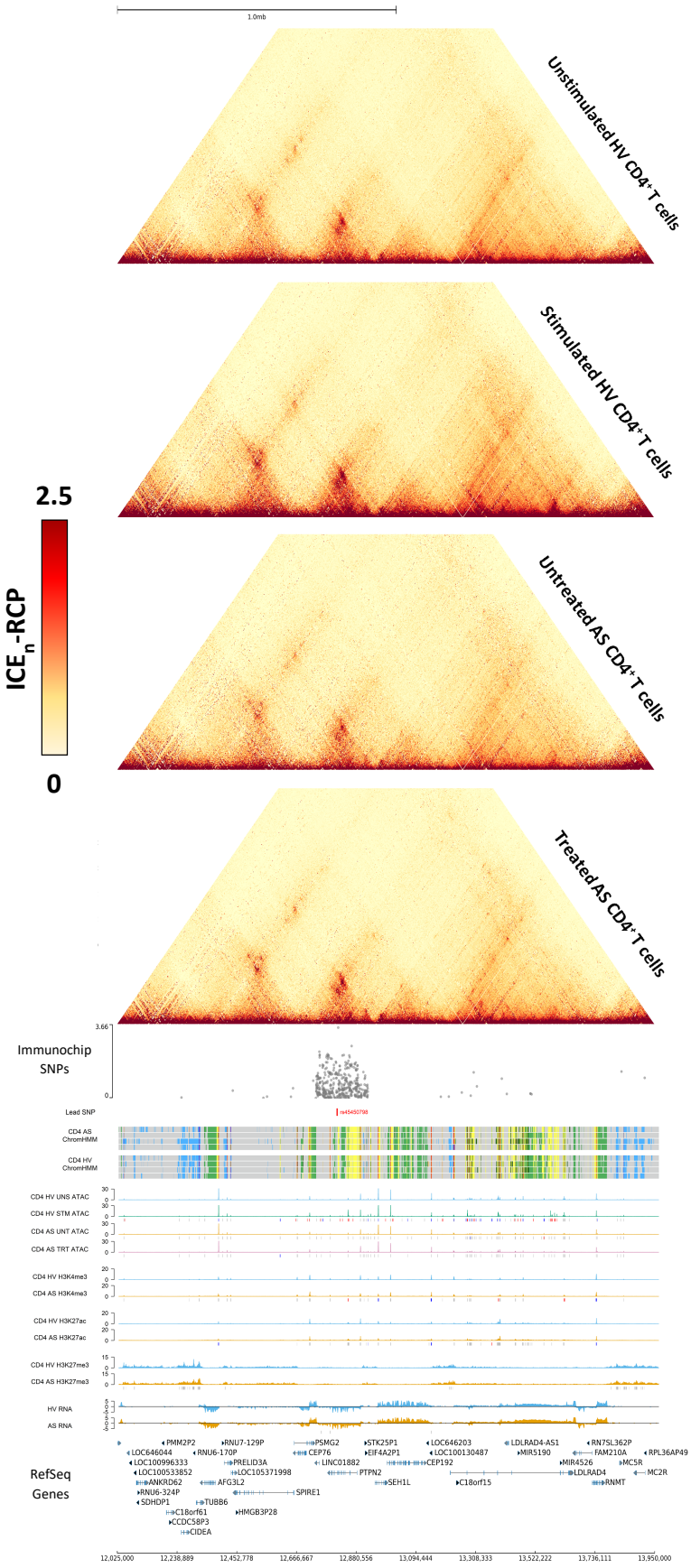

Fig S.40. Tiled-C interactions and the chromatin landscape of the *PTPN2* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr18:12,025,000-13,950,000.

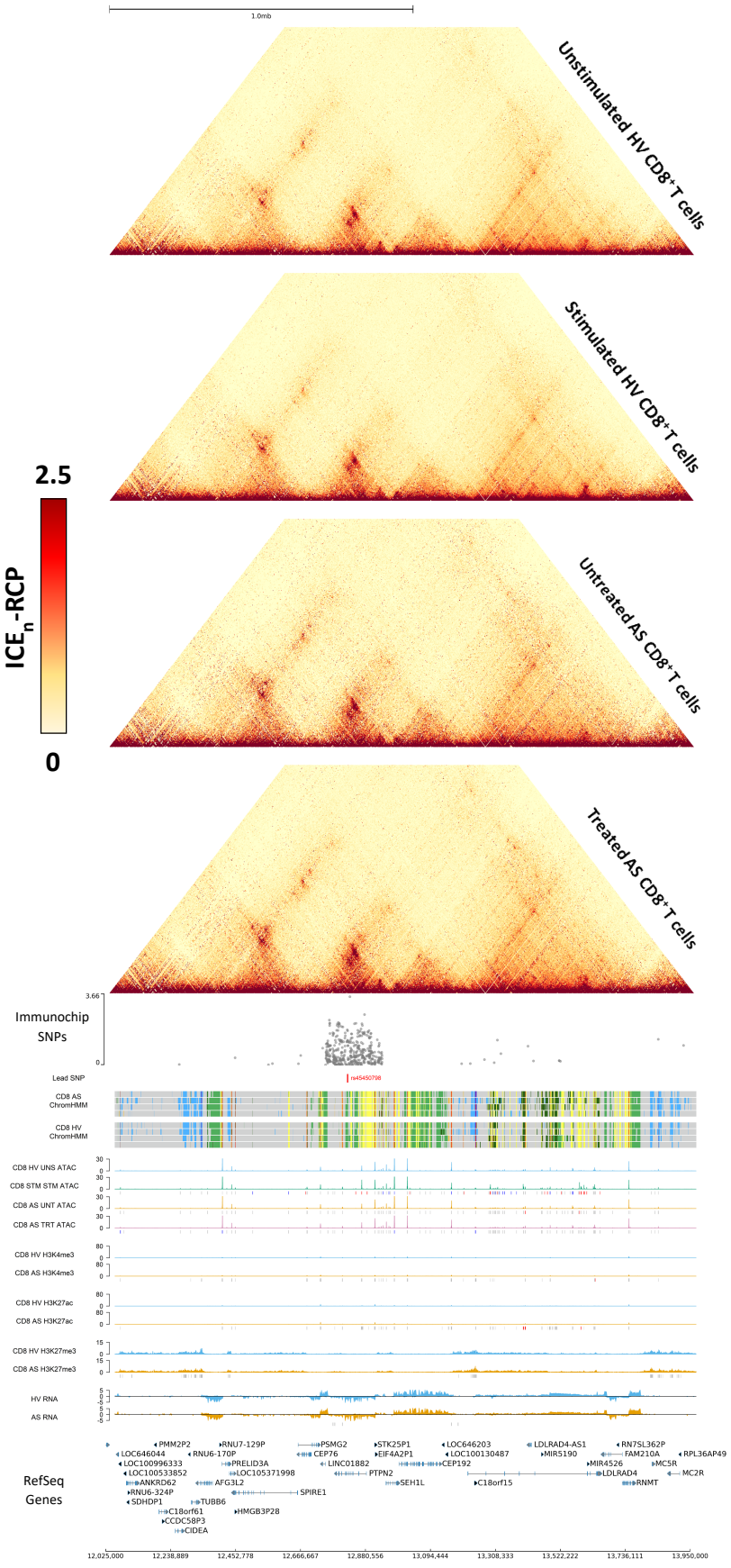

Fig S.41. Tiled-C interactions and the chromatin landscape of the *PTPN2* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr18:12,025,000-13,950,000.

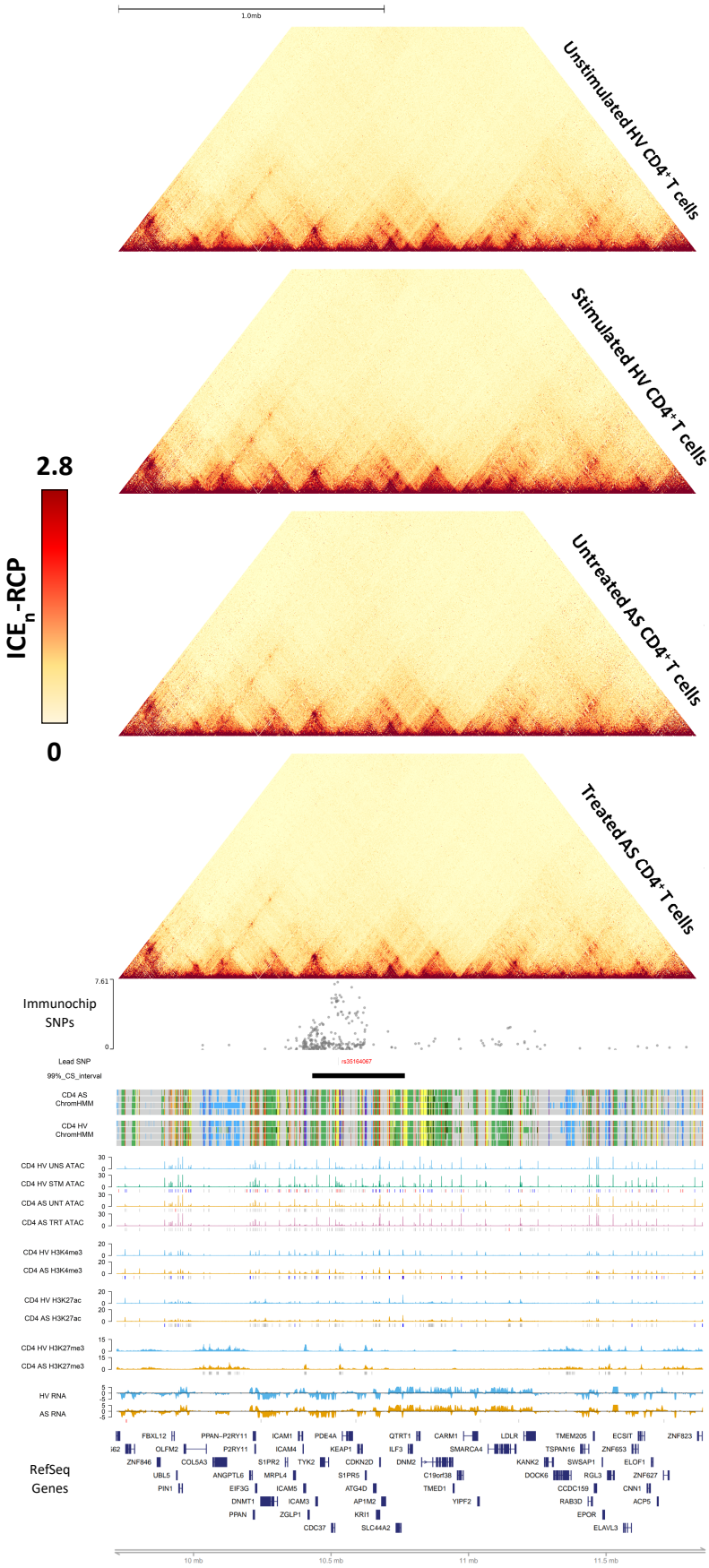

Fig S.42. Tiled-C interactions and the chromatin landscape of the *TYK2-ICAM1-ICAM3* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr19:9,700,000-11,875,000.

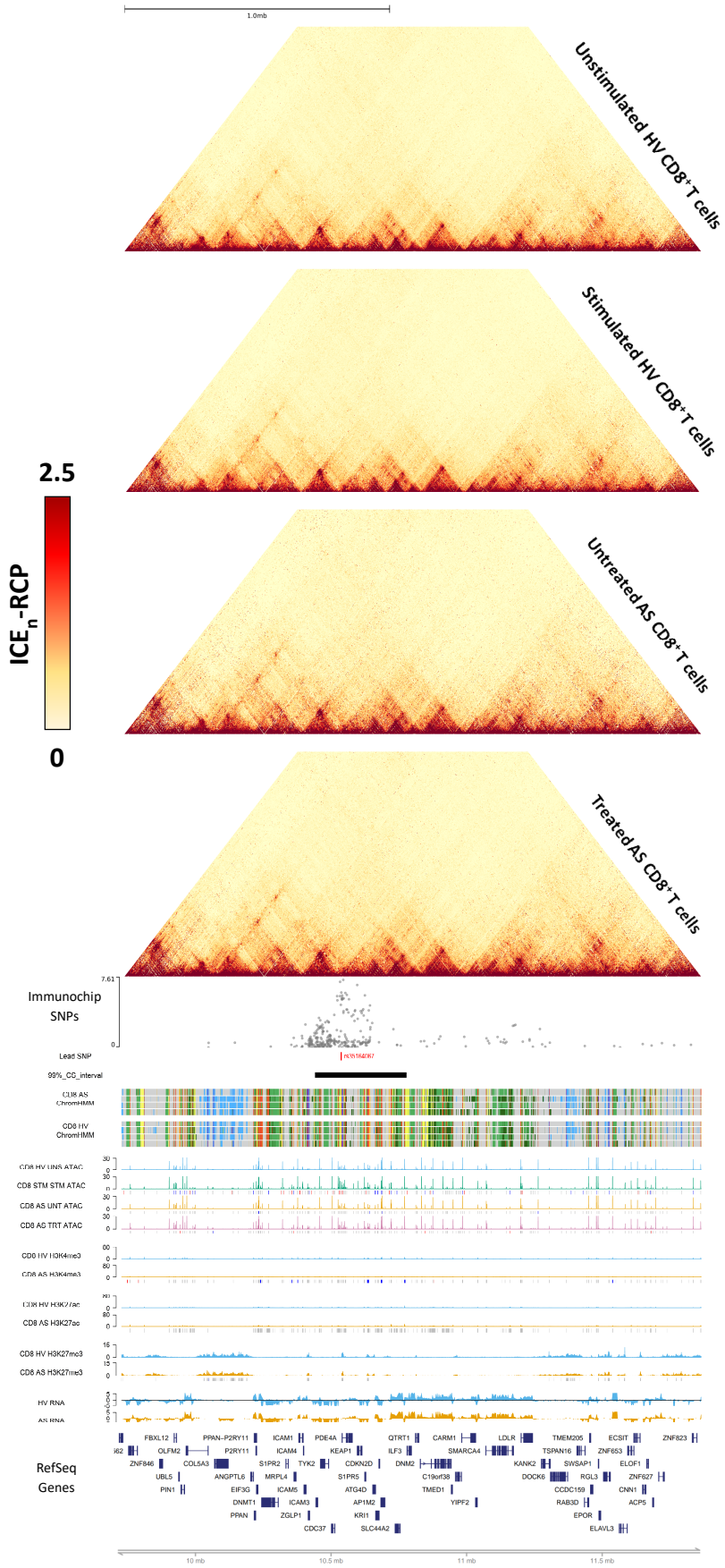

Fig S.43. Tiled-C interactions and the chromatin landscape of the *TYK2-ICAM1-ICAM3* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr19:9,700,000-11,875,000.

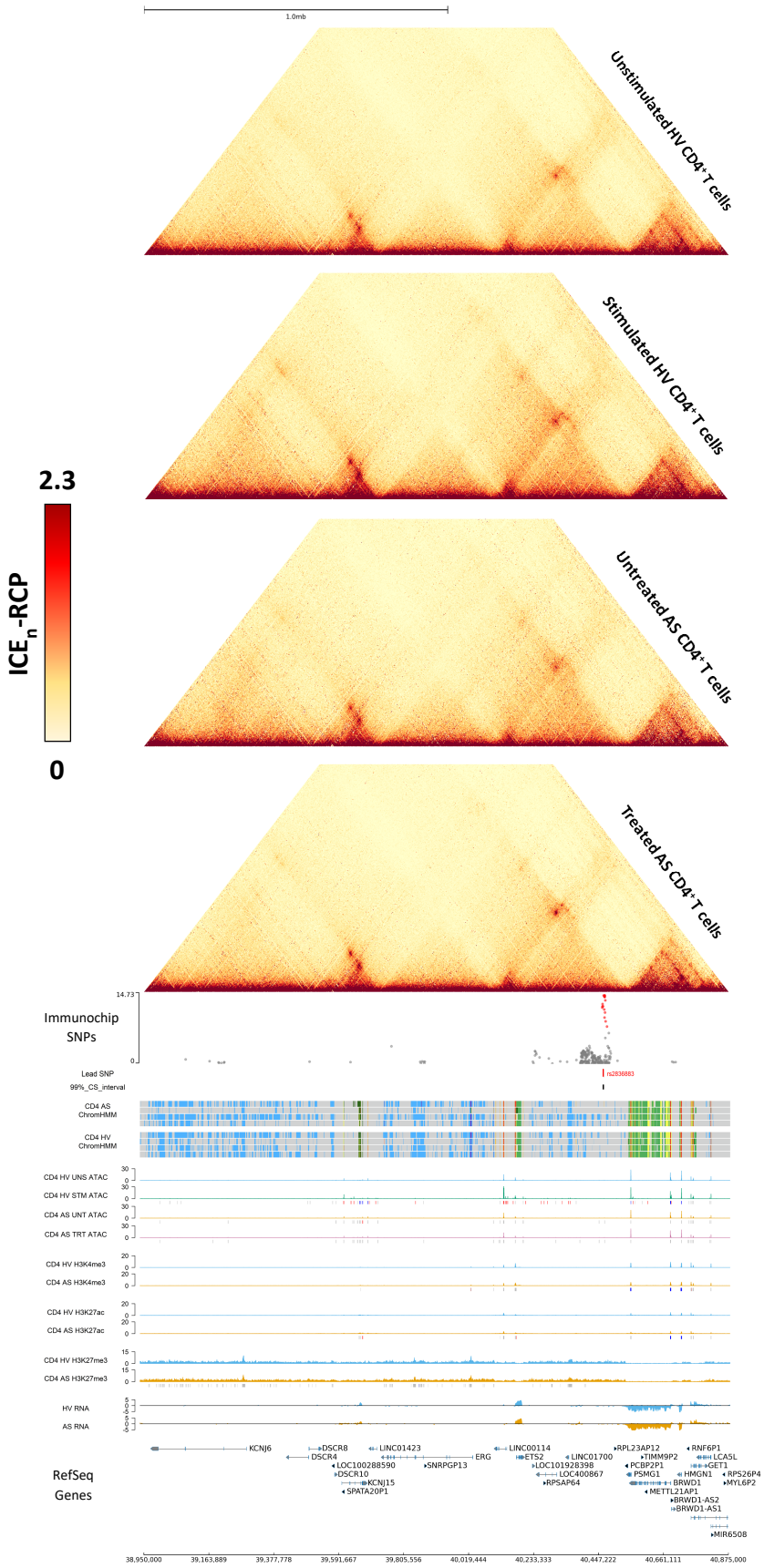

Fig S.44. Tiled-C interactions and the chromatin landscape of the *ETS2* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr21:28,950,000-40,875,000.

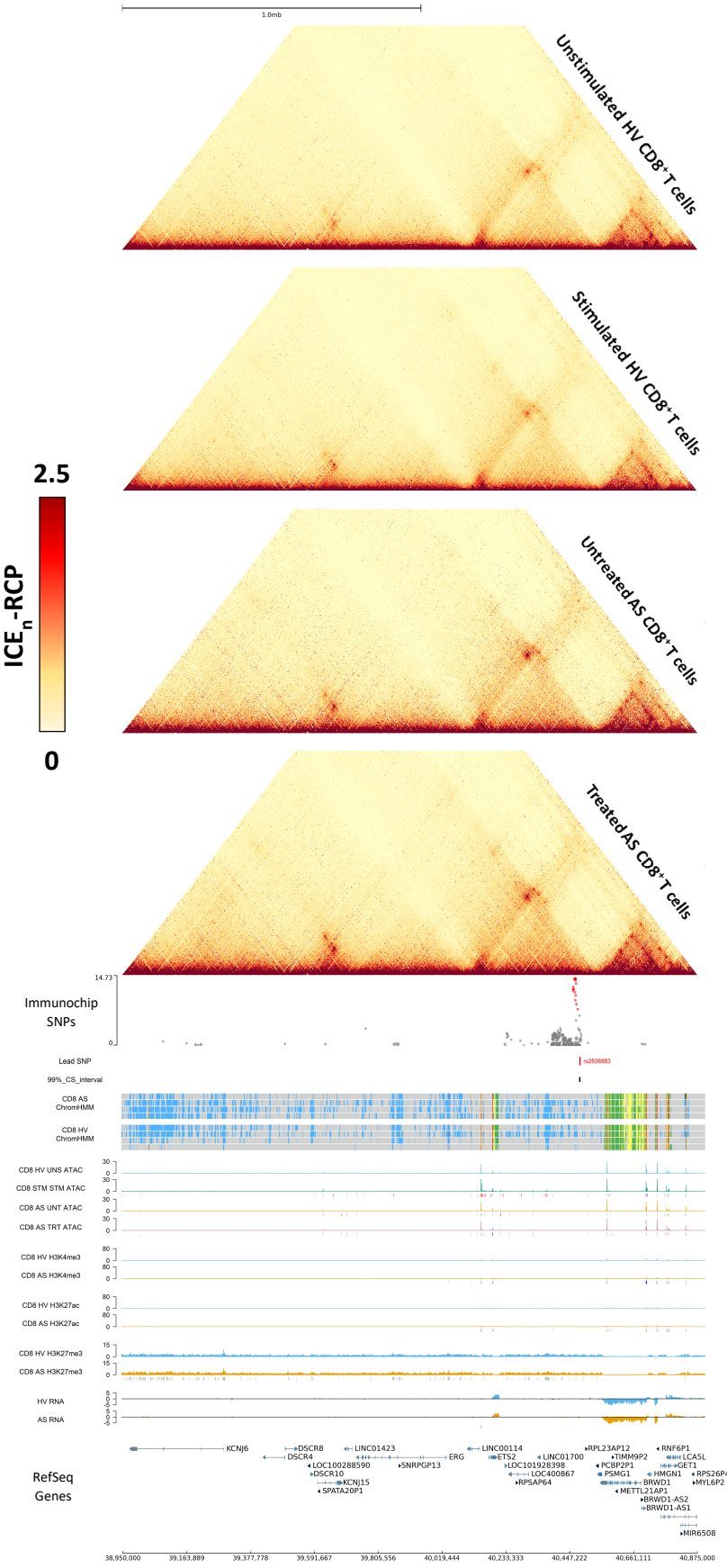

Fig S.45. Tiled-C interactions and the chromatin landscape of the *ETS2* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr21:28,950,000-40,875,000.

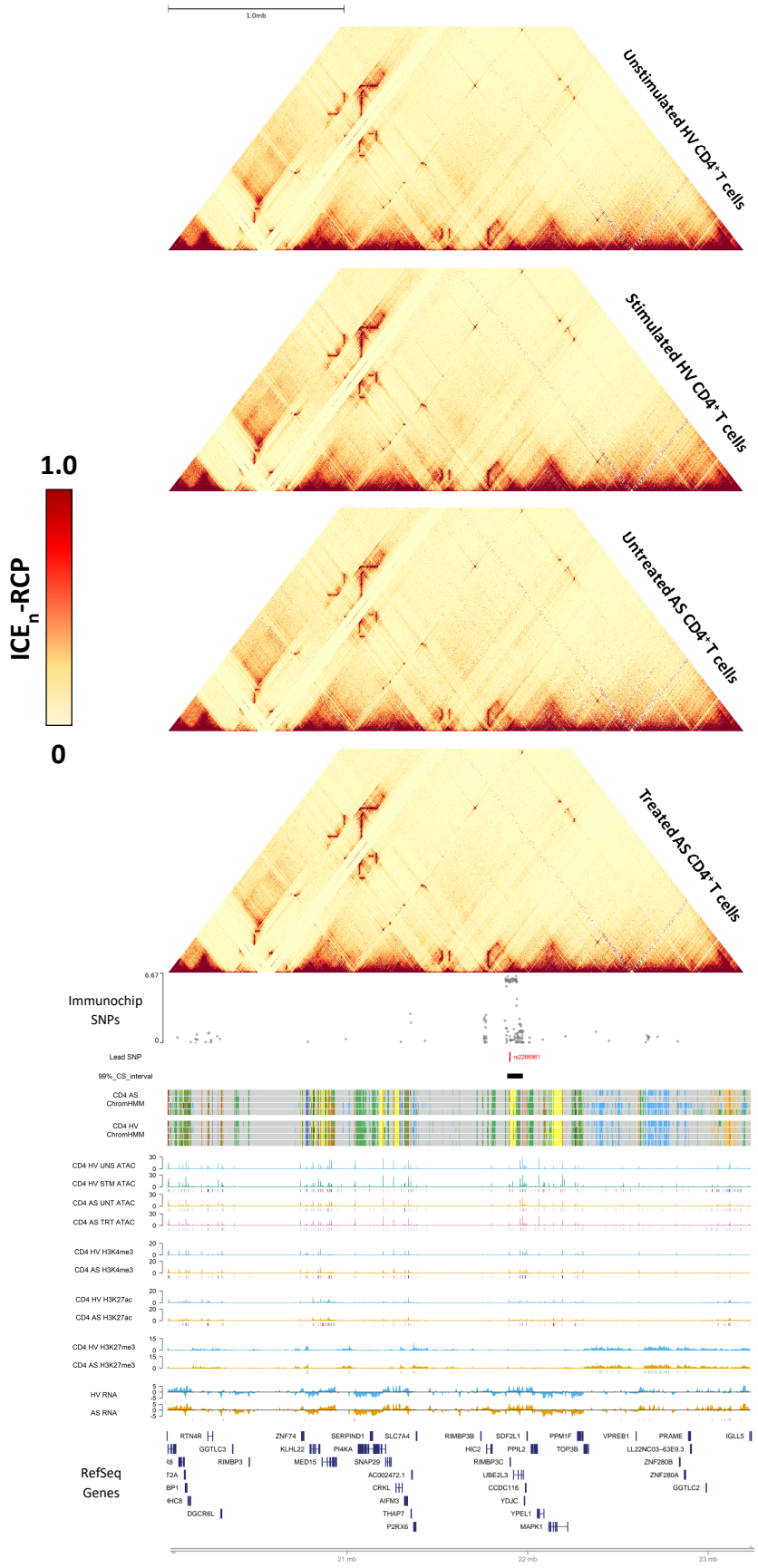

Fig S.46. Tiled-C interactions and the chromatin landscape of the *UBE2E3-MAPK* locus in CD4⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr22:20,001,000-23,275,000.

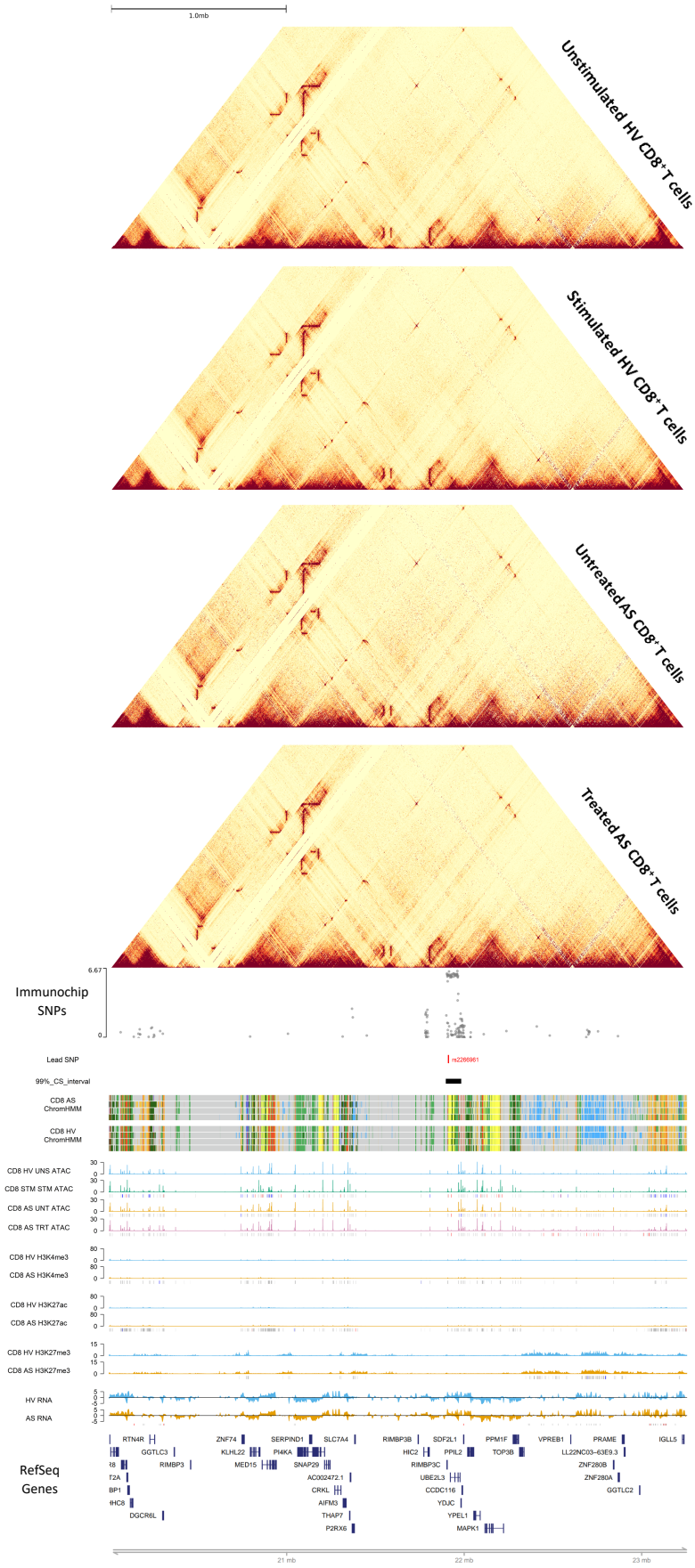

Fig S.47. Tiled-C interactions and the chromatin landscape of the *UBE2E3-MAPK* locus in CD8⁺ T cells. See description of page 308 for complete figure legend. Locus shown between chr22:20,001,000-23,275,000.

"Would it save you a lot of time if I just gave up and went mad now?"

- Douglas Adams, The Hitchhiker's Guide to the Galaxy (1979)

B

BIBLIOGRAPHY

A ----------

Abell N. S., DeGorter M. K., Gloudemans M. J., Greenwald E., Smith K. S., He Z., Montgomery S. B. (2022) "Multiple causal variants underlie genetic associations in humans". *Science*. **375**(6586): pp. 1247-1254. doi: 10.1126/science.abj5117.

Adamopoulos I. E., Tessmer M., Chao C. C., Adda S., Gorman D., Petro M., Chou C. C., Pierce R. H., Yao W., Lane N. E., Laface D., Bowman E. P. (2011) "IL-23 is critical for induction of arthritis, osteoclast formation, and maintenance of bone mass". *J. Immunol*. **187**(2): p. 951-9. doi: 10.4049/jimmunol.1003986.

Al-Jibury E., King J. W. D., Guo Y., Lenhard B., Fisher A. G., Merkenschlager M., Rueckert D. (2023) "A deep learning method for replicate-based analysis of chromosome conformation contacts using Siamese neural networks". *Nat. Commun*. **14**(1): p. 5007. doi: 10.1038/s41467-023-40547-9.

Al-Mossawi H., Yager N., Taylor C. A., Lau E., Danielli S., de Wit J., Gilchrist J., Nassiri I., Mahe E. A., Lee W., Rizvi L., Makino S., Cheeseman J., Neville M., Knight J. C., Bowness P., Fairfax B. P. (2019) "Context-specific regulation of surface and soluble IL7R expression by an autoimmune risk allele". *Nat. Commun*. **10**(1): pp. 4575. doi: 10.1038/s41467-019-12393-1.

Al-Mossawi, M.H., Chen, L., Fang, H., Ridley A., de Wit J., Yager N., Hammitzsh A., Pulyakhina I., Fairfax B. P., Simone D., Yi Y., Bandyopadhyay S., Doig K., Gundle R., Kendrick B., Powrie F., Knight J. C., Bowness P. (2017) "Unique transcriptome signatures and GM-CSF expression in lymphocytes from patients with spondyloarthritis". *Nat Commun* **8**(1): p. 1510. https://doi.org/10.1038/s41467-017-01771-2.

Akgol Oksuz B., Yang L., Abraham S., Venev S. V., Krietenstein N., Parsi K. M., Ozadam H., Oomen M. E., Nand A., Mao H., Genga R. M. J., Maehr R., Rando O. J., Mirny L. A., Gibcus J. H., Dekker J. (2021) "Systematic evaluation of chromosome conformation capture assays". *Nat. Methods*. 2021 Sep;18(9):1046-1055. doi: 10.1038/s41592-021-01248-7.

Alasoo K., Rodrigues J., Danesh J., Freitag D. F., Paul D. S., Gaffney D. J. (2019) "Genetic effects on promoter usage are highly context-specific and contribute to complex traits". *Elife*. **8**(1): e41673. doi: 10.7554/eLife.41673.

Alekseyenko A. A., Walsh E. M., Wang X., Grayson A. R., Hsi P. T., Kharchenko P. V., Kuroda M. I., French C. A. (2015) "The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains". *Genes Dev*. **29**(14): pp. 1507-1523. doi: 10.1101/gad.267583.115.

Aljogol D., Thompson I. R., Osborne C. S., Mifsud B. (2022) “Comparison of Capture Hi-C Analytical Pipelines”. *Front. Genet*. **13**(1): p. 786501. doi: 10.3389/fgene.2022.786501.

Alkan C., Coe B. P., Eichler E. E. “Genome structural variation discovery and genotyping”. *Nat. Rev. Genet*.**12**(5): pp. 363-376. doi: 10.1038/nrg2958.

Allen B. L., Taatjes D. J. (2015) “The Mediator complex: a central integrator of transcription”. *Nat. Rev. Mol. Cell Biol*. **16**(3): pp. 155-66. doi: 10.1038/nrm3951.

Allen R. L., Bowness P., McMichael A. (1999) “The role of HLA-B27 in spondyloarthritis”. *Immunogenetics*. **50**(3-4): pp. 220-227. doi: 10.1007/s002510050596.

Akhtari M., Zargar S. J., Vojdanian M., Jamshidi A., Mahmoudi M. (2021) “Monocyte-derived and M1 macrophages from ankylosing spondylitis patients released higher TNF-α and expressed more IL1B in response to BzATP than macrophages from healthy subjects”. *Sci. Rep*. **11**(1): p. 17842. doi: 10.1038/s41598-021-96262-2.

An Y., Yao J., Niu X. (2021) “The Signaling Pathway of PGE_2 and Its Regulatory Role in T Cell Differentiation”. *Mediators. Inflamm*. **2021**(1): p. 9087816. doi: 10.1155/2021/9087816.

An Z., Li J., Yu J., Wang X., Gao H., Zhang W., Wei Z., Zhang J., Zhang Y., Zhao J., Liang X. (2019) “Neutrophil extracellular traps induced by IL-8 aggravate atherosclerosis via activation NF-κB signaling in macrophages”. *Cell Cycle*. **18**(21): pp. 2928-2938. doi: 10.1080/15384101.2019.1662678.

Anders S., Pyl P. T., Huber W. (2015) “HTSeq—a Python framework to work with high-throughput sequencing data”. *Bioinformatics*. **31**(2): pp. 166–169. doi: 10.1093/bioinformatics/btu638.

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., Bagger F. O., Jørgensen M., Andersen P. R., Bertin N., Rackham O., Burroughs A. M., Baillie J. K., Ishizu Y., Shimizu Y., Furuhata E., Maeda S., Negishi Y., Mungall C. J., Meehan T. F., Lassmann T., Itoh M., Kawaji H., Kondo N., Kawai J., Lennartsson A., Daub C. O., Heutink P., Hume D. A., Jensen T. H., Suzuki H., Hayashizaki Y., Müller F., Forrest A. R. R., Carninci P., Rehli M., Sandelin A. (2014) “An atlas of active enhancers across human cell types and tissues”. *Nature*. **507**(7493): pp. 455-461. doi: 10.1038/nature12787.

Antonelli M.., Khan M. A., Magrey, M. N. (2015) “Differential Adverse Events Between TNF-α Inhibitors and IL-17 Axis Inhibitors for the Treatment of Spondyloarthritis”. *Curr. Treat. Options in Rheum.* **1**(1): pp. 239–254 (2015). doi: 10.1007/s40674-015-0022-7.

Antoniou A. N., Lenart I., Kriston-Vizi J., Iwawaki T., Turmaine M., McHugh K., Ali S., Blake N., Bowness P., Bajaj-Elliott M., Gould K., Nesbeth D., Powis S. J. (2019) “*Salmonella* exploits HLA-B27 and host unfolded protein responses to promote intracellular replication”. *Ann. Rheum. Dis*. **78**(1): pp. 74-82. doi: 10.1136/annrheumdis-2018-213532.

Antoniou A. N., Ford S., Taurog J. D., Butcher G. W., Powis S. J. (2004) “Formation of HLA-B27 homodimers and their relationship to assembly kinetics”. *J. Biol. Chem*. **279**(10): pp. 8895-8902. doi: 10.1074/jbc.M311757200.

Appel H., Maier R., Wu P., Scheer R., Hempfing A., Kayser R., Thiel A., Radbruch A., Loddenkemper C., Sieper J. (2011) “Analysis of IL-17(+) cells in facet joints of patients with

spondyloarthritis suggests that the innate immune pathway might be of greater relevance than the Th17-mediated adaptive immune response". *Arthritis Res. Ther*. **13**(3): e95. doi: 10.1186/ar3370.

Arnold P. R., Wells A. D., Li X. C. (2020)"Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate". *Front. Cell. Dev. Biol*. 7(1): p. 377. doi: 10.3389/fcell.2019.00377.

Arrigoni A., Ranzani V., Rossetti G., Panzeri I., Abrignani S., Bonnal R. J., Pagani M. (2016) "Analysis RNA-seq and Noncoding RNA". *Methods Mol. Biol*. 2016;1480:125-35. doi: 10.1007/978-1-4939-6380-5_11.

Asada R., Kanemoto S., Matsuhisa K., Hino K., Cui M., Cui X., Kaneko M., Imaizumi K. (2015) "IRE1α-XBP1 is a novel branch in the transcriptional regulation of Ucp1 in brown adipocytes". *Sci. Rep*. **5**(1): e16580. doi: 10.1038/srep16580.

Aschenbrenner D., Quaranta M., Banerjee S., Ilott N., Jansen J., Steere B., Chen Y. H., Ho S., Cox K., Arancibia-Cárcamo C. V., Coles M., Gaffney E., Travis S. P., Denson L., Kugathasan S., Schmitz J., Powrie F., Sansom S. N., Uhlig H. H. "Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance". *Gut*. **70**(6): pp. 1023-1036. doi: 10.1136/gutjnl-2020-321731.

Atkin-Smith G. K., Duan M., Zanker D. J., Loh L., Nguyen T. H. O., Koutsakos M., Nguyen T., Jiang X., Carrera J., Phan T. K., Liu C., Paone S., Oveissi S., Hodge A. L., Baxter A. A., Kedzierska K., Mackenzie J. M., Hulett M. D., Bilsel P., Chen W., Poon I. K. H. (2020) "Monocyte apoptotic bodies are vehicles for influenza A virus propagation". **Commun. Biol**. **3**(1): p. 223. doi: 10.1038/s42003-020-0955-8.

B ----------

Bagcivan G., Cinar F. I., Cinar M., Oflaz F., Uzun S., Pay S. (2015) "Living with pain in ankylosing spondylitis: a qualitative study". *Contemp. Nurse*. **51**(2-3): pp. 135-47. doi: 10.1080/10376178.2016.1157028.

Bakland G., Gran J. T., Becker-Merok A., Nordvåg B. Y., Nossent J. C. (2011) "Work disability in patients with ankylosing spondylitis in Norway". *J. Rheumatol*. **38**(3): pp. 479-84. doi: 10.3899/jrheum.100686.

Baraliakos X., Szumski A., Koenig A. S., Jones H. (2019) "The role of C-reactive protein as a predictor of treatment response in patients with ankylosing spondylitis". *Semin. Arthritis Rheum*. **48**(6): pp. 997-1004. doi: 10.1016/j.semarthrit.2018.10.019.

Baraliakos X., Deodhar, A. (2014) "Unanswered questions in the management of axial spondyloarthritis: an opinion piece". *Clin. Rheumatol*. **33**(1): pp. 1359–1365. doi: 10.1007/s10067-014-2740-x.

Barnett R., Ng S., Sengupta R. (2021) "Understanding flare in axial spondyloarthritis: novel insights from daily self-reported flare experience". *Rheumatol. Adv. Pract*. **5**(3): e. rkab082. doi: 10.1093/rap/rkab082.

Barrington C. L., Galindo G., Koch A. L., Horton E. R., Morrison E. J., Tisa S., Stasevich T. J., Rissland O. S. (2023) "Synonymous codon usage regulates translation initiation". *Cell Rep.* **42**(12): p. 113413. doi: 10.1016/j.celrep.2023.113413.

Bartoszewski R. A., Jablonsky M., Bartoszewska S., Stevenson L., Dai Q., Kappes J., Collawn J. F., Bebok Z. A. (2010) "A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein". *J. Biol. Chem*. **285**(37): pp. 28741-28748. doi: 10.1074/jbc.M110.154575.

Baumberger H., Khan M. "Gradual progressive change to equal prevalence of ankylosing spondylitis among males and females in Switzerland: data from the swiss ankylosing spondylitis society (SVMB)". *Ann. Rheum. Dis*. **76**(1): p. 929. doi: 10.1186/s13059-019-1706-3.

Ben Zouari Y., Molitor A. M., Sikorska N., Pancaldi V., Sexton T. (2019) "ChiCMaxima: a robust and simple pipeline for detection and visualization of chromatin looping in Capture Hi-C". *Genome Biol*. **20**(1): p. 102. doi: 10.1186/s13059-019-1706-3.

Benjamin R., Parham P. (1990) "Guilt by association: HLA-B27 and ankylosing spondylitis". *Immunol. Today*. **11**(4): pp. 137-42. doi: 10.1016/0167-5699(90)90051-a.

Benton M. L., Talipineni S. C., Kostka D., Capra J. A. (2019) "Genome-wide enhancer annotations differ significantly in genomic distribution, evolution, and function". *BMC Genomics*. **20**(1): p. 511. doi: 10.1186/s12864-019-5779-x.

Berland M., Meslier V., Berreira Ibraim S., Le Chatelier E., Pons N., Maziers N., Thirion F., Gauthier F., Plaza Oñate F., Furet J. P., Leboime A., Said-Nahal R., Levenez F., Galleron N., Quinquis B., Langella P., Ehrlich S. D., Breban M. (2023) "Both Disease Activity and HLA-B27 Status Are Associated With Gut Microbiome Dysbiosis in Spondyloarthritis Patients". *Arthritis Rheumatol.* **75**(1): pp. 41-52. doi: 10.1002/art.42289.

Blair H. A., Dhillon S. (2016) "Secukinumab: A Review in Ankylosing Spondylitis". *Drugs*. **76**(10): pp. 1023-1030. doi: 10.1007/s40265-016-0598-8.

Blander J. M., Medzhitov R. (2004) "Regulation of phagosome maturation by signals from toll-like receptors". *Science*. **304**(5673): pp. 1014-1018. doi: 10.1126/science.1096158.

Boel A., López-Medina C., van der Heijde D. M. F. M., van Gaalen F. A. (2022) "Age at onset in axial spondyloarthritis around the world: data from the Assessment in SpondyloArthritis international Society Peripheral Involvement in Spondyloarthritis study". *Rheumatology*. **61**(4): pp. 1468-1475. doi: 10.1093/rheumatology/keab544.

Bolt C. C., Lopez-Delisle L., Hintermann A., Mascrez B., Rauseo A., Andrey G., Duboule D. (2022) "Context-dependent enhancer function revealed by targeted inter-TAD relocation". *Nat. Commun*. **13**(1): p. 3488. doi: 10.1038/s41467-022-31241-3.

Bowness P. (2015) "HLA-B27". *Annu. Rev. Immunol*. 2015;33:29-48. doi: 10.1146/annurev-immunol-032414-112110.

Bowness P., Ridley A., Shaw J., Chan A. T., Wong-Baeza I., Fleming M., Cummings F., McMichael A., Kollnberger S. (2011) "Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis". *J. Immunol*. **186**(4): pp. 2672-80. doi: 10.4049/jimmunol.1002653.

Braun J., Baraliakos X., Deodhar A., Poddubnyy D., Emery P., Delicha E. M., Talloczy Z., Porter B. (2019) "Secukinumab shows sustained efficacy and low structural progression in ankylosing spondylitis: 4-year results from the MEASURE 1 study". *Rheumatology*. **58**(5): pp. 859-868. doi: 10.1093/rheumatology/key375.

Bray T. J. P., Lopes A., Fisher C., Ciurtin C., Sen D., Hall-Craggs M. A. (2019) "Sacroiliac Joint Ankylosis in Young Spondyloarthritis Patients Receiving Biologic Therapy: Observation of Serial Magnetic Resonance Imaging Scans". *Arthritis Rheumatol*. **71**(4): pp. 594-598. doi: 10.1002/art.40750.

Brescia C., Audia S., Pugliano A., Scaglione F., Iuliano R., Trapasso F., Perrotti N., Chiarella E., Amato R. (2024) "Metabolic drives affecting Th17/Treg gene expression changes and differentiation: impact on immune-microenvironment regulation". *APMIS*. ePUB ahead of publication. doi: 10.1111/apm.13378..

Brewerton D. A., Hart F. D., Nicholls A., Caffrey M., James D. C., Sturrock R. D. (1973) "Ankylosing spondylitis and HL-A 27". *Lancet.* **1**(7809): pp. 904-907. doi: 10.1016/s0140-6736(73)91360-3.

Bridgewood C., Sharif K., Sherlock J., Watad A., McGonagle D. (2020) "Interleukin-23 pathway at the enthesis: The emerging story of enthesitis in spondyloarthropathy". *Immunol. Rev*. **294**(1): pp. 27-47. doi: 10.1111/imr.12840.

Brown M. A., Wordsworth B. P. (2018) "Genetics in ankylosing spondylitis - Current state of the art and translation into clinical outcomes". *Best Pract. Res. Clin. Rheumatol*. **31**(6): pp. 763-776. doi: 10.1016/j.berh.2018.09.005.

Brown M. A., Kenna T., Wordsworth B. P. (2016) "Genetics of ankylosing spondylitis--insights into pathogenesis". *Nat. Rev. Rheumatol*. **12**(2): pp. 81-91. doi: 10.1038/nrrheum.2015.133.

Brown M. A., Kennedy L. G., MacGregor A. J., Darke C., Duncan E., Shatford J. L., Taylor A., Calin A., Wordsworth P. (1997) "Susceptibility to ankylosing spondylitis in twins: the role of genes, HLA, and the environment". *Arthritis Rheum*. **40**(10): pp. 1823-1828. doi: 10.1002/art.1780401015.

Brown M. A., Pile K. D., Kennedy L. G., Calin A., Darke C., Bell J., Wordsworth B. P., Cornélis F. (1996)"HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom". *Ann. Rheum. Dis*. **55**(4): pp. 268-270. doi: 10.1136/ard.55.4.268.

Bruno P. M., Timms R. T., Abdelfattah N. S., Leng Y., Lelis F. J. N., Wesemann D. R., Yu X. G., Elledge S. J. (2023) "High-throughput, targeted MHC class I immunopeptidomics using a functional genetics screening platform". *Nat. Biotechnol*. **41**(7): pp. 980-992. doi: 10.1038/s41587-022-01566-x.

Buchmuller B. C., Kosel B., Summerer D. (2020) "Complete Profiling of Methyl-CpG-Binding Domains for Combinations of Cytosine Modifications at CpG Dinucleotides Reveals Differential Read-out in Normal and Rett-Associated States". *Sci. Rep*. **10**(1):4053. doi: 10.1038/s41598-020-61030-1.

Buenrostro J. D., Wu B., Chang H. Y., Greenleaf W. J. (2013) "ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide". *Curr. Protoc. Mol. Biol*. **10**(1): pp. 1213–1218. doi: 10.1002/0471142727.mb2129s109.

Bugaj B., Wielińska J., Bogunia-Kubik K., Świerkot J. (2022) "Searching for New Genetic Biomarkers of Axial Spondyloarthritis". *J. Clin. Med*. 11(10): p. 2912. doi: 10.3390/jcm11102912.

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., Flicek P., Burdett T., Hindorff L. A, Cunningham F., Parkinson H. (2019) "The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019". *Nucleic Acids Res*. **47**(1): pp. 1005-1012. doi: 10.1093/nar/gky1120.

Burr M. L., Sparbier C. E., Chan K. L., Chan Y. C., Kersbergen A., Lam E. Y. N., Azidis-Yates E., Vassiliadis D., Bell C. C., Gilan O., Jackson S., Tan L., Wong S. Q., Hollizeck S., Michalak E. M., Siddle H. V., McCabe M. T., Prinjha R. K., Guerra G. R., Solomon B. J., Sandhu S., Dawson S. J., Beavis P. A., Tothill R. W., Cullinane C., Lehner P. J., Sutherland K. D., Dawson M. A. (2019) "An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer". *Cancer Cell*. **36**(4): pp. 385-401.e8. doi: 10.1016/j.ccell.2019.08.008.

Burton P. R., *et al.* (2007) "Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants". *Nat. Genet*. **39**(11): pp. 1329-1337. doi: 10.1038/ng.2007.17.

Butler J. E., Kadonaga J. T. (2002) "The RNA polymerase II core promoter: a key component in the regulation of gene expression". *Genes Dev*. **16**(20): pp. 2583-2592. doi: 10.1101/gad.1026202.

C ----------

Cai Y., Zhang Y., Loh Y. P., Tng J. Q., Lim M. C., Cao Z., Raju A., Lieberman Aiden E., Li S., Manikandan L., Tergaonkar V., Tucker-Kellogg G., Fullwood M. J. (2021) "H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions". *Nat. Commun*. **12**(1): p. 719. doi: 10.1038/s41467-021-20940-y.

Cairns J., Freire-Pritchett P., Wingett S. W., Várnai C., Dimond A., Plagnol V., Zerbino D., Schoenfelder S., Javierre B. M., Osborne C., Fraser P., Spivakov M. (2016) "CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data". *Genome Biol*. **17**(1): p. 127. doi: 10.1186/s13059-016-0992-2.

Cao Z., Guo J., Li Q., Li Y., Wu J. (2022) "Optimal Biologic Drugs for the Treatment of Ankylosing Spondylitis: Results from a Network Meta-Analysis and Network Metaregression". *Biomed. Res. Int*. **6**(1): e:8316106. doi: 10.1155/2022/8316106.

Caparbo V. F., Saad C. G. S., Moraes J. C., de Brum-Fernandes A. J., Pereira R. M. R. (2018) "Monocytes from male patients with ankylosing spondylitis display decreased osteoclastogenesis and decreased RANKL/OPG ratio". *Osteoporos. Int*. **29**(11): pp. 2565-2573. doi: 10.1007/s00198-018-4629-z.

Cardin G. B., Bernard M., Rodier F., Christopoulos A. (2021) "DCBLD1 is associated with the integrin signaling pathway and has prognostic value in non-small cell lung and invasive breast carcinoma". *Sci. Rep*. **11**(1): p. 12753. doi: 10.1038/s41598-021-92090-6.

Cardoneanu A., Cozma S., Rezus C., Petrariu F., Burlui A. M., Rezus E. (2021) "Characteristics of the intestinal microbiome in ankylosing spondylitis". *Exp. Ther. Med*. **22**(1): p. 676. doi: 10.3892/etm.2021.10108.

Carini C., Hunter E.; Scottish Early Rheumatoid Arthritis Inception cohort Investigators; Ramadass A. S., Green J., Akoulitchev A., McInnes I. B., Goodyear C. S. (2018) "Chromosome

conformation signatures define predictive markers of inadequate response to methotrexate in early rheumatoid arthritis". *J. Transl. Med*. **16**(1): p. 18. doi: 10.1186/s12967-018-1387-9.

Carpenter M. D., Fischer D. K., Zhang S., Bond A. M., Czarnecki K. S., Woolf M. T., Song H., Heller E. A. (2022) "Cell-type specific profiling of histone post-translational modifications in the adult mouse striatum". *Nat. Commun*. **13**(1): p. 7720. doi: 10.1038/s41467-022-35384-1.

Casella A. M., Colantuoni C., Ament S. A. (2022) "Identifying enhancer properties associated with genetic risk for complex traits using regulome-wide association studies". *PLoS Comput. Biol*. **18**(9): e1010430. doi: 10.1371/journal.pcbi.1010430.

Castellanos-Rubio A., Fernandez-Jimenez N., Kratchmarov R., Luo X., Bhagat G., Green P. H., Schneider R., Kiledjian M., Bilbao J. R., Ghosh S. (2016) "A long noncoding RNA associated with susceptibility to celiac disease". *Science*. **352**(6281): pp. 91-95. doi: 10.1126/science.aad0467.

Cauli A., Shaw J., Giles J., Hatano H., Rysnik O., Payeli S., McHugh K., Dessole G., Porru G., Desogus E., Fiedler S., Hölper S., Carette A., Blanco-Gelaz M. A., Vacca A., Piga M., Ibba V., Garau P., La Nasa G., López-Larrea C., Mathieu A., Renner C., Bowness P., Kollnberger S. (2013) "The arthritis-associated HLA-B*27:05 allele forms more cell surface B27 dimer and free heavy chain ligands for KIR3DL2 than HLA-B*27:09". *Rheumatology (Oxford)*. **52**(11): pp. 1952-62. doi: 10.1093/rheumatology/ket219.

Ciccia F., Rizzo A., Triolo G. (2016) "Subclinical gut inflammation in ankylosing spondylitis". *Curr. Opin Rheumatol*. **28**(1): pp. 89-96. doi: 10.1097/BOR.0000000000000239. PMID: 26599385.

Chan J., Gladman D. (2018) "Oral treatment options for AS and PsA: DMARDs and small-molecule inhibitors". *Best Pract. Res. Clin. Rheumatol*. **32**(3): pp. 415-426. doi: 10.1016/j.berh.2018.08.003.

Chang S. C., Momburg F., Bhutani N., Goldberg A. L. (2005) "The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism". *Proc. Natl. Acad. Sci. U. S. A.* **102**(47): pp. 17107-17112. doi: 10.1073/pnas.0500721102.

Chang S., Wang J., Zhang K., Wang J. (2015) "Pathway-based analysis for genome-wide association study data of bipolar disorder provides new insights for genetic study". *Protein Cell*. **6**(12): pp. 912-5. doi: 10.1007/s13238-015-0201-1.

Chen C. H., Zheng R., Tokheim C., Dong X., Fan J., Wan C., Tang Q., Brown M., Liu J. S., Meyer C. A., Liu X. S. (2020) "Determinants of transcription factor regulatory range". *Nat. Commun*. **11**(1): p. 2472. doi: 10.1038/s41467-020-16106-x.

Chen G. Y., Nuñez G. (2010) "Sterile inflammation: sensing and reacting to damage". *Nat. Rev Immunol*. **10**(12): pp. 826-837. doi: 10.1038/nri2873.

Chen W. S., Chang Y. S., Lin K. C., Lai C. C., Wang S. H., Hsiao K. H., Lee H. T., Chen M. H., Tsai C. Y., Chou C. T. (2012) "Association of serum interleukin-17 and interleukin-23 levels with disease activity in Chinese patients with ankylosing spondylitis". J. Chin. Med. Assoc. **75**(7): pp. 303-308. doi: 10.1016/j.jcma.2012.05.006.

Chen L., Fischer R., Peng Y., Reeves E., McHugh K., Ternette N., Hanke T., Dong T., Elliott T., Shastri N., Kollnberger S., James E., Kessler B., Bowness P. (2014) "Critical role of endoplasmic

reticulum aminopeptidase 1 in determining the length and sequence of peptides bound and presented by HLA-B27". *Arthritis Rheumatol*. **66**(2): pp. 284-94. doi: 10.1002/art.38249.

Chen Z., Shojaee S., Buchner M., Geng H., Lee J. W., Klemm L., Titz B., Graeber T. G., Park E., Tan Y. X., Satterthwaite A., Paietta E., Hunger S. P., Willman C. L., Melnick A., Loh M. L., Jung J. U., Coligan J. E., Bolland S., Mak T. W., Limnander A., Jumaa H., Reth M., Weiss A., Lowell C. A., Müschen M. (2015) "Signalling thresholds and negative B-cell selection in acute lymphoblastic leukaemia". *Nature*. **521**(7552): pp. 357-361. doi: 10.1038/nature14231.

Chen L., Deshpande M., Grisotto M., Smaldini P., Garcia R., He Z., Gulko P. S., Lira S. A., Furtado G. C. (2020) "Skin expression of IL-23 drives the development of psoriasis and psoriatic arthritis in mice". *Sci. Rep*. **10**(1): p. 8259. doi: 10.1038/s41598-020-65269-6.

Chen H. H. (2022) "The BASDAI Cut-Off for Disease Activity Corresponding to the ASDAS Scores in a Taiwanese Cohort of Ankylosing Spondylitis". *Front. Med.* **16**(1); e9:856654. doi: 10.3389/fmed.2022.856654.

Chen S., Saeed A. F. U. H., Liu Q., Jiang Q., Xu H., Xiao G. G., Rao L., Duo Y. (2023) "Macrophages in immunoregulation and therapeutics". *Signal Transduct. Target Ther*. **8**(1): p. 207. doi: 10.1038/s41392-023-01452-1.

Chen Y., Liu W., Xu X., Zhen H., Pang B., Zhao Z., Zhao Y., Liu H. (2024) "The Role of H3K27me3-Mediated Th17 Differentiation in Ankylosing Spondylitis. Inflammation". ePUB ahead of print. doi: 10.1007/s10753-024-02002-9.

Chen Z. Y., Mortha A. (2024) "INPP4B ensures that ILC1s and NK cells set up a productive home office". *J. Exp. Med*. **221**(3): e20232375. doi: 10.1084/jem.20232375.

Chen Z., Snetkova V., Bower G., Jacinto S., Clock B., Dizehchi A., Barozzi I., Mannion B. J., Alcaina-Caro A., Lopez-Rios J., Dickel D. E., Visel A., Pennacchio L. A., Kvon E. Z. (2024) "Increased enhancer-promoter interactions during developmental enhancer activation in mammals". *Nat. Genet*. **56**(4): pp. 675-685. doi: 10.1038/s41588-024-01681-2.

Chisălău B. A., Crînguș L. I., Vreju F. A., Pârvănescu C. D., Firulescu S. C., Dinescu Ș. C., Ciobanu D. A., Tica A. A., Sandu R. E., Siloși I., Boldeanu M. V., Poenariu I. S., Ungureanu A. M., Boldeanu L., Bărbulescu A. L. (2020) "New insights into IL-17/IL-23 signaling in ankylosing spondylitis". *Exp. Ther. Med*. **20**(4): pp. 3493-3497. doi: 10.3892/etm.2020.8981.

Claussnitzer M., Cho J. H., Collins R., Cox N. J., Dermitzakis E. T., Hurles M. E., Kathiresan S., Kenny E. E., Lindgren C. M., MacArthur D. G., North K. N., Plon S. E., Rehm H. L., Risch N., Rotimi C. N., Shendure J., Soranzo N., McCarthy M. I. (2020) "A brief history of human disease genetics". *Nature*. **577**(7789): pp. 179-189. doi: 10.1038/s41586-019-1879-7.

Cohen C. J., Davidson C., Selmi C., Bowness P., Knight J. C., Wordsworth B. P., Vecellio M. (2021) "Disruption of c-MYC Binding and Chromosomal Looping Involving Genetic Variants Associated With Ankylosing Spondylitis Upstream of the *RUNX3* Promoter". *Front. Genet*. 12(1): p. 741867. doi: 10.3389/fgene.2021.741867.

Conigliaro P., Scrivo R., Valesini G., Perricone R. (2011) "Emerging role for NK cells in the pathogenesis of inflammatory arthropathies". *Autoimmun. Rev*. **10**(10): pp. 577-81. doi: 10.1016/j.autrev.2011.04.017.

Costantino F., Leboime A., Bessalah L., Breban M. (2022) “Correspondence between patient-reported flare and disease activity score variation in axial spondyloarthritis: A 12-months web-based study”. *Joint Bone Spine*. **89**(5): p. 105422. doi: 10.1016/j.jbspin.2022.105422.

Cooper D. N., Youssoufian H. (1988) “The CpG dinucleotide and human genetic disease”. *Hum. Genet*. **78**(2): pp. 151-155. doi: 10.1007/BF00278187.

Cooksey R., Husain M. J., Brophy S., Davies H., Rahman M. A., Atkinson M. D., Phillips C. J., Siebert S. (2015) “The Cost of Ankylosing Spondylitis in the UK Using Linked Routine and Patient-Reported Survey Data”. *PLoS One*. **10**(7): e0126105. doi: 10.1371/journal.pone.0126105. eCollection 2015.

Corradin O., Saiakhova A., Akhtar-Zaidi B., Myeroff L., Willis J., Cowper-Sal lari 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 Res*. **24**(1): pp. 1-13. doi: 10.1101/gr.164079.113.

Cortes A., *et al.* (2013) “Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci”. *Nat. Genet*. **45**(7): pp. 730-738. doi: 10.1038/ng.2667.

Creemers M. C., Franssen M.J., van't Hof M. A., Gribnau F. W., van de Putte L. B., van Riel P. L. (2005) “Assessment of outcome in ankylosing spondylitis: an extended radiographic scoring system”. *Ann. Rheum. Dis.* **64**(1): pp. 127-129. doi: 10.1136/ard.2004.020503.

Cribbs A. P., Terlecki-Zaniewicz S., Philpott M., Baardman J., Ahern D., Lindow M., Obad S., Oerum H., Sampey B., Mander P. K., Penn H., Wordsworth P., Bowness P., de Winther M., Prinjha R. K., Feldmann M., Oppermann U. (2020) “Histone H3K27me3 demethylases regulate human Th17 cell development and effector functions by impacting on metabolism”. *Proc. Natl. Acad. Sci. U. S. A*. **117**(11): pp. 6056-6066. doi: 10.1073/pnas.1919893117.

Cronstein B. N., Sitkovsky M. (2017) “Adenosine and adenosine receptors in the pathogenesis and treatment of rheumatic diseases”. *Nat. Rev. Rheumatol*. **13**(1): p. 41-51. doi: 10.1038/nrrheum.2016.178.

Crowl J. T., Heeg M., Ferry A., Milner J. J., Omilusik K. D., Toma C., He Z., Chang J. T., Goldrath A. W. (2022) “Tissue-resident memory $CD8^+$ T cells possess unique transcriptional, epigenetic and functional adaptations to different tissue environments”. *Nat. Immunol*. **23**(7): pp. 1121-1131. doi: 10.1038/s41590-022-01229-8.

Czimmerer Z., Halasz L., Daniel B., Varga Z., Bene K., Domokos A., Hoeksema M., Shen Z., Berger W. K., Cseh T., Jambrovics K., Kolostyak Z., Fenyvesi F., Varadi J., Poliska S., Hajas G., Szatmari I., Glass C. K., Bacsi A., Nagy L. (2022) “The epigenetic state of IL-4-polarized macrophages enables inflammatory cistromic expansion and extended synergistic response to TLR ligands”. *Immunity*. **55**(11): pp. 2006-2026. doi: 10.1016/j.immuni.2022.10.004.

D ----------

Das A. C., Foroutan A., Qian B., Hosseini Naghavi N., Shabani K., Shooshtari P. (2022) “Single-Cell Chromatin Accessibility Data Combined with GWAS Improves Detection of Relevant Cell Types in 59 Complex Phenotypes”. *Int. J. Mol. Sci*. **23**(19): p. 11456. doi: 10.3390/ijms231911456.

Davidson C., Wordsworth B. P., Cohen C. J., Knight J. C., Vecellio M. (2023) “Chromosome conformation capture approaches to investigate 3D genome architecture in Ankylosing Spondylitis”. *Front. Genet*. **14**(1): p. 1129207. doi: 10.3389/fgene.2023.1129207.

Davidson I. F., Barth R., Zaczek M., van der Torre J., Tang W., Nagasaka K., Janissen R., Kerssemakers J., Wutz G., Dekker C., Peters J. M. (2023) “CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion”. *Nature*. **616**(7958): pp. 822-827. doi: 10.1038/s41586-023-05961-5.

Davies J. O., Oudelaar A. M., Higgs D. R., Hughes J. R. (2017) “How best to identify chromosomal interactions: a comparison of approaches”. *Nat. Methods*. **14**(2): pp. 125-134. doi: 10.1038/nmeth.4146.

de Lima K. A., Donate P. B., Talbot J., Davoli-Ferreira M., Peres R. S., Cunha T. M., Alves-Filho J. C., Cunha F. Q. (2018) “TGFβ1 signaling sustains aryl hydrocarbon receptor (AHR) expression and restrains the pathogenic potential of T_H17 cells by an AHR-independent mechanism”. *Cell Death Dis*. **9**(11): p. 1130. doi: 10.1038/s41419-018-1107-7.

de Vlam K., Mielants H., Cuvelier C., De Keyser F., Veys E. M., De Vos M. (2000) “Spondyloarthropathy is underestimated in inflammatory bowel disease: prevalence and HLA association”. *J. Rheumatol*. **27**(12): pp. 2860-2865.

Dean L. E., Jones G. T., MacDonald A. G., Downham C., Sturrock R. D., Macfarlane G. J. (2014) “Global prevalence of ankylosing spondylitis”. *Rheumatology*. **53**(4): pp. 650-657. doi: 10.1093/rheumatology/ket387.

Dean A., Larson D. R., Sartorelli V. (2021) “Enhancers, gene regulation, and genome organization”. *Genes Dev*. 35(7-8): pp. 427-432. doi: 10.1101/gad.348372.121.

Dekker J., Belmont A. S., Guttman M., Leshyk V. O., Lis J. T., Lomvardas S., Mirny L. A., O'Shea C. C., Park P. J., Ren B., Politz J. C. R., Shendure J., Zhong S.; 4D Nucleome Network. (2017) “The 4D nucleome project”. *Nature*. **549**(7671): pp. 219-226. doi: 10.1038/nature23884.

DeLay M. L., Turner M. J., Klenk E. I., Smith J. A., Sowders D. P., Colbert R. A. (2009) “HLA-B27 misfolding and the unfolded protein response augment interleukin-23 production and are associated with Th17 activation in transgenic rats”. *Arthritis Rheum*. **60**(9): pp. 2633-2643. doi: 10.1002/art.24763.

Demela P., Pirastu N., Soskic B. (2023) “Cross-disorder genetic analysis of immune diseases reveals distinct gene associations that converge on common pathways”. *Nat. Commun*. 2023 May 12;14(1):2743. doi: 10.1038/s41467-023-38389-6.

Deng S., Wang C., Wang Y., Xu Y., Li X., Johnson N. A., Mukherji A., Lo U. G., Xu L., Gonzalez J., Metang L. A., Ye J., Tirado C. R., Rodarte K., Zhou Y., Xie Z., Arana C., Annamalai V., Liu X., Vander Griend D. J., Strand D., Hsieh J. T., Li B., Raj G., Wang T., Mu P. (2022) “Ectopic JAK-STAT activation enables the transition to a stem-like and multilineage state conferring AR-targeted therapy resistance”. *Nat. Cancer*. **3**(9): pp. 1071-1087. doi: 10.1038/s43018-022-00431-9.

Deodhar A., Mittal M., Reilly P., Bao Y., Manthena S., Anderson J., Joshi A. (2016) “Ankylosing spondylitis diagnosis in US patients with back pain: identifying providers involved and factors associated with rheumatology referral delay”. *Clin. Rheumatol*. **35**(1): pp. 1769–1776. doi: 10.1007/s10067-016-3231-z.

Depoil D., Fleire S., Treanor B. L., Weber M., Harwood N. E., Marchbank K. L., Tybulewicz V. L., Batista F. D. (2008) "CD19 is essential for B cell activation by promoting B cell receptor-antigen microcluster formation in response to membrane-bound ligand". *Nat. Immunol*. **9**(1): pp. 63-72. doi: 10.1038/ni1547.

Derakhshan M. H., Pathak H., Cook D., Dickinson S., Siebert S., Gaffney K., NASS and BRITSpA investigators. (2018) "Services for spondyloarthritis: a survey of patients and rheumatologists". *Rheumatology (Oxford)*. ***57***(6): pp. 987-996. doi: 10.1093/rheumatology/kex518.

Deveci H., Turk A. C., Ozmen Z. C., Demir A. K., Say Coskun S. U. (2020) "Biological and genetic evaluation of IL-23/IL-17 pathway in ankylosing spondylitis patients". *Cent. Eur. J. Immunol*. **44**(4): pp. 433-439. doi: 10.5114/ceji.2019.92805.

Di Meglio P., Di Cesare A., Laggner U., Chu C. C., Napolitano L., Villanova F., Tosi I., Capon F., Trembath R. C., Peris K., Nestle F. O. (2011) "The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans". *PLoS One*. **6**(2): e17160. doi: 10.1371/journal.pone.0017160.

Ding M., Guan T. J., Wei C. Y., Chen B. H. (2018) "Identification of pathways significantly associated with spondyloarthropathy/ankylosing spondylitis using the sub-pathway method". *Mol. Med. Rep*.**18**(4): pp. 3825-3833. doi: 10.3892/mmr.2018.9395.

Ding Y., Chen J., Li R., Xue L. (2024) "Inflammatory bowel disease activity threatens ankylosing spondylitis: implications from Mendelian randomization combined with transcriptome analysis". *Front. Immunol*. **15**(1): p. 1289049. doi: 10.3389/fimmu.2024.1289049.

Dhillon S. S., Mastropaolo L. A., Murchie R., Griffiths C., Thöni C., Elkadri A., Xu W., Mack A., Walters T., Guo C., Mack D., Huynh H., Baksh S., Silverberg M. S., Brumell J. H., Snapper S. B., Muise A. M. (2014) "Higher activity of the inducible nitric oxide synthase contributes to very early onset inflammatory bowel disease". *Clin. Transl. Gastroenterol*. **5**(1): e46. doi: 10.1038/ctg.2013.17.

Dhindsa R. S., Wang Q., Vitsios D., Burren O. S., Hu F., DiCarlo J. E., Kruglyak L., MacArthur D. G., Hurles M. E., Petrovski S. (2022) "A minimal role for synonymous variation in human disease". *Am J Hum Genet*. **109**(12): pp. 2105-2109. doi: 10.1016/j.ajhg.2022.10.016.

Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T. R. (2013) "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics*. **29**(1): pp. 15-21. doi: 10.1093/bioinformatics/bts635.

Dong A., Preusch C. B., So W. K., Lin K., Luan S., Yi R., Wong J. W., Wu Z., Cheung T. H. (2020) "A long noncoding RNA, *LncMyoD*, modulates chromatin accessibility to regulate muscle stem cell myogenic lineage progression". *Proc. Natl. Acad. Sci. U. S. A*. **117**(51): pp. 32464-32475. doi: 10.1073/pnas.2005868117.

Dong C. (2021) "Cytokine Regulation and Function in T Cells". *Annu. Rev. Immunol*. **39**(1): pp. 51-76. doi: 10.1146/annurev-immunol-061020-053702.

Dos Santos R. S., Marroqui L., Velayos T., Olazagoitia-Garmendia A., Jauregi-Miguel A., Castellanos-Rubio A., Eizirik D. L., Castaño L., Santin I. (2019) "DEXI, a candidate gene for type 1 diabetes, modulates rat and human pancreatic beta cell inflammation via regulation of the

type I IFN/STAT signalling pathway". *Diabetologia*. **62**(3): pp. 459-472. doi: 10.1007/s00125-018-4782-0.

Downes D. J., Smith A. L., Karpinska M. A., Velychko T., Rue-Albrecht K., Sims D., Milne T. A., Davies J. O. J., Oudelaar A. M., Hughes J. R. (2022) "Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture". *Nat. Protoc*. **18**(1): p. 3155. doi: 10.1038/s41596-021-00651-w.

Dreger H., Ludwig A., Weller A., Baumann G., Stangl V., Stangl K. (2016) "Epigenetic suppression of iNOS expression in human endothelial cells: A potential role of Ezh2-mediated H3K27me3". *Genomics*. **107**(4): pp. 145-9.

Duan J., Wainwright M. S., Comeron J. M., Saitou N., Sanders A. R., Gelernter J., Gejman P. V. (2003) "Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor". *Hum. Mol. Genet*. **12**(3): p. 205-216. doi: 10.1093/hmg/ddg055.

Duan Y. Y., Chen X. F., Zhu R. J., Jia Y. Y., Huang X. T., Zhang M., Yang N., Dong S. S., Zeng M., Feng Z., Zhu D. L., Wu H., Jiang F., Shi W., Hu W. X., Ke X., Chen H., Liu Y., Jing R. H., Guo Y., Li M., Yang T. L. (2023) "High-throughput functional dissection of noncoding SNPs with biased allelic enhancer activity for insulin resistance-relevant phenotypes". *Am. J. Hum Genet*. **110**(8): pp. 1266-1288. doi: 10.1016/j.ajhg.2023.07.002.

Durbin R. M., Abecasis G. R., Altshuler D., Auton A., Brooks L. D., , Gibbs R. A., Hurles M. E., McVean G. A. (2010) "A map of human genome variation from population-scale sequencing". *Nature*. **467**(7319): pp. 1061-1073. doi: 10.1038/nature09534.

E ----------

Eijsbouts C. Q., Burren O. S., Newcombe P. J., Wallace C. (2019) "Fine mapping chromatin contacts in capture Hi-C data". *BMC Genomics*. **20**(1): p. 77. doi: 10.1186/s12864-018-5314-5.

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., D'Amato M., Halfvarson J., Jensen A. B., Lieb W., Degenhardt F., Forstner A. J., Hofmann A.; International IBD Genetics Consortium (IIBDGC); International Genetics of Ankylosing Spondylitis Consortium (IGAS); International PSC Study Group (IPSCSG); Genetic Analysis of Psoriasis Consortium (GAPC); Psoriasis Association Genetics Extension (PAGE); Schreiber S., Mrowietz U., Juran B. D., Lazaridis K. N., Brunak S., Dale A. M., Trembath R. C., Weidinger S., Weichenthal M., Ellinghaus E., Elder J. T., Barker J. N., Andreassen O. A., McGovern D. P., Karlsen T. H., Barrett J. C., Parkes M., Brown M. A., Franke A. (2016) "Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci". *Nat. Genet*. **48**(5): pp. 510-518. doi: 10.1038/ng.3528.

Ernst J., Kellis M. (2017) "Chromatin-state discovery and genome annotation with ChromHMM". *Nat. Protoc*. **12**(12) pp. :2478-2492. doi: 10.1038/nprot.2017.124. doi: 10.1038/nprot.2017.124.

Essex M., Rios Rodriguez V., Rademacher J., Proft F., Löber U., Markó L., Pleyer U., Strowig T., Marchand J., Kirwan J. A., Siegmund B., Forslund S. K., Poddubnyy D. (2024) "Shared and

Distinct Gut Microbiota in Spondyloarthritis, Acute Anterior Uveitis, and Crohn's Disease". *Arthritis Rheumatol*. **76**(1): pp. 48-58. doi: 10.1002/art.42658.

Evans D. M., *et al.* (2011) "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility". *Nat. Genet*. **43**(8): pp. 761-767. doi: 10.1038/ng.873.

Everaert C., Luypaert M., Maag J. L. V., Cheng Q. X., Dinger M. E., Hellemans J., Mestdagh P. (2017) "Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data". *Sci. Rep*. **7**(1): p. 1559. doi: 10.1038/s41598-017-01617-3.

Exarchou S., Lie E., Lindström U., Askling J., Forsblad-d'Elia H., Turesson C., Kristensen L. E., Jacobsson L. T. (2016) "Mortality in ankylosing spondylitis: results from a nationwide population-based study". *Ann. Rheum. Dis*. **75**(8): pp. 1466-1472. doi: 10.1136/annrheumdis-2015-207688.

F ----------

Faham M., Carlton V., Moorhead M., Zheng J., Klinger M., Pepin F., Asbury T., Vignali M., Emerson R. O., Robins H. S., Ireland J., Baechler-Gillespie E., Inman R. D. (2017) "Discovery of T Cell Receptor β Motifs Specific to HLA-B27-Positive Ankylosing Spondylitis by Deep Repertoire Sequence Analysis". *Arthritis Rheumatol*. **69**(4): pp. 774-784. doi: 10.1002/art.40028.

Fahy R. J., Doseff A. I., Wewers M. D. (1999) "Spontaneous human monocyte apoptosis utilizes a caspase-3-dependent pathway that is blocked by endotoxin and is independent of caspase-1". *J. Immunol*. 163(4): pp. 1755-1762.

Fan X., Qi B., Ma L., Ma F. (2019) "Screening of underlying genetic biomarkers for ankylosing spondylitis". *Mol. Med. Rep*. **19**(6): pp. 5263-5274. doi: 10.3892/mmr.2019.10188.

Fang H., ULTRA-DD Consortium, De Wolf H., Knezevic B., Burnham K. L., Osgood J., Sanniti A., Lledó Lara A., Kasela S., De Cesco S., Wegner J. K., Handunnetthi L., McCann F. E., Chen L., Sekine T., Brennan P. E., Marsden B. D., Damerell D., O'Callaghan C. A., Bountra C., Bowness P., Sundström Y., Milani L., Berg L., Göhlmann H. W., Peeters P. J., Fairfax B. P., Sundström M., Knight J. C. (2019) "A genetics-led approach defines the drug target landscape of 30 immune-related traits". *Nat. Genet*. **51**(7): pp. 1082-1091. doi: 10.1038/s41588-019-0456-1.

Fang Y., Liu J., Long Y., Wen J., Huang D., Xin L. (2022) "Knockdown of circular RNA hsa_circ_0003307 inhibits synovial inflammation in ankylosing spondylitis by regulating the PI3K/AKT pathway". *Adv. Clin. Exp. Med*. **31**(7): pp. 781–788. doi: 10.17219/acem/146830.

Farh K. K., Marson A., Zhu J., Kleinewietfeld M., Housley W. J., Beik S., Shoresh N., Whitton H., Ryan R. J., 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., Hafler D. A., Bernstein B. E. (2015) "Genetic and epigenetic fine mapping of causal autoimmune disease variants". *Nature*. **518**(7539): pp. 337-343. doi: 10.1038/nature13835.

Ferrara J. L., Levine J. E., Reddy P., Holler E. (2009) "Graft-versus-host disease". *Lancet*. **373**(9674): pp. 1550-1561. doi: 10.1016/S0140-6736(09)60237-3.

Flores S. K., Cheng Z., Jasper A. M., Natori K., Okamoto T., Tanabe A., Gotoh K., Shibata H., Sakurai A., Nakai T., Wang X., Zethoven M., Balachander S., Aita Y., Young W., Zheng S., Takekoshi K., Nakamura E., Tothill R. W., Aguiar R. C. T., Dahia P. L. M. (2019) "Synonymous

but Not Silent: A Synonymous VHL Variant in Exon 2 Confers Susceptibility to Familial Pheochromocytoma and von Hippel-Lindau Disease". J. Clin. Endocrinol. Metab. **104**(9): pp. 3826-3834. doi: 10.1210/jc.2019-00235.

Fonseca R., Burn T. N., Gandolfo L. C., Devi S., Park S. L., Obers A., Evrard M., Christo S. N., Buquicchio F. A., Lareau C. A., McDonald K. M., Sandford S. K., Zamudio N. M., Zanluqui N. G., Zaid A., Speed T. P., Satpathy A. T., Mueller S. N., Carbone F. R., Mackay L. K. (2022) "Runx3 drives a $CD8^+$ T cell tissue residency program that is absent in $CD4^+$ T cells". *Nat. Immunol*. **23**(8): pp. 1236-1245. doi: 10.1038/s41590-022-01273-4.

Freire-Pritchett P., Ray-Jones H., Della Rosa M., Eijsbouts C. Q., Orchard W. R., Wingett S. W., Wallace C., Cairns J., Spivakov M., Malysheva V. (2021) "Detecting chromosomal interactions in Capture Hi-C data with CHiCAGO and companion tools". *Nat. Protoc*. **16**(9): pp. 4144-4176. doi: 10.1038/s41596-021-00567-5.

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". *Nat. Genet*. **51**(12): p. 1664-1669. doi: 10.1038/s41588-019-0538-0.

G ----------

Gaither J. B. S., Lammi G. E., Li J. L., Gordon D. M., Kuck H. C., Kelly B. J., Fitch J. R., White P. (2021) "Synonymous variants that disrupt messenger RNA structure are significantly constrained in the human population". *Gigascience*. **10**(4): e. 23. doi: 10.1093/gigascience/giab023.

Gagliani N., Amezcua Vesely M. C., Iseppon A., Brockmann L., Xu H., Palm N. W., de Zoete M. R., Licona-Limón P., Paiva R. S., Ching T., Weaver C., Zi X., Pan X., Fan R., Garmire L. X., Cotton M. J., Drier Y., Bernstein B., Geginat J., Stockinger B., Esplugues E., Huber S., Flavell R. A. (2015) "Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation". *Nature*. **523**(7559): pp. 221-225. doi: 10.1038/nature14452.

Gao L., Wu J., Wang H., Yang Y., Zheng Z., Ni B., Wang X., Peng Y., Li Y. (2022) "LMO1 Plays an Oncogenic Role in Human Glioma Associated With NF-kB Pathway". *Front. Oncol*. 12(1): p. 770299. doi: 10.3389/fonc.2022.770299.

Garcia-Montoya L., Gul H., Emery P. (2018) "Recent advances in ankylosing spondylitis: understanding the disease and management". *F1000Res.* **7**(1): eF1000 doi: 10.12688/f1000research.14956.1.

Gateva V., Sandling J. K., Hom G., Taylor K. E., Chung S. A., Sun X., Ortmann W., Kosoy R., Ferreira R. C., Nordmark G., Gunnarsson I., Svenungsson E., Padyukov L., Sturfelt G., Jönsen A., Bengtsson A. A., Rantapää-Dahlqvist S., Baechler E. C., Brown E. E., Alarcón G. S., Edberg J. C., Ramsey-Goldman R., McGwin G. Jr., Reveille J. D., Vilá L. M., Kimberly R. P., Manzi S., Petri M. A., Lee A., Gregersen P. K., Seldin M. F., Rönnblom L., Criswell L. A., Syvänen A. C., Behrens T. W., Graham R. R. (2009) "A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus". *Nat. Genet*. **41**(11): pp. 1228-1233. doi: 10.1038/ng.468.

Gaublomme J. T., Yosef N., Lee Y., Gertner R. S., Yang L. V., Wu C., Pandolfi P. P., Mak T., Satija R., Shalek A. K., Kuchroo V. K., Park H., Regev A. (2015) "Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity". *Cell*. **163**(6): pp. 1400-1412. doi: 10.1016/j.cell.2015.11.009.

Gaulton K. J., Preissl S., Ren B. (2023) "Interpreting non-coding disease-associated human variants using single-cell epigenomics". *Nat. Rev. Genet*. **24**(8): pp. 516-534. doi: 10.1038/s41576-023-00598-6.

Ge X., Frank-Bertoncelj M., Klein K., McGovern A., Kuret T., Houtman M., Burja B., Micheroli R., Shi C., Marks M., Filer A., Buckley C. D., Orozco G., Distler O., Morris A. P., Martin P., Eyre S., Ospelt C. (2021) "Functional genomics atlas of synovial fibroblasts defining rheumatoid arthritis heritability". *Genome Biol*. **22**(1): p. 247. doi: 10.1186/s13059-021-02460-6.

George M. D., Venkatachalam S., Banerjee S., Baker J. F., Merkel P. A., Gavigan K., Curtis D., Danila M. I., Curtis J. R., Nowell W. B. (2021) "Concerns, Healthcare Use, and Treatment Interruptions in Patients with Common Autoimmune Rheumatic Diseases During the COVID-19 Pandemic". *J. Rheumatol*. **48**(4): pp. 603-607. doi: 10.3899/jrheum.201017.

Glass C. K., Natoli G. (2016) "Molecular control of activation and priming in macrophages". *Nat. Immunol*. **17**(1): pp. 26-33. doi: 10.1038/ni.3306.

Glatigny S., Duhen R., Arbelaez C., Kumari S., Bettelli E. (2015) "Integrin alpha L controls the homing of regulatory T cells during CNS autoimmunity in the absence of integrin alpha 4". *Sci. Rep*. **5**(1): p. 7834. doi: 10.1038/srep07834.

Glinos D. A., Soskic B., Williams C., Kennedy A., Jostins L., Sansom D. M., Trynka G. (2020) "Genomic profiling of T-cell activation suggests increased sensitivity of memory T cells to CD28 costimulation". *Genes Immun*. **21**(6-8): pp. 390-408. doi: 10.1038/s41435-020-00118-0.

Greenwald W. W., Li H., Benaglio P., Jakubosky D., Matsui H., Schmitt A., Selvaraj S., D'Antonio M., D'Antonio-Chronowska A., Smith E. N., Frazer K. A. (2019) "Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression". *Nat. Commun*. **10**(1): p. 1054. doi: 10.1038/s41467-019-08940-5.

Gregory A. P., Dendrou C. A., Attfield K. E., Haghikia A., Xifara D. K., Butter F., Poschmann G., Kaur G., Lambert L., Leach O. A., Prömel S., Punwani D., Felce J. H., Davis S. J., Gold R., Nielsen F. C., Siegel R. M., Mann M., Bell J. I., McVean G., Fugger L. (2021) "TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis". *Nature*. **488**(7412): pp. 508-511. doi: 10.1038/nature11307.

Gross T. J., Kremens K., Powers L. S., Brink B., Knutson T., Domann F. E., Philibert R. A., Milhem M. M., Monick M. M. (2014) "Epigenetic silencing of the human NOS2 gene: rethinking the role of nitric oxide in human macrophage inflammatory responses". *J. Immunol*. **192**(5): pp. 2326-2338. doi: 10.4049/jimmunol.1301758.

Gu Z., Gu L., Eils R., Schlesner M., Brors B. (2014) "Circlize Implements and enhances circular visualization in R". *Bioinformatics*. **30**(19): pp. 2811-2812.

Gupta L., Bhattacharya S., Agarwal V., Aggarwal A. (2016) "Elevated levels of serum MRP8/14 in ankylosing spondylitis: associated with peripheral arthritis and active disease". *Clin. Rheumatol*. **35**(12): pp. 3075-3079. doi: 10.1007/s10067-016-3448-x.

H ----------

Haberle V., Stark A. (2018) "Eukaryotic core promoters and the functional basis of transcription initiation". *Nat. Rev. Mol. Cell Biol*. **19**(10): pp. 621-637. doi: 10.1038/s41580-018-0028-8.

Hahn M. A., Li A. X., Wu X., Yang R., Drew D. A., Rosenberg D. W., Pfeifer G. P. (2014) "Loss of the polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors". *Cancer Res*. **7**4(13): pp. 3617-3629. doi: 10.1158/0008-5472.CAN-13-3147.

Hahne F., Ivanek R. (2016) "Visualizing Genomic Data Using Gviz and Bioconductor". Methods Mol. Biol. **1418**(1): pp. 335-351. doi: 10.1007/978-1-4939-3578-9_16.

Halpert M. M., Konduri V., Liang D., Vazquez-Perez J., Hofferek C. J., Weldon S. A., Baig Y., Vedula I., Levitt J. M., Decker W. K. (2020) "MHC class I and II peptide homology regulates the cellular immune response". *FASEB J*. **34**(6): pp. 8082-8101. doi: 10.1096/fj.201903002R.

Halwani R., Al-Kufaidy R., Vazquez-Tello A., Pureza M. A., BaHammam A. S., Al-Jahdali H., Alnassar S. A., Hamid Q., Al-Muhsen S. (2014) "IL-17 Enhances Chemotaxis of Primary Human B Cells during Asthma". *PLoS One*. **9**(12): e114604. doi: 10.1371/journal.pone.0114604.

Hamilton L., Barkham N., Bhalla A., Brittain R., Cook D., Jones G., Mackay K., Marshall D., Marzo-Ortega H., Murphy D., Riddell C., Sengupta R., Siebert S., Van Rossen L., Gaffney K.; BSR and BHPR Standards, Guidelines and Audit Working Group. "BSR and BHPR guideline for the treatment of axial spondyloarthritis (including ankylosing spondylitis) with biologics". *Rheumatology*. **56**(2): pp. 313-316. doi: 10.1093/rheumatology/kew223.

Hammond R. K., Pahl M. C., Su C., Cousminer D. L., Leonard M. E., Lu S., Doege C. A., Wagley Y., Hodge K. M., Lasconi C., Johnson M. E., Pippin J. A., Hankenson K. D., Leibel R. L., Chesi A., Wells A. D., Grant S. F. (2021) "Biological constraints on GWAS SNPs at suggestive significance thresholds reveal additional BMI loci". *Elife*. **10**(1): e62206. doi: 10.7554/eLife.62206.

Han Y., Jia Q., Jahani P.S., Hurrell B. P., Pan C., Huang P., Gukusvan J., Woodward N. C., Eskin E., Gilliland F. D., Akbari O., Hartiala J. A., Allavee H. (2020) "Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat Commun* **11**(1): p. 1776 (2020). doi: 10.1038/s41467-020-15649-3.

Hanssen L. L. P., Kassouf M. T., Oudelaar A. M., Biggs D., Preece C., Downes D. J., Gosden M., Sharpe J. A., Sloane-Stanley J. A., Hughes J. R., Davies B., Higgs D. R. (2017) "Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo". *Nat. Cell. Biol*. **19**(8): pp. 952-961. doi: 10.1038/ncb3573.

Hanson A. L., Sahhar J., Ngian G. S., Roddy J., Walker J., Stevens W., Nikpour M., Assassi S., Proudman S., Mayes M. D., Kenna T. J., Brown M. A. (2022) "Contribution of HLA and KIR Alleles to Systemic Sclerosis Susceptibility and Immunological and Clinical Disease Subtypes". *Front. Genet*. **13**(1): p. 913196. doi: 10.3389/fgene.2022.913196.

Harbour S. N., DiToro D. F., Witte S. J., Zindl C. L., Gao M., Schoeb T. R., Jones G. W., Jones S. A., Hatton R. D., Weaver C. T. (2020) "T_H17 cells require ongoing classic IL-6 receptor signaling to retain transcriptional and functional identity". *Sci. Immunol*. **5**(49): e. aaw2262. doi: 10.1126/sciimmunol.aaw2262.

Harmston N., Ing-Simmons E., Perry M., Barešić A., Lenhard B. (2015) “GenomicInteractions: An R/Bioconductor package for manipulating and investigating chromatin interaction data”. *BMC Genomics* **16**(1): p. 963. doi: 10.1186/s12864-015-2140-x.

Haroon N. N., Paterson J. M., Li P., Inman R. D., Haroon N. (2015) “Patients With Ankylosing Spondylitis Have Increased Cardiovascular and Cerebrovascular Mortality: A Population-Based Study”. *Ann. Intern. Med*. **163**(6): pp. 409-416. doi: 10.7326/M14-2470.

Harris H. L., Gu H., Olshansky M., Wang A., Farabella I., Eliaz Y., Kalluchi A., Krishna A., Jacobs M., Cauer G., Pham M., Rao S. S. P., Dudchenko O., Omer A., Mohajeri K., Kim S., Nichols M. H., Davis E. S., Gkountaroulis D., Udupa D., Aiden A. P., Corces V. G, Phanstiel D. H., Noble W. S., Nir G., Di Pierro M., Seo J. S., Talkowski M. E., Aiden E. L., Rowley M. J. (2023) “Chromatin alternates between A and B compartments at kilobase scale for subgenic organization”. *Nat. Commun*. **14**(1): p. 3303. doi: 10.1038/s41467-023-38429-1.

Harvey D., Pointon J. J., Evans D. M., Karaderi T., Farrar C., Appleton L. H., Sturrock R. D., Stone M. A., Oppermann U., Brown M. A., Wordsworth B. P. (2009) “Investigating the genetic association between ERAP1 and ankylosing spondylitis”. *Hum. Mol. Genet*. **18**(21): pp. 4204-4212. doi: 10.1093/hmg/ddp371.

Hay M., Thomas D. W., Craighead J. L., Economides C., Rosenthal J. (2014) “Clinical development success rates for investigational drugs”. *Nat. Biotechnol*. **32**(1): pp. 40-51. doi: 10.1038/nbt.2786.

Hisanaga T., Romani F., Wu S., Kowar T., Wu Y., Lintermann R., Fridrich A., Cho C. H., Chaumier T., Jamge B., Montgomery S. A., Axelsson E., Akimcheva S., Dierschke T., Bowman J. L., Fujiwara T., Hirooka S., Miyagishima S. Y., Dolan L., Tirichine L., Schubert D., Berger F. (2023) “The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes”. *Curr. Biol*. **33**(20): pp. 4367-4380.e9. doi: 10.1016/j.cub.2023.08.073.

He B., Lu N., Zhou Z. (2009) “Cellular and nuclear degradation during apoptosis”. *Curr. Opin. Cell Biol*. **21**(6): pp. 900-912. doi: 10.1016/j.ceb.2009.08.008.

He Z., Tung N. T. C., Yahara Y., Makino H., Yasuda T., Seki S., Suzuki K., Futakawa H., Kamei K., Kawaguchi Y. (2023) “Association between serum interleukin-17 levels and ectopic bone formation in OPLL patients with DISH”. *Rheumatology (Oxford)*. ePUB ahead of print. doi: 10.1093/rheumatology/kead555.

Heward J. A., Roux B. T., Lindsay M. A. (2015) “Divergent signalling pathways regulate lipopolysaccharide-induced eRNA expression in human monocytic THP1 cells”. *FEBS Lett*. **589**(3): pp. 396-406. doi: 10.1016/j.febslet.2014.12.026.

Hinnebusch A. G., Ivanov I. P., Sonenberg N. (2016) “Translational control by 5'-untranslated regions of eukaryotic mRNAs”. *Science*. **352**(6292): pp. 1413-1416. doi: 10.1126/science.aad9868.

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): pp. 934-947. doi: 10.1016/j.cell.2013.09.053.

Hnisz D., Weintraub A. S., Day D. S., Valton A. L., Bak R. O., Li C. H., Goldmann J., Lajoie B. R., Fan Z. P., Sigova A. A., Reddy J., Borges-Rivera D., Lee T. I., Jaenisch R., Porteus M. H.,

Dekker J., Young R. A. (2016) “Activation of proto-oncogenes by disruption of chromosome neighborhoods”. *Science*. **351**(6280): pp. 1454-1458. doi: 10.1126/science.aad9024.

Hoeksema M. A., de Winther M. P. (2016) “Epigenetic Regulation of Monocyte and Macrophage Function. Antioxid Redox Signal”. **25**(14): pp. 758-774. doi: 10.1089/ars.2016.6695.

Holgersen E. M., Gillespie A., Leavy O. C., Baxter J. S., Zvereva A., Muirhead G., Johnson N., Sipos O., Dryden N. H., Broome L. R., Chen Y., Kozin I., Dudbridge F., Fletcher O., Haider S. (2021) “Identifying high-confidence capture Hi-C interactions using CHiCANE”. *Nat. Protoc*. **16**(4): pp. 2257-2285. doi: 10.1038/s41596-021-00498-1.

Holliday R., Grigg G. W. (1993) “DNA methylation and mutation” Mut. Res. Fund. Mol. Mech. Mut. **285**(1): pp. 61-67. doi: 10.1016/0027-5107(93)90052-H.

Hos B. J., Tondini E., Camps M. G. M., Rademaker W., van den Bulk J., Ruano D., Janssen G. M. C., de Ru A. H., van den Elsen P. J., de Miranda N. F. C. C., van Veelen P. A., Ossendorp F. (2022) “Cancer-specific T helper shared and neo-epitopes uncovered by expression of the MHC class II master regulator CIITA”. *Cell Rep*. **41**(8): p. 111680. doi: 10.1016/j.celrep.2022.111680.

Howie B., Fuchsberger C., Stephens M., Marchini J., Abecasis G. R. (2012) “Fast and accurate genotype imputation in genome-wide association studies through pre-phasing”. *Nat. Genet*. **44**(8): pp. 955-959. doi: 10.1038/ng.2354.

Huang X. F., Li Z., De Guzman E., et al. (2020) “Genomewide Association Study of Acute Anterior Uveitis Identifies New Susceptibility Loci”. *Invest. Ophthalmol. Vis. Sci*. **61**(6): p. 3. doi: 10.1167/iovs.61.6.3.

Hulme K. D., Noye E. C., Short K. R., Labzin L. I. (2021) “Dysregulated Inflammation During Obesity: Driving Disease Severity in Influenza Virus and SARS-CoV-2 Infections”. *Front. Immunol*. **12**(1): p. 770066. doi: 10.3389/fimmu.2021.770066.

Hung T. C., Kingsley D. M., Boettiger A. N. (2024) “Boundary stacking interactions enable cross-TAD enhancer-promoter communication during limb development”. *Nat. Genet*. **56**(2): pp. 306-314. doi: 10.1038/s41588-023-01641-2.

Hunt S. E., Moore B., Amode R. M., Armean I. M., Lemos D., Mushtaq A., Parton A., Schuilenburg H., Szpak M., Thormann A., Perry E., Trevanion S. J., Flicek P., Yates A. D., Cunningham F. (2022) “Annotating and prioritizing genomic variants using the Ensembl Variant Effect Predictor-A tutorial”. *Hum. Mutat.* **43**(8): pp. 986-997. doi: 10.1002/humu.24298.

Hwang J. R., Byeon Y., Kim D., Park S. G. (2020) “Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development”. *Exp. Mol. Med*. **52**(5): pp. 750-761. doi: 10.1038/s12276-020-0435-8.

I ----------

Imakaev M., Fudenberg G., McCord R., Naumova N., Goloborodko A., Lajoie B., Dekker J., Mirny L. A. (2012) “Iterative correction of Hi-C data reveals hallmarks of chromosome organization”. *Nat. Methods* **9**(1): pp. 999–1003. doi: org/10.1038/nmeth.2148.

Inaba A., Tuong Z. K., Zhao T. X., Stewart A. P., Mathews R., Truman L., Sriranjan R., Kennet J., Saeb-Parsy K., Wicker L., Waldron-Lynch F., Cheriyan J., Todd J. A., Mallat Z., Clatworthy

M. R. (2023) “Low-dose IL-2 enhances the generation of IL-10-producing immunoregulatory B cells”. *Nat. Commun*. **14**(1): p. 2071. doi: 10.1038/s41467-023-37424-w.

Iotchkova V., Ritchie G. R. S., Geihs M., Morganella S., Min J. L., Walter K., Timpson N. J., UK10K Consortium; Dunham I., Birney E., Soranzo N. (2019) “GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals”. *Nat. Genet*. **51**(2): pp. 343-353. doi: 10.1038/s41588-018-0322-6.

Ito Y., Inoue S., Nakashima T., Zhang H., Li Y., Kasuya H., Matsukawa T., Wu Z., Yoshikawa T., Kataoka M., Ishikawa T., Kagoya Y. (2024) “Epigenetic profiles guide improved CRISPR/Cas9-mediated gene knockout in human T cells”. *Nucleic Acids Res*. **52**(1): pp. 141-153. doi: 10.1093/nar/gkad1076.

J ----------

Jablonski K. P., Carron L., Mozziconacci J., Forné T., Hütt M. T., Lesne A. (2022) “Contribution of 3D genome topological domains to genetic risk of cancers: a genome-wide computational study”. *Hum. Genomics*. **16**(1): p. 2. doi: 10.1186/s40246-022-00375-2.

Javierre B. M., Burren O. S., Wilder S. P., Kreuzhuber R., Hill S. M., Sewitz S., Cairns J., Wingett S. W., Várnai C., Thiecke M. J., Burden F., Farrow S., Cutler A. J., Rehnström K., Downes K., Grassi L., Kostadima M., Freire-Pritchett P., Wang F.; BLUEPRINT Consortium; Stunnenberg H. G., Todd J. A., Zerbino D. R., Stegle O., Ouwehand W. H., Frontini M., Wallace C., Spivakov M., Fraser P. (2016) “Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters”. *Cell.* **167**(5): pp. 1369-1384.e19. doi: 10.1016/j.cell.2016.09.037.

Jiang W., Jin Y., Zhang S., Ding Y., Huo K., Yang J., Zhao L., Nian B., Zhong T. P., Lu W., Zhang H., Cao X., Shah K. M., Wang N., Liu M., Luo J. (2022) “PGE2 activates EP4 in subchondral bone osteoclasts to regulate osteoarthritis”. *Bone Res*. **10**(1): P. 27. doi: 10.1038/s41413-022-00201-4.

Jilani M., Turcan A., Haspel N., Jagodzinski F. (2022) “Elucidating the Structural Impacts of Protein InDels”. *Biomolecules*. **12**(10): p. 1435. doi: 10.3390/biom12101435.

Jones G. T., Dean L. E., Pathan E., Hollick R. J., Macfarlane G. J. (2020) “Real-world evidence of TNF inhibition in axial spondyloarthritis: can we generalise the results from clinical trials?” *Ann. Rheum. Dis*. **79**(7): pp. 914-919. doi: 10.1136/annrheumdis-2019-216841.

Jonkman T. H., Dekkers K. F., Slieker R. C., Grant C. D., Ikram M. A., van Greevenbroek M. M. J., Franke L., Veldink J. H., Boomsma D. I., Slagboom P. E., Consortium BIOS, Heijmans B. T. (2022) “Functional genomics analysis identifies T and NK cell activation as a driver of epigenetic clock progression”. *Genome Biol*. **23**(1): p. 24. doi: 10.1186/s13059-021-02585-8.

Jostins L., Ripke S., Weersma R. K., et al. (2012) “Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease”. *Nature*. **491**(7422): pp. 119-124. doi: 10.1038/nature11582.

Ju Z., Li M., Xu J., Howell D. C., Li Z., Chen F. E. (2022) “Recent development on COX-2 inhibitors as promising anti-inflammatory agents: The past 10 years”. *Acta. Pharm. Sin. B*. **12**(6): pp. 2790-2807. doi: 10.1016/j.apsb.2022.01.002.

K ----------

Kaikkonen M. 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". *Mol Cell*. **51**(3): pp. 310-325. doi: 10.1016/j.molcel.2013.07.010.

Kao P. Y., Leung K. H., Chan L. W., Yip S. P., Yap M. K. (2017) "Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions". *Biochim. Biophys. Acta. Gen. Subj*. 1**861**(2): pp. 335-353. doi: 10.1016/j.bbagen.2016.11.030.

Karabacak Calviello A., Hirsekorn A., Wurmus R., Yusuf D., Ohler U. (2019) "Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modelling". *Genome Biol*. **20**(1): p. 42. doi: 10.1186/s13059-019-1654-y.

Kasowski M., Kyriazopoulou-Panagiotopoulou S., Grubert F., Zaugg J. B., Kundaje A., Liu Y., Boyle A. P., Zhang Q. C., Zakharia F., Spacek D. V., Li J., Xie D., Olarerin-George A., Steinmetz L. M., Hogenesch J. B., Kellis M., Batzoglou S., Snyder M. (2013) "Extensive variation in chromatin states across humans". *Science*. 342(6159): pp. 750-752. doi: 10.1126/science.1242510.

Kaukonen D., Kaukonen R., Polit L., Hennessy B. T., Lund R., Madden S. F. (2020) "Analysis of H3K4me3 and H3K27me3 bivalent promotors in HER2+ breast cancer cell lines reveals variations depending on estrogen receptor status and significantly correlates with gene expression". *BMC Med. Genomics*. **13**(1): p. 92. doi: 10.1186/s12920-020-00749-2.

Keat A., Barkham N., Bhalla A., Gaffney K., Marzo-Ortega H., Paul S., Rogers F., Somerville M., Sturrock R., Wordsworth P.: BSR Standards, Guidelines and Audit Working group. (2005) "BSR guidelines for prescribing TNF-alpha blockers in adults with ankylosing spondylitis. Report of a working party of the British Society for Rheumatology". *Rheumatology*. **44**(7): pp. 939-947. doi: 10.1093/rheumatology/keh669.

Keegan A. D., Leonard W. J., Zhu J. (2021) "Recent advances in understanding the role of IL-4 signaling". *Fac. Rev*. **10**(1): p. 71. doi: 10.12703/r/10-71.

Kendirli A., de la Rosa C., Lämmle K. F., Eglseer K., Bauer I. J., Kavaka V., Winklmeier S., Zhuo L., Wichmann C., Gerdes L. A., Kümpfel T., Dornmair K., Beltrán E., Kerschensteiner M., Kawakami N. (2023) "A genome-wide in vivo CRISPR screen identifies essential regulators of T cell migration to the CNS in a multiple sclerosis model". *Nat. Neurosci*. **26**(10): pp. 1713-1725. doi: 10.1038/s41593-023-01432-2.

Kent W. J., Zweig A. S., Barber G., Hinrichs A. S., Karolchik D. (2010) "BigWig and BigBed: enabling browsing of large distributed datasets". *Bioinformatics*. **26**(17): pp. 2204-2207. doi: 10.1093/bioinformatics/btq351.

Kenyon M., Maguire S., Rueda Pujol A., O'Shea F., McManus R. (2022) "The genetic backbone of ankylosing spondylitis: how knowledge of genetic susceptibility informs our understanding and management of disease". *Rheumatol. Int*. 2022 Dec;42(12):2085-2095. doi: 10.1007/s00296-022-05174-5.

Kemanetzoglou E., Andreadou E.(2017) "CNS Demyelination with TNF-α Blockers". *Curr. Neurol. Neurosci. Rep*. **17**(4): p. 36. doi: 10.1007/s11910-017-0742-1.

Kempfer R, Pombo A. (2020) “Methods for mapping 3D chromosome architecture*”. Nat. Rev. Genet*. **21**(4): pp. 207-226. doi: 10.1038/s41576-019-0195-2.

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. (2021) “A compendium of uniformly processed human gene expression and splicing quantitative trait loci”. *Nat. Genet*. **53**(9): pp. 1290-1299. doi: 10.1038/s41588-021-00924-w.

Kfoury-Beaumont N., Prakasam R., Pondugula S., Lagas J. S., Matkovich S., Gontarz P., Yang L., Yano H., Kim A. H., Rubin J. B., Kroll K. L. (2022) “The H3K27M mutation alters stem cell growth, epigenetic regulation, and differentiation potential”. *BMC Biol*. **20**(1): p. 124. doi: 10.1186/s12915-022-01324-0.

Kilic G., Kilic E., Ozgocmen S. (2014) “Relationship between psychiatric status, self-reported outcome measures, and clinical parameters in axial spondyloarthritis”. *Medicine*. **93**(29): e337. doi: 10.1097/MD.0000000000000337.

Kim J. Y., Yoon H. K., Song S. T., Park S. R., Shim S. C. (2017) “Expression of activation-induced cytidine deaminase splicing variants in patients with ankylosing spondylitis”. *Autoimmunity*. **50**(8): pp. 435-440. doi: 10.1080/08916934.2017.1385777.

Kitazawa T., Machlab D., Joshi O., Maiorano N., Kohler H., Ducret S., Kessler S., Gezelius H., Soneson C., Papasaikas P., López-Bendito G., Stadler M. B., Rijli F. M. (2021) “A unique bipartite Polycomb signature regulates stimulus-response transcription during development”. *Nat. Genet*. **53**(3): pp. 379-391. doi: 10.1038/s41588-021-00789-z.

Klasen C., Meyer A., Wittekind P. S., Waqué I., Nabhani S., Kofler D. M. (2019) “Prostaglandin receptor EP4 expression by Th17 cells is associated with high disease activity in ankylosing spondylitis”. *Arthritis Res. Ther*. **21**(1): p. 159. doi: 10.1186/s13075-019-1948-1.

Klatt D., Cheng E., Hoffmann D., Santilli G., Thrasher A. J., Brendel C., Schambach A. (2020) “Differential Transgene Silencing of Myeloid-Specific Promoters in the *AAVS1* Safe Harbor Locus of Induced Pluripotent Stem Cell-Derived Myeloid Cells”. *Hum. Gene Ther*. **31**(3-4): pp. 199-210. doi: 10.1089/hum.2019.194.

Klemm S. L., Shipony Z., Greenleaf W. J. (2019) “Chromatin accessibility and the regulatory epigenome*”. Nat. Rev. Genet*. **20**(4): pp. 207-220. doi: 10.1038/s41576-018-0089-8.

Ko C. L., Lin W. Z., Lee M. T., Chang Y. T., Lin H. C., Wu Y. S., Lin J. F., Pan K. T., Chang Y. C., Lee K. H., Lee Y. L., Hsieh T. T., Huang J. H., Wang C. H., Yang S. S., Chen H. C., Chu C. M. (2022) “Genome-wide association study reveals ethnicity-specific SNPs associated with ankylosing spondylitis in the Taiwanese population.]”. *J. Transl. Med*. **20**(1): p. 589. doi: 10.1183/09031936.00104407.

Kochan G., Krojer T., Harvey D., Fischer R., Chen L., Vollmar M., von Delft F., Kavanagh K. L., Brown M. A., Bowness P., Wordsworth P., Kessler B. M., Oppermann U. (2011) “Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming”. *Proc. Natl. Acad. Sci. U. S. A.* 108(19):7745-50. doi: 10.1073/pnas.1101262108. Epub 2011 Apr 20. PMID: 21508329; PMCID: PMC3093473.

Komech E. A., Koltakova A. D., Barinova A. A., Minervina A. A., Salnikova M. A., Shmidt E. I., Korotaeva T. V., Loginova E. Y., Erdes S. F., Bogdanova E. A., Shugay M., Lukyanov S.,

Lebedev Y. B., Zvyagin I. V. (2022) “TCR repertoire profiling revealed antigen-driven CD8+ T cell clonal groups shared in synovial fluid of patients with spondyloarthritis”. *Front. Immunol*. 13(1): p. 973243. doi: 10.3389/fimmu.2022.973243.

Kreitmaier P., Katsoula G., Zeggini E. (2023) “Insights from multi-omics integration in complex disease primary tissues”. *Trends Genet*. **39**(1): pp. 46-58. doi: 10.1016/j.tig.2022.08.005.

Kristofich J., Morgenthaler A. B., Kinney W. R., Ebmeier C. C., Snyder D. J., Old W. M., Cooper V. S., Copley S. D. (2018) “Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme”. *PLoS Genet*. **14**(8): e1007615. doi: 10.1371/journal.pgen.1007615.

Kroon F. P., van der Burg L. R., Ramiro S., Landewé R. B., Buchbinder R., Falzon L., van der Heijde D. (2015) “Non-steroidal anti-inflammatory drugs (NSAIDs) for axial spondyloarthritis (ankylosing spondylitis and non-radiographic axial spondyloarthritis)”. *Cochrane Database Syst. Rev*. **2015**(7): eCD010952. doi: 10.1002/14651858.CD010952.pub2.

Ku C., Loy E., Salim A., Pawitan Y., Chia K. S. (2010) “The discovery of human genetic variations and their use as disease markers: past, present and future”. *J. Hum. Genet.* **55**(1): pp. 403–415. doi: 10.1038/jhg.2010.55.

Kubo N., Ishii H., Xiong X., Bianco S., Meitinger F., Hu R., Hocker J. D., Conte M., Gorkin D., Yu M., Li B., Dixon J. R., Hu M., Nicodemi M., Zhao H., Ren B. (2021) “Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation”. *Nat. Struct. Mol. Biol*. **28**(2): pp. 152-161. doi: 10.1038/s41594-020-00539-5.

Kukurba K. R., Montgomery S. B. (2015) “RNA Sequencing and Analysis”. *Cold Spring Harb. Protoc*. **2015**(11): pp. 951-69. doi: 10.1101/pdb.top084970

Kumosaka N., Knights A. J., Gaffney D. J. (2019) “High-resolution genetic mapping of putative causal interactions between regions of open chromatin”. *Nat. Genet*. **51**(1): pp. 128-137. doi: 10.1038/s41588-018-0278-6.

Kwon O. C., Park J. H., Park M. C. (2021) “Tumour necrosis factor inhibitor tapering in patients with ankylosing spondylitis at low disease activity: factors associated with flare”. *Ther. Adv. Musculoskelet*. Dis. **13**(1): p. 1759720X20986732. doi: 10.1177/1759720X20986732.

L ----------

Lage S. L., Amaral E. P., Hilligan K. L., Laidlaw E., Rupert A., Namasivayan S., Rocco J., Galindo F., Kellogg A., Kumar P., Poon R., Wortmann G. W., Shannon J. P., Hickman H. D., Lisco A., Manion M., Sher A., Sereti I. (2022) “Persistent Oxidative Stress and Inflammasome Activation in $CD14^{high}CD16^{-}$ Monocytes From COVID-19 Patients”. *Front. Immunol.* 12(1): p, 799558. doi: 10.3389/fimmu.2021.799558.

Lan P. D., Nissley D. A., Sitarik I., Vu Q. V., Jiang Y., To P., Xia Y., Fried S. D., Li M. S., O'Brien E. P. (2024) “Synonymous Mutations Can Alter Protein Dimerization Through Localized Interface Misfolding Involving Self-entanglements”. *J. Mol. Biol*. **436**(6): p. 168487. doi: 10.1016/j.jmb.2024.168487.

Landes M. B., Rajaram M. V., Nguyen H., Schlesinger L. S. (2015) “Role for NOD2 in Mycobacterium tuberculosis-induced iNOS expression and NO production in human macrophages”. *J. Leukoc. Biol*. **97**(6): pp. 1111-1119. doi: 10.1189/jlb.3A1114-557R.

Landi M., Maldonado-Ficco H., Perez-Alamino R., Maldonado-Cocco J. A., Citera G., Arturi P., Sampaio-Barros P. D., Flores Alvarado D. E., Burgos-Vargas R., Santos E., Palleiro D., Gutiérrez M. A., Vieyra-Sousa E., Pimentel-Santos F., Paira S. O., Berman A., Barrezueta C. V., Vazquez-Mellado J., Collantes-Estevez E.; RESPONDIA Group. Fundación Reumatológica Argentina; Morteo O. (2016) "Gender differences among patients with primary ankylosing spondylitis and spondylitis associated with psoriasis and inflammatory bowel disease in an iberoamerican spondyloarthritis cohort". *Medicine.* **95**(51): e5652. doi: 10.1097/MD.0000000000005652.

Langmead B., Trapnell C., Pop M., Salzberg S. L. (2009) "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". *Genome Biol*. **10**(3): p. 25. doi: 10.1186/gb-2009-10-3-r25.

Lavarone E., Barbieri C. M., Pasini D. (2019) "Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity". *Nat. Commun*. **10**(1): p. 1679. doi: 10.1038/s41467-019-09624-w.

Lazarevic V., Glimcher L. H., Lord G. M. (2013) "T-bet: a bridge between innate and adaptive immunity". *Nat. Rev. Immunol*. **13**(11): pp. 777-89. doi: 10.1038/nri3536.

Lê S., Josse J., Husson F. (2008) "FactoMineR: An R Package for Multivariate Analysis" J. Stat. Softw., **25**(1): p. 18. doi: 10.18637/jss.v025.i01.

Lee B. H., Wu Z., Rhie S.K . "Characterizing chromatin interactions of regulatory elements and nucleosome positions, using Hi-C, Micro-C, and promoter capture Micro-C". *Epigenetics Chromatin*. **15**(1): p. 41. doi: 10.1186/s13072-022-00473-4.

Lee H. L., Jang J. W., Lee S. W., Yoo S. H., Kwon J. H., Nam S. W., Bae S. H., Choi J. Y., Han N. I., Yoon S. K. (2019) "Inflammatory cytokines and change of Th1/Th2 balance as prognostic indicators for hepatocellular carcinoma in patients treated with transarterial chemoembolization". *Sci. Rep*. **9**(1): p. 3260. doi: 10.1038/s41598-019-40078-8.

Lee J., Aoki T., Thumkeo D., Siriwach R., Yao C., Narumiya S. (2019) "T cell-intrinsic prostaglandin E_2-EP2/EP4 signaling is critical in pathogenic T_H17 cell-driven inflammation". *J. Allergy Clin. Immunol*. **143**(2): pp. 631-643. doi: 10.1016/j.jaci.2018.05.036.

Lee S. A., Kristjánsdóttir K., Kwak H. (2023) "eRNA co-expression network uncovers TF dependency and convergent cooperativity". *Sci. Rep*. **13**(1): p. 19085. doi: 10.1038/s41598-023-46415-2.

Lee T. H., Koo B. S., Nam B., Oh J. S., Park S. Y., Lee S., Joo K. B., Kim T. H. (2020) "Conventional disease-modifying antirheumatic drugs therapy may not slow spinal radiographic progression in ankylosing spondylitis: results from an 18-year longitudinal dataset". *Ther. Adv. Musculoskelet. Dis*. **12**(1): e1759720X20975912. doi: 10.1038/s41563-017-0002-4.

Lehmann B. D., Colaprico A., Silva T. C., Chen J., An H., Ban Y., Huang H., Wang L., James J. L., Balko J. M., Gonzalez-Ericsson P. I., Sanders M. E., Zhang B., Pietenpol J. A., Chen X. S. (2021) "Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes". *Nat. Commun*. **12**(1): p. 6276. doi: 10.1038/s41467-021-26502-6.

Lenart I., Truong L. H., Nguyen D. D., Rasiukienė O., Tsao E., Armstrong J., Kumar P., McHugh K., Pereira B. I., Maan B. S., Garstka M. A., Bowness P., Blake N., Powis S. J., Gould K., Nesbeth D., Antoniou A. N. (2022) "Intrinsic Folding Properties of the HLA-B27 Heavy Chain Revealed

by Single Chain Trimer Versions of Peptide-Loaded Class I Major Histocompatibility Complex Molecules". *Front. Immunol*. **13**(1): p. 902135. doi: 10.3389/fimmu.2022.902135.

Li D., Cao R., Dong W., Cheng M., Pan X., Hu Z., Hao J. (2023) "Identification of potential biomarkers for ankylosing spondylitis based on bioinformatics analysis". *BMC Musculoskelet. Disord*. **24**(1): p. 413. doi: 10.1186/s12891-023-06550-3.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. (2009) "1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools". *Bioinformatics*. **25**(16): pp. 2078-2079. doi: 10.1093/bioinformatics/btp352.

Li K., Liu Y., Cao H., Zhang Y., Gu Z., Liu X., Yu A., Kaphle P., Dickerson K. E., Ni M., Xu J. (2020) "Interrogation of enhancer function by enhancer-targeting CRISPR epigenetic editing". *Nat. Commun*. **11**(1): p. 485. doi: 10.1038/s41467-020-14362-5.

Li P., Zhang Y., Xu Y., Cao H., Li L. (2022) "Characteristics of $CD8^+$ and $CD4^+$ Tissue-Resident Memory Lymphocytes in the Gastrointestinal Tract". *Adv. Gut Micro. Res*. **2022**(1): e 9157455. doi: 10.1155/2022/9157455.

Li Q., Li J., Yu C. P., Chang S., Xie L. L., Wang S. (2021) "Synonymous mutations that regulate translation speed might play a non-negligible role in liver cancer development". *BMC Cancer*. **21**(1): p. 388. doi: 10.1186/s12885-021-08131-w.

Li Y., Zhang S., Zhang C., Wang M. (2020) "LncRNA MEG3 inhibits the inflammatory response of ankylosing spondylitis by targeting miR-146a". *Mol. Cell. Biochem* **466**(1-2): pp. 17–24. doi: 10.1007/s11010-019-03681-x

Li Z., Haynes K., Pennisi D. J., Anderson L. K., Song X., Thomas G. P., Kenna T., Leo P., Brown M. A. (2017) "Epigenetic and gene expression analysis of ankylosing spondylitis-associated loci implicate immune cells and the gut in the disease pathogenesis". *Genes Immun*. **18**(3): pp. 135-143. doi: 10.1038/gene.2017.11.

Li Z., Khan M. K., van der Linden S. M., Winkens B., Villiger P. M., Baumberger H., van Zandwijk H., Khan M. A., Brown M. A. (2023) "HLA-B27, axial spondyloarthritis and survival". *Ann. Rheum. Dis*. **82**(12): pp. 1558-1567. doi: 10.1136/ard-2023-224434.

Liao Y., Smyth G. K., Shi W. (2014) "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". *Bioinformatics*. **30**(7): pp. 923-930. doi: 10.1093/bioinformatics/btt656.

Lichou F., Trynka G. (2020) "Functional studies of GWAS variants are gaining momentum". *Nat. Commun*. **11**(1): p. 6283. doi: 10.1038/s41467-020-20188-y.

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*. **326**(5950): pp. 289-93. doi: 10.1126/science.1181369.

Lin R., Wu W., Chen H., Gao H., Wu X., Li G., He Q., Lu H., Sun M., Liu Z. (2022) "GPR65 promotes intestinal mucosal Th1 and Th17 cell differentiation and gut inflammation through downregulating NUAK2". *Clin. Transl. Med*. **12**(3): e771. doi: 10.1002/ctm2.771.

Liu T., Zhou Y., Ko K. S., Yang H. (2015) “Interactions between Myc and Mediators of Inflammation in Chronic Liver Diseases”. *Mediators Inflamm*. 2015;2015:276850. doi: 10.1155/2015/276850.

Liu C. H., Raj S., Chen C. H., Hung K. H., Chou C. T., Chen I. H., Chien J. T., Lin I. Y., Yang S. Y., Angata T., Tsai W. C., Wei J. C., Tzeng I. S., Hung S. C., Lin K. I. (2019) “HLA-B27-mediated activation of TNAP phosphatase promotes pathogenic syndesmophyte formation in ankylosing spondylitis”. *J. Clin. Invest*. **129**(12): pp. 5357-5373. doi: 10.1172/JCI125212.

Liu L., Yuan Y., Zhang S., Xu J., Zou J. (2021) “Osteoimmunological insights into the pathogenesis of ankylosing spondylitis”. *J. Cell Physiol*. **236**(9): pp. 6090-6100. doi: 10.1002/jcp.30313.

Liu D., Liu B., Lin C., Gu J. (2021) “Imbalance of Peripheral Lymphocyte Subsets in Patients With Ankylosing Spondylitis: A Meta-Analysis”. *Front. Immunol*. **12**(1): p. 696973. doi: 10.3389/fimmu.2021.696973.

Liu Q., Fang L., Wu C. (2022) “Alternative Splicing and Isoforms: From Mechanisms to Diseases”. *Genes (Basel)*. **13**(3): p. 401. doi: 10.3390/genes13030401.

Liu Y., Dekker J. (2022) “CTCF-CTCF loops and intra-TAD interactions show differential dependence on cohesin ring integrity”. *Nat. Cell Biol*. **24**(10): pp. 1516-1527. doi: 10.1038/s41556-022-00992-y.

Liu C., Yu J., Song A., Wang M., Hu J., Chen P., Zhao J., Li G. (2023) “Histone H1 facilitates restoration of H3K27me3 during DNA replication by chromatin compaction”. *Nat. Commun*. **14**(1): p. 4081. doi: 10.1038/s41467-023-39846-y.

Local, A., Huang, H., Albuquerque, C.P. Singh N., Lee A. Y., Wang W., Wang C., Hsia J. E., Shiau A. K., Ge K., Corbett K. D., Wang D., Zhou H., Ren B. (2018) “Identification of H3K4me1-associated proteins at mammalian enhancers”. *Nat Genet* **50**(1): pp. 73–82. doi: 10.1038/s41588-017-0015-6.

Long D., Yang B., Yang M., Xiong F., Zhu X., Tan Y., Wu H. (2023) “Bach2 in $CD4^+$ T cells from SLE patients modulates B-cell differentiation and IgG production”. *Eur. J. Immunol*. **53**(4): e2250109. doi: 10.1002/eji.202250109.

Lonjou C., Zhang W., Collins A., Tapper W. J., Elahi E., Maniatis N., Morton N. E. (2003) “Linkage disequilibrium in human populations”. *Proc. Natl. Acad. Sci. U. S. A*. **100**(10): pp. 6069-6074. doi: 10.1073/pnas.1031521100.

López de Castro J. A. (2018) “How ERAP1 and ERAP2 Shape the Peptidomes of Disease-Associated MHC-I Proteins”. *Front. Immunol*. **9**(1): p. 2463. doi: 10.3389/fimmu.2018.02463.

Love MI, Huber W, Anders S. (2014) “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. *Genome Biol*. **15**(12): p. 550. doi: 10.1186/s13059-014-0550-8.

Lucasson F., Richette P., Aouad K., Ryussen-Witrand A., Wendling D., Fautrel B., Gossec L. (2022) “Prevalence and consequences of psoriasis in recent axial spondyloarthritis: an analysis of the DESIR cohort over 6 years”. *RMD Open*. **8**(1): e001986. doi: 10.1136/rmdopen-2021-001986.

Luger K., Mäder A. W., Richmond R. K., Sargent D. F., Richmond T. J. (1997) "Crystal structure of the nucleosome core particle at 2.8 A resolution". *Nature*. **389**(6648): pp. 251-60. doi: 10.1038/38444.

Luo F., Zhao Z., Zhang J., Leng J. (2019) "Comparison of HLA-B*27 subtypes between Chinese patients with ankylosing spondylitis and non-ankylosing spondylitis carriers". J. Int. Med. Res. **47**(7): pp. 3171-3178. doi: 10.1177/0300060519853929.

Luthra-Guptasarma M., Singh B. (2004) "HLA-B27 lacking associated beta2-microglobulin rearranges to auto-display or cross-display residues 169-181: a novel molecular mechanism for spondyloarthropathies". *FEBS Lett*. **575**(1-3): p. 1-8. doi: 10.1016/j.febslet.2004.08.037.

Luo G., Li F., Li X., Wang Z. G., Zhang B. (2018) "TNF-α and RANKL promote osteoclastogenesis by upregulating RANK via the NF-κB pathway". *Mol. Med. Rep*. 17(5):6605-6611. doi: 10.3892/mmr.2018.8698.

Luchtel R. A., Zhao Y., Aggarwal R. K., Pradhan K., Maqbool S. B. (2022) "ETS1 is a novel transcriptional regulator of adult T-cell leukemia/lymphoma of North American descent". *Blood Adv*. **6**(20): pp. 5613-5624. doi: 10.1182/bloodadvances.2022007725.

Lynch S., Santos S. G., Campbell E. C., Nimmo A. M., Botting C., Prescott A., Antoniou A. N., Powis S. J. (2009) "Novel MHC class I structures on exosomes". *J. Immunol*. **183**(3): pp. 1884-1891. doi: 10.4049/jimmunol.0900798.

M ----------

Ma X., Jiang Z., Li N., Jiang W., Gao P., Yang M., Yu X., Wang G., Zhang Y. (2019) "Ets2 suppresses inflammatory cytokines through MAPK/NF-κB signaling and directly binds to the IL-6 promoter in macrophages". *Aging*. **11**(22): pp. 10610-10625. doi: 10.18632/aging.102480.

Ma Q., Shams H., Didonna A., Baranzini S. E., Cree B. A. C., Hauser S. L., Henry R. G., Oksenberg J. R. (2023) "Integration of epigenetic and genetic profiles identifies multiple sclerosis disease-critical cell types and genes". *Commun. Biol*. **6**(1): p. 342. doi: 10.1038/s42003-023-04713-5.

Ma S., Ming Y., Wu J., Cui G. (2024) "Cellular metabolism regulates the differentiation and function of T-cell subsets". *Cell Mol. Immunol*. **21**(5): pp. 419-435. doi: 10.1038/s41423-024-01148-8.

Macrae T. A., Fothergill-Robinson J., Ramalho-Santos M. (2022) "Regulation, functions and transmission of bivalent chromatin during mammalian development". *Nat. Rev. Mol Cell Biol*. **24**(1) pp. 6-26. doi: 10.1038/s41580-022-00518-2.

Mahmood F., Coates L. C., Helliwell P. S. (2018) "Current concepts and unmet needs in psoriatic arthritis". *Clin. Rheumatol*. **37**(2): pp. 297-305. doi: 10.1007/s10067-017-3908-y.

Mas G., Blanco E., Ballaré C., Sansó M., Spill Y. G., Hu D., Aoi Y., Le Dily F., Shilatifard A., Marti-Renom M. A., Di Croce L. (2018) "Promoter bivalency favors an open chromatin architecture in embryonic stem cells". *Nat. Genet*. **50**(10): pp. 1452-1462. doi: 10.1038/s41588-018-0218-5.

Marchingo J. M., Sinclair L. V., Howden A. J., Cantrell D. A. (2020) "Quantitative analysis of how Myc controls T cell proteomes and metabolic pathways during T cell activation". *Elife*. **9**(1): e53725. doi: 10.7554/eLife.53725.

Martin P., McGovern A., Orozco G., Duffus K., Yarwood A., Schoenfelder S., Cooper N. J., Barton A., Wallace C., Fraser P., Worthington J., Eyre S. (2015) "Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci". *Nat. Commun*. 6(1): p. 10069. doi: 10.1038/ncomms10069.

Martins A. L., Walavalkar N. M., Anderson W. D., Zang C., Guertin M. J. (2018) "Universal correction of enzymatic sequence bias reveals molecular signatures of protein/DNA interactions". *Nucleic Acids Res*. **46**(2): e9. doi: 10.1093/nar/gkx1053.

Martin-Mola E., Balsa A., García-Vicuna R., Gómez-Reino J., González-Gay M. A., Sanmartí R., Loza E. (2016) "Anti-citrullinated peptide antibodies and their value for predicting responses to biologic agents: a review". *Rheumatol. Int*. **36**(8): pp. 1043-1063. doi: 10.1007/s00296-016-3506-3.

Martinez F. O., Gordon S. (2014) "The M1 and M2 paradigm of macrophage activation: time for reassessment". *F1000Prime Rep*. **6**(1): p. 13. doi: 10.12703/P6-13.

Martini V., Silvestri Y., Ciurea A., Möller B., Danelon G., Flamigni F., Jarrossay D., Kwee I., Foglierini M., Rinaldi A., Cecchinato V., Uguccioni M. (2024) "Patients with ankylosing spondylitis present a distinct CD8 T cell subset with osteogenic and cytotoxic potential". *RMD Open*. **10**(1): e003926. doi: 10.1136/rmdopen-2023-003926.

Matsumoto M., Baba A., Yokota T., Nishikawa H., Ohkawa Y., Kayama H., Kallies A., Nutt S. L., Sakaguchi S., Takeda K., Kurosaki T., Baba Y. "Interleukin-10-producing plasmablasts exert regulatory function in autoimmune inflammation". *Immunity.* **41**(6): pp. 1040-51. doi: 10.1016/j.immuni.2014.10.016.

Mattick J. S., Amaral P. P., Carninci P., Carpenter S., Chang H. Y., Chen L. L., Chen R., Dean C., Dinger M. E., Fitzgerald K. A., Gingeras T. R., Guttman M., Hirose T., Huarte M., Johnson R., Kanduri C., Kapranov P., Lawrence J. B., Lee J. T., Mendell J. T., Mercer T. R., Moore K. J., Nakagawa S., Rinn J. L., Spector D. L., Ulitsky I., Wan Y., Wilusz J. E., Wu M. (2023) "Long non-coding RNAs: definitions, functions, challenges and recommendations". *Nat. Rev. Mol. Cell Biol*. 2023 Jun;24(6):430-447. doi: 10.1038/s41580-022-00566-8.

Mau W., Zeidler H., Mau R., Majewski A., Freyschmidt J., Stangel W., Deicher H. (1988) "Clinical features and prognosis of patients with possible ankylosing spondylitis. Results of a 10-year followup". *J. Rheumatol*. **15**(7): pp. 1109-1114.

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., Bates D., Hansen R. S., Neph S., Sabo P. J., Heimfeld S., Raubitschek A., Ziegler S., Cotsapas C., Sotoodehnia N., Glass I., Sunyaev S. R., Kaul R., Stamatoyannopoulos J. A. (2012) "Systematic localization of common disease-associated variation in regulatory DNA". *Science*. **337**(6099): pp. 1190-1195. doi: 10.1126/science.1222794.

Mayr C. (2017) "Regulation by 3'-Untranslated Regions". *Annu. Rev. Genet*. **51**(1): pp. 171-194. doi: 10.1146/annurev-genet-120116-024704.

McCarthy R. L., Kaeding K. E., Keller S. H., Zhong Y., Xu L., Hsieh A., Hou Y., Donahue G., Becker J. S., Alberto O., Lim B., Zaret K. S. (2021) "Diverse heterochromatin-associated proteins repress distinct classes of genes and repetitive elements". *Nat. Cell. Biol*. **23**(8): pp. 905-914. doi: 10.1038/s41556-021-00725-7.

McClure R., Massari P. (2014) "TLR-Dependent Human Mucosal Epithelial Cell Responses to Microbial Pathogens". *Front. Immunol*. **5**(1): p. 386. doi: 10.3389/fimmu.2014.00386.

McCutcheon S. R., Swartz A. M., Brown M. C., Barrera A., McRoberts Amador C., Siklenka K., Humayun L., Ter Weele M. A., Isaacs J. M., Reddy T. E., Allen A. S., Nair S. K., Antonia S. J., Gersbach C. A. (2023) "Transcriptional and epigenetic regulators of human $CD8^+$ T cell function identified through orthogonal CRISPR screens". *Nat. Genet*. **55**(12): pp. 2211-2223. doi: 10.1038/s41588-023-01554-0.

McInnes I. B., Szekanecz Z., McGonagle D., Maksymowych W. P., Pfeil A., Lippe R., Song I. H., Lertratanakul A., Sornasse T., Biljan A., Deodhar A. (2022) "A review of JAK-STAT signalling in the pathogenesis of spondyloarthritis and the role of JAK inhibition". Rheumatology (Oxford). **61**(5): pp. 1783-1794. doi: 10.1093/rheumatology/keab740.

McGeachy M. J., Cua D. J. (2007) "The link between IL-23 and Th17 cell-mediated immune pathologies". Semin Immunol. **19**(6): pp. 372-376. doi: 10.1016/j.smim.2007.10.012.

McGovern A., Schoenfelder S., Martin P., Massey J., Duffus K., Plant D., Yarwood A., Pratt A. G., Anderson A. E., Isaacs J. D., Diboll J., Thalayasingam N., Ospelt C., Barton A., Worthington J., Fraser P., Eyre S., Orozco G. (2016) "Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23". *Genome Biol*. **17**(1): p. 212. doi: 10.1186/s13059-016-1078-x.

McHugh K., Bowness P. (2012) "The link between HLA-B27 and SpA--new ideas on an old problem". *Rheumatology (Oxford)*. **51**(9): pp. 1529-39. doi: 10.1093/rheumatology/kes061.

McVeigh C. M., Cairns A. P. (2006) "Diagnosis and management of ankylosing spondylitis". *BMJ*. **333**(7568): pp. 581-585. doi: 10.1136/bmj.38954.689583.DE.

Mear J. P., Schreiber K. L., Münz C., Zhu X., Stevanović S., Rammensee H. G., Rowland-Jones S. L., Colbert R. A. (1999) "Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies". *J. Immunol*. **163**(12):6665-6670.

Meddens C. A., Harakalova M., van den Dungen N. A., Foroughi Asl H., Hijma H. J., Cuppen E. P., Björkegren J. L., Asselbergs F. W., Nieuwenhuis E. E., Mokry M. (2016) "Systematic analysis of chromatin interactions at disease associated loci links novel candidate genes to inflammatory bowel disease". *Genome Biol*. **17**(1): p. 247. doi: 10.1186/s13059-016-1100-3.

Meijer M., Agirre E., Kabbe M., van Tuijn C. A., Heskol A., Zheng C., Mendanha Falcão A., Bartosovic M., Kirby L., Calini D., Johnson M. R., Corces M. R., Montine T. J., Chen X., Chang H. Y., Malhotra D., Castelo-Branco G. (2022) "Epigenomic priming of immune genes implicates oligodendroglia in multiple sclerosis susceptibility". *Neuron*. **110**(7): pp. 1193-1210.e13. doi: 10.1016/j.neuron.2021.12.034.

Mendenhall E. M., Koche R. P., Truong T., Zhou V. W., Issac B., Chi A. S., Ku M., Bernstein B. E. (2010) "GC-rich sequence elements recruit PRC2 in mammalian ES cells". *PLoS Genet*. **6**(12) :e1001244. doi: 10.1371/journal.pgen.1001244.

Menegatti S., Bianchi E., Rogge L. (2019) “Anti-TNF Therapy in Spondyloarthritis and Related Diseases, Impact on the Immune System and Prediction of Treatment Responses”. *Front. Immunol.* **10**(1): p. 382. doi: 10.3389/fimmu.2019.00382.

Menezes A. C., Dixon C., Scholz A., Nicholson R., Leckenby A., Azevedo A., Baker S., Gilkes A. F., Davies S., Darley R. L., Tonks A. (2022) “RUNX3 overexpression inhibits normal human erythroid development”. *Sci. Rep*. **12**(1): p. 1243. doi: 10.1038/s41598-022-05371-z.

Merino E., Galocha B., Vázquez M. N., López de Castro J. A. (2008) “Disparate folding and stability of the ankylosing spondylitis-associated HLA-B*1403 and B*2705 proteins”. *Arthritis Rheum*. **58**(12): pp. 3693-3704. doi: 10.1002/art.24045.

Meuleman W., Muratov A., Rynes E., Halow J., Lee K., Bates D., Diegel M., Dunn D., Neri F., Teodosiadis A., Reynolds A., Haugen E., Nelson J., Johnson A., Frerker M., Buckley M., Sandstrom R., Vierstra J., Kaul R., Stamatoyannopoulos J. (2020) “Index and biological spectrum of human DNase I hypersensitive sites”. *Nature*. **584**(7820): pp. 244-251. doi: 10.1038/s41586-020-2559-3.

Mihai I. R., Burlui A. M., Rezus I. I., Mihai C., Macovei L. A., Cardoneanu A., Gavrilescu O., Dranga M., Rezus E. (2023) “Inflammatory Bowel Disease as a Paradoxical Reaction to Anti-TNF-α Treatment-A Review”. *Life (Basel)*. **13**(8): p. 1779. doi: 10.3390/life13081779.

Miller M., Rosenthal P., Beppu A., Mueller J. L., Hoffman H. M., Tam A. B., Doherty T. A., McGeough M. D., Pena C. A., Suzukawa M., Niwa M., Broide D. H. (2014) “ORMDL3 transgenic mice have increased airway remodeling and airway responsiveness characteristic of asthma”. *J. Immunol*. **192**(8): pp. 3475-3487. doi: 10.4049/jimmunol.1303047.

Minchin S., Lodge J. (2019) “Understanding biochemistry: structure and function of nucleic acids”. *Essays Biochem*. **63**(4): pp. 433-456. doi: 10.1042/EBC20180038.

Minderjahn J., Fischer A., Maier K., Mendes K., Nuetzel M., Raithel J., Stanewsky H., Ackermann U., Månsson R., Gebhard C., Rehli M. (2022) “Postmitotic differentiation of human monocytes requires cohesin-structured chromatin”. *Nat. Commun*. **13**(1): p. 4301. doi: 10.1038/s41467-022-31892-2.

Minnoye L., Marinov G. K., Krausgruber T., Pan L., Marand A. P., Secchia S., Greenleaf W. J., Furlong E. E. M., Zhao K., Schmitz R. J., Bock C., Aerts S. (2021) “Chromatin accessibility profiling methods”. *Nat. Rev. Methods Primers*. **1**(1): p. 10. doi: 10.1038/s43586-020-00008-9.

Moisan J., Grenningloh R., Bettelli E., Oukka M., Ho I. C. (2007) “Ets-1 is a negative regulator of Th17 differentiation”. *J. Exp. Med*. **204**(12): pp. 2825-2835. doi: 10.1084/jem.20070994.

Moore J. E., Purcaro M. J., Pratt H. E., Epstein C. B., Shoresh N., Adrian J., Kawli T., Davis C. A., Dobin A., Kaul R., Halow J., Van Nostrand E. L., Freese P., Gorkin D. U., Shen Y., He Y., Mackiewicz M., Pauli-Behn F., Williams B. A., Mortazavi A., Keller C. A., Zhang X. O., Elhajjajy S. I., Huey J., Dickel D. E., Snetkova V., Wei X., Wang X., Rivera-Mulia J. C., Rozowsky J., Zhang J., Chhetri S. B., Zhang J., Victorsen A., White K. P., Visel A., Yeo G. W., Burge C. B., Lécuyer E., Gilbert D. M., Dekker J., Rinn J., Mendenhall E. M., Ecker J. R., Kellis M., Klein R. J., Noble W. S., Kundaje A., Guigó R., Farnham P. J., Cherry J. M., Myers R. M., Ren B., Graveley B. R., Gerstein M. B., Pennacchio L. A., Snyder M. P., Bernstein B. E., Wold B., Hardison R. C., Gingeras T. R., Stamatoyannopoulos J. A., Weng Z. (2020) “Expanded

encyclopaedias of DNA elements in the human and mouse genomes". *Nature*. **583**(7818): pp. 699-710. doi: 10.1038/s41586-020-2493-4.

Morova T., Ding Y., Huang C. F., Sar F., Schwarz T., Giambartolomei C., Baca S. C., Grishin D., Hach F., Gusev A., Freedman M. L., Pasaniuc B., Lack N. A. (2023) "Optimized high-throughput screening of non-coding variants identified from genome-wide association studies". *Nucleic Acids Res*. **51**(3): e18. doi: 10.1093/nar/gkac1198.

Mortezaei Z., Tavallaei M. (2021) "Recent innovations and in-depth aspects of post-genome wide association study (Post-GWAS) to understand the genetic basis of complex phenotypes". Heredity (Edinb). **127**(6): pp. 485-497. doi: 10.1038/s41437-021-00479-w.

Mouly E., Chemin K., Nguyen H. V., Chopin M., Mesnard L., Leite-de-Moraes M., Burlen-defranoux O., Bandeira A., Bories J. C. (2010) "The Ets-1 transcription factor controls the development and function of natural regulatory T cells". *J. Exp. Med*. **207**(10): pp. 2113-2125. doi: 10.1084/jem.20092153.

Mounach A., El Maghraoui A. (2014) "Efficacy and safety of adalimumab in ankylosing spondylitis". *Open Access Rheumatol*. **6**(1): pp. 83-90. doi: 10.2147/OARRR.S44550.

Muñoz-Ortego J., Vestergaard P., Rubio J. B., Wordsworth P., Judge A., Javaid M. K., Arden N. K, Cooper C., Díez-Pérez A., Prieto-Alhambra D. (2014) "Ankylosing spondylitis is associated with an increased risk of vertebral and nonvertebral clinical fractures: a population-based cohort study". *J. Bone. Miner. Res*. **29**(8): pp. 1770-1776. doi: 10.1002/jbmr.2217.

N ----------

Nakato R., Sakata T. (2021) "Methods for ChIP-seq analysis: A practical workflow and advanced applications". *Methods*. 2021 Mar;187:44-53. doi: 10.1016/j.ymeth.2020.03.005.

Narula A., Ellis J., Taliaferro J. M., Rissland O. S. (2019) "Coding regions affect mRNA stability in human cells". *RNA*. **25**(12): pp. 1751-1764. doi: 10.1261/rna.073239.119.

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., Collins R. L., Dey K., Pfister H., Price A. L., Epstein C. B., Kundaje A., Xavier R. J., Daly M. J., Huang H., Finucane H. K., Hacohen N., Lander E. S., Engreitz J. M. (2021) "Genome-wide enhancer maps link risk variants to disease genes". *Nature*. **593**(7858): pp. 238-243. doi: 10.1038/s41586-021-03446-x.

Neefjes J., Jongsma M. L., Paul P., Bakke O. (2011) "Towards a systems understanding of MHC class I and MHC class II antigen presentation". *Nat. Rev. Immunol*. **211**(12): pp. 823-36. doi: 10.1038/nri3084.

Nelson M. R., Tipney H., Painter J. L., Shen J., Nicoletti P., Shen Y., Floratos A., Sham P. C., Li M. J., Wang J., Cardon L. R., Whittaker J. C., Sanseau P. (2015) "The support of human genetic evidence for approved drug indications". *Nat. Genet*. **47**(8): pp. 856-860. doi: 10.1038/ng.3314.

Nived P., Pettersson Å., Jönsson G., Bengtsson A. A., Settergren B., Skattum L., Johansson Å., Kapetanovic M. C. (2021) "Methotrexate reduces circulating Th17 cells and impairs plasmablast and memory B cell expansions following pneumococcal conjugate immunization in RA patients". *Sci. Rep*. **11**(1): p. 9199. doi: 10.1038/s41598-021-88491-2.

Nguyen M. L., Jones S. A., Prier J. E., Russ B. E. (2015) “Transcriptional Enhancers in the Regulation of T Cell Differentiation”. *Front. Immunol*. **6**(1): p, 462. doi: 10.3389/fimmu.2015.00462.

Noguchi S., Arakawa T., Fukuda S. *et al.* (2017) “FANTOM5 CAGE profiles of human and mouse samples”. *Sci. Data*. **4**(1): p. 170112. doi: 10.1038/sdata.2017.112.

Nurieva R., Yang X. O., Martinez G., Zhang Y., Panopoulos A. D., Ma L., Schluns K., Tian Q., Watowich S. S., Jetten A. M., Dong C. (2007) “Essential autocrine regulation by IL-21 in the generation of inflammatory T cells”. *Nature*. **448**(7152): pp. 480-483. doi: 10.1038/nature05969.

O ----------

Oelschlaeger P. (2024) “Molecular Mechanisms and the Significance of Synonymous Mutations”. Biomolecules. **14**(1): p. 132. doi: 10.3390/biom14010132.

Ogdie A., Benjamin Nowell W., Reynolds R., Gavigan K., Venkatachalam S., de la Cruz M., Flood E., Schwartz E. J., Romero B., Park Y. (2019) “Real-World Patient Experience on the Path to Diagnosis of Ankylosing Spondylitis”. *Rheumatol. Ther*. **6**(2): pp. 255-267. doi: 10.1007/s40744-019-0153-7.

Oh S., Shao J., Mitra J., Xiong F., D'Antonio M., Wang R., Garcia-Bassets I., Ma Q., Zhu X., Lee J. H., Nair S. J., Yang F., Ohgi K., Frazer K. A., Zhang Z. D., Li W., Rosenfeld M. G. (2021) “Enhancer release and retargeting activates disease-susceptibility genes”. *Nature*. **595**(7869): pp. 735-740. doi: 10.1038/s41586-021-03577-1.

Ohradanova-Repic A., Machacek C., Fischer M. B., Stockinger H. (2016) “Differentiation of human monocytes and derived subsets of macrophages and dendritic cells by the HLDA10 monoclonal antibody panel”. *Clin. Transl. Immunology*. **5**(1): e55. doi: 10.1038/cti.2015.39.

Oram S. H., Thoms J., Sive J. I., Calero-Nieto F. J., Kinston S. J., Schütte J., Knezevic K., Lock R. B., Pimanda J. E., Göttgens B. (2013) “Bivalent promoter marks and a latent enhancer may prime the leukaemia oncogene LMO1 for ectopic expression in T-cell leukaemia. Leukemia”. **27**(6): p. 1348-1357. doi: 10.1038/leu.2013.2.

Ostrovsky O., Baryakh P., Morgulis Y., Mayorov M., Bloom N., Beider K., Shimoni A., Vlodavsky I., Nagler A. (2021) “The HPSE Gene Insulator-A Novel Regulatory Element That Affects Heparanase Expression, Stem Cell Mobilization, and the Risk of Acute Graft versus Host Disease”. *Cells*. **10**(10): p. 2523. doi: 10.3390/cells10102523.

Oudelaar A. M., Beagrie R. A., Gosden M., de Ornellas S., Georgiades E., Kerry J., Hidalgo D., Carrelha J., Shivalingam A., El-Sagheer A. H., Telenius J. M., Brown T., Buckle V. J., Socolovsky M., Higgs D. R., Hughes J. R. (2020) “Dynamics of the 4D genome during in vivo lineage specification and differentiation”. *Nat. Commun*. **11**(1): p. 2722. doi: 10.1038/s41467-020-16598-7.

Owens D. D. G., Anselmi G., Oudelaar A. M., Downes D. J., Cavallo A., Harman J. R., Schwessinger R., Bucakci A., Greder L., de Ornellas S., Jeziorska D., Telenius J., Hughes J. R., de Bruijn M. F. T. R. (2022) “Dynamic Runx1 chromatin boundaries affect gene expression in hematopoietic development”. *Nat. Commun*. **13**(1): p. 773. doi: 10.1038/s41467-022-28376-8.

P ----------

Paczkowska M., Barenboim J., Sintupisut N., Fox N. S., Zhu H., Abd-Rabbo D., Mee M. W., Boutros P. C.; PCAWG Drivers and Functional Interpretation Working Group; Reimand J.; PCAWG Consortium. (2020) "Integrative pathway enrichment analysis of multivariate omics data". *Nat. Commun*. **11**(1): p. 735. doi: 10.1038/s41467-019-13983-9. Erratum in: Nat Commun. **13**(1): p. 7570.

Palmieri E. M., Gonzalez-Cotto M., Baseler W. A., Davies L. C., Ghesquière B., Maio N., Rice C. M., Rouault T. A., Cassel T., Higashi R. M., Lane A. N., Fan T. W., Wink D. A., McVicar D. W. (2020) "Nitric oxide orchestrates metabolic rewiring in M1 macrophages by targeting aconitase 2 and pyruvate dehydrogenase". *Nat. Commun*. **11**(1): p. 698. doi: 10.1038/s41467-020-14433-7.

Pang B., van Weerd J. H., Hamoen F. L. , Snyder M. P. (2022) "Identification of non-coding silencer elements and their regulation of gene expression". *Nat. Rev. Mol. Cell Biol.* **24**(6): pp. 383-395. doi: 10.1038/s41580-022-00549-9.

Pang B., Snyder M. P. (2020) "Systematic identification of silencers in human cells". *Nat. Genet*. **52**(3): pp. 254-263. doi: 10.1038/s41588-020-0578-5.

Papagoras C., Chrysanthopoulou A., Mitsios A., Tsironidou V., Ritis K. (2021) "Neutrophil Extracellular Traps and Interleukin 17 in Ankylosing Spondylitis". *Mediterr. J. Rheumatol*. **32**(2): pp. 182-185. doi: 10.31138/mjr.32.2.182.

Papayannopoulos V. (2018) "Neutrophil extracellular traps in immunity and disease". *Nat. Rev. Immunol.* **18**(2): pp. 134-147. doi: 10.1038/nri.2017.105.

Park L. (2019) "Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants". *Sci. Rep*. **9**(1): p. 11380. doi: 10.1038/s41598-019-47832-y.

Pedersen O. B., Svendsen A. J., Ejstrup L., Skytthe A., Harris J. R., Junker P. (2008) "Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental effectors in disease causation". *Scand. J. Rheumatol*. ;**37**(2): pp. 120-126. doi: 10.1080/03009740701824613.

Pelletier M., Maggi L., Micheletti A., Lazzeri E., Tamassia N., Costantini C., Cosmi L., Lunardi C., Annunziato F., Romagnani S., Cassatella M. A. (2010) "Evidence for a cross-talk between human neutrophils and Th17 cells". *Blood*. **115**(2): pp. 335-343. doi: 10.1182/blood-2009-04-216085.

Perez-Sanchez C., Font-Ugalde P., Ruiz-Limon P., Lopez-Pedrera C., Castro-Villegas M. C., Abalos-Aguilera M. C., Barbarroja N., Arias-de la Rosa I., Lopez-Montilla M. D., Escudero-Contreras A., Lopez-Medina C., Collantes-Estevez E., Jimenez-Gomez Y. (2018)" Circulating microRNAs as potential biomarkers of disease activity and structural damage in ankylosing spondylitis patients". *Hum. Mol. Genet*. **27**(5): pp. 875-890. doi: 10.1093/hmg/ddy008.

Peters J. E., Lyons P. A., Lee J. C., Richard A. C., Fortune M. D., Newcombe P. J., Richardson S., Smith K. G. (2016) "Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease". *PLoS Genet*. **12**(3) :e1005908. doi: 10.1371/journal.pgen.1005908.

Pieterse E., Rother N., Yanginlar C., Hilbrands L. B., van der Vlag J. (2016) "Neutrophils Discriminate between Lipopolysaccharides of Different Bacterial Sources and Selectively Release Neutrophil Extracellular Traps". *Front. Immunol*. 7(1): pp. 484. doi: 10.3389/fimmu.2016.00484.

Pitarresi J. R., Liu X., Sharma S. M., Cuitiño M. C., Kladney R. D., Mace T. A., Donohue S., Nayak S. G., Qu C., Lee J., Woelke S. A., Trela S., LaPak K., Yu L., McElroy J., Rosol T. J., Shakya R., Ludwig T., Lesinski G. B., Fernandez S. A., Konieczny S. F., Leone G., Wu J., Ostrowski M. C. (2016) "Stromal ETS2 Regulates Chemokine Production and Immune Cell Recruitment during Acinar-to-Ductal Metaplasia". *Neoplasia*. **18**(9): pp, 541-552. doi: 10.1016/j.neo.2016.07.006.

Pitzalis C., Jones G. W., Bombardieri M., Jones S. A. (2014) "Ectopic lymphoid-like structures in infection, cancer and autoimmunity". *Nat. Rev. Immunol.* **14**(7): pp. 447-62. doi: 10.1038/nri3700.

Plenge R., Scolnick E., Altshuler D. (2013) "Validating therapeutic targets through human genetics". *Nat. Rev. Drug Discov.* **12**(1): pp. 581–594. do:10.1038/nrd4051.

Poddubnyy D., Vahldiek J., Spiller I., Buss B., Listing J., Rudwaleit M., Sieper J. (2011) "Evaluation of 2 screening strategies for early identification of patients with axial spondyloarthritis in primary care". *J. Rheumatol*. **38**(11): pp. 2452-2460. doi: 10.3899/jrheum.110070.

Pollex T., Rabinowitz A., Gambetta M. C., Marco-Ferreres R., Viales R. R., Jankowski A., Schaub C., Furlong E. E. M. (2024) "Enhancer-promoter interactions become more instructive in the transition from cell-fate specification to tissue differentiation". *Nat. Genet*. **56**(4): pp. 686-696. doi: 10.1038/s41588-024-01678-x.

Pradhananga S., Spalinskas R., Poujade F. A., Eriksson P., Sahlén P. (2020) "Promoter anchored interaction landscape of THP-1 macrophages captures early immune response processes". *Cell Immunol*. **355**(1): p. 104148. doi: 10.1016/j.cellimm.2020.104148.

Prager I., Liesche C., van Ooijen H., Urlaub D., Verron Q., Sandström N., Fasbender F., Claus M., Eils R., Beaudouin J., Önfelt B., Watzl C. (2019) "NK cells switch from granzyme B to death receptor-mediated cytotoxicity during serial killing". *J. Exp. Med*. **216**(9): pp. 2113-2127. doi: 10.1084/jem.20181454.

Preston G. C., Sinclair L. V., Kaskar A., Hukelmann J. L., Navarro M. N., Ferrero I., MacDonald H. R., Cowling V. H., Cantrell D. A. (2015) "Single cell tuning of Myc expression by antigen receptor signal strength and interleukin-2 in T lymphocytes". *EMBO J*. **34**(15): pp. 2008-2024. doi: 10.15252/embj.201490252.

Q ----------

Qaiyum Z., Gracey E., Yao Y., Inman R. D. (2019) "Integrin and transcriptomic profiles identify a distinctive synovial CD8+ T cell subpopulation in spondyloarthritis". *Ann. Rheum. Dis*. **78**(11): pp. 1566-1575. doi: 10.1136/annrheumdis-2019-215349.

Quartuccio L., Zabotti A., Del Zotto S., Zanier L., De Vita S., Valent F. (2018) "Risk of serious infection among patients receiving biologics for chronic inflammatory diseases: Usefulness of administrative data". *J. Adv. Res*. 15(1): pp. 87-93. doi: 10.1016/j.jare.2018.09.003.

Queiro R., Aurrecoechea E., Alonso Castro S., Villa Blanco I., Brandy-Garcia A., Linge R. (2023) “Interleukin-17-targeted treatment in patients with spondyloarthritis and associated cardiometabolic risk profile”. *Front. Immunol*. 14(1): pp. 1203372. doi: 10.3389/fimmu.2023.1203372.

Quevedo M., Meert L., Dekker M. R., Dekkers D. H. W., Brandsma J. H., van den Berg D. L. C., Ozgür Z., van IJcken W. F. J., Demmers J., Fornerod M., Poot R. A. (2019) “Mediator complex interaction partners organize the transcriptional network that defines neural stem cells”. *Nat. Commun*. **10**(1): p. 2669. doi: 10.1038/s41467-019-10502-8.

R ----------

Rahman S., Zorca C. E., Traboulsi T., Noutahi E., Krause M. R., Mader S., Zenklusen D. (2017) “Single-cell profiling reveals that eRNA accumulation at enhancer-promoter loops is not required to sustain transcription”. *Nucleic Acids Res*. **45**(6): p. 3017-3030. doi: 10.1093/nar/gkw1220.

Rashid T., Wilson C., Ebringer A. (2013) “The link between ankylosing spondylitis, Crohn's disease, Klebsiella, and starch consumption”. *Clin. Dev. Immunol*. **2013**(1): p. 872632. doi: 10.1155/2013/872632.

Rasmussen T. K., Andersen T., Hvid M., Hetland M. L., Hørslev-Petersen K., Stengaard-Pedersen K., Holm C. K., Deleuran B. (2010) “Increased interleukin 21 (IL-21) and IL-23 are associated with increased disease activity and with radiographic status in patients with early rheumatoid arthritis”. *J. Rheumatol*. **37**(10): pp. 2014-2020. doi: 10.3899/jrheum.100259.

Razawy W., Alves C. H., Koedam M., Asmawidjaja P. S., Mus A. M. C., Oukka M., Leenen P. J. M., Visser J. A., van der Eerden B. C. J., Lubberts E. (2021) “IL-23 receptor deficiency results in lower bone mass via indirect regulation of bone formation”. *Sci. Rep*. **11**(1): p. 10244. doi: 10.1038/s41598-021-89625-2.

Reimand J., Isserlin R., Voisin V., Kucera M., Tannus-Lopes C., Rostamianfar A., Wadi L., Meyer M., Wong J., Xu C., Merico D., Bader G. D. “Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap”. *Nat. Protoc*. **14**(2): pp. 482-517. doi: 10.1038/s41596-018-0103-9.

Ren C., Li M., Zheng Y., Cai B., Du W., Zhang H., Wu F., Tong M., Lin F., Wang J., Quan R. (2022) “Single-cell RNA-seq reveals altered NK cell subsets and reduced levels of cytotoxic molecules in patients with ankylosing spondylitis”. *J. Cell Mol. Med*. **26**(4): pp. 1071-1082. doi: 10.1111/jcmm.17159.

Ren Y., Chen B., Zhang M. (2023) “Distinct Prognostic and Immunological Roles of ETS1 and ETS2: A Pan-Cancer Analysis”. *Biomed. Res. Int*. **2023**(1): p. 4343350. doi: 10.1155/2023/4343350.

Resorlu M., Gokmen F., Resorlu H., Adam G., Akbal A., Cevizci S., Sariyildirim A., Savas Y., Guven M., Aras A. B. (2015) “Association between apparent diffusion coefficient and intervertebral disc degeneration in patients with ankylosing spondylitis”. *Int. J. Clin. Exp. Med.* **8**(1): pp. 1241-1246.

Reveille J. D., Lee M., Gensler L. S., Ward M. M., Hwang M. C., Learch T. J., Tahanan A., Diekman L., Rahbar M. H., Ishimori M. L., Weisman M. H. (2020) “The changing profile of ankylosing spondylitis in the biologic era”. *Clin. Rheumatol*. **39**(9): pp. 2641-2651. doi: 10.1007/s10067-020-05260-8.

Reveille J. D. (2015) “Biomarkers for diagnosis, monitoring of progression, and treatment responses in ankylosing spondylitis and axial spondyloarthritis”. *Clin. Rheumatol*. **34**(6): pp. 1009-18. doi: 10.1007/s10067-015-2949-3.

Reveille J. D. (2014) “An update on the contribution of the MHC to AS susceptibility”. *Clin. Rheumatol*. **33**(6) pp. :749-757. doi: 10.1007/s10067-014-2662-7.

Reveille J. D., Weisman M. H. (2013) “The epidemiology of back pain, axial spondyloarthritis and HLA-B27 in the United States”. *Am. J. Med. Sci*. **345**(6): pp. 431-436. doi: 10.1097/maj.0b013e318294457f.

Reveille J. D., *et al.* (2010) “Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci”. *Nat. Genet*. **42**(2): pp. 123-127. doi: 10.1038/ng.513.

Revu S., Wu J., Henkel M., Rittenhouse N., Menk A., Delgoffe G. M., Poholek A. C., McGeachy M. J. (2018) “IL-23 and IL-1β Drive Human Th17 Cell Differentiation and Metabolic Reprogramming in Absence of CD28 Costimulation”. *Cell. Rep*. **22**(10): pp. 2642-2653. doi: 10.1016/j.celrep.2018.02.044.

Rezaiemanesh A., Mahmoudi M., Amirzargar A. A., Vojdanian M., Babaie F., Mahdavi J., Rajabinejad M., Jamshidi A. R., Nicknam M. H. (2022) “Upregulation of Unfolded Protein Response and ER Stress-Related IL-23 Production in M1 Macrophages from Ankylosing Spondylitis Patients”. *Inflammation*. **45**(2): pp. 665-676. doi: 10.1007/s10753-021-01575-z.

Rezaiemanesh A., Abdolmaleki M., Abdolmohammadi K., Aghaei H., Pakdel F. D., Fatahi Y., Soleimanifar N., Zavvar M., Nicknam M. H. (2018) “Immune cells involved in the pathogenesis of ankylosing spondylitis”. *Biomed. Pharmacother*. **100**(1): pp. 198-204. doi: 10.1016/j.biopha.2018.01.108.

Richter W. F., Nayak S., Iwasa J., Taatjes D. J. (2022) “The Mediator complex as a master regulator of transcription by RNA polymerase II”. *Nat. Rev. Mol. Cell Biol*. **23**(11): pp. 732-749. doi: 10.1038/s41580-022-00498-3.

Ridley A., Hatano H., Wong-Baeza I., Shaw J., Matthews K. K., Al-Mossawi H., Ladell K., Price D. A., Bowness P., Kollnberger S. (2016) “Activation-Induced Killer Cell Immunoglobulin-like Receptor 3DL2 Binding to HLA-B27 Licenses Pathogenic T Cell Differentiation in Spondyloarthritis”. *Arthritis Rheumatol*. **68**(4): pp. 901-914. doi: 10.1002/art.39515.

Ritchie M. E., Phipson B., Wu D., Hu Y., Law C. W., Shi W., Smyth G. K. (2015) “Limma powers differential expression analyses for RNA-sequencing and microarray studies”. *Nucleic Acids Res*. **43**(7): e47. doi: 10.1093/nar/gkv007.

Robert F., Pelletier J. (2018) “Exploring the Impact of Single-Nucleotide Polymorphisms on Translation”. *Front. Genet*. **9**(1): p. 507. doi: 10.3389/fgene.2018.00507.

Roberts A. R., Appleton L. H., Cortes A., Vecellio M., Lau J., Watts L., Brown M. A., Wordsworth P. (2017) “ERAP1 association with ankylosing spondylitis is attributable to common genotypes rather than rare haplotype combinations”. *Proc. Natl. Acad. Sci. U. S. A*. **114**(3): pp. 558-561. doi: 10.1073/pnas.1618856114.

Roberts C. M., Shahin S. A., Loeza J., Dellinger T. H., Williams J. C., Glackin C. A. (2017) “Disruption of TWIST1-RELA binding by mutation and competitive inhibition to validate the

TWIST1 WR domain as a therapeutic target". *BMC Cancer*. **17**(1): P. 184. doi: 10.1186/s12885-017-3169-9.

Roberts C. M., Tran M. A., Pitruzzello M. C., Wen W., Loeza J., Dellinger T. H., Mor G., Glackin C. A. (2016) "TWIST1 drives cisplatin resistance and cell survival in an ovarian cancer model, via upregulation of GAS6, L1CAM, and Akt signalling". *Sci. Rep*. **6**(1): p. 7652. doi: 10.1038/srep37652.

Robinson P. C., van der Linden S., Khan M. A., Taylor W. J. (2021) "Axial spondyloarthritis: concept, construct, classification and implications for therapy". *Nat. Rev. Rheumatol*. **17**(2): pp. 109-118. doi: 10.1038/s41584-020-00552-4.

Robinson P. C., Costello M. E., Leo P., Bradbury L. A., Hollis K., Cortes A., Lee S., Joo K. B., Shim S. C., Weisman M., Ward M., Zhou X., Garchon H. J., Chiocchia G., Nossent J., Lie B. A., Førre Ø., Tuomilehto J., Laiho K., Jiang L., Liu Y., Wu X., Elewaut D., Burgos-Vargas R., Gensler L. S., Stebbings S., Haroon N., Mulero J., Fernandez-Sueiro J. L., Gonzalez-Gay M. A., Lopez-Larrea C., Bowness P., Gafney K., Gaston J. S., Gladman D. D., Rahman P., Maksymowych W. P., Xu H., van der Horst-Bruinsma I. E., Chou C. T., Valle-Oñate R., Romero-Sánchez M. C., Hansen I. M., Pimentel-Santos F. M., Inman R. D., Martin J., Breban M., Evans D., Reveille J. D., Kim T. H., Wordsworth B. P., Brown M. A. (2015) "ERAP2 is associated with ankylosing spondylitis in HLA-B27-positive and HLA-B27-negative patients". *Ann. Rheum. Dis*. **74**(8): pp. 1627-1629. doi: 10.1136/annrheumdis-2015-207416.

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". *Commun. Biol*. **6**(1): p. 363. doi: 10.1038/s42003-023-04747-9.

Rosine N., Fogel O., Koturan S., Rogge L., Bianchi E., Miceli-Richard C. (2023) "T cells in the pathogenesis of axial spondyloarthritis". *Joint Bone Spine*. **90**(6): p. 105619. doi: 10.1016/j.jbspin.2023.105619.

Roychoudhuri R., Hirahara K., Mousavi K., Clever D., Klebanoff C. A., Bonelli M., Sciumè G., Zare H., Vahedi G., Dema B., Yu Z., Liu H., Takahashi H., Rao M., Muranski P., Crompton J. G., Punkosdy G., Bedognetti D., Wang E., Hoffmann V., Rivera J., Marincola F. M., Nakamura A., Sartorelli V., Kanno Y., Gattinoni L., Muto A., Igarashi K., O'Shea J. J., Restifo N. P. (2013) "BACH2 represses effector programs to stabilize T(reg)-mediated immune homeostasis". *Nature*. **498**(7455): pp. 506-510. doi: 10.1038/nature12199.

Rubin S. J. S., Bai L., Haileselassie Y., Garay G., Yun C., Becker L., Streett S. E., Sinha S. R., Habtezion A. (2019) "Mass cytometry reveals systemic and local immune signatures that distinguish inflammatory bowel diseases". *Nat. Commun*. **10**(1): p. 2686. doi: 10.1038/s41467-019-10387-7.

Rudwaleit M., van der Heijde D., Landewé R., Akkoc N., Brandt J., Chou C. T., Dougados M., Huang F., Gu J., Kirazli Y., Van den Bosch F., Olivieri I., Roussou E., Scarpato S., Sørensen I. J., Valle-Oñate R., Weber U., Wei J., Sieper J. (2011) "The Assessment of SpondyloArthritis International Society classification criteria for peripheral spondyloarthritis and for spondyloarthritis in general". *Ann. Rheum. Dis*. **70**(1): pp. 25-31. doi: 10.1136/ard.2009.108233.

Rudwaleit M., van der Heijde D., Landewé R., Listing J., Akkoc N., Brandt J., Braun J., Chou C. T., Collantes-Estevez E., Dougados M., Huang F., Gu J., Khan M. A., Kirazli Y., Maksymowych

W. P., Mielants H., Sørensen I. J., Ozgocmen S., Roussou E., Valle-Oñate R., Weber U., Wei J., Sieper J. (2009) "The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection". *Ann. Rheum. Dis*. **68**(6): pp. 777-783. doi: 10.1136/ard.2009.108233.

Rudwaleit M., Baeten D. (2006) "Ankylosing spondylitis and bowel disease". *Best Pract. Res. Clin. Rheumatol*. **20**(3): p. 451-471. doi: 10.1016/j.berh.2006.03.010.

Rudwaleit M., van der Heijde D., Khan M, A., Braun J., Sieper J. (2004) "How to diagnose axial spondyloarthritis early". *Ann. Rheum. Dis*. **63**(5): pp. 535-543. doi: 10.1136/ard.2003.011247.

Russ B. E., Tsyganov K., Quon S., Yu B., Li J., Lee J. K. C., Olshansky M., He Z., Harrison P. F., Barugahare A., See M., Nussing S., Morey A. E., Udupa V. A., Bennett T. J., Kallies A., Murre C., Collas P., Powell D., Goldrath A. W., Turner S. J. (2023) "Active maintenance of $CD8^+$ T cell naïvety through regulation of global genome architecture". Cell Rep. **42**(10): p. 113301. doi: 10.1016/j.celrep.2023.113301.

Ryan N. H., Helliwell P. (1998) "A Critical Review of Ankylosing Spondylitis". *Crit. Rev. Phys. Rehabi. Med.* **10**(3): pp. 265-301. doi: 10.1615/CritRevPhysRehabilMed.v10.i3.40.

S ----------

Saadey A. A., Yousif A., Osborne N., Shahinfar R., Chen Y. L., Laster B., Rajeev M., Bauman P., Webb A., Ghoneim H. E. (2022) "Rebalancing TGFβ1/BMP signals in exhausted T cells unlocks responsiveness to immune checkpoint blockade therapy". *Nat. Immunol*. **24**(2): pp. 280-294. doi: 10.1038/s41590-022-01384-y.

Salas A., Hernandez-Rocha C., Duijvestein M., Faubion W., McGovern D., Vermeire S., Vetrano S., Vande Casteele N. (2020) "JAK-STAT pathway targeting for the treatment of inflammatory bowel disease". *Nat. Rev. Gastroenterol. Hepatol*. **17**(6): pp. 323-337. doi: 10.1038/s41575-020-0273-0.

Salim T., Sershen C. L., May E. E. (2016) "Investigating the Role of TNF-α and IFN-γ Activation on the Dynamics of iNOS Gene Expression in LPS Stimulated Macrophages". *PLoS One*. **11**(6): e0153289. doi: 10.1371/journal.pone.0153289.

Sampaio-Barros P. D., Bertolo M. B., Kraemer M. H., Neto J. F., Samara A. M. (2001) "Primary ankylosing spondylitis: patterns of disease in a Brazilian population of 147 patients". *J. Rheumatol*. **28**(3): pp. 560-565.

Sampaio-Barros P. D., Bortoluzzo A. B., Conde R. A., Costallat L. T., Samara A. M., Bértolo M. B. (2010) "Undifferentiated spondyloarthritis: a longterm followup". *J. Rheumatol*. **37**(6): pp. 1195-1199. doi: 10.3899/jrheum.090625.

Sanyal A., Lajoie B. R., Jain G., Dekker J. (20212) "The long-range interaction landscape of gene promoters". *Nature*. **489**(7414): pp. 109-113. doi: 10.1038/nature11279.

Sartorelli V., Lauberth S. M. (2020) "Enhancer RNAs are an important regulatory layer of the epigenome". *Nat. Struct. Mol. Biol*. **27**(6): pp. 521-528. doi: 10.1038/s41594-020-0446-0.

Sassa A., Kanemaru Y., Kamoshita N., Honma M., Yasui M. (2016) "Mutagenic consequences of cytosine alterations site-specifically embedded in the human genome". *Genes Environ*. **38**(1): p. 17. doi: 10.1186/s41021-016-0045-9.

Schaid D. J., Chen W., Larson N. B. (2018) “From genome-wide associations to candidate causal variants by statistical fine-mapping”. *Nat. Rev. Genet*. **19**(8): pp. 491-504. doi: 10.1038/s41576-018-0016-z.

Shi Y., Chen Z., Zhao Z., Yu Y., Fan H., Xu X., Bu X., Gu J. (2019) “IL-21 Induces an Imbalance of Th17/Treg Cells in Moderate-to-Severe Plaque Psoriasis Patients”. *Front. Immunol*. 10(1): p. 1865. doi: 10.3389/fimmu.2019.01865.

Schick S., Rendeiro A. F., Runggatscher K., Ringler A., Boidol B., Hinkel M., Májek P., Vulliard L., Penz T., Parapatics K., Schmidl C., Menche J., Boehmelt G., Petronczki M., Müller A. C., Bock C., Kubicek S. (2019) “Systematic characterization of BAF mutations provides insights into intracomplex synthetic lethalities in human cancers”. *Nat. Genet.* **51**(9): pp. 1399-1410. doi: 10.1038/s41588-019-0477-9.

Schittenhelm R. B., Dudek N. L., Croft N. P., Ramarathinam S. H., Purcell A. W. (2014) “A comprehensive analysis of constitutive naturally processed and presented HLA-C*04:01 (Cw4)-specific peptides”. *Tissue Antigens*. **83**(3): pp. 174-179. doi: 10.1111/tan.12282.

Schlosstein L., Terasaki P. I., Bluestone R., Pearson C. M. (1973) “High association of an HL-A antigen, W27, with ankylosing spondylitis”. *N. Engl. J. Med*. **288**(14): pp. 704-6. doi: 10.1056/NEJM197304052881403.

Schmidl C., Rendeiro A. F., Sheffield N. C., Bock C. (2015) “ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors”. *Nat. Methods*. **12**(10): pp. 963-965. doi: 10.1038/nmeth.3542.

Schmidt D., Wilson M. D., Spyrou C., Brown G. D., Hadfield J., Odom D. T. (2009) “ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions”. *Methods*. **48**(3): pp. 240-8. doi: 10.1016/j.ymeth.2009.03.001.

Schmidt R., Steinhart Z., Layeghi M., Freimer J. W., Bueno R., Nguyen V. Q., Blaeschke F., Ye C. J., Marson A. (2022) “CRISPR activation and interference screens decode stimulation responses in primary human T cells”. *Science*. **2375**(6580): e. eabj4008. doi: 10.1126/science.abj4008.

Schuster S. L., Arora S., Wladyka C. L., Itagi P., Corey L., Young D., Stackhouse B. L., Kollath L., Wu Q. V., Corey E., True L. D., Ha G., Paddison P. J., Hsieh A. C. (2023) “Multi-level functional genomics reveals molecular and cellular oncogenicity of patient-based 3' untranslated region mutations”. *Cell Rep*. **42**(8): p. 112840. doi: 10.1016/j.celrep.2023.112840.

Schuster S. L., Hsieh A. C. (2019) “The Untranslated Regions of mRNAs in Cancer”. *Trends Cancer*. **5**(4): pp. 245-262. doi: 10.1016/j.trecan.2019.02.011.

Seo W., Taniuchi I. (2020) “The Roles of RUNX Family Proteins in Development of Immune Cells”. *Mol. Cells*. **43**(2): p. 107-113. doi: 10.14348/molcells.2019.0291.

Sepriano A., Ramiro S., van der Heijde D., Landewé R. (2021) “Biological DMARDs and disease modification in axial spondyloarthritis: a review through the lens of causal inference”. *RMD Open*. **7**(2): e001654. doi: 10.1136/rmdopen-2021-001654.

Servant N., Varoquaux N., Heard E., Barillot E., Vert J. (2018) “Effective normalization for copy number variation in Hi-C data”. *BMC Bioinformatics* **19**(1): p. 313. doi: 10.1186/s12859-018-2256-5.

Sidwell T., Liao Y., Garnham A. L., Vasanthakumar A., Gloury R., Blume J., Teh P. P., Chisanga D., Thelemann C., de Labastida Rivera F., Engwerda C. R., Corcoran L., Kometani K., Kurosaki T., Smyth G. K., Shi W., Kallies A. (2020) "Attenuation of TCR-induced transcription by Bach2 controls regulatory T cell differentiation and homeostasis". *Nat. Commun*. **11**(1): p. 252. doi: 10.1038/s41467-019-14112-2.

Silberstein M., Nesbit N., Cai J., Lee P. H. (2021) "Pathway analysis for genome-wide genetic variation data: Analytic principles, latest developments, and new opportunities". *J. Genet. Genomics*. **48**(3): pp. 173-183. doi: 10.1016/j.jgg.2021.01.007.

Simone D., Penkava F., Ridley A., Sansom S., Al-Mossawi M. H., Bowness P. (2021) "Single cell analysis of spondyloarthritis regulatory T cells identifies distinct synovial gene expression patterns and clonal fates". *Commun. Biol*. **4**(1): p. 1395. doi: 10.1038/s42003-021-02931-3.

Sharif K., Bridgewood C., Dubash S., McGonagle D. (2020) "Intestinal and enthesis innate immunity in early axial spondyloarthropathy". *Rheumatology (Oxford)*. **59**(4): p. 90. doi: 10.1093/rheumatology/keaa731.

Sharma A., Osato N., Liu H., Asthana S., Dakal T. C., Ambrosini G., Bucher P., Schmitt I., Wüllner U. (2019) "Common genetic variants associated with Parkinson's disease display widespread signature of epigenetic plasticity". *Sci. Rep*. **9**(1): p. 18464. doi: 10.1038/s41598-019-54865-w.

Sharma Y., Miladi M., Dukare S., Boulay K., Caudron-Herger M., Groß M., Backofen R., Diederichs S. (2019) "A pan-cancer analysis of synonymous mutations". *Nat. Commun*. **10**(1): p. 2569. doi: 10.1038/s41467-019-10489-2.

Shen H., Goodall J. C., Hill Gaston J. S. (2009) "Frequency and phenotype of peripheral blood Th17 cells in ankylosing spondylitis and rheumatoid arthritis". *Arthritis Rheum*. **60**(6): pp. 1647-1656. doi: 10.1002/art.24568.

Shen J., Yang L., You K., Chen T., Su Z., Cui Z., Wang M., Zhang W., Liu B., Zhou K., Lu H. (2022) "Indole-3-Acetic Acid Alters Intestinal Microbiota and Alleviates Ankylosing Spondylitis in Mice". *Front. Immunol*. **13**(1): p. 762580. doi: 10.3389/fimmu.2022.762580.

Sherlock J. P., Joyce-Shaikh B., Turner S. P., Chao C. C., Sathe M., Grein J., Gorman D. M., Bowman E. P., McClanahan T. K., Yearley J. H., Eberl G., Buckley C. D., Kastelein R. A., Pierce R. H., Laface D. M., Cua D. J. (2012) "IL-23 induces spondyloarthropathy by acting on ROR-γt+ CD3+CD4-CD8- entheseal resident T cells". *Nat. Med*. **18**(7): pp. 1069-1076. doi: 10.1038/nm.2817.

Sherlock J. P., Cua D. J. (2021) "Interleukin-23 in perspective". *Rheumatology (Oxford)*. **60**(1): p. 1-3. doi: 10.1093/rheumatology/keab461.

Shin B., Zhou W., Wang J., Gao F., Rothenberg E. V. (2023) "Runx factors launch T cell and innate lymphoid programs via direct and gene network-based mechanisms". *Nat. Immunol*. **24**(9): pp. 1458-1472. doi: 10.1038/s41590-023-01585-z.

Shinozaki S., Chang K., Sakai M., Shimizu N., Yamada M., Tanaka T., Nakazawa H., Ichinose F., Yamada Y., Ishigami A., Ito H., Ouchi Y., Starr M. E., Saito H., Shimokado K., Stamler J. S., Kaneki M. (2014) "Inflammatory stimuli induce inhibitory S-nitrosylation of the deacetylase SIRT1 to increase acetylation and activation of p53 and p65". *Sci. Signal*. **7**(351): e106. doi: 10.1126/scisignal.2005375.

Slatkin M. (2016) “Linkage disequilibrium--understanding the evolutionary past and mapping the medical future”. *Nat. Rev. Genet*. **9**(6): pp. 477-485. doi: 10.1038/nrg2361.

Smith J. A. (2018) “Regulation of Cytokine Production by the Unfolded Protein Response; Implications for Infection and Autoimmunity”. *Front. Immunol*. 9(1): p. 422. doi: 10.3389/fimmu.2018.00422.

Sonar S. A., Lal G. (2019) “The iNOS Activity During an Immune Response Controls the CNS Pathology in Experimental Autoimmune Encephalomyelitis”. *Front. Immunol*. **10**(1): p. 710. doi: 10.3389/fimmu.2019.00710.

Sondermann N. C., Faßbender S., Hartung F., Hätälä A. M., Rolfes K. M., Vogel C. F. A., Haarmann-Stemmann T. (2022) “Functions of the aryl hydrocarbon receptor (AHR) beyond the canonical AHR/ARNT signaling pathway”. *Biochem. Pharmacol*. **208**(1): p. 115371. doi: 10.1016/j.bcp.2022.115371.

Song C., Yang X., Lei Y., Zhang Z., Smith W., Yan J., Kong L. (2019) “Evaluation of efficacy on RANKL induced osteoclast from RAW264.7 cells”. *J. Cell Physiol*. **234**(7): pp. 11969-11975. doi: 10.1002/jcp.27852.

Soskic B., Cano-Gamez E., Smyth D. J., Rowan W. C., Nakic N., Esparza-Gordillo J., Bossini-Castillo L., Tough D. F., Larminie C. G. C., Bronson P. G., Willé D., Trynka G. (2019) “Chromatin activity at GWAS loci identifies T cell states driving complex immune diseases”. *Nat. Genet*. **51**(10): pp. 1486-1493. doi: 10.1038/s41588-019-0493-9.

Soukarieh O., Meguerditchian C., Proust C., Aïssi D., Eyries M., Goyenvalle A., Trégouët D. A. (2022) “Common and Rare 5'UTR Variants Altering Upstream Open Reading Frames in Cardiovascular Genomics”. *Front. Cardiovasc. Med*. **9**(1): p. 841032. doi: 10.3389/fcvm.2022.841032.

Spain S. L., Barrett J. C. (2015) “Strategies for fine-mapping complex traits”. *Hum. Mol. Genet*. **24**(1): pp. 111-119. doi: 10.1093/hmg/ddv260.

Spoorenberg A., van der Heijde D., de Klerk E., Dougados M., de Vlam K., Mielants H., van der Tempel H., van der Linden S. (1999) “Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis”. *J. Rheumatol*. **26**(4): pp. 980-984.

Stankey C. T., Bourges C., Haag L. M., Turner-Stokes T., Piedade A. P., Palmer-Jones C., Papa I., Silva Dos Santos M., Zhang Q., Cameron A. J., Legrini A., Zhang T., Wood C. S., New F. N., Randzavola L. O., Speidel L., Brown A. C., Hall A., Saffioti F., Parkes E. C., Edwards W., Direskeneli H., Grayson P. C., Jiang L., Merkel P. A., Saruhan-Direskeneli G., Sawalha A. H., Tombetti E., Quaglia A., Thorburn D., Knight J. C., Rochford A. P., Murray C. D., Divakar P., Green M., Nye E., MacRae J. I., Jamieson N. B., Skoglund P., Cader M. Z., Wallace C., Thomas D. C., Lee J. C. (2024) “A disease-associated gene desert directs macrophage inflammation through ETS2”. *Nature*. **630**(8016): pp. 447-456. doi: 10.1038/s41586-024-07501-1.

Sternes P. R., Brett L., Phipps J., Ciccia F., Kenna T., de Guzman E., Zimmermann K., Morrison M., Holtmann G., Klingberg E., Mauro D., McIvor C., Forsblad-d'Elia H., Brown M. A. (2022) “Distinctive gut microbiomes of ankylosing spondylitis and inflammatory bowel disease patients suggest differing roles in pathogenesis and correlate with disease activity”. *Arthritis Res. Ther*. **24**(1): p. 163. doi: 10.1186/s13075-022-02853-3.

Stone M. A., Pomeroy E., Keat A., Sengupta R., Hickey S., Dieppe P., Gooberman-Hill R., Mogg R., Richardson J., Inman R. D. (2008) "Assessment of the impact of flares in ankylosing spondylitis disease activity using the Flare Illustration". *Rheumatology (Oxford)*. **47**(8): pp. 1213-1218. doi: 10.1093/rheumatology/ken176.

Stuart P. E., Nair R. P., Tsoi L. C., Tejasvi T., Das S., Kang H. M., Ellinghaus E., Chandran V., Callis-Duffin K., Ike R., Li Y., Wen X., Enerbäck C., Gudjonsson J. E., Kõks S., Kingo K., Esko T., Mrowietz U., Reis A., Wichmann H. E., Gieger C., Hoffmann P., Nöthen M. M., Winkelmann J., Kunz M., Moreta E. G., Mease P. J., Ritchlin C. T., Bowcock A. M., Krueger G. G., Lim H. W., Weidinger S., Weichenthal M., Voorhees J. J., Rahman P., Gregersen P. K., Franke A., Gladman D. D., Abecasis G. R., Elder J. T. (2015) "Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture". *Am. J. Hum. Genet*. **97**(6): pp. 816-836. doi: 10.1016/j.ajhg.2015.10.019.

Su Q. Y., Zheng J. W., Yang J. Y., Zhang T. Y., Song S., Zhao R., Di J. K., Zhang S. X., Wang C. H., Gao H. Y. (2022) "Levels of Peripheral Th17 Cells and Th17-Related Cytokines in Patients with Ankylosing Spondylitis: A Meta-analysis". *Adv. Ther*. **39**(10): pp. 4423-4439. doi: 10.1007/s12325-022-02240-z.

Subbarayal B., Chauhan S. K., Di Zazzo A., Dana R. (2016) "IL-17 Augments B Cell Activation in Ocular Surface Autoimmunity". *J. Immunol*. **197**(9): pp. 3464-3470. doi: 10.4049/jimmunol.1502641.

Suchanek O., Ferdinand J. R., Tuong Z. K., Wijeyesinghe S., Chandra A., Clauder A. K., Almeida L. N., Clare S., Harcourt K., Ward C. J., Bashford-Rogers R., Lawley T., Manz R. A., Okkenhaug K., Masopust D., Clatworthy M. R. (2023) "Tissue-resident B cells orchestrate macrophage polarisation and function". *Nat. Commun*. **14**(1): p. 7081. doi: 10.1038/s41467-023-42625-4.

Suhl J. A., Muddashetty R. S., Anderson B. R., Ifrim M. F., Visootsak J., Bassell G. J., Warren S. T. (2015) "A 3' untranslated region variant in FMR1 eliminates neuronal activity-dependent translation of FMRP by disrupting binding of the RNA-binding protein HuR". Proc. Natl. Acad. Sci. U. S. A. 112(47): pp. 6553-6561. doi: 10.1073/pnas.1514260112.

Sun T., Wang M., Liang W., Gao P., Liu Q., Yan X. (2023) "Revealing mechanism of Methazolamide for treatment of ankylosing spondylitis based on network pharmacology and GSEA". *Sci Rep*. **13**(1): p.15370. doi: 10.1038/s41598-023-42721-x.

Sun H, Yu G. (2019) "New insights into the pathogenicity of non-synonymous variants through multi-level analysis". *Sci. Rep*. **9**(1): p. 1667. doi: 10.1038/s41598-018-38189-9.

Sungalee S., Liu Y., Lambuta R. A., Katanayeva N., Donaldson Collier M., Tavernari D., Roulland S., Ciriello G., Oricchio E. (2021) "Histone acetylation dynamics modulates chromatin conformation and allele-specific interactions at oncogenic loci." *Nat. Genet*. **53**(5): pp. 650-662. doi: 10.1038/s41588-021-00842-x.

Swindell W. R., Sarkar M. K., Stuart P. E., Voorhees J. J., Elder J. T., Johnston A., Gudjonsson J. E. (2015) "Psoriasis drug development and GWAS interpretation through in silico analysis of transcription factor binding sites". *Clin. Transl. Med*. **4**(1): p. 13. doi: 10.1186/s40169-015-0054-5.

T ----------

Tagore M., McAndrew M. J., Gjidoda A., Floer M. (2015) “The Lineage-Specific Transcription Factor PU.1 Prevents Polycomb-Mediated Heterochromatin Formation at Macrophage-Specific Genes”. *Mol. Cell. Biol*. **35**(15): pp. 2610-2625. doi: 10.1128/MCB.00027-15.

Tan K., Wong K. H. (2019) “RNA polymerase II ChIP-seq-a powerful and highly affordable method for studying fungal genomics and physiology”. *Biophys. Rev*. **11**(1): pp. 79-82. doi: 10.1007/s12551-018-00497-9.

Tang T., Cheng X., Truong B., Sun L., Yang X., Wang H. (2021) “Molecular basis and therapeutic implications of CD40/CD40L immune checkpoint”. *Pharmacol. Ther*. **219**(1): p. 107709. doi: 10.1016/j.pharmthera.2020.107709.

Tang T., Scambler T. E., Smallie T., Cunliffe H. E., Ross E. A., Rosner D. R., O'Neil J. D., Clark A. R. (2017) “Macrophage responses to lipopolysaccharide are modulated by a feedback loop involving prostaglandin E_2, dual specificity phosphatase 1 and tristetraprolin”. *Sci. Rep*. **7**(1): p. 4350. doi: 10.1038/s41598-017-04100-1.

Tang Y., Yang K., Liu Q., Ma Y., Zhu H., Tang K., Geng C., Xie J., Zhuo D., Wu W., Jin L., Xiao W., Wang J., Zhu Q., Liu J. (2023) “Preosteoclast plays a pathogenic role in syndesmophyte formation of ankylosing spondylitis through the secreted PDGFB - GRB2/ERK/RUNX2 pathway”. *Arthritis Res. Ther*. **25**(1): p. 194. doi: 10.1186/s13075-023-03142-3.

Taube J. H., Sphyris N., Johnson K. S., Reisenauer K. N., Nesbit T. A., Joseph R., Vijay G. V., Sarkar T. R., Bhangre N. A., Song J. J., Chang J. T., Lee M. G., Soundararajan R., Mani S. A. (2017) “The H3K27me3-demethylase KDM6A is suppressed in breast cancer stem-like cells, and enables the resolution of bivalency during the mesenchymal-epithelial transition”. *Oncotarget*. **8**(39): pp. 65548-65565. doi: 10.18632/oncotarget.19214.

Taubmann J., Krishnacoumar B., Böhm C., Faas M., Müller D. I. H., Adam S., Stoll C., Böttcher M., Mougiakakos D., Sonnewald U., Hofmann J., Schett G., Krönke G., Scholtysek C. (2020) “Metabolic reprogramming of osteoclasts represents a therapeutic target during the treatment of osteoporosis”. *Sci. Rep*. **10**(1): p. 21020. doi: 10.1038/s41598-020-77892-4.

Taurog J. D., Chhabra A., Colbert R. A. (2016) “Ankylosing Spondylitis and Axial Spondyloarthritis”. *N. Engl. J. Med*. **374**(26): pp. 2563-74. doi: 10.1056/NEJMra1406182.

Taurog J. D., Richardson J. A., Croft J. T., Simmons W. A., Zhou M., Fernández-Sueiro J. L., Balish E., Hammer R. E. (1994) “The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats”. *J. Exp. Med*. **180**(6): pp. 2359-2564. doi: 10.1084/jem.180.6.2359.

Taylor S., Eyre S., Hinks A., Thomson W., Yarwood A., Martin P., Duffus K., McGovern A. (2017) “Identifying distal interactions between RUNX1 and JIA associated SNP by Chromosome Conformation Capture” *Rheumatology* **56**(1): p. 390. doi: 10.1093/rheumatology/kex390.013.

Tedeschi V., Paldino G., Alba J., Molteni E., Paladini F., Scrivo R., Congia M., Cauli A., Caccavale R., Paroli M., Di Franco M., Tuosto L., Sorrentino R., D'Abramo M., Fiorillo M. T. (2023) “ERAP1 and ERAP2 Haplotypes Influence Suboptimal HLA-B*27:05-Restricted Anti-Viral CD8+ T Cell Responses Cross-Reactive to Self-Epitopes”. *Int. J. Mol. Sci*. **24**(17): p. 13335. doi: 10.3390/ijms241713335.

Telenius J. M., Downes D. J., Sergeant A., Oudelaar M., McGowan S., Kerry J., Hanssen L., Schwessinger R., Eisjbouts C. Q., Davies J. O. J., Taylor S., Hughes J. R. (2020) "CaptureCompendium: a comprehensive toolkit for 3C analysis". *BioRxiv*.

Tewhey R., Kotliar D., Park D. S., Liu B., Winnicki S., Reilly S. K., Andersen K. G., Mikkelsen T. S., Lander E. S., Schaffner S. F., Sabeti P. C. (2016) "Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay". Cell. **165**(6): pp. 1519-1529. doi: 10.1016/j.cell.2016.04.027.

Tolu S., Rezvani A., Hindioglu N., Calkin Korkmaz M. (2018) "Etanercept-induced Crohn's disease in ankylosing spondylitis: a case report and review of the literature". *Rheumatol. Int*. 38(11): pp. 2157-2162. doi: 10.1007/s00296-018-4165-3.

Tomlinson M. J. 4th, Pitsillides A., Pickin R., Mika M., Keene K. L., Hou X., Mychaleckyj J., Chen W. M., Concannon P., Onengut-Gumuscu S. (2014) "Fine mapping and functional studies of risk variants for type 1 diabetes at chromosome 16p13.13". *Diabetes*. **63**(12): pp. 4360-4368. doi: 10.2337/db13-1785.

Thompson S. D., Marion M. C., Sudman M., Ryan M., Tsoras M., Howard T. D., Barnes M. G., Ramos P. S., Thomson W., Hinks A., Haas J. P., Prahalad S., Bohnsack J. F., Wise C. A., Punaro M., Rosé C. D., Pajewski N. M., Spigarelli M., Keddache M., Wagner M., Langefeld C. D., Glass D. N. (2012) "Genome-wide association analysis of juvenile idiopathic arthritis identifies a new susceptibility locus at chromosomal region 3q13". *Arthritis Rheum*. **64**(8): pp. 2781-2791. doi: 10.1002/art.34429.

Thorisson G. A., Smith A. V., Krishnan L., Stein L. D. (2005) "The International HapMap Project Web site". *Genome Res*. **15**(11): pp. 1592-1593. doi: 10.1101/gr.4413105.

Tournadre A., Pereira B., Lhoste A., Dubost J. J., Ristori J. M., Claudepierre P., Dougados M., Soubrier M. (2013) "Differences between women and men with recent-onset axial spondyloarthritis: results from a prospective multicenter French cohort". *Arthritis Care Res*. **65**(9): pp. 1482-1489. doi: 10.1002/acr.22001.

Touzart A., Boissel N., Belhocine M., Smith C., Graux C., Latiri M., Lhermitte L., Mathieu E. L., Huguet F., Lamant L., Ferrier P., Ifrah N., Macintyre E., Dombret H., Asnafi V., Spicuglia S. (2020) "Low level CpG island promoter methylation predicts a poor outcome in adult T-cell acute lymphoblastic leukemia". *Haematologica*. **105**(6): pp. 1575-1581. doi: 10.3324/haematol.2019.223677.

Tremblay M., Tremblay C. S., Herblot S., Aplan P. D., Hébert J., Perreault C., Hoang T. (2010) "Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes". *Genes Dev*. **24**(11): pp. 1093-1105. doi: 10.1101/gad.1897910.

Tserel L., Kolde R., Rebane A., Kisand K., Org T., Peterson H., Vilo J., Peterson P. "Genome-wide promoter analysis of histone modifications in human monocyte-derived antigen presenting cells". *BMC Genomics*. **11**(1): p. 642. doi: 10.1186/1471-2164-11-642.

Tsukamoto H., Takeuchi S., Kubota K., Kobayashi Y., Kozakai S., Ukai I., Shichiku A., Okubo M., Numasaki M., Kanemitsu Y., Matsumoto Y., Nochi T., Watanabe K., Aso H., Tomioka Y. (2018) "Lipopolysaccharide (LPS)-binding protein stimulates CD14-dependent Toll-like receptor 4 internalization and LPS-induced TBK1-IKKϵ-IRF3 axis activation". *J. Biol. Chem*. **293**(26): pp. 10186-10201. doi: 10.1074/jbc.M117.796631.

Tu P. H., Liu Z. H., Yeap M. C., Liu Y. T., Li Y. C., Huang Y. C., Lin T. M., Chen C. C. (2022) "Spinal cord injury and spinal fracture in patients with ankylosing spondylitis". *BMC Emerg. Med*. **22**(1): p. 73. doi: 10.1186/s12873-022-00635-3.

V ----------

van de Bunt M., Cortes A.; IGAS Consortium; Brown M. A., Morris A. P., McCarthy M. I. (2015) "Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci". *PLoS Genet*. **11**(9):e1005535. doi: 10.1371/journal.pgen.1005535.

van der Heijde D., Braun J., Deodhar A., Baraliakos X., Landewé R., Richards H. B., Porter B., Readie A. (2019) "Modified stoke ankylosing spondylitis spinal score as an outcome measure to assess the impact of treatment on structural progression in ankylosing spondylitis". *Rheumatology*. **58**(3): pp. 388-400. doi: 10.1093/rheumatology/key128.

van de Peppel J., Kettelarij N., van Bakel H., Kockelkorn T. T., van Leenen D., Holstege F. C. (2005) "Mediator expression profiling epistasis reveals a signal transduction pathway with antagonistic submodules and highly specific downstream targets". *Mol. Cell*. **19**(4): pp. 511-522. doi: 10.1016/j.molcel.2005.06.033.

VanLiere J. M., Rosenberg N. A. (2008) "Mathematical properties of the r2 measure of linkage disequilibrium". *Theor. Popul. Biol*. **74**(1): pp. 130-137. doi: 10.1016/j.tpb.2008.05.006.

Vihinen M. (2022) "When a Synonymous Variant Is Nonsynonymous". *Genes (Basel)*. **13**(8): p. 1485. doi: 10.3390/genes13081485.

Vecellio M., Chen L., Cohen C. J., Cortes A., Li Y., Bonham S., Selmi C., Brown M. A., Fischer R., Knight J. C., Wordsworth B. P. (2020) "Functional Genomic Analysis of a RUNX3 Polymorphism Associated With Ankylosing Spondylitis". *Arthritis Rheumatol*. **73**(6): pp. 980-990. doi: 10.1002/art.41628.

Vecellio M., Cortes A., Roberts A. R., Ellis J., Cohen C. J., Knight J. C., Brown M. A., Bowness P., Wordsworth B. P. (2018) "Evidence for a second ankylosing spondylitis-associated *RUNX3* regulatory polymorphism". *RMD Open*. **4**(1): e000628. doi: 10.1136/rmdopen-2017-000628.

Vedak P., Kroshinsky D., St John J., Xavier R. J., Yajnik V., Ananthakrishnan A. N. (20160 "Genetic basis of TNF-α antagonist associated psoriasis in inflammatory bowel diseases: a genotype-phenotype analysis". *Aliment Pharmacol. Ther*. **43**(6): pp. 697-704. doi: 10.1111/apt.13542.

Venema W. J., Hiddingh S., van Loosdregt J., Bowes J., Balliu B., de Boer J. H., Ossewaarde-van Norel J., Thompson S. D., Langefeld C. D., de Ligt A., van der Veken L. T., Krijger P. H. L., de Laat W., Kuiper J. J. W. "A cis-regulatory element regulates ERAP2 expression through autoimmune disease risk SNPs". *Cell Genom*. **4**(1): p. 100460. doi: 10.1016/j.xgen.2023.100460.

Vergara-Lope A., Jabalameli M. R., Horscroft C., Ennis S., Collins A., Pengelly R. J. (2019) "Linkage disequilibrium maps for European and African populations constructed from whole genome sequence data". *Sci. Data*. **6**(1): p. 208. doi: 10.1038/s41597-019-0227-y.

Verlaan D. J., Berlivet S., Hunninghake G. M., Madore A. M., Larivière M., Moussette S., Grundberg E., Kwan T., Ouimet M., Ge B., Hoberman R., Swiatek M., Dias J., Lam K. C., Koka V., Harmsen E., Soto-Quiros M., Avila L., Celedón J. C., Weiss S. T., Dewar K., Sinnett D.,

Laprise C., Raby B. A., Pastinen T., Naumova A. K. (2009) "Allele-specific chromatin remodeling in the ZPBP2/GSDMB/ORMDL3 locus associated with the risk of asthma and autoimmune disease". *Am. J. Hum. Genet*. **85**(3): pp. 377-93. doi: 10.1016/j.ajhg.2009.08.007.

Verlouw J. A. M., Clemens E., de Vries J. H., Zolk O., Verkerk A. J. M. H., Am Zehnhoff-Dinnesen A., Medina-Gomez C., Lanvers-Kaminsky C., Rivadeneira F., Langer T., van Meurs J. B. J., van den Heuvel-Eibrink M. M., Uitterlinden A. G., Broer L. (2021) "A comparison of genotyping arrays". *Eur. J. Hum. Genet*. **29**(11): pp. 1611-1624. doi: 10.1038/s41431-021-00917-7.

Verzella D., Pescatore A., Capece D., Vecchiotti D., Ursini M. V., Franzoso G., Alesse E., Zazzeroni F. (2020) "Life, death, and autophagy in cancer: NF-κB turns up everywhere". *Cell Death Dis*. **11**(3): p. 210. doi: 10.1038/s41419-020-2399-y.

Veyrieras J. B., Kudaravalli S., Kim S. Y., Dermitzakis E. T., Gilad Y., Stephens M., Pritchard J. K. (2008) "High-resolution mapping of expression-QTLs yields insight into human gene regulation". *PLoS Genet*. **4**(10): e1000214. doi: 10.1371/journal.pgen.1000214.

Villar J., Cros A., De Juan A., Alaoui L., Bonte P. E., Lau C. M., Tiniakou I., Reizis B., Segura E. (2023) "ETV3 and ETV6 enable monocyte differentiation into dendritic cells by repressing macrophage fate commitment". *Nat. Immunol*. **24**(1): pp. 84-95. doi: 10.1038/s41590-022-01374-0.

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". *Am. J. Hum. Genet*. **101**(1): pp. 5-22. doi: 10.1016/j.ajhg.2017.06.005.

Voruganti A., Bowness P. (2020) "New developments in our understanding of ankylosing spondylitis pathogenesis". *Immunology*. **161**(2): pp. 94-102. doi: 10.1111/imm.13242.

Vu H., Ernst J. (2022) "Universal annotation of the human genome through integration of over a thousand epigenomic datasets". *Genome Biol*. **23**(1) p. 9. doi: 10.1186/s13059-021-02572-z.

W ----------

Wan J., Shan Y., Fan Y., Fan C., Chen S., Sun J., Zhu L., Qin L., Yu M., Lin Z. (2016) "NF-κB inhibition attenuates LPS-induced TLR4 activation in monocyte cells". *Mol. Med. Rep*. **14**(5): pp. 4505-4510. doi: 10.3892/mmr.2016.5825.

Wan L., Li W., Meng Y., Hou Y., Chen M., Xu B. (2022) "Inflammatory Immune-Associated eRNA: Mechanisms, Functions and Therapeutic Prospects". *Front. Immunol*. **13**(1): p. 849451. doi: 10.3389/fimmu.2022.849451.

Wan P. Z., Xu T. H., Tian B. Y., Guo G. Y., Li X. L., Yao L. (2022) "Machine Learning Reveals Ets2 as a Novel Target for Membranous Nephropathy Treatment and Its Role in Immune Infiltration". *Front. Med. (Lausanne)*. **9**(1): p. 813329. doi: 10.3389/fmed.2022.813329.

Wang R., Dillon C. P., Shi L. Z., Milasta S., Carter R., Finkelstein D., McCormick L. L., Fitzgerald P., Chi H., Munger J., Green D. R. (2011) "The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation". *Immunity.* **35**(6): pp. 871-82. doi: 10.1016/j.immuni.2011.09.021.

Wang C. M., Tsai S. C., Lin J. C., Wu Y. J., Wu J., Chen J. Y. (2019) "Association of Genetic Variants of *RANK*, *RANKL*, and *OPG* with Ankylosing Spondylitis Clinical Features in Taiwanese". *Mediators Inflamm*. **2019**(1): p. 8029863. doi: 10.1155/2019/8029863.

Wang Y., Chaffee T. S., LaRue R. S., Huggins D. N., Witschen P. M., Ibrahim A. M., Nelson A. C., Machado H. L., Schwertfeger K. L. (2020) "Tissue-resident macrophages promote extracellular matrix homeostasis in the mammary gland stroma of nulliparous mice". Elife. **9**(1): e57438. doi: 10.7554/eLife.57438.

Wang H., Fan Z., Shliaha P. V., Miele M., Hendrickson R. C., Jiang X., Helin K. (2023) "H3K4me3 regulates RNA polymerase II promoter-proximal pause-release". *Nature*. **615**(7951): pp. 339-348. doi: 10.1038/s41586-023-05780-8.

Wang Z., Luo S., Zhang Z., Zhou T., Zhang J. (2023) "4D nucleome equation predicts gene expression controlled by long-range enhancer-promoter interaction". *PLoS Comput. Biol*. **19**(12): e1011722. doi: 10.1371/journal.pcbi.1011722.

Ward M. M. (1999) "Health-related quality of life in ankylosing spondylitis: a survey of 175 patients". *Arthritis Care Res*. **12**(4): pp. 247-55.

Ward J. R., Francis S. E., Marsden L., Suddason T., Lord G. M., Dower S. K., Crossman D. C., Sabroe I. (2009) "A central role for monocytes in Toll-like receptor-mediated activation of the vasculature". *Immunology*. **128**(1): pp. 58-68. doi: 10.1111/j.1365-2567.2009.03071.x.

Watts L., Karaderi T., Roberts A., Appleton L., Wordsworth T., Cohen C., Wordsworth P., Vecellio M. (2019) "The severity of ankylosing spondylitis and responses to anti-tumour necrosis factor biologics are not influenced by the tumour necrosis factor receptor polymorphism incriminated in multiple sclerosis". *Genes Immun*. **20**(2): pp. 167-171. doi: 10.1038/s41435-018-0017-0.

Way G. P., Youngstrom D. W., Hankenson K. D., Greene C. S., Grant S. F. (2017) "Implicating candidate genes at GWAS signals by leveraging topologically associating domains". *Eur. J. Hum. Genet*. **25**(11): pp. 1286-1289. doi: 10.1038/ejhg.2017.108.

Wen J. T., Zhang D. H., Fang P. F., Li M. H., Wang R. J., Li S. H. (2017) "Role of Th1/Th2 cytokines in the diagnosis and prognostic evaluation of ankylosing spondylitis". *Genet. Mol. Res*. **16**(1). doi: 10.4238/gmr16019322.

Wen Y., Jing Y., Yang L., Kang D., Jiang P., Li N., Cheng J., Li J., Li X., Peng Z., Sun X., Miller H., Sui Z., Gong Q., Ren B., Yin W., Liu C. (2019) "The regulators of BCR signaling during B cell activation". *Blood Sci.* **1**(2): pp. 119-129. doi: 10.1097/BS9.0000000000000026.

Wilbrink[1] R., Spoorenberg A., Arends S., van der Geest K. S. M., Brouwer E., Bootsma H., Kroese F. G. M., Verstappen G. M. (2021) "$CD27^{-}CD38^{low}CD21^{low}$ B-Cells Are Increased in Axial Spondyloarthritis". *Front. Immunol*. **12**(1): p. 686273. doi: 10.3389/fimmu.2021.686273.

Wilbrink[2] R., Spoorenberg A., Verstappen G. M. P. J., Kroese F. G. M. (2021) "B Cell Involvement in the Pathogenesis of Ankylosing Spondylitis". *Int. J. Mol. Sci*. **22**(24): p. 13325. doi: 10.3390/ijms222413325.

Wilczynski B., Liu Y. H., Yeo Z. X., Furlong E. E. (2012) "Predicting spatial and temporal gene expression using an integrative model of transcription factor occupancy and chromatin state". *PLoS Comput Biol*. **8**(12): p. 1002798. doi: 10.1371/journal.pcbi.1002798.

Wiles E. T., Selker E. U. (2017) “H3K27 methylation: a promiscuous repressive chromatin mark”. *Curr. Opin Genet. Dev*. **43**(1): pp. 31-37. doi: 10.1016/j.gde.2016.11.001.

Wilson A. S., Randall K. L., Pettitt J. A., Ellyard J. I., Blumenthal A., Enders A., Quah B. J., Bopp T., Parish C. R., Brüstle A. (2022) “Neutrophil extracellular traps and their histones promote Th17 cell differentiation directly via TLR2”. *Nat. Commun*. **13**(1): p. 528. doi: 10.1038/s41467-022-28172-4.

White M. J., Yaspan B. L., Veatch O. J., Goddard P., Risse-Adams O. S., Contreras M. G. (2019) “Strategies for Pathway Analysis Using GWAS and WGS Data”. *Curr. Protoc. Hum. Genet*. **100**(1): e79. doi: 10.1002/cphg.79.

Wu J. J., Sun Z. L., Liu S. Y., Chen Z. H., Yuan Z. D., Zou M. L., Teng Y. Y., Li Y. Y., Guo D. Y., Yuan F. L. (2022) “The ASIC3-M-CSF-M2 macrophage-positive feedback loop modulates fibroblast-to-myofibroblast differentiation in skin fibrosis pathogenesis”. *Cell Death Dis*. **13**(6): p. 527. doi: 10.1038/s41419-022-04981-9.

Wu L., Wang R., Zhou Y., Zhao D., Chen F., Wu X., Chen X., Chen S., Li J., Zhu J. (2022) “Natural Killer Cells Infiltration in the Joints Exacerbates Collagen-Induced Arthritis”. *Front. Immunol*. **13**(1): p. 860761. doi: 10.3389/fimmu.2022.860761.

Wu V. H., Yung B. S., Faraji F., Saddawi-Konefka R., Wang Z., Wenzel A. T., Song M. J., Pagadala M. S., Clubb L. M., Chiou J., Sinha S., Matic M., Raimondi F., Hoang T. S., Berdeaux R., Vignali D. A. A., Iglesias-Bartolome R., Carter H., Ruppin E., Mesirov J. P., Gutkind J. S. (2023) “The GPCR-Gα_s-PKA signaling axis promotes T cell dysfunction and cancer immunotherapy failure”. *Nat. Immunol*. **24**(8): pp. 1318-1330. doi: 10.1038/s41590-023-01529-7.

Wu X., Wu J., Li X., Wei Q., Lv Q., Zhang P., Zheng X., Chen Z., Cao S., Tu L., Gu J. (2021) “The Clinical Characteristics of Other HLA-B Types in Chinese Ankylosing Spondylitis Patients”. *Front. Med. (Lausanne)*. 7(1): p. 568790. doi: 10.3389/fmed.2020.568790.

X ----------

Xiao F., Zhang H. Y., Liu Y. J., Zhao D., Shan Y. X., Jiang Y. F. (2013) “Higher frequency of peripheral blood interleukin 21 positive follicular helper T cells in patients with ankylosing spondylitis”. *J. Rheumatol*. **40**(12) pp. :2029-37. doi: 10.3899/jrheum.130125.

Xie Y., Chen X., Deng M., Sun Y., Wang X., Chen J., Yuan C., Hesketh T. (2021) “Causal Linkage Between Inflammatory Bowel Disease and Primary Sclerosing Cholangitis: A Two-Sample Mendelian Randomization Analysis”. *Front. Genet*. **12**(1): p. 649376. doi: 10.3389/fgene.2021.649376.

Xing-Rong W., Sheng-Qian X., Wen L., Shan Q., Fa-Ming P., Jian-Hua X. (2018) “Role of TNFRSF1A and TNFRSF1B polymorphisms in susceptibility, severity, and therapeutic efficacy of etanercept in human leukocyte antigen-B27-positive Chinese Han patients with ankylosing spondylitis”. *Medicine (Baltimore)*. **97**(31): e11677. doi: 10.1097/MD.0000000000011677.

Xu H., Yu H., Liu L., Wu H., Zhang C., Cai W., Hong X., Liu D., Tang D., Dai Y. (2021) “Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Peripheral Mononuclear Cells in Patients With Ankylosing Spondylitis”. *Front. Immunol*. **12**(1): p. 760381. doi: 10.3389/fimmu.2021.760381.

Xu S., Ma Y., Wu M., Zhang X., Yang J., Deng J., Guan S., Gao X., Xu S., Shuai Z., Guan S., Chen L., Pan F. (2020) “Neutrophil lymphocyte ratio in patients with ankylosing spondylitis: A systematic review and meta-analysis”. *Mod. Rheumatol*. **30**(1): pp. 141-148. doi: 10.1080/14397595.2018.1564165.

Xu W., Zhong Q., Lin D., Zuo Y., Dai J., Li G., Cao G. (2021) “CoolBox: a flexible toolkit for visual analysis of genomics data”. *BMC Bioinformatics*. **22**(1): p. 489. doi: 10.1186/s12859-021-04408-w.

Xue Q., Yan Y., Zhang R., Xiong H. (2018) “Regulation of iNOS on Immune Cells and Its Role in Diseases”. *Int. J. Mol. Sci.* **19**(12): p. 3805. doi: 10.3390/ijms19123805.

Y ----------

Yang C., Siebert J. R., Burns R., Gerbec Z. J., Bonacci B., Rymaszewski A., Rau M., Riese M. J., Rao S., Carlson K. S., Routes J. M., Verbsky J. W., Thakar M. S., Malarkannan S. (2019) “Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome”. *Nat. Commun*. **10**(1): p. 3931. doi: 10.1038/s41467-019-11947-7.

Yang J. H., Hansen A. S. (2024) “Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation”. *Nat. Rev. Mol. Cell Biol*. ePUB ahead of print. doi: 10.1038/s41580-024-00710-6.

Yang L., Chen S., Zhao Q., Sun Y., Nie H. (2019) “The Critical Role of Bach2 in Shaping the Balance between $CD4^+$ T Cell Subsets in Immune-Mediated Diseases”. *Mediators Inflamm.* **2019**(1): p. 2609737. doi: 10.1155/2019/2609737.

Yang X., Garner L. I., Zvyagin I. V., Paley M. A., Komech E. A., Jude K. M., Zhao X., Fernandes R. A., Hassman L. M., Paley G. L., Savvides C. S., Brackenridge S., Quastel M. N., Chudakov D. M., Bowness P., Yokoyama W. M., McMichael A. J., Gillespie G. M., Garcia K. C. (2022) “Autoimmunity-associated T cell receptors recognize HLA-B*27-bound peptides”. *Nature.* **612**(7941): pp. 771-777. doi: 10.1038/s41586-022-05501-7.

Yang Y., Ochando J. C., Bromberg J. S., Ding Y. (2007) “Identification of a distant T-bet enhancer responsive to IL-12/Stat4 and IFNgamma/Stat1 signals”. *Blood*. **110**(7): pp. 2494-500. doi: 10.1182/blood-2006-11-058271.

Yao C., Sakata D., Esaki Y., Li Y., Matsuoka T., Kuroiwa K., Sugimoto Y., Narumiya S. (2009) “Prostaglandin E2-EP4 signaling promotes immune inflammation through Th1 cell differentiation and Th17 cell expansion”. *Nat. Med*.**15**(6): pp. 633-640. doi: 10.1038/nm.1968.

Yao C., Lou G., Sun H. W., Zhu Z., Sun Y., Chen Z., Chauss D., Moseman E. A., Cheng J., D'Antonio M. A., Shi W., Shi J., Kometani K., Kurosaki T., Wherry E. J., Afzali B., Gattinoni L., Zhu Y., McGavern D. B., O'Shea J. J., Schwartzberg P. L., Wu T. (2021) “BACH2 enforces the transcriptional and epigenetic programs of stem-like $CD8^+$ T cells”. *Nat. Immunol*. **22**(4): p. 530. doi: 10.1038/s41590-021-00906-4.

Yardımcı G. G., Ozadam H., Sauria M. E. G., Ursu O., Yan K. K., Yang T., Chakraborty A., Kaul A., Lajoie B. R., Song F., Zhan Y., Ay F., Gerstein M., Kundaje A., Li Q., Taylor J., Yue F., Dekker J., Noble W. S. (2019) “Measuring the reproducibility and quality of Hi-C data”. *Genome Biol*. **20**(1): p. 57. doi: 10.1186/s13059-019-1658-7.

Yeo J., Kim J. Y., Park J. K., Shin K., Lee E. Y., Kim T. H., Park J. W. (2024) “Flare prediction after tapering the dose of tumour necrosis factor inhibitors in patients with axial spondyloarthritis: A nationwide cohort study”. *Rheumatology (Oxford)*. **13**(1): e. keae138. doi: 10.1093/rheumatology/keae138.

Ying P., Chen C., Lu Z., Chen S., Zhang M., Cai Y., Zhang F., Huang J., Fan L., Ning C., Li Y., Wang W., Geng H., Liu Y., Tian W., Yang Z., Liu J., Huang C., Yang X., Xu B., Li H., Zhu X., Li N., Li B., Wei Y., Zhu Y., Tian J., Miao X. (2023) “Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk”. *Nat. Commun*. **14**(1): p. 5958. doi: 10.1038/s41467-023-41690-z.

Yu H., Wu H., Zheng F., Zhu C., Yin L., Dai W., Liu D., Tang D., Hong X., Dai Y. (2020) “Gene-regulatory network analysis of ankylosing spondylitis with a single-cell chromatin accessible assay”. *Sci. Rep*. **10**(1): pp. 19411. doi: 10.1038/s41598-020-76574-5.

Yu Z., Hong X., Zhang X., Zheng F., Liu F., Xu H., Zhu C., Cai W., Liu D., Yin L., Hu B., Tang D., Dai Y. (2022) “Global Proteomic Analyses Reveals Abnormal Immune Regulation in Patients With New Onset Ankylosing Spondylitis”. *Front. Immunol*. **13**(1): p. 838891. doi: 10.3389/fimmu.2022.838891.

Z ----------

Zboril E., Yoo H., Chen L., Liu Z. (2021) “Dynamic Interactions of Transcription Factors and Enhancer Reprogramming in Cancer Progression”. *Front. Oncol*. 11(1): p. 753051. doi: 10.3389/fonc.2021.753051.

Zhang Y., Liu T., Meyer C.A., Eeckhoute J., Johnson D. S., Bernstein B. E., Nusbaum C., Myers R. M., Brown M., Li W., Liu X. S. (2008) “Model-based Analysis of ChIP-Seq (MACS)”. *Genome Biol.* **9**(1): e137. doi: 10.1186/gb-2008-9-9-r137.

Zhang Z., Cheng Y., Zhou X., Li Y., Gao J., Han J., Quan C., He S., Lv Y., Hu D., Zhu K., Sun L., Yang S., Zhang X. (2011) “Polymorphisms at 16p13 are associated with systemic lupus erythematosus in the Chinese population”. *Med. Gen.* **48**(1)**:** pp. 69-72. doi: 10.1136/jmg.2010.077859.

Zhang J. R., Pang D. D., Dai S. M. (2019) “Non-steroidal Anti-inflammatory Drugs Are Unlikely to Inhibit Radiographic Progression of Ankylosing Spondylitis: A Systematic Review”. *Front. Med*. **6**(1): p. 214. doi: 10.3389/fmed.2019.00214.

Zhang M., Zhou L., Xu Y., Yang M., Xu Y., Komaniecki G. P., Kosciuk T., Chen X., Lu X., Zou X., Linder M. E., Lin H. (2020) “A STAT3 palmitoylation cycle promotes T_H17 differentiation and colitis”. *Nature*. **586**(7829): pp. 434-439. doi: 10.1038/s41586-020-2799-2.

Zhang G., Li Y., Wei G. (2023) “Multi-omic analysis reveals dynamic changes of three-dimensional chromatin architecture during T cell differentiation”. *Commun. Biol*. **6**(1): p. 773. doi: 10.1038/s42003-023-05141-1.

Zhao C., Tan Y. C., Wong W. C., Sem X., Zhang H., Han H., Ong S. M., Wong K. L., Yeap W. H., Sze S. K., Kourilsky P., Wong S. C. (2010) “The CD14(+/low)CD16(+) monocyte subset is more susceptible to spontaneous and oxidant-induced apoptosis than the CD14(+)CD16(-) subset”. *Cell Death Dis*. **1**(11): p. 95. doi: 10.1038/cddis.2010.69.

Zhao D., Zhang L., Zhang M., Xia B., Lv J., Gao X., Wang G., Meng Q., Yi Y., Zhu S., Tomoiaga A. S., Lee M. G., Cooke J. P., Cao Q., Chen K. (2020) "Broad genic repression domains signify enhanced silencing of oncogenes". *Nat. Commun*. **11**(1): p. 5560. doi: 10.1038/s41467-020-18913-8.

Zhao Q., Ma J., Xie F., Wang Y., Zhang Y., Li H., Sun Y., Wang L., Guo M., Han K. (2021) "Recent Advances in Predicting Protein S-Nitrosylation Sites". *Biomed. Res. Int*. 2021(1): p. 5542224. doi: 10.1155/2021/5542224.

Zhao W., Xu Y., Wang Y., Gao D., King J., Xu Y., Liang F. (2021) "Investigating crosstalk between H3K27 acetylation and H3K4 trimethylation in CRISPR/dCas-based epigenome editing and gene activation". *Sci. Rep*. 11(1): p. 15912. doi: 10.1038/s41598-021-95398-5.

Zhao M., Feng D., Hu L., Liu L., Wu J., Hu Z., Long H., Kuang Q., Ouyang L., Lu Q. (2023) "3D genome alterations in T cells associated with disease activity of systemic lupus erythematosus". *Ann. Rheum. Dis*. **82**(2): pp. 226-234. doi: 10.1136/ard-2022-222653.

Zhao Y., Chen J., Andreatta M., Feng B., Xie Y. Q., Wenes M., Wang Y., Gao M., Hu X., Romero P., Carmona S., Sun J., Guo Y., Tang L. (2024) "IL-10-expressing CAR T cells resist dysfunction and mediate durable clearance of solid tumors and metastases". *Nat. Biotechnol*. Epub ahead of print. doi: 10.1038/s41587-023-02060-8.

Zheng H., Xie W. (2019) "The role of 3D genome organization in development and cell differentiation". *Nat. Rev. Mol. Cell Biol*. 2019 Sep;20(9):535-550. doi: 10.1038/s41580-019-0132-4.

Zheng Q., Liu W., Huang Y., Gao Z., Wu Y., Wang X., Cai M., He Y., Chen S., Wang B., Liu L., Chen S., Huang H., Zheng L., Kang R., Zeng X., Chen J., Chen H., Chen J., Li Z., Shi G. (2021) "Predictive Value of Active Sacroiliitis in MRI for Flare Among Chinese Patients with Axial Spondyloarthritis in Remission". *Rheumatol. Ther*. **8**(1): pp. 411-424. doi: 10.1007/s40744-021-00279-y.

Zhong J., Scholz T., Yau A. C. Y., Guerard S., Hüffmeier U., Burkhardt H., Holmdahl R. (2018) "Mannan-induced Nos2 in macrophages enhances IL-17-driven psoriatic arthritis by innate lymphocytes". *Sci. Adv*. **4**(5): e9864. doi: 10.1126/sciadv.aas9864.

Zhu J., Jankovic D., Oler A. J., Wei G., Sharma S., Hu G., Guo L., Yagi R., Yamane H., Punkosdy G., Feigenbaum L., Zhao K., Paul W. E. (2012) "The transcription factor T-bet is induced by multiple pathways and prevents an endogenous Th2 cell program during Th1 cell responses". *Immunity*. **37**(4): pp. 660-673. doi: 10.1016/j.immuni.2012.09.007.

Zhu W., He X., Cheng K., Zhang L., Chen D., Wang X., Qiu G., Cao X., Weng X. (2019) "Ankylosing spondylitis: etiology, pathogenesis, and treatments". *Bone Res*. **7**(1): p. 22. doi: 10.1038/s41413-019-0057-8.

Zuvich R.L., Bush W. S., McCauley J. L., Beecham A. H., De Jager P. L., International Multiple Sclerosis Genetics Consortium, Ivinson A. J., Compston A., Hafler D. A., Hauser S. L., Sawcer S. J., Pericak-Vance M. A., Barcellos L. F., Mortlock D. P., Haines J. L. (2011) "Interrogating the complex role of chromosome 16p13.13 in multiple sclerosis susceptibility: independent genetic signals in the CIITA-CLEC16A-SOCS1 gene complex". *Hum. Mol. Genet*. **20**(17): pp. 3517-3524. doi: 10.1093/hmg/ddr250.

Zwick D., Vo M. T., Shim Y. J., Reijonen H., Do J. S. (2024) “BACH2: The Future of Induced T-Regulatory Cell Therapies”. *Cells*. **13**(11): p. 891. doi: 10.3390/cells13110891.