Functional Characterization of Ankylosing Spondylitis Genetic Risk Variants

Julie A. Osgood



Merton College

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Michaelmas Term, 2019

Abstract

Functional Characterization of Ankylosing Spondylitis Genetic Risk Variants

Julie Osgood, Merton College, Michaelmas Term 2019

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

Ankylosing spondylitis (AS) is a chronic inflammatory arthritis characterized by inflammation and osteoproliferation in the lower spine and sacroiliac joints. This dysregulated bone growth leads to the fusion of affected joints, resulting in chronic pain and disability. It is considered a complex genetic disorder, with a heritability of around 90%, and has over 100 genetic loci associated through genome-wide association studies (GWAS). An inherent challenge with GWAS is that many of the associated genetic variants lie within the non-coding genome and exhibit linkage disequilibrium (LD). Therefore determining their function, and the genes and pathways they modulate, and thus identify the disease causal variant and its mechanism of action is a major roadblock to progress in this field. Functional genomic approaches are emerging to interpret the non-coding genome and its relationship with disease. This thesis aimed to utilize functional genomic approaches to advance the understanding of the genetic aetiology of AS.

CRISPR/Cas9 modulated genome editing was utilized for the functional characterization of two putative enhancers within the AS GWAS associated locus, *TNFRSF1A*. The generation of large deletions using CRISPR/Cas9 presented unanticipated challenges, including the issue of mosaic cellular clones post CRISPR/Cas9 editing. The methodological issues and how these can be addressed for generating homozygous CRISPR/Cas9 edited cells, harbouring deletions of interest, are discussed. Despite unforeseen complications with CRISPR/Cas9 editing, the two putative enhancers were deleted in induced pluripotent stem cells (iPSCs), which were then differentiated in macrophages. The edited cells then underwent assays to map regions of open chromatin, histone modifications, physical chromatin interactions, and gene expression changes in edited and unedited cells. One of the predicted regulatory regions was found to be an enhancer for *TNFRSF1A*, and the other was discovered not to be enhancer, but a potential repressor for *NCAPD2*, located approximately 150kb away.

In addition to functional experiments in a model system, this thesis also examined the genome-wide epigenetic differences between AS patients and healthy controls in primary human immune cells. Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq) was used to define regions of chromatin that were differentially open in AS patients. Chromatin Immunoprecipitation Sequencing (ChIP-Seq) was performed to identify differential regions of H3K27ac, denoting active enhancers, and H3K4me3, denoting active promoters. This data was combined with gene expression data to define putative regulatory regions associated with AS and annotate GWAS variants for future functional studies.

Acknowledgements

First and foremost, I would like to thank my three supervisors, Julian Knight, Ben Davies, and Sally Cowley. I could not have completed this thesis without their guidance, advice, and encouragement. I would like to express my gratitude to Julian for giving me the opportunity to work on this amazing project and for welcoming me into his lab. Also to Ben and Sally for agreeing to co-supervise and providing me with technical training and knowledge that made this project possible.

Thank-you to every member of the Knight group that I have had the pleasure of working with, and to those from Ben and Sally's labs who have helped along the way. All of you have helped me get through the DPhil, providing both technical and moral support. I would especially like to thank Andrew Brown, Alicia Lledó Lara, Giuseppe Scozzafava, Katie Burnham, Olga Mielczarek, Félicie Costantino, Gabriele Migliorini, Hai Fang, Chwen Tay, Anna Sanniti, and Bogdan Knezevic. All of you have been instrumental in putting together this thesis and in making my DPhil so enjoyable.

I would like to thank all my friends and family back in Canada who supported me from afar. Your encouragement has helped me get through all the trials and tribulations of doing a DPhil. Thank-you as well to all the friends I have made here in Oxford. My DPhil has been such a fun and an enriching experience and I am so glad that I got to share it with all of you.

Lastly I want to say a special thank-you to the University of Oxford Dancesport Team. The Dancesport Team has been such integral part of my DPhil experience and I want to express my gratitude to all the team members that I have competed and trained with over the past four years. Dancing was my break from failed experiments and could always make me smile. I will always remember the competitions, early morning practices, long coach trips, bad poetry, catchy music, and most importantly, the opportunity to wear sparkly dresses.

Anything's possible if you've got enough nerve.

J.K. Rowling

Declarations

I declare that, unless otherwise stated, all work presented in this thesis is my own. Several aspects of the study relied upon collaboration where part of the work was conducted with or by others.

Karyotyping of K562 and K562-Cas9 cells was performed using a multiplex fluorescence in situ hybridization (M-FISH) by Daniela Moralli in Cath Green's core cytogenetics facility at the WHG. The FISH experiment used to determine the copy number of the regions containing the *RUNX1* enhancer and the *RPP30* gene on the parental K562 cell line and the edited Clones was also performed by Daniela Moralli.

All of the iPSC work was done at the Dunn School of Pathology under the supervision of Sally Cowley and she assisted with the general care of the cells. Sally also monitored the cells at different stages of differentiation to ensure they were maturing correctly. The Illumina BeadChip array on my edited iPSCs was carried out by Jane Vowles. RNA-Seq library preparation and sequencing on the iPSC derived macrophages was performed by Core Genomics at the Wellcome Centre for Human Genetics. RNA-Seq data analysis was carried out with the assistance of Katie Burnham. The Capture-C protocol was carried out alongside Olga Mielczarek and the sequencing data was analyzed by Olga Mielczarek and Gabriele Migliorini. The data processing for ATAC-Seq and ChIP-Seq on iPSC derived macrophages was carried out by Gabriele Migliorini and Andrew Brown. FACS on iPSC derived macrophages was carried out with the assistance of Chwen Tay and Elena Navarro-Guerrero at the Target Discovery Institute (TDI). The RNA-Seq iPSC derived macrophage dataset used for comparison in Chapter 4 was generated by Chwen Tay and Harindra Amarasinghe at the WHG.

The profiling of the AS patient and healthy control cohorts was part of a large collaborative effort that started before I joined the group. I assisted in processing and sequencing of 21 samples for ChIP-Seq and ATAC-Seq. Patient recruitment, sample processing, and data analysis of the larger cohort has been carried out at different times by Anna Sanniti, Alicia Lledó Lara, Katharine Plant, Evelyn Lau, Giuseppe Scozzafava, Olga Mielczarek, Gabriele Migliorini, Andrew Brown, and Félicie Costantino. The AS ATAC-Seq and ChIP-Seq data presented in this thesis was analyzed with the support of Alicia Lledó Lara, Gabriele Migliorini, Andrew Brown, and Félicie Costantino. Fine mapping of AS GWAS loci was carried out by Anna Sanniti and Félicie Costantino. The Capture-C data used in Chapter 5 was generated by Olga Mielczarek. The list of differentially expressed genes between patients and controls used in Chapter 5 was generated as part of a broader study involving Anna Sanniti, Gabriele Migliorini, and Félicie Costantino using data from the AS cohort.

Submitted Abstracts

The use of CRISPR/Cas9 in an iPSC derived macrophage model to investigate ankylosing spondylitis genetic risk loci

Poster: Immunogenomics of Disease: Accelerating to Patient Benefit (2019) - Winner of the Best Poster Prize

Julie Osgood, Andrew Brown, Katie Burnham, Olga Mielczarek, Gabriele Migliorini, Ben Davies, Sally Cowley, Julian C. Knight

Associated Publications

A genetics-led approach defines the drug target landscape of 30 immune-related traits
Nature Genetics (2019)
Hai Fang, The ULTRA-DD Consortium, Hans De Wolf, Bogdan Knezevic, Katie L. Burnham, **Julie Osgood**, Anna Sanniti, Alicia Lledó Lara, Silva Kasela, Stephane De Cesco, Jörg K. Wegner, Lahiru Handunnetthi, Fiona E. McCann, Liye Chen, Takuya Sekine, Paul E. Brennan, Brian D. Marsden, David Damerell, Chris A. O'Callaghan, Chas Bountra, Paul Bowness, Yvonne Sundström, Lili Milani, Louise Berg, Hinrich W. Göhlmann, Pieter J. Peeters, Benjamin P. Fairfax, Michael Sundström & Julian C. Knight

Translating GWAS in rheumatic disease: approaches to establishing mechanism and function for genetic associations with ankylosing spondylitis
Briefings in Functional Genomics (2018)
Julie A Osgood, Julian C Knight

Contents

Abstract i

Acknowledgements ii

Dedication iii

Declarations iv

Submitted Abstracts v

Associated Publications vi

Contents vii

List of Figures xiii

List of Tables xv

Abbreviations xvi

1 Introduction 1

1.1 Goal of the Thesis . . . 1
1.2 Ankylosing Spondylitis . . . 2
1.2.1 Clinical Presentations . . . 3
1.2.2 Diagnosis and Biomarkers . . . 5
1.3 The Pathophysiology of Ankylosing Spondylitis . . . 8
1.3.1 Ankylsoing Spondylitis and HLA-B27 . . . 8
1.3.2 Gut Inflammation and Ankylosing Spondylitis . . . 10
1.3.3 Disease Relevant Cell Types . . . 11
1.3.4 Therapeutics . . . 15
1.4 Gene Regulation and Disease . . . 18
1.4.1 Genome-Wide Association Studies . . . 18
1.4.2 Fine Mapping of GWAS Variants . . . 19
1.4.3 Understanding the Role of Non-Coding Variants . . . 20
1.5 Genetics of Ankylosing Spondylitis . . . 26
1.5.1 Ankylosing Spondylitis GWAS . . . 26
1.5.2 Notable GWAS Findings . . . 28
1.5.3 The Limitations of GWAS . . . 31
1.6 Functional Genomic Approaches . . . 32
1.6.1 Defining Chromatin Accessibility . . . 34
1.6.2 Defining Chromatin Interactions . . . 35
1.6.3 Defining Histone Modifications . . . 37
1.6.4 Assaying Gene Expression and Mapping eQTLs . . . 38
1.7 Functional Validation of Genetic Variants . . . 40
1.7.1 Genome Editing . . . 41
1.7.2 CRISPR/Cas9 Mediated Genome Editing . . . 42

1.8 Specific Aims and Objectives 47

2 Materials and Methods **50**
2.1 Cell Culture 50
2.1.1 K562 Cell Culture 50
2.1.2 iPS Cell Culture 51
2.2 Macrophage Differentiation 52
2.2.1 iPSC Derived Macrophage Culture 52
2.2.2 LPS and IFNγ Stimulation 53
2.2.3 Validation of Macrophage Phenotype 55
2.3 CRISPR/Cas9 Mediated Genome Editing 55
2.3.1 Guide RNA Design 55
2.3.2 Transfection 56
2.3.3 Clone Picking 59
2.3.4 Clone Screening 60
2.4 Expression Analysis 66
2.4.1 RNA Extraction 66
2.4.2 Quantitative PCR 66
2.4.3 RNA-Seq 68
2.5 Functional Genomic Assays 68
2.5.1 Capture-C 68
2.5.2 ATAC-Seq 70
2.5.3 ChIP-Seq 72
2.6 Additional Analysis 74
2.6.1 Fine Mapping 74
2.6.2 Pathway Analysis 75

3 Determining the Suitability of Utilizing CRISPR/Cas9 to Investigate AS Genetic Risk Loci in Disease Relevant Cell Types **76**
3.1 Introduction 76
3.1.1 Prioritization and Functional Validation of GWAS Variants 76
3.1.2 CRISPR/Cas9 Delivery Methods 77
3.1.3 Model Systems for Ankylosing Spondylitis 80
3.1.4 Aims 83
3.2 Results: CRISPR/Cas9 in K562 Cells 84
3.2.1 K562 Cell Karyotype 84
3.2.2 Regulatory Element Selection 87
3.2.3 CRISPR/Cas9 Approach and Validation 88
3.2.4 Bulk Cell Population Expression Analysis 90
3.2.5 Clonal Isolation and Expansion 91
3.2.6 Screening Clonal Populations for Homozygotes 92
3.2.7 Copy Number Analysis of Edited K562 Cell Clones 93
3.2.8 Expression Analysis of Edited K562 Cell Clones 96
3.3 Results: CRISPR/Cas9 in iPSCs 98
3.3.1 Priority Index for Variant Prioritization 98
3.3.2 Further Prioritization of *TNFRSF1A* 101
3.3.3 Experimental Design 105
3.3.4 CRISPR/Cas9 Validation 106

3.3.5 Mosaicism . . . 111
3.3.6 Generating Homozygous Cell Lines . . . 112
3.3.7 Macrophage Differentiation . . . 118
3.4 Discussion . . . 120
3.4.1 Variant Prioritization for CRISPR/Cas9 Work . . . 121
3.4.2 CRISPR/Cas9 in K562 Cells . . . 121
3.4.3 Mosaicism in CRISPR/Cas9 . . . 124
3.4.4 Future Experiments . . . 127
3.4.5 High Throughput CRISPR/Cas9 Assays . . . 129
3.5 Conclusions . . . 130

4 Functional Characterization of Putative Enhancers Within *TNFRSF1A* Using iPSC Derived Macrophages 131
4.1 Introduction . . . 131
4.1.1 TNFRSF1A and Inflammation . . . 131
4.1.2 TNFRSF1A and Disease . . . 132
4.1.3 Characterizing Putative Enhancers . . . 135
4.1.4 Aims . . . 138
4.2 Results . . . 139
4.2.1 Samples . . . 139
4.2.2 Expression Analysis of CRISPR/Cas9 Edited Cells . . . 141
4.2.3 Interrogating Physical Chromatin Interactions . . . 151
4.2.4 Assaying Regions of Chromatin Accessibility . . . 156
4.2.5 Assaying Histone Modifications . . . 166
4.2.6 Examination of the *TNFRSF1A* Locus in Primary Cells . . . 180
4.3 Discussion . . . 182
4.3.1 iPSC Derived Macrophages as a Model . . . 182
4.3.2 Functional Genomics to Investigate Enhancers . . . 183
4.3.3 Data Integration to Characterize Putative Enhancers . . . 189
4.3.4 Validation of the *TNFRSF1A* Locus . . . 191
4.4 Conclusions . . . 192

5 Defining Putative Regulatory Elements in Ankylosing Spondylitis 193
5.1 Introduction . . . 193
5.1.1 Epigenetic Regulation of Gene Expression . . . 193
5.1.2 Assaying Chromatin Accessibility Using Omni-ATAC-Seq . . . 194
5.1.3 Assaying Histone Modifications Using ChIPmentation . . . 195
5.1.4 Integrating -OMICS Data to Define Regulatory Regions . . . 196
5.1.5 Aims . . . 197
5.2 Results . . . 198
5.2.1 Patients and Healthy Controls Recruited . . . 198
5.2.2 The Chromatin Accessibility Landscape in AS . . . 200
5.2.3 The Histone Modification Landscape in AS . . . 222
5.2.4 Data Integration . . . 247
5.3 Discussion . . . 251
5.3.1 Important Cell Types in AS Pathology . . . 251
5.3.2 Cell Type Specific Pathway Enrichment . . . 251
5.3.3 Uncovering Putative Enhancers in AS Patients . . . 254

5.3.4 Limitations of Patient Vs Control Differential Analysis . . . 257
5.4 Conclusions . . . 257

6 General Discussion **259**
6.1 Model Systems for AS Biology . . . 259
6.2 CRISPR/Cas9 as a Functional Genomics Tool . . . 261
6.2.1 CRISPR/Cas9 to Generate Disease Associated Alleles . . . 262
6.2.2 CRISPR/Cas9 to Correct Disease Associated Alleles . . . 264
6.2.3 Challenges of CRISPR/Cas9 Mediated Genome Editing . . . 265
6.3 Integration of -Omics Data in AS Patients . . . 266
6.4 Future Work . . . 267
6.4.1 CRISPR/Cas9 Experiments . . . 267
6.4.2 Additional Cell Types . . . 268
6.4.3 Continued Patient Recruitment . . . 268
6.5 Conclusions . . . 269

Appendices **270**

A Supplementary Data from Chapter 3 **270**

B Supplementary Data from Chapter 4 **284**

C Supplementary Data from Chapter 5 **305**

Bibliography **320**

List of Figures

1.1 AS with comorbidities and major complications. 4
1.2 Cells of the immune system. 12
1.3 AS GWAS associated loci. 27
1.4 Approaches used to functionally characterize GWAS variants. . . . 33
1.5 Overview of CISPR/Cas9 technique. 44

2.1 Images of iPSCs undergoing macrophage differentiation. 53
2.2 Overview of the CRISPR and macrophage differentiation process. 54
2.3 iPSC cell colonies growing on MEFs. 60
2.4 Schematic of screening primers. 64

3.1 K562 Cell Karyotype. 85
3.2 Schematic of CRISPR/Cas9 deletion and screening approach. . . . 88
3.3 Agarose gel indicating deletion success in K562 cells. 89
3.4 Expression of *GATA1* and *RUNX1* in K562-Cas9 cells. 91
3.5 Schematic of CRISPR/Cas9 zygosity screening approach. 92
3.6 FISH of K562 cells showing *RUNX1* enhancer and *RPP30* CNV. . . 95
3.7 *RUNX1* enhancer copy number in K562 edited clones. 96
3.8 *RUNX1* expression in K562 edited clones. 97
3.9 Overview of the Pi pipeline. 99
3.10 Prioritization interaction plot for *TNFRSF1A*. 101
3.11 *TNFRSF1A* fine mapped SNPs. 103
3.12 Overview of PROMO predicted transcription factor binding. . . . 104
3.13 Schematic of the TNFRSF1A gene. 106
3.14 Schematic showing how mosacism can occur. 112
3.15 FACS profile of iPSC derived macrophages. 120

4.1 Overview of the TNFRSF1A/TNFR1 signaling pathway. 133
4.2 Schematic showing the *TNFRSF1A* gene. 140
4.3 PCA of primary immune cells and iPSC derived macrophages. . . 142
4.4 Illustration of the TAD containing *TNFRSF1A*. 143
4.5 *TNFRSF1A* expression. 144
4.6 *NCAPD2* expression. 146
4.7 PCA of stimulated iPSC derived macrophages. 148
4.8 Correlation between primary and iPSC derived macrophages. . . . 149
4.9 Volcano plot of differentially expressed genes. 150
4.10 Capture-C data for control regions. 153
4.11 $CD14^{+}$ Capture-C data for the TAD containing *TNFRSF1A*. 156
4.12 Total reads following Omni-ATAC on iPSC derived macrophages. . 157
4.13 Omni-ATAC read percentages from iPSC derived macrophages. . . 158
4.14 Differentially open chromatin peaks in edited and unedited clones. 160
4.15 Differential open chromatin over the Intronic Enhancer. 161
4.16 chromatin accessibility over the Intergenic Enhancer region. . . . 162
4.17 Intronic Enhancer Deletion ATAC peaks (unstimulated vs LPS). . . 164
4.18 Intergenic Enhancer Deletion ATAC peaks (unstimulated vs LPS). 165

4.19 H3K27ac ChIPm reads from iPSC derived macrophages. 167
4.20 H3K27ac ChIPm read percentages from iPSC derived macrophages. 168
4.21 Regions of H3K27ac in iPSC derived macrophages. 170
4.22 Differential H3K27ac peak over the Intronic Enhancer. 171
4.23 H3K27ac peak over the Intergenic Enhancer Deletion region. . . . 173
4.24 Differential H3K27ac between stimulated states (Intronic). 176
4.25 Differential H3K27ac peaks between stimulation states (Intergenic). 179
4.26 H3K27ac peak in unstimulated and LPS stimulated clones. 180
4.27 Chromatin accessibility and H3K27ac at the *TNFRSF1A* locus. . . 181

5.1 Total reads following ATAC on AS patients and controls. 201
5.2 Read percentages following ATAC on AS patients and controls. . . 201
5.3 ATAC TSS fold enrichment for AS patient samples and controls. . 202
5.4 Comparison between fresh and frozen ATAC samples. 203
5.5 Principal component analysis of ATAC batches by cell type. 204
5.6 Corrected PCA of ATAC batches by cell type. 205
5.7 ATAC-Seq peak cell type distribution. 207
5.8 ATAC-Seq differential analysis between cell types. 207
5.9 Pathway enrichment analysis on genes overlapping ATAC peaks. . 209
5.10 $CD8^+$ cell type specific peak within *ATXN2*. 211
5.11 Capture-C and ATAC-Seq data for the *IL6R* AS GWAS locus. . . . 212
5.12 Region of differential open chromatin in $CD8^+$ cells. 221
5.13 Total reads following ChIPm on AS patients and controls. 223
5.14 Read percentages following ChIPm on AS patients and controls. . 224
5.15 ChIPm TSS fold enrichment for AS patient samples and controls. . 225
5.16 H3K27ac ChIPm peak cell type distribution. 226
5.17 H3K27ac ChIPm differential analysis between cell types. 227
5.18 Pathway enrichment analysis on genes overlapping H3K27ac. . . . 229
5.19 $CD14^+$ H3K27ac ChIPm peaks overlapping *IL1R1*. 230
5.20 H3K4me3 ChIPm peak cell type distribution. 236
5.21 H3K4me3 ChIPm differential analysis between cell types. 236
5.22 Pathway analysis on genes overlapping H3K4me3 peaks. 237
5.23 $CD14^+$ H3K4me3 ChIPm peaks overlapping *IL1R1*. 240
5.24 Capture-C, ATAC-Seq, and ChIPm data for the *ETS2* gene. 241
5.25 Region of differential H3K4me3 in $CD8^+$ cells. 246
5.26 *PTCH1* gene in AS patients and controls. 247
5.27 Chromatin accessibility and histone modifications over *THBS1*. . . 250

A.1 Agarose gel for *MYC* and *LMO2* enhancer deletions. 270
A.2 Agarose gel for *RUNX1* enhancer deletion). 271
A.3 Enhancer deletion success in K562 edited cell clones. 272
A.4 Agarose gel indicating zygosity of K562 edited cell clones. 273
A.5 Agarose gel showing deletion region of K562 edited cell clones. . . 274
A.6 Agarose gel showing gRNA test in HEK-293 cells. 274
A.7 Agarose gel indicating Exon Deletion success. 275
A.8 Agarose gel indicating zygosity of the Exon Deletion. 275
A.9 Alternative agarose gel indicating zygosity of the Exon Deletion. . 276
A.10 Agarose gel indicating Intronic Enhancer Deletion success. 276

A.11 Agarose gel indicating zygosity of the Intronic Enhancer Deletion. 277
A.12 Alternative agarose gel for Intronic Enhancer Deletion zygosity. . . 277
A.13 Agarose gel indicating Intergenic Enhancer Deletion success. . . . 278
A.14 Agarose gel indicating zygosity of the Intronic Enhancer Deletion. 278
A.15 Alternative agarose gel for Intergenic Enhancer Deletion zygosity. 279
A.16 Karyogram for Intronic Enhancer CRISPR Control 1. 279
A.17 Karyogram for Intronic Enhancer CRISPR Control 2. 280
A.18 Karyogram for Intronic Enhancer Deletion clone 1. 280
A.19 Karyogram for Intronic Enhancer Deletion clone 2. 281
A.20 Karyogram for Intronic Enhancer Deletion clone 3. 281
A.21 Karyogram for Intergenic Enhancer CRISPR Control 1. 282
A.22 Karyogram for Intergenic Enhancer CRISPR Control 2. 282
A.23 Karyogram for Intergenic Enhancer Deletion clone 1. 283
A.24 Karyogram for Intergenic Enhancer Deletion clone 2. 283

B.1 RNA-Seq mapping efficiency by sample type. 284
B.2 ATAC TSS fold enrichment on iPSC derived macrophages. 291
B.3 Intergenic Enhancer Deletion open chromatin (unstimulated). . . 292
B.4 Intronic Enhancer Deletion open chromatin (LPS). 293
B.5 differential ATAC peak over Intronic Enhancer Deletion (LPS). . . 294
B.6 Intergenic Enhancer Deletion open chromatin (LPS). 295
B.7 ATAC reads over the Intergenic Enhancer Deletion (LPS). 296
B.8 ChIPm TSS fold enrichment on iPSC derived macrophages. 297
B.9 Intergenic Enhancer Deletion differential H3K27ac (unstimulated). 298
B.10 Differential H3K27ac over the Intronic Enhancer (LPS). 300
B.11 Intronic Enhancer Deletion H3K27ac (LPS). 301
B.12 H3K27ac over the Intergenic Enhancer Deletion region (LPS). . . . 303
B.13 Intergenic Enhancer Deletion H3K27ac (LPS). 304

C.1 Chromatin accessibility and histone modifications over *ANO10*. . 318
C.2 Chromatin accessibility and histone modifications over *CABIN1*. . 319

List of Tables

1.1 Diagnostic Criteria 6
1.2 BASDAI Scoring 7
1.3 Diseases associated with a single nucleotide variant change. 23

2.1 K562 CRISPR/Cas9 gRNAs 56
2.2 *TNFRSF1A* CRISPR/Cas9 gRNAs 57
2.3 Screening primers used to examine deletion regions. 62
2.4 *TNFRSF1A* deletion screening primers. 63
2.5 Primers and probes used in ddPCR analysis. 65
2.6 qPCR primers used to examine gene expression. 67
2.7 Capture-C bait regions 69

3.1 Percentage of non-triploid K562 cells. 86
3.2 Percentage of K562 cells containing recurring rearrangements. . . 86
3.3 *RUNX1* enhancer region copy number per cell. 94
3.4 *RPP30* gene copy number per cell. 94
3.5 Output from the preliminary Pi Pipeline. 100
3.6 Expected amplicon sizes following CRISPR/Cas9 deletion events. 107
3.7 iPSC zygosity after CRISPR/Cas9 transfections with a plasmid. . . 108
3.8 Copy number of Exon Deletion edited iPSCs. 109
3.9 Copy number of Intronic Enhancer Deletion edited iPSCs. 109
3.10 Copy number of Intergenic Enhancer Deletion edited iPSCs. . . . 110
3.11 iPSC zygosity after transfections with RNPs. 113
3.12 iPSC colonies selected for sub-cloning. 114
3.13 iPSC zygosity after CRISPR/Cas9 after sub-cloning approach. . . . 114
3.14 Copy number of enhancer deletions in iPSC sub-clones. 115
3.15 Copy number of Exon Deletion in iPSCs after re-transfection. . . . 117
3.16 iPSCs taken forward to macrophage differentiation. 118

4.1 Edited and unedited clones used in each experiment. 140
4.2 Intronic Enhancer differentially expressed genes (unstimulated). . 145
4.3 Intergenic Enhancer differentially expressed genes (unstimulated). 147
4.4 Capture-C interactions in primary $CD14^+$ cells. 155
4.5 Differential Omni-ATAC peaks between edited and unedited clones. 159
4.6 Intergenic Enhancer Deletion open chromatin (unstimulated). . . 162
4.7 Differential H3K27ac peaks between edited and unedited clones. . 169
4.8 Intergenic Enhancer Deletion differential H3K27ac (unstimulated). 172
4.9 Differential H3K27ac between stimulation states (Intronic). 175
4.10 Differential H3K27ac between stimulation states (Intergenic). . . . 178

5.1 Patients Recruited. 199
5.2 Cell type specific ATAC-Seq peaks overlapping AS GWAS loci. . . 213
5.3 Regions of differential open chromatin in $CD8^+$ and $CD14^+$ cells. . 219
5.4 Cell type specific H3K27ac peaks overlapping AS GWAS loci. . . . 231
5.5 Regions of differential H3K27ac histone modifications. 234
5.6 Cell type specific H3K4me3 peaks overlapping AS GWAS loci. . . 242

5.7 Regions of differential H3K4me3 histone modifications. 244
5.8 Differentially expressed genes in CD14$^+$ cells. 249

B.1 Intronic Enhancer differentially expressed genes (IFNγ) 284
B.2 Intergenic Enhancer differentially expressed genes (unstimulated) 285
B.3 Intergenic Enhancer differentially expressed genes (LPS). 287
B.4 Intergenic Enhancer differentially expressed genes (IFNγ) 287
B.5 Capture-C sequenced fragments for CD14$^+$ cells. 289
B.6 Intronic Enhancer Deletion Capture-C sequenced fragments. . . . 289
B.7 Intergenic Enhancer Deletion Capture-C sequenced fragments. . . 290
B.8 Intronic Enhancer Deletion open chromatin (LPS). 290
B.9 Intergenic Enhancer Deletion open chromatin (LPS). 294
B.10 Intronic Enhancer Deletion open chromatin (unstimulated vs LPS). 296
B.11 Intergenic Enhancer Deletion open chromatin (unstimulated vs LPS).297
B.12 Intronic Enhancer Deletion differential H3K27ac (LPS). 299
B.13 Intergenic Enhancer Deletion H3K27ac (LPS).. 302

C.1 Patient Cell Types 305
C.2 Control Cell Types 307
C.3 Top 25 CD4$^+$ cell type specific ATAC peaks. 309
C.4 Top 25 CD8$^+$ cell type specific ATAC peaks. 310
C.5 Top 25 CD14$^+$ cell type specific ATAC peaks. 311
C.6 Top 25 CD4$^+$ cell type specific H3K27ac ChIPm peaks. 312
C.7 Top 25 CD8$^+$ cell type specific H3K27ac ChIPm peaks. 313
C.8 Top 25 CD14$^+$ cell type specific H3K27ac ChIPm peaks. 314
C.9 Top 25 CD4$^+$ cell type specific H3K4me3 ChIPm peaks. 315
C.10 Top 25 CD8$^+$ cell type specific H3K4me3 ChIPm peaks. 316
C.11 Top 25 CD14$^+$ cell type specific H3K4me3 ChIPm peaks. 317

Abbreviations

AS	Ankylosing spondylitis
ATAC-Seq	Assay for transposase-accessible chromatin sequencing
bFGF	Basic fibroblast growth factor
BMP4	Bone morphogenetic protein 4
Cas9	CRISPR associated protein 9
cDNA	Complimentary DNA
ChIP-Seq	Chromatin immunoprecipitation and DNA sequencing
CNV	Copy number variation
CRISPR	Clustered regularly interspaced short palindromic repeats
CRP	C-reactive protein
crRNA	CRISPR RNA
CTCF	CCCTC-binding factor
ddPCR	Droplet digital polymerase chain reaction
EB	Embryoid body
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EOR	ER-overload response
eQTL	Expression quantitative trait loci
ER	Endoplasmic reticulum
eRNA	Enhancer RNA
ERS	Erythrocyte sedimentation rate
FACS	Fluorescence-activated cell sorting
FBS	Foetal Bovine Serum
FDR	False Discovery Rate
FISH	Fluorescence in situ hybridization
gDNA	Genomic DNA
GFP	Green fluorescent protein
gRNA	Guide RNA
GWAS	Genome-wide association study
HDR	Homology directed repair
HLA-B27	Human Leukocyte Antigen B27
IBD	Inflammatory bowel disease
IDT	Integrated DNA technologies
INDELs	Insertions and deletions
IL-3	Interleukin 3
iPSC	Induced pluripotent stem cell
IFNγ	Interferon-gamma
K562-Cas9	K562 cells constitutively expressing Cas9
LPS	Lipopolysaccharide
M-CSF	Macrophage colony-stimulating factor
MEF	Mouse embryonic fibroblast
M-FISH	Multiplex fluorescence in situ hybridization
2-ME	β-mercaptoethanol
MHC	The major histocompatibility complex

MMP3	matrix metalloproteinase-3
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
mtDNA	Mitochondrial DNA
NF-κB	nuclear factor kappa B
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NK cells	Natural killer cells
nr-axSpA	Non-radiographic axial spondyloarthritis
NRF	Non-redundant fraction
NSAIDs	Nonsteroidal anti-inflammatory drugs
SCF	Stem cell factor
SNP	Single nucleotide polymorphism
SpA	Spondyloarthropathies
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RA	Rheumatoid arthritis
RNP	Ribonucleoprotein
TAD	Topologically associated domain
TALENs	Transcription activator-like effector nucleases
TDI	Target Discovery Institute
TLR	Toll like receptor
TNF	Tumour Necrosis Factor-alpha
TNFRSF1A	Tumor Necrosis Factor Receptor Superfamily, Member 1A
tracrRNA	Trans-activating crRNA
TRAPS	Tumor necrosis factor receptor associated periodic syndrome
TSS	Transcription start site
UC	Ulcerative colitis
UPR	Unfolded protein response
VEGF	Vascular endothelial growth factor
WHG	Wellcome Centre for Human Genetics
ZFNs	Zinc-finger nucleases

1

Introduction

Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less. -Marie Curie

1.1 Goal of the Thesis 1
1.2 Ankylosing Spondylitis 2
1.3 The Pathophysiology of Ankylosing Spondylitis 8
1.4 Gene Regulation and Disease 18
1.5 Genetics of Ankylosing Spondylitis 26
1.6 Functional Genomic Approaches 32
1.7 Functional Validation of Genetic Variants 40
1.8 Specific Aims and Objectives 47

1.1 Goal of the Thesis

The overall goal of this thesis is to utilize functional genomic approaches to advance the understanding of the genetic aetiology of ankylosing spondylitis (AS). Genome-wide association studies (GWAS) have identified many genes and genetic variants associated with AS, but the majority reside in the non-coding genome and have not been examined in a biological context. In this thesis, functional genomic techniques are employed to look at the overall epigenetic landscape in AS including histone modifications and chromatin accessibility, as well as changes in gene expression. Genome editing is also utilized to functionally examine specific GWAS associated variants to evaluate their mechanism of action and how they may relate to disease. Insight into the genes

and pathways modulated by GWAS associated variants could lead to a better understanding of AS pathology to improve both diagnostic and therapeutic options. In this chapter, an overview is provided of the current understanding of AS pathology, genetic associations, functional genomic techniques used to assay gene expression and the chromatin landscape, and methods of genome editing that could be used for functional validation of AS risk variants.

1.2 Ankylosing Spondylitis

AS is a heritable, immune-mediated, chronic inflammatory disorder that dates back to ancient Egypt (Feldtkeller et al. 2003), although it was not properly characterized until the 1800s (Taurog et al. 2016). It is characterized by inflammation and osteoproliferation resulting in fusion of the spine and sacroiliac joints, causing chronic pain and loss of mobility. It is associated with significant morbidity and mortality, and patients often suffer from cardiovascular and gastrointestinal complications in addition to pain and disability, highlighting the systemic manifestations of this disease. AS is part of a group of diseases collectively known as spondyloarthropathies (SpA), which include axial SpA and peripheral arthritis such as, reactive arthritis, psoriatic arthritis, and arthritis associated with inflammatory bowel disease (IBD), also known as enteropathic arthritis (Raychaudhuri and Deodhar 2014). The global prevalence of AS is between 0.1-2%, with the highest frequency being in Caucasian populations (Dean et al. 2013). It had been believed that AS affects men two to three times more often than women, but recent studies suggest that this sex difference is not as large as once believed (Taurog et al. 2016). The chronic pain and disability associated with AS can have a detrimental impact on a patient's quality of life and increases their risk of developing depression and anxiety, as well as experiencing severe fatigue and weight loss (Mahmood and Helliwell 2017). The highest incidence of disease is in young adults, with most cases presenting themselves

before the age of 45. This has a huge socioeconomic burden due to healthcare costs, lower productivity in the work place, and in some cases, the inability to work at all (Palla et al. 2012; Mahmood and Helliwell 2017).

1.2.1 Clinical Presentations

As mentioned, SpA is a collection of immune-mediated inflammatory diseases that includes axial SpA, which can be further subdivided into radiographic axial SpA, often referred to as AS, and non-radiographic axial SpA (nr-axSpA) (Taurog et al. 2016; Van Der Heijde et al. 2017a). These classifications are based on the presence or absence of changes in the sarcoliliac joints (Reed et al. 2008; Brown and Bradbury 2017). The sacroiliac joints are the most prominently affected in AS (Mahmood and Helliwell 2017). Patients with nr-axSpA do not have the characteristic X-ray changes seen in AS, but still have axial inflammatory disease and experience pain and disability, with approximately half of them developing AS within 10 years (Brown and Bradbury 2017; Mau et al. 1988). As AS progresses, secondary bone growth can lead to fusion of the axial skeleton (Martín-Mola et al. 2010). In some cases, total spinal fusion can occur leading to the characteristic "bamboo spine". This results in reduced spinal mobility, significant pain, and restricted lung function which can result in respiratory failure (Ryall and Helliwell 1998). Additionally, AS patients also suffer from bone fragility and have an increased risk of osteoporosis and vertebral fractures (Taurog et al. 2016).

In addition to classical symptoms of axial arthritis, AS patients can suffer from a number of different co-morbidities including peripheral arthritis, anterior uveitis, psoriasis, IBD, renal disease, and cardiovascular disease (aortitis, aortic valve disease, conduction disturbances, cardiomyopathy and ischaemic heart disease) (Figure 1.1) (Raychaudhuri and Deodhar 2014; Ryall and Helliwell 1998). For example, approximately 5% of SpA patients develop significant gut

inflammation and IBD (Ciccia et al. 2016). Many of these co-existing diseases are also chronic inflammatory disorders caused by a dysregulated immune response. This overlap of diseases and phenotypes suggest a shared pathophysiology among immune traits.

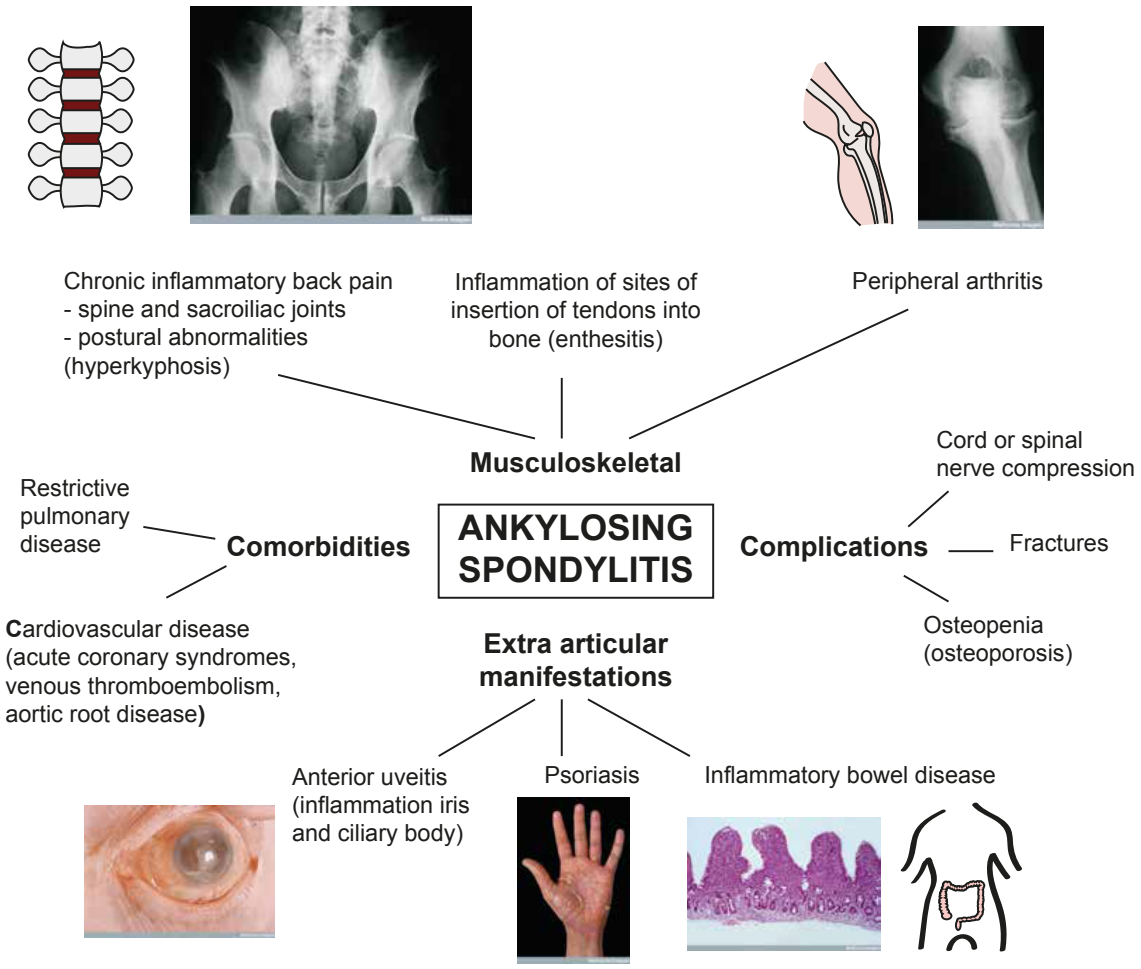


Figure 1.1: Overview of the musculoskeletal and extra-articular manifestations of AS together with comorbidities and major complications.
Figure reproduced from Osgood and Knight 2018.

Though there is controversy in the proportion of males and females affected, there does seem to be a difference in disease manifestation based on sex. A recent study looking at 2044 patients found that women experience an increased prevalence in complications such as peripheral arthritis, dactylitis, and enthesitis, and do not respond as well to anti-Tumour Necrosis Factor-alpha (TNF) treatments (Landi et al. 2017; Rusman et al. 2018). They also found that

men have a lower disease activity and a better quality of life, but have worse spinal mobility and radiographic progression. Moreover, they experience a delay in diagnosis compared to women, though it is believed that men have earlier disease onset. These changes can be explained by immunological, hormonal, and genetic differences between men and women (Rusman et al. 2018). Multiple studies have uncovered differences in how women respond to TNF and IL-17A during an immune response, which could explain the differing success rates of anti-TNF therapy (Gracey et al. 2016; Huang et al. 2012). Studies have also found sex differences in gene expression profiles (Gracey et al. 2016). One notable change is that different loci of the gene, *ANKH*, which is involved in with joint ankylosis, have been associated with men and women with AS and could help explain different radiological progression between the sexes (Tsui et al. 2005).

1.2.2 Diagnosis and Biomarkers

Because AS shares so many features with other immune-mediated diseases, it is incredibly hard to diagnose and predict disease outcome. The diagnosis of AS is based on the modified New York classification criteria (Table 1.1) and requires the presence of X-ray changes in the sacroiliac joints (Van Der Linden et al. 1984). X-rays can be beneficial, but they cannot detect early disease as radiographic changes can take 10 years before they are visible. Magnetic resonance imaging (MRI) scanning has greatly improved AS diagnosis by enabling the diagnosis of the disease at an early stage and can detect less severe cases. It also has enabled the diagnosis of nr-axSpA, which was previously un-characterized, and could help identify those at risk of developing AS (Rudwaleit et al. 2009).

Disease activity is frequently measured by the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) (Garrett et al. 1994). A BASDAI score is given based on self assessment of six questions relating to a patient's pain and inflammation, as seen in Table 1.2 . A patient will answer each question on a

AS Diagnostic Criteria
Clinical Criteria
Lower back pain of greater than 3 months, improved by exercise and not relieved by rest
Limitation of lumbar spine in sagittal and frontal planes
Limitation of chest expansion (relative to normal values corrected for age and sex)
Radiological Criteria
Bilateral grade 2-4 sacroiliitis, or;
Unilateral 3-4 sacroiliitis

Table 1.1: The modified New York classification criteria for AS.
This is the modified New York classification criteria for the diagnosis of AS as presented by Van Der Linden et al. 1984.

scale from 1 to 10, with the responses being added up for a final score indicating disease severity. There are other types of questionnaires in use to assess disease activity and quality of life, but all of them rely on a patient's perception of pain and do not relate to any clinical parameters.

In conjunction with MRIs and disease activity scores, biomarkers are also used in the diagnosis of AS. C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR) are used to assess inflammation levels in patients, as they are acute-phase reactants and increase in response to inflammation. However, these measurements have low sensitivity and specificity and are limited in their ability to detect nr-axSPA (Spoorenberg et al. 1999; Wallis et al. 2013). Research is ongoing for improved biomarkers to assess disease. Recently, matrix metalloproteinase-3 (MMP3), vascular endothelial growth factor (VEGF), and citrullinated vimentin fragments have been found to have a relationship between serum concentration and radiographic spinal progression and could become clinical biomarkers for predicting inflammation and disease progression (Maksymowych et al. 2007; Poddubnyy et al. 2014a; Bay-Jensen et al. 2013).

BASDAI
Please indicate your level of ability with each of the following activities during the past week.
How would you describe the overall level of...
Fatigue/tiredness you have experienced?
AS neck, back or hip pain you have had?
Pain/swelling in joints other than neck, back or hips you have had?
Discomfort you have had from any areas tender to touch or pressure?
Discomfort you have had from the time you wake up?
How long does your morning stiffness last from the time you wake up?
Pain on a ten point scale (1 is none and 10 is the worst)

Table 1.2: The Bath Ankylosing Spondylitis Disease Activity Index (BASDAI).
The questions used to evaluate AS patient disease activity (Garrett et al. 1994).

Levels of adipokine visfatin and the Wnt pathway modulators, sclerostin and dickkopf-1 have been uncovered as potential biomarkers to assess bone growth (Syrbe et al. 2015; Appel et al. 2009; Heiland et al. 2012). IL-6 was also a strong biomarker candidate owing to its role in inflammation, but studies are conflicted over the relationship between IL-6 levels and disease severity (Gratacos et al. 1994; Romero-Sanchez et al. 2011; Sveaas et al. 2015).

Biomarkers identified in related diseases are also under investigation. Calprotectin is a dimer of calcium binding proteins and is increased in the presence of gut inflammation, making it useful as a marker. It can be measured in a patient's serum and stool and is therefore used for monitoring IBD disease activity (Lee et al. 2019). Due to the overlap between IBD and AS, researchers have examined calprotectin as a biomarker for AS. It was found that increased calprotectin levels in a patient's stool correlated to a higher disease activity and better predicted those AS patients at a higher risk of developing IBD (Duran et al. 2016; Klingberg et al. 2016). Researchers are also looking into genetics informed biomarkers. Single nucleotide polymorphisms (SNPs) found within

IL-12B and *IL-6R* have been associated with an increased risk of developing AS and have been suggested as future biomarkers to improve diagnosis and better predict disease progression (Ruan et al. 2018).

1.3 The Pathophysiology of Ankylosing Spondylitis

1.3.1 Ankylsoing Spondylitis and HLA-B27

AS is an immune-mediated inflammatory disorder. One of the crucial components of the immune response is the major histocompatibility complex (MHC), which comprises of MHC molecules that bind and display antigens for cell surface presentation to T-cells. MHC Class I molecules are found on the surface of every nucleated cell and are involved in regulating the innate immune response through interactions with immunoregulatory cell receptors (McHugh and Bowness 2012). HLA-B27 is the most common MCH Class I B allele among Caucasians, and there exist over 100 subtypes of HLA-B27, with B*2705 being the most prevalent (Bowness 2015; Brown et al. 1996). The first breakthrough in understanding AS aetiology was made in the 1970's with the association between HLA-B27 and AS (Brewerton et al. 1973). Approximately 89% of AS patients are HLA-B27 positive, but only a small proportion of the general population (1-5%) who are HLA-B27 positive go on to develop disease (Taurog et al. 2016; Brown et al. 1997; Brown et al. 2000; Linden et al. 1984; Calin et al. 1983). A recent study in a Spanish cohort determined that AS patients who were HLA-B27 positive experienced earlier onset of AS, but they did not have a more debilitating disease than those who were HLA-B27 negative. Moreover, those who were HLA-B27 negative had an increased incidence of dactylitis and peripheral arthritis (Arévalo et al. 2018). Despite the long standing association between HLA-B27 and AS, it is still unclear as to how it contributes to disease, and how being positive or negative can affect disease manifestations. HLA-B27's

primary function in the immune response is to present antigens to cytotoxic $CD8^+$ T cells and this has led to a number of hypotheses for HLA-B27 mediated disease pathogenesis, as discussed below.

Arthritogenic Peptide Hypothesis

The arthritogenic peptide hypothesis was one of the first hypotheses about AS aetiology. It suggests that HLA-B27 has the ability to present unique peptides that could trigger disease (Benjamin and Parham 1992). This includes self-peptides presented by HLA-B27 that resemble peptides produced by foreign pathogens that are targeted by autoreactive $CD8^+$ T cells. It could also be caused by certain bacterial peptides or self-mimicking bacterial peptides being presented on HLA-B27 that initiate an immune response, resulting in inflammation. This could explain disease variation due to the number of different HLA-B27 subtypes that may bind different peptides. However, some studies disagree with this hypothesis (Schittenhelm et al. 2015; May et al. 2003). Schittenhelm et al. identified thousands of HLA–B27 specific peptides, but did not uncover any considered to be arthritogenic (Schittenhelm et al. 2015). However, they suggest that genetic polymorphisms relating to the peptide-binding cleft of HLA–B27 could relate to disease.

Unfolded Protein Response Hypothesis

The unfolded protein response hypothesis revolves around the theory that HLA-B27 frequently misfolds (Mear et al. 1999). When proteins misfold they cannot be correctly transported to their intended location and will accumulate in the endoplasmic reticulum (ER). When proteins accumulate in the ER it causes a stress response via two pathways, the unfolded protein response (UPR) and the ER-overload response (EOR) (Pahl 1999). These pathways trigger the immune response through the activation of nuclear factor kappa B (NF-κB), a key immunomodulatory transcription factor. NF-κB activation results in the

release of pro-inflammatory cytokines including TNF, IL-6, and IL-23, leading to inflammation (Smith and Colbert 2014; Pahl and Baeuerle 1997).

HLA-B27 Homodimer Model

The HLA-B27 homodimer model suggests that HLA-B27 homodimers interact differently with natural killer (NK) cells and $CD4^+$ T cells compared to its heterodimer (Allen et al. 1999). Killer cell immunoglobulin-like receptors (KIRs) are expressed on NK and T cells and recognize cells which have been infected with a pathogen and some can trigger an immune response. The HLA-B27 homodimer binds KIRs and causes a release of IL-17, which in turn stimulates fibroblasts and macrophages to release IL-6 and TNF, and causes T cell priming (Kollnberger et al. 2002; Dong 2008; Yen et al. 2006). HLA-B27 positive patients have been found to have an increase in the number of cells expressing KIR3DL2 in their blood and synovial fluid (Bowness et al. 2011). KIR-3DL2 is upregulated on the cell surface of $CD4^+$ T cells and when bound results in T helper 17 (Th17) cell differentiation, the main producers of IL-17 (Ridley et al. 2016).

1.3.2 Gut Inflammation and Ankylosing Spondylitis

An estimated 70% of AS patients suffer from some degree of gut mucosal inflammation, including Crohn's disease and ulcerative colitis, with 5% of patients having IBD (Costello et al. 2015; Ciccia et al. 2016). It has been suggested that HLA-B27 can influence and alter the gut microbiome, which could affect gut inflammation and explain the large overlap between IBD and AS (Lin et al. 2014). Th17 cells are thought to play an important role in AS pathogenesis and they are often found within the intestinal lamina propria of the gut (Ghoreschi et al. 2010). It has been found that interaction with bacteria can cause transition of T helper cells into Th17 cells, which accumulate in the bowel under the effects of IL-23 (Gagliani et al. 2015; Maxwell et al. 2015). Moreover, an inhibition of IL-23 results in an increase in Treg cell populations in the colon to facilitate an

immunosupressive response (Maxwell et al. 2015). Some gut bacteria have been found to be associated with AS including *Klebsiella pneumoniae*, *Lachnospiraceae*, *Prevotellaceae*, *Rikenellaceae*, *Porphyromonadaceae*, and *Bacteroidaceae* (Ebringer 1989; Garcia-Montoya et al. 2018). Gut permeability also appears to be increased in AS patients, which could increase exposure to bacteria (Martinez-Gonzalez et al. 1994). It has also been suggested that a relationship could exist between gut inflammation and AS where gut inflammation is actually a driver of disease pathogenesis. Transgenic rats have been used to investigate this hypothesis. HLA-B27 is considered the biggest factor in AS development, but when HLA-B27 positive rats were raised in a germ free environment they did not develop AS (Taurog et al. 1994). This suggests the HLA-B27 alone is not sufficient for disease development.

1.3.3 Disease Relevant Cell Types

Various cell types involved in both the innate and adaptive immune system are thought to be involved in autoimmune disease, and likely in AS. Cell types including T cells, NK cells, mast cells, neutrophils, dendritic cells, monocytes, macrophages, and B cells all have evidence suggesting their contribution to AS pathogenesis (Zambrano-Zaragoza et al. 2013). Unfortunately, it is unknown which of these cell types is the most relevant in disease manifestation. Evidence for the involvement for some of the major cell types is outlined below.

T Cells

T cells play a range of roles in an immune response. As described above, the arthritogenic peptide hypothesis relies on the involvement of $CD8^+$ T cells to initiate an immune response, and $CD4^+$ T cells express KIR3DL2 which are implicated in the homodimer model (Allen et al. 1999; Benjamin and Parham 1992). These hypotheses suggest a connection between T cells and HLA-B27 that may trigger disease, but these are not the only ways T cells may be

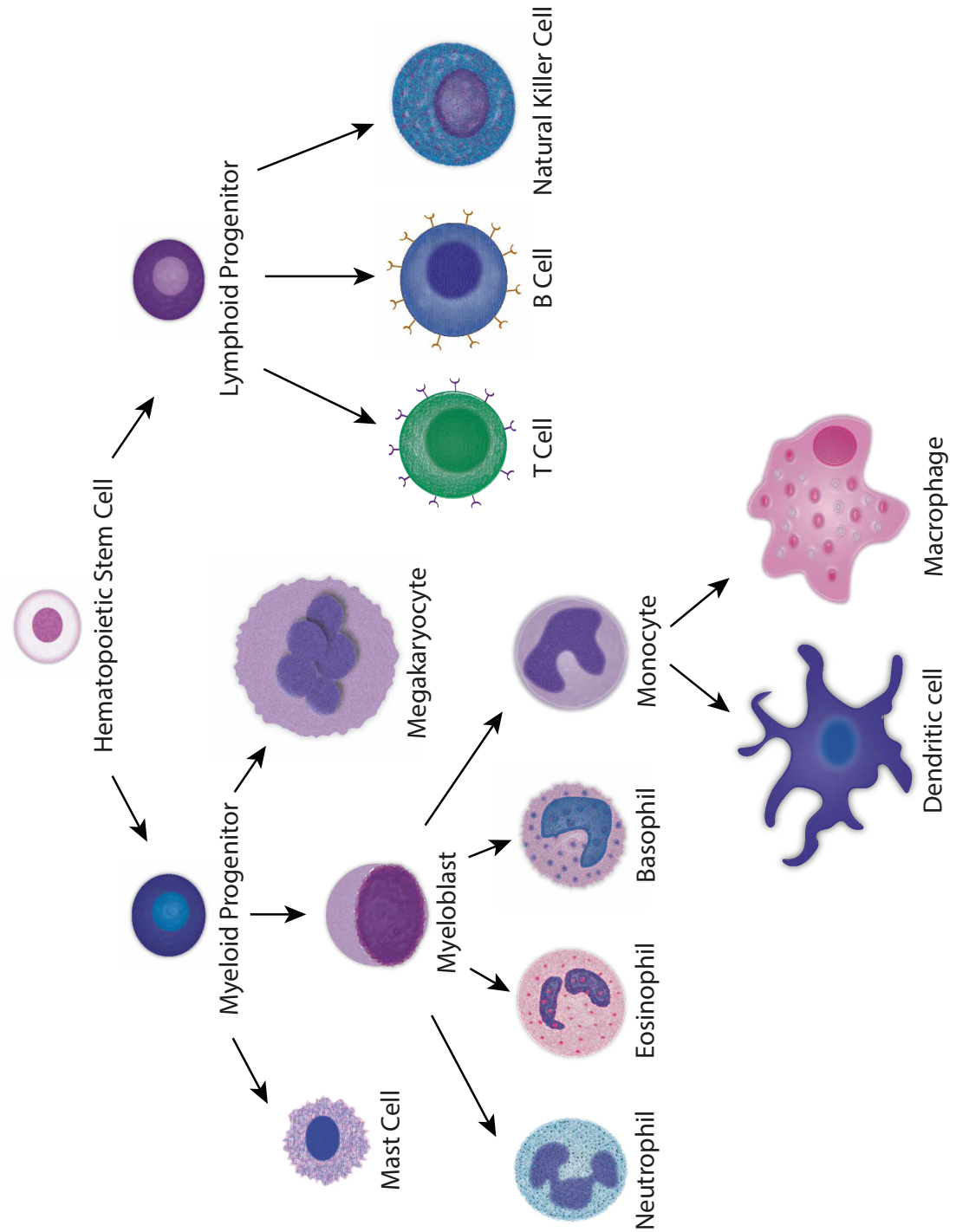


Figure 1.2: Cells of the immune system.

involved in AS pathogenesis. IL-23 is a pro-inflammatory cytokine involved in various autoimmune diseases, including AS (Cortes et al. 2013; Burton et al. 2007). IL-23 is secreted by antigen presenting cells and will bind IL-23R during an inflammatory response to trigger the expansion of Th17 cells and IL-17 expression (Yago et al. 2017). IL-17, in turn, induces IL-6 and IL-8 production, which are cytokines heavily involved in inflammation. IL-17 is secreted primarily by Th17 cells, but NK cells, mast cells, $CD4^+$ T cells, and neutrophils are also contributors (Miossec and Kolls 2012; Noordenbos et al. 2012). Neutrophils and monocytes are recruited to a site of inflammation by Th17 cells. They also encourage the development of osteoclasts, responsible for bone resorption, which is thought to be one of the triggers for bone growth to occur in later stages of AS (Limón-Camacho et al. 2012). Moreover, IL-17 is responsible for promoting osteoclastogenesis and works with TNF to suppress bone growth through over-expression of *DKK1* in the Wnt signaling cascade (Chabaud and Miossec 2001; Lubberts et al. 2001; Sato et al. 2006). Further evidence implicating T cells in disease is that high levels of Th17 cells have been found in the peripheral blood of AS patients and some AS biologic treatments have been found to reduce this number (Xueyi et al. 2013). One study also found evidence for the involvement of Th0 and Th1 cells, as well as Th17 cells (Farh et al. 2015). They examined SNPs across immune diseases and compared their locations with chromatin maps outlining regions of cis-regulatory elements in various cell types. Within AS, they found enrichment in Th17, Th0, Th1 cells and monocytes, suggesting they may play role in pathogenesis.

Monocytes and Macrophages

Monocytes are key players in the innate immune system and are the precursors for macrophages and dendritic cells, which are all involved in chemotaxis and antigen presentation, and therefore could be an important cell type in AS. Macrophages can also be sub-categorized as M1 or M2, which are predominately

pro-inflammatory and anti-inflammatory, respectively (Mantovani et al. 2004; Martinez and Gordon 2014). Macrophages differentiate from blood monocytes at the site of inflammation in response to different signals, such as IL-10 and Lipopolysaccharide (LPS), and are important in phagocytosis of pathogens and cleaning up cellular debris (Martinez and Gordon 2014). They are also present in most tissues under normal physiological conditions as resident tissue macrophages, which proliferate independently from blood monocytes (Honold and Nahrendorf 2018; Yona et al. 2013). Genes involved in crucial inflammation pathways are known to be upregulated in monocytes from SpA patients. Proteins involved in leukocyte recruitment, and important signaling pathways such as, VEGF, Janus kinases/Signal Transducer and Activator of Transcription proteins (JAK/STAT), and toll-like receptor (TLR) were found to be upregulated in AS monocytes following a study using label-free quantitative expression profiling to examine protein expression (Wright et al. 2009). Moreover, genes involved in the ubiquitin proteasome pathway (UPP) were shown to be up-regulated in AS monocytes, which is responsible for the formation of peptides to be presented by MHC Class I proteins, like HLA-B27 (Wright et al. 2009).

B Cells

B cells are primarily known for their role in secreting antibodies to combat infection, but they are also important in cytokine secretion. B cells work to suppress T cells by secreting IL-10, an anti-inflammatory cytokine (Mauri et al. 2003). IL-10 also drives macrophage differentiation to an anti-inflammatory M2 state (Martinez and Gordon 2014). One study found that B cell populations increased in AS patients, but those B cells secreting IL-10 were reduced. The authors also noted that the healthy controls were better able to suppress memory $CD8^+$ T cells and naïve T cells (Chen et al. 2015). There is also evidence that B cells are receptive to IL-17 which, as mentioned, is a key regulator of inflammation and contributor to AS pathogenesis (Miossec and Kolls 2012).

Other Cell Types

NK cells are a component of the innate immune response and work to destroy cells infected by pathogens. Like T cells, they express KIR3DL2 which drives the homodimer model of AS pathogenesis. In SpA patients, NK cells have been found to produce pro-inflammatory cytokines, such as IL-12 and IFNγ, and are involved in activating fibroblast-like synoviocytes and dendritic cells (Conigliaro et al. 2011). There is also evidence they are involved in bone damage, further highlighting their potential importance in AS (Conigliaro et al. 2011).

Neutrophils are the most abundant immune cell type in humans and make up an important part of the innate immune response to regulate and resolve inflammation (Kruger et al. 2015). They are recruited by IL-17 and CXCL8 secreting Th17 cells to a site of infection, but this can create a positive feedback loop as neutrophils secrete the chemokines CCL20 and CCL2, which recruit Th17 cells and chronic inflammation can ensue (Pelletier et al. 2010). Studies have shown that AS patients have increased levels of neutrophils compared to healthy controls and are thought to play a role in autoimmune disease (Mercan et al. 2016).

1.3.4 Therapeutics

The first line of treatment for AS patients is nonsteroidal anti-inflammatory drugs (NSAIDs), which include drugs such as ibuprofen and aspirin (Mahmood and Helliwell 2017). NSAIDs have proven to be effective at disease management, with one study finding that the majority of patients benefited from NSAID use, but at the conclusion of the study, 44% of participants still had a high disease score (Baraliakos et al. 2016). Moreover, it has been found that continued, long-term use of NSAIDs can have adverse effects that include hypertension, renal disease, and gastrointestinal complications (Baraliakos et al. 2016; Mahmood and Helliwell 2017). A new drug, B-D-mannuronic acid is a potential improvement

over other NSAIDs because there are no known renal complications with treatment and has increased gastrointestinal tolerability, but it is still under debate as to whether NSAIDs improve radiographic progession (Fattahi et al. 2018; Wanders et al. 2005; Sieper et al. 2016). Opioids and steroids are sometimes used to relieve symptoms as well. Steroids can be injected directly into the affected joints to reduce inflammation, but it has been found that long-term use can result in osteoporosis (Ryall and Helliwell 1998).

When these treatments have not proven successful for an AS patient, then biologic drugs are prescribed. TNF inhibitors were the first biologics to be approved and are the most commonly prescribed (Garcia-Montoya et al. 2018). Common anti-TNF drugs include infliximab, etanercept, adalimumab, certolizumab pegol, and golimumab. Most patients respond well to anti-TNF treatment and report a decrease in their levels of pain, with a reduction in spinal inflammation as seen by MRI (Martín-Mola et al. 2010; McVeigh and Cairns 2006). However, anti-TNF therapy does not work in all instances. Approximately 30-40% of SpA patients do not respond or become tolerant to anti-TNF drugs (Menegatti et al. 2019). In addition to non-responders, some patients develop complications with tolerance, and some patients experience a decrease in efficacy with long-term use (Antonelli et al. 2015). There is also some debate as to whether anti-TNFs are indeed disease modifying (Tan et al. 2015).

This has driven researchers to develop new drugs to combat AS. Unfortunately disease modifying anti-rheumatic drugs such as methotrexate that are used to treat rheumatoid arthritis (RA) have been found to be ineffective in AS (Van Der Heijde et al. 2017a; Haibel et al. 2007). As IL-17 is thought to play an important role in disease pathogenesis, secukinumab was introduced as the first IL-17A inhibitor for use in AS (Blair 2019). Secukinumab is a human monoclonal antibody that binds selectively to IL-17A which inhibits its interaction with its receptor, and therefore stops the release of pro-inflammatory cytokines (Blair

and Dhillon 2016). Secukinumab has proven to be successful in AS by improving signs and symptoms and is recommended as an alternative for those who are not suited to NSAIDs or anti-TNF treatment (Robertson et al. 2018; Gentileschi et al. 2018; Blair 2019). It was associated with a low rate of radiographic progression, and reduced inflammation in the sacroiliac joints, with long-term efficacy and tolerability (Braun et al. 2019; Blair and Dhillon 2016). Other drugs targeting IL-17 are undergoing trials in AS and are being used in other inflammatory diseases. Despite their success, IL-17 inhibitors have been found to exacerbate existing IBD and potentially contribute to new-onset IBD in AS patients (Mahmood and Helliwell 2017). Moreover, treatments with biologic drugs such as anti-TNF and anti-IL-17 can increase a patient's risk of infection due to the drugs' anti-inflammatory properties (Blair 2019).

Drugs targeting other AS-relevant cytokines are under investigation (Garcia-Montoya et al. 2018). The IL-23 signaling cascade is thought to be an important part of AS pathogenesis and a potential target for treatment. Ustekinumab is a human monoclonal antibody for the p40 subunit of IL-12 and IL-23 that has been successful in psoriasis disease management, but has had conflicting results in AS trials (McInnes et al. 2013; Poddubnyy et al. 2014b). Tofacitinib is a JAK inhibitor involved in the IL-17/IL-23 signaling cascade and has so far been successful in AS clinical trials (Van Der Heijde et al. 2017b). These biologic drugs represent new treatment options for AS patients, however they are also very expensive.

Regardless of treatment, it is recommended that AS patients maintain a high level of physical activity and exercise frequently (Ryall and Helliwell 1998). This helps improve flexibility and range of movement, decreases joint stiffness, and helps improve the mood of those struggling with depression or anxiety as a result of chronic pain and disability. Physical therapy is a common way to improve disease management. In severe cases, patients may need to undergo surgery for spinal correction or hip replacements (Mahmood and Helliwell 2017).

1.4 Gene Regulation and Disease

Changes in a DNA sequence can have significant effects on gene expression, and can sometimes lead to disease (Ku et al. 2010; Ramagopalan et al. 2007; Haraksingh and Snyder 2013). Genetic variation can be single nucleotide polymorphisms (SNPs), structural variants such as small insertions and deletions (INDELs), or larger copy number variants (CNVs). Changing a single base within a coding region can lead to a synonymous mutation in which the resulting amino acid sequence remains unchanged, or a non-synonymous change which results in missense and nonsense mutations, altering the final protein product. INDELs and CNVs within a coding region can lead to frameshifts and non-functional proteins. Variants that fall within the non-coding region can have an impact on gene expression by disrupting regulatory regions, but their effects are much harder to ascertain. Understanding the underlying genetics of disease is crucial in finding better diagnostic and treatment options for patients suffering from genetic disorders.

1.4.1 Genome-Wide Association Studies

Disease can result from genetic changes within a single gene or be the result of a combination of variants across multiple genes. GWAS have been instrumental in the past 10 years in uncovering genetic risk for complex traits by associating common variants with disease (Visscher et al. 2017). GWAS involve understanding genetic variation in a population, including informative 'tag' SNPs, and using a genotyping array to genotype thousands of individuals, including cases and controls. At least one SNP is tagged per haplotype block, based on the assumption that all the genetic variants within that region are in linkage disequilibrium (LD) and are inherited together (Visscher and Hill 2009). Therefore, knowing one SNP allows researchers to know other SNPs

in LD with it by performing genotype imputation. Some researchers are also attempting whole-genome sequencing (WGS) as an alternative to SNP arrays to capture more variation to help identify disease causing mutations (Brody et al. 2017). However, WGS remains costly. Bioinformatic improvements on the analysis of low-depth WGS genotypes to identify variants are emerging (Gilly et al. 2018), but traditional SNP array based GWAS remain the more common way of examining genetic variation across large populations. Variants associated with disease through GWAS can involve hundreds of genomic loci, located anywhere in the genome, making GWAS interpretation a challenge.

1.4.2 Fine Mapping of GWAS Variants

Due to the nature of GWAS, it is not always the SNP observed to be associated with the disease that is the causal variant. SNPs lie in haplotype blocks and are inherited together, making it difficult to discern which SNP is driving the association signal. Some variants in LD with the lead GWAS marker(s) may not even be genotyped and are therefore missed during downstream analysis. Existing functional annotations of loci and variants along with statistics are combined in fine mapping approaches. Fine mapping attempts to narrow down an extensive list of variants from a GWAS signal to a select few that are likely to be associated with disease. Genotyping arrays developed to study a specific set of SNPs, such as the Immunochip for immune traits can be used for fine mapping as well as statistical analysis to determine a credible set, which is a minimal set of SNPs with the highest chance of being causal (Cortes et al. 2013; Spain and Barrett 2015). Variants from these analyses can then be annotated using publicly available datasets and in silico predictors.

To fine map a locus, imputation must be carried out to generate a list of all the variants within a particular genomic region by providing all the known SNPs within the same haplotype block as the lead GWAS SNP (Spain and Barrett

2015; Howie et al. 2012), commonly using data from the International HapMap Project (International HapMap Consortium 2003; Altshuler et al. 2010), or the 1000 Genomes Project (Durbin et al. 2010). Typically, loci being fine mapped undergo targeted re-sequencing using next-generation sequencing (NGS), which serves to verify the presence of SNPs already within a locus and to find any rare or missed variants that may be present (Salomon et al. 2016). A stepwise conditional analysis is then performed to define additional associated signals within a locus (Spain and Barrett 2015; Bunt et al. 2015). To narrow down the list of associated variants, a Bayesian approach is commonly used to define credible sets by calculating a posterior probability (Spain and Barrett 2015). This posterior probability is the probability of each SNP within an associated loci to be the causal variant, with the assumption that there is only one causal variant at a given locus. Variants that collectively represent 99% of the posterior probability make up the 99% credible set. Hopefully, credible sets will contain a manageable number of variants, that will include the causal SNP, for experimentation or annotation using data from ENCODE (Dunham et al. 2012) and NIH Roadmap studies (Bernstein et al. 2010).

1.4.3 Understanding the Role of Non-Coding Variants

The impact of variants within coding regions of the genome are relatively straightforward to predict using in-silico modeling of non-synonymous mutations within genes and proteins. However, mutations found outside of exons present a more considerable challenge. With 98% of the human genome being non-coding and the majority of GWAS SNPs for diseases found within introns and intergenic regions, it is clear that these non-coding regions are key in gene regulation and understanding the variety of ways in which small and large genomic variation can impact gene expression is essential (Maurano et al. 2012). Complex processes govern gene regulation. These include the physical bending and looping of chromatin to bring enhancer elements nearer their

regulated genes, microRNA activity, protein binding of transcription factors, alternative splicing, and post-translational modifications, such as methylation and acetylation (Ward and Kellis 2012; Knight 2012; Rivera and Ren 2013; Kilpinen et al. 2013; McVicker et al. 2013). Single nucleotide changes or small INDELs do not need to be near a gene to impact their expression, making understanding the effects of these variants complex. Context specificity can also occur in these processes. For example, an enhancer may only be active in a certain cell type or in the presence of a particular stimulus. Regions of open chromatin, locations of histone markers, and chromatin interactions all vary between cell types to govern cell and tissue specific gene expression (Jaenisch and Bird 2003; Agelopoulos et al. 2012; Lelli et al. 2012). Interactions of less than 1Mb in distance are known as local interactions and are often cis-acting, and if they are greater than 1Mb apart, this is known as distant interaction, and can be cis-acting through physical interactions of chromatin, or can be trans-acting through protein recruitment and signaling cascades (Fairfax and Knight 2014). Elucidating the role of non-coding variants is made more complex by the fact that a single non-coding SNP may be involved in the regulation of multiple genes through a given enhancer and that multiple SNPs may exert a combined (haplotypic) effect on a given gene. This was demonstrated by Corradin et al. when they used data from HapMap and B lymphoblasts across six immune diseases and concluded that SNPs in LD can act on multiple enhancers, all near each other, and collectively contribute to expression of a gene (Corradin et al. 2014).

Non-coding variants have the potential to disrupt splice sites, transcription factor binding motifs, physical chromatin interactions, or chromatin accessibly and have a detrimental effect on gene expression, leading to disease. Some examples of non-coding variants resulting in disease are outlined in Table 1.3. One such disease is δ-thalassemia. It is an autosomal recessive disorder and

can result from a point mutation 77 nucleotides upstream of the transcription start site for the gene δ-globin (Matsuda et al. 1992). When DNA becomes hypo- or hyper-methylated, genes can be silenced or over expressed, due to variants within methylation sites (El-Maarri 2003). Proteins may no longer be able to access DNA to initiate transcription, or conversely, enhancers and promoters could be exposed, when under regular physiological conditions, they would be inaccessible. Changes in methylation have been associated with several types of cancers including gastric (Akiyama et al. 2003), colon (Nakamura and Takenaga 1997), and breast cancers (Gupta et al. 2003). As well as methylation, dysregulation of histone acetylation can also cause disease. Rubinstein–Taybi syndrome (RSTS), characterized by cognitive impairment and congenital abnormalities, is the result of mutations in the histone acetyl transferase, CREB-binding protein (CBP) (Petrif et al. 1995). This causes impaired function of the acetyl transferase and a decrease in transcription. Splicing defects have also been associated with autoimmune disease, including diabetes (Ueda et al. 2003). Understanding the non-coding genome is imperative in characterizing GWAS SNPs and finding disease causing variants.

In the last few years, large scale projects such as the Encyclopedia of DNA Elements (ENCODE) Project Consortium (Dunham et al. 2012), the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010) and BLUEPRINT Consortium (Adams et al. 2012), have worked to annotate the non-coding regions of the genome in different cell types, using a variety of assays, to better understand the effects of variants within these regions. Their publicly available datasets can help make in silico annotations for GWAS hits.

Table 1.3: Examples of diseases associated with a single nucleotide variant change to illustrate the diversity of mechanisms that may be involved.
A single nucleotide change in the sequence of a gene or regulatory region can have detrimental effects on gene expression and function, and lead to disease.

SNP Effect	Disease	Mechanism
Transcription factor binding	δ-thalassemia	A mutation upstream of the transcription start site for the gene δ-globin results in an altered recognition site for the transcription factor, GATA-1, causing a decrease in binding affinity and therefore a change in gene expression (Matsuda et al. 1992).
Transcription factor binding	Coronary heart disease	A SNP disrupts the transcription factor, activator protein 1 (AP-1), which in turn decreases expression of *TCF21*, another transcription factor, leading to improper gene regulation (Miller et al. 2013).
mRNA stability	Psoriasis	A synonymous point mutation in *CDSN* causes an increase in RNA stability, resulting in a decrease in the affinity of a RNA binding protein, altering expression levels (Capon et al. 2004).

Continued on following page

Table 1.3 - *continued from previous page*

SNV Effect	Disease	Mechanism
Nucleosome Structure	Rubinstein–Taybi syndrome	CREB-binding protein (*CBP*) is a histone acetyl transferase that is responsible for acetylating core histones that make up nucleosomes. Mutations in *CBP* inhibit its function and result in either a decrease of histone acetylation or incorrect acetylation of other proteins (Hendrich and Bickmore 2001; Petrif et al. 1995).
Non-synonymous change	Phenylketonuria (PKU)	A C>T mutation changes an arginine residue (CGA) into a stop codon (TGA) resulting in termination of translation. This produces a truncated, non-functional protein (Erlandsen and Stevens 1999).
Splicing defect	Nager syndrome	*SF3B4* encodes SAP49, which is a component of the U2 pre-mRNA spliceosomal complex. Frameshift mutations in *SF3B4* inhibit its ability to bind to the pre-mRNA for splicing (Bernier et al. 2012).

Continued on following page

Table 1.3 - *continued from previous page*

SNV Effect	Disease	Mechanism
Splicing defect	Hutchinson–Gilford progeria syndrome	A C>T mutation in *LMNA* exon 11 results in the activation of a normally unused splice site, producing a truncated version of the protein. This truncated version is dominant to the full version of the protein, causing disease (De Sandre-Giovannoli and Lévy 2006; Tazi et al. 2009).
Splicing defect	Type 1 diabetes, Graves' disease	Multiple Splicing variants found in the *CTLA* gene have been associated with autoimmune disease and are thought to impact T cell activity (Ueda et al. 2003).
Methylation	Cancer	Many mutations in cancers have been known to alter methylation recognition sites of enzymes and result in the DNA becoming hypo- or hyper-methylated. This is associated with a deleterious up- or down-regulation in gene expression (Akiyama et al. 2003; Nakamura and Takenaga 1997; Gupta et al. 2003).

1.5 Genetics of Ankylosing Spondylitis

AS is a complex genetic disorder with an overall heritability of 90% (Brown et al. 1997; Pedersen et al. 2008). AS has a parent–child recurrence risk of 7.9%, and first-degree relatives are >50 times more likely to develop the disease than unrelated individuals (Brown et al. 1997; Pedersen et al. 2008; Brown et al. 2000). The more distantly related the individuals, the lower the recurrence risk becomes, suggesting that AS is a polygenic disorder with multiple genes contributing to its pathogenesis (Brown et al. 2016). Early evidence of genetic association involving HLA-B27 has provided an explanation of some of the heritability of AS, but not all HLA-B27 positive individuals develop disease, and recent evidence suggests that it accounts for only 20% of the risk (Chen et al. 2017). Monozygotic twins have a disease concordance rate of 63%, compared to 12.5% for dizygotic, even when they are both HLA-B27 positive, suggesting that there are other genes playing a role in the disease (Brown et al. 2000).

1.5.1 Ankylosing Spondylitis GWAS

To uncover genes contributing to AS pathogenesis, various GWAS analyses have been carried out. The first AS GWAS was done by the Welcome Trust Case Consortium and the Australo-Anglo-American Spondylitis Consortium in 2007 (Burton et al. 2007). This initial study used a set of 14 436 non-synonymous SNPs genome-wide, and an additional 897 SNPs within the MHC. It identified two major new susceptibility loci, endoplasmic reticulum aminopeptidase 1 (*ERAP1*) and *IL23R*. In 2010 a larger GWAS was undertaken using 2053 AS patients with 5140 controls and identified AS associations at two loci in gene deserts, suggestive of the involvement of regulatory elements (Reveille et al. 2010). They also found evidence of association involving *ANTXR2* and *IL1R*. A further study in 2011 identified more novel associations including *RUNX3*,

LTBR-TNFRSF1A, *IL12B*, *PTGER4*, *TBKBP1*, and *CARD9* (Evans et al. 2011). Two years later an Immunochip GWAS was performed that identified 13 new loci and confirmed 12 previously associated loci (Cortes et al. 2013). In 2016, a meta-analysis study of GWAS compared high-density genotyping data from across five different inflammatory diseases and identified another 17 loci associated with AS (Ellinghaus et al. 2016). A total of 48 genomic loci have been identified following all these studies (Figure 1.3) and have highlighted the role of non-MHC genes in AS pathogenesis.

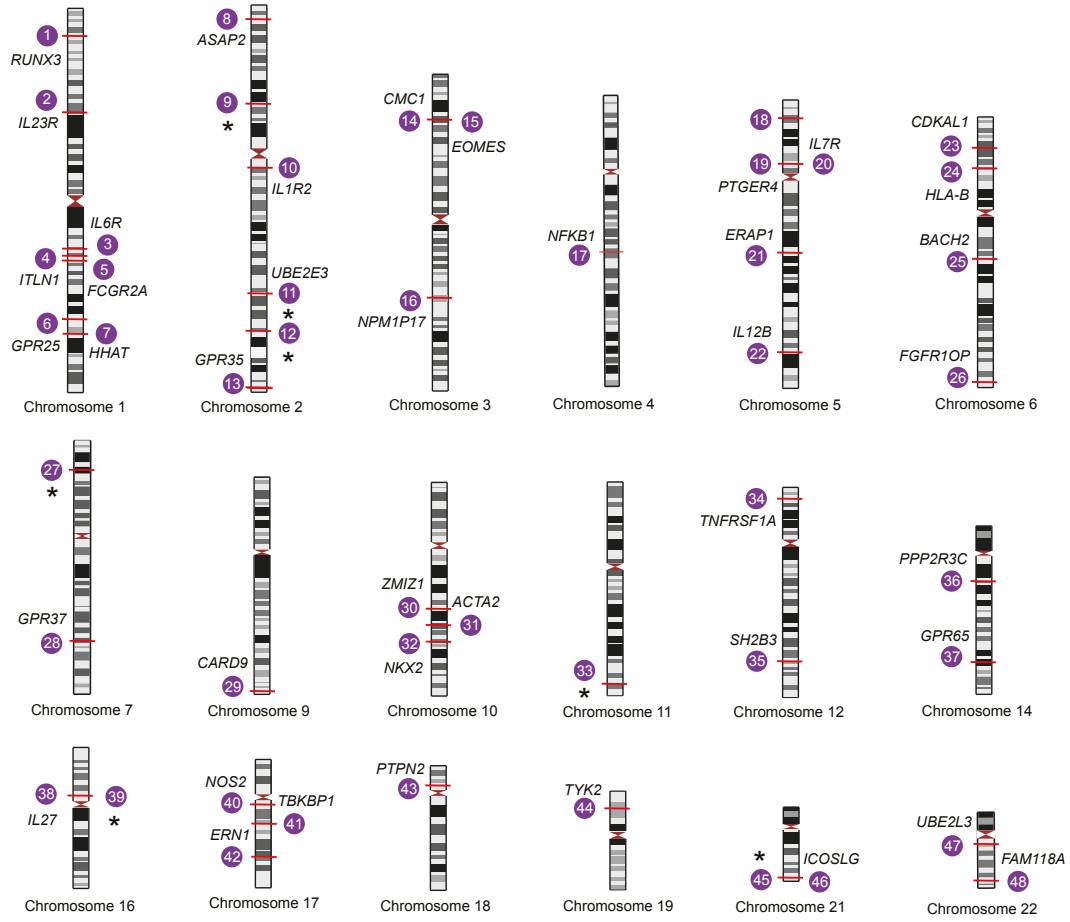


Figure 1.3: AS GWAS associated loci.
Loci associated with AS at genome-wide significance following the meta-anaylsis by (Ellinghaus et al. 2016) are plotted (numbered 1–48). The implicated gene (by proximity) and chromosomal position (marked in red) are shown for each locus; where there is no reported gene, this is indicated by an asterisk. In the majority of cases, the genes listed are candidates based on proximity and biological plausibility, and causality has not been established except in a small number of instances. Figure reproduced from Osgood and Knight 2018.

1.5.2 Notable GWAS Findings

Aminopeptidases

ERAP1 was one of the first genes to be associated with AS from GWAS and has been associated with disease across populations of both European and Asian ancestry (Burton et al. 2007; Robinson and Brown 2014). It codes for an aminopeptidase of the ER and is highly polymorphic (Sanz-Bravo et al. 2018). It is responsible for peptide cleavage prior to presentation by MHC Class I molecules (Saric et al. 2002; Serwold et al. 2002). Proteins are processed and then transported from the cytoplasm of a cell to the ER where ERAP1 and ERAP2 are located to cleave any long N-terminally extended peptides. ERAP1 can cleave almost any type of peptide bond and will trim peptides that are 10 amino acids or longer, and becomes inactive the shorter the peptide (Serwold et al. 2001; Chang et al. 2005). This feature is known as the molecular ruler mechanism.

Variants within *ERAP1* could interfere with peptide processing and presentation on MHC Class I molecules. *ERAP1* is an important disease candidate because it has been associated with disease only in HLA-B27 positive individuals (Evans et al. 2011). Given HLA-B27 is an MHC Class I molecule reliant on peptide cleavage by ERAP1 this relationship could help explain pathogenesis. However, ERAP1 is highly polymorphic and it is important to functionally characterize the effect of these polymorphisms on ERAP1 function and how they change the peptides presented by MHC Class I molecules. *ERAP1* has 10 major haplotypes across human populations that can either be protective against disease, or increase disease risk (Ombrello et al. 2015). Evidence exists that an AS associated SNP, rs31087, is protective due to a decreased ability to cleave peptides prior to MHC presentation, which would slow down the cleavage rate and may help moderate an immune response (Evans et al. 2011; Kochan et al. 2011). A fine mapping study performed to examine *ERAP1* variants in AS found two missense

SNPs, rs30187 and rs10050860, that are likely to be involved in disease risk (Bunt et al. 2015). Another study found a decreased stability of HLA-B27 correlated with a decrease in ERAP1 activity and the N and C terminals of HLA-B27 epitopes are both extended in *ERAP1* silencing experiments (García-Medel et al. 2012; Chen et al. 2014). HLA-B27 free heavy chain expression on the cell surface could be controlled by ERAP1 enzymatic activity and, therefore, could be highly susceptible to *ERAP1* polymorphisms (Chen et al. 2016a).

There is also evidence that a loss of ERAP1 could suppress Th17 cell expansion, one of the disease relevant cell types involved in regular immune responses (Chen et al. 2016a). ERAP2 is a similar aminopeptidase highlighted through GWAS that has the ability to form a heterodimer with ERAP1 (Cortes et al. 2013; Burton et al. 2007; Evnouchidou et al. 2014). It is also involved in peptide cleavage for presentation by MHC Class I molecules and therefore thought to be involved in AS pathogenesis (Martín-Esteban et al. 2016). NPEPPS and LNPEP are other aminopeptidases that have been associated with AS (Cortes et al. 2013). Peptide cleavage and presentation is an important part of initiating an immune response and understanding the role of these genetically associated aminopeptidases may help uncover AS pathology.

IL-23 Signaling Cascade

The initial AS GWAS highlighted the importance of the IL-23 signaling cascade as a driver of disease, with the gene encoding its receptor, *IL-23R*, being strongly associated (Burton et al. 2007). IL-23 is a pro-inflammatory cytokine that stimulates both the NF-κB and IL-17 pathways. IL-23R is highly expressed on T cells and NK cells and is made of two subunits, IL-12p40 and IL23p19, with the former being encoded by *IL-12B*, another GWAS-associated gene (Evans et al. 2011; Danoy et al. 2010). IL-23 binds IL-23R which results in JAK/STAT pathway signaling involving TYK2, JAK2, and STAT3, which have also been

associated with AS (Danoy et al. 2010; Cortes et al. 2013). STAT3 stimulates the IL-17 signaling pathway, resulting in the expansion of Th17 cell populations from $CD4^+$ cells. Variants within *IL-23R* can be either risk or protective alleles. Protective variants have been associated in AS that reduce the sensitivity of IL-23R, thus stimulation by IL-23 has a reduced effect (Di Meglio et al. 2011; Sarin et al. 2011). One protective variant, rs11209026, is a missense mutation that inhibits IL-23R signaling and results in a decrease in IL-17 and IL-22 production, as well as a reduction in Th17 cells (Di Meglio et al. 2011). Mouse models have shown that when IL-23 is over-expressed, it is enough to cause murine SpA by inducing bone loss and increasing myelopoiesis and osteoclast differentiation (Adamopoulos et al. 2011). One study found that over-expression of IL-23 caused inflammation of the entheses, which are areas where ligaments connect to bone and are key in arthritis (Sherlock et al. 2012). Other immune disorders have also been associated with IL-23R including Behcet's disease (Mizuki et al. 2010), ulcerative colitis (McGovern et al. 2010), Crohn's disease (Barrett et al. 2008), IBD (Liu et al. 2015), psoriasis (Nair et al. 2009), and psoriatic arthritis (Stuart et al. 2015).

TNF Signaling Cascade

TNF is a pro-inflammatory cytokine produced mainly by monocytes and macrophages (Idriss and Naismith 2000). It binds to its cell surface receptors, TNFRSF1A or TNFRSF1B, to signal other pro-inflammatory cytokines and pathways, such as NF-κB, and can also trigger apoptosis. TNF is involved in bone degradation and its levels are elevated in synovial tissue and serum from AS patients, suggesting this pathway could play a role in disease pathogenesis(Baron and Rawadi 2007; Sato et al. 2006; Martín-Mola et al. 2010). TNF inhibitors also remain the most effective treatment option for AS (discussed in section 1.3.4). SNPs near/in three major genes in the TNF pathway have been associated with AS through GWAS. These include the TNF receptor *TNFRSF1A* (Evans et al.

2011; Burton et al. 2007; Corona-Sanchez et al. 2012; Davidson et al. 2011; Karaderi et al. 2012), a NF-κB inducer *TBKBP1* (Cortes et al. 2013), and the apoptosis mediators *TRADD* (Pointon et al. 2010) and TNFSF15 (Zinovieva et al. 2009). A signal from AS GWAS at 12p13 has been associated with *TNFRSF1A* and is in LD with a multiple sclerosis (MS) associated SNP. This MS SNP results in the loss of exon 6 and the transmembrane domain of TNFRSF1A due to a splicing defect, resulting in a reduction in TNF signaling (Gregory et al. 2012). Interestingly, anti-TNF therapy has been shown to exacerbate MS cases (Arnason et al. 1999; Van Oosten et al. 1996). It is unclear how TNF, or its modulated genes, fully contribute to AS pathogenesis, but it is a major modulator of the immune response and inflammation, making it a strong candidate for future investigations.

1.5.3 The Limitations of GWAS

GWAS have been invaluable in uncovering risk genes and variants across many complex diseases, but they do not provide the whole story for a variant's mechanism of action and how it pertains to disease aetiology. Finding a disease causing SNP can be challenging as many variants are in LD with each other making it difficult to discern disease relevant variants. Furthermore, GWAS are limited by the informativeness of genotyping and often only detect common SNPs, so rare genetic variants will not be associated with a disease (Wu et al. 2017). Studies may also struggle to identify causal variants due to restrictive arrays, unknown epistatic interactions, or errors in phenotyping individuals when recruiting cases and controls (Blanco-Gómez et al. 2016; Eichler et al. 2010; Glazier et al. 2002; McCarthy et al. 2008). Moreover, most GWAS SNPs (93%) are not found in exons and can be scattered throughout introns and intergenic regions, and this is true of AS GWAS (Maurano et al. 2012; Schaub et al. 2012). Mutations in regulatory regions of DNA can affect gene expression, and GWAS alone lack the ability to characterize such variants. To combat these

issues, functional genomic technologies are being implemented to determine the regulatory genomic landscape within which genetic variants may act, the consequences of allelic variation, and the genes and pathways they modulate to aid in uncovering the underlying genetic etiology of AS.

1.6 Functional Genomic Approaches

As discussed, there are many challenges faced when attempting to find a causal relationship between a variant and disease. GWAS provide no information about how or why a variant is associated with disease, only that it is shared by affected individuals. This need for further characterization has resulted in the post-GWAS era, to bring new understanding to GWAS results. Post-GWAS analyses often begin with fine mapping approaches to prioritize variants, followed by in silico analysis. In silico analysis has improved over the years and is highly informative for coding variants, but for non-coding variants its informativeness is limited. Non-coding variants can have an effect on transcription factor binding, so often these variants are assessed to see if they fall within a transcription factor binding motif. However, only 10% to 20% of predicted causal variants from autoimmune GWAS are thought to lie within known motifs (Farh et al. 2015). Researchers also attempt to characterize GWAS variants through experiments such as reporter assays to identify variants that disrupt enhancers or promoters. However, these types of experiments are limited by the fact promoters and enhancers may only be active in a specific cell type or under certain physiological conditions, which are not replicated by the assay (Gallagher and Chen-Plotkin 2018; Heinz et al. 2015). The importance of context specificity is highlighted by a variant in a gene desert at 9p21. It is associated with coronary artery disease and has been shown to impact STAT1 binding by disrupting an enhancer in vascular endothelial cells that is only active through chromatin looping in IFNγ activated cells (Harismendy et al. 2011). NGS has revolutionized the field

of genetics and has enabled the development of new technologies to map and annotate chromatin. Chromatin accessibility and interactions, as well as histone modifications and resulting gene expression changes can be examined in high-throughput ways to bring new understanding to the non-coding genome and the disease causing variants situated within it (Figure 1.4).

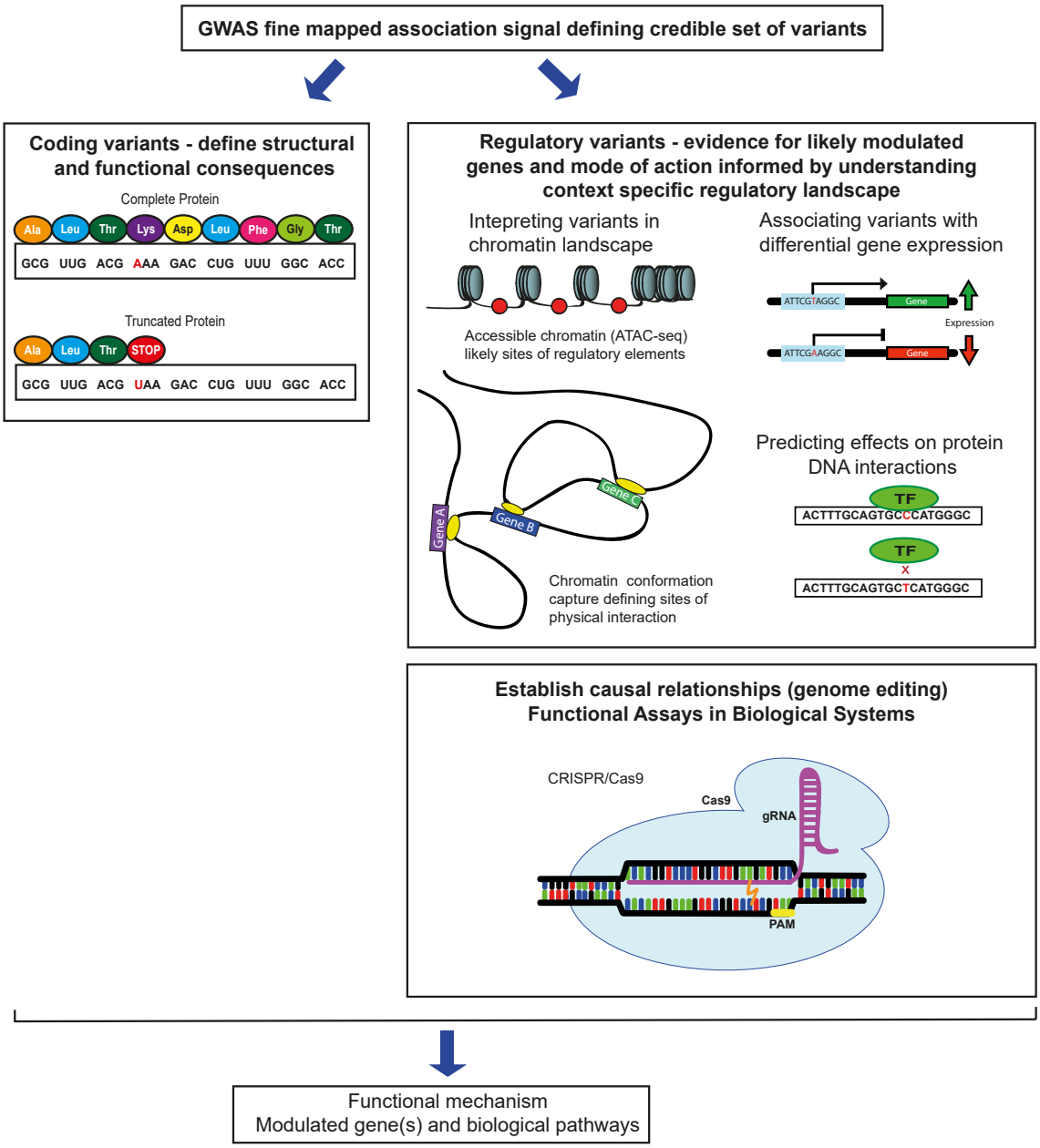


Figure 1.4: Approaches enabling understanding of the functional basis of GWAS. GWAS highlights many variants associated with disease. Fine mapping followed by functional genomic techniques and experimental validation can enable the understanding of the role of these variants in a biological system and how they may lead to disease. Figure reproduced from Osgood and Knight 2018.

1.6.1 Defining Chromatin Accessibility

Chromatin is organized by winding DNA around histones to package it into nucloesomes. DNA accessibility is determined by post translational modifications and protein binding, which condense it into heterochromatin to inhibit transcription, or open it up into euchromatin to encourage transcription. Important regulatory elements are thought to be situated in regions of open chromatin so that they are accessible to transcription factors, co-factors, and other regions of chromatin for physical interactions. Chromatin accessibility is cell type specific and can be context dependent. Traditionally, DNase I hypersensitivity mapping has been used to define regions of accessible chromatin and is considered to be the gold standard (John et al. 2013; Tsompana and Buck 2014). It was used by the ENCODE consortium for mapping chromatin accessibility genome-wide and their resulting datasets are publicly available (Dunham et al. 2012). The DNase I enzyme is used to generate double-stand breaks, but can only do so in regions where the chromatin is not tightly wound around histones and inaccessible (Weintraub and Groudine 1976). This generates fragments of DNA from open chromatin that can be amplified and analyzed using real-time quantitative polymerase chain reaction (qPCR), Southern blotting, or NGS to provide a visual map of chromatin accessibility (Keene et al. 1981; Crawford et al. 2005).

In 2013, Buenrostro et al. introduced a new method of mapping open chromatin called, the assay for transposase-accessible chromatin using high-throughput sequencing (ATAC-Seq) (Buenrostro et al. 2013). ATAC-Seq is able to evaluate transcription factor binding sites, nucleosome positions, and chromatin accessibility with substantially fewer cells than DNase I hypersensitivity, and is a more robust assay (Buenrostro et al. 2013; Tsompana and Buck 2014). It relies on two main components, fragmentation of DNA, and high-throughput sequencing. ATAC-Seq employs the use of a special enzyme, Tn5 transposase,

that simultaneously fragments DNA and tags it with adapters for sequencing in a process deemed, 'tagmentation'. The transposase can only fragment the DNA where is it readily accessible, like in DNase I assays, and therefore produces fragments that undergo NGS to provide information about nucleosome packing, transcription factor binding, and regions of open chromatin. Variants highlighted through GWAS can be annotated by overlapping their location with chromatin accessibility maps produced from ATAC-Seq. SNPs found within densely packed chromatin are unlikely to be of relevance while those situated within open chromatin could have a regulatory function and be disease causing. ATAC-Seq has had recent success in schizophrenia, a neurological disorder with many associated variants. Forrest et al. used ATAC-Seq to map the regions of open chromatin in human induced pluripotent stem cell (iPSC) derived neurons (Forrest et al. 2017). By overlapping schizophrenia GWAS risk variants with genome-wide chromatin accessibilty maps, the researchers found a single SNP from a collection of potential causal variants at an associated risk locus, *MIR137*, that was located in open chromatin. They then functionally validated this SNP through the use of clustered regularly interspaced short palindromic repeats (CRISPR) to confirm that the variant resulted in a decreased expression of *MIR137*.

1.6.2 Defining Chromatin Interactions

DNA, and the chromatin which packages it, does not exist as a straight line of nucleotides. It can bend and loop to cause physical interactions between cis-acting regulatory elements and promoters to modulate gene expression, therefore creating a 3D landscape. Assaying what and where interactions are occurring can provide valuable information to understand gene regulation and prioritize non-coding disease variants. Regions of chromatin that preferentially interact with one another form topologically associated domains (TADs) (Dixon et al. 2012). Chromatin within a TAD will frequently form physical interactions with

chromatin within the same TAD, but will unlikely interact with chromatin in a different TAD (Phillips-Cremins et al. 2013; Nora et al. 2012). TAD boundaries are created by cohesin and the CCCTC-binding factor (CTCF), a transcriptional repressor (Cubeñas-Potts and Corces 2015). They work to regulate gene expression by bringing enhancer elements and promoters together to facilitate interactions. Insight into the genomic 3D landscape can provide evidence for the role of non-coding variants and better assess potential disease risk variants. Various techniques have been created to assay physical interactions, the first being chromatin conformation capture (3C) (Dekker et al. 2002). 3C is used to uncover interactions between regulatory regions and promoters by crosslinking the DNA within a cell followed by DNA digestion and re-ligation. Regions of chromatin that form physical interactions will come together during crosslinking and join during re-ligation to produce fragments that can be examined using qPCR and sequencing. Variations of this method have also been introduced, including Circular 3C (4C) and Carbon-copy 3C (5C) (Werken et al. 2012; Dostie et al. 2006). These methods are highly useful for investigating cis-acting regulatory variants, however, for diseases like AS with increasingly large numbers of variants to investigate, they can be costly and tedious, sometimes lacking the resolution necessary. They are limited by their ability to only examine one region of interest at a time and, as such, are not very high-throughput.

Hi-C was developed as a major advancement on these techniques as it implements NGS to find interactions occurring across the whole genome, making it much higher throughput, and has been used for mapping interactions such as TAD boundaries (Belton et al. 2012; Lieberman-Aiden et al. 2009). However, due to the large area being assayed and sequencing costs, it can be low resolution and miss interactions (Zhang et al. 2018). Hughes et al. established Capture-C, a new method similar to Hi-C, but utilizes oligonucleotide capture technology to improve throughput and resolution (Hughes et al. 2014). In Capture-C, a 3C

library is generated by crosslinking the DNA and digesting it with a restriction enzyme, followed by re-ligation and addition of adapters. Specific baits are designed and then used for an oligonucleotide capture to pull out regions of interest with a known sequence. Once the capture is complete the library is amplified and sequenced to visualize cis-interactions between areas of interest and other regions of chromatin. In a single assay, many cis interactions can be mapped to functionally annotate GWAS risk variants.

Smemo et al. was able to successfully implement chromatin conformation capture technology in the study of obesity (Smemo et al. 2014). The gene Alpha-Ketoglutarate Dependent Dioxygenase (*FTO*) encodes an enzyme important in body weight and metabolism maintenance and variants within its introns have been associated with obesity in various GWAS studies. Smemo et al. performed a 3C experiment in mice to look for interactions between these variants and *FTO* expression, but found that the intronic variants were interacting with the transcription factor, *Irx3*, located over 500kb away. Following further investigation, it was found that Irx3 silencing resulted in a 30% reduction in body weight and the variants correlated with expression levels of *IRX3* in human samples.

1.6.3 Defining Histone Modifications

Post translational modifications of histones can alter DNA accessibility by condensing it into heterochromatin, or opening it up into euchromatin. Heterochromatin can be constitutive or facultative and change based on the cell's function or needs. Acetylation, methylation, phosphorylation, and ubiquitination are all the result of chemical modifications moderated by enzymes to change nucleosome structure and change chromatin exposure (Rothbart and Strahl 2014; Bannister and Kouzarides 2011). Chromatin immunoprecipitation and DNA sequencing (ChIP-Seq) is an assay designed to examine the genomic

position of post translational modifications of histones as well as binding of transcription factors and other proteins. In ChIP-Seq, samples are crosslinked to preserve DNA-protein interactions, followed by sonication and selection using an antibody for the protein in question (Schmidt et al. 2009). NGS is then used to reveal the sequence of DNA where the particular protein is bound. The type of histone modification present at chromosomal location provides clues as to the function of that segment of DNA, including whether it is a regulatory element, promoter, or coding region. The enrichment of H3K9ac and/or H3K27ac is indicative of an active enhancer element while H3K4me3 suggests promoter activity (Hon et al. 2009; Creyghton et al. 2010; Heintzman et al. 2007). Constitutive or facultative heterochromatin can be assessed by the presence of H3K9me3 and H3K27me3, respectively (Hansen et al. 2008; Bannister et al. 2001). Enhancers are marked with H3K4me1, regardless of their activation state, and repressors are enriched for H3K27me3 (Hon et al. 2009; Hansen et al. 2008; Heintzman et al. 2007; Pan et al. 2007). All of these histone modifications provide evidence for how a segment of chromatin can influence expression and shape the epigenetic chromatin landscape. Co-localization of histone markers and GWAS hits can be an indicator of a variant's function and how it may modulate gene expression changes, making ChIP-Seq for histone modifications a valuable tool. ChIP-Seq can also be used to study transcription factor binding to assess how a variant may have altered a binding motif, or to find CTCF binding regions to provide clues as to TAD boundaries and chromatin interactions.

1.6.4 Assaying Gene Expression and Mapping eQTLs

Knowing the relationship between a variant and gene expression is vital for understanding its mechanism of action and the role it may play in disease pathogenesis. Genes that are differentially expressed between patients and healthy controls provide valuable insight into disease aetiology and help

characterize causal GWAS variants. Over the years there have been a multitude of methods developed to study gene expression, most of which focused on measuring mRNA levels. Initial assays included northern blots, qPCR, and Sanger sequencing on complimentary DNA (cDNA) (Kukurba and Montgomery 2015; Adams et al. 1991). However, all these methods are low-throughput and limited in their capabilities. Improvements on these technologies include hybridization-based microarray technologies. RNA microarrays enable transcriptome analysis by allowing more samples and genes to be analyzed in a cost effective manner (Schena et al. 1995). However, this type of assay lacks the necessary sensitivity to quantify genes that are very lowly or highly expressed and cross reactions between similar sequences can occur due to the hybridization method (Casneuf et al. 2007). It is also limited by the fact that to develop a microarray, prior knowledge of sequences are required. To overcome these challenges, RNA-Seq was developed, which uses NGS on cDNA to interrogate the entire genome (Wang et al. 2009). It is better able to quantify varying levels of gene expression and provide information on allele-specific expression and alternative splicing, as well as non-coding RNAs.

A specific polymorphism that associates with differential gene expression is referred to as an expression quantitative trait locus (eQTL), and understanding them can be used to better functionally annotate GWAS hits (Battle and Montgomery 2014; Rockman and Kruglyak 2006). eQTL mapping combines genotyping and RNA-Seq data to establish associations between variants and gene expression levels and can detect both cis- and trans-interactions. eQTL studies have shown them to be sometimes context specific, making mapping difficult (Westra and Franke 2014). Fairfax et al. noted differential expression between B cells and monocytes (Fairfax et al. 2012). They found, for example, that the presence of the SNP, rs2223286, was associated with an increased expression of the gene, *SELL*, in B cells, but the same allele showed a reduced

expression in monocytes. Another study also examined the effect of context specific eQTL and found eQTLs specific to either T cells or monocytes (Raj et al. 2014). Crucially, when studying disorders of the immune system it is important to not only consider cell type, but also activations states from stimuli and other factors that affect cell behavior in an immune setting. eQTLs can vary based on cell environment and exposures, for example the presence of LPS. Fairfax et al. also uncovered variants in monocytes that only appear as an eQTL in a LPS stimulated state (33.4%), further emphasizing the importance of examining variants under the relevant conditions (Fairfax et al. 2014). A database for tissue eQTLs has been generated called the Genotype-Tissue Expression (GTEx) Portal as a resource for genetic studies and may improve our understanding of context specific eQTLs (Lonsdale et al. 2013; GTEx Consortium 2015). They compiled WGS data and RNA-Seq data from 960 individuals across multiple tissue types including skin, brain and lungs to better understand the effect of genetic variation on gene expression. Extensive eQTL mapping could help evaluate GWAS hits and lead to pinpointing causal disease variants.

1.7 Functional Validation of Genetic Variants

There is often little experimental evidence for understanding how GWAS variants may change gene expression to cause disease. The use of the previously described techniques on relevant cell lines can be predictive of a variant's function, but without experimental validation there is no proof as to its mechanism of action or ability to cause disease. Reporter assays have been a popular method of distilling the function of putative regulatory regions of DNA (Dailey 2015). They commonly involve the presence of an inducible reporter gene, such as green fluorescent protein (GFP) or luciferase, contained in a plasmid with the sequence in question. Plasmid based assays have been used to examine potential promoters, enhancers, and repressors in a variety of cell types, but they have their

limitations. They are not necessarily an accurate representation of the cellular context in which the regulatory elements are normally present, and therefore may respond differently. To combat this, genome editing techniques have been developed to assay regulatory elements and variants of interest in their native context.

1.7.1 Genome Editing

Genome editing enables specific sequence changes to be made within a cell line or animal model for the functional characterization of variants and regulatory regions. These changes can be the introduction of a single specific variant or sequence, deletion of a region, or non-specific INDELs to disrupt a sequence. Many genome editing technologies rely on DNA repair mechanisms by introducing double-strand breaks into a region of interest. Once a double-strand break occurs, the cell will try to repair the damage, either though non-homologous end joining (NHEJ) or by homology directed repair (HDR) (Wyman and Kanaar 2006). In NHEJ the blunt ends of the DNA are joined together regardless of sequence, which can sometimes leads to INDELs being generated. This mechanism can be used in genome editing to generate deletions and disrupt genes by introducing frameshift or nonsense mutations. If a homologous sequence is present, a cell will repair a double-strand break by HDR, in which it recognizes sequence similarities with an introduced DNA segment and uses it as a template to repair the damage. This can be valuable for introducing specific mutations by providing a template harbouring the desired variant(s) along with a DNA cleavage element.

One of the first major technologies developed for genome editing was zinc-finger nucleases (ZFNs). ZFNs are synthetic restriction enzymes that have a DNA binding domain specific to the targeted area and a DNA cleavage domain to produce a double-strand break (Carroll 2011). ZFNs utilize the non-specific

cleavage domain of the Fok1 enzyme, which was shown to have no sequence specificity in its cleavage domain and could therefore be engineered to find and cut specific sequences by changing the DNA-binding domain (Li et al. 1992; Kim and Chandrasegaran 1994; Kim et al. 1996). Transcription activator-like effector nucleases (TALENs) are another class of site-specific nucleases used in gene editing, designed to improve upon ZFN technology. Transcription activator-like effectors (TALEs) are proteins found naturally in pathogenic bacteria that are secreted and bind DNA to affect transcription in the host cell (Gaj et al. 2013). TALEs are composed of repeat domains with 33-35 amino acids which each recognize a base pair and can be modified to bind a region of interest.

1.7.2 CRISPR/Cas9 Mediated Genome Editing

The most recent advance in genome editing is CRISPR/Cas9 technology. CRISPR is a defense mechanism used by bacteria and archaea against invading pathogens (Wiedenheft et al. 2012; Javed et al. 2018). Upon invasion of a bacterial cell by a viral pathogen, the bacterium will uptake segments of the viral DNA into its own to use as a tracker and will recognize and cleave the viral DNA upon subsequent infection to destroy the pathogen. This mechanism has now been exploited by scientists for site-specific genome editing. It is composed of a guide RNA (gRNA), which is made up of a crRNA (CRISPR RNA) and a tracrRNA (trans-activating CRISPR RNA), and a DNA cleavage enzyme, CRISPR associated protein (Cas9) (Javed et al. 2018). The 5' end of the gRNA is complimentary to a specific sequence, roughly 20 nucleotides long, and can be modified to recognize different targets. The gRNA will complex with Cas9 and bind to the desired genomic location. Once there, the Cas9 will recognize a protospacer adjacent motif (PAM) sequence and create a double-strand break three basepairs upstream. The most commonly used PAM sequence is NGG, compatible with Cas9 from *Streptococcus pyogenes*, but other sequences such as NGA are used with different Cas enzymes (Kleinstiver et al. 2015). Once the double-strand break is introduced, the cell will

repair the damage via HDR with the use of a template, or by NHEJ (Figure 1.5). Cas9 enzymes can also be modified and used as a nuclease to cleave both strands of DNA, or as a nickase to cleave just one to reduce off-targets (Gopalappa et al. 2018).

Successes and Shortcomings

The CRISPR/Cas9 system is preferable to the aforementioned ZFNs or TALENs for several reasons. The major advantage held by CRISPR/Cas9 is its ease of use and design. Only a handful of labs in the world have been able to generate efficient ZFNs, but CRISPR/Cas9 use is widespread (Javed et al. 2018). For each target sequence, a new ZFN or TALEN protein needs to be synthesized, but with CRISPR/Cas9, only the gRNA needs to be modified and can be administered into a cell as a plasmid or as complexed with Cas9 as a ribonucleotide protein complex (Javed et al. 2018). The gRNAs for CRISPR/Cas9 also recognize a shorter DNA sequence than ZFNs or TALENs, making design easier and allowing more regions to be targeted (Cradick et al. 2013). Moreover, multiple gRNAs can be used simultaneously to introduce more than one double-strand break, which is useful in deletion generation (Jiang et al. 2015; Mans et al. 2015). CRISPR/Cas9 is not without limitations, however. Like other editing techniques, CRISPR/Cas9 can also produce undesirable off-target effects and gRNAs can be difficult to design for certain genomic loci, such as repeat regions (Gaj et al. 2013). The cell type being used is also a major factor in any genome editing experiment. Some cells are easily transfectable, while some refuse to uptake foreign DNA, or die in the process. AS relevant cells, such as monocytes, macrophages, and T cells, are desirable for understanding disease pathophysiology, but they are not easily transfectable and therefore represent a challenge for CRISPR/Cas9 experiments.

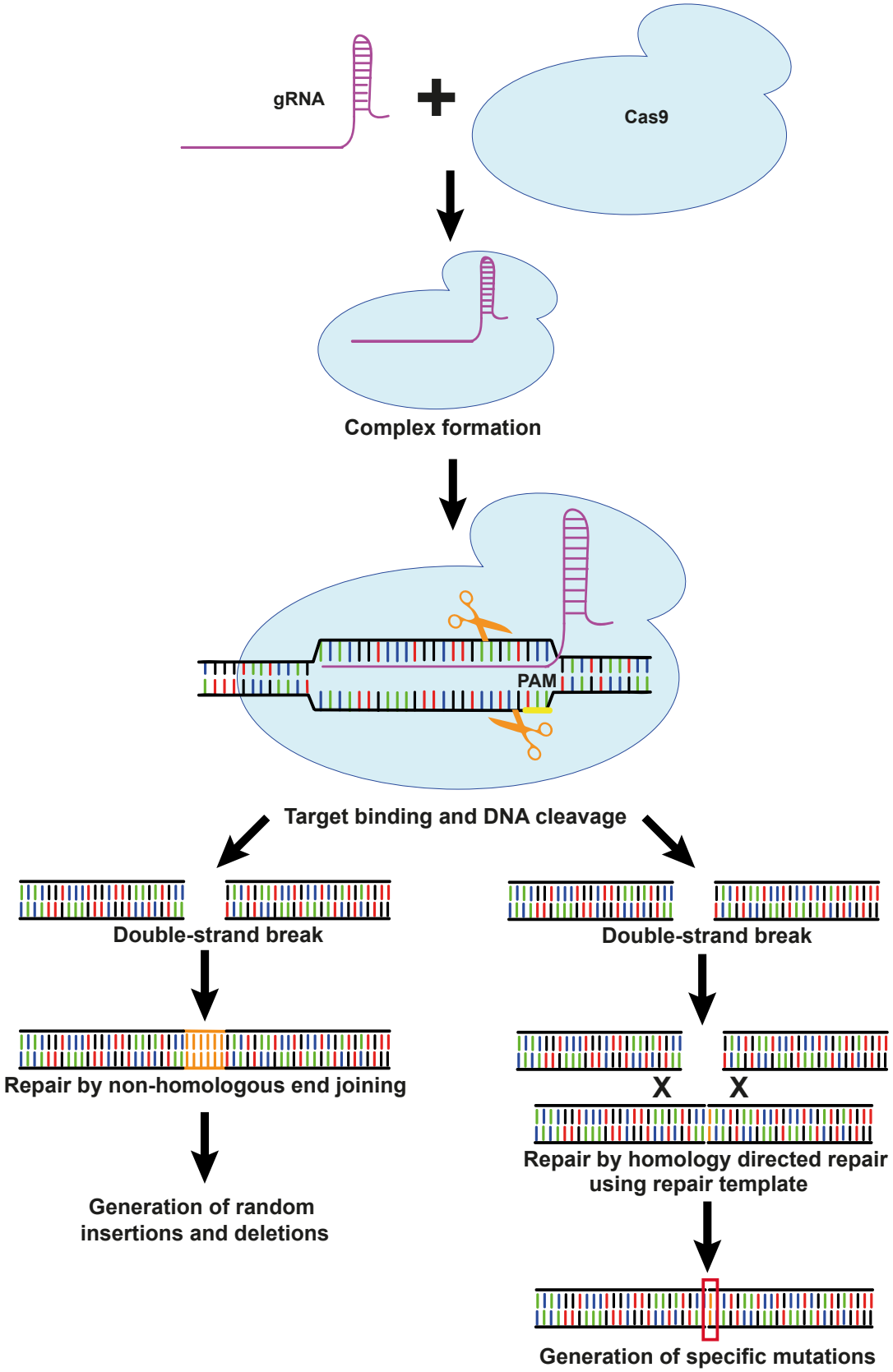


Figure 1.5: Overview of the CRISPR/Cas9 technique.
CRISPR/Cas9 can be used to introduce indels through non-homologous end joining and specific variants through homology directed repair with the use of a repair template.

CRISPR/Cas9 to Validate GWAS Hits

Functional genomic techniques such as ATAC-Seq and ChIP-Seq can be used to develop hypotheses about GWAS variants, which can then be tested using CRISPR/Cas9 to identify causal variants. Soldner et al. have used this approach successfully in the study of Parkinson's disease (Soldner et al. 2016). Using CRISPR/Cas9 to generate multiple SNPs associated with disease, the authors discovered a risk variant, rs356168, located in an enhancer element responsible for the regulation of the gene Alpha-Synuclein (*SNCA*). *SNCA* had already been associated with Parkinson's disease, but factors influencing its expression were unknown. Both protective and risk alleles were uncovered for rs356168. The A allele allows for the binding of the transcription factors, EMX2 and NKX6-1, to regulate expression levels of *SNCA*, while the G allele alters the transcription factor motif to inhibit binding and up-regulate gene expression. Without CRISPR/Cas9, the researchers could not have discerned the causal SNP as it was in LD with other potential risk variants, therefore highlighting the significance of this new technology. Experimental validation of GWAS SNPs can be a challenge if there are too many variants to examine, even with fine mapping approaches. The methods employed by Soldner et al. in Parkinson's disease are highly informative, but are also low-throughput, which has led to the development of high-throughput ways to functionally characterize regulatory regions and causal disease variants.

CRISPR/Cas9 Screens

Genome wide CRISPR/Cas9 screens have been rapidly gaining momentum as a high-throughput way to examine variants. Within a single experiment they allow for the generation of thousands of variant cells, each harbouring a knock-out of a different gene. CRISPR/Cas9 screening experiments begin with designing a library of gRNAs for a specific set of genes or loci, or can be done genome-wide

(Shalem et al. 2014). The library is then packaged into viral DNA to transduce cells at a low multiplicity of infection (MOI) (Wang et al. 2014). This ensures that only one gRNA will be present in a single cell. Selection markers, such as a drug, are then used to select cells with the desired phenotype, followed by NGS to reveal the gRNA enrichment for the cells of interest. CRISPR/Cas9 screens have proven initially successful in understanding the epigentic landscape in cancer. A recent study used a CRISPR/Cas9 screen to examine non-coding regions and their influence on melanoma drug resistance (Shalem et al. 2014). They designed gRNAs surrounding three known vemurafenib drug resistance genes to investigate regulatory regions in these loci. Vemurafenib is a commonly used melanoma drug and works by inhibiting BRAF, which is mutated in melanoma cells. The researchers transduced cells with their gRNA library and then selected cells that survived vemurafenib treatment. An enrichment for gRNAs surrounding the *CUL3* gene was found in cells that exhibited vemurafenib resistance, suggesting the potential importance of *CUL3* and proximal regulatory regions in melanoma drug resistance.

Alternative CRISPR/Cas9 Approaches

Recently, Cas9 is being used beyond its conventional use as a nuclease and is being modified in alternative types of CRISPR/Cas9 screens. CRISPR activation (CRISPRa) is a new technique to uncover and characterize putative enhancers by fusing Cas9 to transcriptional enhancers, such as VP64, a viral protein used for activation of transcription. Enhancers can be challenging to find because they often exhibit context specificity and require certain conditions to be active. Simeonov et al. utilized CRISPRa and uncovered novel elements within the *IL2RA* super-enhancer (Simeonov et al. 2017). These elements are involved in the inducible expression of *IL2RA* and would have required experimentation with different conditions or stimuli to be found without the use of CRISPRa. CRISPR interference (CRISPRi) is a converse method implemented in genome-

wide screens. CRISPRi also utilizes a modified Cas9 that is catalytically inactive, but will cause steric hindrance wherever the gRNA is bound, thus inhibiting gene transcription (Larson et al. 2013). Additionally, inactive Cas9 proteins can be modified to methylate or deacetylate histones at a region of interest to alter chromatin accessibility and gene transcription (Lei et al. 2017; Kwon et al. 2017).

1.8 Specific Aims and Objectives

The aims of this thesis are to:

1. **Determine the suitability of a CRISPR/Cas9 edited iPSC derived macrophage model to investigate AS GWAS relevant enhancers (Chapter 3).**

   GWAS studies are limited in their informativeness of disease associated variants and genes. To fully understand how associated loci contribute to disease pathology and undercover the underlying genetic aetiology of AS, GWAS associated variants or genes must be characterized in a relevant biological system. To do this, loci of interest must be generated in an appropriate model system. This work aims to:

   - Prioritize AS GWAS hits using in silico tools to find initial targets for functional characterization.
   - Utilize CRISPR/Cas9 in a cell line to generate deletions of validated enhancers to determine the suitability of this approach to generate and screen AS associated loci.
   - Utilize CRISPR/Cas9 in iPSCs to generate deletions of putative enhancers, spanning SNPs that have been associated with AS through GWAS.
   - Differentiate of CRISPR/Cas9 edited iPSCs into macrophages to assess the impact of the deleted enhancers using functional genomic assays.

2. **Characterize two putative enhancers within the AS associated *TNFRS1A* locus by utilizing functional genomics techniques on the CRISPR/Cas9 edited iPSC derived macrophages generated (Chapter 4).**

   Functional genomics can be used to assess chromatin accessibility, histone modifications, chromatin interactions, and other genetic and epigenetic traits to characterize GWAS associated loci. Once GWAS associated variants or genes are generated in a model system, they must be assessed to establish a causal relationship with disease. This work aims to:

   - Perform gene expression profiling on CRISPR/Cas9 edited iPSC derived macrophages to determine the effect of putative enhancers on gene expression.
   - Assess the impact on chromatin accessibility of the removal of two putative enhancers in CRISPR/Cas9 edited iPSC derived macrophages.
   - Determine H3K27ac enhancer histone mark locations and the impact of the removal of two putative enhancers in edited iPSC derived macrophages.
   - Establish the function of the putative enhancers by integrating generated functional genomic datasets.

3. **Understand the epigenetic landscape in $CD4^+$ and $CD8^+$ T cells, and $CD14^+$ monocytes, and how it changes between AS patients and healthy controls to define disease associated regulatory elements (Chapter 5).**

   Annotation of GWAS variants requires knowledge of gene expression, chromatin accessibility, and histone modifications, and how they change between patients and healthy individuals. These functional genomic datasets can be integrated to hypothesize the function of a GWAS variant and its role in disease pathology. This work aims to:

- Identify regions of open chromatin in primary immune cell types, and determine which are differential between AS patients and healthy controls.
- Map the landscape of histone marks H3K27ac and H3K4me3, that are indicative of active promoters and enhancers, and determine which are differential between AS patients and healthy controls in primary immune cells.
- Integrate evidence for regions of differentially open chromatin, and regions of H3K27ac and H3K4me3 with differential gene expression and AS GWAS data to annotate and generate hypotheses about AS associated loci.

2

Materials and Methods

If we knew what it was we were doing, it wouldn't be called research, would it?

-Albert Einstein

2.1 Cell Culture 50
2.2 Macrophage Differentiation 52
2.3 CRISPR/Cas9 Mediated Genome Editing 55
2.4 Expression Analysis 66
2.5 Functional Genomic Assays 68
2.6 Additional Analysis 74

2.1 Cell Culture

2.1.1 K562 Cell Culture

K562 cells (Catalog# 89121407, Sigma-Aldrich) were cultured in suspension in T75 flasks (Catlog# 658175, Greiner) in RPMI 1640 growth medium (Catalog# R0883, Sigma-Aldrich) supplemented with 10% FBS (Catalog# F6178, Sigma-Aldrich), 2mM L-glutamine (Catalog# G7513, Sigma-Aldrich), 100U/mL penicillin and 0.1mg/mL streptomycin (Catalog# P0781, Sigma-Aldrich). They were grown at 37ºC in 5% CO_2. Once cells reached a density of approximately $1x10^6$ cells/mL they were split using a 1:4 dilution in fresh media. Media was completely changed once every 7 days by centrifugation at 400g for 5 min at room temperature.

CRISPR Cas9 K562/Cas9-hyg Stable cell line cells (Catalog# SL552, GeneCopoeia)

were cultured in T75 flasks (catlog# 658175, Greiner) in RPMI 1640 growth medium (Catalog# R0883, Sigma-Aldrich) supplemented with 10% FBS (Catalog# F6178, Sigma-Aldrich), 2mM L-glutamine (Catalog# G7513, Sigma-Aldrich), 100U/mL penicillin and 0.1mg/mL streptomycin (Catalog# P0781, Sigma-Aldrich). Cells expressing Cas9 were selected for by using 100µg/ml hygromycin (Catalog# H3274, Sigma-Aldrich) in the media. Hygromycin was removed from the media 24 hours before transfection. They were grown at 37°C in 5% CO_2. Once cells reached a density of approximately $1x10^6$/mL they were split using a 1:4 dilution in fresh media. Media was completely changed once every 7 days by centrifugation at 400g for 5 min at room temperature.

2.1.2 iPS Cell Culture

The iPSC line used for this project was SFC840-03-03 (AKA AH017-13). It has been previously published (Fernandes et al. 2016) and is deposited in EBiSC https://cells.ebisc.org/. It was derived from dermal fibroblasts from a male healthy donor recruited through the Oxford Parkinson's Disease Centre: participants were recruited to this study having given signed informed consent, which included derivation of iPSC lines from skin biopsies (Ethics Committee: National Health Service, Health Research Authority, NRES Committee South Central, Berkshire, UK) (REC 10/H0505/71). iPSCs were cultured in 6-well plates, pre-coated with Matrigel (Catalog# 356234, Corning), in mTeSR media (Catalog# 05850, STEMCELL Technologies) containing 100U/mL penicillin and 0.1mg/mL streptomycin (Catalog# P0781, Sigma-Aldrich). Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) was added during the thawing of the cells to prevent apoptosis. They were grown at 37°C in 5% CO_2 and underwent a full media change everyday. The cells were passaged every two days using TrypLE (Catalog# 12604013, Gibco). This work was carried out at the James Martin Stem Cell Facility at the Sir William Dunn School of Pathology.

2.2 Macrophage Differentiation

2.2.1 iPSC Derived Macrophage Culture

Macrophages were differentiated at the Dunn School of Pathology under the supervision of Sally Cowley following the protocol laid out by Wilgenburg et al (Wilgenburg et al. 2013). Ten cell lines were differentiated: three Intronic Enhancer Deletion clones, two Intergenic Enhancer Deletion clones, two Intronic Enhancer CRISPR controls, two Intergenic Enhancer CRISPR controls, and one wild-type parental control (AH017-13/SFC840-03-03). A summarized version of the method is as follows. iPSCs were cultured for seven days. Approximately $4.0x10^6$ cells were then plated into wells on an AggreWell plate (Catalog# 34815, STEMCELL Technologies) which allow for small clumps of cells to form and become embryoid bodies (EBs) (Figure 2.1). They were fed daily with E8 medium (Catalog# A1517001, Gibco) supplemented with 0.1mg/mL streptomycin (Catalog# P0781, Sigma-Aldrich), 50ng/mL BMP4 (Catalog# PHC9531, Gibco), 50ng/mL VEGF (Catalog# PHC9391, Gibco), and 20ng/mL SCF (Catalog# PHC2111, Gibco). After a week in the AggreWell plate, the EBs were transferred to T175 flasks for maturation into macrophage precursor producing factories. The flasks were fed weekly with supplemented XVIVO-15 (Catalog# BE04-418F, Lonza) medium containing 25ng/mL IL-3 (Catalog# PHC0033, Gibco), 0.055nM β-mercaptoethanol (2-ME) (Catalog# 31350-010, Gibco), and 100ng/mL M-CSF (Catalog# PHC9501, Gibco). In two to three weeks after the EBs were transferred, they became fully mature, adherent factories producing $CD14^+$ macrophage precursors (Figure 2.1). These cells were harvested by removing supernatant from the flasks every week for 15 weeks. The precursor cells were set up for further differentiation. They were plated at a density of $1.5x10^6$ cells per well in a 6-well plate. They were grown at 37^oC with 5% CO_2 in X-VIVO media (Catalog# BE04-418, Lonza) containing

2nM Glutamax (Catalog# 35050, Gibco), 0.055nM 2-ME (Catalog# 31350-010, Gibco), and 100ng/mL M-CSF (Catalog# PHC9501, Gibco). They were left to differentiate for one week with a 50% media change on day 3. A visual overview of the CRISPR/Cas9 and macrophage differentiation protocol can be seen in Figure 2.2. This protocol produces primitive macrophages that differ from blood derived macrophages (Buchrieser et al. 2017).

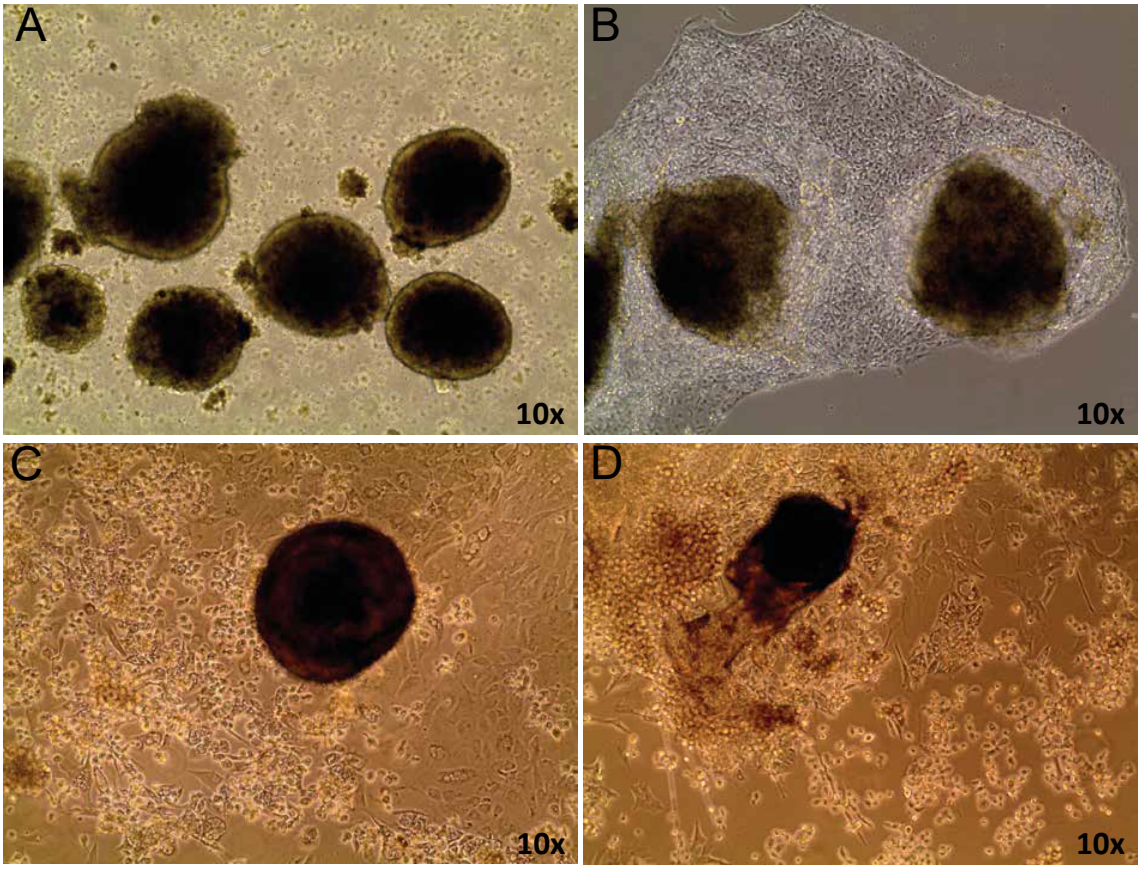


Figure 2.1: Images of iPSCs undergoing macrophage differentiation.
A. Embryoid bodies from cell line SFC-856-03-04. **B**. Early factories (1 week) from cell line AH017-13 showing adhered embryoid bodies and stromal projections. **C**. Mature factory from cell line OX1-61 showing adhered embryoid body, stromal projections, and macrophage precursors in suspension. **D**. Mature factory from cell line OX1-61 showing adhered embryoid body, stromal projections, and macrophage precursors in suspension.

2.2.2 LPS and IFNγ Stimulation

After the 7 day final differentiation step, iPSC derived macrophages were stimulated to mimic an immune response. Cells were stimulated with either

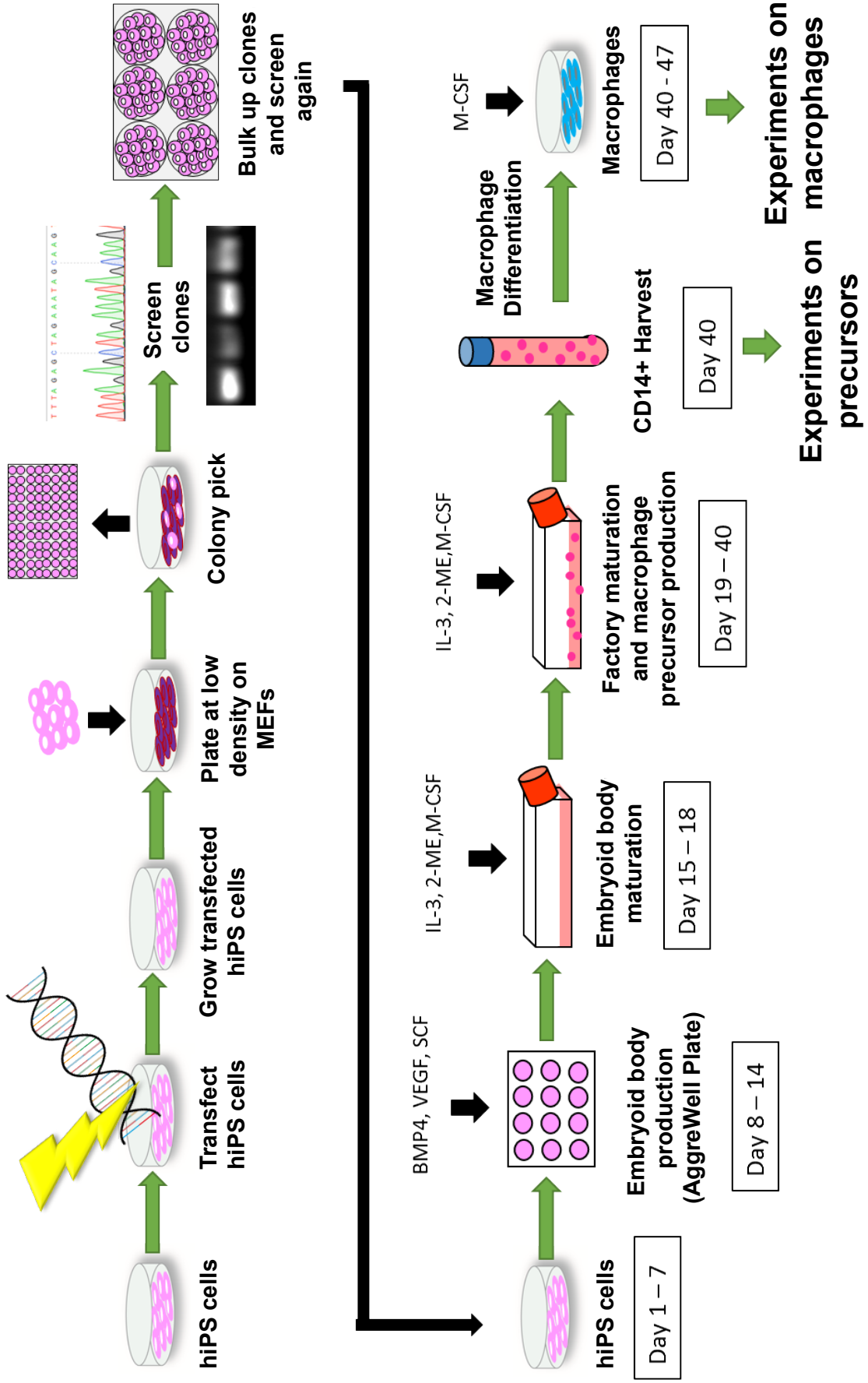


Figure 2.2: Overview of the CRISPR and macrophage differentiation process.

100U/mL IFNγ (Catalog# 285IF100, R&D), 100ng/mL LPS (Catalog# L2762, Sigma-Aldrich), or a combination of both 100U/mL IFNγ and 100ng/mL LPS. Stimulated cells were left for 48 hours and then lysed with RLT or RIPA buffer or detached using PBS containing 5mM EDTA.

2.2.3 Validation of Macrophage Phenotype

FACS analysis was performed on each cell type differentiated into macrophages (Intronic Enhancer Deletion clones, Intergeinc Enhancer Deletion clones, CRISPR controls, and parental control). Approximately 1.5x10^6 cells from each cell line were detached using PBS containing 5mM EDTA and stained for analysis using antibodies for CD14 and CD11b, or their respective isotype controls. FACS staining was carried out using the following antibodies and controls: PE Mouse anti-human CD14 (HCD14) (Catalog# 325605, Biolegend), PE Mouse IgG1 K Isotype Control (MOPC21) (Catalog# 400113, Biolegend), AlexaFluor-647 Mouse anti-human CD11b (ICRF44) (catlog# 301319, Biolegend), and AlexaFluor-647 Mouse IgG1 K Isotype Control (MOPC21) (Catalog# 400130, Biolegend). Samples were examined at the Target Discovery Institute (TDI) with the assistance of Chwen Tay and Elena Navarro-Guerrero.

2.3 CRISPR/Cas9 Mediated Genome Editing

2.3.1 Guide RNA Design

CRISPR/Cas9 gRNAs were designed using two online tools. The CRISPR Guide Picker tool designed by Desktop Genetics (https://www.deskgen.com/landing/) was utilized, along with CRISOR by TEFOR (http://crispor.tefor.net/crispor.cgi). An approximately 100bp sequence in the desired region of the double stranded break was uploaded onto the software. Both programs selected provided options for gRNAs of around 20bp in length based on presence of a protospacer adjacent

motif (PAM) sequence of NGG, which is required for the Cas9 nuclease to cut the DNA. Both softwares ranked the gRNAs by predicted efficiency using a Doench/Fusi score (Doench et al. 2016). These scores are between 0 and 100 and the higher the number, the higher the predicted efficiency. The software also provided information on off target effects by indicating how many other sites in the genome it may bind to, including if one for more nucleotide changes occurred in the genome. The CRISPR/Cas9 gRNAs selected for this project are found below in Tables 2.1 and 2.2. The gRNAs used for the *GATA1* Enhancer deletion and the 5' *MYC* Enhancer gRNA were from Fulco et al. (Fulco et al. 2016).

gRNA Name	Sequence
GATA1 Enhancer Deletion 5'	5'-GGGCAGATAAGGGAATCAGT-3'
GATA1 Enhancer Deletion 3'	5'-GCGCTCAGGAAGAGAAGCCA-3'
RUNX1 Enhancer Deletion 5'	5'-TTGCTCAAGAGCCCCGTGAT-3'
RUNX1 Enhancer Deletion 3'	5'-TTGGCATTACAGGCCTCATG-3'
LMO2 Enhancer Deletion 5'	5'-AAGGTTTTAGGCCGAGCCTT-3'
LMO2 Enhancer Deletion 3'	5'-GCAGACAGACATGCATGCGG-3'
HBG1 Enhancer Deletion 5'	5'-AGCACTTGAGAGGTAGAAGT -3'
HBG1 Enhancer Deletion 3'	5'-TTTCACATCAATGTCTATAT-3'
MYC Enhancer Deletion 5'	5'-ATGACAAAACTACGACGTAC-3'
MYC Enhancer Deletion 3'	5'-CTCTTCGTGCACCGCGTTCA-3'

Table 2.1: Control CRISPR/Cas9 gRNAs for pilot study.
The gRNAs designed to generate deletions of known enhancers from the literature. These gRNAs were used in K562 cells.

2.3.2 Transfection

Plasmid Method

Complementary oligos of the gRNA sequences (Table 2.2) were ordered and annealed together to generate a small linker with overhangs suitable for cloning into the BbsI site of plasmid pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene plasmid# 42230) (Cong et al. 2013). These gRNAs were prepared by transforming them into chemically competent *E.coli* cells. DNA was

gRNA Name	Sequence
TNFRSF1A Exon Deletion 5'	5'-CTTAGTGAGTGGCGTGGCCT-3'
TNFRSF1A Exon Deletion 3'	5'-GGCATCTGTTGCCCAGCTAA-3'
TNFRSF1A Intron Enhancer Deletion 5'	5'-ACCACCAGGACGGAATGTAG-3'
TNFRSF1A Intron Enhancer Deletion 3'	5'-TCCCCAGGAGTGGTGCTCAC-3'
TNFRSF1A Intergenic Enhancer Deletion 5'	5'-CCCCACTCCAGCATAAGACC-3'
TNFRSF1A Intergenic Enhancer Deletion 3'	5'-TATGGGTTACTCCTGGACCC-3'

Table 2.2: TNFRSF1A CRISPR/Cas9 gRNAs.
The gRNAs designed to generate deletions of four exons, an intronic enhancer, and an intergenic enhancer in and around *TNFRSF1A* . These gRNAs were used in iPSCs.

prepared from multiple colonies using the GeneJET Plasmid Miniprep Kit (Catalog# K0502, Thermo Fisher). Colonies were sequenced to verify correct construct insertion and then underwent a midiprep to increase yield using the GeneJET Plasmid Midiprep Kit (Catalog# K0481, Thermo Fisher). gRNAs were tested by transfecting HEK293 cells (Catalog# 85120602, Sigma-Aldrich) and screening for the deletion of interested. Successful gRNAs were carried forward into iPSCs. iPSCs were transfected at the Dunn School of Pathology using a Neon Transfection System (Thermo Fisher) with the 100μl transfection kit (Thermo Fisher). Prior to transfection, iPSCs were pre-treated with Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) and lifted using TrypLE (Catalog# 12604013, Gibco). Cells were counted and 3x10^6 cells were centrifuged at 300g for 5 min and resuspended in 150μL of Buffer R and mixed with 15μg of DNA for transfection. Cells were transfected with the following conditions: 1250V/20ms/1 pulse. Transfected iPSCs were immediately placed in a 24-well plate containing fresh, pre-warmed mTeSR media (Catalog# 05850, STEMCELL Technologies) and Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) that had been pre-coated with Matrigel (Catalog# 356234, Corning). Transfected cells were selected by adding puromycin (Catalog# A1113802, Thermo Fisher) 48 hours post transfection at concentrations of 0μg/mL, 0.1μg/mL, 0.25μg/mL, and 0.5μg/mL for two days.

Ribonucleoprotein Method

gRNAs (Table 2.1) were ordered as Alt-R CRISPR-Cas9 crRNAs, from IDT for the transfection of K562 cells. K562 cells were transfected using a Neon Transfection System (Thermo Fisher) with the 10μL transfection kit (Thermo Fisher). The crRNAs and Alt-R® CRISPR-Cas9 tracrRNA (Catalog# 1072532, IDT) were mixed and heated for 5 min at 95^{o}C, followed by the addition of the Alt-R® S.p. HiFi Cas9 Nuclease (Catalog# 1081060, IDT). The RNP transfection mixture was left to incubate at room temperature for 20 min. K562 cells were counted and 2.0x10^{5} cells per transfection were centrifuged at 300g for 5 min and resuspended in Buffer R. To transfect, RNP transfection solution, 2.0x10^{5} K562 cells and Alt-R® Cas9 Electroporation Enhancer (Catalog# 1075915, IDT) were mixed, added to the 10μL tip and transfected with the following conditions: 1700V/20ms/1 pulse (Liang et al. 2015). Transfected K562 cells were immediately placed in a 24-well plate containing fresh, pre-warmed RPMI 1640 growth medium (Catalog# R0883, Sigma-Aldrich). The K562-Cas9 cells were transfected using exactly the same conditions, but no Alt-R® S.p. HiFi Cas9 Nuclease was added to the reaction.

gRNAs (Table 2.2) were ordered as Alt-R CRISPR-Cas9 crRNAs, from Integrated DNA Technologies (IDT) for the transfection of iPSCs. iPSCs were transfected at the Dunn School of Pathology using a Neon Transfection System (Thermo Fisher) with the 10μL transfection kit (Thermo Fisher). Prior to transfection, iPSCs were pre-treated with Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) and lifted using TrypLE (Catalog# 12604013, Gibco). The crRNAs and Alt-R® CRISPR-Cas9 tracrRNA (Catalog# 1072532, IDT) were mixed and heated for 5 min at 95^{o}C, followed by the addition of the Alt-R® S.p. HiFi Cas9 Nuclease (Catalog# 1081060, IDT). The transfection mixture was left to incubate at room temperature for 20 min. iPSCs were counted and 2.2x10^{5} cells per transfection were centrifuged at 300g for 5 min and resuspended in

Buffer R. To transfect, RNP transfection solution, 2.2x10^5 iPSCs and Alt-R® Cas9 Electroporation Enhancer (Catalog# 1075915, IDT) were mixed, added to the 10μL tip and transfected with the following conditions: 1250V/20ms/1 pulse. Transfected iPSCs were immediately placed in a 24-well plate containing fresh, pre-warmed mTeSR media (Catalog# 05850, STEMCELL Technologies) and Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies).

2.3.3 Clone Picking

K562 Cells

K562 cells were counted and diluted to a concentration of 1 cell/100μL in RPMI 1640 growth medium (Catalog# R0883, Sigma-Aldrich). Diluted cells were plated into each well of two 96-well plates and left to grow up. Cells were given conditioned media from K562 cells after one week to improve cell growth. After one month, surviving colonies were transferred to 24-well plates for expansion.

iPSCs

iPSCs were plated on irradiated mouse embryonic fibroblasts (MEFs) at densities of 500 cells/mL, 1000 cells/mL, and 4500 cells/mL in 2mL of hES media consisting of KnockOut™ DMEM (Catalog# 10829-018, Life Technologies), KnockOut™ SR (Catalog# 10828-028, Life Technologies), Glutamax (Catalog# 35050, Gibco), NEAA (Catalog# 11140035, Life Technologies), 2-ME (Catalog# 31350-010, Gibco), bFGF (Catalog# 4114-TC-01M, R&D Systems), with Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) in a 6-well plate. Once formed, colonies were picked using a pipette and each colony was transferred to one well of a 96-well plate containing fresh, pre-warmed mTeSR media (Catalog# 05850, STEMCELL Technologies) and Rho kinase inhibitor (Catalog# 72302, STEMCELL Technologies) that had been pre-coated with Matrigel (Catalog# 356234, Corning). An example of colony picking can be seen

in 2.3. Each plasmid transfection was repeated twice, and 96 colonies were selected for screening from each. Approximately 60 clones were selected for screening from each RNP transfection.

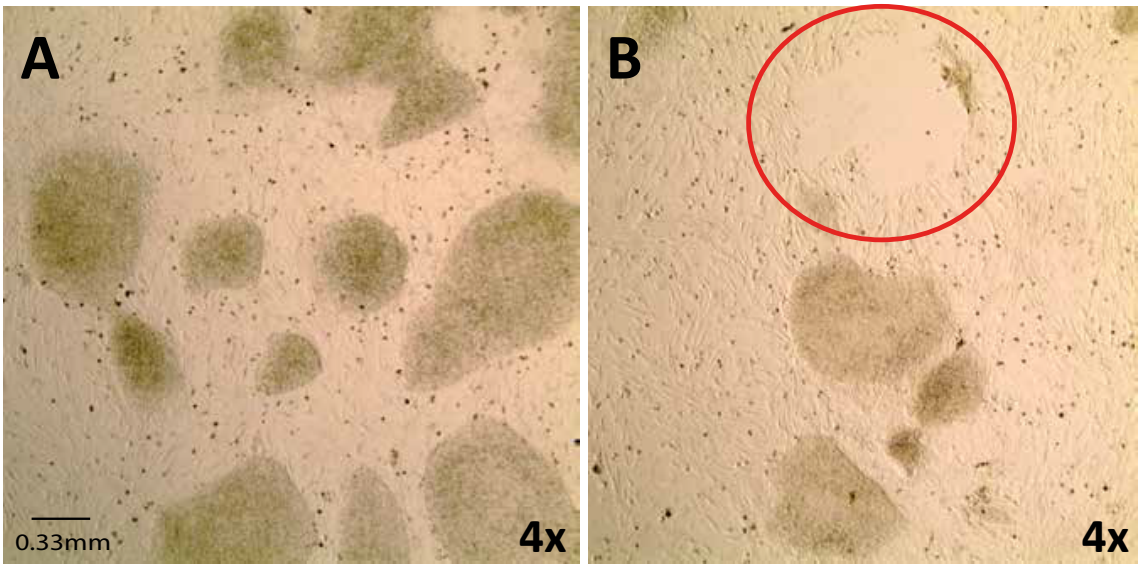


Figure 2.3: iPSC cell colonies growing on MEFs.
A. iPSC cell clones from cell line AH017-13 grown on MEFs, shown before colonies are picked. Colonies that are too large or look as if multiple colonies have merged together are not selected. **B**. iPSC cell clones from cell line AH017-13 grown on MEFs, shown after colonies are picked. The area outlined by the red circle indicates where a colony has been removed by a pipette and transferred to a well of a 96-well plate.

2.3.4 Clone Screening

DNA Extraction

DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit (Catalog# 69504) or the Qiagen AllPrep DNA/RNA Kit (Catalog# 80204), following the protocol provided. Briefly, cells were homogenized using the Qiagen QIAshredder (Catalog# 79654) and the nucleic acid was bound to a spin column, where it was washed and reconstituted in Elution buffer. Some large scale, crude DNA extractions were carried out by adding lysis buffer to a plate of colonies and precipitating the DNA using ethanol and NaCl.

Polymerase Chain Reaction

To determine which clones had incorporated the deletion of interest, PCR was used followed by visualization by gel electrophoresis. The same method was used in both K562 cells and iPSCs to determine which cells incorporated the deletion of interest. The PCR reaction was performed using Platinum Taq DNA Polymerase (Catalog# 10966026, Invitrogen) following the protocol provided by the manufacturer. The desired region was amplified by PCR using primers with binding sites on either side of the desired deletion region to generate PCR products of different sizes, depending on the presence of the deletion (Tables 2.3 and 2.4) . This size difference could then be visualized on a 1% agaraose gel. Because smaller PCR products are preferentially amplified, the zygosity of the cells cannot be determined from this PCR because one product is substantially larger. Therefore, another set of primers was designed. Primers sitting just inside the deletion region were designed to amplify the segments of DNA at either side of the deletion. These primers amplify the DNA if a cell population is heterozygous for a deletion but not if it is homozygous. These PCR products are small and similar in size, so amplify with equal efficiency. Additionally, INDELS produced as an artifact of CRISPR/Cas9 editing could encompass the original primer binding site and bias the results, therefore these additional primer sets were used. Primers with binding sites completely internal to the deletion were also designed as an extra check for homozygosity, as they would not amplify in a cell line homozygous for the deletion. A schematic showing the primer positions relative to the deletion can be seen in Figure 2.4.

In the K562 cells, only one cell line (*RUNX1* enhancer deletion) was clonally expanded and screened for homozygous cell lines, so only this cell line has primers to check for zygosity. The primer sequences used for screening K562 cells can be seen in Table 2.3. The primer sequences used for screening in iPSCs can be seen in Table 2.4.

Primer Name	Sequence
GATA1 Enhancer Deletion	**F** 5'-AAGGCTTCCGAGAAGAGAGG-3' **R** 5'-TCGAGGCATGTCTAAGAGCT-3'
RUNX1 Enhancer Deletion	**F** 5'-ATCAGGCCAGGTCAACTCTC -3' **R** 5'-TTGGACTGGCAGCATCACTA-3'
LMO2 Enhancer Deletion	**F** 5'-CCCTCTGAGTCCTGAAGTCC-3' **R** 5'-TGAACAATTAGCCGGGAAAACA-3'
HBG1 Enhancer Deletion	**F** 5'-GTCCACGTGCTTGTCTTTGT-3' **R** 5'-CTCAACCCTACAGTCACCCA-3'
MYC Enhancer Deletion	**F** 5'-GGAAGCAGGGAATTGGCATC-3' **R** 5'-ACAACACACAGCACCGAATC-3'
RUNX1 Enhancer 5' Het	**F** 5'-ATCAGGCCAGGTCAACTCTC-3' **R** 5'-CTTTCTGCGTCGCTCTTTCA-3'
RUNX1 Enhancer 3' Het	**F** 5'-AGCATCAAGAAGGGCAGGG-3' **R** 5'-TTGGACTGGCAGCATCACTA-3'
RUNX1 Enhancer Internal	**F** 5'-TGAAAGAGCGACGCAGAAAG-3' **R** 5'-CCGAAATGTTGGCCTCTCTG-3'
RUNX1 Enhancer Long Range	**F** 5'- ATCTAGGGTGCTTGCTCGTT-3' **R** 5'-AGGCCCAAGAATCTGCATCT-3'

Table 2.3: Screening primers used to examine deletion regions in K562 Cells.
Primers flanking the deletion region were used to amplify the DNA to generate PCR products of a different size to be visualized via gel electrophoresis. Cells with the correct segment of DNA removed would generate a smaller PCR product that those with the region still unedited. To examine the zygosity of the cells, two additional sets of primers were generated. One set of primers sit just within the deleted region and would amplify in a heterozygous cell line but not a homozygous one. They are labeled as 'Het' primers. Primers were also designed that bind fully within the deletion region and are labeled as 'Internal Primers'. They also will not amplify in a cell line homozygous for the deletion of interest.

Primer Name	Sequence
TNFRSF1A Exon Deletion	**F** 5'-GAGGCCTGGACAGTGACTAA-3' **R** 5'-TAAGGTGGGAGGACTGCTTG-3'
TNFRSF1A Intron Enhancer Deletion	**F** 5'-AAAGATTGGGGCGGTGTTTC-3' **R** 5'-ATCTCAGATGGGTGCTGTGG-3'
TNFRSF1A Intergenic Enhancer Deletion	**F** 5'-GGGCAGACTAGGGATGACAG-3' **R** 5'-GTGGTGGTGGAGGATGATGA-3'
TNFRSF1A Exon 5' Het	**F** 5'-ACCAGTCCAATAACCCCTGA-3' **R** 5'-TAAGGTGGGAGGACTGCTTG-3'
TNFRSF1A Exon 3' Het	**F** 5'-GAGGCCTGGACAGTGACTAA-3' **R** 5'-CTGCCATGCAGGTTTCTTTC-3'
TNFRSF1A Intron Enhancer 5' Het	**F** 5'-ACTTCCACAGATGCCTCACT-3' **R** 5'-ATCTCAGATGGGTGCTGTGG-3'
TNFRSF1A Intron Enhancer 3' Het	**F** 5'-AAAGATTGGGGCGGTGTTTC-3' **R** 5'-CTAGCCTTGGCCAGCTCTAT-3'
TNFRSF1A Intergenic Enhancer 5' Het	**F** 5'-CTCCTTCTCCTCTTTCCCCG-3' **R** 5'-GTGGTGGTGGAGGATGATGA-3'
TNFRSF1A Intergenic Enhancer 3' Het	**F** 5'-GGGCAGACTAGGGATGACAG-3' **R** 5'-CCTGCTTCTCTTGCCTCCT-3'
TNFRSF1A Exon Deletion Internal	**F** 5'-AGAGATCTCCACCTGACCCA-3' **R** 5'-GGCAGGAAGGTGTGTGTTTT-3'
TNFRSF1A Intron Enhancer Deletion Internal	**F** 5'-AAGCACGTGAACTGACCCTA-3' **R** 5'-GAGGAGCACACCAGACTCTT-3'
TNFRSF1A Intergenic Enhancer Deletion Internal	**F** 5'-CAGTCATCCGTAATACATCTGGC-3' **R** 5'-ACGACTGCTGACTTCCCTTT-3'

Table 2.4: Screening primers used to examine *TNFRSF1A* deletion region in iPSCs. Primers flanking the deletion region were used to amplify the DNA to generate PCR products of a different size to be visualized via gel electrophoresis. Cells with the correct segment of DNA removed would generate a smaller PCR product that those with the region still unedited. To examine the zygosity of the cells, two additional sets of primers were generated. One set of primers sit just within the deleted region and would amplify in a heterozygous cell line but not a homozygous one. They are labeled as 'Het' primers. Primers were also designed that bind fully within the deletion region and are labeled as 'Internal Primers'. They also will not amplify in a cell line homozygous for the deletion of interest.

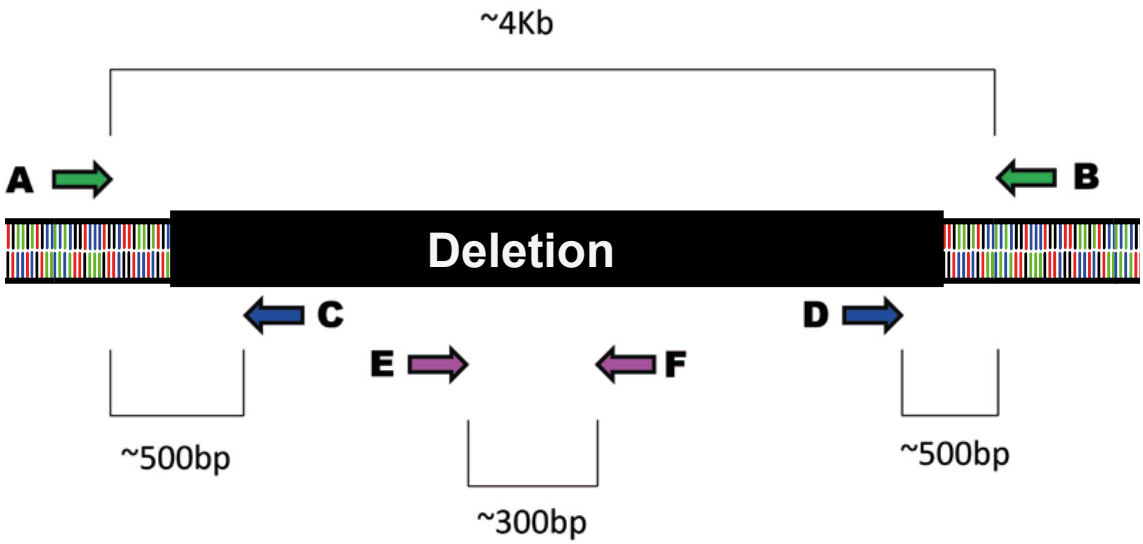


Figure 2.4: Schematic of PCR screening primers used to examine the success of deletion generation via CRISPR/Cas9.
Primers flanking the deletion region were used to amplify the DNA to generate PCR products of a different size to be visualized via gel electrophoresis. Cells with the correct segment of DNA removed would generate a smaller PCR product that those with the region still unedited. Primers 'A' and 'B' were used to amplify the entire deletion region. If the deletion had occurred the PCR product would be around 500bp and if it had not it would be between 2kb and 4kb. To examine the zygosity of the cells, an additional set of primers were designed. These primers are labeled as 'C' and 'D' and sit just within the deleted region. These are the 'Het' primers in Table 2.4. When used in conjunction with primers 'A' and 'B' they would amplify an approximately 500bp product in a heterozygous cell line and nothing in a homozygous one. Primers 'E' and 'F' bind fully within the deletion region and are 'Internal Primers'. They also will not amplify in a cell line homozygous for the deletion of interest.

Sanger Sequencing

Sanger sequencing was performed to validate that the correct deletion had occurred after CRISPR/Cas9 editing and to examine how the regions surrounding the deletion joined together. Sequencing was carried out using Mix2Seq kits from Eurofins. The primers used for sequencing were the same primers used for PCR analysis and are listed in Tables 2.4 and 2.3.

Droplet Digital PCR

Droplet digital PCR (ddPCR) was used to investigate the deletion copy number in expanded cellular clones. Samples were prepared and analyzed following the

Copy Number Variation (CNV) Analysis Protocol provided by Bio-Rad. Sample droplets were generated using the QX100 Droplet Generator (Bio-Rad), then amplified using the C1000 Touch Thermal Cycler (Bio-Rad). Visualization of sample droplets was carried out using the QX100 Droplet Reader (BioRad) and analyzed using the QuantaSoft Software (Bio-Rad). *RPP30* was used as a copy number reference for all samples. The primers and probes used to examine each deletion are listed in Table 2.5 and were designed using IDT's PrimerQuest Tool.

Region of Interest	Oligo	Sequence
TNFRSF1A Exon Deletion	F Primer	5'-CAGGAGAGACAGCAAAGTTAGG-3'
	R Primer	5'-GCAAGTGGAAACGGTGAATG-3'
	Probe	5'-AGAAGCAGAGAAAGAAGCAGCACCC-3'
TNFRSF1A Intron Enhancer Deletion	F Primer	5'-GGGTCATGAGTCAGCAGAAA-3'
	R Primer	5'-GGCCATAGGAAGAAGGGATTAG-3'
	Probe	5'-CCTGCATGCAAGTGGGACTCCAT-3'
TNFRSF1A Intergenic Enhancer Deletion	F Primer	5'-ACCAGCTCCTTCTCCTCTT-3'
	R Primer	5'-CCCTGCCACCTTAGTCTTATTC-3'
	Probe	5'-TCTCTTCCTCCTTTGGCTCCTCAGA-3'
RUNX1 Enhancer	F Primer	5'-AAACTGATTCCACTACCCCATC-3'
	R Primer	5'-CTAAATCTCCCGCCTCCAAG-3'
	Probe	5'-TGTCCCCGAAATGTTGGCCTCT-3'
RPP30 (Reference)	F Primer	5'-GATTTGGACCTGCGAGCG-3'
	R Primer	5'-GCGGCTGTCTCCACAAGT-3'
	Probe	5'-CTGACCTGAAGGCTCT-3'

Table 2.5: Primers and probes used to examine deletion copy number in CRISPR/Cas9 edited cells.
Primers and probes binding within and around the deletion region were designed to examine deletion copy number on cell colonies after CRISPR/Cas9 editing. Cell populations homozygous for the deletion would have no probe binding and no fluorescent signal, resulting in a copy number of zero. The *RUNX1* Enhancer set was used only in K562 cells and other sets were used only in iPSCs. The *RPP30* gene was used as a reference to calculate copy number in both cell lines. The fluorophores used were FAM for the deletion regions and HEX for RPP30. ZEN and IBFQ were used a quenchers.

Genome Integrity Check of Edited Cells

Genome integrity was assessed by Illumina Human OmniExpress24 BeadChip array, evaluating approximately 710,000 markers. DNA from CRISPR/Cas9 edited iPSCs was extracted using the Qiagen DNeasy Blood & Tissue Kit (Catalog# 69504). Analysis was performed using GenomeStudio and Karyostudio software from Illumina using default detection cutoffs. All edited samples were compared to their parental controls and showed no gross abnormalities indicative of effecting their growth or ability to differentiate. This assay detects genomic changes of 1MB or greater around common variants and an analysis at a higher resolution was not carried out.

2.4 Expression Analysis

2.4.1 RNA Extraction

RNA was extracted using the Qiagen RNeasy Kit (Catalog# 74104) or the Qiagen AllPrep DNA/RNA Kit (Catalog# 80204), following the protocol provided. Briefly, cells were homogenized using the Qiagen QIAshredder (Catalog# 79654) and the RNA was bound to a spin column, where it was washed and reconstituted in RNase free water. RNA was quantified with a NanoDrop (Thermo Fisher).

2.4.2 Quantitative PCR

Single strand cDNA was synthesized using Superscript III (Catalog# 18080093, Thermo Fisher), following the protocol provided by the manufacturer. cDNA was synthesized from 1μg of extracted RNA. qPCR was performed using IQ SybrGreen Supermix (Catalog# 1708880, Bio-Rad), following the protocol provided by the manufacturer. Primers used for qPCR analysis are listed in Table 2.6. The qPCR was carried out and visualized on the CFX96 Real-Time PCR Detection System (Bio-Rad). Both β-actin and *TBP* were used as reference genes

to normalize C_t values. The β-actin primer set was developed by Doerner and Zuraw 2009, and the *TBP* primer set was designed using the qPrimerDepot web site (https://primerdepot.nci.nih.gov/). The *TNFRSF1A* primer set was designed using the Primer3 tool (http://bioinfo.ut.ee/primer3-0.4.0/). The primers used in K562 cells were found in the literature and are listed as follows: *HBG1* from Ramakrishnan and Pace 2011, *LMO2* from Deleuze et al. 2007, *GATA1* and *MYC* from Fulco et al. 2016, and *RUNX1* from Markova et al. 2011. Controls prepared without the reverse transcriptase enzyme were also used to check for genomic DNA contamination.

Primer Name	Sequence	Amplicon Size
β-actin	**F** 5'-TGCGTGACATTAAGGAGAAG-3' **R** 5'-GTCAGGCAGCTCGTAGCTCT-3'	107
TBP	**F** 5'-TTCTGAATAGGCTGTGGGGT-3' **R** 5'-GATCAGAACAACAGCCTGCC-3'	131
TNFRSF1A	**F** 5'-CTGCAATTGAAGCACTGGAA-3' **R** 5'-TGGAGATCTCTTCTTGCACAGT-3'	106
GATA1	**F** 5'-TGCGGCCTCTATCACAAGATG-3' **R** 5'-CTGCCCGTTTACTGACAATCA-3'	82
RUNX1	**F** 5'-AGCCCAGGCAAGATGAGC-3' **R** 5'-GTAGGCAGCACGGAGCAG-3'	170
LMO2	**F** 5'-GCCGCCTGTCAGAAGCAT-3' **R** 5'-CACCCTCAAACCCCCAAAGT-3'	210
HBG1	**F** 5'-GGCAACCTGTCCTCTGCCTC-3' **R** 5'-GAAATGGATTGCCAAAACGG-3'	219
MYC	**F** 5'-TCCCTCCACTCGGAAGGAC-3' **R** 5'-CTGGTGCATTTTCGGTTGTTG-3'	96

Table 2.6: qPCR primers used to examine gene expression in iPSC derived macrophages and K562 cells.

.

2.4.3 RNA-Seq

Polyadenylated RNA was sequenced to examine changes in gene expression. RNA-Seq library preparation including cDNA synthesis and indexing was carried out by the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (WHG). Sequencing was performed on the Illumina HiSeq4000. RNA-Seq data analysis was carried out with the assistance of Katie Burnham. RNA-Seq reads were mapped to hg19 using STAR (Dobin et al. 2013), and sorted and deduplicated with SAMtools (Li et al. 2009) and Picard tools (Broad Institute) to generate unique reads. HTSeq-count (Anders et al. 2015) was then used to generate gene counts. Differential analysis was carried out using DESeq2 (Love et al. 2014). Significance was determined by a Wald test and the Benjamini and Hochberg method to correct for multiple testing. Genes with an FDR <0.05 were considered significant. RNA-Seq data generated from fresh $CD14^+$, $CD4^+$, and $CD8^+$ cells from AS patients and healthy controls was carried out by Anna Sanniti.

2.5 Functional Genomic Assays

2.5.1 Capture-C

Capture-C was carried out on iPSC derived macrophages that had been fixed with 2% formaldehyde and kept at -80°C following the protocol described by Davies et al. 2016. Capture-C usually requires an excess of 10 million cells, however, due to harvest numbers only around 2.5 million CRISPR edited cells were available for the experiment. Around 10 million blood-derived $CD14^+$ cells were used as a control. Biotinylated baits were designed using CapSequm from the University of Oxford (Hughes et al. 2014) based on DpnII cut sites. Each targeted DpnII fragment containing a promoter of interest had two 120bp bait oligos hybridising to the left and right ends, each starting at GATC. The locations of the genomic

regions targeted by the baits can be found in Table 2.7. The fixed cells were thawed and digested with DpnII for 24 hours, followed by proximity ligation and reverse cross-linking. DNA was then extracted and purified, and the resulting 3C libraries were sonicated on the Covaris sonicator. Illumina sequencing adapters were then added by PCR. Two rounds of capture using biotinylated baits were completed. The first was done for 72 hours followed by washing and recovery of material, and then a second capture was done for 24 hours to improve on-target capture and minimize carryover of un-baited 3C material. Samples were then prepared for sequencing.

Name	Bait Region
PLEKHG6 Promoter	chr12:6419891-6421888
TNFRSF1A Promoter	chr12:6451598 6451790
SCNN1A Promoter	chr12:6483803-6484342
LTBR Promoter	chr12:6493145-6493668
CD27 Promoter	chr12:6553414-6554570
TAPBPL Promoter	chr12:6561057-6562119
VAMP1 Promoter	chr12:6578933-6580036
NCAPD2 Promoter	chr12:6602122-6603315
TNFRSF1A rs887477	chr12:6443904-6444146
TNFRSF1A rs11064145	chr12:6455388-6456236
RUNX1 Promoter	chr21:36260222-36260588
TAL1 CTCF	chr1:47658265-47658443
LMO2 CTCF	chr11:33879368-33879846

Table 2.7: Capture-C bait regions.
Genomic coordinates of DpnII fragments targeted by Capture-C bait oligos. The coordinates stated are in UCSC notation based on genome build hg19.

Sequencing was carried out on the Illumina NextSeq500 using 150bp paired-end reads. The resulting sequences were analyzed using the CCBasicCS5 pipeline (Davies et al. 2016). TrimGalore was used to to remove Illumina adapters from the reads and forward and reverse reads were paired using FLASH (Magoč and Salzberg 2011) and in silico digested at DpnII restriction sites. These sequences were aligned to the hg19 reference genome using Bowtie (Langmead et al. 2009)

and PCR duplicates were removed. The fragments were then classified as either a capture fragment (contains the biotinylated oligo) or a reporter fragment, which represents the 'other end' of the interaction. Only interaction read pairs containing a single capture fragment and at least one reporter fragment were kept for analysis. Reporter fragment counts were normalized to the total number of reporters within each sample dataset. Fragments within 1kb around the capture fragment (exclusion fragments) were removed for clarity. Capture-C interactions were visualized on the WashU Epigenome Browser (Zhou and Wang 2012), and the UCSC Genome Browser (Kent et al. 2002). The significance of chromatin interactions was measured using Peaky (Eijsbouts et al. 2019).

2.5.2 ATAC-Seq

Omni-ATAC-Seq was carried out on fresh iPSC derived macrophages and on formaldehyde fixed $CD14^+$, $CD4^+$, and $CD8^+$ cells from AS patients and healthy controls following the protocols described by Buenrostro et al. 2013 and Corces et al. 2017. Traditional ATAC-Seq (Buenrostro et al. 2013) and Fast-ATAC-Seq (Corces et al. 2016) was carried out on fresh $CD14^+$, $CD4^+$, and $CD8^+$ cells from AS patients and healthy controls by Andrew Brown, Alicia Lledó Lara, and Anna Sanniti. In brief, 50,000 cells were used of either EDTA lifted iPSC derived macrophages or freshly thawed formaldehyde fixed human samples from 2012 and 2013. The formaldehyde fixed samples are referred to in this thesis as historic samples, due to their age. Cells were lysed for 3 min and the lysis buffer was washed out followed by a 30 min transposition reaction using the Nextera Tn5 transposase (Illumina). Samples were purified using the MinElute PCR Purification Kit (Catalog# 28004, Qiagen). Formaldehyde fixed samples underwent overnight reverse cross-linking prior to PCR purification (Chen et al. 2016b). To determine the appropriate cycle number, qPCR was used prior to the addition of adapters. Libraries were indexed using Nextera primers from Illumina and purified using the MinElute PCR Purification Kit. A library

quantification kit from Kapa Biosystems was used to quantify the final libraries. Sequencing was carried out using a high-output kit on the Illumina NextSeq500 to a depth of 35 million paired-end reads, assuming 70% usable reads to leave 25 million paired-end reads.

Omni-ATAC-Seq analysis was performed using an in-house pipeline, run with the assistance of Gabriele Migliorini, Andrew Brown, Félicie Costantino, and Alicia Lledó Lara. Sequencing reads were aligned to the hg19 genome build using Bowtie2 (Langmead et al. 2009). Picard tools (Broad Institute) and SAMtools (Li et al. 2009) were used to filter reads by finding and removing duplicated reads and poor quality reads. Mitochondrial DNA (mtDNA) reads were also removed from the final analysis. BigWig files were generated to show read pileups for visualization on the WashU Epigenome Browser (Zhou and Wang 2012) and the UCSC Genome Browser (Kent et al. 2002) using bedtools genomeCoverageBed (Quinlan and Hall 2010). Peak calling and filtering was done using MACS2 callpeak (Zhang et al. 2008) (parameters: --nomodel, --shift -100, --extsize 200, keep-dup all, call-summits) with a p-value threshold of 10^{-8} in narrow peak mode. Peaks were removed if they overlapped blacklisted features from ENCODE (https://www.encodeproject.org/annotations). The median of the summits was used in instances where more than one summit was identified for a peak. Peak summits were extended ±250bp to create a non-overlapping homogenous 500bp peak list. Peaks were merged to generate a masterlist and peaks not occurring in at least 20% of samples were not included.

Differential open chromatin analysis was carried out using DESeq2 (Love et al. 2014) in a paired designed. Reads overlapping masterlist peaks were retrieved for each sample using the HTSeq-count algorithm. Significance was determined by a Wald test and the Benjamini and Hochberg method to correct for multiple testing. Peaks with an FDR <0.05 were considered significant. Batch correction was performed using DESeq2 by including batch effect as a covariate. DESeq

batch correction was visualized on PCA by Limma. Sample quality was measured by determining the fold enrichment of ATAC-Seq reads ± 1000bp around transcription start sites (TSSs), as they are known regions of open chromatin. TSSs for hg19 were determined by Ensembl (Hunt et al. 2018). For comparisons between iPSC derived macrophages, and AS patient cells and healthy control cells, a peak was considered significantly differential if it had an FDR of <0.05. For the cell type specific differential analysis in Chapter 5, an ANOVA was run, with the assistance of Félicie Costantino, and Alicia Lledó Lara, comparing all three cell types ($CD4^+$, $CD8^+$, and $CD14^+$), followed by a pairwise comparison comparing each cell pairing. A peak had to have an adjusted p-value (Bonferroni corrected) of <0.01 in both analyses to be considered significant for that cell type.

2.5.3 ChIP-Seq

ChIPm was carried out on formaldehyde fixed iPSC derived macrophages and on $CD14^+$, $CD4^+$, and $CD8^+$ cells from AS patients, and healthy controls following the ChIPmentation protocol described by Schmidl et al. 2015. Some of the samples were processed by Andrew Brown and Connor Short. Briefly, $6x10^5$ iPSC derived macrophages were fixed with 1% formaldehyde and stored at -80°C. Historic samples were formaldehyde fixed in 2012 and 2013 using an unspecified cell number. Chromatin was thawed, aliquoted and sonicated for 8 min using the Covaris sonicator (5% duty factor) and was snap frozen and stored at -80°C. Input samples were reverse cross-linked, tagmented and stored at -20°C. ChIP Samples were given either 1μg of H3K4me3 (Catalog# C15410003, Diagenode) to investigate promoters, or 2μg of H3K27ac (Catalog# C15410196, Diagenode) to investigate enhancers. Samples were left to incubate overnight at 4°C with the antibodies. Following this incubation, protein G beads (Catalog# 10003D, Invitrogen) were added to each sample and washed 8 times in the appropriate buffers. Samples were then tagmented for 10 min at 37°C and reverse cross-linked overnight. Cleanup was carried out using the MinElute PCR Purification

Kit (Catalog# 28004, Qiagen). Sequencing of H3K27ac samples was carried out using a high-output kit on the Illumina NextSeq500 to a depth of 35 million paired-end reads, assuming 70% usable reads. Sequencing of H3K4me3 samples was performed on the Illumina HiSeq4000 by the Oxford Genomics Centre at the WHG to the same depth.

ChIP-Seq analysis was carried out using an in-house ChIPm pipeline with the assistance of Gabriele Migliorini and Andrew Brown. Sequencing reads were aligned to the hg19 genome build using Bowtie2 (Langmead et al. 2009). Picard tools (Broad Institute) and SAMtools (Li et al. 2009) were used to filter reads by finding and removing duplicated reads and poor quality reads. Mitochondrial DNA (mtDNA) reads were also removed from the final analysis. BigWig files were generated to show read pileups for visualization on the WashU Epigenome Browser (Zhou and Wang 2012) and the UCSC Genome Browser (Kent et al. 2002) using bedtools genomeCoverageBed (Quinlan and Hall 2010). Peak calling and filtering was done using MACS2 callpeak (Zhang et al. 2008) (parameters: --bw200, -p 0.1, keep-dup all, call-summits) with a p-value threshold of 10^{-8} in narrow peak mode. Uniform peak was size was adjusted to ±500bp around the peak summit. MACS2 used the average library fragment size (bw) to determine the shift parameter and to find a model to optimally represent locations of precise protein-DNA interactions. Peaks were merged to generate a masterlist and peaks not occurring in at least 20% of samples were not included.

Differential H3K27ac and H3K4me3 analysis was carried out using DESeq2 (Love et al. 2014). Significance was determined by a Wald test and the Benjamini and Hochberg method to correct for multiple testing. Peaks with an FDR <0.05 were considered significant. Batch correction was performed using DESeq2 by including batch effect as a covariate. DESeq batch correction was visualized on PCA by Limma. Sample quality was measured by determining the fold enrichment of ChIP-Seq reads ± 1000bp around TSSs, as H3K27ac and H3k4me3

mark promoters and overlap TSSs. TSSs for hg19 were determined by Ensembl (Hunt et al. 2018). For comparisons between iPSC derived macrophages, and AS patient cells and healthy control cells, a peak was considered significantly differential if it had an FDR of <0.05. For the cell type specific differential analysis in Chapter 5, an ANOVA was run, with the assistance of Félicie Costantino, and Alicia Lledó Lara, comparing all three cell types ($CD4^+$, $CD8^+$, and $CD14^+$), followed by a pairwise comparison comparing each cell pairing. A peak had to have an adjusted p-value (Bonferroni corrected) of <0.01 in both analyses to be considered significant for that cell type.

2.6 Additional Analysis

2.6.1 Fine Mapping

Fine mapping of AS GWAS loci was carried out by Anna Sanniti and Félicie Costantino using a subset of GWAS data from Cortes et al. 2013 and Ellinghaus et al. 2016. PLINK 1.9 was used to select 2MBb genomic region surrounsing each lead GWAS SNP (Chang et al. 2015) and genotypes were prephased using SHAPEIT (Delaneau et al. 2012). Missing SNPs were imputed using IMPUTE2 (Buske et al. 2011) and the 1000 Genome Project (Consortium 2015). A Bayesian approach was applied to generate a posterior probability for each SNP, based on effect size. For each association signal at a particular locus, credible sets were calculated to contain the SNPs most likely driving the association signal. SNPs were ranked based on their approximate Bayes factor and their posterior probabilities were cumulatively added until they reached 0.95 or 0.99 to determine SNPs within the 95% and 99% credible sets, respectively (Bunt et al. 2015).

2.6.2 Pathway Analysis

The pathway enrichment analysis in Chapter 5 was carried out using the web app version of the R package, XGR (http://galahad.well.ox.ac.uk/XGR) (Fang et al. 2016). The gene-based enrichment analysis was selected to identify enriched ontology terms/pathways from the gene list being examined. $CD4^+$, $CD8^+$, and $CD14^+$ cell type specific gene lists were generated from ATAC-Seq and ChIP-Seq peaks that were within 5Kb of the start or end of a gene. The gene ontology sets used for the analysis were from Canonical, KEGG, REACTOME and BioCarta pathways from the Molecular Signatures Database (MSigDB) Collections (Subramanian et al. 2005) using a hypergeometric distribution, with redundant terms removed. Genes associated with non cell type specific ATAC-Seq and ChIP-Seq peaks were used as the background set of genes, in their respective analyses.

3

Determining the Suitability of Utilizing CRISPR/Cas9 to Investigate AS Genetic Risk Loci in Disease Relevant Cell Types

I was taught that the way of progress was neither swift nor easy. -Marie Curie

3.1 Introduction 76
3.2 Results: CRISPR/Cas9 in K562 Cells 84
3.3 Results: CRISPR/Cas9 in iPSCs 98
3.4 Discussion 120
3.5 Conclusions 130

3.1 Introduction

3.1.1 Prioritization and Functional Validation of GWAS Variants

GWAS have revolutionized the field of complex trait genetics and have provided a new level of understanding to the genetic basis of disease. However with complex diseases, like AS, causative variants cannot be determined through GWAS alone. There are often many GWAS variants associated with a disease, and they lie within haplotype blocks with other SNPs, so determining the

variant driving the disease association is challenging. Researchers have been working on ways of prioritizing variants and genes including, statistics, in silico analysis, and literature reviews. Combining fine mapping results, publicly available data through ENCODE (Dunham et al. 2012) and NIH Roadmap studies (Bernstein et al. 2010), known gene interactions, gene function, and sequence knowledge can help to prioritize GWAS variants. If variants have been prioritized successfully, then there should be a manageable number of possible candidate genes or variants to take forward into a model system for validation using functional genomics. In silico analyses have been instrumental in GWAS interpretation, but the majority of AS associated GWAS variants lie within the non-coding region of the genome, with little annotation for reliable predictions (Maurano et al. 2012; Schaub et al. 2012). Even variants situated within an intron of a gene may not necessarily impact that gene and could be acting at a distance on another gene (Smemo et al. 2014). Functional genomics studies provide a way to examine risk variants in a biological system to evaluate the mechanism of action and relationship with disease. Utilizing CRISPR/Cas9 technology in a model system allows for the introduction of variants, or other desired genomic changes, to study how they may be impacting gene expression through genetic or epigenetic changes, in a causal manner. Variants associated with AS through GWAS can be narrowed down using fine mapping and in silico predictors, and then introduced in a biological model system to fully understand their effect and functionally validate their role in disease. This chapter describes the implementation and optimization of CRISPR/Cas9 to examine putative regulatory regions associated with AS.

3.1.2 CRISPR/Cas9 Delivery Methods

CRISPR/Cas9 technology is changing the way genetic studies are performed and has become a major tool in the post-GWAS era. There are several options to package and deliver the CRISPR/Cas9 system into a cell or organism. Commonly

used approaches include creating a plasmid encoding the desired gRNA and the Cas9 protein, providing an mRNA version of the Cas9 protein and a gRNA, or introducing the Cas9 as a protein with a gRNA (Lino et al. 2018). These packaging methods can be introduced into the cell physically, chemically, or with the use of a virus.

Plasmid Based Methods

Delivering the CRISPR/Cas9 complex in a plasmid is a traditional and common way to edit the genome of a cell or organism. Plasmids are constructed and cloned to encompass all the necessary components of the CRISPR/Cas9 system. This methods relies on the cellular machinery of the transfected cell to transcribe the gRNA and transcribe and translate the Cas9 protein before they can form a complex together and generate a double-strand break at the target sequence. Plasmids can be introduced in a variety of ways. Chemical transfections are popular due to their ease of use, but they are ineffective on a number of cell types (Lino et al. 2018). For difficult to transfect cells, like immune cell types, electroporation is routinely used. Electroporation allows cells to uptake nucleic acids by using electrical currents to create nanometer sized pores on the membrane of the cells. These tiny pores allow for the nucleic acids to move into the cell. Electroporation works on a variety of cell types, but can cause significant cell death if the cell line is particularly sensitive. Due to the size of the Cas9 sequence, CRISPR/Cas9 plasmids can be bulky, which is sometimes challenging for cells to uptake. Once a plasmid has been incorporated by a cell it can remain for several days and become lost or inactive at different times during cell proliferation (Gaj et al. 2012). This can increase off-target effects and sometimes result in mosaicism, where cell clones do not have identical genotypes.

Viral Vectors

The use of viral vectors as a nucleic acid delivery system is another common approach. Lentivirus and adenovirus vectors are frequently used to introduce the CRISPR/Cas9 complex into difficult to transfect cell types, and are used in large scale CRISPR/Cas9 screens (Lino et al. 2018). The virus is engineered to express the gRNA of interest and the Cas9 nuclease. It will then interact with a cell and transfer the CRISPR/Cas9 components. Viral transductions have been successful in genome editing in a variety of studies, but do have some limitations. They can stimulate an immune response, making them ineffective for delivery in some organisms and immune cell types, and can cause off-targets through genome integration and persistent expression of the CRISPR machinery (Follenzi et al. 2007; Papapetrou and Schambach 2016). Plasmid size can also be a concern in viral transductions. Sometimes a plasmid is too bulky for viral packaging and very large recombinant viruses can be difficult to purify at a sufficiently high titre. To circumvent this, researchers have generated cell lines constitutively expressing Cas9 so it can be removed from the viral vector for effective packaging and delivery.

Ribonucleoprotein Method

The ribonucleoprotein (RNP) method is gaining momentum in the field of CRISPR/Cas9 genome editing and becoming preferable to plasmids (Kim et al. 2014). In this method, the gRNA is introduced into the cell already as an RNA, and the Cas9 protein is already a translated protein. This eliminates the need for the cell to transcribe and translate a plasmid, and individually these components are much smaller and easier to uptake when compared to a plasmid. Additionally, as an RNP, the Cas9 and the gRNA are already complexed to achieve optimal convergence, which may be difficult to achieve with a translated Cas9 and transcribed gRNA. They also have an immediacy of action and a

relatively short life within a cell once delivered, which reduces off-target effects and mosaicism. Moreover, it is reported to be less stressful on sensitive cell types, such as embryonic stem cells (Kim et al. 2014). This method has even been successful in hard to transfect primary human T cells. One study was able to knockout the HIV co-receptor, *CXCR4*, and saw a 40% reduction in its expression using the RNP method (Schumann et al. 2015).

3.1.3 Model Systems for Ankylosing Spondylitis

Animal Models

There exist several model system options to study AS. There have been a number of different mouse and rat models developed and used over the years to understand AS pathophysiology. A transgenic rat model was used to provide initial evidence in the involvement of HLA-B27 in disease through over-expression of HLA-B27, resulting in inflammation (Hammer et al. 1990). These different mouse and rat models exhibit different disease features and an appropriate one can be selected depending on the desired study outcome (Zhang et al. 2002; Smith et al. 2006). However, each animal model only represents specific aspects of AS and does not fully recapitulate the human disease. Moreover, the generation of transgenic mouse and rat models can also be a lengthy and complicated process, depending on the desired genetic changes, and are not well suited for high throughput genetic studies. Studies in human tissue can provide more valuable information about disease pathophysiology.

Human Primary Cells and Cell Lines

Immortalized cell lines and primary cell culture are an alternative to complicated animal models. Immortalized human cell lines are the easiest model to work with when studying disease as they grow well and are robust under different experimental conditions. However, because many are derived from cancer they

often differ from healthy human cells by exhibiting aneuploidy, translocations, and unusual gene copy numbers, making them difficult for functional genomic studies (Pan et al. 2009; Alge et al. 2006). They do not necessarily share epigenomic features with normal human cells, and genome editing with varying copy numbers is challenging. Primary cells are the ideal model as they can be taken directly from AS patients and healthy individuals, and work well for functional genomic studies involving ATAC-Seq or ChIP-Seq. However, they are more challenging for CRISPR/Cas9 experiments. Once isolated from the blood, they are very difficult to culture and do not respond well to transfections. Many cells die in response to nucleofection or viral transductions and other transfection methods are ineffective (Peng et al. 2016; LaFleur et al. 2019). Moreover, some immune cell types struggle to uptake foreign DNA introduced using CRISPR/Cas9, as their role is to recognize and destroy it (Wilgenburg et al. 2013). Therefore, finding an appropriate cellular model for the study of AS, and autoimmune diseases, remains challenging.

Induced Pluripotent Stem Cells

Another option for the study of disease is the use of human iPSCs. iPSCs are somatic cells that have been reprogrammed to regain pluripotency, and unlike embryonic stem cells, they are easier to procure and are more ethically acceptable (Takahashi and Yamanaka 2006; Dimos et al. 2008). When derived from healthy individuals they have normal chromosome and gene numbers, and have the ability to be genetically manipulated (Jehuda et al. 2018; Guenther et al. 2010). They can also be differentiated into a variety of different cell types specific to the disease of interest. Moreover, they can be derived from AS patients and control groups to suit the study design. iPSCs can be beneficial for functionally validating AS associated GWAS variants. CRISPR/Cas9 can be used to generate desired genetic changes, including SNP generation, INDEL generation, or the creation of large insertions or deletions. Once these have been produced in

iPSCs, the cells can be differentiated into a cell type of interest to evaluate the genetic effects in a disease relevant model. This approach improves upon the issues of abnormal karytopes from immortalized human cell lines and they are easier to culture and transfect compared to human primary cells. Another benefit of iPSCs as a disease model is the ability to generate isogenic controls or isogenic mutated cell lines (Jehuda et al. 2018). This helps uncover a causative relationship between genotype and effect, rather than question if an observed effect is due to other genetic variation present beyond the region of interest. CRISPR/Cas9 can be used to create mutations of interest in a healthy cell line, or to correct genetic defects in a disease cell line to rescue a phenotype. The use of iPSCs in disease modeling has been successful in a multitude of studies and its popularity is increasing, but this approach is not without limitations (Jehuda et al. 2018). iPSCs require a lot of care and attention to culture successfully and transefction protocols need to be well optimized to avoid substantial cell death. It is also limited by differentiation protocols. Protocols for the differentiation of stem cells into certain cell types are not currently available and some require substantial time investments, and optimization.

Selected Approach

As a proof of concept project, and to establish CRISPR/Cas9 gene editing within the laboratory, the cell line K562 was selected for CRISPR/Cas9 experimentation. Regulatory regions were selected to be deleted using CRISPR/Cas9 to determine if removing these regions has any effect on the expression of regulated genes. K562 cells are of the myeloid cell lineage and are an immortalized leukemia cell line (Lozzio and Lozzio 1975). They are characterized as being multipotent hematopoietic cells that spontaneously differentiate into progenitors of erythrocyte, granulocyte, and monocyte cell lineages. They were selected because they are easy to culture and transfect, and are one of the cell lines used in the ENCODE Project Consortium (Dunham et al. 2012), so there is publicly

available data to aid in the selection of targets. Being a myeloid cell type was also desirable because of their suggested involvement in AS pathogenesis. Both traditional K562 cells and K562 cells constitutively expressing Cas9 (K562-Cas9) were used.

Following validation work with K562 cells, the differentiation of iPSCs into macrophages was the selected approach for functional validation of AS associated regions for this thesis. Macrophages are thought be a key cell type in the aetiology of AS and an important immune cell type (Section 1.3.3 in Chapter 1). The Dunn School of Pathology at the University of Oxford has published an efficient and effective protocol for the differentiation of iPSCs into macrophages and was a realistic approach for this thesis (Wilgenburg et al. 2013).

3.1.4 Aims

The overall aim of this chapter is to assess the feasibility and suitability of CRISPR/Cas9 edited K562 cells and iPSC derived macrophages as a model to investigate AS GWAS relevant enhancers. The work outlined in this chapter includes using CRISPR/Cas9 genome editing in an immortalized cell line to determine its suitability as a model system for the study of AS relevant enhancers. It also involves using CRISPR/Cas9 in iPSCs to generate deletions of putative enhancers that have been associated with AS through GWAS. Furthermore, it includes the differentiation of the edited cell lines into macrophages so these enhancers may be assessed in a biologically relevant cell type. The specific aims for this chapter are:

1. To determine the suitability of using CRISPR/Cas9 in an immortalized cell line, K562, for screening AS associated putative enhancers by generating deletions of validated enhancers and examining a change in expression of each corresponding gene, as proof of principle experiment.
2. To utilize publicly available data and in silico tools to prioritize regions

that have been associated with AS through GWAS in order to select initial regions for CRISPR/Cas9 editing and functional characterization.

3. To determine the suitability of using iPSCs as an AS disease model for CRISPR/Cas9 edits of disease relevant loci by deleting two enhancers found at the AS associated *TNFRSF1A* locus, as proof of concept for the approach.

4. To differentiate edited iPSCs into macrophages so the impact of the deleted enhancers may be assessed using functional genomic assays (Chapter 4).

3.2 Results: CRISPR/Cas9 in K562 Cells

3.2.1 K562 Cell Karyotype

Before CRISPR/Cas9 editing could begin, the karyotype of the K562 cells being used had to be ascertained. Karyotyping was carried out on K562 and K562-Cas9 cells to verify the cell line and to determine chromosome number and interchromosomal rearrangements. Karyotyping was performed using a multiplex fluorescence in situ hybridization (M-FISH) (Figure 3.1) by Daniela Moralli in Cath Green's Group at the WHG. Both cell lines were found to be essentially triploid with the majority of chromosomes present in three copies. Some chromosomes were recurrently found in just two copies or less, while others were present in four or more copies (Table 3.1). Both cell lines had a high variability in chromosome number with the modal total cellular chromosome number being 69. Moreover, several recurring chromosome rearrangements were found in most cells (Table 3.2), and acentric fragments were detected with a high frequency. This is consistent with the expected karyotype for K562 cells as they are an immortalized leukemia cell line, and differ from regular human cells.

Despite the chromosome abnormalities present in this cell line, K562 cells were still taken forward for CRISPR/Cas9 editing. Immortalized cell lines typically

have abnormal karyotypes and genetic mutations as this is what allows them to endlessly proliferate (Pan et al. 2009; Alge et al. 2006). Having four copies of chromosomes can be problematic for genome editing as more genetic material needs to be modified and more editing events need to occur. However, if the CRISPR/Cas9 system being used is efficient enough, then it should be able to produce the desired edits, even if extra copies of the locus of interest are present.

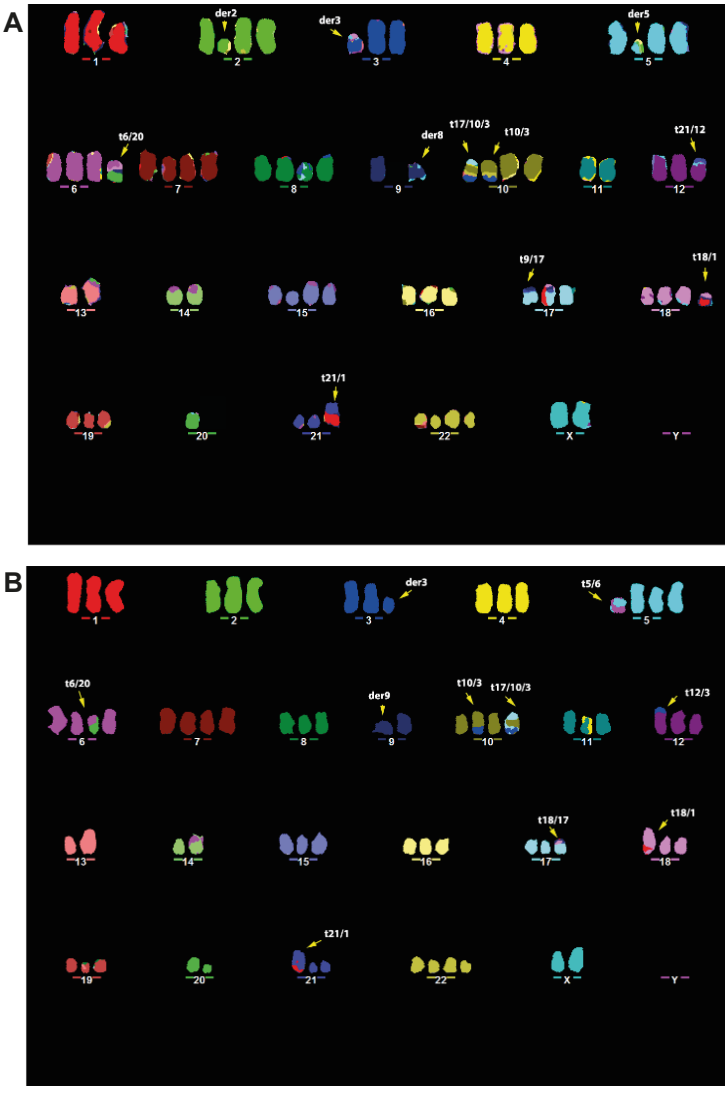


Figure 3.1: K562 and K562-Cas9 Cell Karyotype.
K562 and K562-Cas9 cells, were analysed by M-FISH. A minimum of 20 metaphases were analysed per cell line. Both cell lines were found to be essentially triploid with the majority of chromosomes present in three copies. Both cell lines had a high variability in chromosome number with modal total cellular chromosome number being 69. Several recurring chromosome rearrangements were found in most cells and acentric fragments were detected with a high frequency.

	K562 Cells	K562-Cas9 Cells
Two or less copies of chr 13	64%	90%
Two or less copies of chr 14	92%	80%
Two or less copies of chr 20	92%	90%
Two or less copies of chr X	92%	90%
Four or more copies of chr 7	72%	90%
Four or more copies of chr 22	56%	70%

Table 3.1: Percentage of K562 cells and K562-Cas9 cells showing more or less than 3 copies of the chromosomes.
Both cell lines were found to be essentially triploid with the majority of chromosomes present in three copies. However, some chromosome had more or less than three copies. The chromosomes that were recurrently found with just two copies or less, or four or more copies are listed here along with the percentage of the cell population that this occurred in.

	K562 Cells	K562-Cas9 Cells
Derivative chr 3 (missing p or q arm)	100%	55%
Derivative chr 5 (missing most of the q arm)	24%	20%
Translocation between chr 5 and 6	60%	35%
Translocation between chr 6 and 20	80%	65%
Derivative chr 9 (missing all of the p arm)	40%	60%
Translocation between chr 17, 10 and 3	68%	90%
Translocation between chr 10 and 3	100%	80%
Translocation between chr 12 and 21	76%	80%
Translocation between chr 12 and 3	-	60%
Translocation between chr 17 and 9	44%	15%
Translocation between chr 17 and 18	24%	35%
Translocation between chr 18 and 1	60%	45%
Translocation between chr 21 and 1	92%	85%

Table 3.2: Percentage of K562 and K562-Cas9 cells containing recurring chromosomal rearrangements.
Several recurring chromosome rearrangements were found in most cells in both cell lines. The M-FISH technique cannot determine with certainty that the rearrangements identified as the same have identical breakpoints, and it is possible that the rearrangements grouped together could be slightly different. Other translocations are found at lower frequency.

3.2.2 Regulatory Element Selection

As a proof of principle, known and validated enhancers were selected to determine if deleting an enhancer in a cell line would have a large enough effect to be seen through expression changes in a bulk population of edited cells. Using control enhancers allows for the assessment of the feasibility and informativeness of this approach. If deleting known enhancers shows the expected reduction in gene expression, then this method of CRISPR/Cas9 editing in K562 cells followed by qPCR could be repeated for AS relevant loci that are presently uncharacterized.

For the initial proof of principle set of CRISPR/Cas9 edits, five validated enhancers were selected from the literature. All the chosen enhancers were found to be active in K562 cells, based on publicly available data (Dunham et al. 2012), and regulate known genes. The first selected enhancer was the element -25, which regulates LIM Domain Only 2 (*LMO2*) gene expression (Bonadies et al. 2015). This enhancer has two components, with a myeloid specific enhancer at the 3', end and a T-cell repressor at the 5' end. The myeloid specific enhancer activity is dependent on the binding of the transcription factor, GATA1. The second selected enhancer was element, E2, regulating Myc Proto-Oncogene Protein (*MYC*) expression (Fulco et al. 2016). E2 was uncovered during a pooled CRISPRi screen in K562 cells. This enhancer physically interacts with *MYC* through chromatin looping and results in a 40% decrease in *MYC* expression when deleted. Another element selected was an enhancer regulating GATA Binding Protein 1 (*GATA1*) expression (Fulco et al. 2016). This enhancer is located 3.6 kb upstream of *GATA1* and when deleted or disrupted it reduces *GATA1* expression by 44%. The fourth element selected was RE1, which regulates Runt-Related Transcription Factor 1 (*RUNX1*) expression (Markova et al. 2011). *RUNX1* has two promoters, P1 and P2. Element RE1 was shown to act as a tissue specific enhancer in erythroid cells to stimulate P2. In K562 cells, the

majority of the mRNA transcribed from *RUNX1* is from P2 and stimulating RE1 increases *RUNX1* expression. The fifth enhancer element selected was HS2, which regulates γGlobin (*HGB1*) expression, as well as the expression of other globin genes within the human β-globin locus(Amrolia et al. 1998). It contains a 400-base pair core element with AP1/NF-E2 binding motifs and HS2 enhances *HGB1* promoter activity in K562 cells.

3.2.3 CRISPR/Cas9 Approach and Validation

To delete each enhancer, gRNAs were designed to target both directly upstream and downstream of the enhancer. They were designed as such to cause a double-strand break to occur on either side of the enhancer in the hope that the cells would repair the damage by NHEJ and remove the intervening enhancer entirely (Figure 3.2).

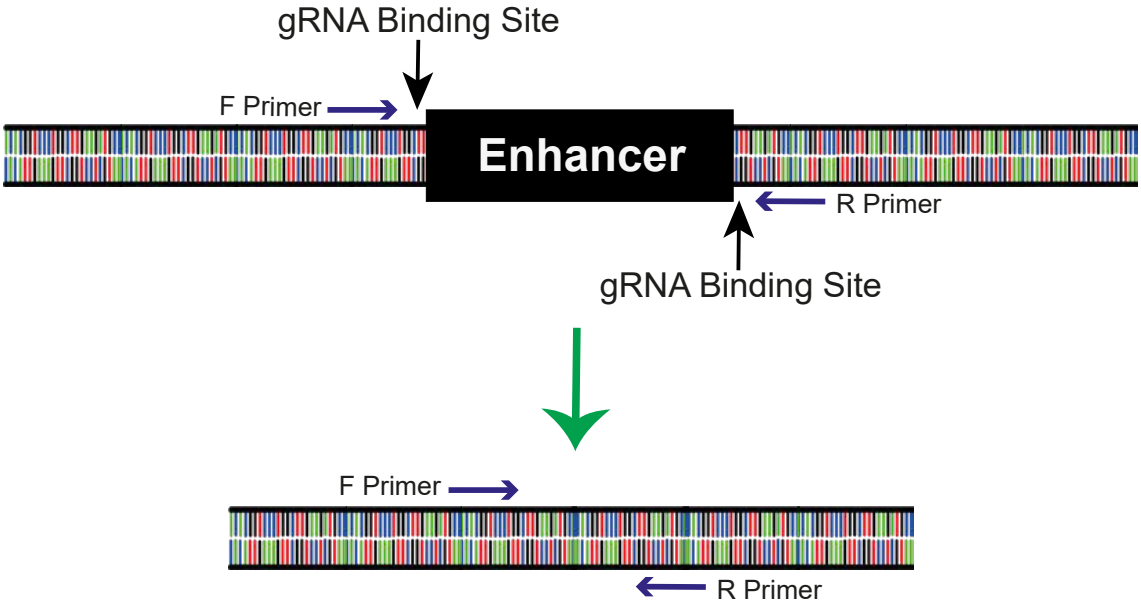


Figure 3.2: Schematic of CRISPR/Cas9 deletion and screening approach.
gRNAs were designed to recognize and bind both upstream and downstream of each desired deletion to induce double-strand breaks and remove the enhancer during DNA repair. PCR primers were designed to amplify the deleted region, labeled as Forward (F) primer and Reverse (R) primer. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel.

The RNP CRISPR/Cas9 approach was used for K562 cell editing. Both K562 and K562-Cas9 cells were transfected and harvested for DNA and RNA at three and five days post transfection. The success of the genome editing was determined by PCR analysis and gel electrophoresis. Primers were designed to amplify the deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel (Figure 3.2). For example, a correct deletion would show a size of 500bp on a gel, while a non-deletion event would show a fragment of 2000bp. The transfected cells were screened as a bulk population and were not single cell sorted or clonally expanded.

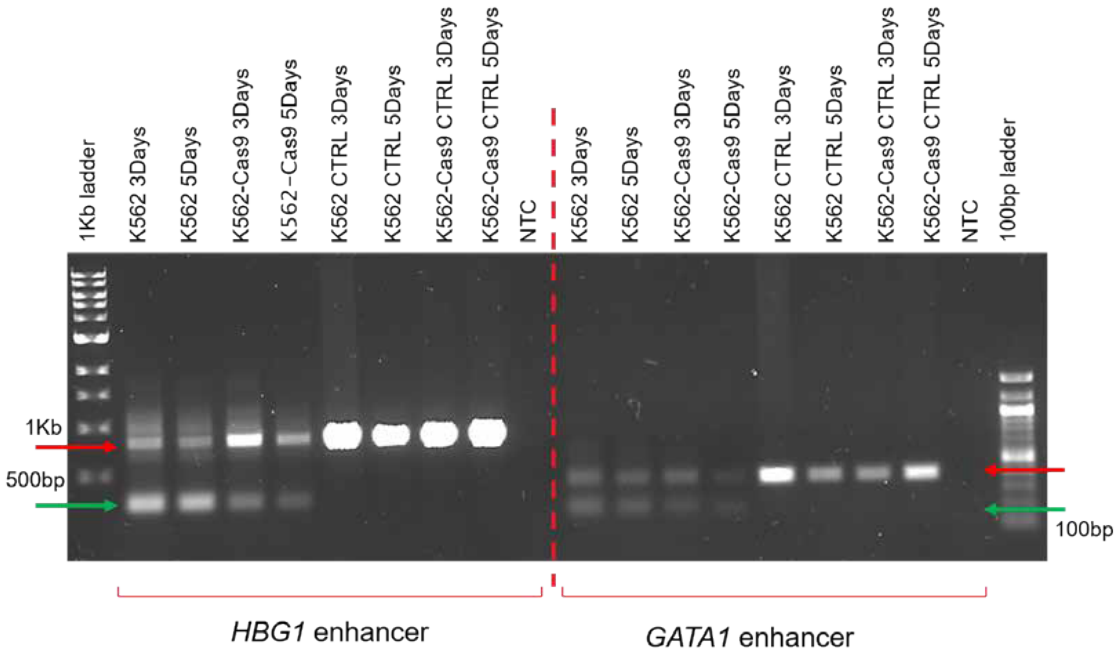


Figure 3.3: Agarose gel indicating CRISPR/Cas9 deletion success in K562 cells (*HGB1* and *GATA1*).
Enhancers for *HGB1* and *GATA1* are shown from a bulk edited cell population. Transfections were performed in both K562 and K562-Cas9 cells and were harvested at three days and five days post transfection. Unedited control cells for both cell lines were harvested and processed at the same time as the transfected cells, labeled here as CTRL. If the *HGB1* enhancer was successfully deleted it would show a band at approximately 300bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 835bp (red arrow). If the *GATA1* enhancer was successfully deleted it would show a band at approximately 220bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 400bp (red arrow). Two bands in a lane showing both the deleted and undeleted product suggest multiple genotypes exist in the cell population.

The gel in Figure 3.3 shows two of the five deletions, enhancers for *HBG1* and *GATA1*. The gel images for the other three enhancers can be seen in Figures A.1 and A.2 in Appendix A. All five enhancers were successfully deleted to some extent in K562 and K562-Cas9 cells, as shown by gel electrophoresis. However, this method cannot quantify the amount of cells in the bulk cell population, it only shows that the desired deletion event has occurred.

3.2.4 Bulk Cell Population Expression Analysis

To determine if a bulk cell population of edited cells would exhibit an effect on gene expression, RNA was extracted for qPCR analysis. The relative expression of the gene associated with each deleted enhancer was evaluated in K562 and K562-Cas9 cells at three and five days post transfection. No significant changes in expression levels were observed. A small down-regulation of gene expression was observed for *GATA1* and *RUNX1* in the K562-Cas9 cells for deletions of their respective enhancers at five days post transfection. Those transfections were repeated a further two times and changes in gene expression were measured again by qPCR. No significant changes in *GATA1* or *RUNX1* expression were observed following the three CRISPR/Cas9 deletion replicates (Figure 3.4).

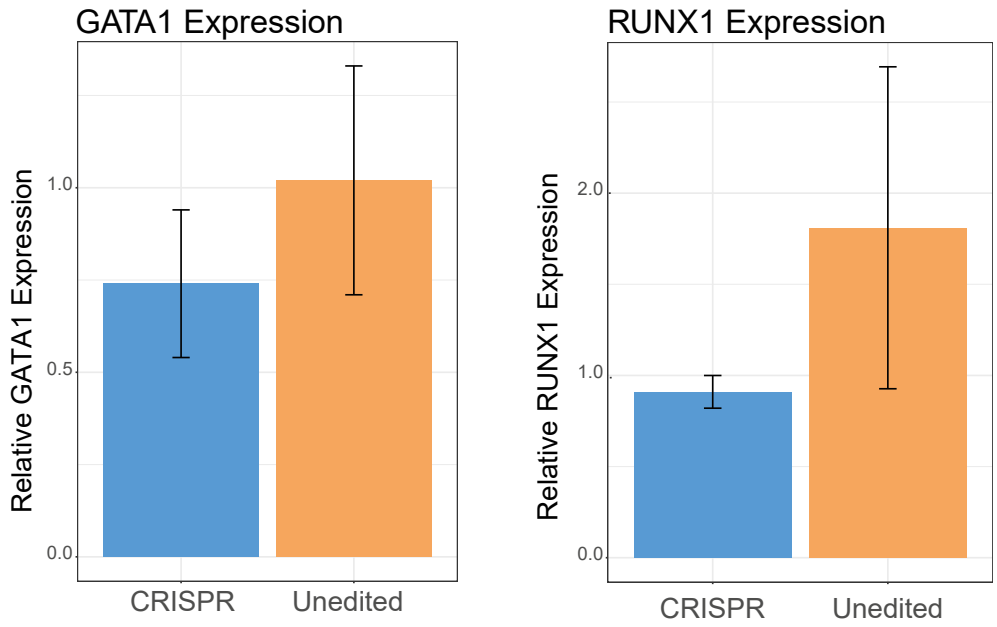


Figure 3.4: Expression of *GATA1* and *RUNX1* in K562-Cas9 cells.
The relative expression of *GATA1* and *RUNX1* following qPCR analysis on bulk edited cell lines harbouring deletions for the *GATA1* and *RUNX1* enhancers. Expression levels are relative to the gene *TBP*. The t-test p-values for the *GATA1* and *RUNX1* expression changes were 0.24 and 0.37, respectively. All replicates were harvested at five days post transfection and analysis was carried out on the bulk edited cell population. Error bars represent SEM from three replicates.

3.2.5 Clonal Isolation and Expansion

The efficiency of the enhancer deletion in the bulk edited pool was not ascertained, but it was potentially too low to detect a change in gene expression. Therefore, single clone isolation was performed to assess the success of the deletion. Cells transfected with the *RUNX1* enhancer gRNAs were individually isolated and separated out for clonal expansion. A total of 66 colonies were successfully grown and screened for the presence of the *RUNX1* enhancer deletion using PCR and gel electrophoresis (Figure A.3 in Appendix A). A total of 46 colonies (70%) contained cells harbouring the deletion. Therefore, approximately 70% of the bulk edited cell population harboured the desired deletion. Even with a high deletion efficiency, no gene expression changes were discernible, suggesting the enhancer may need to be deleted completely to see any visible change.

3.2.6 Screening Clonal Populations for Homozygotes

To maximize signal and determine if cell populations homozygous for an enhancer deletion would have a significant effect on gene expression, clones showing signs of a deletion event were screened further. To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region, and one binding just outside (Figure 3.5) (Primers *RUNX1* Enhancer 3' and 5' Het in Chapter 2). Amplification should only occur if the deletion region is present in a heterozygous state, as binding would not occur if the enhancer region has been removed.

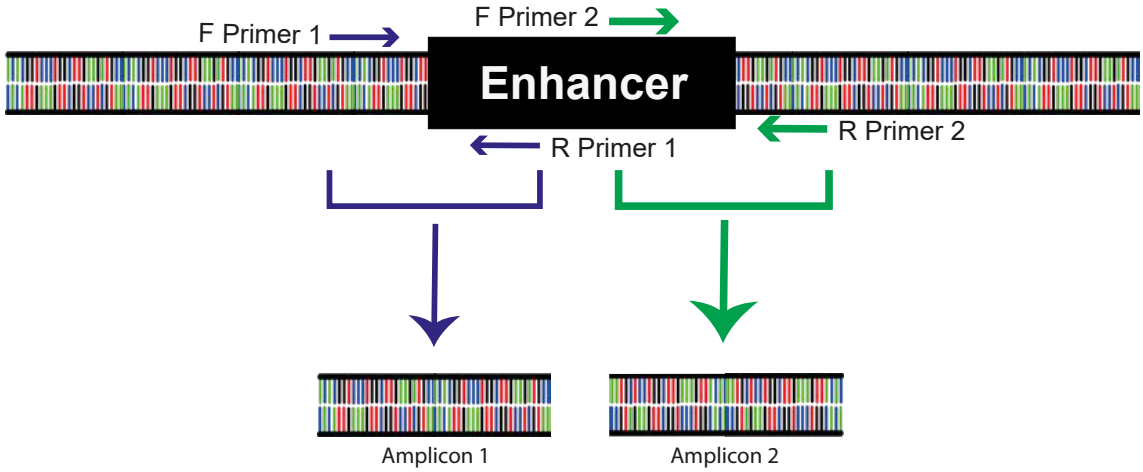


Figure 3.5: Schematic of CRISPR/Cas9 zygosity screening approach.
To determine if a cell population was homozygous or heterozygous for the enhancer deletion, two primer sets were designed for PCR and gel electrophoresis. One primer from each set binds within the deletion region and one binds just outside. Amplification should only occur if the deletion region is present in a heterozygous state as binding would not occur if the enhancer region has been removed.

All clones that showed signs of a deletion event were screened in this way and all of them appeared to have at least one unedited copy of the *RUNX1* enhancer, except for one sample that did not amplify (Clone 30) (Figure A.4 in Appendix A). For further investigation, a small subset of four clones, including Clone 30, were taken forward. Primers binding fully within the deletion region (Primers *RUNX1* Enhancer Internal in Chapter 2) and long range primers binding further away from the deletion region (Primers *RUNX1* Enhancer Long Range in Chapter

2) were designed and used because of the risk of INDELs disrupting the primer binding site (Hilton and Gersbach 2015) (Figure A.5 in Appendix A). Most clones showed that the desired deletion event had occurred, but none of the clones appeared homozygous for the deletion.

3.2.7 Copy Number Analysis of Edited K562 Cell Clones

FISH Analysis

Because K562 cells have an unusual karyotype, the copy number of the *RUNX1* enhancer had to be determined in the parental cell line and in the edited clones. Fluorescence in situ hybridization (FISH) was used to determine the copy number of the regions containing the *RUNX1* enhancer and the *RPP30* gene on the parental K562 cell line and the edited Clones. This was performed by Daniela Moralli in Cath Green's Group at the WHG. Only large regions can be probed in FISH, so the areas containing the *RUNX1* enhancer and *RPP30* gene were analyzed. *RPP30* was analyzed here as it was selected as a reference gene for downstream ddPCR analysis. A minimum of 10 metaphases, and 50 nuclei were analyzed per cell line. Table 3.3 and Table 3.4 show the percentage of cells with varying copy numbers for the *RUNX1* enhancer region and *RPP30*, respectively. The majority of all the K562 cells examined have four copies for the *RUNX1* enhancer and two copies for *RPP30*. However, the parental cell line, Clone 12, and Clone 30 have a large percentage of cells with one copy for RPP30. A large percentage of cells also have three copies of the *RUNX1* enhancer region. All the cell lines, including the parental K562 cells, exhibit a high amount of genetic heterogeneity. Example metaphases of each cell line are shown in Figure 3.6.

	No Copies	1 Copy	2 Copies	3 Copies	4 Copies	More than 4 Copies
K562 Parental	0%	5.0%	6.7%	21.7%	63.3%	3.3%
K562 Clone 12	0%	8.3%	5.0%	23.3%	60.0%	3.3%
K562 Clone 17	0%	3.3%	3.3%	25.0%	56.7%	11.7%
K562 Clone 30	0%	0%	8.8%	19.3%	64.9%	7.0%
K562 Clone 57	0%	0%	5.0%	20.0%	66.7%	8.3%

Table 3.3: *RUNX1* enhancer region copy number per cell.
Percentages of cells containing the listed copy numbers of the *RUNX1* enhancer region as determined by FISH. The region examined was chr10:92585413-92729845 from the hg19 genome build. The majority of the analyzed cells from the K562 parental cell line and Clones have four copies for the *RUNX1* enhancer region. A large porportion of the five cell lines have three copies for the *RUNX1* enhancer region. All the cell lines have at least one copy of the *RUNX1* enhancer region.

	No Copies	1 Copy	2 Copies	3 Copies	4 Copies	More than 4 Copies
K562 Parental	6.7%	23.3%	68.3%	1.7%	0%	0%
K562 Clone 12	11.7%	18.3%	68.3%	0%	1.7%	0%
K562 Clone 17	1.7%	11.7%	60.0%	18.3%	6.7%	1.7%
K562 Clone 30	3.5%	14.0%	71.9%	3.5%	7.0%	0%
K562 Clone 57	1.7%	10.0%	70.0%	18.3%	0%	0%

Table 3.4: *RPP30* gene copy number per cell.
Percentages of cells containing the listed copy numbers of the *RPP30* gene as determined by FISH. The region examined was chr21:36304205-36473556 from the hg19 genome build. The majority of the analyzed cells from the K562 cell lines have two copies of *RPP30*. The parental cell line, Clone 12, and Clone 30 have a large percentage of cells with only one copy for *RPP30*. Clone 17 has a considerate number of cells with one or three copies for *RPP30*, and Clone 57 has a considerate number of cells with three *RPP30 copies*.

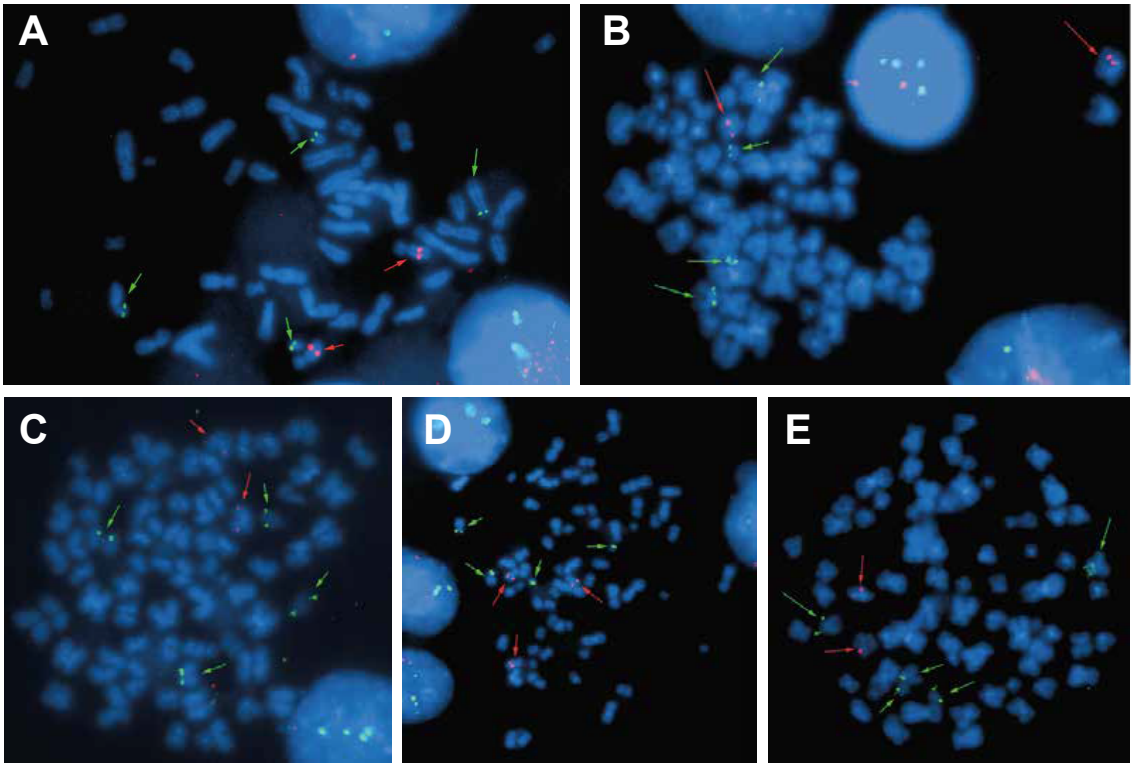


Figure 3.6: FISH of K562 cells showing *RUNX1* enhancer region and *RPP30* copy numbers.
Example metaphase FISH of each cell line. The chromosomes are counter-stained in DAPI (blue). Probes for the *RUNX1* enhancer region (green) and *RPP30* (red) were used. **A**. K562 parental cell line. Four copies of the *RUNX1* enhancer region (green) and two copies of *RPP30* (red) are identified. **B**. K562 Clone 30. Four copies of the *RUNX1* enhancer region (green) and two copies of *RPP30* (red) are identified. **C**. K562 Clone 12. Four copies of the *RUNX1* enhancer region (green) and two copies of *RPP30* (red) are identified. **D**. K562 Clone 17. Four copies of the *RUNX1* enhancer region (green) and three copies of *RPP30* (red) are identified. **E**. K562 Clone 57. Five copies of the *RUNX1* enhancer region (green) and two copies of *RPP30* (red) are identified.

ddPCR Analysis

To examine the exact number of *RUNX1* enhancer copies successfully deleted from the K562 clones, droplet digital polymerase chain reaction (ddPCR) was used. Primers and probes were designed to target the exact *RUNX1* enhancer, rather than the broader surrounding area as with FISH. *RPP30* was used as the reference gene to calculate *RUNX1* enhancer copy number. The determined copy number of the *RUNX1* enhancer for each clone can be seen in Figure 3.7. The K562 parental cell line had approximately 4 copies of the *RUNX1* enhancer, which is consistent with the FISH analysis. The edited cell clones and the bulk

edited cell population all show a reduction in the *RUNX1* enhancer copy number following CRISPR/Cas9. None of the cell lines are homozygous for the deletion, which is consistent with the gel electrophoresis analysis.

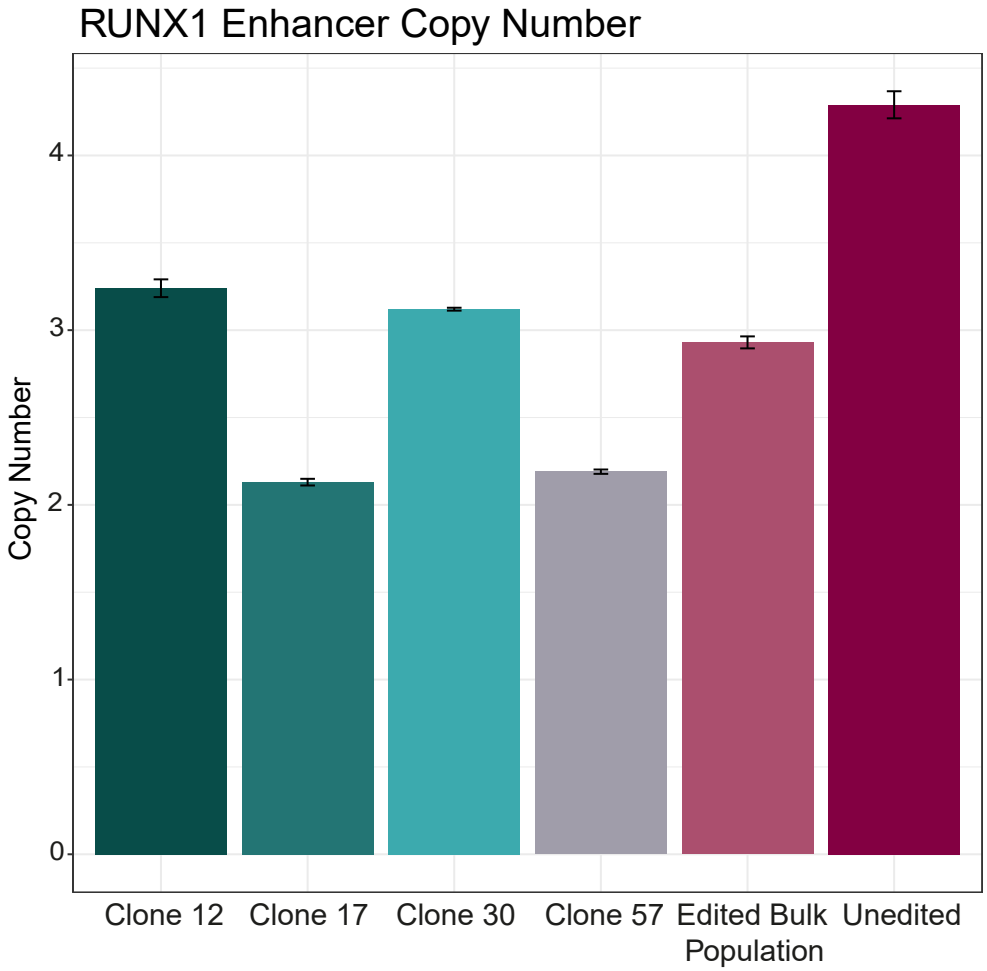


Figure 3.7: *RUNX1* enhancer copy number in K562 cell CRISPR/Cas9 edited clones. The copy number of the *RUNX1* enhancer following ddPCR analysis in the CRISPR/Cas9 edited cell clones and the K562 parental cell line. The bulk population of edited cells with no clonal isolation was also examined. *RPP30* was used as the reference gene. Error bars represent SEM from three replicates.

3.2.8 Expression Analysis of Edited K562 Cell Clones

To assess if removing copies of the *RUNX1* enhancer from the cells had any effect on *RUNX1* gene expression, qPCR was used. The relative expression of *RUNX1* was evaluated in each clone and in the K562 parental cell line (Figure 3.8). No significant changes in expression levels were observed, with the exception of

Clone 12. There was an up-regulation of RUNX1 expression in Clone 12 (p-value of 0.0315), when removing the enhancer is expected to result in a down-regulation of expression. There does not appear to be a correlation between the *RUNX1* enhancer copy number and the gene expression of RUNX1 in the K562 edited cell clones. The variability of copy number in these clones makes it difficult to discern and compare gene expression changes.

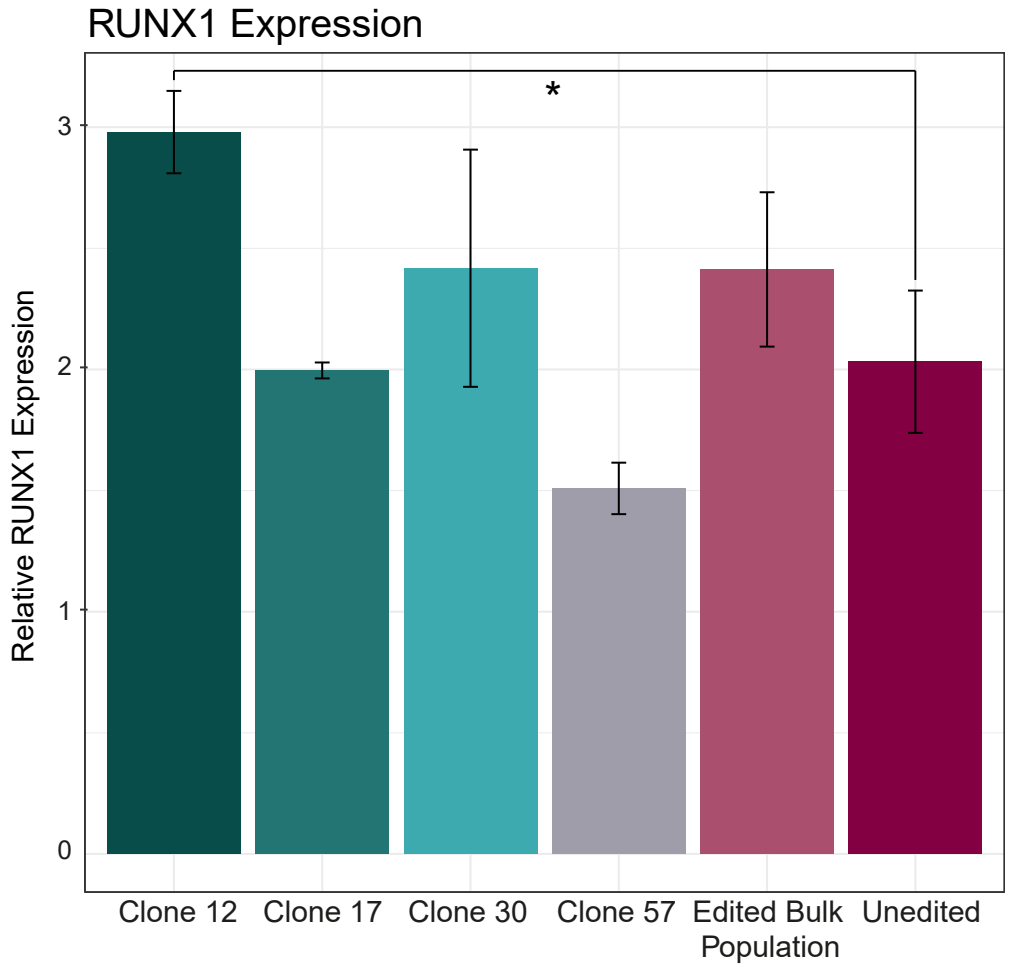


Figure 3.8: *RUNX1* expression in K562 edited clones.
The relative expression of *RUNX1* from qPCR analysis of four cell clones following CRISPR/Cas9 editing to delete the *RUNX1* enhancer. The bulk population of edited cells with no clonal isolation and the K562 unedited parental cell line were also examined. Expression levels are relative to the gene *TBP*. Only Clone 12 exhibited a significant change in expression. The p-values for Clone 12, Clone 17, Clone 30, Clone 57, and the edited bulk population are as follows: 0.0315, 0.907, 0.524, 0.145, and 0.414. Error bars represent SEM from three replicates.

3.3 Results: CRISPR/Cas9 in iPSCs

3.3.1 Priority Index for Variant Prioritization

Hai Fang in Julian Knight's Group developed the Priority Index (Pi) computational pipeline as a way to help researchers prioritize genes and variants in autoimmune diseases (Fang et al. 2019). An early version of Pi was used in this thesis (2015-16) to prioritize genes associated with AS through GWAS. Pi works by using genomic predictors on GWAS variants to score genes based on their likelihood of being responsible for the disease association (Figure 3.9). Each GWAS gene is scored using: (1) genomic proximity to a disease-associated SNP (nGene), (2) physical chromatin interactions in immune cells (cGene), and (3) modulation of gene expression in immune cells as determined by eQTL analysis (eGene). Genes that have a score in at least one of these categories are known as seed genes. The seed genes are then further scored using annotation predictors: (1) immune function (fGene), (2) immune phenotype (pGene), and (3) rare genetic diseases related to immunity (dGene). Genes could also be given a score if they are known to interact with seed genes or highly ranked genes, for example being in the same signaling pathway. When used for AS, a ranked list of around 15,000 genes is generated. Pi has two modes, discovery mode and supervised mode. Discovery mode ranks genes that may play a role in disease aetiology, whereas supervised mode uses machine learning on approved drug lists to rank genes based on their potential as a drug target for a disease.

Preliminary Version of Pi

There are many loci currently associated with AS, and of those, one needed to be prioritized and selected as the starting point for this thesis work. A very early version of the Pi pipeline was used to aid in variant prioritization. This initial Pi pipeline scored each gene using similar parameters and metrics as the current

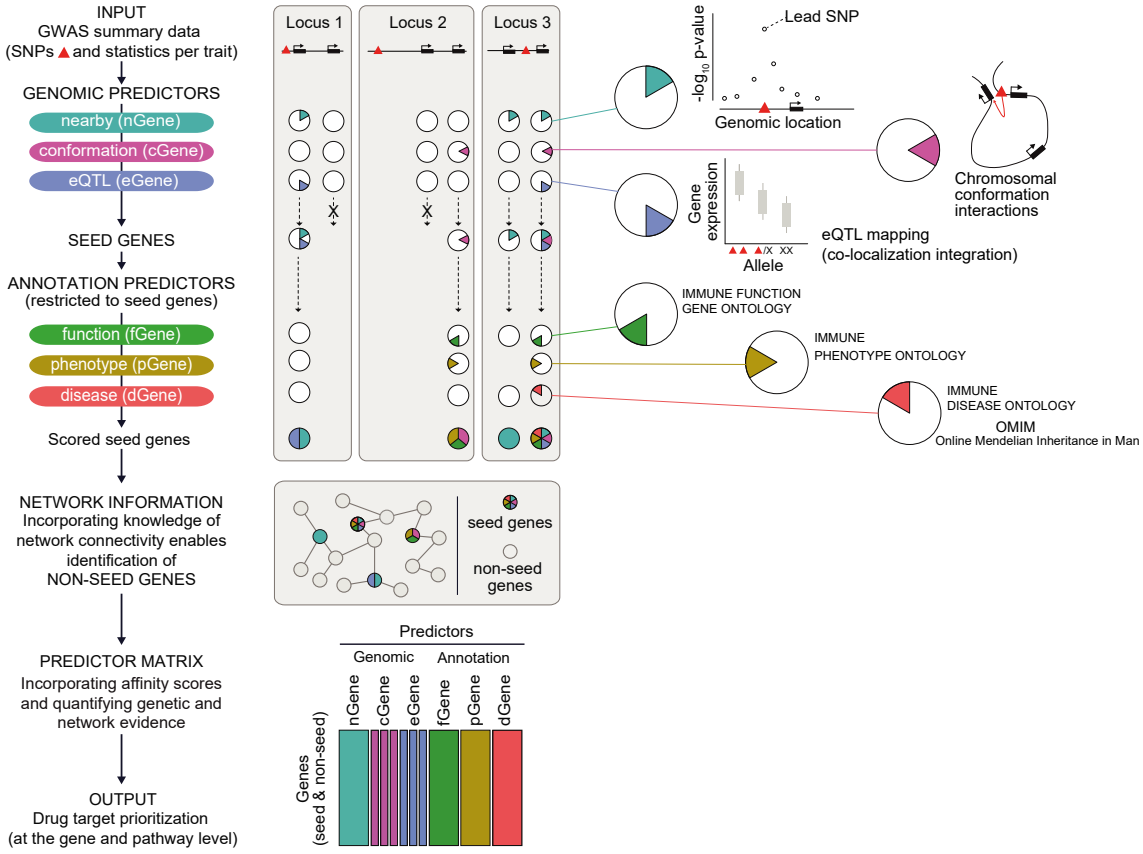


Figure 3.9: Overview of the Pi pipeline.
Schematic showing how genes are ranked. Genomic predictors, based on proximity, conformation, and expression (represented as a different pie segment), are used to define seed genes. Annotation predictors (immune function/phenotype/disease) are then only applied to seed genes. Non-seed genes are identified through network connectivity with seed genes. These factors are used to rank approximatley 15,000 genes per disease trait. Figure adapted from Fang et al. 2019.

version of Pi, but with fewer datasets and annotations. The output showing the top 20 genes from the initial Pi analysis can be seen in Table 3.5. This was the version used to initially pick a locus for functional characterization work.

Tumor Necrosis Factor Receptor Superfamily Member 1A (*TNFRSF1A*) was the highest ranked gene and is involved in the immune system and inflammation, with genetic variants within it also associated with other diseases. The most recent version of Pi has additional datasets including promoter capture Hi-C and multiple eQTL datasets for different cell types, and ranks genes based on GWAS-eQTL colocalization and associations with other highly ranked genes, which the initial version did not do. Using the current version of Pi, *TNFRSF1A* is ranked

Rank	Gene	GWAS	eGene	Variant	dGene	pGene	fGene	Pathway
1	*TNFRSF1A*	1	1	0	1	1	1	0
2	*ERAP2*	1	1	1	1	0	0	0
3	*IL6R*	1	1	1	0	0	0	0
4	*IL12B*	1	0	0	0	0	1	1
5	*NOS2*	1	0	1	1	0	0	0
6	*PTGER4*	1	1	0	0	0	1	0
7	*ERAP1*	1	0	1	1	0	0	0
8	*CARD9*	1	1	0	1	0	0	0
9	*IL23R*	1	0	1	1	0	0	0
10	*C5*	0	0	0	0	1	1	1
11	*IL2*	0	0	0	1	0	1	1
12	*IL6*	0	0	0	1	0	1	1
13	*IL10*	0	0	0	1	0	1	1
14	*ITGB2*	0	0	0	0	1	1	1
15	*STAT3*	0	0	0	1	1	0	1
16	*TNF*	0	0	0	1	0	1	1
17	*NOD2*	0	0	0	1	0	1	0
18	*RUNX3*	1	0	1	0	0	0	0
19	*ETS2*	1	1	0	0	0	0	0
20	*FCGR2A*	1	1	0	0	0	0	0

Table 3.5: Output from the preliminary Pi Pipeline ranking genes as drug targets in AS from genetic evidence.
The top 20 ranked genes from an early version of Pi are shown along with how they scored in each annotation section. Each gene was given an overall score relating disease relevance based on scores of 1 or 0 in individual categories. GWAS: based on proximity to a disease-associated variant in an AS GWAS. eGene: gene where the GWAS variant is associated with differential gene expression. Variant: Gene with an associated SNP found in protein coding regions predicted to be missense, stop gained or lost, or a splice donor or acceptor site. dGene: Gene with experimental data linking with AS pathogenesis. pGene: associated with a human phenotype of "abnormality of immune system physiology. fGene: Gene annotated on Gene Ontology as being related to an inflammatory response. Pathway: Gene in an inflammatory pathway.

at 13 for AS, out of around 15,000 genes. It scored highly due to the presence of genomic and annotation predictors, including nGene, eGene, cGene, fGene, pGene, and dGene. It is also suggested to interact with other genes thought to be important in disease pathogenesis, including *ERAP1* and *NF-κB* (Figure 3.10). Along with *TNFRSF1A*, other genes that were highly ranked using the preliminary version of Pi and are still highly ranked using the current version include *ERAP2*, *IL10*, ETS2, IL23R, and *PTGER4*.

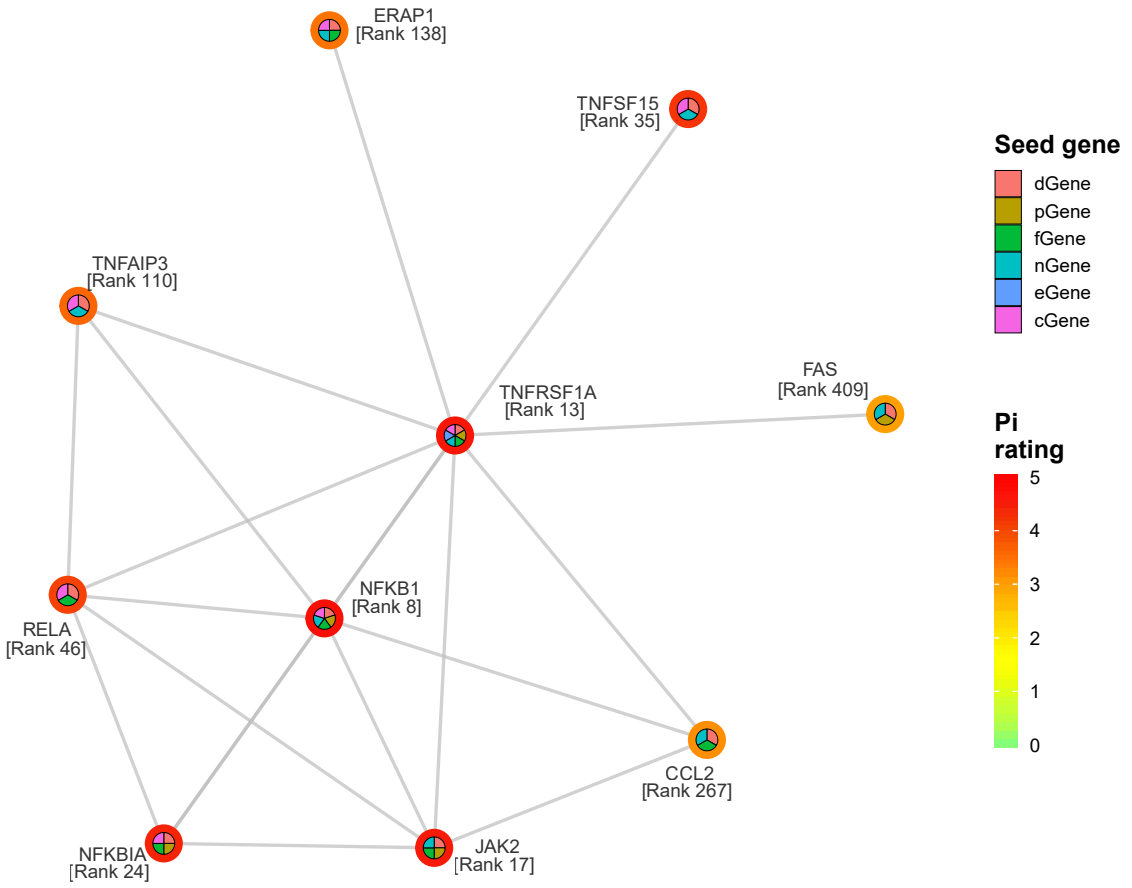


Figure 3.10: Prioritization interaction plot for *TNFRSF1A*.

3.3.2 Further Prioritization of *TNFRSF1A*

TNFRSF1A was the gene selected to target for CRISPR/Cas9 experimentation. In addition to being highly ranked by Pi, it plays a key role in the modulation of the inflammatory response and is associated with multiple immune disorders

(reviewed further in Chapter 4). The lead GWAS associated SNP for this region is rs1860545, which is located in the middle of the first intron of *TNFRSF1A*. To further examine this region, fine mapping was performed by Anna Sanniti in Julian Knight's Group using a Bayesian approach on a subset of the data from the International Genetics of Ankylosing Spondylitis Consortium (IGAS) (Cortes et al. 2013). The 95% credible set generated included 6 SNPs in very strong LD: rs1860545, rs1800693, rs887477, rs4149577, rs34822098, and rs11064145. One of these SNPs is already implicated in disease. rs1800693 is associated with a splicing defect found in MS patients (Gregory et al. 2012), but is not thought to be causal in AS based on direction of association and TNF response in MS and AS (Cortes et al. 2013). Not much is known about the other five co-inherited SNPs, leaving open the question of which SNP(s) are causally related to the disease. Therefore, all six fine mapped variants were interrogated using the UCSC genome browser with publicly available datasets (Figure 3.11) (Kent et al. 2002).

Two SNPs rs887477, and rs11064145 appeared to be promising candidates for further study in immune cell types. They are located in regions of open chromatin (DNase I hypersensitive sites) in $CD4^+$, $CD14^+$, $CD20^+$, and K562 cells based on ENCODE data (Dunham et al. 2012) . They are also situated within predicted enhancers/ regulatory elements (histone marks H3K27ac, H3K4me1) in GM12878 and K562 cells also based on ENCODE data (Dunham et al. 2012). They are also located within predicted enhancers as annotated by ORegAnno (Lesurf et al. 2015). Moreover, the regions show evidence of transcription factor binding based on ChIP data from ENCODE (Dunham et al. 2012) and region conservation across 44 vertebrate species, as annotated by the phastCons program (Pollard et al. 2010).

rs887477 is an A>C variant with a minor allele frequency of 0.46(A) in Europeans, and rs11064145 is a T>G variant with a minor allele frequency of 0.44(G) in Europeans (Consortium 2015). Multiple transcription factors are

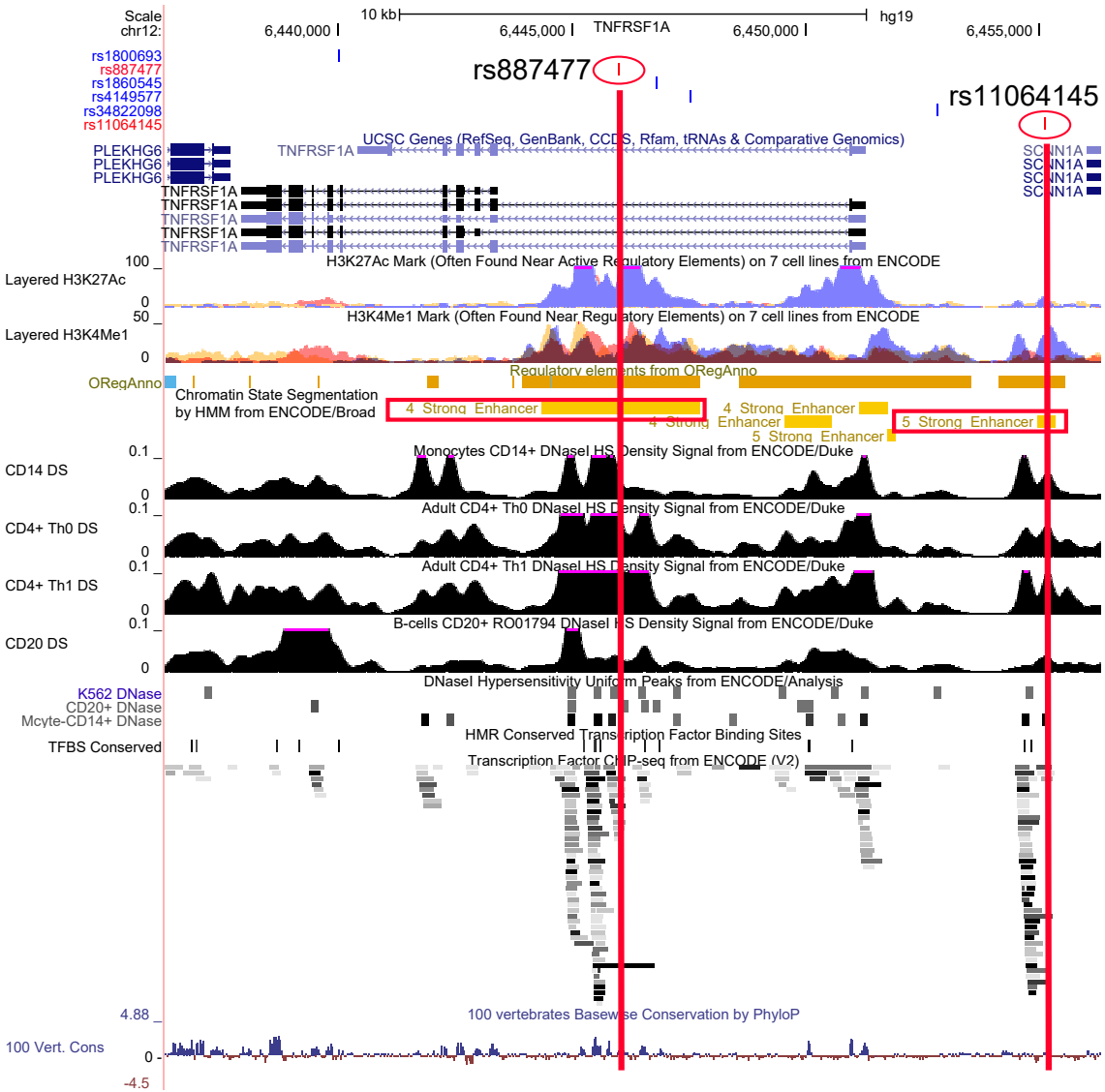


Figure 3.11: *TNFRSF1A* fine mapped SNPs.
Region of TNFRSF1A showing 6 fine mapped GWAS SNPs for CD14, CD4, CD20, and K562 (monocyte like) cells. The two SNPs selected for this project, rs887477 and rs11064145, are highlighted. The tracks show the SNPs to be in predicted regulatory regions, in DNase I hypersensitivity sites in multiple relevant cell types, in predicted enhancers, and regions of transcription factor binding and sequence conservation. Image of TNFRSF1A and tracks taken from the UCSC Genome Browser (Kent et al. 2002).

predicted to bind the regions spanning these SNPs when queried using the PROMO software (Messeguer et al. 2002). The A allele of rs887477 is situated within the predicted binding motif for the transcription factors, TFII-I and PPARα, while the C allele alters the binding motif to bind those two factors as well as Myc-associated zinc finger protein (MAZ) (Figure 3.12A). MAZ both up- and down-regulates transcription of a variety of genes including, MYC (Bossone et al. 1992), CD4 (Duncan et al. 1995), VEGF (Ray et al. 2011), and many others (Yu et al. 2017). It is ubiquitously expressed and has been associated with cancer (Yu et al. 2017). The T allele for rs11064145 is situated within the binding motif for Pax-5 and p53, while the G allele changes the binding motif to include GCF, as well as Pax-5 and p53 (Figure 3.12B). In vitro studies have shown that GCF, which stands for GC-Rich Sequence DNA-Binding Factor 2, is a transcriptional repressor and has been found to repress the epidermal growth factor (*EGF*) receptor, β-actin, and the protease CANP (Beguinot et al. 1995).

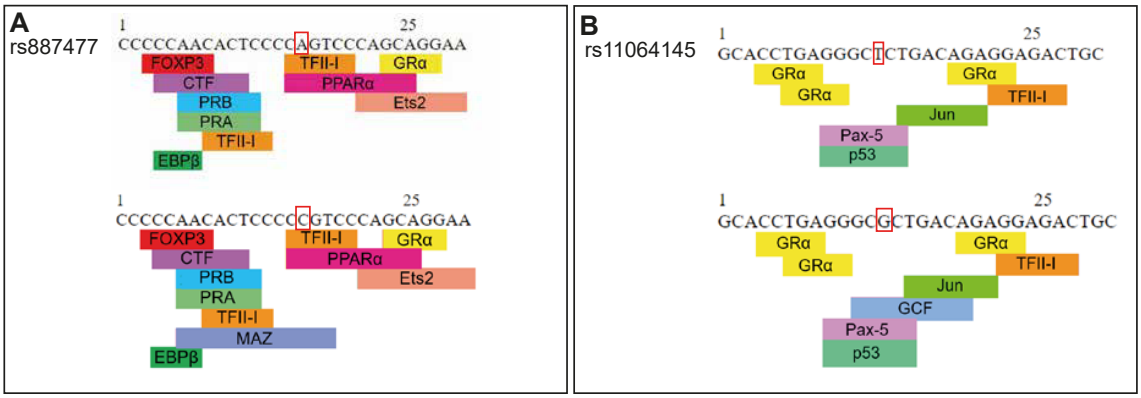


Figure 3.12: Overview of PROMO predicted transcription factor binding.
A. The A allele of rs887477 is situated within the predicted binding motif for the transcription factors, TFII-I and PPARα, while the C allele alters the binding motif to bind those two factors as well as MAZ. **B**. The T allele for rs11064145 is situated within the binding motif for Pax-5 and p53, while the G allele changes the binding motif to include GCF, as well as Pax-5 and p53.

3.3.3 Experimental Design

An initial AS associated GWAS locus had to be selected to optimize CRISPR/Cas9 in iPSCs and to determine the overall suitability of this approach. The hope was to start with one locus to create a robust system for characterizing variants using CRISPR/Cas9 in iPSCs, and then to expand to more loci. Both rs887477 and rs11064145 fall within accessible chromatin regions, with many transcription factors predicted to bind, and are situated within predicted enhancers, based on histone marks. Therefore, these SNPs, and the putative enhancers they are situated within, were the initial selected regions for this project.

The aim of this work was to perform genome editing to generate a gene knockout of *TNFRSF1A* by removing four exons, and disrupt two putative enhancers spanning the AS GWAS SNPs, rs887477 and rs11064145 (Figure 3.13). The exons selected for removal were exons 2, 3, 4, and 5. They are present in every transcript of the gene, and deleting them should be sufficient to cause total knockout of the gene. This gene silencing, by deletion of critical exons, was needed as an experimental control to compare the effects of enhancer removal with the effects of a gene knockout. Silencing a gene should have the same or similar affects as disrupting its enhancer, and therefore, this knockout could aid in understanding the role of the putative enhancers. Removing the four exons shall be referred to as Exon Deletion, removing the putative enhancer containing rs887477 shall be referred to as 'Intronic Enhancer Deletion', and removing the putative enhancer containing rs11064145 shall be referred to as 'Intergenic Enhancer Deletion'. The Exon Deletion region is 1177bp, the Intronic Deletion region is 3346 bp, and the Intergenic Enhancer Deletion region is 402bp. The first approach used to achieve CRISPR/Cas9 edits was a plasmid based method for CRISPR/Cas9 complex delivery into iPSCs. Several gRNA constructs were tested in HEK-293 cells for editing efficiency, and the most effective were selected to take forward into iPSCs (Figure A.6 in Appendix A).

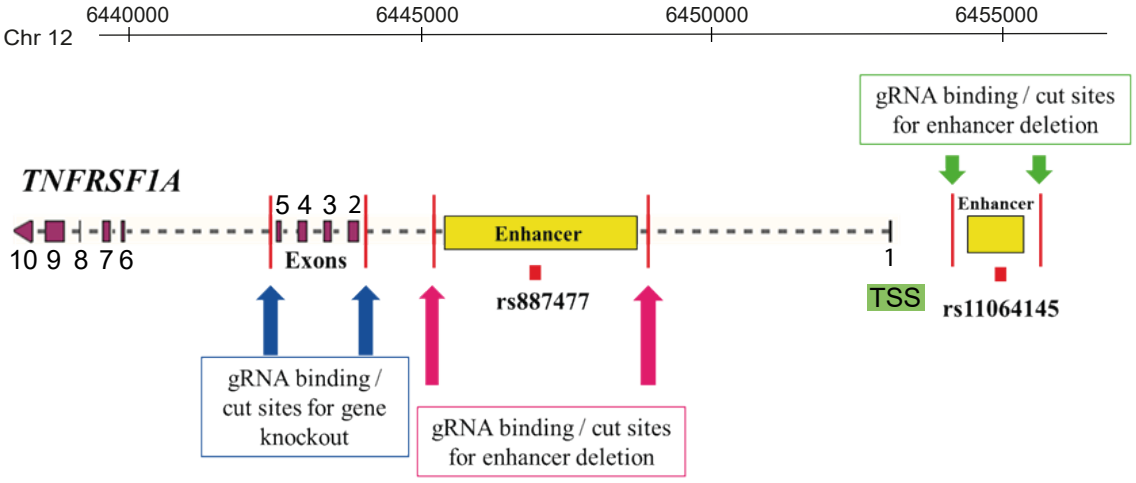


Figure 3.13: Schematic of the TNFRSF1A gene showing introns, exons, and broader genomic context.
The SNPs, rs887477 and rs11064145, are depicted along with the putative enhancers they lie in. The numbers denote exons and the transcription start site (TSS) is indicated. The gRNA binding sites where the double-stranded break will occur are highlighted. The genome editing will remove four exons for gene knockout and two enhancers.

3.3.4 CRISPR/Cas9 Validation

Gel Electrophoresis Analysis

The iPSCs were transfected with plasmids containing expression cassettes for both Cas9 and for the desired gRNAs, as well as puromycin for selection. In order to generate cell lines harboruing a specific edit, they must be separated into clonally derived cell colonies and screened. The bulk edited population will contain cells of varying genotypes, so by separating them into single cells they can proliferate and create a population of cells with the same CRISPR/Cas9 edit. To accomplish this, iPSCs were plated at low density on MEFs, and colonies were picked and expanded for screening. Colonies that had merged together or were too close in proximity were not picked due to their increased chance of not being derived of a single cell. A total of 192 colonies were picked for each each desired deletion and were initially screened using PCR and gel electrophoresis to look for a deletion event (Figures A.7, A.10, and A.13 in Appendix A). Primers were designed to amplify the deleted region using PCR. If the correct deletion had

occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel (Figure 3.2). For example, a correct deletion would show a size of 500bp on a gel, while a non-deletion event would show a fragment of 2000bp. The expected sizes following each deletion are outlined in Table 3.6.

Deletion Region	Size Without Deletion	Size With Deletion
Exon Deletion	2005bp	592bp
Intronic Enhancer Deletion	4005bp	414bp
Intergenic Enhancer Deletion	1065bp	552bp

Table 3.6: Expected amplicon sizes following CRISPR/Cas9 deletion events.
All three deletions would show a large size difference when amplified by PCR and run on an agarose gel.

To properly characterize these putative enhancer regions, cell colonies homozygous for the deletion of interest were required. Heterozygous deletion cells may not show a large enough effect to understand the role of these regulatory regions, so it was imperative to screen cells for zygosity and find colonies with both alleles deleted.

To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region and one binding just outside (Figure 3.5). Amplification should only occur if the deletion region is present in a heterozygous state as binding would not occur if the enhancer region has been removed. All the cell colonies showing signs of the correct deletion were screened again in this way to test for zygosity (Figures A.8, A.11, and A.14 in Appendix A). Primers binding entirely within the deletion region were also used to double check the genotype, as they would bind and produce an amplicon in cell population homozygous for the deletion (Figures A.9, A.12, and A.15 in Appendix A). The colonies were also sequenced to verify that the correct deletion had taken place. Multiple cell colonies for each desired deletion were heterozygous, but none were homozygous for the deletion of interest (Table 3.7).

Deletion Region	Homozygous Cells	Heterozygous Cells
Exon Deletion	0	6
Intronic Enhancer Deletion	0	8
Intergenic Enhancer Deletion	0	20

Table 3.7: iPSC zygosity after CRISPR/Cas9 transfections using a plasmid approach method.
Number of cells either heterozygous or homozygous for each deletion of interest. A total of 192 colonies were screened per deletion.

ddPCR Analysis

To verify the heterozygous status of the isolated clones, ddPCR was implemented to examine the copy number of the deletion present in each of the cells. Primers and probes were designed to target each of the enhancers and the exon region, and *RPP30* was used as the reference gene to calculate the copy number. The iPSCs used here were derived from healthy human cells and have a normal karyotype. All cell colonies that appeared heterozygous when visualized on an agarose gel were examined using ddPCR. The copy number was expected to be 0, 1 or 2, however, the results ranged from 0.69 to 2.0 across the three deletions. None of the colonies were homozygous for any of the deletions, but some did have a very low copy number. The copy number in most of the colonies was not the expected 0, 1 or 2, but a range of unusual numbers, which would not be possible in a pure clonal population and is therefore indicative of mosaicism.

Sample	Deletion Copy Number
Exon Deletion B1	Undetermined
Exon Deletion B2	1.87 ± 0.029
Exon Deletion D11	2.08 ± 0.052
Exon Deletion E7	1.56 ± 0.0058
Exon Deletion E1	Undetermined
Exon Deletion H3	Undetermined
Unedited	2.01 ± 0.038

Table 3.8: Copy number of Exon Deletion CRISPR/Cas9 edited iPSC clones.
Copy number was determined by ddPCR for each cell colony. Colonies are named based on the well of the 96 well plate they were selected from. The copy number was expected to be 0, 1 or 2, however the results show many of the colonies had a copy number of 1.5. The error is represented by SEM from three replicates. Some of the samples did not amplify, possibly due to low quality DNA.

Sample	Deletion Copy Number
Intronic Enhancer Deletion A6	1.62 ± 0.0033
Intronic Enhancer Deletion C10	2.10 ± 0.041
Intronic Enhancer Deletion C3	1.58 ± 0.041
Intronic Enhancer Deletion C4	2.09 ± 0.0033
Intronic Enhancer Deletion C9	2.10 ± 0.047
Intronic Enhancer Deletion D10	2.04 ± 0.023
Intronic Enhancer Deletion D3	1.73 ± 0.24
Intronic Enhancer Deletion D4	2.05 ± 0.023
Unedited	2.10 ± 0.077

Table 3.9: Copy number of Intronic Enhancer Deletion CRISPR/Cas9 edited iPSC clones.
Copy number was determined by ddPCR for each cell colony. Colonies are named based on the well of the 96 well plate they were selected from. The copy number was expected to be 0, 1 or 2, however the results show many of the colonies had a copy number between 1.5 and 2. The error is represented by SEM from three replicates.

Sample	Deletion Copy Number
Intergenic Enhancer Deletion A3	1.87 ± 0.029
Intergenic Enhancer Deletion B10	1.72 ± 0.026
Intergenic Enhancer Deletion C10	1.41 ± 0.029
Intergenic Enhancer Deletion C3	2.01 ± 0.021
Intergenic Enhancer Deletion C5	1.03 ± 0.13
Intergenic Enhancer Deletion C7	1.19 ± 0.0057
Intergenic Enhancer Deletion C8	0.69 ± 0.074
Intergenic Enhancer Deletion D7	1.85 ± 0.014
Intergenic Enhancer Deletion D8	1.78 ± 0.025
Intergenic Enhancer Deletion D9	0.79 ± 0.064
Intergenic Enhancer Deletion E10	1.22 ± 0.020
Intergenic Enhancer Deletion E7	0.78 ± 0.030
Intergenic Enhancer Deletion E8	1.89 ± 0.0058
Intergenic Enhancer Deletion F1	1.06 ± 0.044
Intergenic Enhancer Deletion F9	1.82 ± 0.024
Intergenic Enhancer Deletion G1	1.92 ± 0.073
Intergenic Enhancer Deletion G3	1.39 ± 0.033
Intergenic Enhancer Deletion G8	1.59 ± 0.041
Intergenic Enhancer Deletion H1	2.04 ± 0.026
Intergenic Enhancer Deletion H2	1.49 ± 0.044
Unedited	2.02 ± 0.015

Table 3.10: Copy number of Intergenic Enhancer Deletion CRISPR/Cas9 edited iPSC clones.
Copy number was determined by ddPCR for each cell colony. Colonies are named based on the well of the 96 well plate they were selected from. The copy number was expected to be 0, 1 or 2, however the results show colonies had copy numbers ranging from 0.69 to 2. The error is represented by SEM from three replicates.

3.3.5 Mosaicism

The unusual copy numbers found, such as 0.79 or 1.5, are indicative of mosaicism occurring. In this context, mosaicism, means the presence of multiple genotypes within a cell clone, and can arise from different factors. When colonies are expanded it is assumed that each cell in the colony is a clone from one individual cell and should therefore all share the same genotype. Mosaicism can occur when multiple cells merge to generate colonies and were therefore not derived of a single cell (Figure 3.14). It can also occur when the CRISPR/Cas9 enzyme complex remains active in the cell past the first cleavage division, thus producing multiple genotypes within a clone. This phenomenon is prevalent when the CRISPR/Cas9 complex is delivered to the cell as a plasmid, as plasmids can have different life spans within individual cells. A CRISPR/Cas9 plasmid may remain active and continue to make edits in some cells of a colony, but not in others. These scenarios could produce unusual copy numbers such as 0.79 or 1.5, as seen here. Moreover, mosaicism can be a serious issue when the desired CRISPR/Cas9 genome edited cells are out-competed by the non edited cells and they form the dominant genotype of the colony. This results in a loss of the desired genotype and can occur even if the edited cells make up a large proportion initially. This is more common when editing genes or regions that affect cell survival and proliferation. *TNFRSF1A* is not known to be an essential gene (Wang et al. 2015; Liao and Zhang 2008; Blomen et al. 2015). Removing it, however, could give those cells a small competitive disadvantage.

Single cell isolation in iPSCs is susceptible to mixed clone formation, since there is a bias towards the survival of small clusters of cells. To obtain a proper clone, a single cell must expand to form a colony, but when groups of cells have a higher success of survival this creates a bias in the system. Though steps are taken to reduce this bias, it is possible that isolated colonies are commonly derived from more than one cell and the chances of picking a purely clonal colony are low.

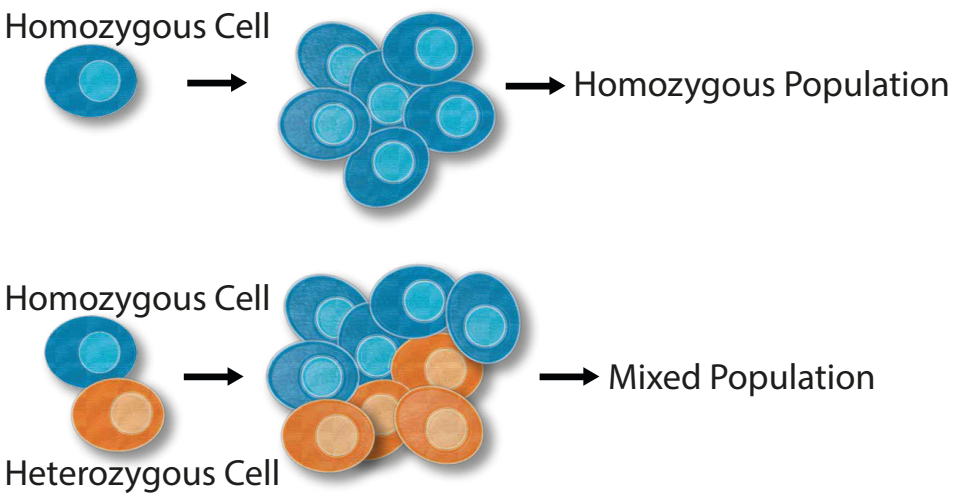


Figure 3.14: Schematic showing how mosacism can occur during clonal expansion following CRISPR/Cas9.
One way mosaicism can occur following CRISPR/Cas9 is if multiple cells of a different genotype join together to form a colony. This means the colony is not a clone of an individual cell, but from two cells with different genotypes. This can result in unusual copy numbers, such as 1.5. Other methods in which mosacism can arise is due to the CRISPR/Cas9 plasmid, as plasmids can have different life spans within individual cells.

3.3.6 Generating Homozygous Cell Lines

To overcome the issue of mosaicism and generate cell lines homozygous for the deletions, three approaches were used. The first was to change the CRISPR/Cas9 delivery method from a plasmid to an RNP. The second approach was to sub-clone the previously transfected populations to separate out the putative homozygous cells, and the third was to re-transfect the heterozygous cells in an attempt to delete the other allele.

CRISPR/Cas9 RNP Transfections

As described above, the CRISPR/Cas9 RNP method involves providing cells with CRISPR/Cas9 components that have already been transcribed and translated, and eliminates the need for cells to uptake a bulky plasmid. It is suggested to be more efficient at genome editing, so could perhaps produce homozygous deletion

cell lines more effectively (Kim et al. 2014; Schumann et al. 2015). Importantly, it could reduce mosaicism by eliminating the potential issue of a plasmid being active and creating edits in some cells of a clonal population, but not others. RNPs are cleared by the cell within 24-48 hours, but plasmids can remain active in a cell for several days (Liang et al. 2015; Gaj et al. 2012).

Fresh iPSCs were transfected using the RNP approach for all three deletions using the same gRNA sequences as used initially. Only one replicate was performed for each. No editing occurred at all in the Exon Deletion cells or the Intergenic Enhancer Deletion cells (Table 3.11). This was probably due to an error during the transfection process, but the experiment was not repeated due to time constraints. Intronic Enhancer Deletion colonies did, however, show signs of editing when analyzed using PCR. The zygosity of the cell colonies were, again, determined through PCR, gel electrophoresis, and ddPCR. The RNP transfection yielded three homozygous cell lines and eleven heterozygous cell lines for the Intronic Enhancer deletion.

Deletion Region	Homozygous Cells	Heterozygous Cells
Exon Deletion	0	0
Intronic Enhancer Deletion	3	11
Intergenic Enhancer Deletion	0	0

Table 3.11: iPSC zygosity after CRISPR/Cas9 transfections using an RNP approach. Number of cells either heterozygous or homozygous for each deletion of interest. A total of 60 colonies were screened per deletion.

Sub-Cloning Transfected iPSCs

Because some of the original plasmid transfected iPSC colonies had very low deletion copy numbers, such as 0.69, it is possible that cells homozygous for the deletion of interest were present in those putatively mixed cell populations. By re-plating those previously transfected cells at low-density and picking new clones, it may be possible to pull out homozygous cells. Three cell colonies were

selected for this approach due to their low copy number, and are listed in Table 3.12. Most of the cell lines selected contained the Intergenic Enhancer Deletion because they had a lower copy number than the other two deletions, and therefore had a higher chance of success.

Cell Line	Deletion Copy Number
Exon Deletion E7	1.56 ± 0.0058
Intergenic Enhancer Deletion C8	0.69 ± 0.074
Intergenic Enhancer Deletion D9	0.79 ± 0.064

Table 3.12: iPSC colonies selected for sub-cloning.
Cell colonies were selected for this approach due to their low copy number. The error is represented by SEM from three replicates.

A total of 11 colonies were selected and screened from the three Intergenic Enhancer Deletion sub-cloned colonies, and 18 from the Exon Deletion cell colony. The zygosity of the cell colonies were, again, determined through gel electrophoresis and ddPCR. The copy number results from each screened colony can be seen in Table 3.14. Two of the clones were homozygous for the Intergenic Enhancer Deletion and were derived from the original colony, Intergenic Enhancer Deletion C8 (copy number of 0.69) (Table 3.13). None of the colonies were homozygous for the Exon Deletion, but the copy number was reduced from the original 1.56.

Deletion Region	Homozygous Cells	Heterozygous Cells
Exon Deletion	0	18
Intronic Enhancer Deletion	NA	NA
Intergenic Enhancer Deletion	2	9

Table 3.13: iPSC zygosity after CRISPR/Cas9 after sub-cloning approach.
Number of cells either heterozygous or homozygous for each deletion of interest.

Deletion Region	Copy Number
Exon Deletion E7 subclone 1	0.68 ± 0.0047
Exon Deletion E7 subclone2	0.35 ± 0.0054
Exon Deletion E7 subclone 3	0.80 ± 0.0081
Exon Deletion E7 subclone 4	1.67 ± 0.013
Exon Deletion E7 subclone 5	0.99 ± 0.0004
Exon Deletion E7 subclone 6	1.10 ± 0.019
Exon Deletion E7 subclone 7	0.93 ± 0.0044
Exon Deletion E7 subclone 8	0.46 ± 0.0043
Exon Deletion E7 subclone 9	1.62 ± 0.0084
Exon Deletion E7 subclone 10	0.64 ± 0.015
Exon Deletion E7 subclone 11	1.07 ± 0.012
Exon Deletion E7 subclone12	1.19 ± 0.015
Exon Deletion E7 subclone 13	0.82 ± 0.0090
Exon Deletion E7 subclone 14	0.80 ± 0.0086
Exon Deletion E7 subclone 15	0.89 ± 0.0044
Exon Deletion E7 subclone 16	0.78 ± 0.0099
Exon Deletion E7 subclone 17	0.64 ± 0.0045
Exon Deletion E7 subclone 18	0.91 ± 0.0094
Intergenic Enhancer Deletion C8 subclone 1	1.73 ± 0.36
Intergenic Enhancer Deletion C8 subclone 2	2.06 ± 0.066
Intergenic Enhancer Deletion C8 subclone 3	0.00 ± 0.0001
Intergenic Enhancer Deletion C8 subclone 4	0 .00 ± 0.0006
Intergenic Enhancer Deletion C8 subclone 5	1.50 ± 0.18
Intergenic Enhancer Deletion D9 subclone 1	1.58 ± 0.015
Intergenic Enhancer Deletion D9 subclone 2	2.03 ± 0.032
Intergenic Enhancer Deletion D9 subclone 3	2.03 ± 0.11
Intergenic Enhancer Deletion D9 subclone 4	1.72 ± 0.015
Intergenic Enhancer Deletion D9 subclone 5	1.76 ± 0.023
Intergenic Enhancer Deletion D9 subclone 6	1.95 ± 0.13

Table 3.14: Copy number of Intergenic Enhancer Deletion and Exon Deletion CRISPR/Cas9 edited iPSC sub-clones.
Copy number was determined by ddPCR. Two of the Intergenic Enhancer Deletion colonies had a copy number of zero, meaning they are homozygous for the deletion. None of the Exon Deletion colonies are homzygous for the deletion, but the copy number was reduced from the original 1.56. The error is represented by SEM from three replicates.

Re-Transfections of Edited iPSCs

The final approach to generate homozygous deletion cell lines was to re-transfect a previously edited cell line. Cell colony, Exon Deletion E7, was selected. It had a deletion copy number of 1.56, which was the lowest determined copy number of Exon Deletion cell lines. It was re-transfected using the RNP approach with the same gRNA sequences as were used initially. It was then plated at low-density and cell colonies were selected and screened. The zygosity of the cell colonies were, again, determined through gel electrophoresis and ddPCR. A total of 29 colonies were selected and screened. The copy number was reduced in nearly all the samples, but none of them were homozygous for the deletion (Table 3.15).

Genome Integrity Check of Edited Cells

Genome integrity was assessed by an Illumina BeadChip array, evaluating approximately 710,000 markers. This assay detects genomic changes of 1MB or greater around common variants and an analysis at a higher resolution was not carried out. All edited samples were compared to their parental controls and showed no gross abnormalities indicative of affecting their growth or ability to differentiate. The resulting karyograms can be seen in Figures A.16 to A.24 in Appendix A.

Successful Cell Line Generation

Following CRISPR/Cas9 RNP transfections and sub-cloning of previously edited cells, three homozygous colonies for the Intronic Enhancer Deletion and two colonies for the Intergenic Enhancer Deletion were generated (Table 3.16). No homozygous cell lines were generated for the Exon Deletion. The Exon Deletion was not pursued further due to time constraints.

Sample	Deletion Copy Number
Exon Deletion B1	1.04 ± 0.025
Exon Deletion C1	1.81 ± 0.013
Exon Deletion D1	1.24 ± 0.017
Exon Deletion E1	0.95 ± 0.0060
Exon Deletion F1	0.72 ± 0.0040
Exon Deletion A2	0.71 ± 0.014
Exon Deletion B3	1.79 ± 0.029
Exon Deletion C4	1.46 ± 0.011
Exon Deletion C5	2.11 ± 0.0072
Exon Deletion B2	1.01 ± 0.0097
Exon Deletion C3	1.34 ± 0.011
Exon Deletion D4	1.07 ± 0.015
Exon Deletion C2	1.99 ± 0.084
Exon Deletion D3	1.08 ± 0.015
Exon Deletion D2	1.00 ± 0.010
Exon Deletion E3	1.86 ± 0.0079
Exon Deletion E4	1.03 ± 0.016
Exon Deletion F2	1.06 ± 0.045
Exon Deletion F3	2.14 ± 0.0054
Exon Deletion G4	1.69 ± 0.015
Exon Deletion G2	1.47 ± 0.012
Exon Deletion G3	1.06 ± 0.0054
Exon Deletion H4	0.92 ± 0.029
Exon Deletion C6	0.81 ± 0.061
Exon Deletion H2	0.81 ± 0.0044
Exon Deletion B4	1.33 ± 0.0066
Exon Deletion B5	1.07 ± 0.032
Exon Deletion E6	0.92 ± 0.019
Exon Deletion G6	1.11 ± 0.0002
Unedited	2.04 ± 0.015

Table 3.15: Copy number of Exon Deletion CRISPR/Cas9 edited iPSC clones after re-transfection.
Copy number was determined by ddPCR. Cell line Exon Deletion E7 with a copy number of 1.56 was re-transfected using an RNP approach. The copy number was reduced in nearly all the samples, but none of them were homozygous for the deletion. The error is represented by SEM from three replicates.

Cell Colony Number	Genotype	Thesis Nomenclature
Intronic Enhancer Deletion B6	Edited	Intronic Enhancer Deletion 1
Intronic Enhancer Deletion C5	Edited	Intronic Enhancer Deletion 2
Intronic Enhancer Deletion F7	Edited	Intronic Enhancer Deletion 3
Intergenic Enhancer Deletion C8 subclone 3	Edited	Intergenic Enhancer Deletion 1
Intergenic Enhancer Deletion C8 subclone 4	Edited	Intergenic Enhancer Deletion 2
Intronic Enhancer A1	Unedited CRISPR Control	Intronic Enhancer CRISPR Control 1
Intronic Enhancer B1	Unedited CRISPR Control	Intronic Enhancer CRISPR Control 2
Intergenic Enhancer C2	Unedited CRISPR Control	Intergenic Enhancer CRISPR Control 1
Intergenic Enhancer G1	Unedited CRISPR Control	Intergenic Enhancer CRISPR Control 2
SFC840-03-03	Unedited Parental Control	Parental Control

Table 3.16: iPSCs taken forward to macrophage differentiation.
Three edited Intronic Enhancer Deletion colonies, two Intergenic Enhancer Deletion colonies, two CRISPR controls for each enhancer, and the unedited parental cell line were differentiated into macrophages. The original colony name for each cell line is listed above along with its name used for the remainder of the this thesis.

3.3.7 Macrophage Differentiation

Once the desired cell lines had been generated, they were differentiated into macrophages for experimentation. For details on the macrophage differentiation protocol see Section 2.2 in Chapter 2. Along with the CRISPR/Cas9 edited cells, control cell lines were also taken forward for differentiation. Cell lines henceforth referred to as CRISPR controls are cells that underwent the

CRISPR/Cas9 transfection process, including gRNA exposure, but remained unedited at the target deletion site (Table 3.16). The cells that underwent macrophage differentiation were three homozygous deletion colonies for the Intronic Enhancer Deletion, two CRISPR Controls for the Intronic Enhancer Deletion, two homozygous deletion colonies for the Intergenic Enhancer Deletion, two CRISPR controls for the Intergenic Enhancer Deletion, and the unedited iPSC parental cell line. The parental iPSC line was not exposed to any of the CRISPR/Cas9 protocol. The experimental results of the edited iPSC derived macrophages can be found in Chapter 4.

To validate the macrophage phenotype, fluorescence-activated cell sorting (FACS) was performed. CD11b and CD14 were used as markers of successful differentiation into macrophages. CD11b is a cell surface marker found on monocytes, macrophages, neutrophils, and NK cells, that is responsible for leukocyte adhesion and cell migration during an inflammatory response. CD11b forms a heterodimer with CD18 to create a transmembrane receptor involved in TLR signaling and the LPS response (Ling et al. 2014; Triantafilou and Triantafilou 2002). It also has a role in the phagocytosis of apoptotic cells (Chen et al. 2008). CD14 is strongly expressed by monocytes and macrophages. It is a cell surface endotoxin receptor that binds directly to LPS and chaperones it to TLR4 to induce a signaling cascade that results in NF-κB activation and initiation of an inflammatory response (Wright et al. 1990; Silva Correia et al. 2001; Zanoni et al. 2011). These markers are not expressed on iPSCs. After differentiation, the majority of cells expressed both CD11b and CD14, indicating that the iPSCs were successfully differentiated into macrophages. An example of one of the FACS profiles can be seen in Figure 3.15. The genotypes of the cells were re-confirmed following differentiation. The deletions are not believed to have affected the differentiation process.

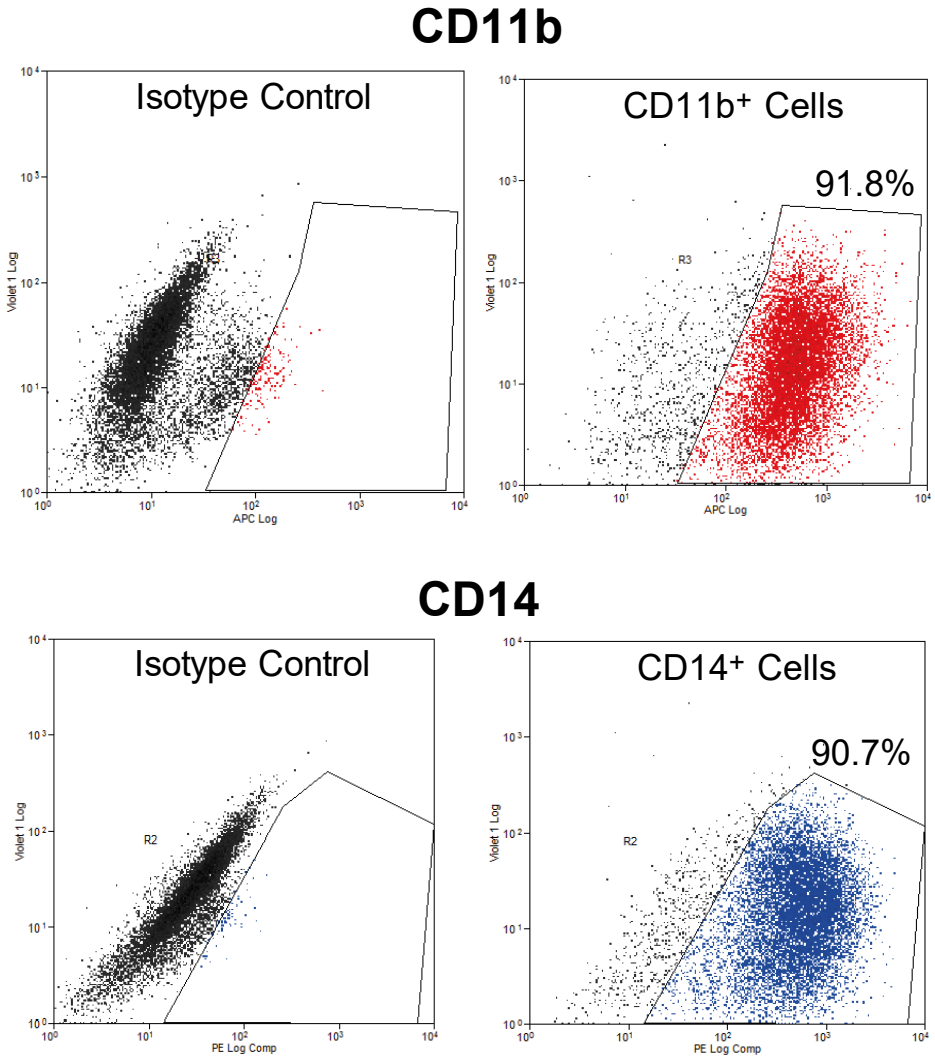


Figure 3.15: FACS profile of iPSC derived macrophages.
One of the cell lines of iPSC derived macrophages (CRISPR Control) following FACS analysis with CD11b (red) and CD14 (blue) antibodies. Over 90% of the cells were $CD11b^+$ and over 90% were $CD14^+$, indicating that the iPSCs were successfully differentiated into macrophages. The other cell lines examined shared a similar profile.

3.4 Discussion

CRISPR/Cas9 technology is an invaluable tool in the field of functional genomics and allows for new understanding of GWAS variants and regions. However, as this chapter has illustrated, utilizing CRISPR/Cas9 is complex and requires significant optimization. Despite the unexpected challenges, it can still be used to generate cell lines homozygous for deletions of interest and interrogate putative regulatory regions.

3.4.1 Variant Prioritization for CRISPR/Cas9 Work

Pi is a useful tool for variant prioritization in autoimmune diseases and can be used in conjunction with manual curation of regions and in-house datasets. It pulls together many datasets and gene annotations to rank known GWAS hits as well as novel genes, ranked through interactions. It could provide post-GWAS studies with a realistic and feasible way of prioritizing genes and SNPs for functional experiments.

A preliminary version of Pi ranked *TNFRSF1A* first for AS based on its role in inflammation and immunity, and its association through GWAS. The current version of Pi ranks it at 13 out of around 15,000 genes. It is no longer the highest ranked gene, but is still strongly prioritized in AS using a more comprehensive version of Pi. Therefore, its predicted importance in AS is robust based on genetic studies and annotation predictors, and remains an important AS GWAS hit.

3.4.2 CRISPR/Cas9 in K562 Cells

CRISPR/Cas9 in K562 cells proved to be a challenge. The experiment was carried out as a proof of principle to determine if deleting an enhancer in a cell line would have a large enough effect to be seen through expression changes in a bulk population of edited cells. No significant changes in expression were observed across any of the five enhancer deletions. This could be because removing these enhancers does not create a large enough change in expression to be seen by qPCR, and perhaps RNA-Seq would be a better method to measure expression changes. However, it is likely due to the fact that K562 cells are roughly triploid, so it is difficult to remove enough copies of the enhancer to see any effect. Moreover, studies report that when completely removed or disrupted, the *GATA1* and *MYC* enhancers result in a reduction of approximately 40% in gene expression (Fulco et al. 2016). In a population with so many wild-type copies still present, this small reduction may be impossible to see.

To try and improve upon this problem, *RUNX1* enhancer deletion clones were isolated and expanded. Around 70% of the selected clones showed evidence of having a deletion, which suggests that the bulk edited population had a similar percentage of cells harbouring the deletion. Despite this high efficiency of deletion, the impact on gene expression was not found to be significant. Therefore, clones were screened for zygosity to determine if fully removing the enhancer would have a significant effect on gene expression. However, no homozygous clones were produced. All of the clones that underwent ddPCR analysis still had at least two copies of the *RUNX1* enhancer and only one clone showed a slight change in gene expression, but it was in the opposite direction as expected. It is also difficult to determine how accurate the ddPCR results were in this cell line. The copy number of the *RUNX1* enhancer was calculated by comparing enhancer copies to copies of the reference gene, *RPP30*. The calculations are based on the assumption that the cells have two copies of *RPP30*, but this was not necessarily the case. As shown in Table 3.4 following the FISH analysis, the largest proportion of the cells did have two copies of the gene, but a significant amount had only one copy and some had none at all. This could have affected the calculated *RUNX1* enhancer copy number and made it inaccurate. *RPP30* is the recommended reference gene according to Bio-Rad, the equipment suppliers, but perhaps a thorough investigation of the K562 genome would reveal an alternative with a more consistent copy number. Although, K562 cells are very genetically heterogeneous with many translocations and chromosomal abnormalities so finding a robust control gene may not be realistic.

This experimental approach could have potentially benefited from an alternative set of enhancer controls. None of the five controls used here, enhancers for *LMO2*, *MYC*, *GATA1*, *RUNX1*, and *HGB1* had any effect on gene expression when deleted. Both *MYC* and *GATA1* are involved in K562 cellular proliferation (Fulco et al. 2016). It is possible that the cells harbouring deletions for their enhancers

were out-completed in culture by unedited cells and, therefore, no effect on gene expression was observed. This experimental approach was unsuccessful in finding gene expression changes associated with the deletion of enhancers known to regulate them, and would therefore not be suited for characterizing putative AS enhancers.

A different CRISPR/Cas9 approach may be more successful in future experiments. Instead of using a gRNA on either side of the deletion, editing success rates could be increased by just using one gRNA to disrupt an enhancer. However, prior knowledge of the enhancer would be required to know where to introduce INDELs to have an effect. Disrupting enhancers is better suited in CRISPR/Cas9 screens with gRNAs covering an entire region to increase the chance of finding a significant result (Canver et al. 2015; Sanjana et al. 2016; Korkmaz et al. 2016). Other cell lines may prove more successful for this type of analysis, but immortalized cell lines typically have have more than 46 chromosomes, making CRISPR/Cas9 experiments challenging.

Primary cells may be an alternative to immortalized cell lines. They can be difficult to culture and transfections are often inefficient and cause cellular stress, but they have a normal human karyotype and better recapitulate human biology (Peng et al. 2016; LaFleur et al. 2019). Multiple studies have successfully edited primary immune cells using CRISPR/Cas9 (Gundry et al. 2016; Hendel et al. 2015; Schumann et al. 2015). However, these protocols can be difficult to optimize and are expensive if trying to increase the throughput for a CRISPR/Cas9 screen. Moreover, it is ideal for the cell line to already be expressing Cas9 for large-scale studies, which is more easily achieved using an immortalized cell line. A novel method was recently developed to perform CRISPR/Cas9 screens in primary human T cells, called Guide Swap (Ting et al. 2018). Guide Swap combines both RNP transfections and lentiviral transductions. They found that they could effectively edit primary T cells by

delivering the Cas9 to the cells as a non-targeting RNP, and directing the editing events through gRNAs delivered by a lentivirus. The researchers found that their unconventional merge of two delivery techniques was more effective than using either one alone, and found it to be robust and reproducible across multiple targets and cell types. Primary cells are preferred to cell lines as they are more biologically relevant, but their use in genome editing is limited by their difficulty to effectively transfect and edit. With more novel CRISPR/Cas9 techniques being brought to light, editing primary cells using CRISPR/Cas9 could present a better alternative to immortalized cell lines.

3.4.3 Mosaicism in CRISPR/Cas9

RNPs

One of the biggest issues uncovered whilst editing iPSCs was mosaicism in isolated cell clones after CRISPR/Cas9 transfections. Mosaicism is an emerging problem in CRISPR/Cas9 genome editing. It is an issue when trying to isolate and expand cells of a particular genotype and can affect animal model generation by creating an animal with cells of different genotypes (Nakade et al. 2014; Guo and Li 2015; Sato et al. 2015; Tu et al. 2017). One potential solution is to eliminate the use of plasmids in CRISPR/Cas9 editing. Plasmids can remain in cells for days and cause active cutting and repairing of DNA (Gaj et al. 2012). Not only does this increase the chance of off-target effects, but it also allows for mosaicism to occur.

When cells are proliferating, some cells may retain the plasmid while others will lose it. Moreover, the enzymes the plasmids produce can persist past the cell division. Within one cell, a single copy of the gene may be edited while the other remains wild-type, producing a heterozygous cell. If the CRISPR/Cas9 complex persists after cell division, there is chance that the wild-type copy will be edited differently in the resulting daughter cells and produce different mutations. This

can cause cells in a clonal population to exhibit different genotypes. The strategy employed here to combat this when using CRISPR/Cas9 plasmids in iPSCs was to culture the cells as a pool for several days before plating at low density plating for clone generation. This was expected to minimize the effects of the varying duration of Cas9/gRNAs and reduce mosaicism. However, using an RNP method is a further improvement on CRISPR/Cas9 methods. RNPs are cleared by the cell within 24-48 hours, making it less likely to keep cutting beyond the desired edits (Liang et al. 2015). Additionally, using RNPs removes laborious gRNA cloning prior to transfection, works faster because it is not relying on cellular machinery, and is much easier for cells to uptake, which is important when working with sensitive cell types like primary cells or iPSCs. The RNP method was successful in generating three homozygous deletion cell colonies for the Intronic Enhancer Deletion after just one transfection. One downside to RNPs is that you cannot easily determine transfection efficiency or select for that cells have been edited. When working with plasmids, they can be engineered to contain cassettes for a reporter such as GFP and/or a selection marker such as puromycin. This allows for the assessment and optimization of transfection conditions and to select those cells that have taken up the CRISPR/Cas9 components through puromycin treatment. One study attributed mosaicism to prolonged expression of Cas9 in cells and embryos (Tu et al. 2017). They successfully reduced the half-life of Cas9 by tagging it with the ubiquitin-proteasome degradation signal on the N-terminus and they found that this reduced the number of mosaic mutations in monkey embryos. This approach could become more mainstream and brought into cell culture to combat mosaicism.

Sub-Cloning

Sub-cloning as an approach to combat mosaicism was successful, but only in cells that already had a very low copy number of the enhancer. The colony that was successfully sub-cloned to produce homozygous deletion colonies for

the Intergenic Enhancer Deletion had a copy number of 0.69. This approach was unsuccessful for the Exon Deletion cell line, which had a copy number of 1.56 prior to sub-cloning. Sub-cloning did lower the copy number of the Exon Deletion cell line, but none of the isolated sub-clones were homozygous for the deletion. It is possible that sub-cloning again could have produced cell lines that were homozygous for the deletion. However, this is a very costly, slow and laborious process, which increases the passage number of these critical stem cells, and was not feasible to continue, given time constraints.

Re-transfecting

Re-transfecting was not successful in generating homozygous deletion colonies. It did produce colonies with a lower deletion copy number than the original 1.56, but no sub-clones were isolated with a copy number of zero. This could be because the gRNA recognition sites and cut sites were damaged during the first round of editing, and therefore the CRISPR/Cas9 complex was unable to find and edit the correct segment of DNA. Thorough sequencing of the cell line intended for use could avoid this issue by ensuring that the cut sites are still unaltered. However, when the cell line is a heterogeneous population with some cells containing a deletion, it is very hard to tell by Sanger sequencing if the cells are suitable for re-transfection. One possible workaround is to clone a PCR product spanning the region in question into a pGEM-T easy vector to sequence many clones to confirm the sequence of both alleles. Additionally, another way to avoid this issue would be to design new gRNAs that still target the same area, but recognize a different sequence that should be unedited from the first transfection. The gRNAs used for re-transfections in the is experiment were identical to the ones used in the first round of editing.

Other Solutions

Another issue that may lead to mosaicism is during clonal isolation. Cells need to be separated out from each other to grow into a truly clonal population. If cells of different genotypes join together to form a colony, then a mixed population forms. This can be difficult to get around and is dependent on cell type constraints. iPSCs are a very sensitive cell type and need to be plated at low density on mouse embryonic fibroblasts (MEFs) to form colonies that are individually picked by hand. iPSCs rely on cell-cell interactions for their survival and will die in isolation. This can be partly overcome by addition of Rho kinase inhibitor and growing the cells on MEFs. However, plating the cells at low density on MEFs means it is hard to track whether a clone originated from a single cell. There are other methods to separate out single cells that are frequently used on other cell lines, such as FACS or limiting dilutions (Hu et al. 2016). A limiting dilution method was used for the K562 clonal isolation in this thesis. No homozygous deletion cell lines were generated using this method, although it is uncertain whether this is due to improper separation into single cells, or simply because the cell line is so heterogeneous and with a complicated karyotype. The development of new techniques for clonal isolation and expansion could help eliminate mosaicism and make CRISPR/Cas9 editing more effective.

3.4.4 Future Experiments

Unfortunately Exon Deletion homozygous clones were not produced. With more time, however, it could hopefully have been achieved, assuming that a cell line with a *TNFRSF1A* knockout is not competitively disadvantageous. Repeating the RNP transfection may be efficient enough to produce homozygous clones after one transfection, as it was with the Intronic Enhancer Deletion. No editing for the exon deletion occurred when the RNP method was tried on fresh iPSCs. This is likely due to human error during the experimental procedure and repeating

the experiment would likely yield different results. Further sub-cloning of the already sub-cloned cells may have also been successful if tried. After sub-cloning and re-transfection of a previously edited cell line, the deletion copy number of several of the clones was less than one. One of the clones had a copy number as low as 0.35. Further sub-cloning would likely result in the generation of cell colonies homozygous for the Exon Deletion. Changing the deletion approach could also generate the desired edited clones, as there are a number of factors that effect the efficiency of CRISPR/Cas9 editing (Campenhout et al. 2019). Perhaps the gRNAs selected were not that efficient. One study found that gRNAs are highly variable in efficiency and can effect the success of a gene knockout (Wroblewska et al. 2018). New gRNAs could be designed to target the same four exons, but recognize slightly different regions to improve efficiency. Completely changing the region being targeted could also work to knockout the gene. Targeting a different segment of the gene to remove or disrupt may prove more effective than the originally selected four exon gene segment.

The two specific SNPs, rs887477 and rs11064145, that are situated within the enhancers targeting for deletion were initially part of the plan to be edited, using homology directed repair to achieve allelic substitution. Constructs for these edits were designed and tested in HEK-293 cells and prepped for use in iPSCs. However, they were never used to generate the SNPs in iPSCs. The deletion edits were carried out first to stagger the work and proved to be much more challenging and time consuming than originally thought. Due to the timeline of this thesis, the SNP generation had to be put on hold. Time permitting, it would have been ideal to generate cell lines homozygous for the two SNPs and compare the effects of the SNPs on gene expression and chromosomal organization and see how they differ from the enhancer deletion cell lines. Soldner et al. was able to generate SNPs associated with Parkinson's disease using CRISPR/Cas9 in iPSCs and characterize a risk variant, rs356168 (Soldner et al. 2016). Another

study used CRISPR/Cas9 to examine the effect of prostate cancer GWAS SNPs and found that two SNPs in LD with each other, rs7098889 and rs10993994, independently altered the expression of *MSMB*, which is strongly associated with prostate cancer (Wang et al. 2018b). Generating rs887477 and rs11064145 in iPSCs for differentiation would help functionally characterize the SNPs, as well as the enhancers they are situated within.

3.4.5 High Throughput CRISPR/Cas9 Assays

This method of functionally characterizing risk variants for AS is low-throughput. It was thought that this approach could be used to generate a number of SNPs and regulatory regions associated with AS to better understand its pathology. However, this method was slow and required a lot of work to generate cell lines with the desired deletions. Pooled CRISPR/Cas9 screens would be an improvement upon the method employed in this thesis. CRISPR/Cas9 screens allow for the interrogation of many genetic risk loci associated with disease through GWAS. They were not attempted in this thesis due to time constraints, but would be a promising approach for future work. They have already proven successful in understanding the non-coding genome by using large libraries of gRNAs to associate specific regions with involvement in disease (Shalem et al. 2014). They are also enhancing the understanding of immune cells in autoimmune disease and cancer. A recent study used a genome-wide retroviral CRISPR knockout library combined with functional genomics techniques, such as RNA-seq, ATAC-seq, and ChIP-seq, to look for factors influencing T helper type 2 (Th2) cell activation and differentiation (Henriksson et al. 2019). From this, they were able to validate known factors involved in activation and differentiation and uncover novel ones for a more compressive understanding of their regulatory network. Variations of the traditional pooled CRISPR/Cas9 screen are also becoming popular. CRISPR activation (CRISPRa) is used to uncover and characterize putative enhancers

by fusing enzyme-inactive Cas9 to VP64, a viral protein used for activation of transcription. CRISPRa screens are gaining momentum and they have been used in a multitude of studies, including uncovering genes and mechanisms involved in melanoma drug resistance (Konermann et al. 2015), and identifying novel factors and mechanisms involved in cellular reprogramming (Yang et al. 2019). A CRISPRa screen could present a promising alternative to the work presented here. In this thesis, two putative enhancers were deleted, but in a time consuming way. CRISPRa would allow for the interrogation of many enhancers, across AS associated risk loci, in a single CRISPR/Cas9 experiment. It even has the ability to find enhancers that are only active under certain conditions and has been used successfully in the study of autoimmune disease (Simeonov et al. 2017).

3.5 Conclusions

Variant prioritization of GWAS remains an ongoing challenge, but new tools such as Pi can be useful in combining evidence from different sources to enable post-GWAS functional studies. The selection of a model system for AS research remains difficult. K562 cells are not a suitable cell line for examining putative regulatory regions associated with AS. That approach was unsuccessful in showing an effect when validated enhancers were deleted using CRISPR/Cas9, and it is likely that other immortalized cell lines would suffer from the same limitations. CRISPR/Cas9 genome editing in iPSCs was successful in generating deletions of putative enhancers associated with AS, after much troubleshooting. Three cell clones homozygous for the Intronic Enhancer Deletion and two cell clones homozygous for the Intergenic Enhancer deletion were generated, and they were successfully differentiated into macrophages. iPSC derived macrophages provide a biologically relevant system to examine the effects of these enhancers on gene expression, and chromatin accessibility and interactions (experimental results are discussed in Chapter 4).

4

Functional Characterization of Putative Enhancers Within *TNFRSF1A* Using iPSC Derived Macrophages

Imagination is more important than knowledge. -Albert Einstein

4.1 Introduction 131
4.2 Results 139
4.3 Discussion 182
4.4 Conclusions 192

4.1 Introduction

4.1.1 TNFRSF1A and Inflammation

Tumor Necrosis Factor Receptor Superfamily Member 1A (*TNFRSF1A*), also commonly known as Tumor Necrosis Factor Receptor Type 1 (*TNFR1*) was the selected gene for initial validation of the CRISPR/Cas9 edited iPSC derived macrophage model (Chapter 3). It is a key regulator of the inflammatory response and anti-TNF biologic therapy is the most commonly prescribed treatment for AS (Garcia-Montoya et al. 2018), emphasizing the importance of the TNF family and its signaling pathways in inflammation. *TNFRSF1A* encodes a 55KDa

transmembrane receptor for TNF that is constitutively expressed on almost all cell types (Tracey et al. 2008). Most of the pro-inflammatory functions of TNF are mediated through TNFRSF1A (Ksontini et al. 1998; Pfeffer et al. 1993; Peschon et al. 1998). TNF is found in both soluble and membrane-bound forms, as is its receptor, TNFRSF1A. Binding of extracellular TNF results in a signaling cascade through the recruitment of adapter proteins that interact with the cytoplasmic domain of the receptor (Figure 4.1) (Tracey et al. 2008). The cytoplasmic domain of membrane-bound TNFRSF1A contains a death domain that allows for the binding of the adapter protein, TRADD, which initiates two signaling pathways. One pathway results in apoptosis of the cell, while the more prominent pathway results in the activation of NF-κB to increase the expression of pro-inflammatory genes. Under normal circumstances, apoptosis is inhibited by FADD-like IL-1β-converting enzyme-inhibitory protein, which is induced by NF-κB for cell survival (Safa 2012; Tracey et al. 2008). If the cell becomes compromised and NF-κB is no longer inhibiting the apoptosis pathway, then the TNF/TNFRSF1A complex is internalized into endocytic vesicles for the aggregation of TRADD and FADD, to initiate cell death (Schneider-Brachert et al. 2004). TNFRSF1A can also cause natural cellular desensitization to TNF to inhibit inflammation. The extracellular domain of TNFRSF1A can be cleaved and shed from the cell surface through the TNF converting enzyme, which results in a decreased number of available membrane-bound receptors for TNF to interact with (Ermert et al. 2003; Xanthoulea et al. 2004). Moreover, soluble TNFRSF1A can bind to free TNF and act as a natural inhibitor by preventing TNF binding with membrane-bound receptors (Xanthoulea et al. 2004).

4.1.2 TNFRSF1A and Disease

TNFRSF1A is an important regulator of TNF and the inflammatory response, and has been associated with multiple immune disorders. GWAS have uncovered associations between *TNFRSF1A* and AS (Cortes et al. 2013), primary sclerosing

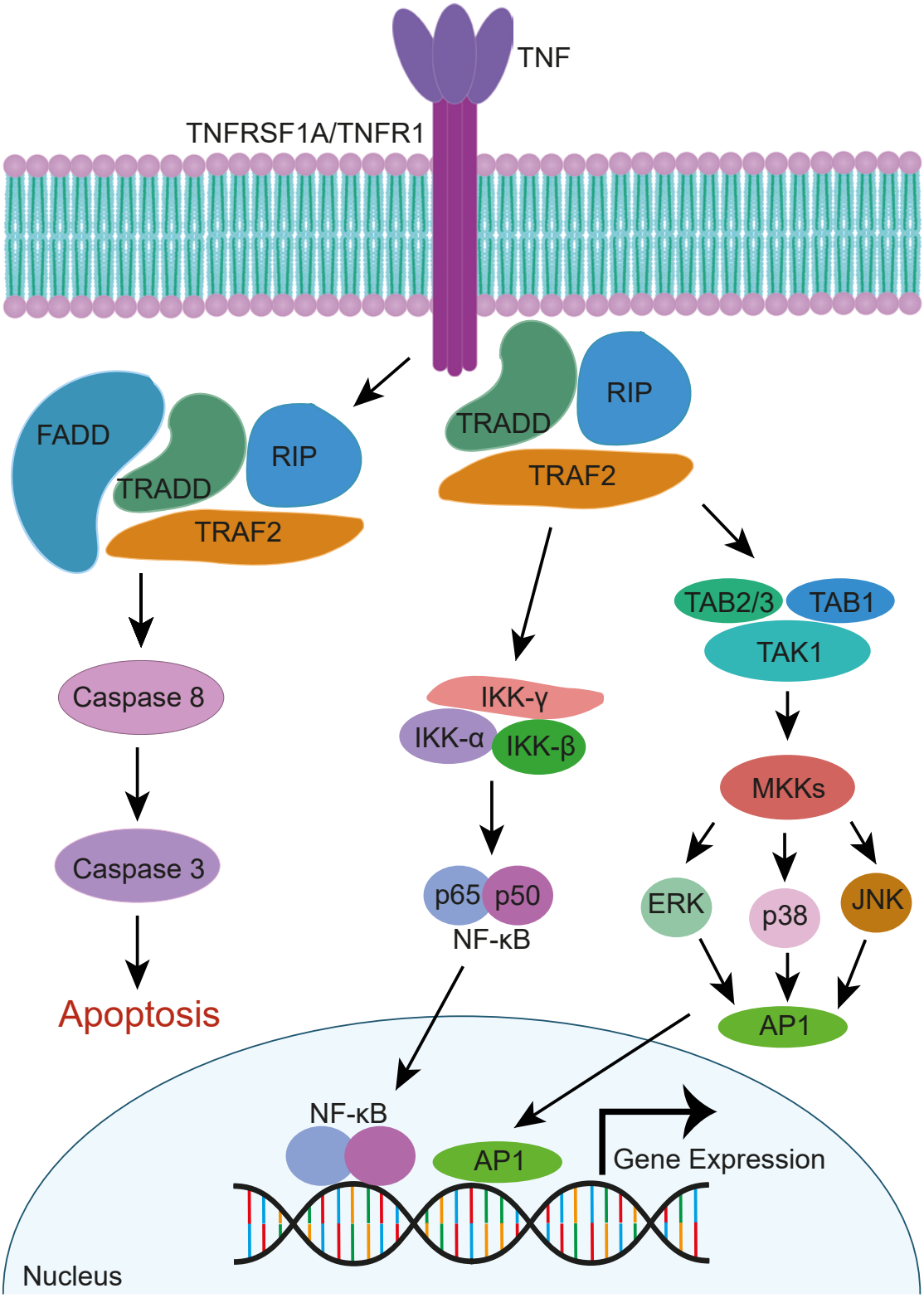


Figure 4.1: Overview of the TNFRSF1A/TNFR1 signaling pathway.

cholangitis (Ellinghaus et al. 2016), ulcerative colitis (Ellinghaus et al. 2016), primary biliary cholangitis (Mells et al. 2011), psoriasis (Ellinghaus et al. 2016), Crohn's disease (Liu et al. 2015), and MS (Sawcer et al. 2011; Beecham et al. 2013; De Jager et al. 2009). It has also been associated with cancer and RA (Egusquiaguirre et al. 2018; Roux-Lombard et al. 1993).

One of the early diseases linked with *TNFRSF1A* was Tumor necrosis factor receptor associated periodic syndrome (TRAPS) (McDermott et al. 1999). TRAPS is an inflammatory disorder characterized by fever, abdominal pain, vomiting, conjunctivitis, myalgia, arthralgia, and chest pain (Dodé et al. 2002; Hull et al. 2002). Multiple genetic variants within *TNFRSF1A* have been associated with TRAPS. The original association study, which named the disease, suggested that TRAPS was a result of defective shedding of the extracellular domain of TNFRSF1A, due to genetic defects (McDermott et al. 1999). Mutations within the coding region could disrupt the conserved extracellular disulfide bonds. One particular mutation they identified, C52F, inhibits post-stimulatory TNFRSF1A cleavage and clearance, which reduces the amount of the soluble TNFRSF1A available to naturally inhibit TNF, leading to increased inflammation. However, not all mutations associated with TRAPS affect clearance, but it is evident that TNFRSF1A is important in correct TNF signaling and inflammation modulation.

Another disease strongly associated with *TNFRSF1A* is MS. The most strongly associated MS SNP within the *TNFRSF1A* locus is rs1800693 (Sawcer et al. 2011). It is situated at the boundary of exon 6 and intron 6, and results in a splicing mutation that causes skipping of exon 6 to produce a shortened protein isoform (Gregory et al. 2012; Chen and Manley 2009). This isoform lacks the extracellular c-terminal portion, the transmembrane domain, and the intracellular region, including the death domain, compared to the full length protein, which are important in protein localization, signaling, and function (Gregory et al. 2012; Schütze et al. 2008). High levels of this truncated protein exist as a soluble

protein, where it binds to TNF to dampen the inflammatory response (Gregory et al. 2012). A recent study screened 29 childhood onset MS patients and found that 38% of them had mutations within exons of *TNFRSF1A* (Blaschek et al. 2018). One hypothesis is that deleterious mutations within the gene cause the improperly formed receptor to accumulate in the ER and result in a stress response and inflammation (Caminero et al. 2011; Probert et al. 2000). One coding SNP, rs4149584 (R92Q), was found in 20% of their patient population and a previous study had found that MS patients with this mutation had an increased amount of soluble TNFRSF1A in their serum (Comabella et al. 2013). This mutation is also associated with TRAPS, which can sometimes coexist in MS patients (Kümpfel et al. 2007; Kümpfel et al. 2008). Interestingly, anti-TNF drugs are not successful in treating MS, and often exacerbate disease (Kemanetzoglou and Andreadou 2017). Patients with R92Q have been found to experience more side effects when treated with biologics (Kümpfel et al. 2008). It is possible that this mutation is impacting the cleavage and shedding of TNFRSF1A, and therefore dysregulating the cellular response to TNF. Whatever the mechanism of action may be, it is evident that TNFRSF1A is crucial to the normal functioning of the inflammatory response and that mutations in this gene can be detrimental and contribute to disease.

4.1.3 Characterizing Putative Enhancers

As described in Chapter 3, deletions of two putative enhancers at the AS associated locus, *TNFRSF1A*, were generated in iPSCs and differentiated into macrophages. Examining chromatin interactions, epigenetic marks, and gene expression levels are vital in understanding and characterizing enhancers, and can be assayed using a variety of techniques in edited cell lines. ATAC-Seq, ChIP-Seq, Capture-C, and RNA-Seq are all valuable functional genomic tools and can be used to characterize putative enhancers.

Regulatory elements are found in regions of the genome with accessible chromatin so they may be bound by proteins to modulate gene expression, or interact with other regions of chromatin, and these can be assayed using ATAC-Seq. ATAC-Seq is a relatively new technique designed to examine regions of open chromatin in less time and with fewer cells than classical DNase hypersensitivity assays by using 'tagmentation' to simultaneously tag and fragment the DNA (Buenrostro et al. 2013). It utilizes the transposase enzyme to cut the DNA where it is accessible and produces sequencing reads that can be used to map open chromatin. ATAC-Seq was used successfully by Davie et al. when they screened for regulatory regions that become active during oncogenesis (Davie et al. 2015). The researchers discovered that the transcription factors AP-1 and Stat92E were key regulators in tumorigenesis by binding to over-activated regions and inducing the JAK/STAT signaling cascade. To examine the impact of genetic variation on chromatin accessibility, Gate et al. used ATAC-Seq on $CD4^+$ T cells (Gate et al. 2018). They found that stimulation had a large affect on regions of open chromatin and, when stimulated, a larger proportion of accessible regions overlapped with SNPs associated with autoimmune disease. They found that roughly 15% of SNPs within defined ATAC peaks affected the accessibility of that peak, and only 5% of these variants resided in transcription factor binding motifs. These studies highlight the complexity of variation driving chromatin accessibility and the importance of understanding its role in complex disorders.

As well as understanding which regions of chromatin are accessible, it is important to know the factors influencing chromatin state. ChIP-Seq is an assay used to examine DNA-protein interactions including transcription factors and histone marks. ChIP-Seq involves the crosslinking of DNA to DNA-binding proteins by formaldehyde fixation, followed by sonication and addition of an antibody for the protein of interest for immunoprecipitatation. The sample is then reverse crosslinked to dissociate the DNA and protein so the DNA can be

sequenced. ChIP-Seq can be used to map histone modifications to determine chromatin states. The methylation and acetylation of particular amino acids on histones can alter the chromatin state to increase or decrease its accessibility and regulate gene expression. For example, the presence of H3K27ac is indicative of an active enhancer element, and is therefore useful in interpreting the role of a putative regulatory element (Creyghton et al. 2010). H3K4me3 is a histone mark for active promoters and can be assayed to determine gene expression and activation (Hon et al. 2009; Heintzman et al. 2007). The first genome-wide study to examine histone modifications was in 2007 (Barski et al. 2007). Since then, many studies and projects, including ENCODE (Dunham et al. 2012) and the NIH Roadmap Epigenomics Mapping Consortium (Bernstein et al. 2010), have used ChIP-Seq to interrogate transcription factor binding and histone modifications across the genome to enhance the understanding of epigenetic regulation of gene expression.

Chromatin conformation methods are also being used in distilling the function of a non-coding genomic region. Regulatory elements may physically interact with other regions of chromatin to regulate gene expression. Knowing the physical interactions occurring between genomic regions can aid in the understanding and characterization of a putative regulatory region, making experiments like Capture-C incredibly valuable. In Capture-C, cells are cross-linked with formaldehyde to preserve physical interactions and then regions of interest can be baited and sequenced to reveal chromatin interactions (Hughes et al. 2014; Davies et al. 2016). This technique can reveal an enhancer-gene interaction, or the targets of other regulatory elements, and can be used to validate RNA-Seq or ChIP-Seq data. Capture-C can also provide information about topologically associated domains (TADs) to better understand which segments of DNA preferentially interact with one another (Dixon et al. 2012). One study used a chromatin conformation assay (Hi-C) to annotate the target genes of type

2 diabetes associated putative enhancers in pancreatic islets (Greenwald et al. 2019). This evidences that chromatin conformation experiments can be used to validate and characterize disease associated regions.

Understanding the effect of genetic changes at a transcriptomics level is also important in characterizing putative enhancers. RNA-Seq can be used to determine which genes are differentially expressed in cell lines containing enhancer deletions (Wang et al. 2009). Knowing what genes are impacted by the presence or absence of putative regulatory regions can provide clues as their function and biological role. Combining all of these techniques in a model with putative enhancers removed can provide information about those regions as potential enhancers, modulators of expression, and their contribution to disease.

4.1.4 Aims

The overall aim of this chapter is to characterize two putative enhancers within the AS associated *TNFRS1A* locus by utilizing functional genomic techniques on the CRISPR/Cas9 edited iPSC derived macrophages generated in Chapter 3. The work outlined in this chapter includes the assessment of the chromatin landscape and interactions, and gene expression changes associated with the removal of the two putative enhancers. The specific aims for this chapter are:

1. To carry out gene expression profiling on the CRISPR/Cas9 edited iPSC derived macrophages using RNA-Seq to determine the effect of putative enhancers on gene expression.

2. To map physical chromatin interactions occurring near *TNFRSF1A* by using Capture-C in iPSC derived macrophages and primary $CD14^{+}$ cells, and determine if removing putative enhancers has altered these interactions.

3. To investigate chromatin accessibility near *TNFRSF1A* by using ATAC-Seq in iPSC derived macrophages, and determine if removing putative

enhancers has impacted these regions.

4. To investigate enhancer associated histone marks near *TNFRSF1A* in iPSC derived macrophages using ChIP-Seq, and determine the impact of removing putative enhancers.

5. To determine the function of the putative Intronic and Intergenic Enhancers by integrating the generated functional genomic datasets.

6. To determine the success and future utility of iPSC derived macrophages for functional genomic studies in the context of AS and other diseases.

7. To validate the *TNFRSF1A* GWAS locus as candidate for further functional studies using integrated datasets.

4.2 Results

4.2.1 Samples

The samples used in this Chapter are the edited and unedited iPSC derived macrophages generated in Chapter 3. As mentioned in Chapter 3, the putative enhancer situated within the intron of *TNFRSF1A* is henceforth referred to as 'Intronic Enhancer' (Figure 4.2). The putative enhancer situated upstream of *TNFRSF1A*, and near *SCNN1A* is henceforth referred to as 'Intergenic Enhancer' (Figure 4.2). The iPSC derived macrophages harbouring a deletion of the Intronic Enhancer region are henceforth referred to as Intronic 'Enhancer Deletion clones' and the cell lines harbouring a deletion of the Intergenic Enhancer region are henceforth referred to as 'Intergenic Enhancer Deletion clones'. The samples used for each experiment are outlined in Table 4.1.

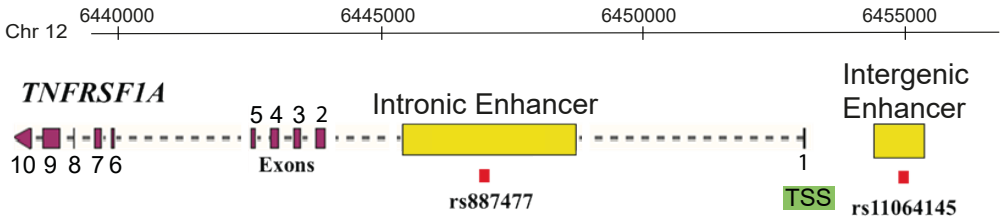


Figure 4.2: Schematic showing the *TNFRSF1A* gene with the Intronic Enhancer and Intergenic Enhancer.
Both enhancers are highlighted in yellow. The numbers denote exons and the transcription start site (TSS) is indicated. The location of the SNPs, rs887477 and rs11064145, are marked in red.

iPSC Derived Macrophage Clone	RNA-Seq			ATAC-Seq		ChIP-Seq	
	Unstim	LPS	IFNγ	Unstim	LPS	Unstim	LPS
Intronic Enhancer Deletion 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intronic Enhancer Deletion 2	Yes	Yes	No	No	No	Yes	Yes
Intronic Enhancer Deletion 3	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intergenic Enhancer Deletion 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intergenic Enhancer Deletion 2	Yes	No	No	Yes	Yes	Yes	Yes
Intronic Enhancer CRISPR Control 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intronic Enhancer CRISPR Control 2	Yes	Yes	Yes	No	No	Yes	Yes
Intergenic Enhancer CRISPR Control 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intergenic Enhancer CRISPR Control 2	Yes	Yes	Yes	No	No	Yes	Yes
Parental Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4.1: Edited and unedited clones used in each experiment.
Unstim = unstimulated. Deletion clones are edited. Parental control is unedited iPSC derived macrophage line from which edited clones were derived.

4.2.2 Expression Analysis of CRISPR/Cas9 Edited Cells

RNA-seq Quality Control

RNA from both edited and unedited iPSC derived macrophages was sequenced and mapped using STAR (Dobin et al. 2013). Read counts were generated using HTSeq-count (Anders et al. 2015) and differential analysis was performed using DESeq2 (Love et al. 2014), as outlined in Chapter 2. All but five of 27 samples had at least 20 million uniquely mapped reads (Figure B.1 in Appendix B).

A principal component analysis (PCA) was carried out using RNA-Seq data from primary immune cells and iPSC derived macrophages to validate their expression profile. $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and two iPSC derived macrophage cell lines were compared. One of the iPSC derived macrophage datasets was the one generated in this thesis, and the other was from an independently produced set of iPSC derived macrophages, generated from another group member using the same protocol. Figure 4.3 illustrates the first two principal components from the analysis, showing clear clusters between cell types. The two iPSC derived macrophage cell lines cluster together, but separately from the primary immune cell types. Though they are similar to $CD14^+$ cells, they are not the same cell type and do not completely share expression profiles, and therefore cluster independently of them (Buchrieser et al. 2017).

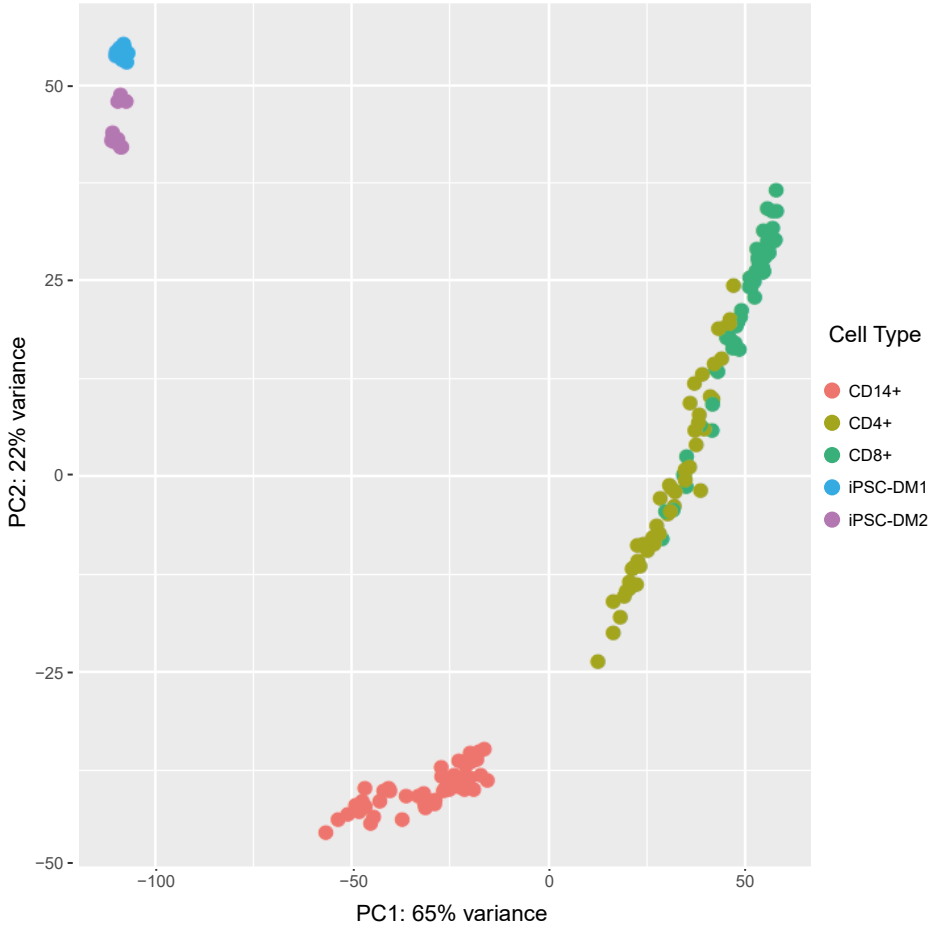


Figure 4.3: PCA of primary immune cells and iPSC derived macrophages.
The first two principal components of primary $CD4^+$ T cells, $CD8^+$ T cells, $CD14^+$ monocytes, and two iPSC derived macrophage cell lines are plotted based on expression profiles from RNA-Seq. iPSC-DM1 refers to the iPSC derived macrophages produced and experimented on in this thesis. iPSC-DM2 refers to another set of iPSC derived macrophages, produced independently by another group member. Each dot represents one sample and are coloured based on sample type.

Gene Expression Analysis in Edited iPSC Derived Macrophages

The expression level of *TNFRSF1A* and other genes was measured between edited cells and controls using RNA-Seq. Edited clones were compared against CRISPR Control clones and the Parental cell line for differential expression analysis (Table 4.1). The analysis was restricted to the TAD containing *TNFRSF1A*, as removing this putative enhancer would likely have the largest effect on nearby genes found within the same TAD, and other local genes that might be regulated by the putative enhancers. The TAD selected (*TNFRSF1A* TAD) is around 200kb in size

and contains nine protein coding genes, and two non-coding RNA genes (Figure 4.4) (Schmitt et al. 2016; Rao et al. 2014). It was defined using Hi-C in the B cell-like cell line, GM12878 (Rao et al. 2014). Although the TAD used was not defined in macrophages or macrophage-like cells, TAD boundaries are found to be relatively invariable across cell types (Gong et al. 2018).

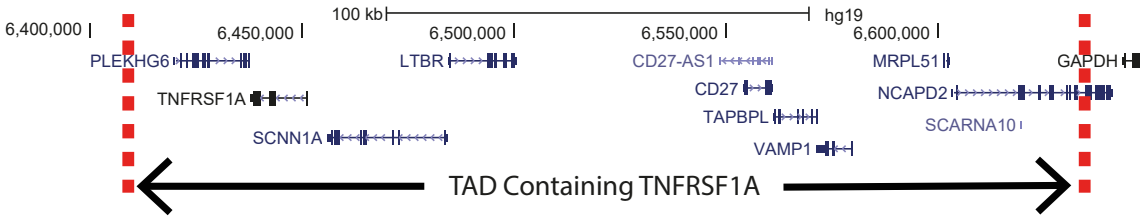


Figure 4.4: Illustration of the TAD containing *TNFRSF1A*.
The TAD selected (*TNFRSF1A* TAD) on chromosome 12 is around 200kb in size and contains 11 genes. It was defined using Hi-C in GM12878 cells (Rao et al. 2014).

The Intronic Enhancer Regulates *TNFRSF1A*

The analysis comparing the Intronic Enhancer Deletion edited clones to controls showed that in unstimulated cells, *TNFRSF1A*, was differentially expressed with a Log_2 fold change of -0.27 and an FDR of 0.0034 (Figure 4.5A). Deleting the Intronic Enhancer resulted in a down regulation of *TNFRSF1A*, consistent with this putative enhancer acting as a positive regulator of *TNFRSF1A*, as removing it decreases the mRNA levels for this gene. No other genes were differentially expressed.

Differential expression analysis between the Intronic Enhancer Deletion edited clones, and CRISPR Controls and Parental cell line was also performed genome-wide. Three genes were differentially expressed between Intronic Enhancer Deletion clones and the Control cell lines (Table 4.2). However, due to the low sample size, the analysis has very limited power, so such results would need further validation.

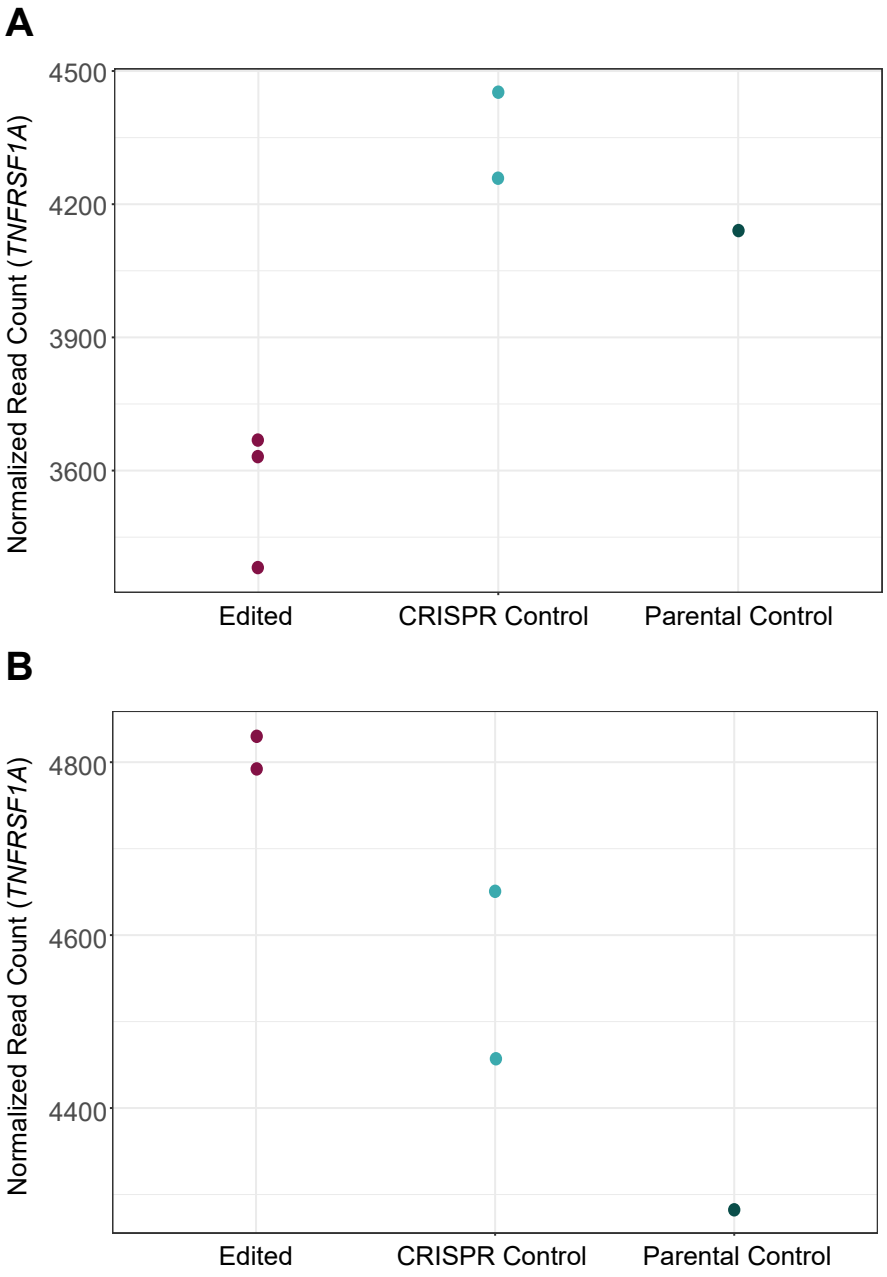


Figure 4.5: *TNFRSF1A* expression in iPSC derived macrophages.
DESeq2 differential analysis restricted to the genes within the TAD containing the deletion. **A**. Intronic Enhancer Deletion analysis. *TNFRSF1A* expression in edited and unedited iPSC derived macrophages following a differential analysis between three Intronic Enhancer Deletion edited clones, two CRISPR Controls, and one Parental Control. *TNFRSF1A* was down-regulated in Intronic Enhancer Deletion edited cells compared to the Control cell lines with a Log_2 fold change of -0.27 and an FDR of 0.0034. **B**. Intergenic Enhancer Deletion analysis. *TNFRSF1A* expression in edited and unedited iPSC derived macrophages following a differential analysis between two Intergenic Enhancer Deletion edited clones, two CRISPR Controls, and one Parental Control. *TNFRSF1A* was not significantly differential with a Log_2 fold change of 0.073 and an FDR of 0.94.

Gene Name	Log_2 fold Change	FDR
TRIM4	0.389	$1.60x10^{-4}$
HELZ2	-0.350	$3.45x10^{-3}$
CCL8	-0.275	$3.20x10^{-2}$

Table 4.2: Differentially expressed genes between Intronic Enhancer Deletion clones and controls (unstimulated).
Summary of differentially expressed genes following a genome-wide analysis following a differential analysis between three Intronic Enhancer Deletion edited cell lines, two CRISPR Control cell lines, and one Parental Control cell line in an unstimulated state. This table lists the differentially expressed genes following a genome-wide gene comparison and was not restricted to TADs. Only those genes with an FDR of <0.05 (DESeq2) are shown.

The Intergenic Enhancer Regulates *NCAPD2*

The differential analysis between Intergenic Enhancer Deletion edited clones and controls showed that in unstimulated cells, Non-SMC Condensin I Complex Subunit D2 (*NCAPD2*), was differentially expressed, with a Log_2 fold change of 0.4 and an FDR of 0.012 (Figure 4.6). Deleting the Intergenic Enhancer did not appear to have any impact on the expression of *TNFRSF1A*, as hypothesized (Figure 4.5B). Instead, it resulted in an up-regulation of *NCAPD2*, which is consistent with this putative enhancer being a silencer or repressor for *NCAPD2*. *NCAPD2* is roughly 150Kb away from the Intergenic Enhancer, suggesting that could be acting at a distance though co-factors or chromatin looping.

Differential expression analysis between the Intergenic Enhancer Deletion edited clones, and CRISPR Controls and Parental cell line was also performed genome-wide. A total of 75 genes were differentially expressed between Intergenic Enhancer Deletion clones and the controls. The top 10 most differential genes following the analysis are listed Table 4.3 and all the significantly differentially genes are listed in Table B.2 in Appendix B.

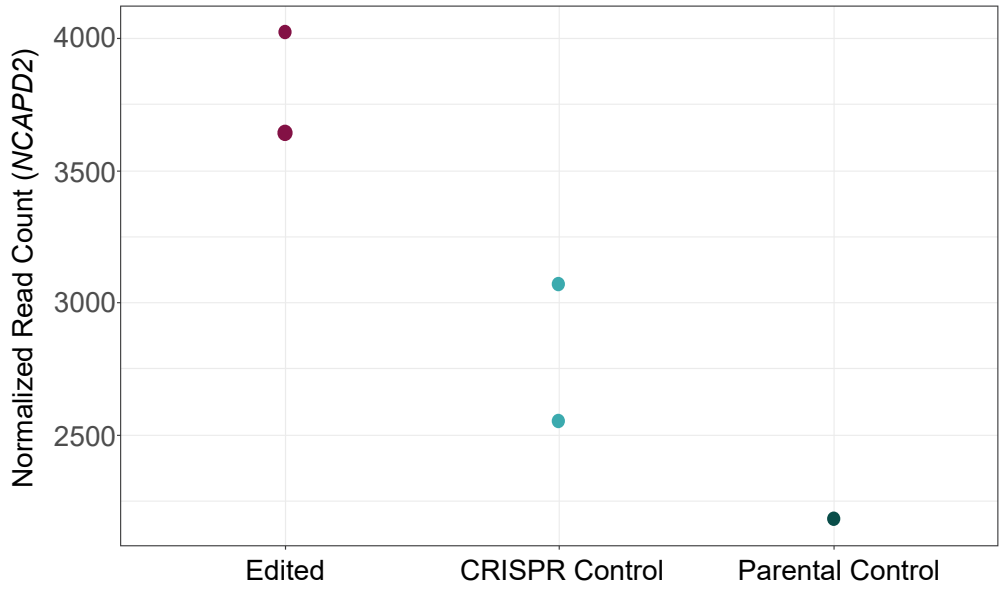


Figure 4.6: *NCAPD2* expression in iPSC derived macrophages.
NCAPD2 expression in edited and unedited iPSC derived macrophages following a differential analysis between two Intergenic Enhancer Deletion edited clones, two CRISPR Controls, and one Parental Control. *NCAPD2* was up-regulated in edited cells compared to the controls with a Log_2 fold change of 0.4 and an FDR of 0.012. DESeq2 differential analysis restricted to the genes within the TAD containing the deletion.

LPS and IFNγ Stimulations

The iPSC derived macrophages were also stimulated for 48 hours using LPS, and IFNγ. Both LPS and IFNγ interact with immune cell types to trigger an immune response and are commonly used to stimulate $CD14^+$ monocytes and macrophages. LPS binds to CD14 on the cell surface and is then recognized by TLR4 to induce a signaling cascade that results in NF-κB activation and initiation of an inflammatory response (Wright et al. 1990; Silva Correia et al. 2001; Zanoni et al. 2011). LPS is also attributed as being one of the main driving factors in pushing macrophages towards an M1, or pro-inflammatory, phenotype (Martinez and Gordon 2014). IFNγ is produced by natural killer cells and T cell subsets, including $CD4^+$ and $CD8^+$ as a response to immune system activation (Schoenborn and Wilson 2007). It is one of the main cytokines associated with the M1 macrophage phenotype as it induces signaling cascades in macrophages resulting in the up-regulation of pro-inflammatory genes including

Gene Name	Log_2 fold Change	FDR
LY6E	-1.41	$1.33x10^{-31}$
LPL	-0.688	$7.76x10^{-8}$
CD93	0.709	$7.76x10^{-8}$
ADAMDEC1	-0.754	$7.26x10^{-7}$
ZFP41	-0.764	$9.06x10^{-6}$
EGR2	-0.759	$1.47x10^{-5}$
GALNT12	-0.765	$1.70x10^{-5}$
RP11-455F5.3	0.581	$4.56x10^{-5}$
ITGA3	-0.743	$4.56x10^{-5}$
CD9	-0.607	$6.13x10^{-5}$

Table 4.3: Top 10 of the differentially expressed genes between Intergenic Enhancer Deletion clones and controls (unstimulated).
Differentially expressed genes following the analysis between two Intergenic Enhancer Deletion edited clones, two CRISPR Controls, and Parental Control in an unstimulated state. This table lists the differentially expressed genes following a genome-wide comparison and was not restricted to TADs. Only the 10 most significantly differentially expressed genes are shown. FDR from DESeq2.

other interferons, and genes encoding MHC class II molecules (Martinez and Gordon 2014; Tau and Rothman 1999).

The LPS and IFNγ stimulations on iPSC derived macrophages were performed to investigate any context specific properties of the enhancers, as gene expression profiles can change when macrophages encounter immunostimulants. As macrophages play an important role in the immune system and inflammation, it was important to investigate changes in gene expression associated with pro-inflammatory signals.

A PCA was performed on RNA-Seq samples to examine the variation between stimulated and unstimulated iPSC derived macrophages (Figure 4.7). Due to limited cell numbers, not every cell line received both treatments. Table 4.1 lists the stimulations received by each sample. Ten unstimulated samples, nine LPS stimulated samples, and eight IFNγ stimulated samples were sequenced and

compared to examine the effect of the stimulation. LPS had the largest effect and showed the most variance. IFNγ showed a much smaller effect as indicated by clustering together with the unstimulated samples.

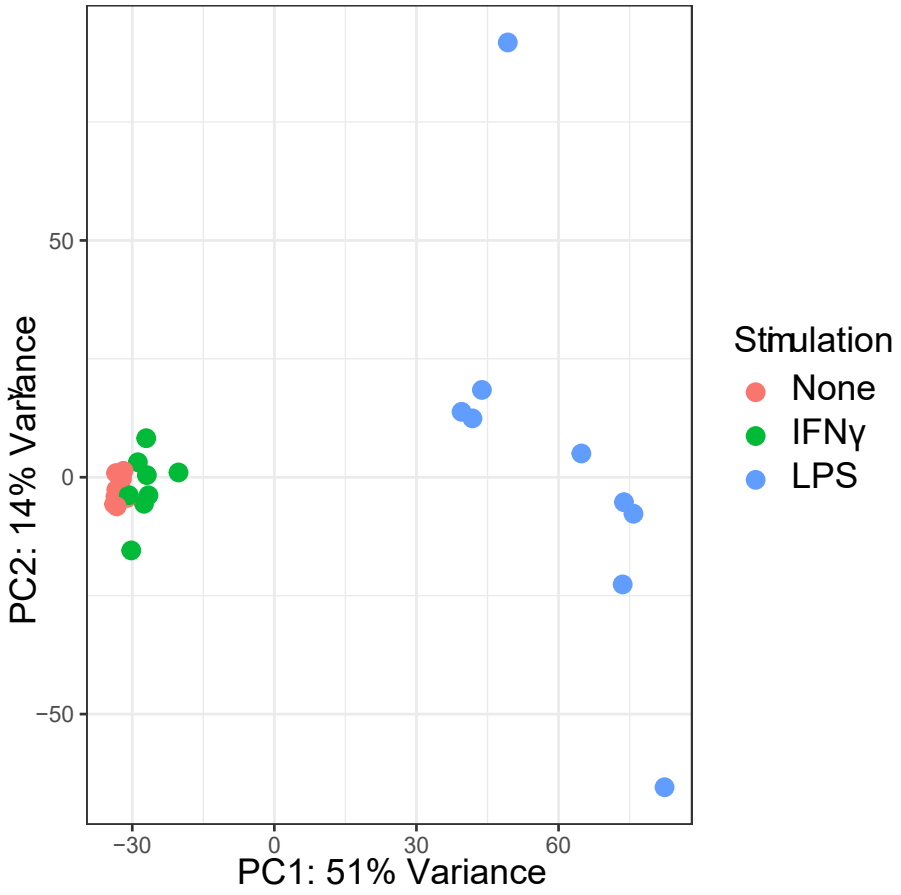


Figure 4.7: PCA of stimulated iPSC derived macrophages.
The first two principal components of the stimulated iPSC derived macrophage gene expression data are plotted, showing variation between stimulated states. Ten unstimulated samples, nine LPS stimulated samples, and eight IFNγ stimulated samples were sequenced and compared. Points are coloured based on whether they were unstimulated, LPS stimulation or IFNγ stimulated. The LPS treatment exhibited the most variance.

Validating LPS Induced iPSC Derived Macrophages as Model

The iPSC derived macrophage expression data produced here were compared against a dataset produced by Fairfax et al. for evidence of correlation (Fairfax et al. 2014). The Fairfax dataset consists of primary human $CD14^{+}$ cells stimulated for 24 hours with LPS. The iPSC derived macrophage dataset

presented here consists of a much smaller sample size and cells were stimulated for 48 hours with LPS. The expression datasets are reasonably well correlated, despite the different cell type and stimulation time, with a Pearson's correlation coefficient of 0.478 and share a large proportion of differentially expressed genes.

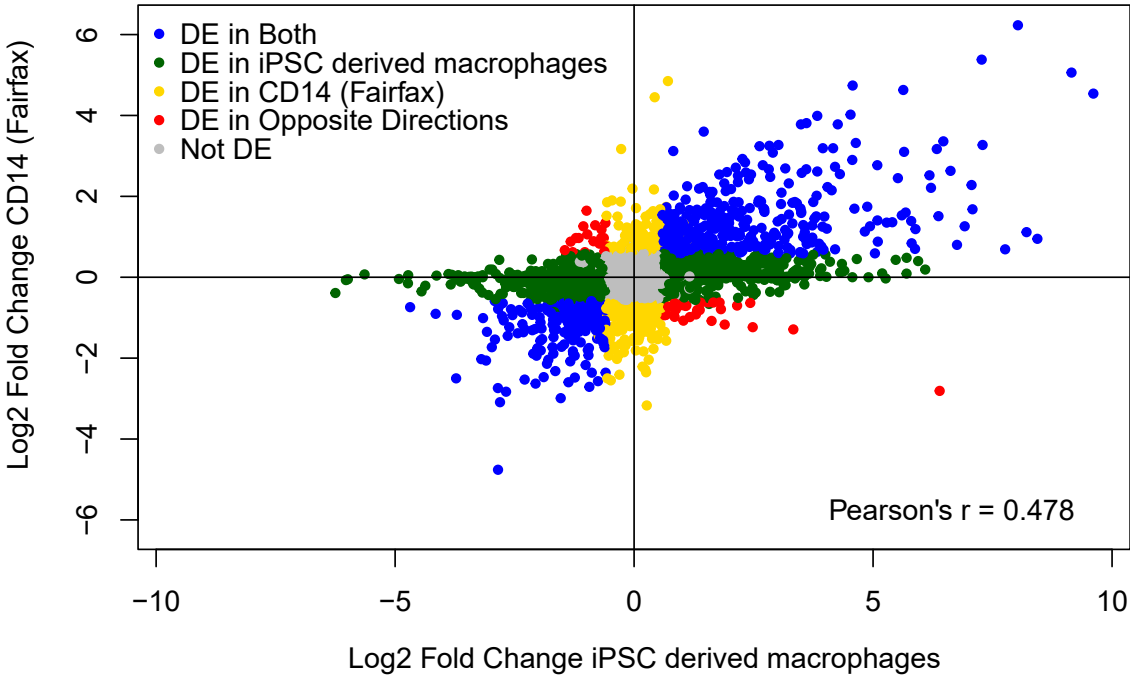


Figure 4.8: Correlation between LPS stimulated iPSC derived macrophages and CD14$^+$ cells.
The differentially expressed genes from 48 hour LPS stimulated iPSC derived macrophages were compared with the differentially expressed genes from 24 hour CD14$^+$ cells from Fairfax et al. (Fairfax et al. 2014). They were plotted based on their Log_2 fold change. DE = differentially expressed.

Because LPS had the largest impact on gene expression, differentially expressed genes between unstimulated and LPS stimulated states were examined. A volcano plot indicating the up- and down-regulated differentially expressed genes with an FDR <0.05 and the Log_2 fold change values are shown in Figure 4.9. The most significantly differentially expressed genes with a -log10(FDR) of greater than 100 are listed on the plot.

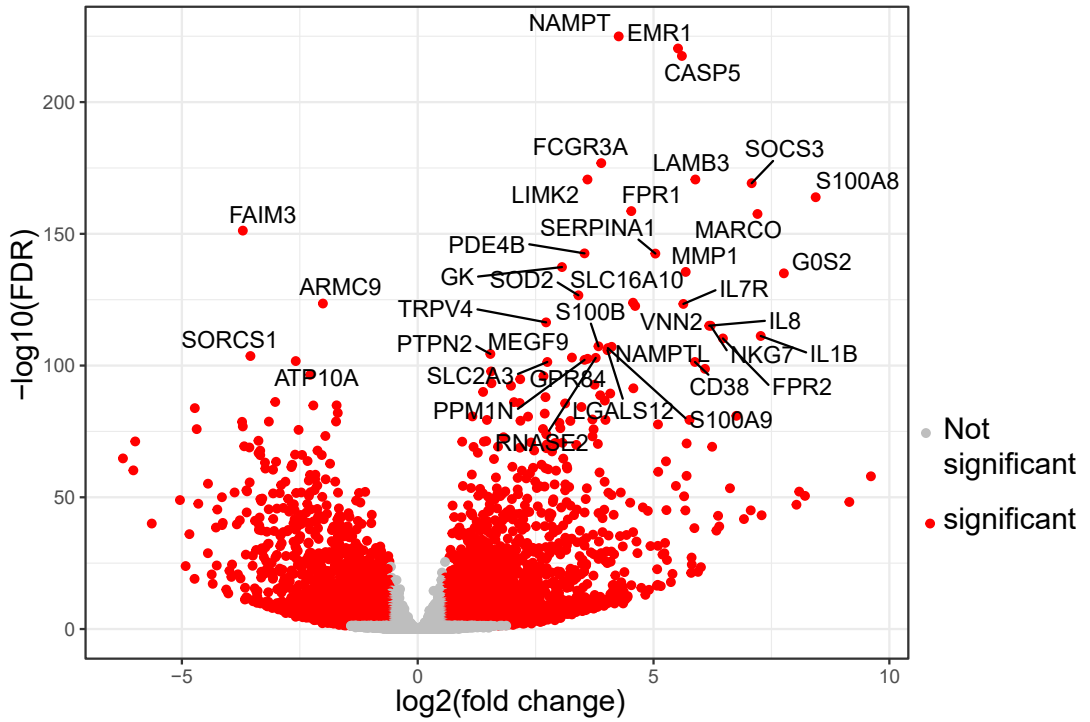


Figure 4.9: Volcano plot of differentially expressed genes between LPS treated and unstimulated iPSC derived macrophages.
Differential expression is shown by plotting the FDR (DESeq2) against the Log_2 fold change. Ten unstimulated and nine LPS treated samples were compared. Significant genes are shown in red with an FDR of <0.05. The most significantly differentially expressed genes are labeled.

Gene Expression Analysis in Stimulated Edited iPSC Derived Macrophages

A differential analysis was performed comparing unstimulated and stimulated cells, regardless of being edited or not. This analysis was restricted to the TAD containing *TNFRSF1A*. There were no differentially expressed genes between unstimulated and LPS treated cells, or IFNγ treated cells.

Another TAD specific differential analysis was performed, comparing edited and unedited cells in stimulated states. Intronic Enhancer Deletion edited clones were compared against the CRISPR Controls and the Parental control. No genes were differentially expressed between edited cells and controls when stimulated with LPS, or IFNγ. This analysis was repeated for the Intergenic Enhancer Deletion edited clones and also yielded no differentially expressed genes.

Differential expression analysis on stimulated cells, comparing between edited and unedited, was also performed genome-wide. Two genes were differentially expressed between Intronic Enhancer Deletion clones and controls, when stimulated with IFNγ, but none were differentially expressed following LPS stimulation (Table B.1 in Appendix B). When this genome-wide analysis was performed using the Intergenic Enhancer Deletion edited clones, three genes were differentially expressed between edited cells and controls when stimulated with LPS, and 59 genes were differentially expressed when stimulated with IFNγ (Tables B.3 and B.4 in Appendix B).

4.2.3 Interrogating Physical Chromatin Interactions

Enhancers can act on a promoter through the physical bending and looping of chromatin, and removing enhancers may impact how the chromatin interacts to modulate gene expression. Chromatin within a TAD will frequently form physical interactions, but will unlikely interact with chromatin in a different TAD (Phillips-Cremins et al. 2013; Nora et al. 2012). To better understand the interplay between the genes in this region, the chromatin landscape, and the putative enhancers, the TAD containing *TNFRSF1A* was examined for evidence of physical interactions using Capture-C.

Capture-C Quality Control

Two Intronic Enhancer Deletion clones, two Intergenic Enhancer Deletion clones, two CRISPR Controls, and primary $CD14^+$ cells were fixed and prepped for Capture-C. The primary $CD14^+$ cells were isolated from a healthy individual and were included in this analysis as a positive control, as they are the closest primary cell type to the iPSC derived macrophages. Unfortunately, due to the low cell numbers harvested from the iPSC derived macrophages for this analysis, most clones failed during processing and were not sequenced. Capture-C requires around 10 million cells, and only around 2 million cells were available from the

iPSC derived macrophage clones. One Intronic Enhancer Deletion clone, one Intergenic Enhancer Deletion clone, and the primary $CD14^+$ cells progressed to the sequencing stage of the experiment. However, the Intronic Enhancer Deletion clone and the Intergenic Enhancer Deletion clone both had substantially fewer cells than is optimal and not enough ligation events were sequenced to accurately visualize interactions. Baits for each promoter were used to create viewpoints to examine the chromosomal interactions across the whole TAD and their coordinates are listed in Table 2.7 in Chapter 2. Sequenced fragments are classified as either a capture fragment (contains the biotinylated oligo) or a reporter fragment, which represents the 'other end' of the interaction. Only interaction read pairs containing a single capture fragment and at least one reporter fragment are kept for analysis. A minimum of 10,000 reporters per viewpoint are required. The number of fragments obtained for the Intronic Enhancer Deletion clone and the Intergenic Enhancer Deletion clone ranged from around 200 to 2000 and are listed in Tables B.6 and B.7 in Appendix B. The primary $CD14^+$ cells had many more fragments per capture site than the required 10,000 (Table B.5 in Appendix B), so this library was used to interrogate physical chromatin interactions and the iPSC derived macrophage data were disregarded.

Three control regions were added to the baits to test the quality of the Capture-C data (Table 2.7 in Chapter 2). The control regions were an interaction between a CTCF motif in *LMO2* and a CTCF motif in an intergenic non-coding region, an interaction between the *RUNX1* promoter and the *RUNX1* super-enhancer, and an interaction between a CTCF motif near *TAL1* and another CTCF motif within the intron of *STIL*. These capture sites show known interactions that were previously mapped by ChIA-PET in Jurkat cells (Hnisz et al. 2016), and subsequently confirmed in-house using Capture-C in $CD14^+$ cells. The $CD14^+$ libraries used for this thesis work are consistent with other in-house $CD14^+$ cell libraries and clearly show the control interactions (Figure 4.10).

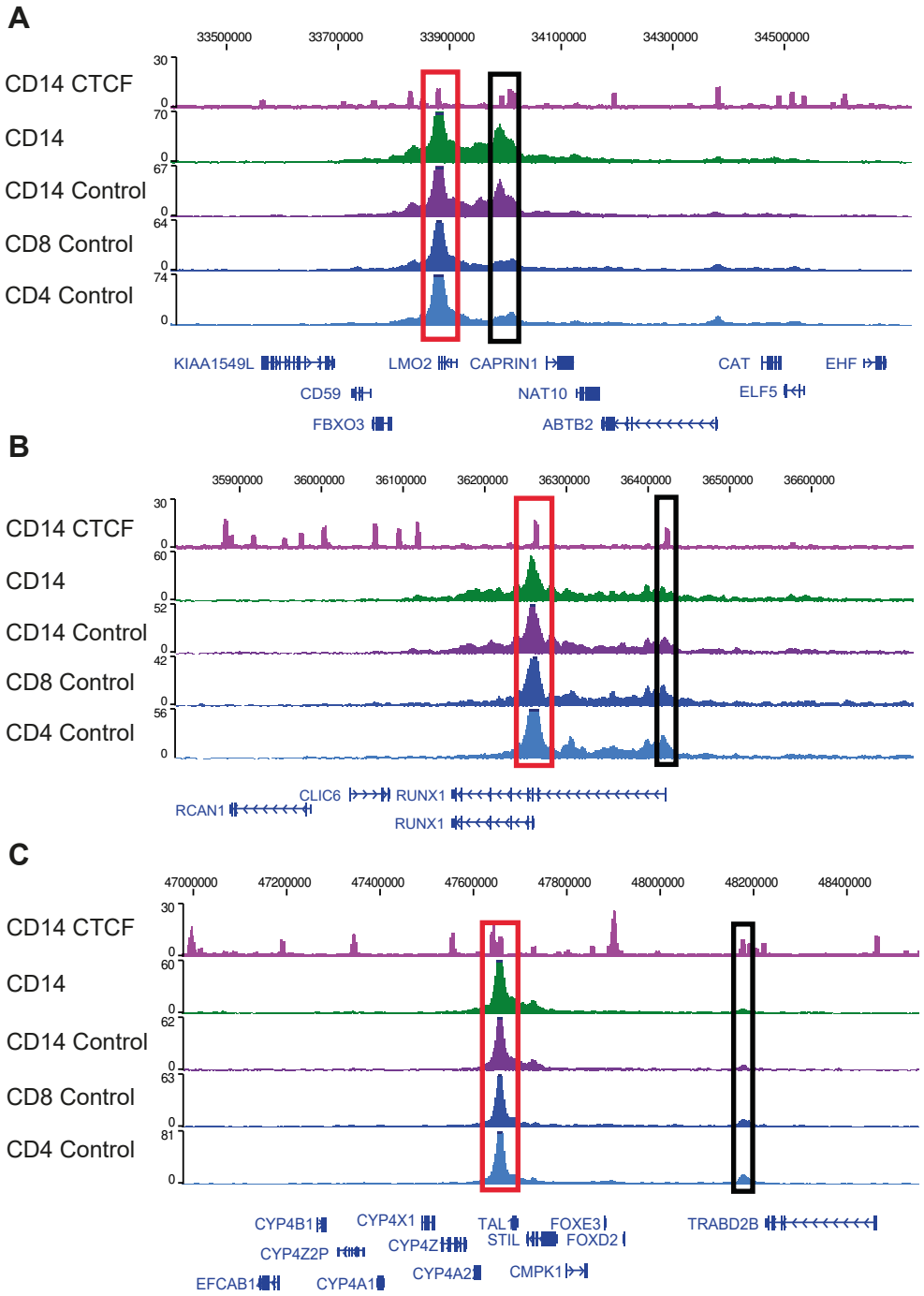


Figure 4.10: Capture-C data for control regions.
Capture sites showing known interactions in primary cells to validate data quality. The captured fragments were visualized using the WashU Epigenome Browser (Li et al. 2019). The $CD14^+$ cell Capture-C data (green track) was compared to previously collected in-house data (purple, dark blue and light blue for $CD14^+$, $CD8^+$, and $CD4^+$ libraries, respectively). CTCF binding regions in $CD14^+$ cells were obtained from ENCODE (top purple track). Viewpoints are outlined in red boxes and reporter fragments are outlined in black boxes. **A**. Capture data showing an interaction between a CTCF motif in *LMO2* and a CTCF motif in an intergenic non-coding region. **B**. Capture-C data showing a control interaction between the *RUNX1* promoter and the *RUNX1* super-enhancer. **C**. Capture-C data showing a control interaction between a CTCF motif near *TAL1* and another CTCF motif 500kb away.

Chromatin Interactions in CD14$^+$ Monocytes Within the *TNFRSF1A* TAD

The chromatin interactions occurring within the *TNFRSF1A* TAD were interrogated in CD14$^+$ cells by baiting all nine gene promoters within it (*PLEKHG6*, *TNFRSF1A*, *SCNN1A*, *LTBR*, *CD27*, *TAPBPL*, *VAMP1*, and *NCAPD2*). Additionally, baits near the two putative enhancers being examined were also used. One bait was near rs887477 for the Intronic Enhancer and the other was near rs11064145 for the Intergenic Enhancer. Peaky (Eijsbouts et al. 2019) was used to determine which interactions were statistically significant. The Peaky analysis called no interactions in the TAD as being significant using p-values of both 0.01 and 0.05. However, there is no well established method for calling significant peaks in Capture-C data. The signal to noise varies between viewpoints and only one replicate was available for analysis. Therefore, definitive significance cannot be determined from this analysis, so interactions were annotated by visually examining peaks.

All the interactions occurring within the TAD, from each viewpoint, can be seen in Figure 4.11 and are summarized in Table 4.4. The genes tend to show the most frequent chromatin interactions (high peaks) with their closest neighboring genes. Most genes show interactions with *TNFRSF1A* and *VAMP1*, with the *VAMP1* promoter viewpoint showing interactions with every gene in the TAD. The viewpoints near rs887477 and rs11064145 show the same interactions as the *TNFRSF1A* promoter viewpoint. They are both near the *TNFRSF1A* promoter, so this is expected. *CD27*, also known as *TNFRSF7* is primarily expressed on T cells and B cells (Mack et al. 2009), but was still included in this analysis and shows interactions with its neighboring genes. Though it is not expressed in this cell type, its surrounding chromatin may help regulate other genes. The *NCAPD2* gene stretches across the TAD boundary and the RNA-Seq data presented above suggest there may be an interaction between *TNFRSF1A* and *NCAPD2*. The *NCAPD2* promoter viewpoint shows interactions with its closest

genes, *VAMP1* and *TAPBPL*, but it also shows an interaction with the more distant, *TNFRSF1A*. This interaction is not visible from the reciprocal *TNFRSF1A* promoter viewpoint.

Viewpoint	**Frequent Interactions**	**Infrequent Interactions**
PLEKHG6 promoter	*PLEKHG6*, *TNFRSF1A*	None
TNFRSF1A promoter	*SCNN1A*, *LTBR*, *TNFRSF1A*, *VAMP1*	*PLEKHG6*
rs887477	*SCNN1A*, *LTBR*, *TNFRSF1A*, *VAMP1*	*PLEKHG6*
rs11064145	*SCNN1A*, *LTBR*, *TNFRSF1A*, *VAMP1*	*PLEKHG6*
SCNN1A promoter	*TNFRSF1A*, *LTBR*	*PLEKHG6*, *VAMP1*
LTBR promoter	*TNFRSF1A*, *SCNN1A*	*VAMP1*
CD27 promoter	*TAPBPL*, *VAMP1*	*TNFRSF1A*
TAPBPL promoter	*CD27*, *VAMP1*	*TNFRSF1A*, *PLEKHG6*, *NCAPD2*
VAMP1 promoter	*TAPBPL*, *CD27*, *NCAPD2*	*PLEKHG6*, *TNFRSF1A*, *SCNN1A*, *LTBR*
NCAPD2 promoter	*VAMP1*, *TAPBPL*	*TNFRSF1A*

Table 4.4: Capture-C interactions in primary CD14$^+$ cells in the *TNFRSF1A* TAD. Frequent interactions are determined by large peaks and infrequent interactions are determined by small peaks.

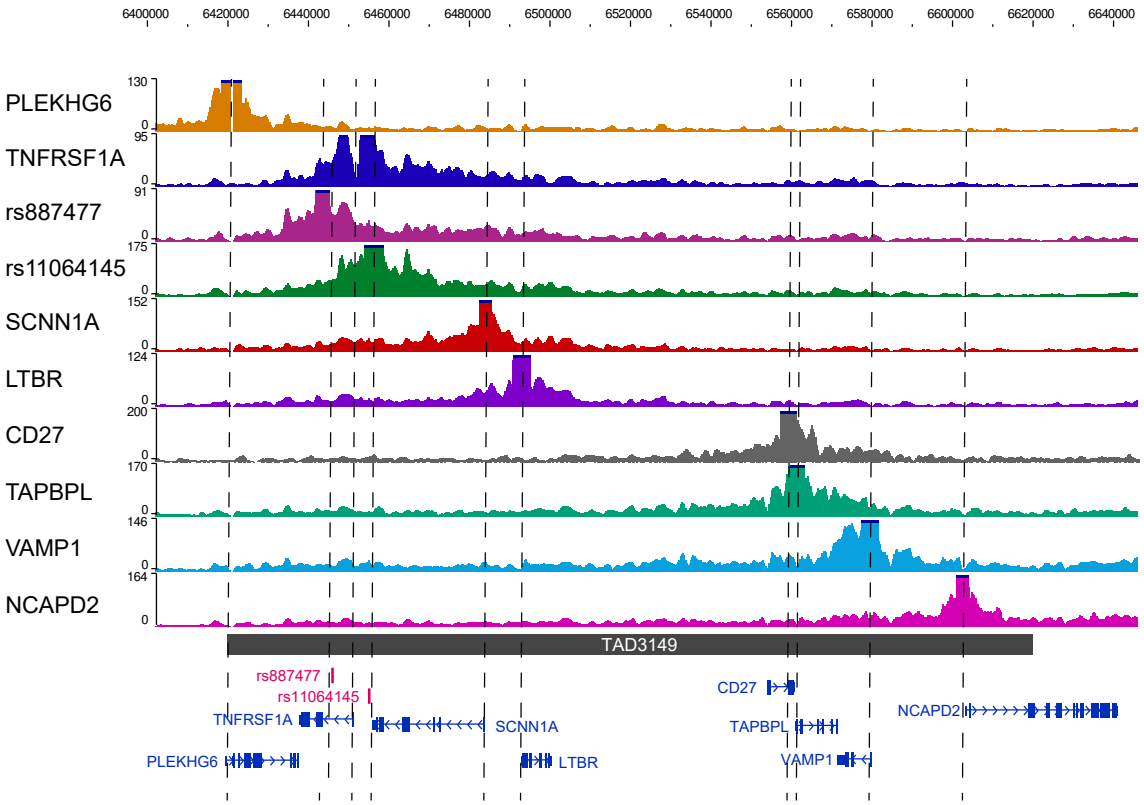


Figure 4.11: CD14⁺ monocyte Capture-C data for the TAD containing *TNFRSF1A* from different viewpoints.
All the captured fragments were visualized using the WashU Epigenome Browser (Li et al. 2019) and the scales were corrected based on the total number of fragments captured for each sample. The gene name listed beside each track denotes the promoter viewpoint for that gene. The two tracks labeled with the SNPs, rs887477 and rs11064145, denote viewpoints around those SNPs and the putative enhancer they are situated within. Peaks represent interactions between the capture viewpoint and another location. The dotted lines represent the promoters and the two SNPs.

4.2.4 Assaying Regions of Chromatin Accessibility

To investigate regions of open chromatin within iPSC derived macrophages and the impact of removing putative enhancers on chromatin accessibility, ATAC-Seq was performed. Chromatin accessibility can be context specific so both unstimulated and LPS treated cells were examined. LPS treated cells were stimulated for 48 hours to mimic an immune response. The analysis was restricted to the TAD containing *TNFRSF1A* as this showed more promising results in the RNA-Seq analysis, and changes in chromatin accessibility due to the removal of regulatory elements are more likely to be seen locally.

Omni-ATAC-Seq Quality Control

Omni-ATAC-Seq was the selected protocol for this work (Corces et al. 2017) (discussed in Chapter 5). Two Intronic Enhancer Deletion edited clones, two Intergenic Enhancer Deletion edited clones, two CRISPR Controls, and the Parental cell line were used (Table 4.1). Both unstimulated and LPS stimulated samples for each line were sequenced and run through the in-house pipeline (described in Chapter 2). ENCODE recommends 50 million reads (25 million fragments) per sample for ATAC-Seq experiments. Half of the samples reached this threshold (Figure 4.12). Reads are filtered out due to low quality, PCR duplicates, and mapping to the mitochondrial genome. All samples except one had at least 75% usable reads (Figure 4.13).

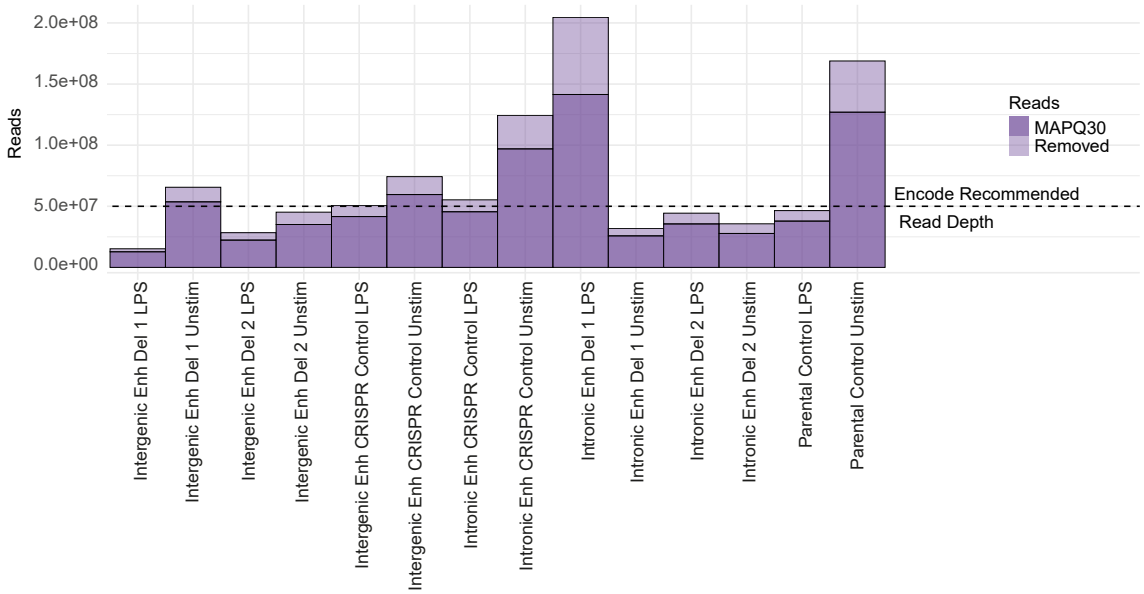


Figure 4.12: Total reads following OmniATAC-Seq on iPSC derived macrophages. Unstimulated and LPS stimulated samples were sequenced for samples containing an enhancer deletion (Enh Del) and control samples. MAPQ30 reads represent usable, high quality reads. ENCODE recommends 50 million aligned reads per sample.

Plotting enrichment of ATAC-Seq reads around the transcription start site (TSS) is a way to measure signal to noise, as TSSs are regions of known open chromatin. ENCODE recommends samples having a TSS fold enrichment of >5. All samples demonstrated a fold enrichment below five, suggesting a low signal to noise ratio,

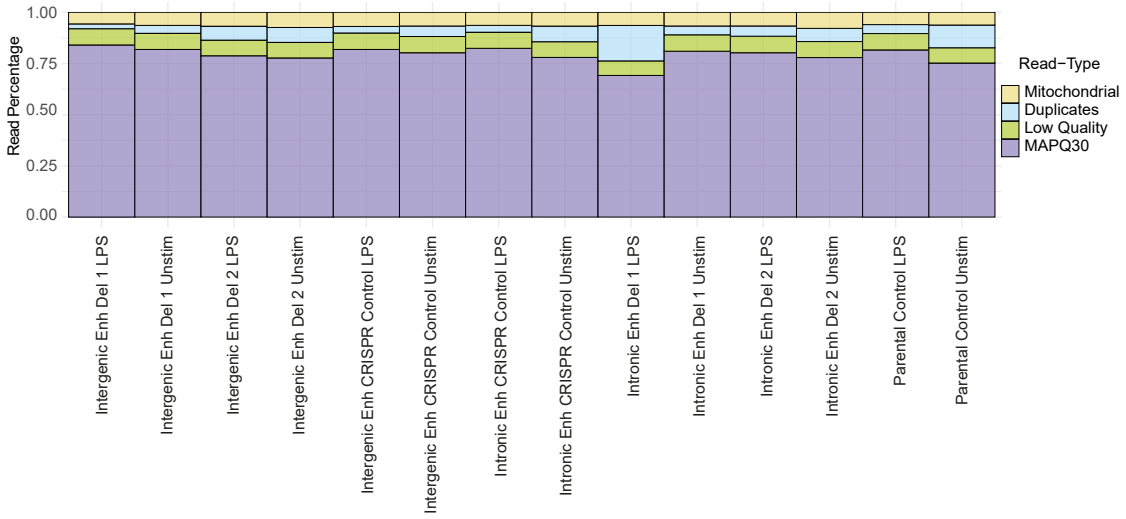


Figure 4.13: Read percentages following Omni-ATAC on iPSC derived macrophages. Samples containing an enhancer deletion (Enh Del) and control samples were sequenced in both an unstimulated and LPS stimulated state. MAPQ30 reads represent usable, high quality reads. Reads are lost due to low quality, PCR duplicates, and mapping to the mitochondrial genome.

which may impact peak calling and differential analysis (Figure B.2 in Appendix B). As described in Chapter 2, MACS2 was used for peak calling, based on the regions with the highest number of reads. A total of 12 peaks were called within the *TNFRSF1A* TAD. DESeq2 was used for differential analysis.

Chromatin Accessibility in Edited iPSC Derived Macrophages

Chromatin accessibility following the removal of the putative Enhancers was interrogated by differential analysis between edited clones and unedited controls in an unstimulated state. Due to the low sample quality, the CRISPR Controls and Parental Control were combined into one sample group for all ATAC analysis.

When the Intronic Enhancer was removed, there was one differential peak between the edited cells and controls (Table 4.5). The peak covering chr12:6444927-6446228 exhibits differential open chromatin between edited and unedited cells (FDR of $3.01 \text{x} 10^{-8}$). This region corresponds to the peak directly over the Intronic Enhancer that was deleted, which is expected. There were no other differential peaks. Figure 4.15 is a WashU Epigenome Browser view

of the significant differential peak over the Intronic Enhancer, clearly showing the area that was deleted. Figure 4.14 illustrates the peaks for each sample across the entire TAD. The peak over the *TNFRSF1A* promoter is visually much larger in the Intronic Enhancer Deletion clones, as is the peak over exons 2-5 (outlined in Figure 4.15). This could be a compensatory mechanism by the cell to open the promoter further due to the loss of the enhancer. However, the promoter peak was not significantly different in the differential analysis, nor was the peak situated over the central exons. With so few samples and low quality, the significance of these peak changes is not discernible.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6444927-6446228	*TNFRSF1A*	8.701	$3.01x10^{-8}$
chr12:6579176-6580574	*VAMP1*	-0.420	0.482
chr12:6421670-6422903	*PLEKHG6*	0.583	0.701
chr12:6432144-6433445	*PLEKHG6*	0.330	0.807
chr12:6525452-6526822	NA	0.226	0.885
chr12:6532164-6533793	NA	0.242	0.885
chr12:6602133-6603649	*NCAPD2*	-0.185	0.886
chr12:6492518-6493983	*SCNN1A*	-0.108	0.904
chr12:6539162-6540634	NA	0.154	0.904
chr12:6570317-6572085	*TAPBPL*	0.064	0.973
chr12:6502999-6504310	NA	0.0239	0.992
chr12:6560238-6561919	*TAPBPL*	0.002	0.992

Table 4.5: Regions of differential open chromatin between Intronic Enhancer Deletion clones and unedited controls (unstimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. Only one peak was significantly different (FDR of <0.05), which corresponds to the region that was deleted. FDR from DESeq2. The coordinates stated are in UCSC notation based on genome build hg19.

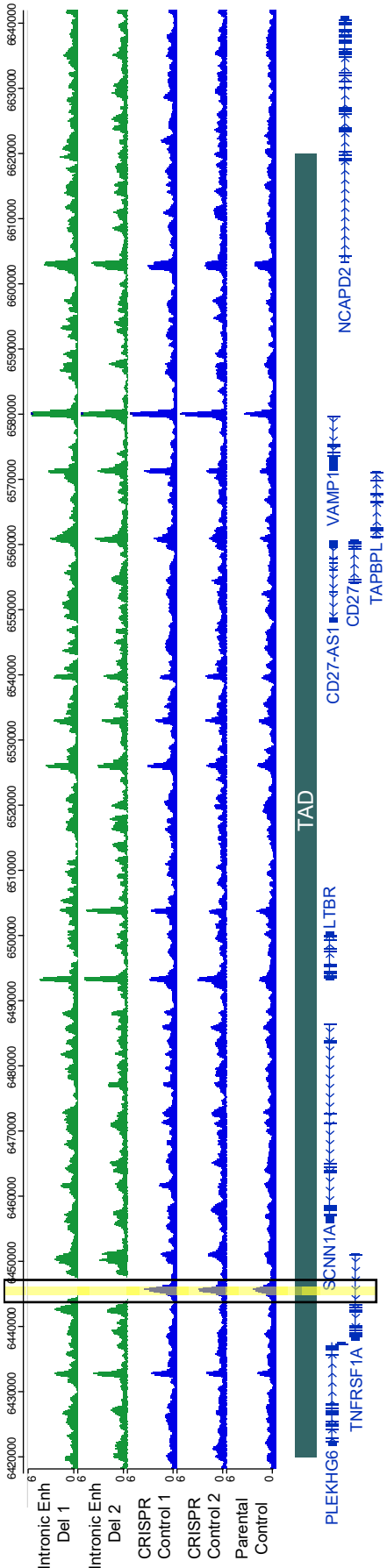


Figure 4.14: Differentially open chromatin peaks in Intronic Enhancer Deletion clones and unedited controls in an unstimulated state. WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 3.01×10^{-8}).

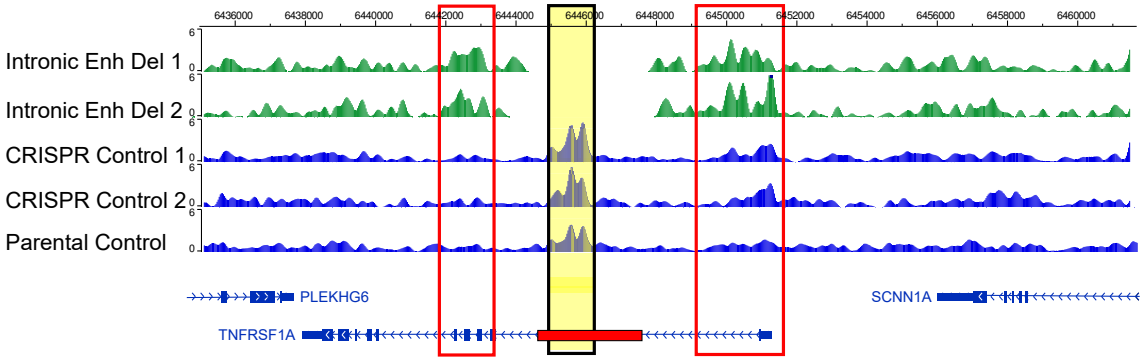


Figure 4.15: Close up of the differentially open chromatin peak over the Intronic Enhancer Deletion in edited cells and unedited controls in an unstimulated state. WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 3.01×10^{-8}). The deletion region is represented by the red box. The peak over the *TNFRSF1A* promoter and exons 2-5 are outlined in red. These peaks are not differential, but visually appears larger in edited cells.

There were no significant differential peaks between Intergenic Enhancer edited samples and controls (Table 4.6 in Appendix B). There was no peak called within the Intergenic Enhancer Deletion region, so deleting this region showed no effect on chromatin accessibility. However, the low TSS enrichment suggests poor signal to noise in these samples, and may have interfered with peak calling. Figure 4.16 illustrates the Intergenic Enhancer Deletion region and Figure B.3 in Appendix B shows the peaks for each sample across the entire TAD. The Intergenic Enhancer region is clearly deleted in the edited samples, but the controls do not show strong peaks in this region. Deleting the Intergenic Enhancer region did not appear to impact chromatin accessibility at any region within the TAD. However, with so few samples and low quality, there may have been more subtle changes that were missed in the analysis.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6421670-6422903	*PLEKHG6*	0.404	0.874
chr12:6432144-6433445	*PLEKHG6*	0.186	0.874
chr12:6502999-6504310	NA	0.155	0.874
chr12:6525452-6526822	NA	0.201	0.874
chr12:6532164-6533793	NA	0.141	0.874
chr12:6539162-6540634	NA	-0.239	0.874
chr12:6560238-6561919	*TAPBPL*	-0.172	0.874
chr12:6579176-6580574	*VAMP1*	-0.262	0.874
chr12:6602133-6603649	*NCAPD2*	-0.101	0.874
chr12:6492518-6493983	*SCNN1A*	-0.0689	0.918
chr12:6444927-6446228	*TNFRSF1A*	0.008	0.962
chr12:6570317-6572085	*TAPBPL*	0.028	0.962

Table 4.6: Regions of differential open chromatin between Intergenic Enhancer Deletion clones and unedited controls (unstimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significantly different (FDR of <0.05) between samples. The coordinates stated are in UCSC notation based on genome build hg19.

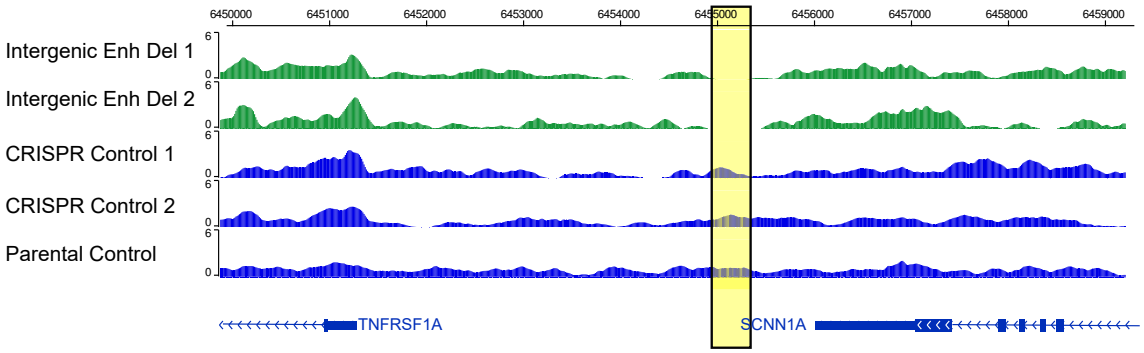


Figure 4.16: Close up of the Intergenic Enhancer Deletion region in edited cells and unedited controls in an unstimulated state, showing chromatin accessibility.
WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages. Two Intergenic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The Intergenic Enhancer is highlighted in yellow, but no open chromatin peak was called in this region. None of the peaks were differential.

Chromatin Accessibility in Stimulated Edited iPSC Derived Macrophages

Intronic Enhancer Deletion clones were compared to unedited controls after stimulation with LPS. Once again, only the peak over the Intronic Enhancer was significantly different with a Log_2 fold Change of 9.351 and an FDR of $1.96x10^{-9}$ (Table B.8 in Appendix B). The figures outlining the peaks produced from LPS stimulated cells appear visually similar to those from the unstimulated analysis and can be seen in Appendix B. Figure B.4 illustrates the peaks for each sample across the entire TAD after LPS stimulation. Figure B.5 is a close up image of this peak, clearly showing the area that was deleted. The peak over the *TNFRSF1A* promoter also appears much larger in the Intronic Enhancer Deletion clones after LPS stimulation.

This analysis was repeated for the Intergenic Enhancer Deletion clones, and none of the peaks were differential between edited cells and controls (Table 4.6). Again, there was no peak called spanning the Intergenic Enhancer Deletion region. The figures outlining peaks produced from LPS stimulated cells appear visually similar to those from the unstimulated analysis and can be seen in Figures B.7 and B.6 in Appendix B.

A differential analysis where samples were grouped as either unstimulated or LPS stimulated, not as edited or unedited, was also carried out to determine the effects of LPS stimulation. There were no differential peaks due to LPS stimulation, suggesting it had no effect on chromatin accessibility (Tables B.10 and B.11 in Appendix B). Figures 4.17 and 4.18 illustrate the peaks between the unstimulated and LPS stimulated samples.

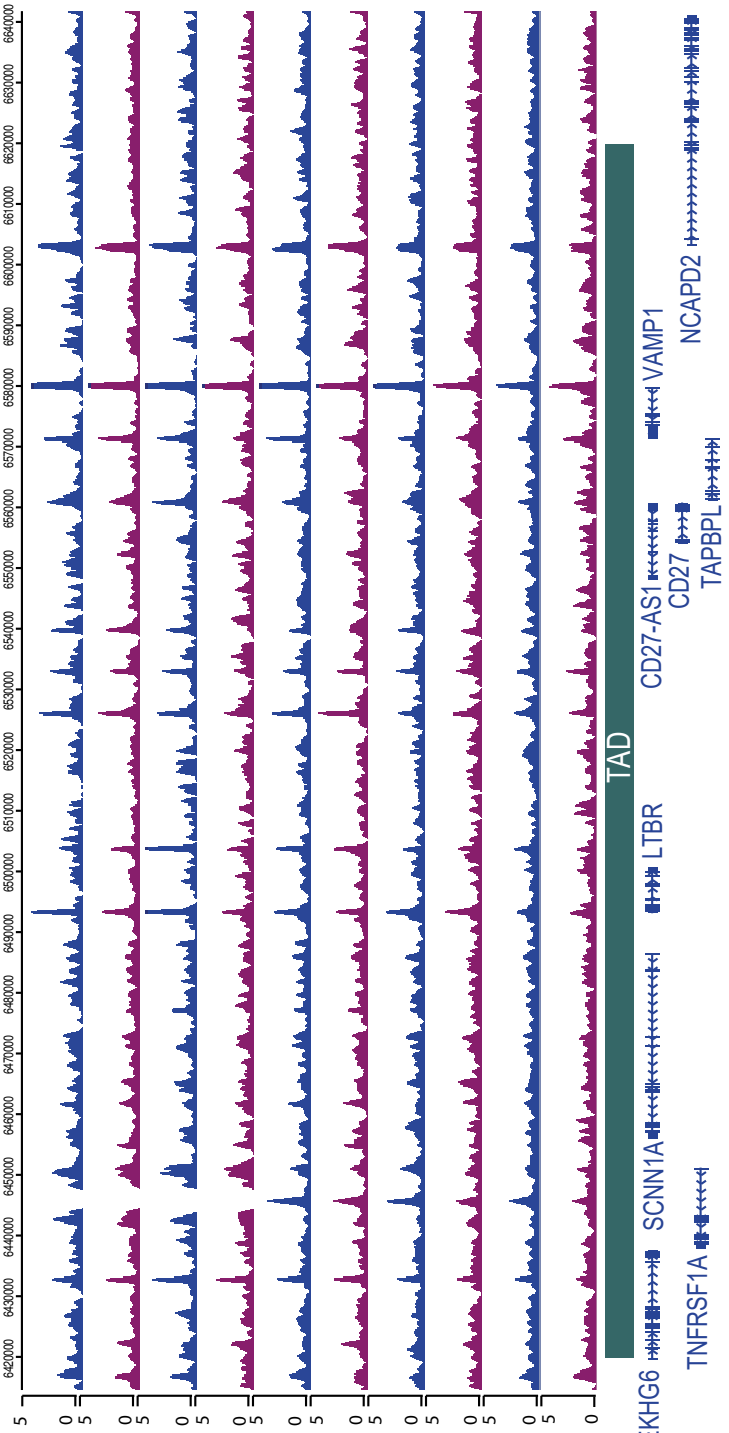


Figure 4.17: Open chromatin peaks in Intronic Enhancer Deletion clones and unedited controls in an unstimulated and LPS stimulated state.

WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The unstimulated cells are shown in blue and the LPS stimulated cells are shown in pink. None of the peaks were differential.

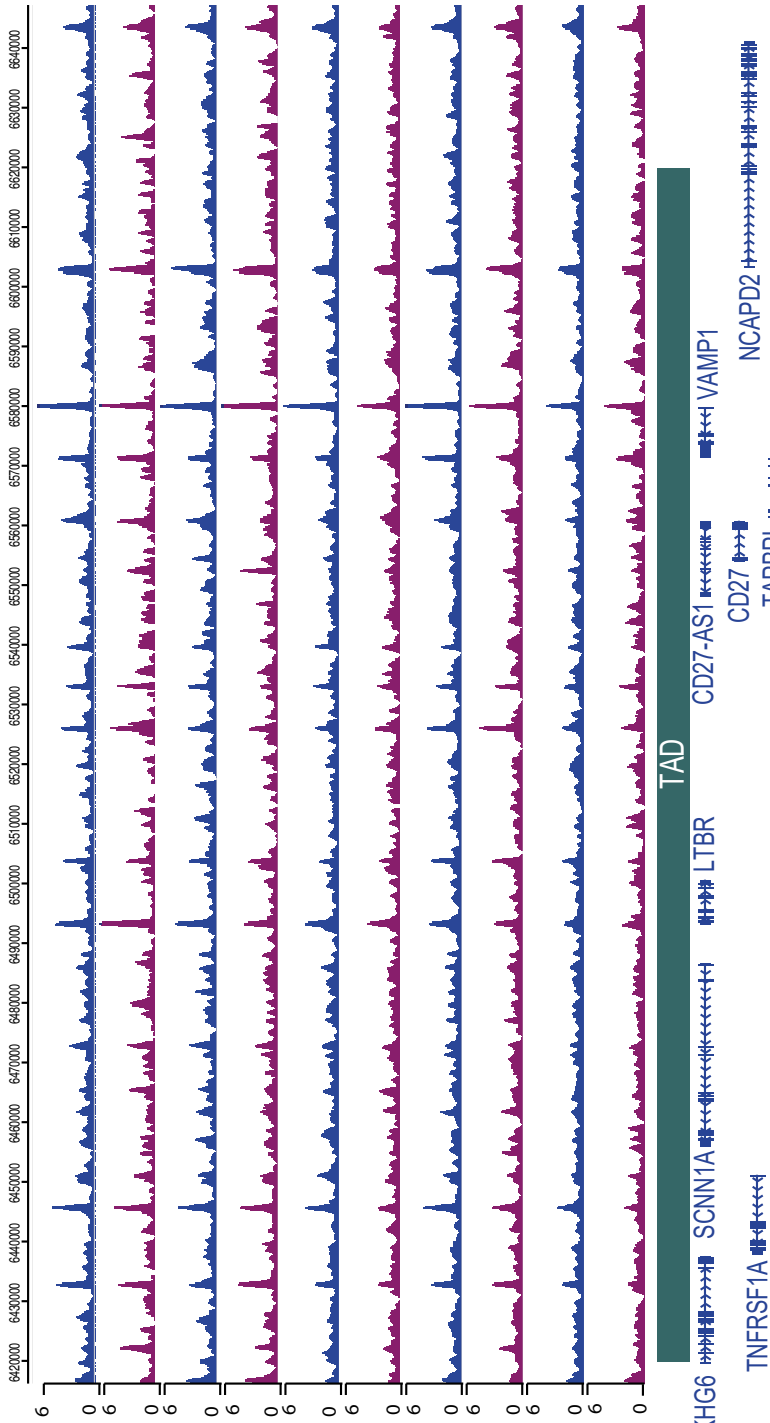


Figure 4.18: Open chromatin peaks in Intergenic Enhancer Deletion clones and unedited controls in an unstimulated and LPS stimulated state.

WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intergenic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The unstimulated cells are shown in blue and the LPS stimulated cells are shown in pink. None of the peaks were differential.

4.2.5 Assaying Histone Modifications

Active enhancers were assayed in the CRISPR/Cas9 edited iPSC derived macrophages using ChIP-Seq for H3K27ac, as it is a mark for active promoters and enhancers. This was carried out to validate that the selected putative enhancers have the H3K27ac histone modification in iPSC derived macrophages, and to determine if deleting these enhancers had any impact on the H3K27ac landscape. Histone modifications can be context specific so both unstimulated and LPS treated cells were examined. This analysis was also restricted to the TAD containing *TNFRSF1A*.

ChIPm Quality Control

ChIPmentation (ChIPm) (details in Chapter 5) was the selected protocol for this work (Schmidl et al. 2015). Three Intronic Enhancer Deletion clones, two Intergenic Enhancer Deletion clones, four CRISPR Controls, and one Parental control were used (Table 4.1). A negative control (Input) was also sequenced for each sample, which was prepared and sonicated at the same time as the ChIPm samples, but was not immuno-precipitated with the H3K27ac antibody. These controls were used to improve peak calling by normalizing peaks against background enrichment and to control for the shearing of chromatin during sonication. Both unstimulated and LPS stimulated samples were sequenced and run through the in-house pipeline (described in Chapter ch:MandM). ENCODE recommends 40 million reads (20 million fragments) per sample for ChIP-Seq experiments. Eight of 20 samples reached this threshold (Figure 4.19). Approximately half of the samples had at least 75% usable reads after filtering (Figure 4.20). Many of the samples suffered from having many PCR duplicates (25%-75%), which reduces the complexity of the sample data. The samples may have been over amplified or not enough of the original sample was sequenced. Plotting the enrichment of H3K27ac ChIPm reads around the TSS was used to

measure signal to noise, as H3K27ac is often found at gene promoters which coincide with TSSs. All samples had a TSS fold enrichment below six, suggesting a low signal to noise ratio, which may impact peak calling and differential analysis (Figures B.8 in Appendix B). MACS2 was used for peak calling, based on regions with the highest number of reads, and DESeq2 was used for differential analysis (Chapter 2). A total of 24 H3K27ac ChIP peaks were called within the *TNFRSF1A* TAD.

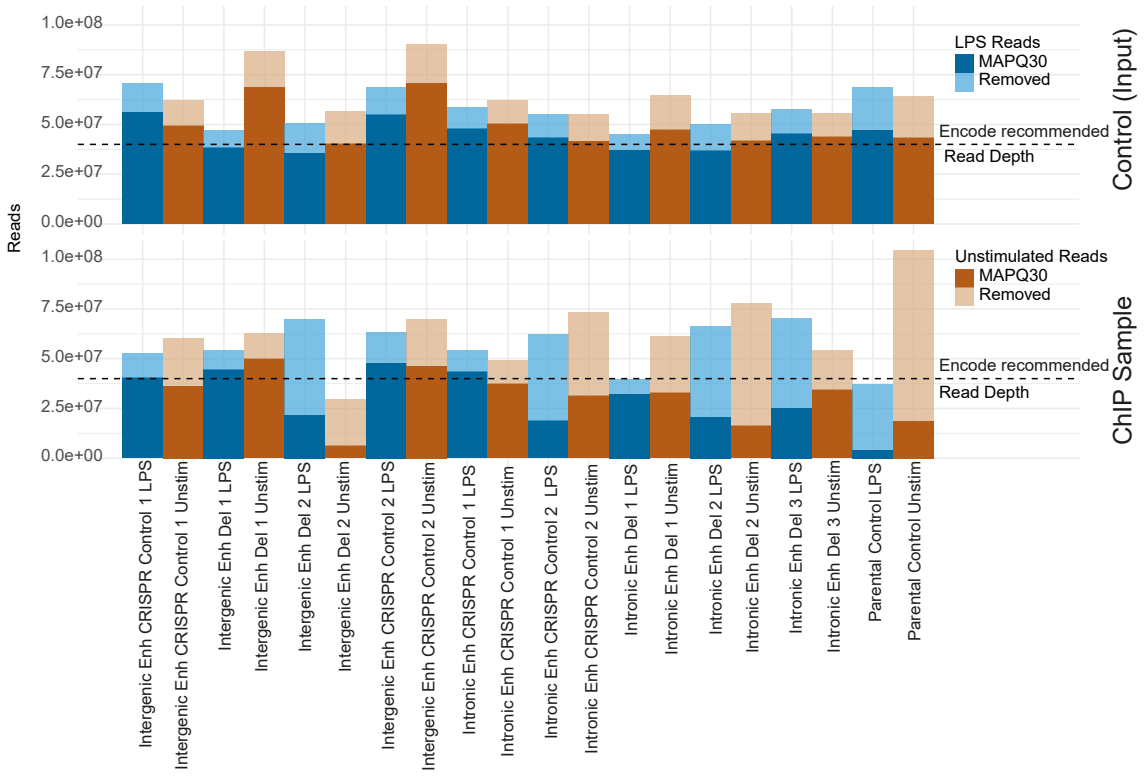


Figure 4.19: Total reads following H3K27ac ChIPm on iPSC derived macrophages. Samples containing an enhancer deletion (Enh Del) and control samples were sequenced in both an unstimulated and LPS stimulated state. MAPQ30 reads represent usable, high quality reads. ENCODE recommends 40 million reads per sample. Most samples reached this threshold, or were just under.

The H3K27ac Landscape in Edited iPSC Derived Macrophages

To determine if removing putative enhancers impacted H3K27ac, a differential analysis between edited clones and unedited controls was performed. Due to the low sample quality, the CRISPR Controls and Parental Control were combined

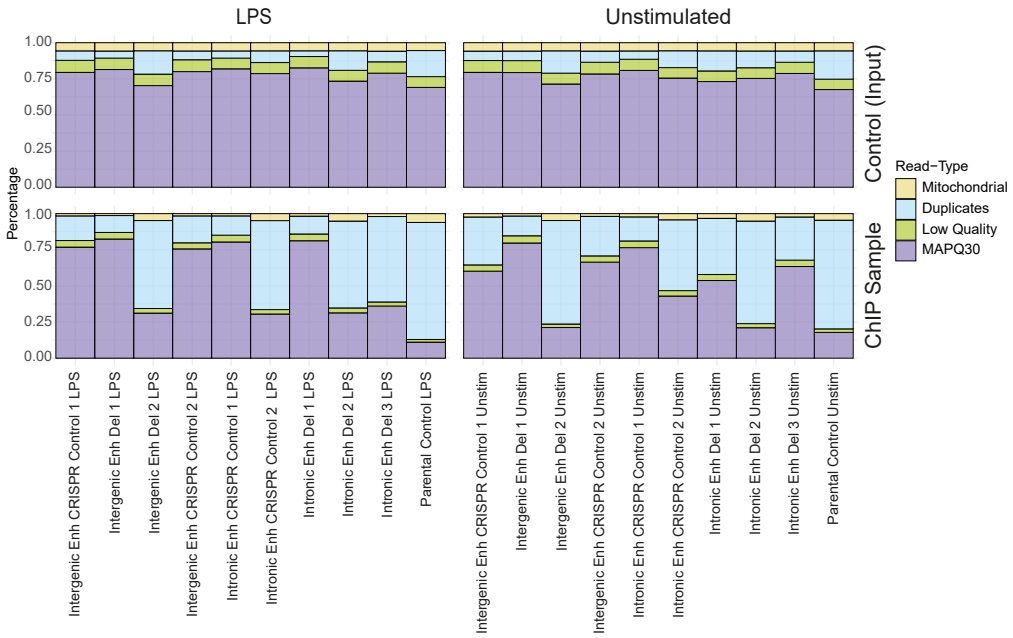


Figure 4.20: Read percentages following H3K27ac ChIPm on iPSC derived macrophages.
Samples containing an enhancer deletion (Enh Del) and control samples were sequenced in both an unstimulated and LPS stimulated state. MAPQ30 reads represent usable, high quality reads. Reads are lost due to low quality, PCR duplicates (identical reads), and mapping to the mitochondrial genome. Most samples had at least 50% MAPQ30 reads, with some having a high percentage of PCR duplicates.

into one sample group for the analysis. Three Intronic Enhancer Deletion clones, Two Intergenic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control underwent analysis (Table 4.1).

The peak situated at chr12:6443654-6447439 (Figure 4.22) was significantly reduced in Intronic Enhancer Deletion edited samples, following the differential analysis between edited and unedited cells (FDR of 7.31×10^{-25}). This region corresponds to the peak directly over the Intronic Enhancer that was deleted, which is expected. There were no other differential peaks (Table 4.7). Figure 4.21 illustrates the peaks for each sample across the entire TAD.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6443654-6447439	*TNFRSF1A*	3.638	$7.31x10^{-25}$
chr12:6492222-6494533	*LTBR*	-0.365	0.472
chr12:6532784-6533655	NA	-0.545	0.472
chr12:6476454-6478580	*SCNN1A*	0.351	0.497
chr12:6487632-6489030	*LTBR*	-0.451	0.497
chr12:6501613-6502113	*LTBR*	0.755	0.598
chr12:6603193-6604157	*NCAPD2*	0.518	0.598
chr12:6460247-6460887	*SCNN1A*	-0.367	0.682
chr12:6559861-6562469	*CD27/TAPBPL*	-0.217	0.682
chr12:6495594-6496638	*LTBR*	0.442	0.688
chr12:6432218-6433603	*PLEKHG6*	0.218	0.689
chr12:6538819-6540518	NA	-0.203	0.689
chr12:6441496-6442515	*TNFRSF1A*	0.254	0.716
chr12:6460934-6462955	*SCNN1A*	-0.187	0.716
chr12:6471971-6473762	*SCNN1A*	0.185	0.716
chr12:6578918-6580886	*VAMP1*	-0.150	0.716
chr12:6467539-6468401	*SCNN1A*	0.193	0.739
chr12:6503005-6504920	*LTBR*	0.124	0.739
chr12:6524686-6527917	NA	-0.103	0.769
chr12:6453927-6455189	*TNFRSF1A*	-0.124	0.839
chr12:6448993-6452438	*TNFRSF1A*	-0.038	0.862
chr12:6481586-6482482	*SCNN1A*	0.080	0.862
chr12:6568210-6573444	*TAPBPL/VAMP1*	-0.042	0.862
chr12:6601568-6603078	*NCAPD2*	-0.003	0.988

Table 4.7: Regions of differential chromatin with the H3K27ac histone mark between Intronic Enhancer Deletion clones and unedited controls (unstimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. Only one peak was significantly differential (FDR of <0.05), which corresponds to the region that was deleted. FDR from DESeq2. The coordinates stated are in UCSC notation based on genome build hg19.

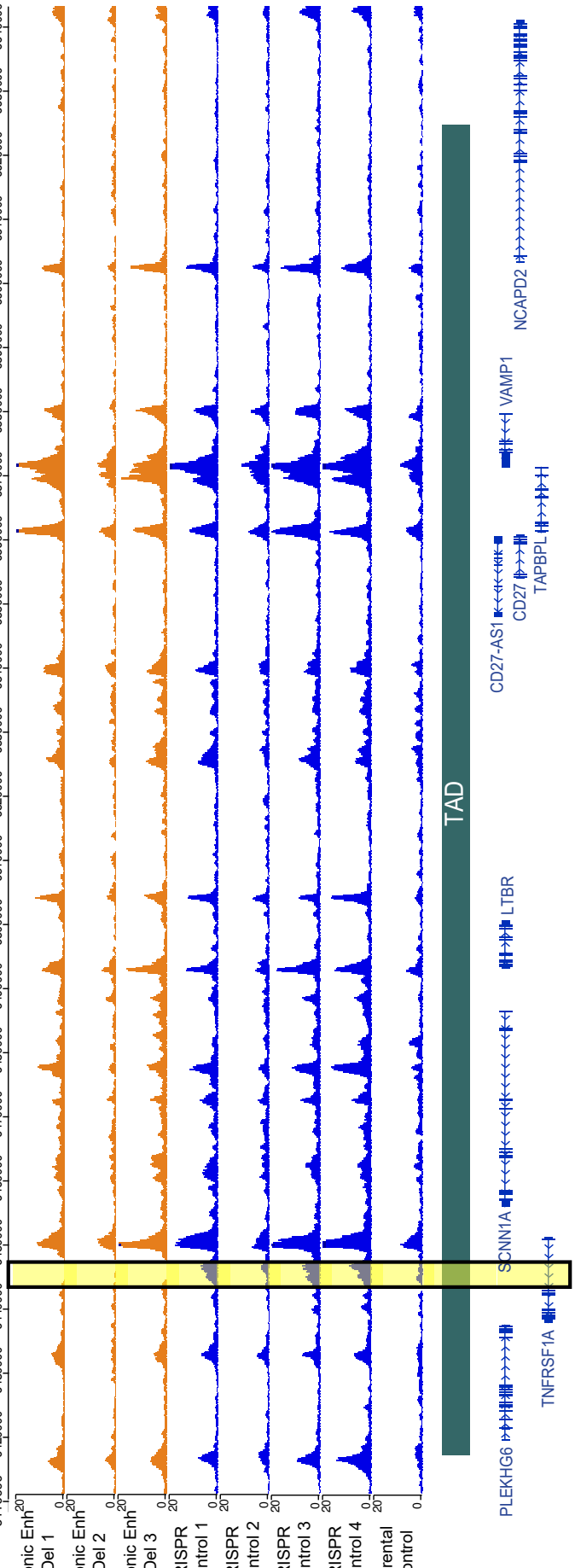


Figure 4.21: Regions of H3K27ac enhancer histone mark modifications in Intronic Enhancer Deletion clones and unedited controls in an unstimulated state.
WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Three Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 7.31×10^{-25}).

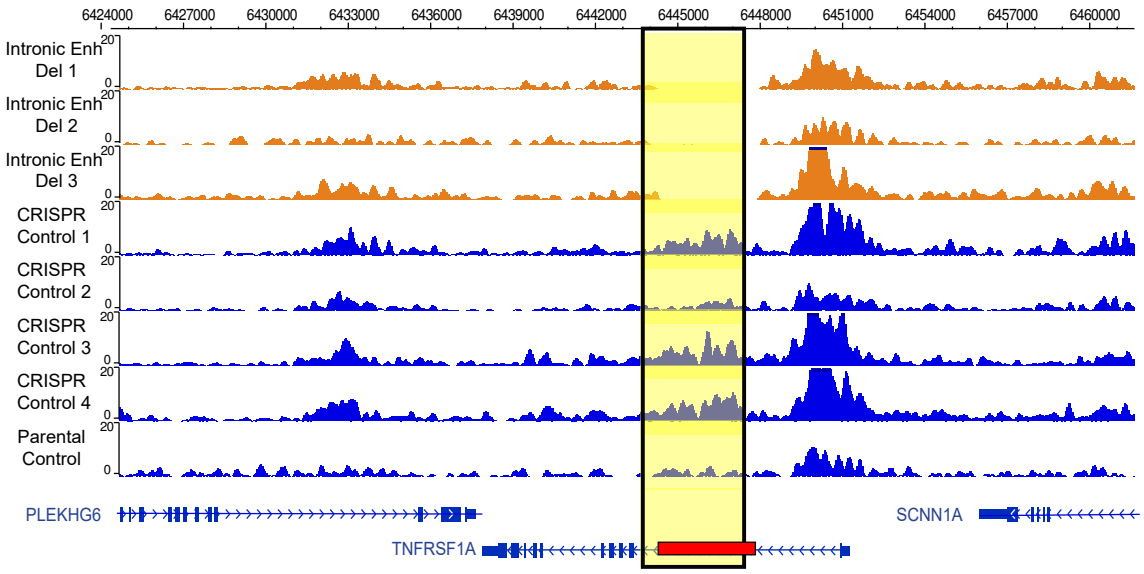


Figure 4.22: Close up of the differential H3K27ac modification peak over the Intronic Enhancer Deletion in edited clones and unedited controls in an unstimulated state. WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages. Three Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 7.31×10^{-25}). The deletion region is indicated by the red box.

The presence of the H3K27ac histone mark, was also interrogated in Intergenic Enhancer Deletion edited clones and unedited controls. None of the peaks were differential between the edited and unedited samples (Table 4.8). There was no discernible peak over the Intergenic Enhancer Deletion region (Figure 4.23), but it does partially overlap a peak that was called (chr12:6453927-6455189). However, this peak is small and was not differential in the analysis. Figure B.9 in Appendix B illustrates the peaks for each sample across the entire TAD.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6432218-6433603	*PLEKHG6*	0.353	0.751
chr12:6448993-6452438	*TNFRSF1A*	-0.131	0.751
chr12:6471971-6473762	*SCNN1A*	0.328	0.751
chr12:6476454-6478580	*SCNN1A*	0.478	0.751
chr12:6481586-6482482	*SCNN1A*	-0.613	0.751
chr12:6487632-6489030	*LTBR*	-0.440	0.751
chr12:6492222-6494533	*LTBR*	-0.252	0.751
chr12:6503005-6504920	*LTBR*	0.649	0.751
chr12:6559861-6562469	*CD27/TAPBPL*	-0.298	0.751
chr12:6578918-6580886	*VAMP1*	-0.402	0.751
chr12:6441496-6442515	*TNFRSF1A*	0.187	0.870
chr12:6443654-6447439	*TNFRSF1A*	-0.119	0.870
chr12:6453927-6455189	*TNFRSF1A*	0.332	0.870
chr12:6460247-6460887	*SCNN1A*	0.167	0.870
chr12:6460934-6462955	*SCNN1A*	0.083	0.870
chr12:6467539-6468401	*SCNN1A*	0.246	0.870
chr12:6495594-6496638	*LTBR*	0.211	0.870
chr12:6524686-6527917	NA	-0.073	0.870
chr12:6532784-6533655	NA	-0.184	0.870
chr12:6538819-6540518	NA	-0.103	0.870
chr12:6568210-6573444	*TAPBPL/VAMP1*	0.082	0.870
chr12:6501613-6502113	*LTBR*	0.065	0.920
chr12:6601568-6603078	*NCAPD2*	0.519	0.920
chr12:6603193-6604157	*NCAPD2*	-0.193	0.920

Table 4.8: Regions of differential chromatin with the H3K27ac histone mark between Intergenic Enhancer Deletion clones and unedited controls (unstimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significantly differential (FDR of <0.05) between samples. The coordinates stated are in UCSC notation based on genome build hg19.

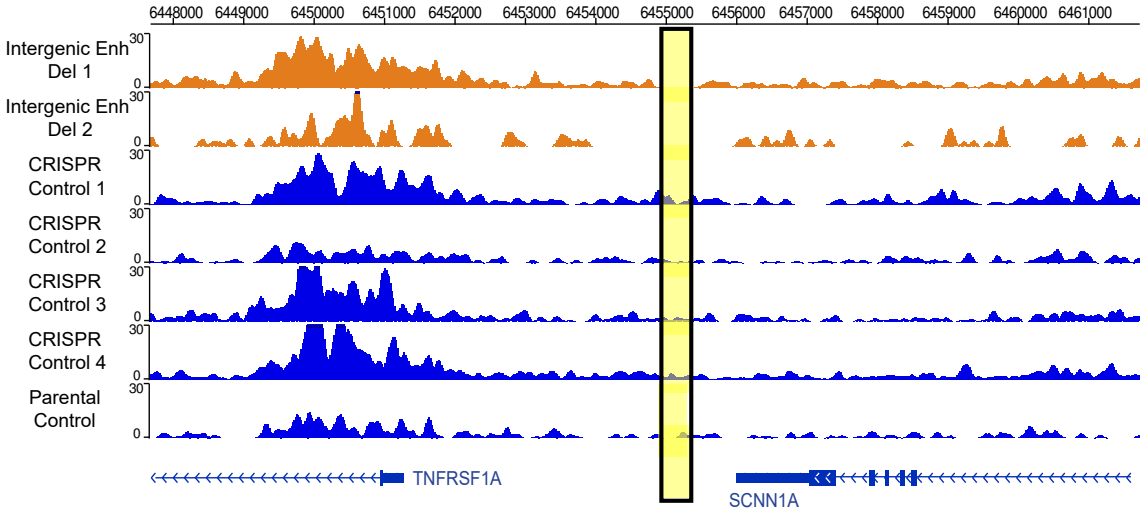


Figure 4.23: Close up of the H3K27ac modification peaks over the Intergenic Enhancer Deletion in edited cells and unedited controls in an unstimulated state.
WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages. Two Intergenic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. None of the peaks were significantly differential between edited and unedited samples. The deleted region is highlighted in yellow.

The H3K27ac Landscape in Stimulated Edited iPSC Derived Macrophages

Stimulation with LPS to mimic an immune response can change the activity of enhancers and promoters. Therefore, the presence of the H3K27ac histone mark was interrogated in edited clones and unedited controls after stimulation with LPS. The results of the differential analysis between the Intronic Enhancer Deletion clones and the unedited controls post LPS treatment was the same as the unstimulated analysis. The peak over the Intronic Enhancer (chr12:6443654-6447439) was the only significantly differential peak with a Log_2 fold Change of 4.798 and an FDR of $1.12x10^{-26}$ (Figure B.10 and Table B.12 in Appendix B). The Log_2 fold Change for this peak in the LPS stimulated cells was 4.798 compared to 3.638 in the unstimulated analysis, suggesting that this peak is larger in stimulated cells. Figure B.11 in Appendix B illustrates the peaks for each sample across the entire TAD.

The differential analysis was repeated for Intergenic Enhancer Deletion edited clones and unedited controls following LPS stimulation. None of the peaks were

significantly differential following this analysis and there was no discernible peak over the deleted region (Figures B.12 and B.13 and Table B.13 in Appendix B).

H3K27ac Between Stimulated and Unstimulated iPSC Derived Macrophages

To determine how LPS affects the H3K27ac histone mark landscape within the *TNFRSF1A* TAD, unstimulated samples were compared to LPS stimulated cells. A differential analysis between unstimulated and LPS stimulated states was carried out, with samples being grouped as either unstimulated or LPS stimulated, not as edited or unedited. However, the Intronic Enhancer Deletion clones and Intergenic Enhancer Deletion clones were grouped separately for two independent analyses, but the same control samples were used for both.

The first analysis included Intronic Enhancer Deletion clones and controls. There were eleven differential peaks in this sample group between the unstimulated samples and the LPS treated samples (Table 4.9). Figure 4.24 is a visualization of the peaks over the whole TAD with the differential peaks highlighted. The LPS treatment affected the peaks in both directions, with some becoming smaller, and some larger. The peak located at chr12:6568210-6573444, which spans parts of both *TAPBPL* and *VAMP1*, was reduced with LPS stimulation and was the most differential (FDR $7.21\text{x}10^{-14}$). The peak over chr12:6495594-6496638, within *LTBR*, had a large fold change (-1.381) and was the second most significant (FDR $7.34\text{x}10^{-7}$), suggesting that it is a context dependent enhancer and is more active under LPS stimulation. The peak located at chr12:6532784-6533655 does not overlap with any genes but has the highest fold change (1.472) and is the third most differential (FDR $7.34\text{x}10^{-7}$). It could be an intergenic enhancer that is not active, or less active under immune stimulation. The peak spanning chr12:6443654-6447439 which corresponds to the Intronic Enhancer in *TNFRSF1A* was not differential in this analysis despite the peaks visually appearing larger in the LPS samples. However, the peak located at

chr12:6453927-6455189, which partially overlaps the Intergenic Enhancer, is differential in this analysis (FDR $7.38x10^{-4}$) and appears to be more active in the presence of LPS.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6568210-6573444	*TAPBPL/VAMP1*	0.648	$7.21x10^{-14}$
chr12:6495594-6496638	*LTBR*	-1.381	$7.34x10^{-7}$
chr12:6532784-6533655	NA	1.472	$7.34x10^{-7}$
chr12:6441496-6442515	*TNFRSF1A*	-1.063	$2.74x10^{-4}$
chr12:6460247-6460887	*SCNN1A*	1.084	$2.74x10^{-4}$
chr12:6453927-6455189	*TNFRSF1A*	-0.815	$7.38x10^{-4}$
chr12:6538819-6540518	NA	0.557	$1.49x10^{-3}$
chr12:6501613-6502113	*LTBR*	-0.862	$9.37x10^{-3}$
chr12:6603193-6604157	*NCAPD2*	-0.704	0.012
chr12:6559861-6562469	*CD27/TAPBPL*	0.367	0.014
chr12:6601568-6603078	*NCAPD2*	-0.342	0.021
chr12:6460934-6462955	*SCNN1A*	0.236	0.354
chr12:6492222-6494533	*LTBR*	0.224	0.354
chr12:6476454-6478580	*SCNN1A*	-0.187	0.432
chr12:6524686-6527917	NA	-0.145	0.432
chr12:6578918-6580886	*VAMP1*	0.161	0.432
chr12:6467539-6468401	*SCNN1A*	-0.206	0.509
chr12:6443654-6447439	*TNFRSF1A*	-0.848	0.543
chr12:6448993-6452438	*TNFRSF1A*	-0.057	0.712
chr12:6487632-6489030	*LTBR*	0.102	0.712
chr12:6481586-6482482	*SCNN1A*	-0.116	0.717
chr12:6432218-6433603	*PLEKHG6*	-0.015	0.938
chr12:6471971-6473762	*SCNN1A*	-0.013	0.938
chr12:6503005-6504920	*LTBR*	-0.026	0.938

Table 4.9: Regions of differential chromatin with the H3K27ac histone mark between unstimulated and LPS stimulated Intronic Enhancer Deletion clones and controls. This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significant different in cells treated with LPS. FDR from DESeq2. The Log_2 fold Change represents the difference between LPS stimulated and unstimulated samples. The coordinates stated are in UCSC notation based on genome build hg19.

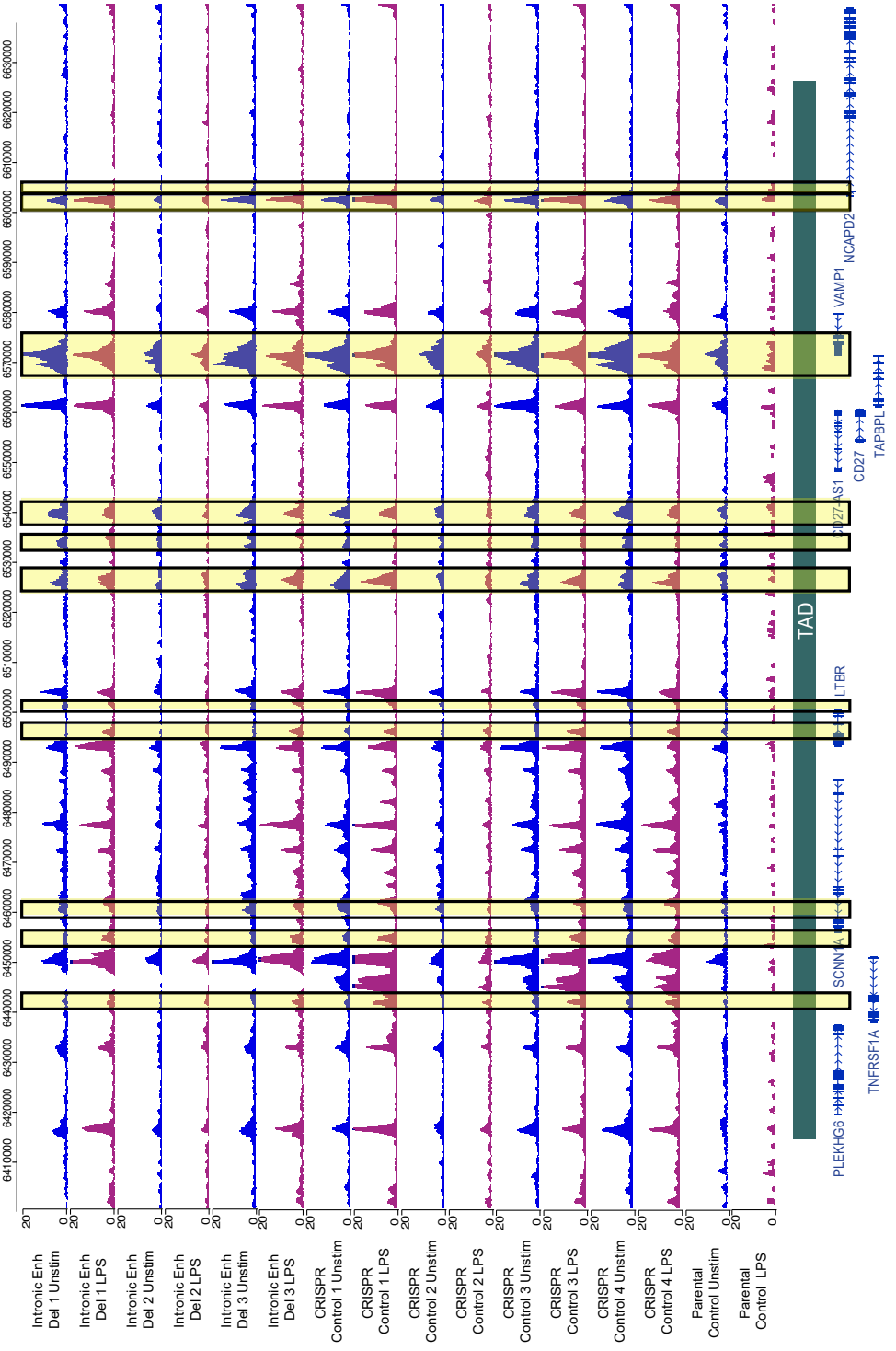


Figure 4.24: Differential H3K27ac enhancer histone mark peaks between unstimulated and LPS stimulated Intronic Enhancer Deletion clones and controls.

WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages across the entire TAD either unstimulated or LPS stimulated. Unstimulated samples are coloured blue while LPS stimulated are coloured magenta. Three Intronic Enhancer Deletion edited clones, four CRISPR Controls, and the Parental Control are shown. The differential paks are highlighted in yellow.

There were ten differential peaks in the analysis between stimulated and unstimulated Intergenic Enhancer samples (Table 4.10). Figure 4.25 is a visualization of the peaks over the whole TAD with the differential peaks highlighted. The peak located at chr12:6443654-6447439, within *TNFRSF1A* was the most differential with an FDR of $1.41x10^{-14}$ (Figure 4.26). This peak is situated over the Intronic Enhancer region and appears to represent an enhancer that is much more active when stimulated with LPS. This peak was not significantly differential between unstimulated and stimulated samples during the previous analysis using the Intronic Enhancer Deletion edited clones. Because three of the samples in that analysis had this peak deleted, it may have interfered with the analysis. The peak slightly overlapping the Intergenic Enhancer Deletion (chr12:6453927-6455189) was also differential in this analysis (FDR $6.79x10^{-3}$) and appears to be more active when stimulated with a pro-inflammatory factor. In the RNA-Seq analysis, *NCAPD2* appeared to be differentially expressed in samples where the Intergenic Enhancer was removed. The peak near the promoter of the *NCAPD2* gene (chr12:6601568-6603078) is differential for H3K27ac when stimulated (FDR 0.0466). Stimulation with LPS appears to increase promoter activity and therefore *NCAPD2* could play a role in the immune response.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6443654-6447439	*TNFRSF1A*	-1.057	$1.41x10^{-14}$
chr12:6568210-6573444	*TAPBPL/VAMP1*	0.756	$2.47x10^{-10}$
chr12:6441496-6442515	*TNFRSF1A*	-1.418	$1.94x10^{-8}$
chr12:6495594-6496638	*LTBR*	-1.388	$8.02x10^{-6}$
chr12:6532784-6533655	NA	1.426	$2.61x10^{-5}$
chr12:6559861-6562469	*CD27/TAPBPL*	0.428	$6.72x10^{-3}$
chr12:6453927-6455189	*TNFRSF1A*	-0.906	$6.79x10^{-3}$
chr12:6460247-6460887	*SCNN1A*	1.001	$6.79x10^{-3}$
chr12:6538819-6540518	NA	0.516	$8.04x10^{-3}$
chr12:6601568-6603078	*NCAPD2*	-0.348	0.0466
chr12:6487632-6489030	*LTBR*	0.393	0.151
chr12:6492222-6494533	*LTBR*	0.265	0.186
chr12:6501613-6502113	*LTBR*	-0.514	0.372
chr12:6603193-6604157	*NCAPD2*	-0.289	0.502
chr12:6467539-6468401	*SCNN1A*	-0.171	0.807
chr12:6503005-6504920	*LTBR*	-0.100	0.807
chr12:6460934-6462955	*SCNN1A*	0.075	0.871
chr12:6524686-6527917	NA	-0.067	0.871
chr12:6432218-6433603	*PLEKHG6*	-0.069	0.885
chr12:6448993-6452438	*TNFRSF1A*	-0.028	0.915
chr12:6476454-6478580	*SCNN1A*	-0.039	0.915
chr12:6481586-6482482	*SCNN1A*	0.079	0.915
chr12:6471971-6473762	*SCNN1A*	-0.025	0.930
chr12:6578918-6580886	*VAMP1*	-0.021	0.930

Table 4.10: Regions of differential chromatin with the H3K27ac histone mark between unstimulated and LPS stimulated Intergenic Enhancer Deletion cells and controls. This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significant different in cells treated with LPS. FDR from DESeq2. The Log_2 fold Change represents the difference between LPS stimulated and unstimulated samples. The coordinates stated are in UCSC notation based on genome build hg19.

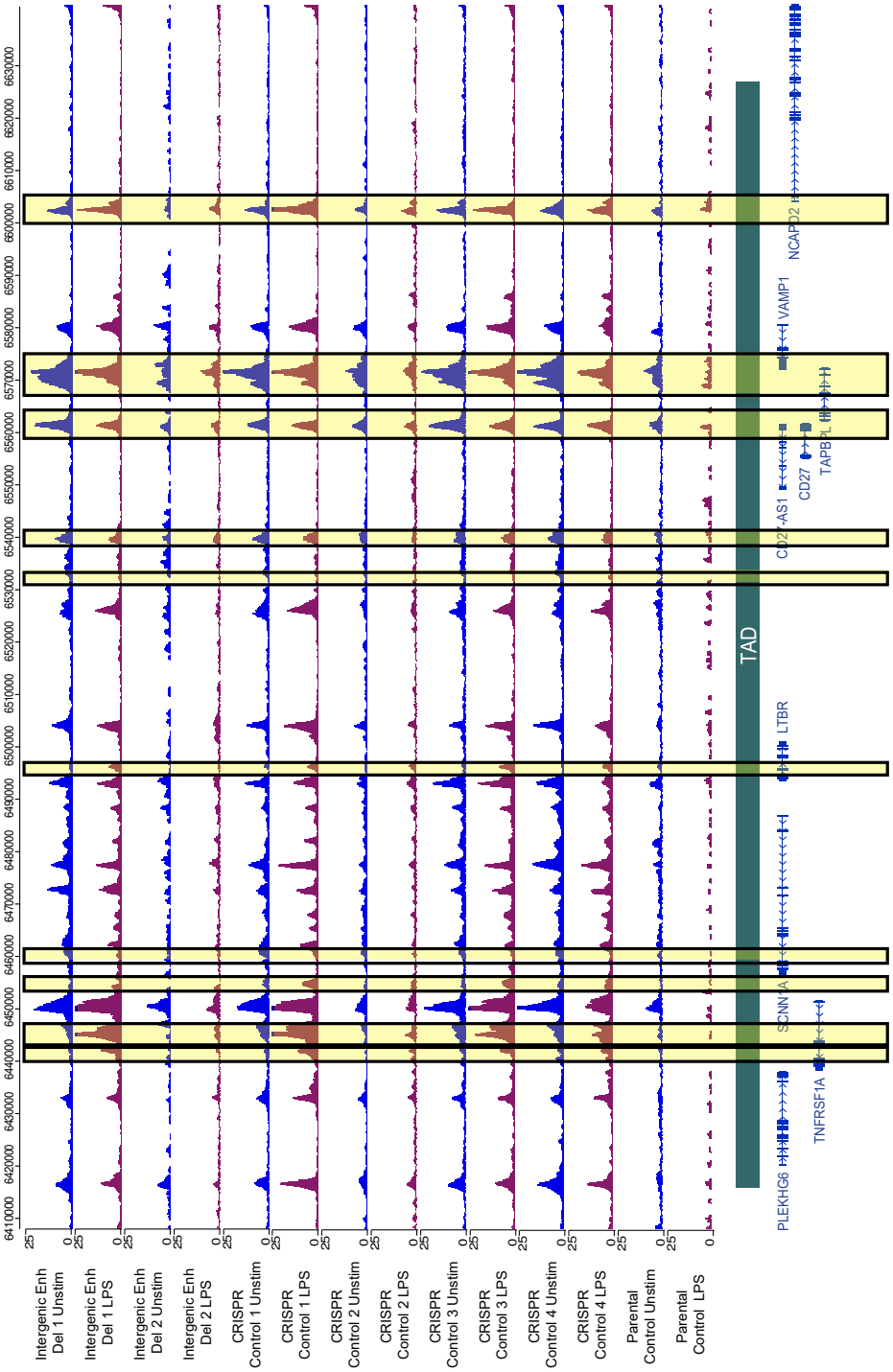


Figure 4.25: Differential H3K27ac enhancer histone mark peaks between unstimulated and LPS stimulated Intergenic Enhancer Deletion cells and controls.

WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages across the entire TAD, either unstimulated or LPS stimulated. Unstimulated samples are coloured blue while LPS stimulated are coloured magenta. Two Intergenic Enhancer Deletion edited clones, four CRISPR Controls, and the Parental Control are shown. The differential peaks are highlighted in yellow.

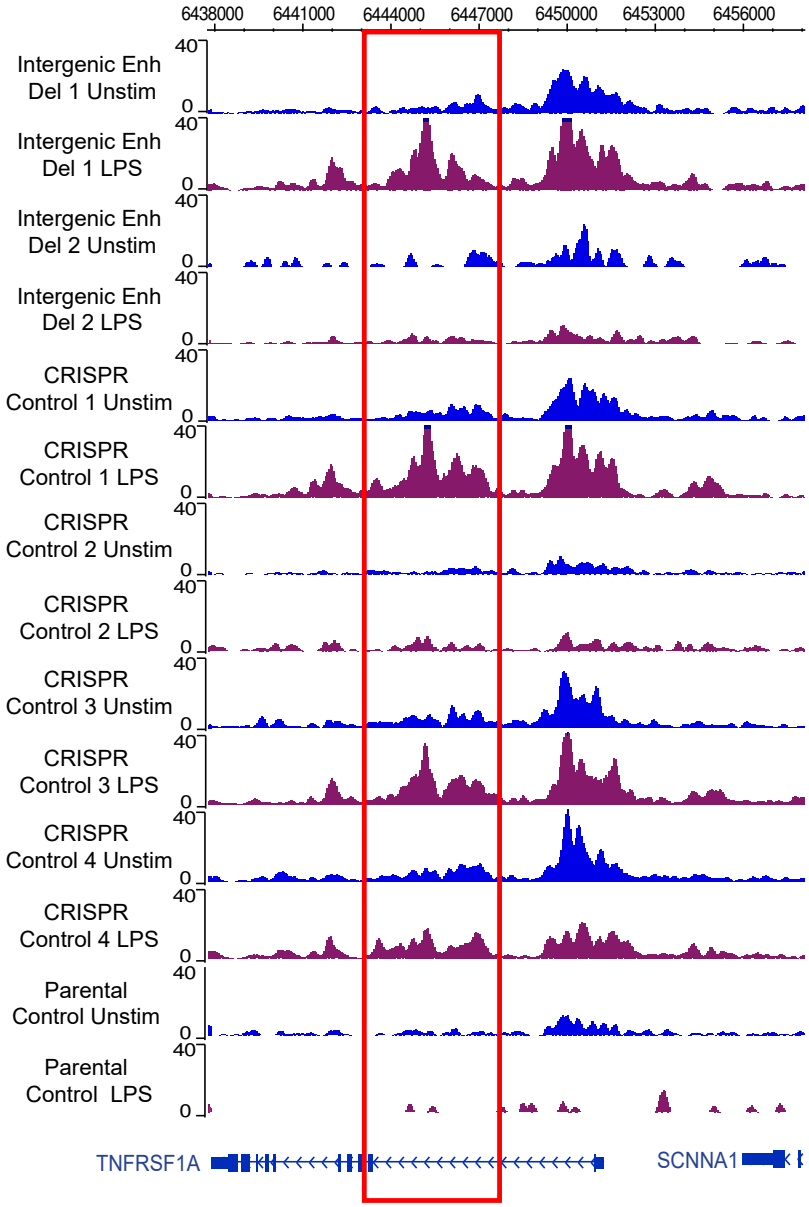


Figure 4.26: Close up of the H3K27ac histone modification peak between unstimulated and LPS stimulated Intergenic Enhancer cells.
Peak chr12:6443654-6447439 with a Log_2 fold Change of -1.057 and an FDR of $1.41x10^{-14}$ was the most differential peak and is outlined in red. Unstimulated samples are coloured blue while LPS stimulated are coloured magenta. Two Intronic Enhancer Deletion edited clones, four CRISPR Controls, and the Parental Control are shown.

4.2.6 Examination of the *TNFRSF1A* Locus in Primary Cells

To further validate the findings here and to illustrate its relevance in AS, *TNFRSF1A* was examined using epigenetic data across $CD4^+$, $CD8^+$, and $CD14^+$ primary cells in AS patients and controls (Figure 4.27). The samples used for this analysis are outlined in Chapter 5.

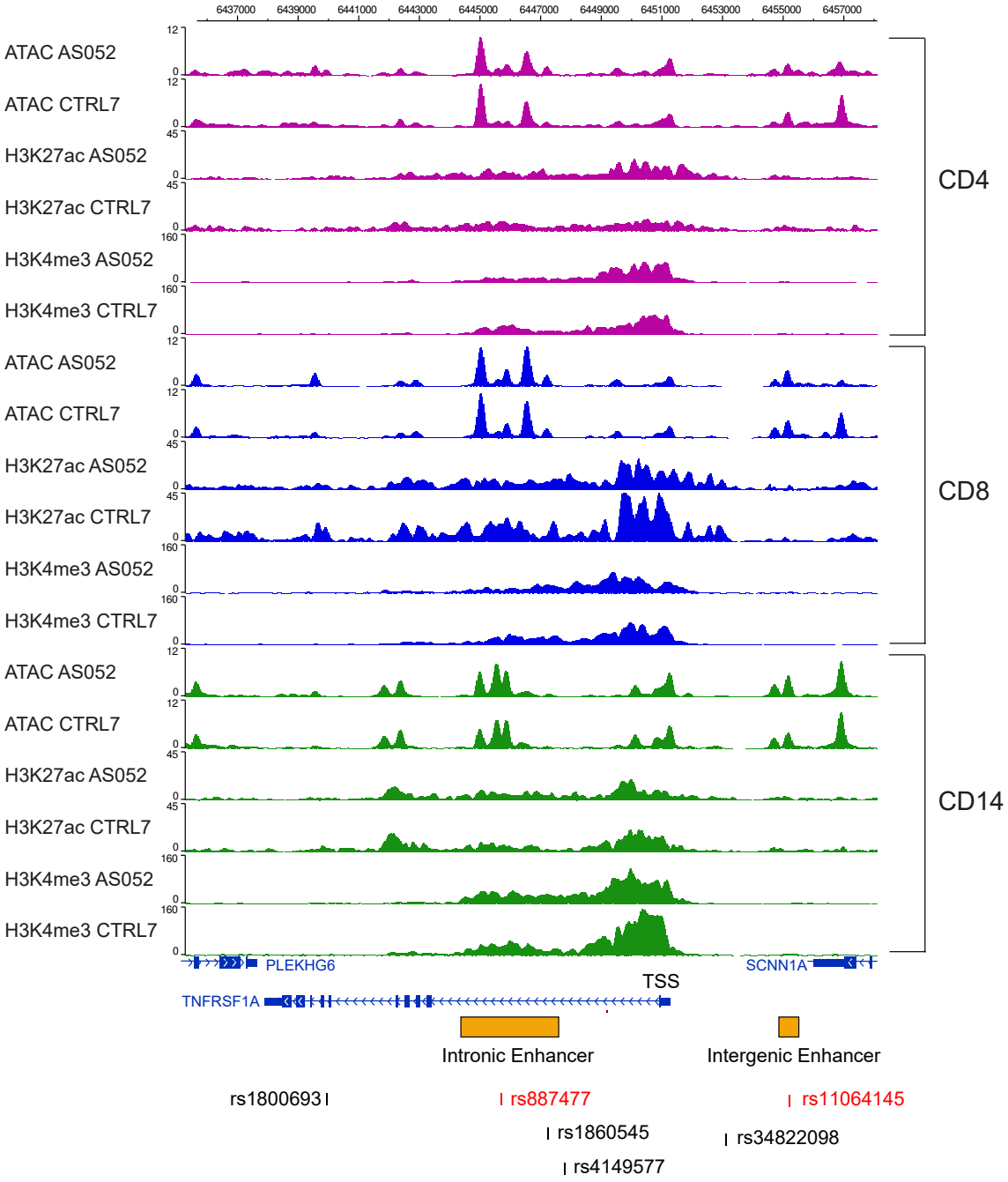


Figure 4.27: Chromatin accessibility and histone modifications at the *TNFRSF1A* locus.
WashU Epigenome Browser view of the results of ATAC, H3K27ac ChIP, and H3K4me3 ChIP across $CD4^+$, $CD8^+$, and $CD14^+$ cells at the *TNFRSF1A* locus. One AS patient (AS052) and one control sample (CTRL7) are shown for each technique and cell type. The locations of the Intronic Enhancer and Intergenic Enhancer investigated are indicated. The two SNPs also investigated are highlighted in red. All six fine mapped SNPs at this locus are indicated.

TNFRSF1A is accessible and active in all three primary cell types as indicated by the ATAC-Seq peaks showing open chromatin over the gene promoter, and the presence of both H3K27ac and H3K4me3 (both markers of active promoters) over the start of the gene. There are large ATAC-Seq and H3K27ac ChIPm peaks over the Intronic Enhancer in all three cell types. The H3K4me3 ChIPm peaks are the strongest in $CD14^+$ monocytes at both the gene promoter and over the Intronic Enhancer. The Intergenic Enhancer also falls within ATAC peaks across all three cell types, but it did not overlap any regions with H3K27ac or H3K4me3.

4.3 Discussion

4.3.1 iPSC Derived Macrophages as a Model

iPSC derived macrophages have advantages and disadvantages as a model system for genetic studies. Studies have found that although they are not the same as human primary macrophages, iPSC derived macrophages are similar in their function, phenotype, and transcriptome to their blood derived and tissue-resident counterparts and recapitulate macrophage biology reasonably well (Alasoo et al. 2015; Wilgenburg et al. 2013; Karlsson et al. 2008; Zhang et al. 2015; Buchrieser et al. 2017; Cao et al. 2019). Their differentiation protocol is well established and produces $CD14^+$ cells for harvest every week for around three to four months (Wilgenburg et al. 2013). iPSCs are easier to transfect for CRISPR/Cas9 experiments than primary monocytes or macrophages, and they have the advantage of having an isogenic background to compare specific genetic manipulations.

There are, of course, some caveats to this model, however. Though iPSC derived macrophages are very similar to blood derived macrophages or tissue-resident macrophages, they are not the same. iPSC derived macrophages also do not proliferate so it can sometimes be difficult to acquire a large number of cells

for an experiment. However, this can be countered by setting up a large volume of differentiations for a single experiment. One of the main limitations of using these cells as a model is the differences between cells from different harvests. $CD14^+$ cells are harvested every week for experimentation and cells from older factories may give different results than cells from fresher ones. Recent thesis work by Julian Buchrieser at the Dunn School of Pathology has shown that cells from harvest one and two from the factories are still proliferative, so it was suggested that those cells should be used for pilot experiments only (Buchrieser 2016). A transcriptome analysis across harvests revealed that genes involved in the cell cycle are down-regulated between the first and fifth harvests, and genes involved in adaptive immunity are progressively up-regulated with each harvest (Buchrieser 2016). Based on these findings, cells from harvests three to six are considered optimal for experimentation. Cells from harvest five were used for RNA-Seq in this thesis, while cells from Harvests nine and ten were used for Omni-ATAC-Seq and ChIPm. This may explain differences in sample quality. The exact differences between cells across harvests and what drives them is not fully understood and under investigation.

4.3.2 Functional Genomics to Investigate Enhancers

The functional genomics work carried out here on edited iPSC derived macrophages is novel. Very little has been done in the way of epigenetic profiling in iPSC derived macrophages and no work has been carried out to characterize the two enhancers examined here. There have been multiple transcriptome analyses performed on iPSC derived macrophages (Alasoo et al. 2015; Zhang et al. 2015), but no published work on combined analysis of ATAC-Seq, ChIP-Seq or Capture-C in these cells. Although, these types of experiments have been carried out in other types of iPSC derivatives (Wang et al. 2018a; Smith et al. 2019). Smith et al. used an approach similar to the one presented in this chapter to examine GWAS variants for age-related macular degeneration (Smith et al.

2019). They used iPSC derived retinal pigment epithelium for their model and carried out RNA-Seq, ATAC-Seq, and ChIP-Seq for H3K27ac. Following their experiments, they were able to determine that a GWAS variant, rs943080, was the probable causal variant by associating it with altered chromatin accessibility and a decreased expression of *VEGFA*, which had previously been associated with disease. To fully understand the role of non-coding genomic regions it is important to consider their impact on gene expression as well as the epigenetic landscape. Therefore, RNA-Seq, Capture-C, ATAC-Seq, and ChIP-Seq were carried out on edited iPSC derived macrophages to understand the role of these putative enhancers in a biologically relevant model system.

Transcriptomic Analysis of Edited iPSC Derived Macrophages

Enhancers exist to regulate gene expression, and therefore to characterize the two putative enhancers that had been deleted, a gene expression analysis was carried out. Due to gene proximity, it was hypothesized that both the putative enhancers would act on *TNFRSF1A*. RNA-Seq was successfully performed on edited and unedited iPSC derived macrophages.

The differential analysis between the Intronic Enhancer Deletion clones and unedited controls, when restricted to the *TNFRSF1A* TAD, resulted in one gene being significantly differentially expressed, *TNFRSF1A* (FDR 0.0034). The expression of *TNFRSF1A* decreased in cells where the enhancer had been deleted with a Log_2 fold change of -0.27, indicating that the Intronic Enhancer is indeed an enhancer for *TNFRSF1A*. However this fold change is very small and with so few samples compared against each other, this analysis may not be reliable. To show this change in expression is robust, the analysis should be repeated on a larger scale.

The differential analysis between the Intergenic Enhancer Deletion clones and unedited controls, when restricted to the *TNFRSF1A* TAD, also resulted in one

gene being significantly differentially expressed, *NCAPD2* (FDR 0.012). In cells with the Intergenic Enhancer deleted, the expression level of *NCAPD2* increased (Log_2 fold change of -0.4). This suggests that this putative enhancer may not be an enhancer, but more likely a repressor. It may function to inhibit the normal expression of *NCAPD2* and in cells with this region removed, its expression is increased.

NCAPD2 is a subunit of condensin I, which is a protein complex responsible for the mediation of chromosome condensation and assembly for mitosis (Martin et al. 2016). It is also thought that condensin I works with topoisomerase II to untangle sister chromatids during cellular division (Baxter et al. 2011; Charbin et al. 2013). Recent studies have associated mutations in *NCAPD2* with microcephaly (Martin et al. 2016; Lin et al. 2019). They found that these mutations relate to condensin-deficient cells which cause chromosome segregation errors due to the inability of the cell to correct sister chromatin entanglement. The resultant increase in aneuploidy in daughter cells reduced cerebral cortex size. *NCAPD2* has also been linked to cancer, which is unsurprising given its role in cell division. Whole-exome sequencing and karyotypic analysis on Hodgkin lymphoma cells revealed defects in *NCAPD2* and a high amount of aneuploidy, suggesting *NCAPD2* may be causing abnormal chromosome segregation during cellular division (Hudnall et al. 2016). *NCAPD2* has also recently been associated with breast cancer, with defects being linked to poor disease progression and prognosis (Zhang et al. 2019). There is no clear link between *NCAPD2* and inflammation or the immune system, but one may exist. It has been associated with Hodgkin lymphoma which is thought to involve T and B cells, so links between *NCAPD2* and immune mediated disorders may be uncovered. The connection between *NCAPD2* and a putative regulatory region associated with AS through GWAS remains unclear. The Intergenic Enhancer may be a suppressor of *NCAPD2*, as the RNA-Seq analysis suggests, or an artifact

of the analysis with a small sample size. As with the Intronic Enhancer analysis, the fold change in expression is very small between edited and unedited samples and with so few samples compared against each other, this analysis may not be reliable. To show this change in expression is robust, the analysis should be repeated on a larger scale.

Mapping Chromatin Interactions occurring within and around *TNFRSF1A*

Capture-C was used to map the physical chromatin interactions occurring within the *TNFRSF1A* TAD to better understand the relationship between the genes and regulatory regions. Unfortunately, Capture-C was unsuccessful in CRISPR/Cas9 edited iPSC derived macrophages. Due to the necessity of large cell numbers for this experiment (≥ 10 million) it would be difficult to the perform this experiment on iPSC derived macrophages without explicitly setting up large quantities of macrophage factories solely for this purpose. Capture-C did work well on the $CD14^+$ cells isolated from blood. These cells are not identical to iPSC derived macrophages, but give a close representation of the chromatin interactions occurring. *TNFRSF1A* appears to be an important gene in the TAD and has many interactions with the other genes. Moreover, it may be interacting with *NCAPD2*, which the Intergenic Enhancer may be regulating, as evidenced by the RNA-Seq analysis. Unfortunately the two SNPs in question, rs887477 and rs11064145 and the putative enhancers they are situated within lie very close to gene promoters which make it hard to discern which interactions they may be responsible for.

Assaying Chromatin Accessibility in Edited iPSC Derived Macrophages

ATAC-Seq was used to understand the chromatin accessibility in and around *TNFRSF1A*. The TSS enrichment in these samples was very low (<5) and did not meet the recommended threshold as set by ENCODE (Dunham et al. 2012). This suggests that the samples are quite 'noisy' and it is more difficult to discern peaks

from background. Cell death during the experiment can reduce TSS enrichment, as floating DNA from dead cells is readily accessible for the transposase. Because iPSC derived macrophages are adherent cells, some may die in the process of detaching them from their dish, and this could result in an excess of dead cells. A potential way to reduce this issue is to use ficoll at the beginning of the experiment. Ficoll is a solution that will separate out cells by density and should enable for the selection of living cells (Corces et al. 2017). The cells could also be pre-treated with DNase to digest any floating DNA from dead cells (Corces et al. 2017). The samples had a large percentage of high quality reads (>75%) and a low percentage of reads mapping to the mitochondrial genome. One of the major issues faced in our lab when using other ATAC-Seq protocols was losing around 50% of sequencing reads to the mitochondrial genome. This protocol successfully solved this problem with only around 10% of reads being lost to mitochondria.

Despite the low TSS enrichment, clear peaks were visible throughout the *TNFRSF1A* TAD. Active genes should be marked by accessible chromatin to allow for the binding of transcription factors. The data here showed large peaks representing open chromatin over the promoters for all the expressed genes in the TAD, which is expected. Importantly, in control samples there was a very large peak over the Intronic Enhancer suggesting that this region is readily accessible for protein or chromatin binding and is indicative of it being a regulatory element. The Intronic Enhancer Deletion clones very clearly show a deletion of this peak, confirming that this region had been successfully deleted and that the surrounding chromatin was closed off. Interestingly the peak over the *TNFRSF1A* promoter appeared much larger in size in the Intronic Enhancer Deletion clones, suggesting that the cell increased chromatin accessibility over the promoter. This could be a compensatory mechanism by the cell. The RNA-Seq data from the Intronic Enhancer Deletion clones suggest that this enhancer regulates *TNFRSF1A* expression. With this enhancer removed, a drop in TNFRSF1A levels

could cause the cell to increase promoter accessibility in order to encourage transcription factor binding and increase transcription. However, although this peak is visually larger when viewing read pile-ups, it was not found to be significantly larger following the differential analysis between edited clones and controls.

There was no strong peak over the Intergenic Enhancer, suggesting that it may not be a region of open chromatin. However, the deletion was still clearly visible as no reads mapped to that region. The analysis between edited Intergenic Enhancer Deletion clones and unedited controls showed that there were no differential peaks due to enhancer removal, and LPS treatment did not effect this.

Assaying Enhancer Marks in Edited iPSC Derived Macrophages

ChIP-Seq was performed to reveal the chromatin binding location of the active enhancer and promoter histone mark, H3K27ac. The TSS enrichment in these samples was very low (<5) suggesting they may have background noise which can effect peak calling. As with ATAC-Seq, this could be improved by using ficoll to separate the living cells. Dying cells in the experiment may lose their nucleosome structure and increase background noise. There is no ENCODE standard for H3K27ac TSS because it is primarily a marker for enhancers, but it is still indicative of the signal to noise in the sample. Many of the samples also did not meet the suggested read depth as many sequencing reads were lost due to PCR duplicates, likely a result of not having enough starting material. This could be due to cell loss during their detachment from the dish. The chromatin input control samples were high quality indicating that the protocol worked well, but more starting material may have been helpful for the H3K27ac ChIPm samples.

Despite the low sample quality, peaks were discernible and provided insight into relevant histone modifications. The iPSC derived macrophages showed a large peak over the Intronic Enhancer that was being investigated, suggesting

that is a strong active enhancer. When stimulated with LPS, the peak over this enhancer increased in size, and in one of the analyses, was the most significantly increased peak between the unstimulated cells and those stimulated with LPS. Macrophages respond to LPS to initiate an immune response and therefore the epigenetic landscape within the cells should change upon stimulation to meet the needs of the cell. Chromatin regions where the H3K27ac histone mark increased suggest that the gene under regulation by that enhancer or promoter is involved in the immune response. This suggests that the Intronic Enhancer is an enhancer that regulates inflammatory or immunomodulatory gene(s) and could be relevant in AS pathogenesis.

There was no H3K27ac peak corresponding to the Intergenic Enhancer region. A portion of a small peak overlapped with the region that was deleted, but this peak was not significant in any of the differential analyses between edited and unedited samples. The peak did increase significantly when stimulated with LPS, suggesting it may be a context dependent enhancer. The samples exhibited a high amount of noise so it is difficult to determine if the small peak indicates an enhancer or if it is just background. With each deleted enhancer, however, it was very clear to see that they had been successfully deleted as there were no reads mapping to those areas.

4.3.3 Data Integration to Characterize Putative Enhancers

Based on the experiments performed on edited iPSC derived macrophages, the Intronic Enhancer is likely an enhancer for *TNFRSF1A*. It is situated within a region of open chromatin which means that it is accessible to transcription factors or other proteins, and available for chromatin interactions. It is also within a strong peak for H3K27ac, which denotes active enhancers. Moreover, when the Intronic Enhancer was deleted in a cell line, the expression of *TNFRSF1A* was reduced, suggesting that this region modulates its transcription. However,

the context dependency of this enhancer is inconclusive. Stimulation with LPS significantly increased H3K27ac over the enhancer suggesting it increased in activity, and the ATAC-Seq peak also appeared larger, suggesting increased accessibility. However, *TNFRSF1A* was only differentially expressed between edited cells and controls in an unstimulated state. No genes were differentially expressed in the analysis comparing Intronic Enhancer Deletion clones to unedited controls in a stimulated state. Because this region is likely an enhancer for *TNFRSF1A*, the GWAS SNP, rs887477, that lies within this enhancer could have a functional role in the immune response and relate back to AS.

The Intergenic Enhancer is unlikely to be an enhancer, but could possibly be a repressor or silencer, as evidenced by the experiments performed on edited iPSC derived macrophages. It was not found in an open chromatin peak, indicating that it may not be accessible for protein binding. Enhancers are found in regions of open chromatin, but silencer sequences are not necessarily (Kolovos et al. 2012). The Intergenic Enhancer also did not coincide with the H3K27ac histone mark, further suggesting that it is not an enhancer. Repressors are enriched for H3K27me3, so a ChIP-Seq experiment looking for this histone mark would help characterize this regulatory element. The RNA-Seq analysis highlighted *NCAPD2* as being positively differentially expressed in deletion clones. This provides evidence that this region of DNA may be responsible for the suppression of *NCAPD2* under normal physiological conditions. Though Capture-C was unsuccessful in edited iPSCs, $CD14^{+}$ cells illustrated a potential a physical interaction occurring between *TNFRSF1A*, where this region lies, and the *NCAPD2* promoter at approximately 150kb away. Though the significance of this interaction is not clear, the promoter for *NCAPD2* did become significantly stronger with LPS stimulation, as evidenced by ChIPm, so there may be an uncovered link between *NCAPD2* and the immune response. The GWAS SNP, rs11064145, that lies within this region could have a functional role but its

relation to inflammation or AS remains unclear. Rs11064145 was found to be an eQTL for *TNFRSF1A* in another study (Jansen et al. 2017), but the experiments here do not link the immediate region surrounding this SNP with *TNFRSF1A*.

4.3.4 Validation of the *TNFRSF1A* Locus

TNFRSF1A, was investigated in $CD4^+$, $CD8^+$, and $CD14^+$ primary cells to validate the findings from this chapter, and to determine if it would be worth pursuing for future functional studies in AS. It was concluded above that the Intronic Enhancer is likely an enhancer for *TNFRSF1A* and that the Intergenic Enhancer is perhaps a silencer or repressor, possibly for *NCAPD2*, both in iPSC derived macrophages. These findings mimic those found in primary immune cells. The Intronic Enhancer appears to be an enhancer in $CD4^+$, $CD8^+$, and $CD14^+$ cells, as it was found to be situated within a region of open chromatin and overlaps the H3K27ac and H3K4me3 active enhancer/promoter marks. The Intergenic Enhancer does not appear to be an enhancer in $CD4^+$, $CD8^+$, or $CD14^+$ cells as this region is not within H3K27ac or H3K4me3 regions, but it is within a region of chromatin accessibility. This is consistent with the finding that it is likely not an enhancer, but could be a silencer.

These findings were not exclusive to $CD14^+$ cells, which is the most similar cell type to iPSC derived macrophages. Many of the peaks were consistent across all three cell types, but some of them appeared larger in $CD8^+$ cells. It could be worth examining this locus in $CD8^+$ T cells and perhaps other cell types to better understand the GWAS SNPs and regulatory regions that lie within it. There were no significant differences between AS patients and controls at this locus, but regulatory elements in this region could be acting at a distance, or there may be differences in downstream pathways involving *TNFRSF1A*. It is also possible that a larger sample size is required. This locus remains a promising candidate to help understand AS pathogenesis and the genetic variants associated with disease.

4.4 Conclusions

iPSC derived macrophages are a good, disease relevant, immune cell model. They were used successfully for performing a transcriptomic and an epigenetic analysis to characterize putative enhancers and could be used for future experiments in AS or other immune-mediated disorders. RNA-Seq, ATAC-Seq, and ChIP-Seq were all successfully performed in CRISPR/Cas9 edited iPSC derived macrophages, but Capture-C was not due to cell number restrictions. The Intronic Enhancer is likely to be an enhancer for *TNFRSF1A*, as it is situated within open chromatin and an active enhancer mark, and its deletion results in down-regulation of *TNFRSF1A*. The Intergenic Enhancer potentially acts to repress *NCAPD2*, as it was not found within open chromatin or an active enhancer mark, and its deletion resulted in up-regulation of *NCAPD2*. However, to validate these findings the experiments should be repeated with more replicates and sample numbers, preferably with higher quality.

5

Defining Putative Regulatory Elements in Ankylosing Spondylitis

Somewhere, something incredible is waiting to be known. -Carl Sagan

5.1 Introduction 193
5.2 Results 198
5.3 Discussion 251
5.4 Conclusions 257

5.1 Introduction

5.1.1 Epigenetic Regulation of Gene Expression

Only 2% of the entire human genome codes for proteins. The vast majority of GWAS associated SNPs are found within the non-coding genome, suggesting they play a important role in the regulation of gene expression. Supporting this, 60% of of these non-coding GWAS variants are situated within DNase I hypersensitivity sites, with an additional 20% in LD with DNase I hypersensitivity sites (Maurano et al. 2012). These regions of open chromatin are predictive of regulatory elements and represent the potential for DNA-protein interactions to occur and facilitate gene transcription (Ward and Kellis 2012; Kilpinen et al. 2013; Knight 2012). Chromatin accessibility can be modulated by histone acetylation and methylation, as well as DNA methylation. Gene expression is controlled through complex processes involving proteins, non-

coding RNAs, and physical DNA interactions. Mutations impacting these can have detrimental effects on chromatin organization and gene regulation. For example, one SNP, rs174556, in LD with a Crohn's disease variant, was found to disrupt a binding site for CTCF, which in turn altered chromatin conformation and accessibility (Gate et al. 2018). Understanding the epigenetic landscape in the context of AS is an important to step in understanding its pathology and the genetic mutations which drive it.

5.1.2 Assaying Chromatin Accessibility Using Omni-ATAC-Seq

Transcription factors and other regulatory proteins can bind promoters, enhancers, and insulators to drive transcription, but only where the chromatin is readily accessible, and these regions can be mapped using ATAC-Seq. There have been several variations of the ATAC technique published, with the most recent being Omni-ATAC (Buenrostro et al. 2013; Corces et al. 2016; Corces et al. 2017). The initially reported method for ATAC-Seq suffers from a high number of unusable sequencing reads due mapping to the mitochondrial genome, and it is not well suited to all cell or tissue types (Buenrostro et al. 2013). Corces et al. made improvements to this protocol, creating Omni-ATAC, to maximize reads and improve signal to noise across multiple cell types (Corces et al. 2017). Omni-ATAC uses multiple detergents to increase cell permeabilization and has an additional post-lysis wash step to reduce the number of mitochondria being sequenced. It also alters the transposition reaction to improve the signal-to-background ratio to allow for more accurate peak calling. It reportedly produces a 13-fold decrease in mitochondrial reads, drastically reducing the cost of sequencing and improving data quality by increasing the percentage of reads that map to peaks. Traditional ATAC-Seq, Fast-ATAC, and Omni-ATAC have all been used in the Knight Group, with Omni-ATAC being the preferred method, as it maximizes on reads due to mitochondria reduction and can be performed with only a small amount of cells, and was therefore selected for this thesis.

All methods of ATAC-Seq are typically carried out on fresh samples. Most of the data from AS patients and healthy controls used in this thesis were derived from ATAC-Seq experiments performed on fresh cells, the same day they were isolated from blood. However, Omni-ATAC performed on a smaller number of AS patients and healthy controls was carried out on frozen formaldehyde fixed samples, as they were collected and stored in 2012/2013. The ATAC-Seq analysis presented here consists of multiple batches of AS patients and healthy controls. Data from some batches were generated using traditional ATAC-Seq and Fast-ATAC, as well as Omni-ATAC and one batch, deemed "historic", consists of Omni-ATAC performed on frozen formaldehyde fixed samples. Three disease-relevant, primary cell types were examined, $CD8^+$, $CD4^+$, and $CD14^+$ cells.

5.1.3 Assaying Histone Modifications Using ChIPmentation

ChIP-Seq provides knowledge of DNA-binding proteins and can be used to look for the presence of transcription factors or histone modifications to characterize regulatory regions and assay levels and regulation of gene activation. Despite its informativeness, ChIP-Seq can be unsuitable for a number of studies because it requires a substantial amount of cells, and is laborious and expensive. ChIPmentation (ChIPm) was recently introduced by Schmidl et al. to overcome some of the limitations of traditional ChIP-Seq, but uses the same fundamental procedure (Schmidl et al. 2015). Similar to ATAC-Seq, ChIPm utilizes tagmentation to fragment and tag DNA in a single step, which decreases the protocol time. It has also been optimized for low input and has been successful in assaying both transcription factors and histone marks. ChIPm was selected for this thesis to assay the histone marks, H3K27ac and H3K4me3, in AS patients and healthy controls. H3K27ac primarily indicates the presence of an active enhancer and H3K4me3 is primarily found at the location of active gene promoters, but both marks can overlap at either enhancers or promoters (Hon et al. 2009; Creyghton et al. 2010; Heintzman et al. 2007). All samples

in this analysis were formaldehyde fixed the same day the sample was taken from the AS patient or healthy volunteer and frozen. For most samples, ChIPm and subsequent analysis followed within a year. However, a small batch deemed "historic", consists of ChIPm performed on frozen formaldehyde fixed samples that had been in the freezer since 2012/2013 before processing in 2018/2019. Three cell types were examined, $CD8^+$, $CD4^+$, and $CD14^+$ cells.

5.1.4 Integrating -OMICS Data to Define Regulatory Regions

Regulatory elements in the genome can best be interrogated by integrating informative datasets about the structure and function of the non-coding genome. The Knight lab has generated RNA-Seq data from AS patients and healthy controls to examine which genes are differentially expressed in disease. These differentially expressed genes may overlap with differentially open chromatin and differential histone modifications in and around the genes. These regions may also overlap GWAS associated SNPs, providing evidence as to their role in disease. When combined, these datasets can be used to uncover putative regulatory elements that may be involved in AS pathogenesis. These regulatory elements can then be investigated further in a biological system and provide a new understanding of gene regulation in disease. A recent study on type 2 diabetes performed Hi-C in pancreatic islet samples to map physical chromatin interactions, and integrated these data with ATAC-Seq and ChIP-Seq datasets (Greenwald et al. 2019). They found putative genes corresponding to thousands of islet enhancers and that chromatin looping of enhancers is context specific and effects gene expression only in islet cells. They further overlapped these data with fine mapped GWAS hits and eQTL data sets to implicate associated non-coding SNPs with roles in chromatin conformation and enhancer function to reduce the number of potential causative variants. A similar approach of integrating functional genomic datastes is implemented in this chapter to understand the genetic and epigenetic drivers of AS.

5.1.5 Aims

The overall aim of this chapter is to understand the epigenetic landscape in $CD4^+$ and $CD8^+$ T cells, and $CD14^+$ monocytes, and how it differs between AS patients and healthy controls, allowing disease associated regulatory elements to be defined. The work outlined in this chapter includes ATAC-Seq to examine regions of chromatin accessibility, ChIP-Seq for H3K27ac to map regions of active enhancers, and ChIP-Seq with H3K4me3 to map regions of active promoters. The specific aims for this chapter are:

1. To identify regions of open chromatin in primary $CD4^+$, $CD8^+$, and $CD14^+$ cells.
2. To determine differential regions of open chromatin between AS patients and healthy controls in the three immune cell types.
3. To identify regions of H3K27ac, denoting active enhancers, and H3K4me3, denoting active promoters, in primary $CD4^+$, $CD8^+$, and $CD14^+$ cells.
4. To determine differential regions of H3K27ac and H3K4me3 between AS patients and healthy controls in the three immune cell types.
5. To integrate the analysis of differentially open chromatin, and regions of H3K27ac and H3K4me3 with differential gene expression and AS GWAS data to define disease associated regulatory elements.

5.2 Results

5.2.1 Patients and Healthy Controls Recruited

A total of 19 AS patients and 42 healthy controls, of British or European ancestry, were recruited and included in the analysis for this thesis work. Six of the patients were female and 13 were male. The patients recruited had an average BASDAI score of 4.97 ± 2.94 (range 0.66 to 8.68), an average age of diagnosis of 34 ± 12 (range 14 to 61), an average age at recruitment of 40 ± 14 (range 22 to 73), and an average disease duration of 7.83 years ± 7.17 (range 0.5 to 25). Healthy controls had an average age of 47 ± 8 (range 31 to 67), with 25 females and 17 males. All patients were recruited with active disease when seen in clinic prior to starting biologic therapy (anti-TNF). Patients were assessed based on questions about their lifestyle, family disease history, and comorbidities (Table 5.1). A BASDAI score was determined for each patient (see Table 1.2 in Chapter 1 for criteria), and patients with both low and high BASDAI scores were included in the analysis as it is a subjective measure of pain and joint stiffness. Up to three cell types were analyzed per patient or control, which included $CD4^+$ and $CD8^+$ T cells, and $CD14^+$ monocytes. Patient samples were were used to examine regions of open chromatin (ATAC-Seq), histone modifications (ChIP-Seq), and gene expression (RNA-Seq done by previous group members). Not all patient samples underwent all three analyses. The cell types, method of ATAC-Seq used, and the batch of recruitment for each patient and control are outlined in Tables C.1 and C.2, respectively in Appendix C.

Patient ID	Age of Onset	Smoker	Family History of AS	Extra-Articular Manifestations	Other Complications	Current Treatment
AS004	39	Yes	No	PJI	None	Ibuprofen, Diclofenac
AS014	27	Yes	Great Grandmother	Uveitis	None	Naproxen, Tramadol
AS017	38	Previously	No	Uveitis	Psoriasis, fractures	Ibuprofen, Naproxen
AS018	27	Yes	Sibling	None	Psoriasis	Naproxen
AS019	18	Previously	Yes	PJI	Psoriasis, RA, UC, stroke	Arthrotec
AS035	23	No	Father	Uveitis	Fractures	Arcoxia
AS036	35	Previously	Aunt	PJI, uveitis	None	None
AS037	42	No	No	PJI, uveitis	Psoriasis	Naproxen, Diclofenac
AS038	21	No	Father	Uveitis	None	None
AS041	32	No	Father	None	None	Etoricoxib
AS043	52	Yes	No	PJI	None	Naproxen
AS044	14	No	Unknown	None	None	Naproxen
AS045	27	Previously	No	PJI	None	Arcoxia
AS046	31	No	Unknown	Vertebral fractures	Fractures	Naproxen
AS047	35	Yes	Parent	PJI	None	Naproxen
AS049	38	Yes	No	None	UC	Naproxen
AS050	42	Previously	No	PJI	Fractures	None
AS051	52	No	Unknown	Uveitis	None	None
AS052	61	No	No	Uveitis	Cancer, fractures	None

Table 5.1: Patients Recruited.
RA = rheumatiod arthritis, UC = ulcerative colitis, PJI = Peripheral joint involvement.

5.2.2 The Chromatin Accessibility Landscape in AS

ATAC-Seq Quality Control

A total of 14 patients and 23 healthy controls were included in the ATAC-Seq analysis. With 11 traditional ATAC-Seq samples (five patients and six controls), six Fast-ATAC samples (three patients and three controls), 12 Omni-ATAC samples (one patient and 11 controls), and 8 historic samples (five patients and three controls). A subset of the samples were processed by Andrew Brown, Alicia Lledó Lara, and Anna Sanniti. All samples were sequenced and run through the in-house pipeline as described in Chapter 2. ENCODE recommends 50 million reads per sample (25 million fragments) for ATAC-Seq experiments and all but ten samples reached this threshold (Figure 5.1). The total reads were filtered, and low quality reads, PCR duplicates, and reads mapping to the mitochondrial genome were removed. Many samples achieved a high amount of total reads, but lost many reads to mitochondria, leaving them with 30%-70% useable reads (Figure 5.2). This was found to be an inherent problem with traditional ATAC-Seq and Fast-ATAC because mitochondrial DNA is nucleosome free and readily accessible to the transposase. Samples processed using Omni-ATAC lost far fewer reads to the mitochondrial genome because of the use of dual detergents and washes in the protocol (Corces et al. 2017). Although historic samples were processed using Omni-ATAC, they also lost around 40%-50% of reads to mitochondria. This could be because the formaldehyde treatment made it much harder for mitochondrial DNA to be separated and removed during the protocol.

MACS2 callpeak (Zhang et al. 2008) was used for peak calling (methodology detailed in Chapter 2), based on the regions with the highest amount of reads, owing to transposase integration. As mentioned in Chapter 4, signal to noise is measured by plotting the enrichment of ATAC-Seq reads around the TSS. The

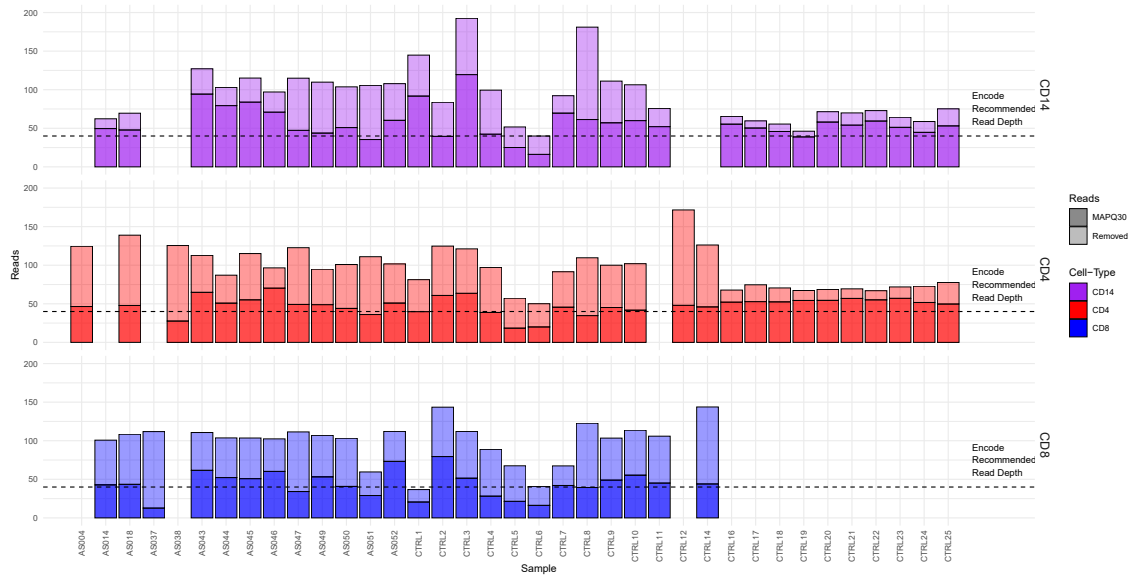


Figure 5.1: Total reads following ATAC on AS patients and controls in CD4$^+$, CD8$^+$, and CD14$^+$ cells.
MAPQ30 reads represent usable, high quality reads. ENCODE recommends 50 million reads per sample.

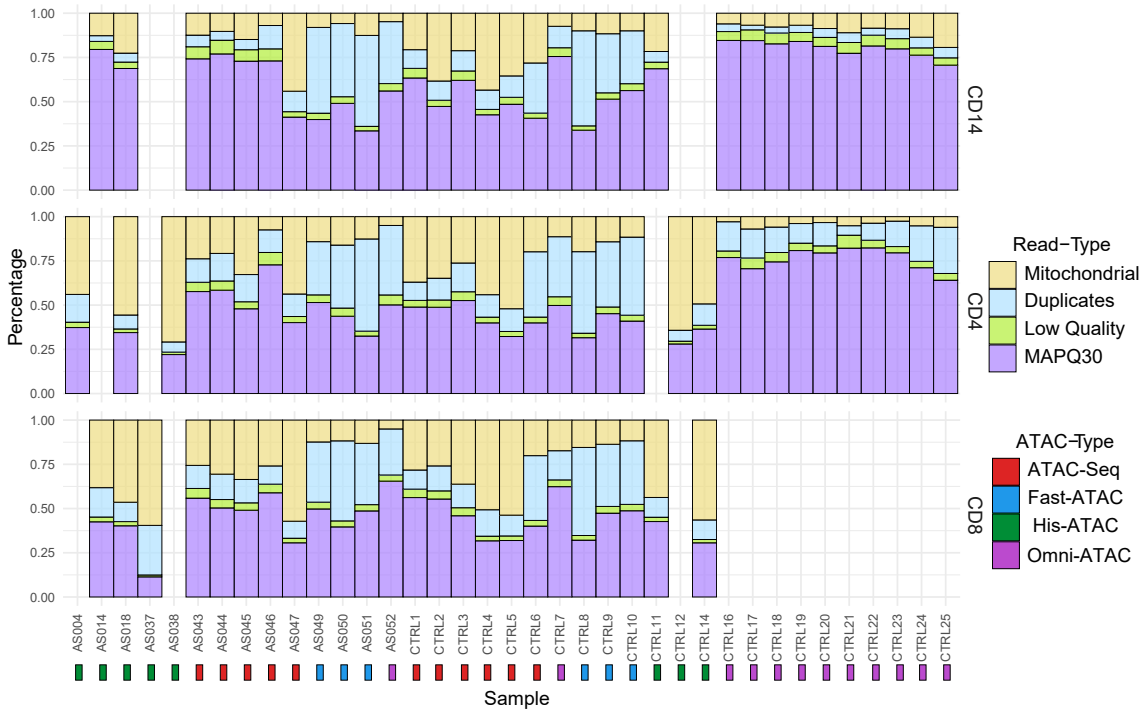


Figure 5.2: Read percentages following ATAC on AS patients and controls in CD4$^+$, CD8$^+$, and CD14$^+$ cells.
MAPQ30 reads represent usable, high quality reads. Reads are lost due to low quality, PCR duplicates, and mapping to the mitochondrial genome. Many samples lost reads to the mitochondrial genome because of the ATAC-Seq protocol.

TSS plot for AS patients and controls across $CD4^+$, $CD8^+$, and $CD14^+$ cells can be seen in Figure 5.3. Samples processed using Omni-ATAC had the highest TSS enrichment. ENCODE recommends samples having a TSS fold enrichment of >5, which only five samples did not reach. DESeq2 (Love et al. 2014) was used for differential analysis to determine how peaks changed between AS patients and controls across the three cell types.

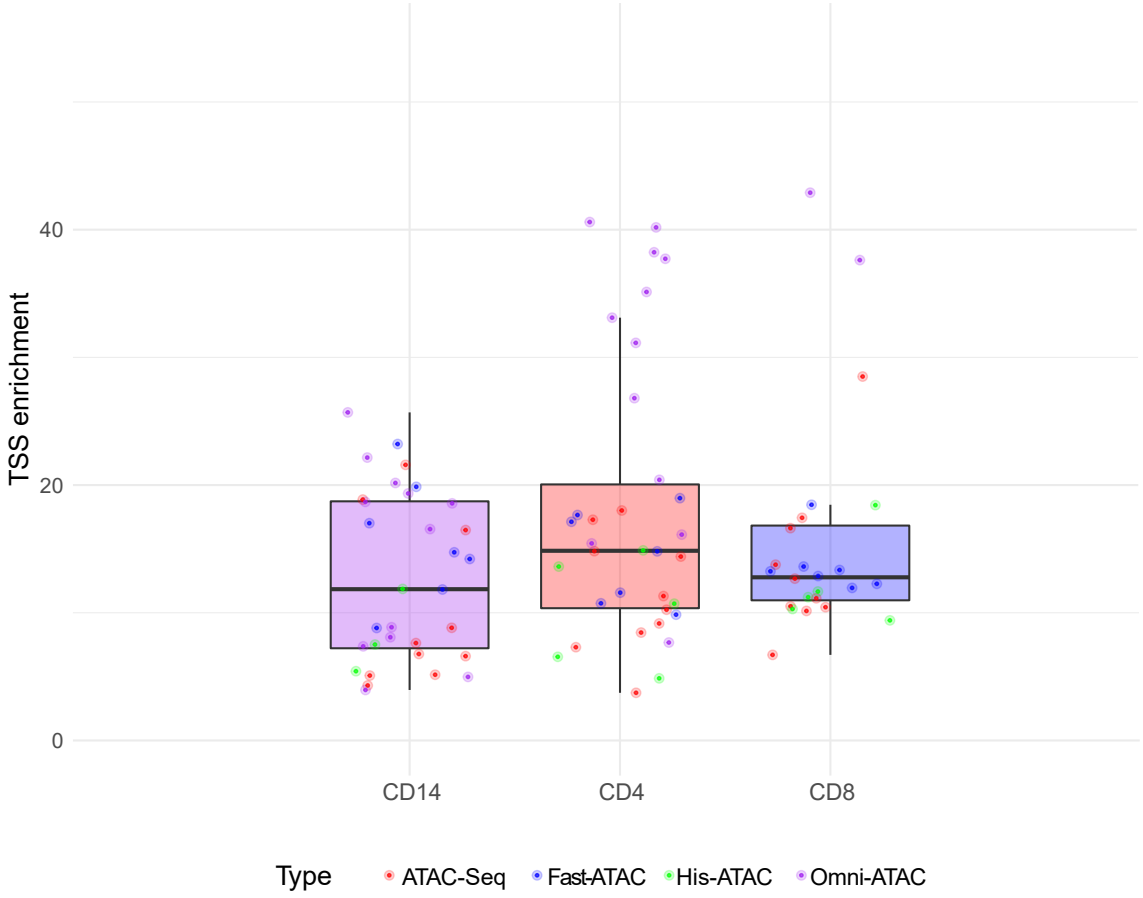


Figure 5.3: ATAC TSS fold enrichment for AS patient samples and controls in $CD4^+$, $CD8^+$, and $CD14^+$ cell types.
ENCODE recommends samples having a TSS fold enrichment of >5. His = Historic.

ATAC-Seq experiments are typically carried out on fresh cells and not on ones that have been fixed or frozen. The Historic ATAC samples here underwent formaldehyde fixation and were stored in the freezer for approximately five years prior to processing with Omni-ATAC. All other samples were prepared fresh on the same day that blood was harvested from a patient or control. Before including

them in this thesis, their quality was validated and compared to fresh samples. When comparing read percentages following sequencing, the historic samples lost many more reads to mitochondrial DNA than fresh Omni-ATAC samples (Figure 5.4A). Samples processed using traditional ATAC-Seq or Fast-ATAC had this same problem due to the nature of the protocol. Although losing half of the sequencing to mitochondrial DNA is not ideal because it increases the sequencing costs to reach the desired number of reads, it does not effect the quality of the reads. The TSS of the historic samples reached the ENCODE recommended threshold and was comparable to the fresh ATAC samples (Figure 5.4B).

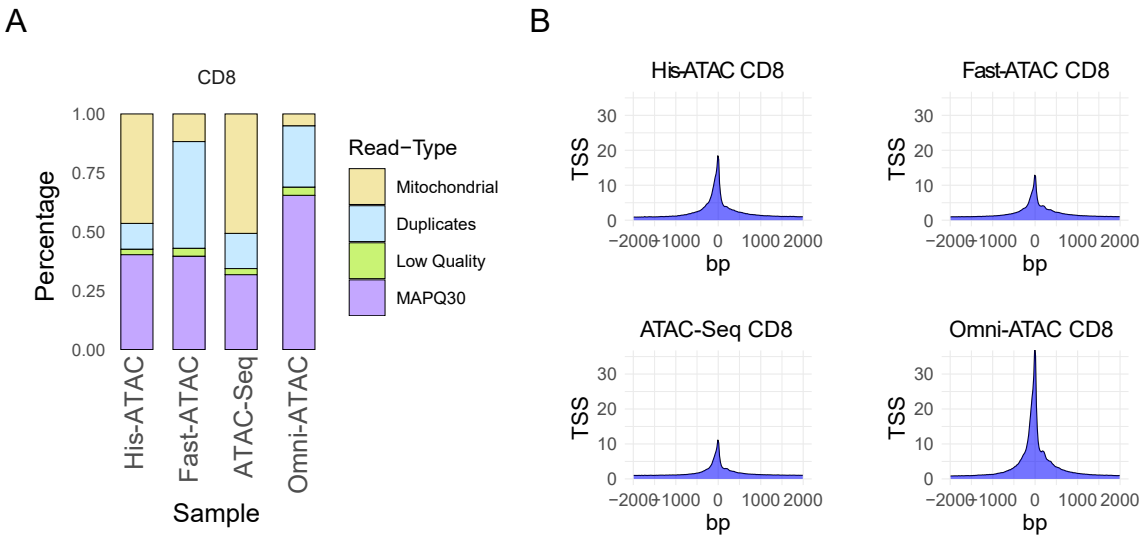


Figure 5.4: Comparison between fresh and frozen ATAC samples.
A. Read percentages of Historic ATAC (formaldehyde fixed and frozen), Fast-ATAC (fresh), ATAC-Seq (fresh), and Omni-ATAC (fresh) in $CD8^+$ cells. **B**. TSS enrichment of Historic ATAC (formaldehyde fixed and frozen), Fast-ATAC (fresh), ATAC-Seq (fresh), and Omni-ATAC (fresh) in $CD8^+$ cells.

Because three different ATAC protocols were used in four batches, a batch effect was seen following principal component analysis (PCA) of variance in the data (Figure 5.5). The samples were found to cluster based on ATAC protocol. Before the differential analysis could be performed, the batch effect was corrected using DESeq2. DESeq2 batch correction was visualized on PCA by Limma (Figure 5.6), demonstrating that the batch correction was successful.

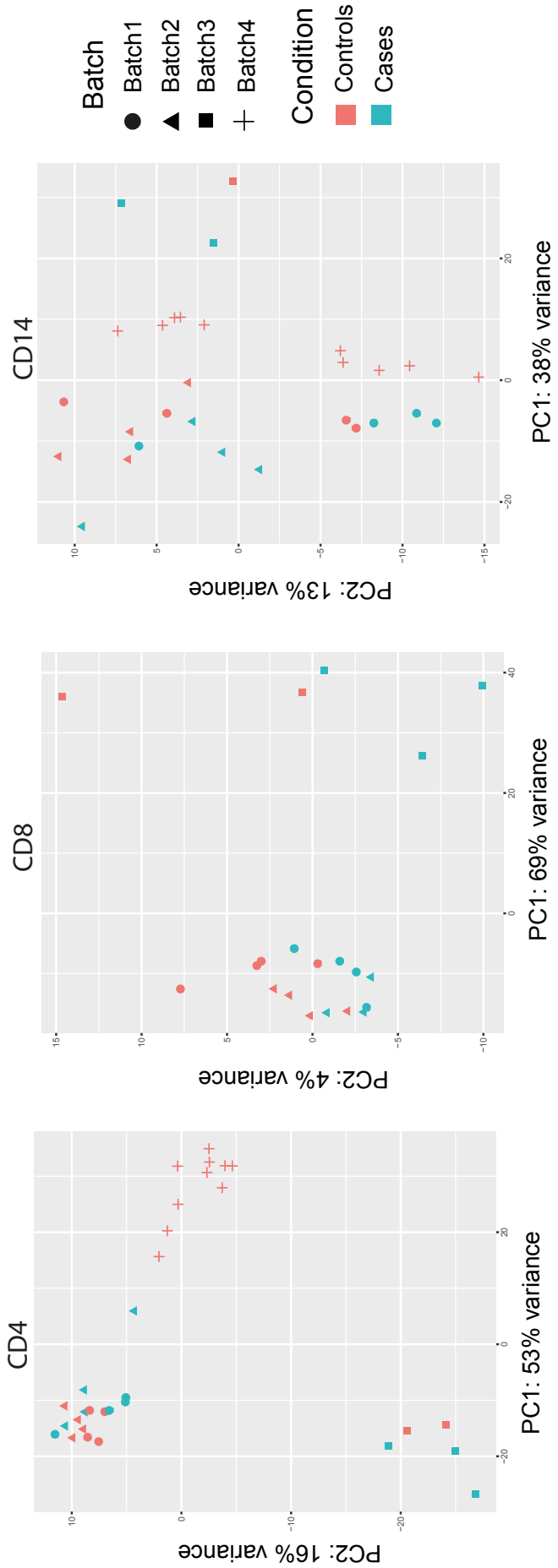


Figure 5.5: Principal component analysis of ATAC batches by cell type.
Non-corrected PCA plots were generated to show the variation between each ATAC batch within $CD4^+$, $CD8^+$, and $CD14^+$ cells. Principal component 1 and 2 are shown. Batch 1 consists of samples processed using traditional ATAC-Seq. Batch 2 consists of samples processed using both Fast-ATAC and Omni-ATAC. Batch 3 consists of the historic samples processed using Omni-ATAC on formaldehyde fixed cells. Batch 4 consists of samples processed using Omni-ATAC. This analysis shows a batch effect.

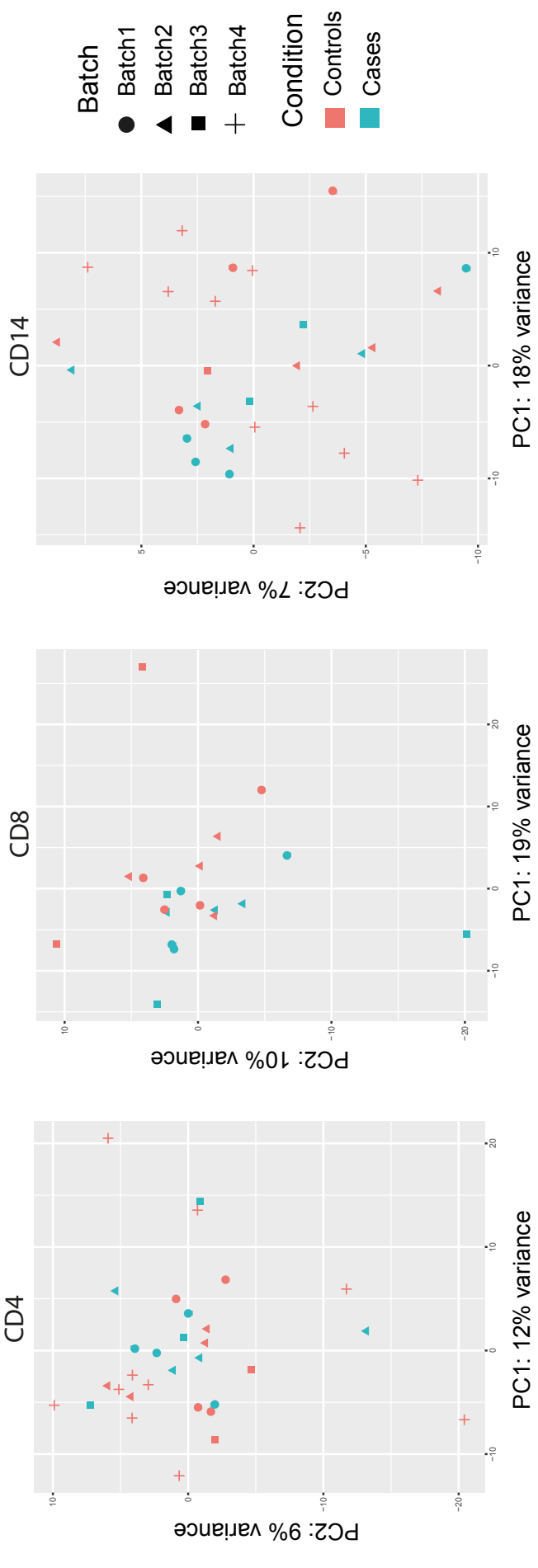


Figure 5.6: Corrected principal component analysis of ATAC batches by cell type. Batch and sex corrected PCA plots were generated to show the variation between each ATAC batch within $CD4^+$, $CD8^+$, and $CD14^+$ cells. Principal component 1 and 2 are shown. Batch 1 consists of samples processed using traditional ATAC-Seq. Batch 2 consists of samples processed using both Fast-ATAC and Omni-ATAC. Batch 3 consists of the historic samples processed using Omni-ATAC on formaldehyde fixed cells. Batch 4 consists of samples processed using Omni-ATAC. Sample variance is reduced after correction.

Cell Type Specific Regions of Chromatin Accessibility

Different cell types have unique roles and functions and as such, each cell type has a unique chromatin accessibility landscape. Before comparing regions of chromatin accessibility between patient samples and controls, it is important to validate the method by visualizing expected changes as a proof of principle. Therefore, a chromatin accessibility differential analysis was carried out, comparing between three different cell types, $CD4^+$, $CD8^+$, and $CD14^+$. For this analysis, an ANOVA was run comparing all three cell types, followed by a pairwise comparison comparing each cell pairing. A peak had to have an adjusted p-value of <0.01 in both analyses to be considered significant for that cell type.

Figure 5.7 illustrates the peak distribution, representing regions of open chromatin between $CD4^+$, $CD8^+$, and $CD14^+$ cell types. AS patient samples and controls were combined for this analysis. The largest number of shared peaks were between $CD4^+$ and $CD8^+$ T cells, with $CD14^+$ monocytes having the highest number of unique peaks. This is unsurprising as $CD4^+$ and $CD8^+$ T cells are a similar cell type and are expected to share a large proportion of accessible chromatin regions.

Figure 5.8 illustrates the number of differential peaks from all 58,096 called peaks in the masterlist following the pairwise analysis between each cell type. The highest number of differential peaks were between $CD14^+$ monocytes and $CD4^+$ T cells, and $CD14^+$ monocytes and $CD8^+$ T cells.

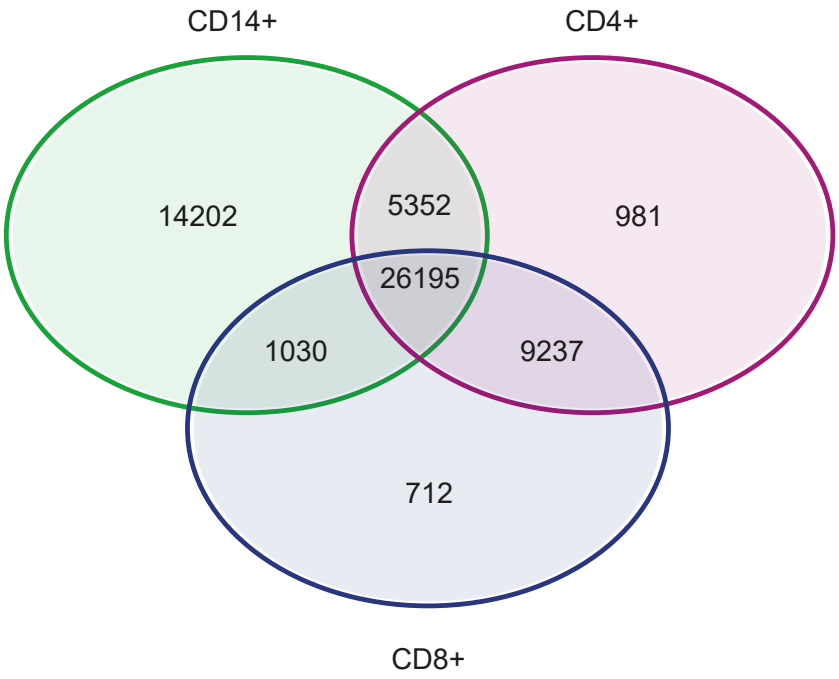


Figure 5.7: ATAC-Seq peak cell type distribution.
Venn diagram showing number of peaks (regions of open chromatin) and their distribution between $CD4^+$, $CD8^+$, and $CD14^+$ cell types. ATAC-seq data from 14 AS patients and 23 healthy controls.

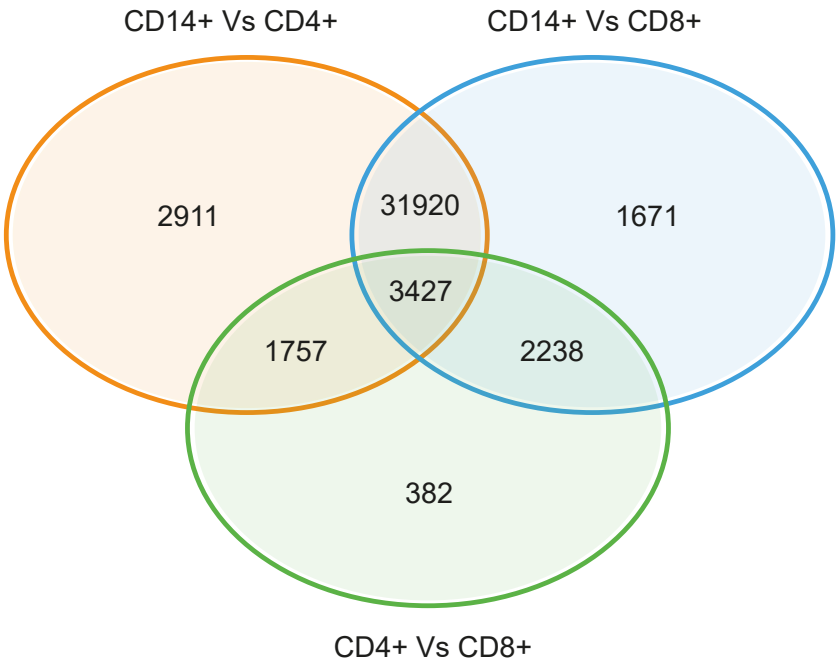


Figure 5.8: ATAC-Seq differential analysis between cell types.
Venn diagram summarizing the number of differential ATAC-Seq peaks (regions of open chromatin) between $CD14^+$ monocytes, $CD4^+$ T cells, and $CD8^+$ T cells from all all 58,096 called peaks in the masterlist. Differential analysis was performed using a pairwise comparison between each cell type.

Pathway Enrichment Analysis on Cell Type Specific Peaks

To validate the cell type specific peaks as being relevant to the cells being examined, pathway enrichment analysis was performed on the genes overlapping $CD4^+$, $CD8^+$, and $CD14^+$ cell type specific ATAC peaks using XGR (FDR<0.05) (Fang et al. 2016) (Figure 5.9). Peaks were considered to be associated with a gene if they were within 5kb of the start or end of the gene. The gene sets used for the analysis were from Canonical, KEGG, REACTOME and BioCarta pathways (Subramanian et al. 2005). Genes associated with ATAC peaks from all three cell types were used as the background set of genes. The top 25 peaks for each cell type, as determined by their global p-value (p-value provided by the ANOVA, FDR <0.01), are listed in Tables C.3, C.4, and C.5 in Appendix C.

In total there were 475 $CD4^+$ cell type specific peaks which were associated with, based on proximity (within 5kb), 280 genes. The most significantly enriched pathways were 'TCR signaling in naive $CD8^+$ T cells' and 'TCR signaling in naive $CD4^+$ T cells'. The genes overlapping cell type specific ATAC peaks involved in these pathways were *CBL*, *CD247*, *CD28*, *FYN*, *IKBKB*, *MALT1*, *PRKCB*, and *RAP1A*. Some of the other top enriched pathways included, 'TAK1 activation of NF-κB', 'Tob in T-cell activation', 'T cell receptor signaling pathway', 'canonical NF-κB pathway', and the 'immune system'.

There were 690 $CD8^+$ cell type specific peaks associated with 390 genes. The most significantly enriched pathway was 'signaling by Rho GTPases'. Rho GTPases are involved in signal transduction in a variety of biological pathways, including the immune system. The genes with cell type specific open chromatin included in this pathway were, *ABR*, *ARAP2*, *ARHGAP10*, *ARHGAP15*, *ARHGAP18*, *ARHGAP25*, *ARHGAP26*, *ARHGAP44*, *ARHGEF4*, *CHN2*, *PREX1*, *VAV3*. Various NOTCH signaling pathways were also highly enriched.

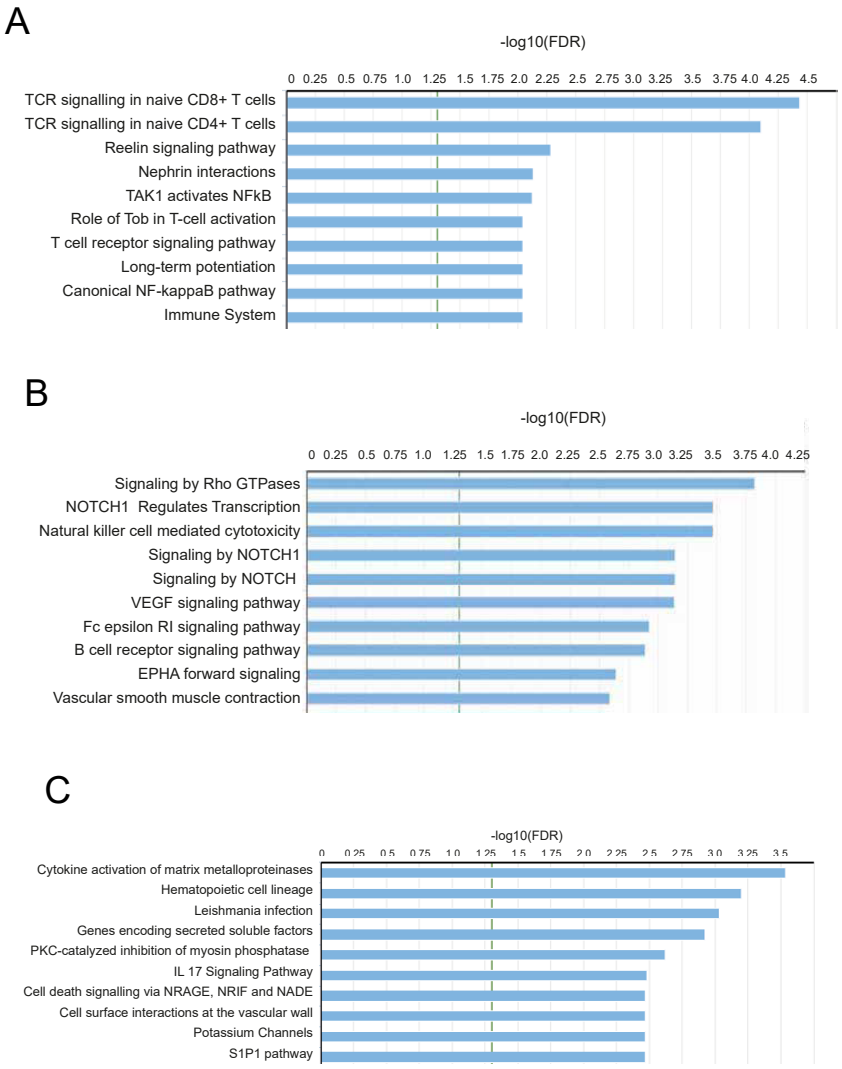


Figure 5.9: Pathway enrichment analysis on genes overlapping cell type specific ATAC peaks.
Top 10 pathways from the enrichment enrichment analysis performed on genes overlapping cell type specific ATAC peaks using XGR. Green line represents FDR<0.05. **A**. $CD4^+$ cells. **B**. $CD8^+$ cells. **C**. $CD14^+$ cells.

$CD14^+$ cells had 13928 specific peaks which were associated with 4446 genes. Pathway analysis was performed on the top 300 genes. The most significantly enriched pathway was 'cytokine activation of matrix metalloproteinases' which included *CSF1*, *FCGR3A*, *IL6R*, and *TNFRSF1B*. Another highly enriched pathway was 'hematopoietic cell lineage', involving *CD1A*, *CD34*, *CR1*, *CSF1*, and *IL6R*. Other highly enriched pathways included, 'genes encoding secreted soluble factors', 'IL-17 signaling pathway', and 'cell death signaling'.

GWAS Overlap with Cell Type Specific Peaks

Understanding the chromatin landscape in specific immune cells and their link to disease associated regions may help uncover new mechanisms relating to disease pathology and help identify disease relevant cell types. Therefore, peaks associated with cell type specific open chromatin were analyzed for overlap with AS GWAS loci. GWAS loci locations were determined by using the 95% credible sets following fine mapping by Félicie Costantino. This was performed on $CD4^+$, $CD8^+$, and $CD14^+$ specific ATAC peaks. The results of this analysis are listed in Table 5.2.

There were two $CD4^+$ cell type specific ATAC peaks overlapping AS GWAS regions. Both peaks are within an intron of *MAP4K4*, which is in LD with the lead GWAS SNP, rs4851529, associated with *IL1R1*. One SNP, rs2192752, in LD with the lead SNP at this locus is an eQTL for *IL1R1* in LPS stimulated $CD14^+$ cells (Fairfax et al. 2014), but not known to be in $CD4^+$ cells. Neither peak is over a predicted enhancer based on data from NIH Roadmap (Bernstein et al. 2010) and FANTOM (Andersson et al. 2014).

Five $CD8^+$ cell type specific ATAC peaks overlapped four different AS GWAS regions. None of the peaks are over coding regions or gene promoters, suggesting they represent regulatory elements. One peak, chr12:111942729-111944397, is within an intron of *ATXN2* and is predicted to be an enhancer for *ATXN2* by NIH Roadmap (Bernstein et al. 2010) (Figure 5.10). It overlaps the 95% credible set associated with rs11065898 (*SH2B3*). This locus is not found within any Hi-C datasets (Schofield et al. 2016), so its target gene is only predicted by proximity. NIH Roadmap also predicts that it could be an enhancer for *BRAP* or *SH2B3*, which are the genes immediately upstream and downstream of *ATXN2*. This is the only peak within a gene, with all others located in intergenic regions. None of the other peaks overlapping GWAS loci are predicted to be enhancers, have Hi-C interactions, or are associated with eQTLs using publicly available data.

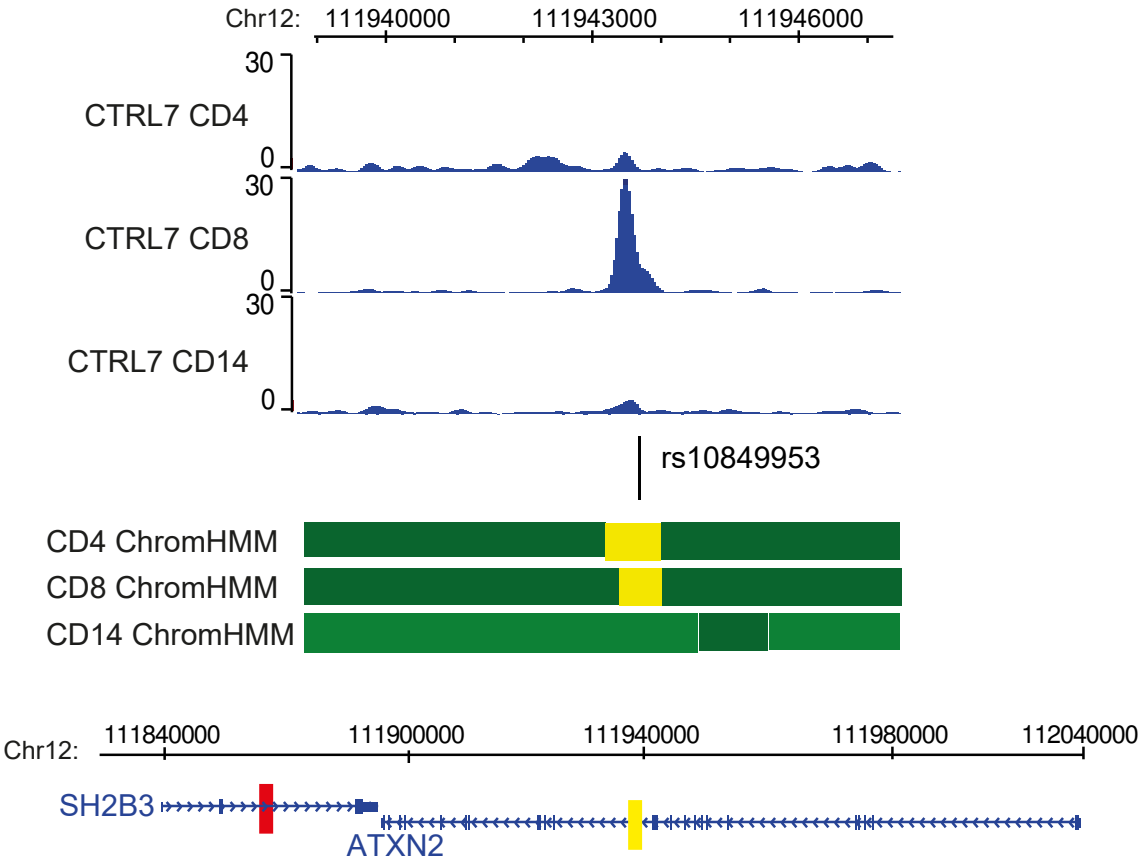


Figure 5.10: CD8$^+$ cell type specific peak within *ATXN2*.
WashU Epigenome Browser view of the ATAC-seq peak, chr12:111942729-111944397, in CD4$^+$, CD8$^+$, and CD14$^+$ cells. This peaks overlaps enhancer predictions based on histone marks in CD4$^+$ and CD8$^+$ cells, represented by the yellow box in the ChromHMM tracks. A full view of *ATXN2* and *SH2B3* are shown, with the peak region highlighted in yellow. The lead GWAS SNP for this locus within *SH2B3* (rs11065898) is highlighted in red.

A total of 51 CD14$^+$ cell type specific ATAC peaks overlapped 13 different AS GWAS loci. The peaks at chr1:161474662-161476522, chr16:28620456-28622023, chr2:102607544-102608971, and chr9:139266583-139268379 are located over the promoters for *FCGR2A*, *SULT1A1*, *IL1R2*, and *CARD9*, respectively. This suggests that these genes may be differentially expressed in CD14$^+$ cells. SNPs in this region have not been found as eQTLs. Peak chr1:154426850-154427915 is very near the lead GWAS SNP, rs4129267, for the *IL6R* locus and is within an intron *IL6R*. In-house Capture-C data has found an interaction occurring between the intronic region of this SNP and the promoter for *IL6R*, but only in CD14$^+$ cells (Figure 5.11). The presence of a cell type specific

ATAC-Seq peak near this SNP suggests this may represent a CD14$^+$ specific regulatory element modulating the expression of *IL6R*. Moreover, rs4129267 was found to be a an eQTL in peripheral blood regulating the expression of *IL6R* (Revez et al. 2013; Westra et al. 2013).

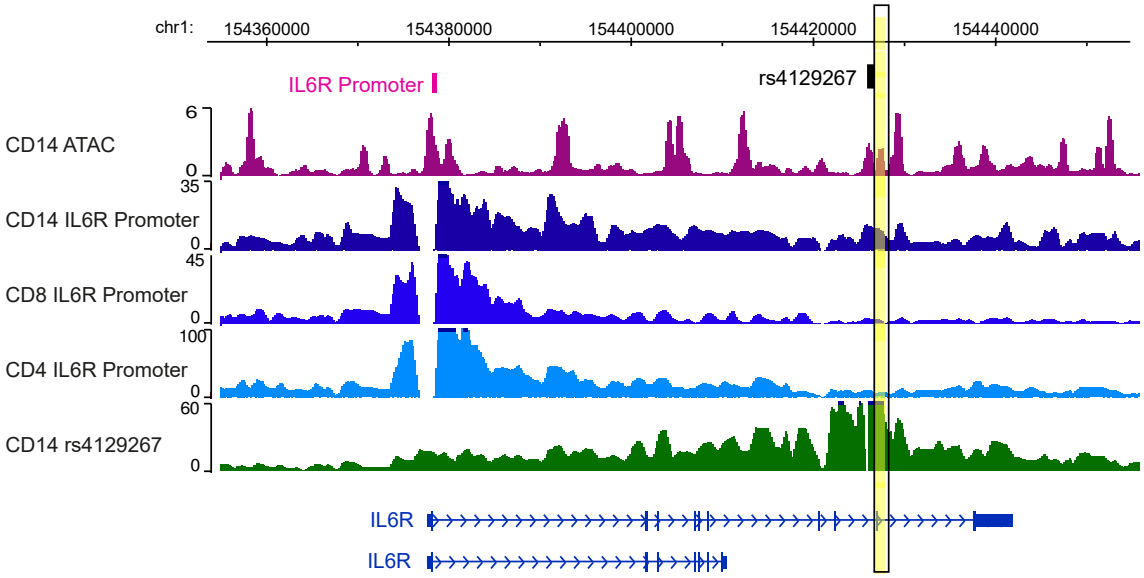


Figure 5.11: Capture-C and ATAC-Seq data for the *IL6R* AS GWAS locus.
WashU Epigenome Browser view of the ATAC-seq peak, chr1:154426850-154427915 (highlighted in yellow) in CD14$^+$ cells (top track). Capture-C data for CD4$^+$, CD8$^+$, and CD14$^+$ cells are shown (lower four tracks). Tracks listed as promoter denote the promoter Capture-C viewpoint for that gene, while the rs4129267 track denotes the Capture-C baited viewpoint near that SNP. The *IL6R* promoter and lead SNP are indicated. The Capture-C data illustrate an interaction between the *IL6R* promoter and region around rs4129267 only in CD14$^+$ cells.

Table 5.2: Cell type specific ATAC-Seq peaks overlapping AS GWAS loci.
The coordinates stated are in UCSC notation based on genome build hg19. GWAS coordinates represent the 95% credible set.

ATAC-Seq Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	CD4 vs CD8 p-value	Proximal Gene	GWAS Loci
$CD4^+$ Specific Peaks						
chr2:102345508-102347221	$2.50x10^{-17}$	$1.20x10^{-18}$	1	0.009	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102384524-102385878	$1.50x10^{-19}$	$3.70x10^{-20}$	1	0.0022	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
$CD8^+$ Specific Peaks						
chr12:111942729-111944397	$1.00x10^{-18}$	1	$1.30x10^{-14}$	$1.10x10^{-11}$	*ATXN2*	rs11065898 (SH2B3) chr12:111844217-112208838
chr19:10633395-10634593	$1.50x10^{-33}$	1	$5.30x10^{-29}$	$3.90x10^{-23}$	*S1PR5*	rs3516406 (TYK2) chr19:10496620-10771941
chr2:102597464-102598736	$3.60x10^{-37}$	1	$5.50x10^{-24}$	$6.50x10^{-34}$	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102869220-102870717	$2.60x10^{-34}$	0.075	$6.40x10^{-32}$	$4.50x10^{-13}$	*IL1RL2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:182010364-182011802	$1.10x10^{-83}$	0.099	$7.20x10^{-72}$	$1.20x10^{-44}$	NA	rs12615545 (UBE2E3) chr2:182007799-182095265
$CD14^+$ Specific Peaks						
chr1:154426850-154427915	$2.1x10^{-62}$	$9x10^{-42}$	$1.4x10^{-38}$	1	*IL6R*	rs4129267 (IL6R) chr1:154417828-154507888

chr1:154428515-154429836	$7.2x10^{-110}$	$1.8x10^{-71}$	$2x10^{-75}$	1	*IL6R*	rs4129267 (IL6R) chr1:154417828-154507888
chr1:154462015-154463455	$5.3x10^{-43}$	$3.6x10^{-32}$	$4x10^{-27}$	1	*SHE*	rs4129267 (IL6R) chr1:154417828-154507888
chr1:161468918-161470241	$3x10^{-43}$	$1.4x10^{-33}$	$1x10^{-23}$	1	*FCGR2A*	rs1801274 (FCGR2A) chr1:161469053-161479745
chr1:161474662-161476522	$8.8x10^{-150}$	$4.4x10^{-99}$	$2.3x10^{-100}$	1	*FCGR2A*	rs1801274 (FCGR2A) chr1:161469053-161479745
chr1:200876294-200877695	$2.2x10^{-25}$	$1.9x10^{-20}$	$2.5x10^{-14}$	1	*C1orf106*	rs41299637 (GPR25) chr1:200874228-200881595
chr10:81033406-81034828	$7x10^{-56}$	$6.9x10^{-39}$	$1.9x10^{-37}$	1	*ZMIZ1*	rs1250550 (ZMIZ1) chr10:81032531-81066982
chr10:81054966-81056066	$1.1x10^{-97}$	$1.1x10^{-66}$	$2x10^{-63}$	1	*ZMIZ1*	rs1250550 (ZMIZ1) chr10:81032531-81066982
chr10:81058137-81059566	$1.1x10^{-42}$	$8.7x10^{-29}$	$9x10^{-30}$	1	*ZMIZ1*	rs1250550 (ZMIZ1) chr10:81032531-81066982
chr12:111899456-111900873	$1.4x10^{-58}$	$2.7x10^{-42}$	$3.9x10^{-36}$	1	*ATXN2*	rs11065898 (SH2B3) chr12:111844217-112208838
chr12:112194958-112196231	$7.8x10^{-36}$	$8.7x10^{-29}$	$2.6x10^{-21}$	1	*ACAD10*	rs11065898 (SH2B3) chr12:111844217-112208838
chr14:88465294-88466486	$8.8x10^{-126}$	$2.2x10^{-90}$	$6.3x10^{-75}$	1	*GPR65*	rs11624293 (GPR65) chr14:88426296-88488821
chr14:88468434-88469624	$1.5x10^{-52}$	$1.6x10^{-47}$	$2.4x10^{-19}$	1	*GPR65*	rs11624293 (GPR65) chr14:88426296-88488821

chr16:28535251-28536773	$6.7x10^{-54}$	$5.5x10^{-34}$	$2.2x10^{-39}$	1	*SULT1A1*	rs26528 (IL27) chr16:28483060-28895130
chr16:28620456-28622023	$1.9x10^{-48}$	$5.3x10^{-36}$	$3.5x10^{-33}$	1	*SULT1A1*	rs26528 (IL27) chr16:28483060-28895130
chr19:10593656-10595083	$9.33x10^{-8}$	$2.2x10^{-5}$	$7.6x10^{-5}$	1	*KEAP1*	rs35164067 (TYK2) chr19:10496620-10771941
chr19:10745481-10746955	$4.3x10^{-34}$	$3.9x10^{-24}$	$1.1x10^{-24}$	1	*SLC44A2*	rs35164067 (TYK2) chr19:10496620-10771941
chr19:10750167-10751203	$2.2x10^{-25}$	$1.8x10^{-20}$	$9.9x10^{-15}$	1	*SLC44A2*	rs35164067 (TYK2) chr19:10496620-10771941
chr2:102329827-102331340	$2.6x10^{-62}$	$2.2x10^{-43}$	$9.9x10^{-39}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102338942-102340471	$4.1x10^{-105}$	$2.9x10^{-70}$	$7.9x10^{-63}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102352647-102354069	$7.8x10^{-35}$	$1.9x10^{-23}$	$1.2x10^{-22}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102356748-102357966	$1.4x10^{-51}$	$2.5x10^{-36}$	$1.3x10^{-29}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102397126-102398693	$4.1x10^{-119}$	$3x10^{-82}$	$5.9x10^{-68}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102528418-102530037	$1.7x10^{-86}$	$3.6x10^{-69}$	$6.2x10^{-43}$	1	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102573991-102575217	$7.1x10^{-82}$	$1.5x10^{-57}$	$8.4x10^{-53}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491

chr2:102577286-102578946	$1.5x10^{-142}$	$1.2x10^{-105}$	$3.1x10^{-99}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102588920-102590218	$2.4x10^{-85}$	$2x10^{-62}$	$3.6x10^{-52}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102607544-102608971	$3.6x10^{-100}$	$2.9x10^{-68}$	$5.4x10^{-65}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102626905-102628627	$1x10^{-62}$	$1.9x10^{-40}$	$1.4x10^{-46}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102641571-102643003	$1.5x10^{-97}$	$4.4x10^{-65}$	$1.6x10^{-66}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102657135-102658952	$2.3x10^{-107}$	$1.5x10^{-74}$	$4.6x10^{-69}$	1	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102668779-102670297	$3.2x10^{-72}$	$3.7x10^{-56}$	$6.1x10^{-44}$	1	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102679654-102681155	$2.9x10^{-147}$	$1.5x10^{-93}$	$2.2x10^{-110}$	0.026	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102690092-102691572	$3.6x10^{-55}$	$4.1x10^{-38}$	$2.7x10^{-37}$	1	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102710768-102712332	$5.3x10^{-41}$	$4.5x10^{-29}$	$1.6x10^{-27}$	1	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102758501-102760110	$5.1x10^{-105}$	$3.1x10^{-69}$	$6.7x10^{-71}$	1	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102771614-102773381	$1.3x10^{-51}$	$1.7x10^{-37}$	$2.5x10^{-32}$	1	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491

chr2:102889956-102891331	$3.6x10^{-51}$	$3.2x10^{-38}$	$5x10^{-31}$	1	*IL1RL2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102894798-102896043	$2x10^{-46}$	$3.9x10^{-36}$	$3.5x10^{-26}$	1	*IL18R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102898767-102900243	$2.3x10^{-55}$	$4.9x10^{-38}$	$2x10^{-37}$	1	*IL18R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102907466-102908809	$2.4x10^{-72}$	$3.9x10^{-51}$	$6.3x10^{-44}$	1	*IL18R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102911267-102912726	$2.3x10^{-31}$	$1.2x10^{-23}$	$5.3x10^{-20}$	1	*IL18R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102973937-102975259	$1.53x10^{-5}$	$3.5x10^{-3}$	$4.1x10^{-4}$	1	*IL18R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:103124983-103126906	$3.6x10^{-128}$	$3.3x10^{-88}$	$3.6x10^{-74}$	1	*SLC9A4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr21:40466230-40467317	$9.3x10^{-109}$	$1.2x10^{-66}$	$2.1x10^{-81}$	0.067	NA	rs9977672 (21q22) chr21:40463282-40466744
chr5:158770477-158771663	$1.5x10^{-151}$	$3.1x10^{-102}$	$2.7x10^{-100}$	1	*IL12B*	rs6871626 (IL12B) chr5:158767332-158818745
chr5:158772037-158773086	$3.9x10^{-78}$	$1.1x10^{-53}$	$7.8x10^{-51}$	1	*IL12B*	rs6871626 (IL12B) chr5:158767332-158818745
chr5:158786087-158787328	$5.2x10^{-86}$	$1.1x10^{-59}$	$1.2x10^{-51}$	1	*IL12B*	rs6871626 (IL12B) chr5:158767332-158818745
chr5:158788579-158790010	$3.3x10^{-134}$	$8.4x10^{-90}$	$6.7x10^{-87}$	1	*IL12B*	rs6871626 (IL12B) chr5:158767332-158818745

chr5:96126187-96127743	$3.6x10^{-51}$	$8.9x10^{-39}$	$1.2x10^{-31}$	1	*ERAP1*	rs30187 (ERAP1) chr5:96121151-96126308
chr9:139266583-139268379	$8.5x10^{-116}$	$3.8x10^{-80}$	$1.1x10^{-81}$	1	*CARD9*	rs1128905 (CARD9) chr9:139240629-139301583

Differential Chromatin Accessibility Between AS Patients and Controls

To better understand the chromatin landscape associated with disease, a differential ATAC analysis was carried out between AS patients and controls in $CD4^+$, $CD8^+$, and $CD14^+$ cells. The presence of differential chromatin between AS patients and healthy controls indicates disease-specific differences in gene regulation. Differential open chromatin regions that overlap with GWAS SNPs provide evidence for them having a functional role in inflammation or immunity. Furthermore, overlapping of these regions with differential histone marks or expressed genes may reveal disease relevant regulatory regions. The differential analysis was carried out using DESeq2, correcting for batch, with an FDR of <0.05 being considered significant. Following the differential analysis, there were no significant differential ATAC peaks in $CD4^+$ cells, however a number of differential ATAC peaks were found in $CD8^+$ T cells and $CD14^+$ monocytes (Table 5.3).

ATAC Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
$CD8^+$ Differential Peaks			
chr3:111430661-111431969	*PLCXD2*	-2.10	$6.43x10^{-4}$
chr1:239664289-239666134	*CHRM3*	0.935	$8.65x10^{-3}$
chr21:46339198-46341399	*ITGB2*	-0.569	0.0121
chr3:56766115-56767492	*ARHGEF3*	-0.759	0.0181
chr1:179783641-179785018	*FAM163A*	0.483	0.0334
chr6:16510956-16513174	*ATXN1*	0.591	0.0334
chr8:131315920-131317608	*ASAP1*	-0.728	0.0370
$CD14^+$ Differential Peaks			
chr19:20348445-20349819	*ZNF486*	0.849	0.0249
chr9:133004993-133006146	*NCS1*	0.981	0.0495

Table 5.3: Regions of differential open chromatin between AS patients and controls in $CD8^+$ and $CD14^+$ cells.
FDR of <0.05 (DESeq2) was considered differential. The coordinates stated are in UCSC notation based on genome build hg19.

Following the differential analysis between patients and controls in $CD8^+$ cells, there were seven differential peaks. The most differential peak is located in an intron of *PLCXD2* (Figure 5.12). *PLCXD2* is not associated with immunity or inflammation, but has been associated with cancer (Gameiro et al. 2012). It is a phospholipase responsible for the hydrolysis of phospholipids into fatty acids and other molecules and is involved in signal transduction. This region of open chromatin could represent a regulatory element acting on *PLCXD2*, or it could regulate a different gene farther away. It is not situated over a predicted enhancer based on data from NIH Roadmap (Bernstein et al. 2010) and FANTOM (Andersson et al. 2014), and is not shown to interact with *PLCXD2* or other genes in Hi-C data (Schofield et al. 2016). However, the GeneHancer database predicts this region to interact with the TSS of *PLCXD2* (Fishilevich et al. 2017). Another differential peak is situated over the promoter for *ITGB2*, meaning that this gene may be regulated or expressed differentially in AS patients, however it was found to be when examining in-house RNA-Seq data. ITGB2 is an integrin which is a cell surface protein involved in cell signaling and adhesion (Evans et al. 2009) and genetic variants within this gene have been associated with leukocyte adhesion deficiency type 1 (Parvaneh et al. 2010; Schmidt et al. 2013). SNPs within this peak have also been identified as eQTLs for *ITGB2* in $CD4^+$ and $CD8^+$ cells (Jansen et al. 2017; Kasela et al. 2017). Peak chr3:56766115-56767492, within an intron of *ARHGEF3* is predicted to overlap an enhancer for *FAM208A* in $CD8^+$ cells by NIH Roadmap (Bernstein et al. 2010). *FAM208A* and *ARHGEF3* are situated next to each other. None of the differential peaks overlap AS GWAS regions.

There were two differential ATAC peaks indicating open chromatin in $CD14^+$ cells between patients and controls. Neither significant peak is within a gene, suggesting they represent regulatory elements. Peak chr19:20348445-20349819 is predicted by NIH Roadmap (Bernstein et al. 2010) to overlap an enhancer for

*ZNF486*in CD14$^+$ cells, but does not coincide with an eQTL. The other peak is near *NCS1*. NCS1 is a Neuronal calcium sensor which has not been associated with autoimmune disease. Neither peak has been associated to a target with Hi-C data (Schofield et al. 2016), nor do they overlap AS GWAS regions.

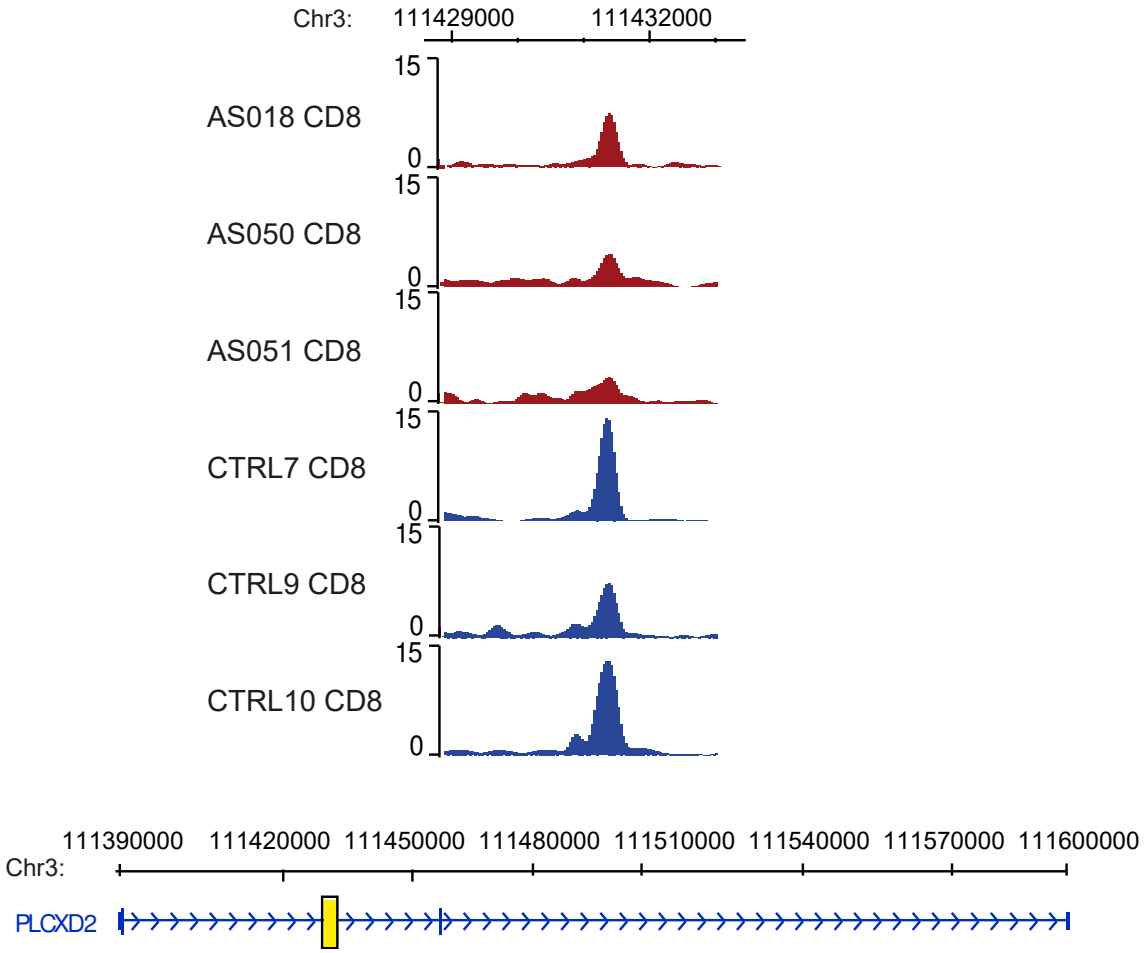


Figure 5.12: Peaks showing one region of differential open chromatin between AS patients and controls in CD8$^+$ cells.
ATAC-Seq peak chr3:111430661-111431969 within an intron of *PLCXD2* (Log_2 fold Change -2.10, FDR 6.43x10^{-4}) in three representative AS patient samples and three healthy control samples. The Log_2 fold Change is based on an average of all the AS samples. A full view of *PLCXD2* is shown, with the peak region highlighted in yellow.

5.2.3 The Histone Modification Landscape in AS

H3K27ac and H3K4me3 ChIPm Quality Control

A total of 10 AS patients and 9 healthy controls were included in the H3K27ac and H3K4me3 ChIP-Seq analysis. All samples were processed using the ChIPm protocol and were formaldehyde fixed and frozen prior to processing. A subset of the samples were processed by Andrew Brown and Connor Short. All samples were sequenced and run through the in-house pipeline as described in Chapter 2. ENCODE recommends 40 million reads per sample (20 million fragments) for ChIP-Seq experiments and all but three samples and six input controls reached this threshold for H3K27ac and H3K4me3 (Figure 5.13). The total reads were filtered, with some removed due to low quality, PCR duplicates, and mapping to the mitochondrial genome. A total of 22 input controls and one sample had less than 75% usable reads due to low quality reads and PCR duplicates, possibly from over amplification (Figure 5.14). MACS2 was used for peak calling (described in Chapter 2). ENCODE does not have a recommended TSS enrichment value for ChIP-Seq experiments, but for ATAC it recommends samples having a TSS fold enrichment of >5. H3K27ac samples had TSS values between 5 and 15 and H3K4me3 samples ranged between 10 and 30 (Figure 5.15). H3K4me3 is localized more frequently around promoters than H3K27ac and is expected to have higher TSS enrichment. DESeq2 was used for differential analysis.

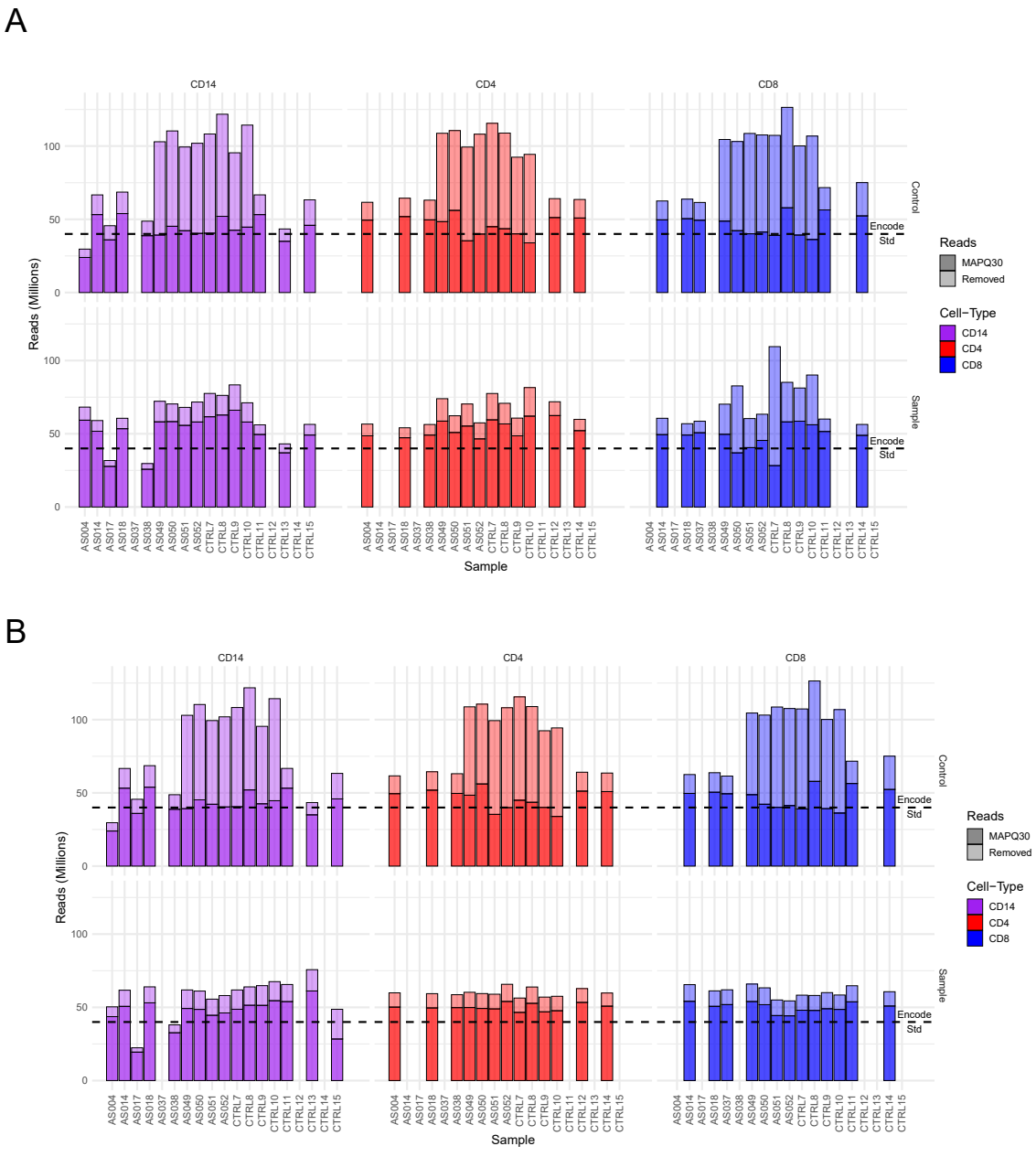


Figure 5.13: Total reads following ChIPm on AS patients and controls in CD4$^+$, CD8$^+$, and CD14$^+$ cells.
MAPQ30 reads represent usable, high quality reads. ENCODE recommends 40 million reads per sample. Control denotes Input and Sample denotes ChIP sample with antibody. **A**. H3K27ac ChIPm. **B**. H3K4me3 ChIPm.

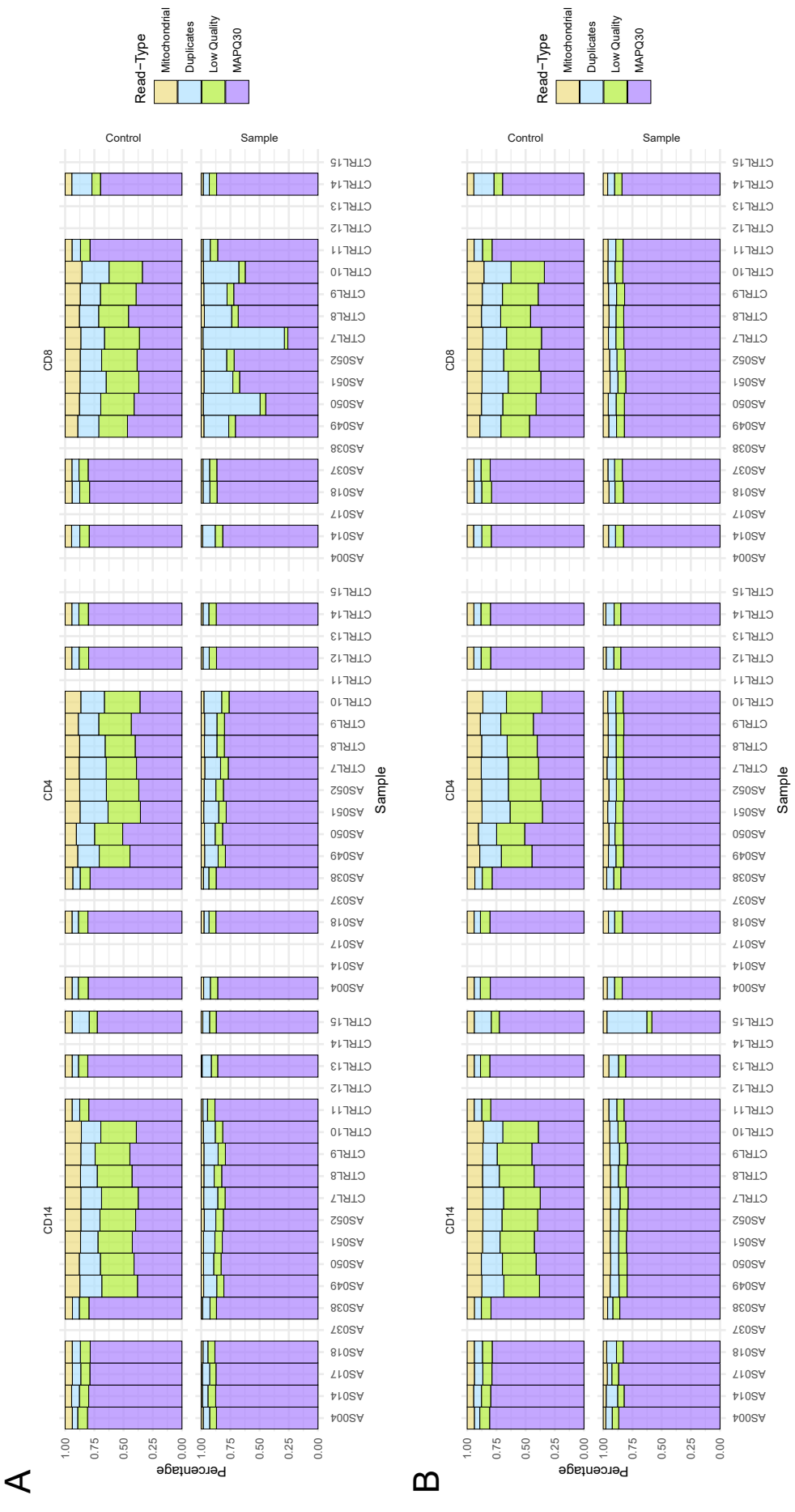


Figure 5.14: Read percentages following ChIPm on AS patients and controls in $CD4^+$, $CD8^+$, and $CD14^+$ cells. MAPQ30 reads represent usable, high quality reads. Reads are removed due to low quality, PCR duplicates, and mapping to the mitochondrial genome. **A**. H3K27ac ChIPm. **B**. H3K4me3 ChIPm.

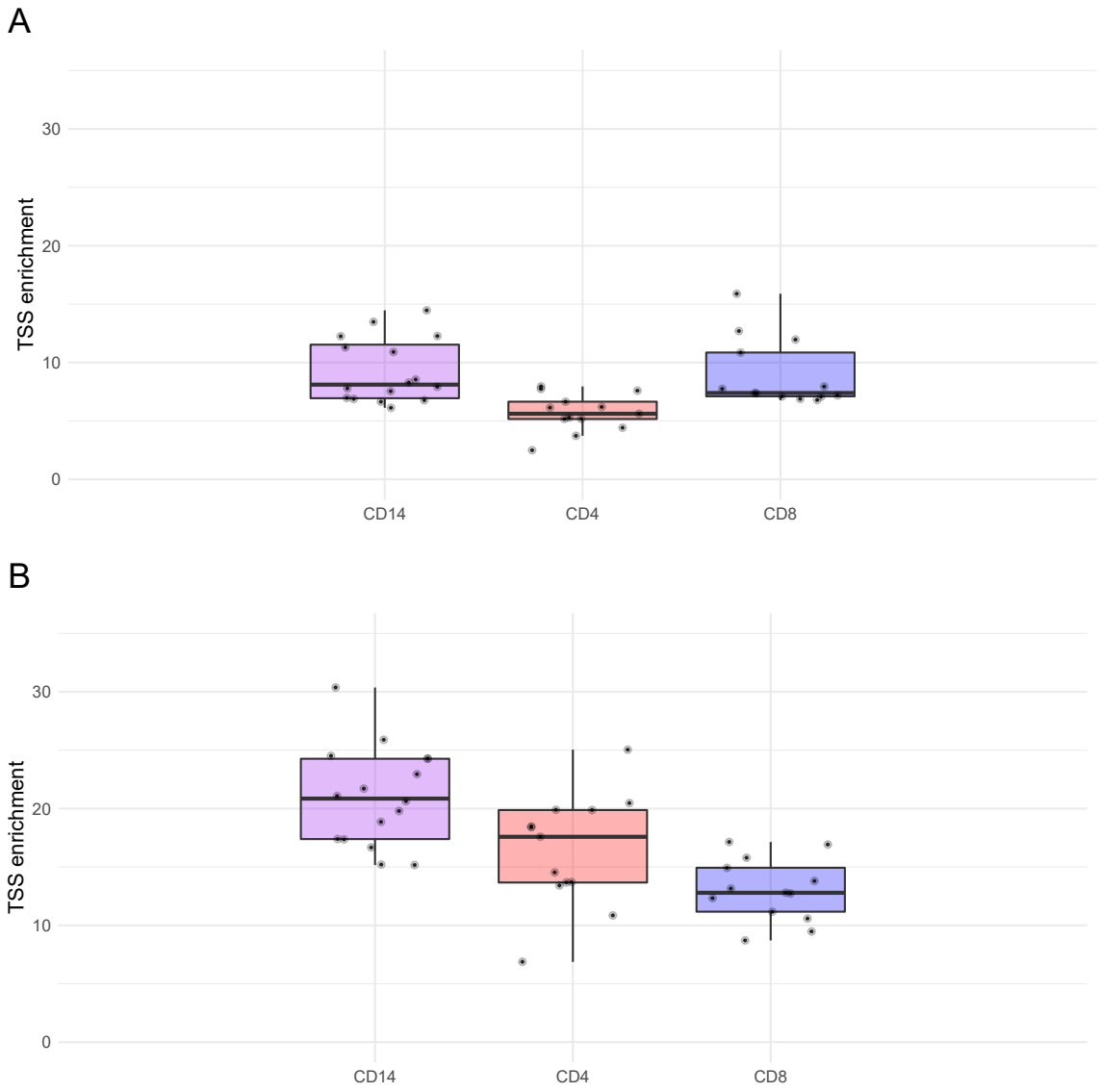


Figure 5.15: ChIPm TSS fold enrichment for AS patient samples and controls in CD4$^+$, CD8$^+$, and CD14$^+$ cell types.
A. H3K27ac ChIPm. **B**. H3K4me3 ChIPm.

Cell Type Specific Regions of H3K27ac

Enhancers can be context and cell type specific, therefore, a differential analysis on H3K27ac peaks was carried out, comparing between three cell types, $CD4^+$, $CD8^+$, and $CD14^+$. For this analysis, an ANOVA was run comparing all three cell types, followed by a pairwise comparison for each cell pairing. A peak had to have an adjusted p-value of <0.01 in both analyses to be considered significant for that cell type.

Figure 5.16 illustrates the peak distribution, representing regions of bound H3K27ac between $CD4^+$, $CD8^+$, and $CD14^+$ cell types. Figure 5.17 illustrates the number of differential peaks from the 34,412 called peaks from the masterlist following the pairwise analysis between each cell type. AS patient samples and controls were combined for this analysis. The most shared peaks are between $CD4^+$ and $CD8^+$ T cells, with $CD14^+$ monocytes having the highest number of unique peaks.

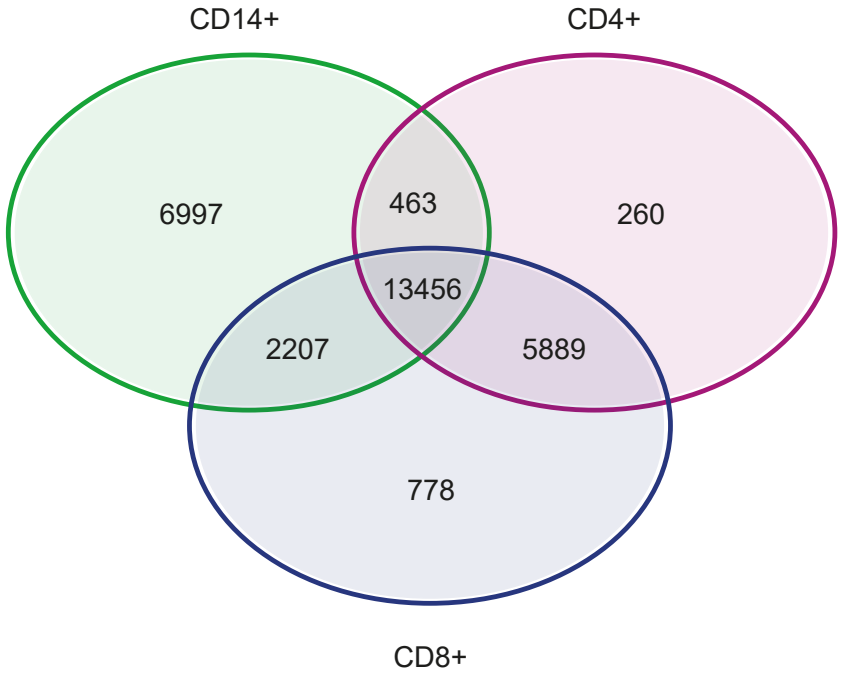


Figure 5.16: H3K27ac ChIPm peak cell type distribution.
Venn diagram showing number of H3K27ac peaks (regions of enhancers) and their distribution between $CD4^+$, $CD8^+$, and $CD14^+$ cell types. ChIPm data from 10 AS patients and 9 healthy controls.

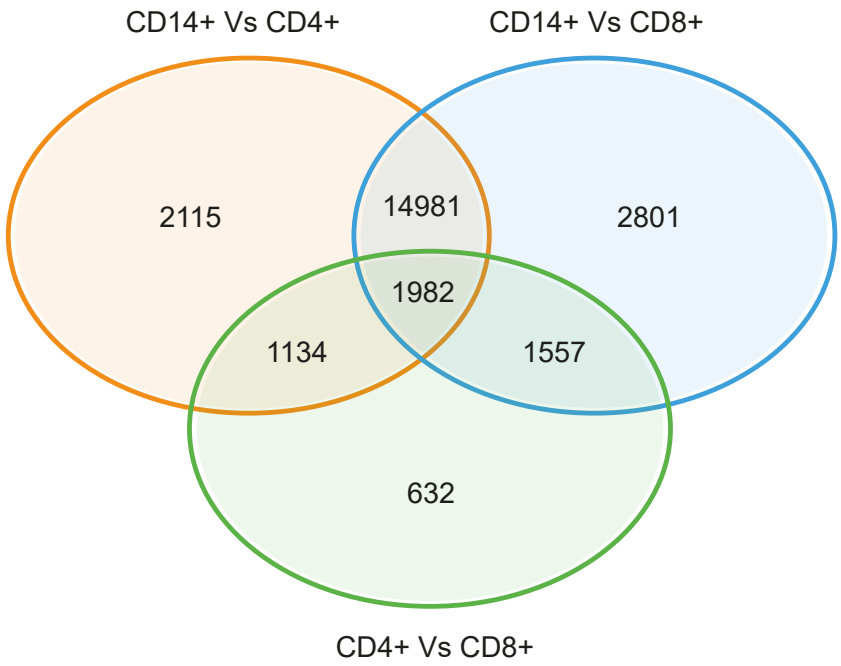


Figure 5.17: H3K27ac ChIPm differential analysis between cell types.
Venn diagram summarizing the number of differential ChIPm peaks (regions of H3K27ac) between CD14$^+$ monocytes, CD4$^+$ T cells, and CD8$^+$ T cells from all all 34,412 called peaks in the masterlist. Differential analysis was performed using a pairwise comparison between each cell type.

Pathway Enrichment Analysis on Cell Type Specific H3K27ac Peaks

To validate the cell type specific peaks as being relevant to the cells being examined, pathway enrichment analysis was performed on the genes overlapping CD4$^+$, CD8$^+$, and CD14$^+$ cell type specific H3K27ac peaks using XGR (FDR<0.05) (Fang et al. 2016) (Figure 5.18). Peaks were considered to be associated with a gene if they were within 5kb of the start or end of the gene. The gene sets used for the analysis were from Canonical, KEGG, REACTOME and BioCarta pathways (Subramanian et al. 2005). Genes associated with H3K27ac peaks from all three cell types were used as the background set of genes. The top 25 peaks as determined by their global p-value (ANOVA) are listed in Tables C.6, C.7, and C.8 in Appendix C.

In total there were 688 CD4$^+$ cell type specific H3K27ac peaks which coincided with 475 genes. The most enriched pathway was 'ERBB2 in Signal Transduction and Oncology' involving *ESR1*, *GRIP1*, *IL6*, *IL6ST*, and *PIK3CA*. Other ERBB2

related pathways were also enriched. The second most significantly enriched pathway was the 'canonical NF-κB pathway', which is a key inflammatory pathway. The genes involved in this pathway were *BIRC2*, *CYLD*, *NOD2*, *PRKCA*, and *XPO1*. Another significantly enriched pathway was 'VEGFR1 specific signals', which also plays a role in inflammation. The genes involved in this pathway were *HIF1A*, *NRP1*, *PIK3CA*, *PRKCA*, and *RASA1*.

$CD8^+$ cells had 394 specific H3K27ac peaks which were associated with 350 genes. Multiple significantly enriched pathways involved the extracellular matrix and extracellular matrix-associated proteins. The ECM is an important modulator of the inflammatory response. The most enriched pathway, 'genes encoding ECM and ECM-associated proteins' involved 34 of the genes. Some of the genes involved included *ADAM33*, *ADAMTS1*, *BMP1*, *CCL5*, *EGFL7*, *FGFBP2*, *IL11*, *IL13*, *KCP*, *LTBP2*,*PDGFD*, and *WNT11*. Some other pathways related to the immune response that were enriched were 'immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell', 'genes encoding secreted soluble factors', and 'Graft-versus-host disease'.

There were 7161 $CD14^+$ cell type specific H3K27ac peaks which were associated with 3389 genes. The top 300 genes were used for the pathway analysis. The two most enriched pathways were 'GPCR downstream signaling' and 'signaling by GPCR', which involved 21 and 22 genes, respectively. Some of the genes in these pathways included *AKT3*, *C5AR1*, *CXCL1*, *ITPR2*, *RHOB*, and *ROCK1*. 'Endogenous TLR signaling' was another significantly enriched pathway which included the genes, *CD14*, *TLR2*, and *TLR4*.

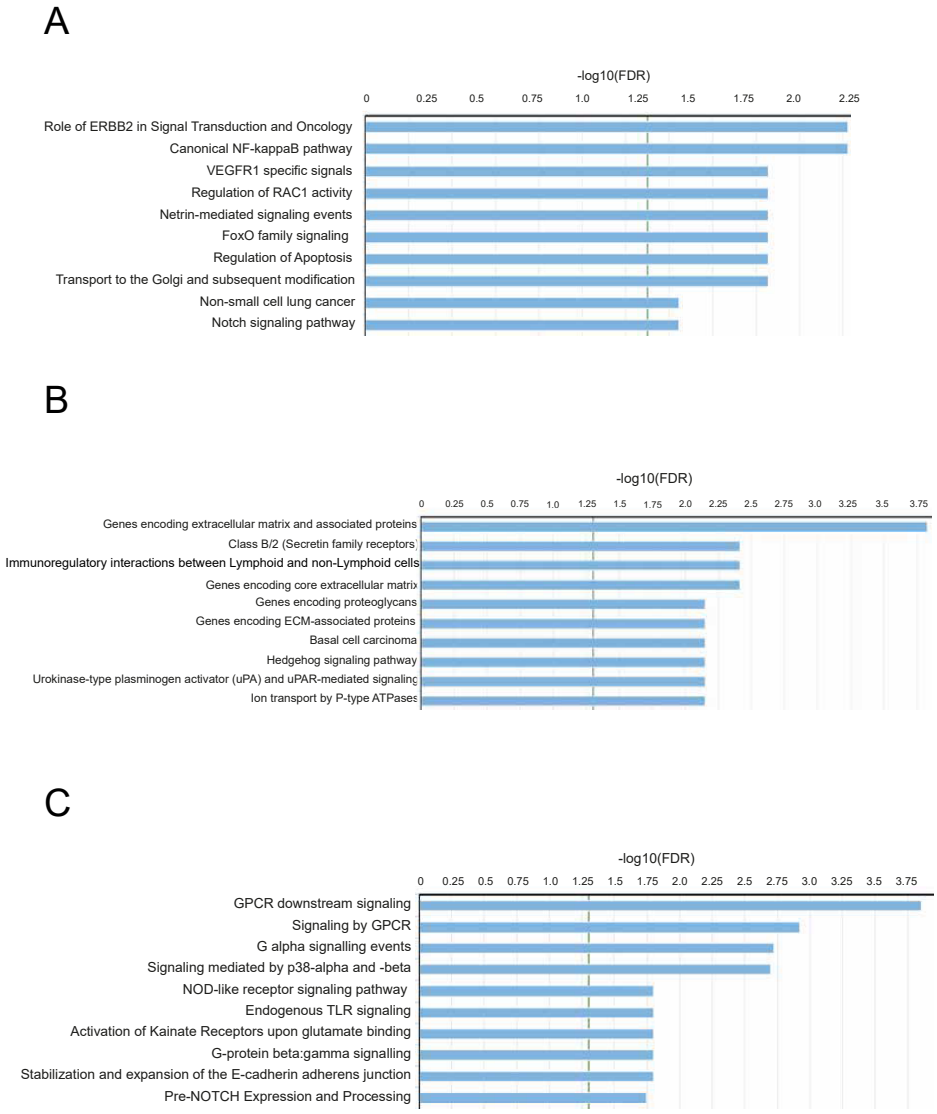


Figure 5.18: Pathway enrichment analysis on genes overlapping cell type specific H3K27ac ChIPm peaks.
Top 10 pathways from the enrichment enrichment analysis performed on genes overlapping cell type specific H3K27ac ChIPm peaks using XGR. Green line represents FDR<0.05. **A**. $CD4^+$ cells. **B**. $CD8^+$ cells. **C**. $CD14^+$ cells.

GWAS Overlap with Cell Type Specific H3K27ac Peaks

Mapping the H3K27ac landscape in immune relevant cell types can uncover enhancers involved in cell type specific immunity and can annotate disease associated non-coding variants. Therefore, the 95% credible sets of AS GWAS loci were overlapped with cell type specific H3K27ac regions. This was performed on $CD4^+$, $CD8^+$, and $CD14^+$ specific H3K27ac peaks. No $CD4^+$ cell type specific peaks overlapped GWAS loci, but peaks specific to the other two cell types did.

There was one CD8$^+$ cell type specific H3K27ac peak, chr19:10622857-10637235, overlapping an AS GWAS region and was associated with rs35164067 (Table 5.6). This peak is within the *S1PR5* gene, which has been implicated in inflammation (Aoki et al. 2016). The location of this peak is a predicted enhancer for *S1PR5* in CD8$^+$ and CD14$^+$ cells by FANTOM (Andersson et al. 2014). SNPs in this region are annotated as eQTLs for *S1PR5* (Kasela et al. 2017), and Hi-C data indicates there is an interaction between this peak and the promoter for *S1PR5* in all three cell types (Javierre et al. 2016).

There were 18 CD14$^+$ cell type specific H3K27ac peaks overlapping six AS GWAS regions (Table 5.6). Ten of the peaks overlap the *IL1R1* locus, with rs4851529 as the lead GWAS SNP (Figure 5.19). The ATAC-Seq analysis also showed a large number of CD14$^+$ specific peaks overlapping this locus. The *IL1R1* GWAS locus could be an important disease driver in CD14$^+$ cells. Two of the H3K27ac peaks are located in a gene desert located at 21q22 suggesting there may be important enhancers at this loci. Neither have predicted target genes based on Hi-C data and these regions are not predicted enhancers by NIH Roadmap or FANTOM. However, the peak at chr21:40464272-40471800 overlaps rs2836882, which is annotated as an eQTL in CD14$^+$ for *ETS2*, over 250kb away (Fairfax et al. 2014).

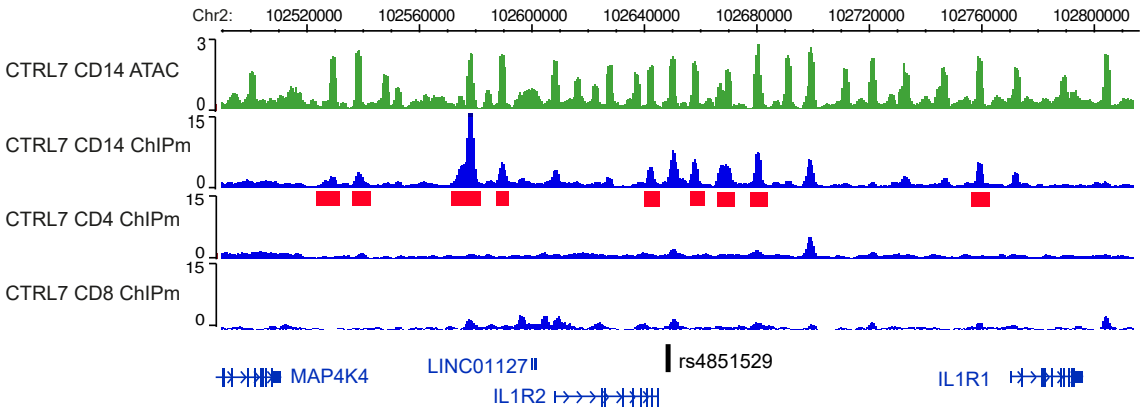


Figure 5.19: CD14$^+$ cell type specific H3K27ac ChIPm peaks overlapping the *IL1R1* GWAS locus.
WashU Epigenome Browser view of nine of the H3K27ac ChIPm peaks that overlap the rs4851529 (IL1R1) GWAS locus. CD14$^+$ ATAC-Seq peaks are shown. Each H3K27ac CD14$^+$ specific peak is underlined in red.The lead GWAS SNP, rs4851529, is indicated.

Table 5.4: Cell type specific H3K27ac peaks overlapping AS GWAS loci.
The coordinates stated are in UCSC notation based on genome build hg19. GWAS coordinates represent the 95% credible set.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	CD4 vs CD8 p-value	Proximal Gene	GWAS Loci
$CD8^+$ Specific Peaks						
chr19:10622857-10637235	$1.04x10^{-81}$	0.0529	$9.85x10^{-67}$	$3.11x10^{-55}$	*S1PR5*	rs35164067 (TYK2) chr19:10496620-10771941
$CD14^+$ Specific Peaks						
chr2:102572212-102581447	$1.07x10^{-81}$	$1.65x10^{-56}$	$8.06x10^{-61}$	0.587	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102655202-102660177	$1.24x10^{-55}$	$4.77x10^{-29}$	$2.31x10^{-50}$	0.000	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr21:40464272-40471800	$1.48x10^{-49}$	$1.03x10^{-36}$	$2.97x10^{-37}$	0.942	NA	rs9977672 (21q22) chr21:40463282-40466744
chr2:102526801-102531268	$9.89x10^{-49}$	$1.68x10^{-25}$	$2.19x10^{-42}$	0.003	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102664728-102671765	$4.92x10^{-48}$	$4.52x10^{-29}$	$2.27x10^{-42}$	0.014	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102536090-102540816	$1.38x10^{-46}$	$4.20x10^{-30}$	$5.74x10^{-38}$	0.123	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr12:111897937-111903362	$9.08x10^{-46}$	$9.90x10^{-15}$	$2.17x10^{-48}$	0.005	*ATXN2*	rs11065898 (SH2B3) chr12:111844217-112208838

chr2:102587355-102592324	$3.51x10^{-44}$	$9.97x10^{-34}$	$1.62x10^{-32}$	0.838	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr1:161473335-161478686	$1.30x10^{-36}$	$3.67x10^{-29}$	$7.67x10^{-28}$	0.802	*FCGR2A*	rs1801274 (FCGR2A) chr1:161469053-161479745
chr17:45648535-45651933	$1.83x10^{-32}$	$7.73x10^{-18}$	$1.61x10^{-31}$	0.001	*NPEPPS*	rs12943464 (NPEPPS) chr17:45539116-45786452
chr2:103124004-103127377	$7.14x10^{-30}$	$7.02x10^{-18}$	$1.14x10^{-28}$	0.009	*SLC9A4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr21:40460244-40464272	$3.62x10^{-25}$	$2.30x10^{-21}$	$7.11x10^{-21}$	0.927	NA	rs9977672 (21q22) chr21:40463282-40466744
chr2:102770419-102773863	$7.78x10^{-21}$	$1.47x10^{-14}$	$1.86x10^{-21}$	0.070	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr12:112098269-112103538	$3.31x10^{-16}$	$7.44x10^{-11}$	$2.86x10^{-19}$	0.015	*BRAP*	rs11065898 (SH2B3) chr12:111844217-112208838
chr9:139264574-139270459	$9.31x10^{-16}$	$1.87x10^{-19}$	$4.09x10^{-09}$	0.002	*CARD9*	rs1128905 (CARD9) chr9:139240629-139301583
chr2:102640062-102644346	$2.08x10^{-14}$	$4.95x10^{-09}$	$7.37x10^{-18}$	0.007	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102677415-102683799	$2.33x10^{-12}$	$3.90x10^{-10}$	$5.50x10^{-15}$	0.137	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491
chr1:161467761-161471750	$2.14x10^{-09}$	$5.61x10^{-10}$	$4.89x10^{-11}$	0.730	*FCGR2A*	rs1801274 (FCGR2A) chr1:161469053-161479745

Differential H3K27ac Landscape Between AS Patients and Controls

The differences in epigenetic landscape between patients and controls could provide insight into which regulatory elements may be driving disease pathogenesis. Differing H3K27ac histone modifications indicate a change in enhancer activity which could help annotate AS associated GWAS SNPs that fall within these regions. Moreover, they can potentially highlight genes or pathways that may be involved in disease if the target gene(s) for these enhancers are known or can be discerned. Therefore, a differential analysis comparing H3K27ac ChIP peaks was performed between AS patients and healthy controls. Because histone modifications can be cell type specific, this was carried out in $CD4^+$, $CD8^+$, and $CD14^+$ cells. The differential analysis was carried out using DESeq2, correcting for batch, with an FDR of < 0.05 being considered significant. There were differential H3K27ac peaks in $CD4^+$ and $CD14^+$ cells between patients and controls, but not in $CD8^+$ cells (Table 5.5). None of the differential peaks overlap AS GWAS regions.

Following the differential analysis between patients and controls in $CD4^+$ cells, there were four differential peaks, listed in Table 5.5. The fold change for each peak is modest, with all of them having a Log_2 fold Change of <1. The most differential peak is located in an intron of *PTCH1*. *PTCH1* encodes the receptor for sonic hedgehog, the binding of which initiates a signaling cascade for embryonic development and tumorigenesis. This peak region has been found to interact with *PTCH1* using Hi-C data (Javierre et al. 2016), and is therefore likely an enhancer for that gene. One of the peaks is within an intron of *MYO1D*, but is a FANTOM (Andersson et al. 2014) predicted enhancer for the Alzheimer's risk gene, *CDK5R1*, around 30kb away (Spreafico et al. 2018). Moreover, multiple SNPs within this peak region are annotated as eQTLs for *CDK5R1* in $CD4^+$ and $CD8^+$ T cells (Kasela et al. 2017). Peak chr17:43129300-43130342 is situated within an intron of *DCAKD*, and is a predicted enhancer for

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
$CD4^+$ Differential Peaks			
chr9:98256542-98258220	*PTCH1*	-0.5041	0.0175
chr17:30845143-30848447	*MYO1D*	-0.4944	0.0175
chr17:43129300-43130342	*DCAKD*	-0.5280	0.0307
chr7:102329801-102330538	NA	-0.5391	0.0307
$CD14^+$ Differential Peaks			
chr15:39871868-39874004	*THBS1*	0.552	0.0132
chr18:13136470-13138895	*CEP192*	-0.466	0.0132
chr3:184209679-184211402	NA	-0.508	0.0309
chr22:24540738-24544080	*CABIN1*	0.361	0.0309
chr3:43430117-43432624	*ANO10*	0.332	0.0322
chr6:18165546-18168272	*KDM1B*	0.430	0.0387
chr2:216592037-216594486	NA	0.470	0.0387
chr18:74771500-74781814	*MBP*	0.308	0.0387
chr6:134733106-134737142	NA	0.433	0.0487
chr20:57738229-57739642	*ZNF831*	0.441	0.0487
chr20:5653951-5657713	NA	0.404	0.0487
chr2:48576902-48582164	*FOXN2*	0.313	0.0487
chr1:185535500-185538980	*HMCN1*	0.472	0.0487

Table 5.5: Regions of differential H3K27ac histone modifications between AS patients and controls.
An FDR of <0.05 (DESeq2) was considered differential. The coordinates stated are in UCSC notation based on genome build hg19.

this gene by FANTOM (Andersson et al. 2014) annotation. Peak chr7:102329801-102330538 is not found within a gene but is within a pseudogene for *RASA4*, which is Ras GTPase-activating protein involved in immune cytokine signaling. However, none of these genes were differently expressed following RNA-Seq analysis between patients and controls.

The differential analysis between patients and controls in $CD14^+$ cells yielded 13 differential H3K27ac peaks (Table 5.5). As with the $CD4^+$ cell analysis, the fold change for each peak is modest, with all of them having a Log_2 fold Change of <1. The most differential peak is situated over the promoter for *THBS1*,

suggesting that *THBS1* gene activity may be different in AS patients. *THBS1* was up-regulated in AS patients compared to controls in $CD14^+$ cells in a differential RNA-Seq analysis, confirming this hypothesis (discussed in Section 5.2.4). The other most differential peak is located just downstream of *CEP192*, suggesting it may denote an enhancer for this or another gene, but it is not a predicted enhancer and there are no known physical interactions with this region and a target gene. The peak located at chr22:24540738-24544080 is situated within an intron of *CABIN1* and Hi-C data has uncovered a physical interaction between this region and the *CABIN1* promoter (Javierre et al. 2016). This peak also overlaps eQTLs for the *UPB1* and *SUSD2*, but only in $CD4^+$ cells (Kasela et al. 2017). One peak, chr18:74771500-74781814, is situated within an intron of *MBP*, and there is also a Hi-C interaction between this peak and the promoter for *MBP*, indicating that it is likely an enhancer for this gene (Javierre et al. 2016). It is also predicted by FANTOM (Andersson et al. 2014) to be an enhancer for *MBP* in $CD14^+$ cells. The peaks within introns for *KDM1B* and *ZNF831* are predicted enhancers for each of those genes, respectively (NIH Roadmap and FANTOM), and the peak within *ZNF831* also overlaps an eQTL for *SLMO2* in $CD14^+$ cells (Fairfax et al. 2014). Additionally, the peak, chr2:48576902-48582164, within *FOXN2* overlaps another $CD14^+$ specific eQTL, but for *MSH6* (Fairfax et al. 2014).

Cell Type Specific Regions of H3K4me3

H3K4me3 is primarily a mark for promoters and can indicate gene expression and activity. As with H3K27ac, the presence of H3K4me3 is context and cell type specific, so cell type specific differential analysis was performed on $CD4^+$, $CD8^+$, and $CD14^+$ cells using the same methodology and criteria as the H3K27ac ChIPm analysis (Section 5.2.3). Figure 5.20 illustrates the distribution of H3K4me3 peaks between $CD4^+$, $CD8^+$, and $CD14^+$ cells, and Figure 5.21 illustrates the number of differential peaks from the 28,511 called H3K4me3 peaks, following the pairwise analysis between each cell type.

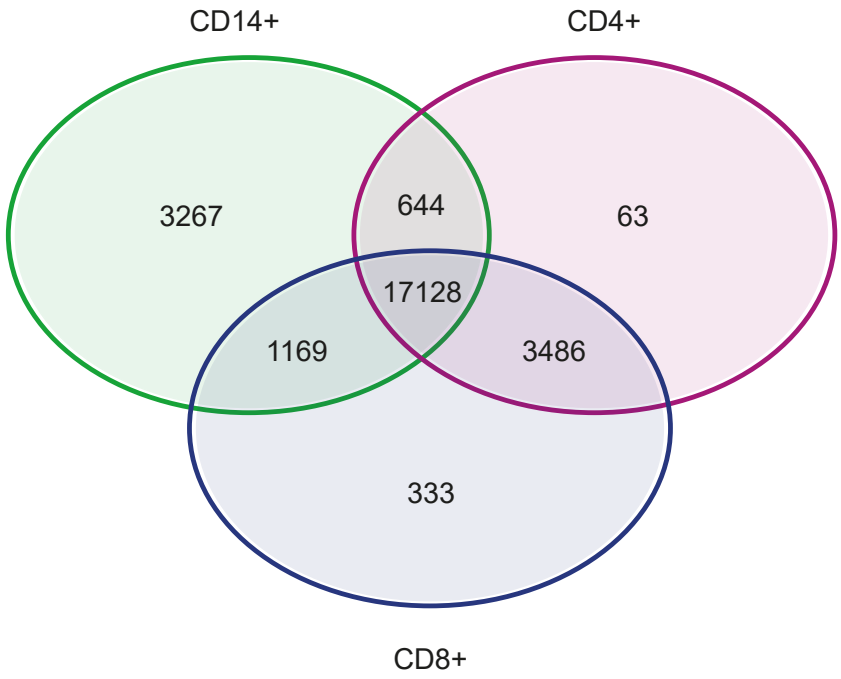


Figure 5.20: H3K4me3 ChIPm peak cell type distribution.
Venn diagram showing number of H3K4me3 peaks (regions of promoters) and their distribution between $CD4^+$, $CD8^+$, and $CD14^+$ cell types. ChIPm data from 10 AS patients and 9 healthy controls.

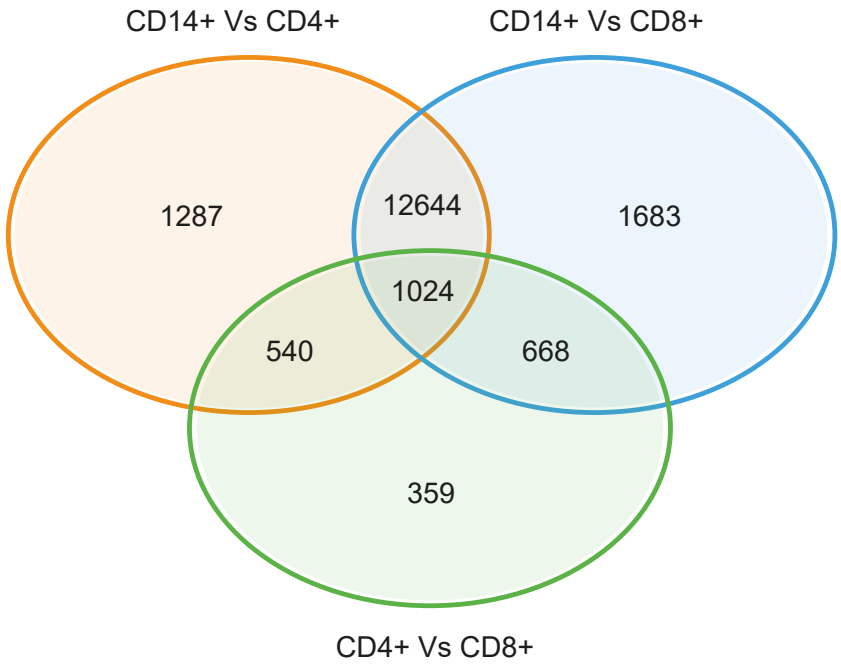


Figure 5.21: H3K4me3 ChIPm differential analysis between cell types.
Venn diagram summarizing the number of differential ChIPm peaks (regions of H3K4me3) between $CD14^+$ monocytes, $CD4^+$ T cells, and $CD8^+$ T cells from all all 28,511 called peaks from the masterlist. Differential analysis was carried out using a pairwise comparison between each cell type.

Pathway Enrichment Analysis on Cell Type Specific H3K4me3 Peaks

Pathway enrichment analysis was performed on the genes overlapping $CD4^+$, $CD8^+$, and $CD14^+$ cell type specific H3K4me3 peaks using the same methodology and parameters as the H3K27ac ChIPm analysis (Section 5.2.3) (Figure 5.22). The top 25 peaks for each cell type can be seen in Tables C.9, C.10, and C.11 in Appendix C.

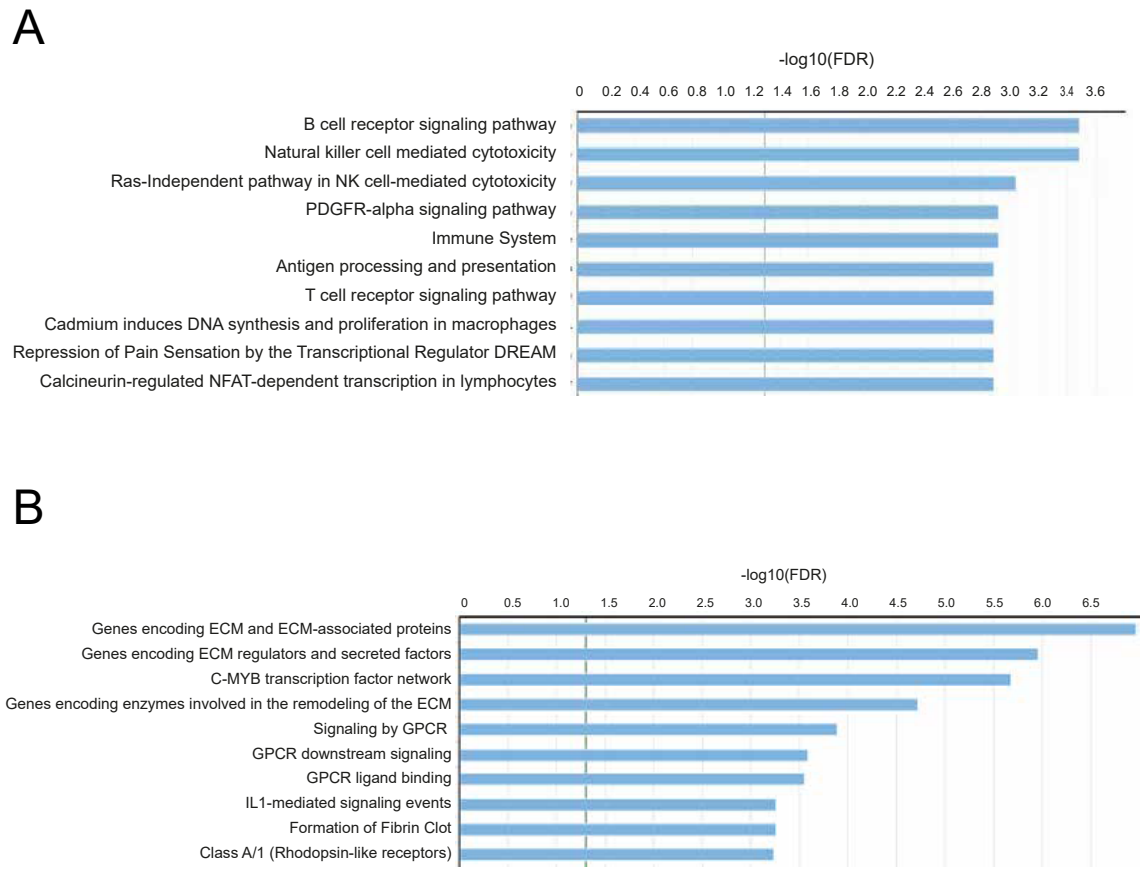


Figure 5.22: Pathway enrichment analysis on genes overlapping cell type specific H3K4me3 ChIPm peaks.
Top 10 pathways from the enrichment enrichment analysis performed on genes overlapping cell type specific H3K4me3 peaks using XGR. Green line represents FDR<0.05. **A**. $CD8^+$ cells. **B**. $CD14^+$ cells.

There were 63 $CD4^+$ cell type specific H3K4me3 peaks coinciding with 48 genes, but pathway analysis was not performed in this cell type as there were not enough genes.

There were 333 $CD8^+$ cell type specific H3K4me3 peaks associated by proximity

with 264 genes. The most significantly enriched pathways were related to the immune system and the inflammatory response. They included 'B cell receptor signaling pathway', 'natural killer cell mediated cytotoxicity', 'Ras-Independent pathway in NK cell-mediated cytotoxicity', 'immune system', 'antigen processing and presentation', and 'T cell receptor signaling pathway'. The genes overlapping H3K4me3 peaks involved in T cell receptor signaling were *AKT3*, *CARD11*, *CD8B*, *FOS*, *JUN*, *NFKBIA*, *VAV3*, *ZAP70*.

A total of 2154 genes coincided with the 3267 $CD14^+$ cell type specific H3K4me3 peaks. Pathway analysis was performed on the top 300 genes. ECM pathways were three of the most significantly enriched which included 38, 29 and 13 genes. 'Signaling by GPCR', 'GPCR downstream signaling', and 'GPCR ligand binding' were also highly enriched pathways. GPCRs are regulators of many cellular functions and responses, including immunity. Another enriched pathway related to immunity was 'IL-1 mediated signaling events' which included the genes *IL1B*, *IL1R2*, *IL1RN*, *IRAK3*, *MAP2K6*.

GWAS Overlap with Cell Type Specific H3K4me3 Peaks

There was one $CD4^+$ cell type specific H3K4me3 peak overlapping an AS GWAS region (Table 5.6). This peak overlaps the final intron and exon of *IL6R*. This region is not predicted to be a enhancer for any gene by NIH Roadmap or FANTOM (Bernstein et al. 2010; Andersson et al. 2014), and has no interacting partners in Hi-C data (Schofield et al. 2016). The lead GWAS SNP, rs4129267, at the *IL6R* locus is an eQTL for *IL6R* in peripheral blood (Revez et al. 2013), but this peak does not specifically overlap any SNPs that have been annotated as eQTLs. A different region of this GWAS loci was highlighted in the ATAC-Seq $CD14^+$ cell analysis and showed $CD14^+$ specific interactions (Section 5.2.2). There may be yet uncovered significance of this particular region specific to $CD4^+$ cells.

Peak chr1:25295136-25298173 was the only H3K4me3 $CD8^+$ cell type specific

peak overlapping an AS GWAS region (Table 5.6) and is about 3.5kb upstream of *RUNX1*. This region is associated with the GWAS SNP, rs6602047. This peak is not a predicted enhancer (Bernstein et al. 2010; Andersson et al. 2014), has no known physical interactions in Hi-C data (Schofield et al. 2016), and does not coincide with any known eQTLs.

There were 12 $CD14^+$ cell type specific H3K4me3 ChIPm peaks overlapping five AS GWAS regions (Table 5.6). Six of the peaks overlap the *IL1R1* locus, with rs4851529 as the lead GWAS SNP (Figure 5.23). This is similar to the $CD14^+$ analysis of H3K27ac peaks as well as the ATAC-Seq analysis. This suggests that the *IL1R1* GWAS locus is prominent in $CD14^+$ cells and could implicate this cell type in disease. *CARD9*, *FCGR2A*, *IL27*, and *TYK2* are other GWAS loci over lapping H3K4me3 $CD14^+$ peaks. Peak chr9:139263567-139270258 is situated over the promoter for *CARD9* and overlaps a *CARD9* eQTL (Fairfax et al. 2014). One peak, chr21:40464261-40468314, was found within the gene desert at 21q22 and overlaps an eQTL for *ETS2* (Fairfax et al. 2014), suggesting it may represent an enhancer for this gene. This H3K4me3 peak also coincides with a $CD14^+$ cell type specific H3K27ac peak. There is no publicly available Hi-C data linking these regions, but in-house Capture-C data has uncovered a $CD14^+$ specific physical interaction between the promoter of *ETS2* and the region around the eQTL SNP, rs2242944 (Figure 5.24). This interaction is visible from the *ETS2* promoter viewpoint and from the viewpoint around rs2242944, and does not occur in $CD4^+$ or $CD8^+$ cells.

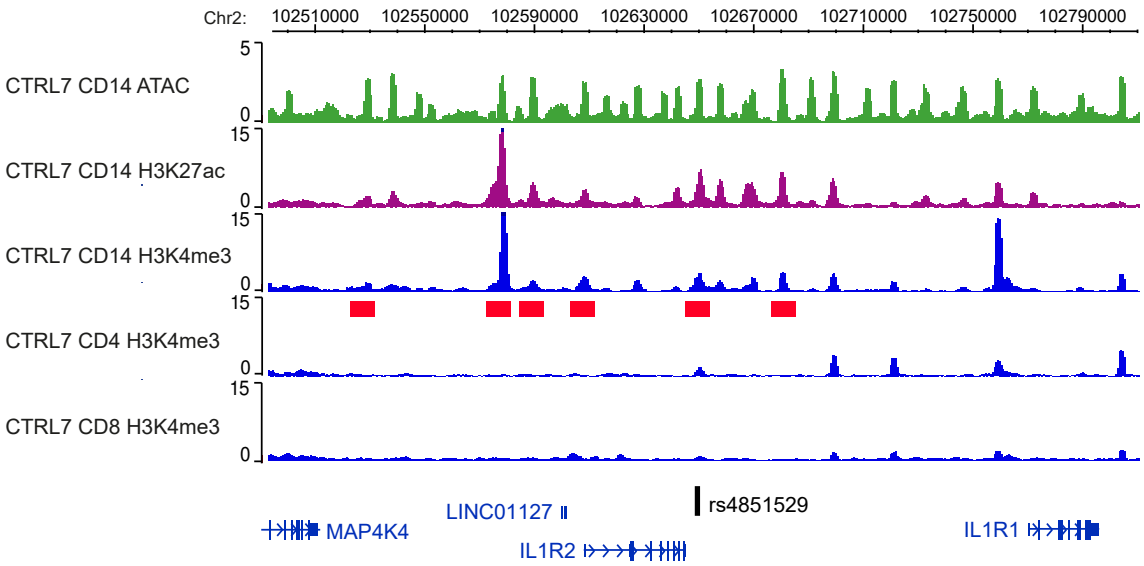


Figure 5.23: CD14$^+$ cell type specific H3K4me3 ChIPm peaks overlapping the *IL1R1* GWAS locus.
WashU Epigenome Browser view of the six H3K27ac ChIPm peaks that overlap the rs4851529 (IL1R1) GWAS locus. CD14$^+$ ATAC-Seq and H3K27ac peaks are shown. Each H3K4me3 CD14$^+$ specific peak is underlined in red.The lead GWAS SNP, rs4851529, is indicated.

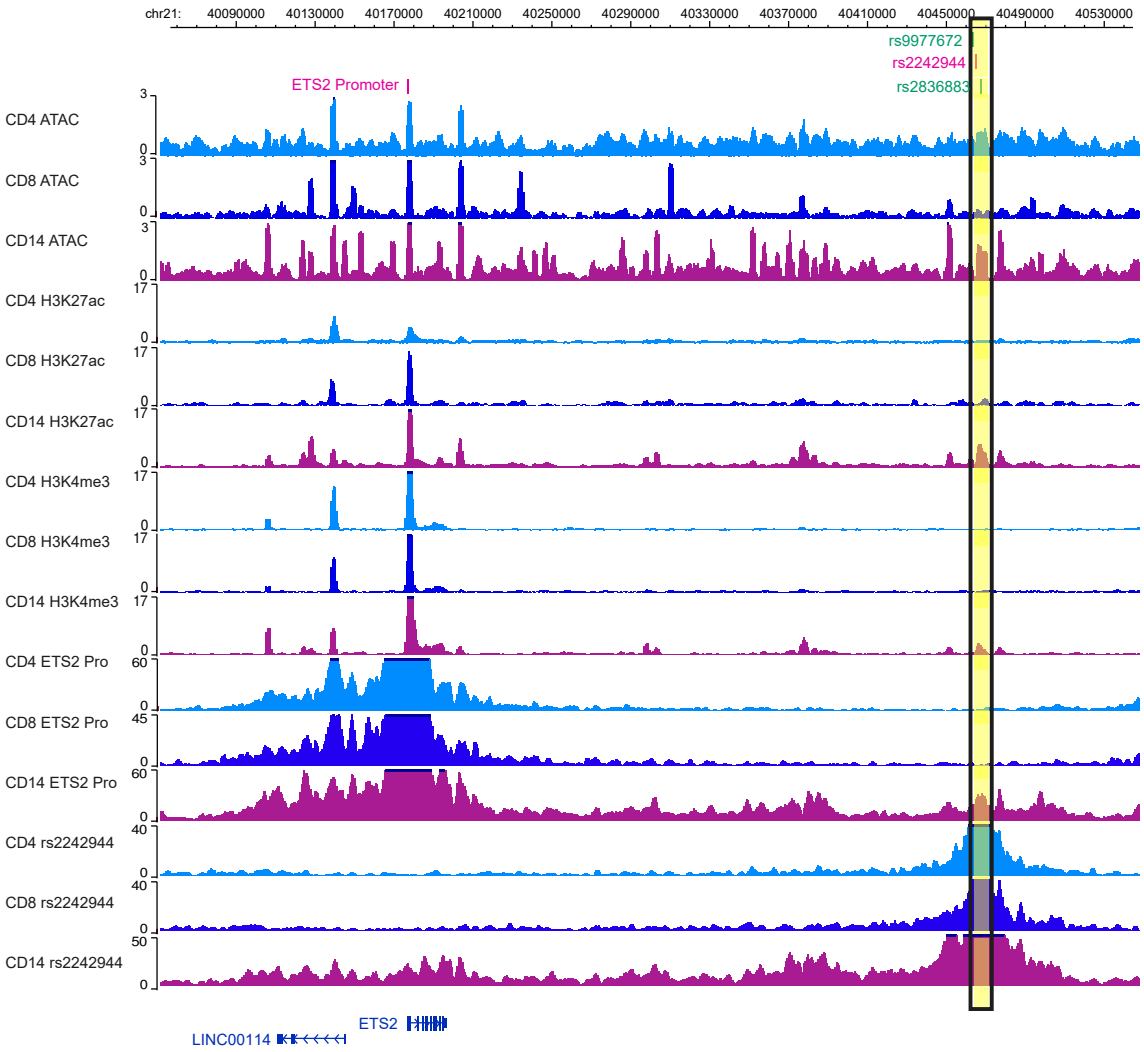


Figure 5.24: Capture-C, ATAC-Seq, and ChIPm data for the *ETS2* gene and 21q22 AS GWAS locus.
WashU Epigenome Browser view of the *ETS2* locus and eQTL SNP, rs2242944, around 300kb away. $CD4^+$ (light blue), $CD8^+$ (dark blue), and $CD14^+$ (magenta) cells are shown. The $CD14^+$ cell type specific H3K27ac and H3K4me3 ChIPm peak is highlighted in yellow. The Capture-C tracks denote the viewpoint around the *ETS2* Promoter or the viewpoint around rs2242944. The interaction between *ETS2* and rs2242944 is visible only in $CD14^+$ cells.

Table 5.6: Cell type specific H3K4me3 peaks overlapping AS GWAS loci.
The coordinates stated are in UCSC notation based on genome build hg19. GWAS coordinates represent the 95% credible set.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	CD4 vs CD8 p-value	Proximal Gene	GWAS Loci
$CD4^+$ Specific Peaks						
chr1:154434754-154441500	$7.26x10^{-27}$	$4.08x10^{-28}$	0.0021	$7.20x10^{-15}$	*IL6R*	rs4129267 (IL6R) chr1:154417828-154507888
$CD8^+$ Specific Peaks						
chr1:25295136-25298173	$2.93x10^{-26}$	$1.87x10^{-08}$	$1.03x10^{-31}$	$4.26x10^{-10}$	*RUNX3*	rs6602047 (RUNX3) chr1:25291009-25305783
$CD14^+$ Specific Peaks						
chr9:139263567-139270258	$1.02x10^{-251}$	$6.43x10^{-163}$	$5.64x10^{-174}$	0.228	*CARD9*	rs1128905 (CARD9) chr9:139240629-139301583
chr1:161473341-161483370	$6.70x10^{-178}$	$1.03x10^{-112}$	$8.47x10^{-138}$	0.009	*FCGR2A*	rs1801274 (FCGR2A) chr1:161469053-161479745
chr2:102575758-102581512	$4.18x10^{-177}$	$1.58x10^{-118}$	$1.79x10^{-119}$	0.703	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr16:28618080-28624034	$3.20x10^{-76}$	$6.11x10^{-57}$	$4.68x10^{-51}$	0.513	*NPIPB7*	rs26528 (IL27) chr16:28483060-28895130
chr21:40464261-40468314	$5.87x10^{-34}$	$6.60x10^{-31}$	$1.20x10^{-21}$	0.090	NA	rs9977672 (21q22) chr21:40463282-40466744
chr2:102678612-102682451	$3.59x10^{-33}$	$1.46x10^{-32}$	$2.64x10^{-18}$	0.004	*IL1R1*	rs4851529 (IL1R1) chr2:102322810-103142491

chr2:102587580-102591085	$1.05x10^{-26}$	$1.40x10^{-28}$	$2.52x10^{-13}$	0.001	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr2:102647923-102652167	$2.00x10^{-25}$	$2.88x10^{-18}$	$1.04x10^{-24}$	0.114	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491
chr16:28552648-28556240	$2.00x10^{-24}$	$6.59x10^{-21}$	$2.04x10^{-19}$	0.953	*NPIPB7*	rs26528 (IL27) chr16:28483060-28895130
chr2:102527735-102531257	$1.71x10^{-18}$	$9.70x10^{-20}$	$1.35x10^{-12}$	0.087	*MAP4K4*	rs4851529 (IL1R1) chr2:102322810-103142491
chr19:10647947-10651391	$6.20x10^{-17}$	$3.07x10^{-19}$	$8.65x10^{-11}$	0.030	*ATG4D*	rs35164067 (TYK2) chr19:10496620-10771941
chr2:102606382-102610738	$1.41x10^{-16}$	$5.52x10^{-18}$	$2.98x10^{-12}$	0.163	*IL1R2*	rs4851529 (IL1R1) chr2:102322810-103142491

Differential H3K4me3 Landscape Between AS Patients and Controls

A differential analysis comparing H3K4me3 ChIP peaks between AS patients and controls was performed in $CD4^+$, $CD8^+$, and $CD14^+$ cells. The same methodology and significance thresholds as the H3K27ac analysis were used (Section 5.2.3). There were no differential H3K4me3 peaks in $CD14^+$ cells, but the other two cell types exhibited regions of differential H3Kme4 histone modifications between patients and controls (Table 5.7). None of the differential peaks overlap AS GWAS regions.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
$CD4^+$ Differential Peaks			
chr17:41400392-41401195	NA	0.526	0.0021
chr19:48756787-48762183	*CARD8*	0.368	0.0060
chr1:234461683-234462739	*SLC35F3*	-0.391	0.0421
$CD8^+$ Differential Peaks			
chr7:70158104-70161988	*AUTS2*	-0.370	0.0034
chr9:80907509-80908542	*PSAT1*	0.468	0.0065
chr2:96807603-96814973	*DUSP2*	-0.191	0.0127
chr2:10571894-10573187	*HPCAL1*	-0.433	0.0163
chr16:85368341-85369262	NA	0.341	0.0247
chr9:98266837-98274542	*PTCH1*	-0.216	0.0297
chr19:36397972-36399472	*TYROBP*	-0.439	0.0333
chr19:52627188-52627892	*ZNF616*	0.419	0.0413
chr8:11662777-11664468	*FDFT1*	0.421	0.0413

Table 5.7: Regions of differential H3K4me3 histone modifications between AS patients and controls.
FDR of <0.05 (DESeq2) was considered differential. The coordinates stated are in UCSC notation based on genome build hg19.

The differential analysis between patients and controls in $CD4^+$ cells revealed three differential H3K4me3 peaks. As seen with the H3K27ac analyses, the fold change for the H3K4me3 $CD4^+$ peaks are modest, with all of them having a Log_2 fold Change of <1. The most differential peak is located in the middle of an

intergenic region and likely denotes an enhancer and not a promoter. FANTOM predicts this region to be an enhancer for *TMEM106A*, and *NBR1* (Andersson et al. 2014), but it does not overlap an eQTL and there is no Hi-C data physically associating it with a target gene (Schofield et al. 2016). The second differential peak is situated over the promoter for *CARD8*. This suggests that there could be changes in the expression of this gene in AS patients, but no such changes were found when examining in-house RNA-Seq data between patients and controls.

Following the differential analysis between patients and controls in $CD8^+$ cells, there were nine differential H3K4me3 peaks, all with a modest fold change. The most differential peak can be seen in Figure 5.25, but the fold change is too marginal to be visualized in all samples. This peak is situated within an intron of *AUTS2*, but is not a predicted enhancer by FANTOM or NIH Roadmap. One of the differential peaks is located over the promoter for *PTCH1*. A peak over an intron of *PTCH1* was also significantly differential in the analysis of H3K27ac in patients and controls in $CD4^+$ cells. Figure 5.26 illustrates the *PTCH1* gene with associated ATAC-Seq and ChIPm peaks. However, *PTCH1* was not differentially expressed between patients and controls.

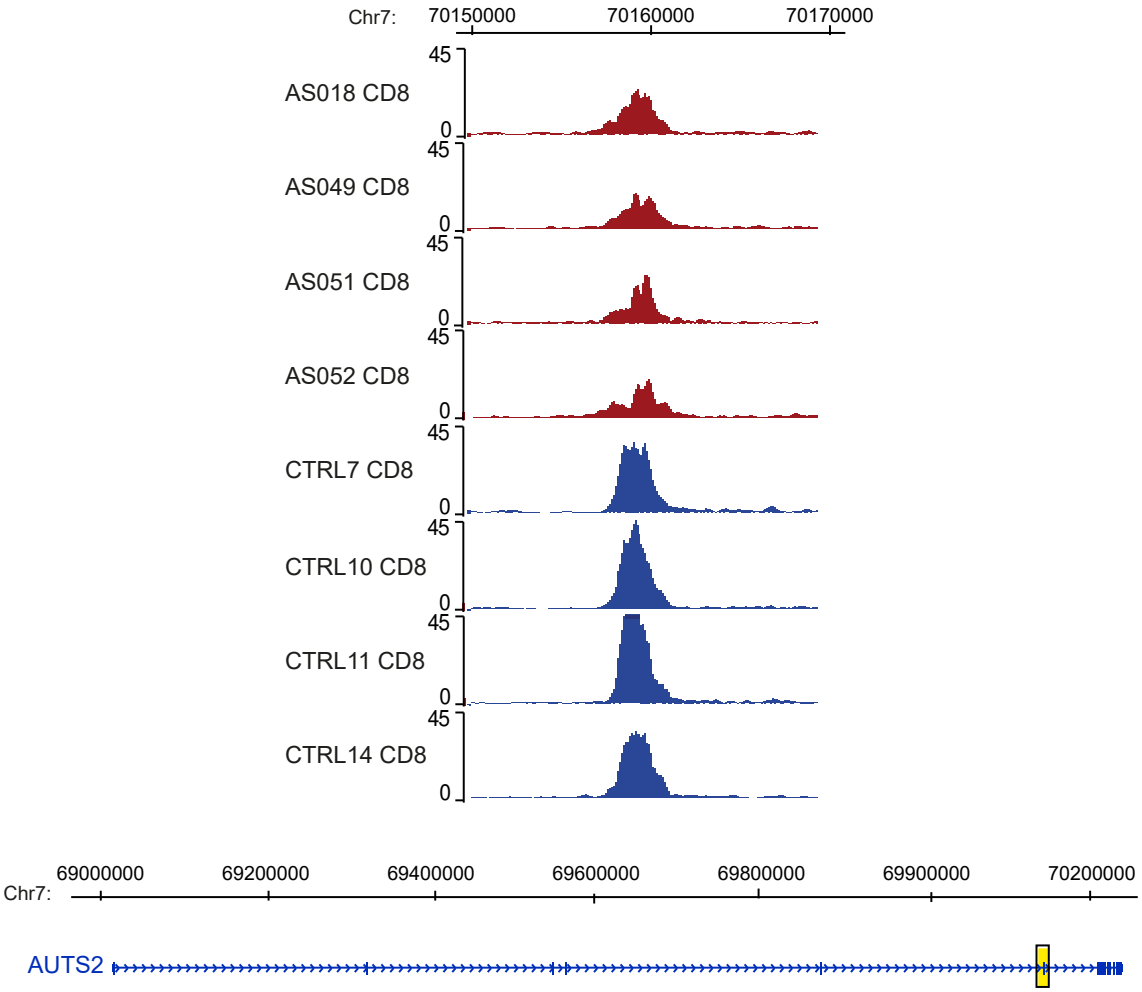


Figure 5.25: Peaks showing a region of differential H3K4me3 between AS patients and controls in CD8$^+$ cells.
Peak chr7:70158104-70161988 (Log_2 fold Change -0.370, FDR 0.0034) in four representative AS patient samples and healthy control samples. The Log_2 fold Change is based on an average of all the AS samples. A full view of *AUTS2* is shown, with the peak region highlighted in yellow.

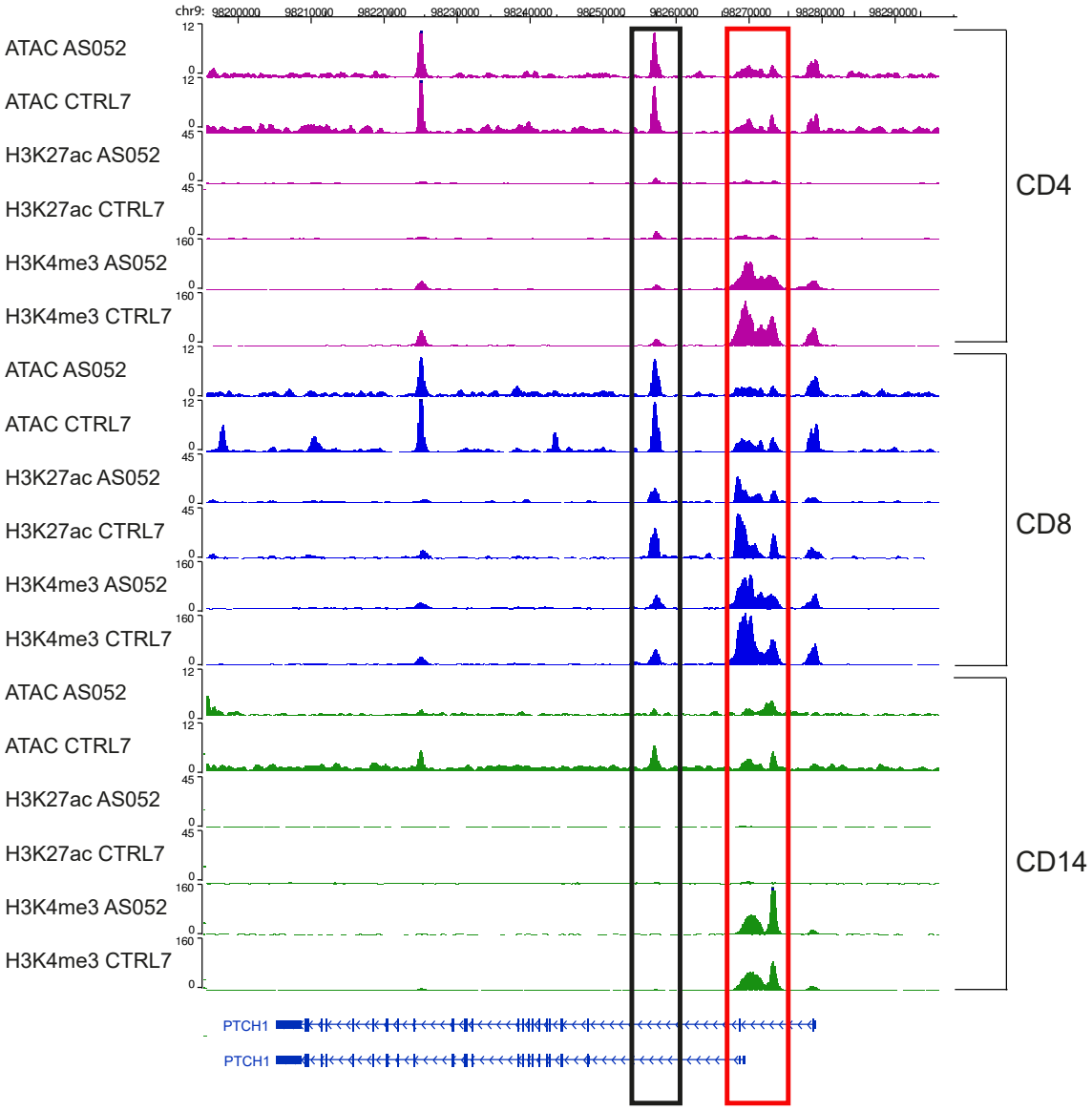


Figure 5.26: *PTCH1* gene in AS patients and controls.
WashU Epigenome Browser view of the results of ATAC, H3K27ac ChIP, and H3K4me3 ChIP across $CD4^+$, $CD8^+$, and $CD14^+$ cells over the *PTCH1* gene. One representative AS patient and control sample are shown for each technique and cell type. The $CD8^+$ differential H3K4me3 peak is outlined in red (Log_2 fold Change -0.216) and the $CD4^+$ H3K27ac peak is outlined in black (Log_2 fold Change -0.5041).

5.2.4 Data Integration

To uncover regulatory regions associated with AS patients or healthy controls, differential regions of open chromatin and hisone modifications were compared within each cell type. The coordinates of each differential ATAC-Seq peak and H3K27ac and H3K4me3 ChIPm peaks were overlapped to see if they coincided.

Genes that these peaks were associated with were also compared to see if any of them shared differential epigenetic marks within a particular cell type. None of the differential regions of open chromatin, H3K27ac locations, or H3K4me3 locations coincided with each other in $CD4^+$, $CD8^+$, or $CD14^+$ cells. All differential regions between AS patients and controls from the ATAC and ChIP experiments were also compared against GWAS regions. However, no GWAS loci overlap any regions of differential chromatin accessibility or differential H3K27 and H3K4me3 regions in any of the three cell types.

In-house RNA-Seq data was available on the studied AS patients and controls and a differential expression analysis was performed in $CD4^+$, $CD8^+$, or $CD14^+$. For this analysis, genes were considered significant with a FDR of <0.05. A total of 241 genes, 242 genes, and 1818 genes were differentially expressed in $CD4^+$, $CD8^+$, and $CD14^+$ cells, respectively (data not shown). These differentially expressed gene lists were compared against the genes associated with differential chromatin accessibility and H3K27 and H3K4me3 regions. Genes were associated with differential regions based on proximity, if a peak region had been annotated as enhancer for a particular gene, had evidence of a physical interaction, or eQTL overlap. No differentially expressed genes overlapped these epigenetic regions in $CD4^+$ or $CD8^+$ cells, but three genes (*ANO10*, *THBS1*, *CABIN1*) overlapped regions of differential H3K27ac between patients and controls in $CD14^+$ cells. These genes were all up-regulated in AS patients (Table 5.8) and all play a role in inflammation and immunity. The differential H3K27ac peaks overlapping these three genes also correspond with regions of open chromatin and H3K4me marks.

The H3K27ac peak in *THBS1* (chr15:39871868-39874004) was the most differential in the ChIPm analysis with a Log_2 fold Change of 0.552, and an FDR of 0.0132. This peak is situated over the promoter for *THBS1*, and is differentially up-regulated in AS patients, suggesting that *THBS1* is more active.

This is consistent with *THBS1* expression being up-regulated in AS patients. This peak, across all cell types, can be seen in Figure 5.27. The H3K4me3 peak over the *THBS1* visually appears larger in some of the AS patient samples, but was not statistically significant in the differential analysis. It is possible that with a larger sample population, the H3K4me3 *THBS1* promoter peak would become significant. There is a region of open chromatin over this gene in all three cell types, but appears most active in $CD14^+$ cells.

Gene	Log_2 fold Change	FDR
ANO10	0.25	0.00016
THBS1	2.05	0.00025
CABIN1	0.19	0.02710

Table 5.8: Differentially expressed genes in $CD14^+$ cells between AS patients and controls.
Three differentially expressed genes overlapped differential H3K27ac peaks between AS patients and controls. FDR from DESeq2.

ANO10 was also differentially expressed between patients and controls in $CD14^+$ cells. The differential H3K27ac peak, chr3:43430117-43432624, corresponds with the last intron of *ANO10* and was up-regulated in AS patients (Figure C.1 in Appendix C). This is consistent with this peak denoting an enhancer for *ANO10* that is more active in AS patients to cause an increase in mRNA levels. It also coincides with an ATAC-Seq peak, but not an H3K4me3 peak.

The differential H3K27ac peak, chr22:24540738-24544080, corresponds with an intronic region of *CABIN1* and is up-regulated in AS patients (Figure C.2 in Appendix C). *CABIN1* was also differentially up-regulated in AS patients in $CD14^+$ cells. This peak likely indicates the presence of enhancer that positively regulates *CABIN1*. It also overlaps an ATAC peak, further supporting this theory. This putative enhancer appears more active in AS patients and correlates with an increase *CABIN1* mRNA levels. It does not, however, overlap an H3K4me3 peak.

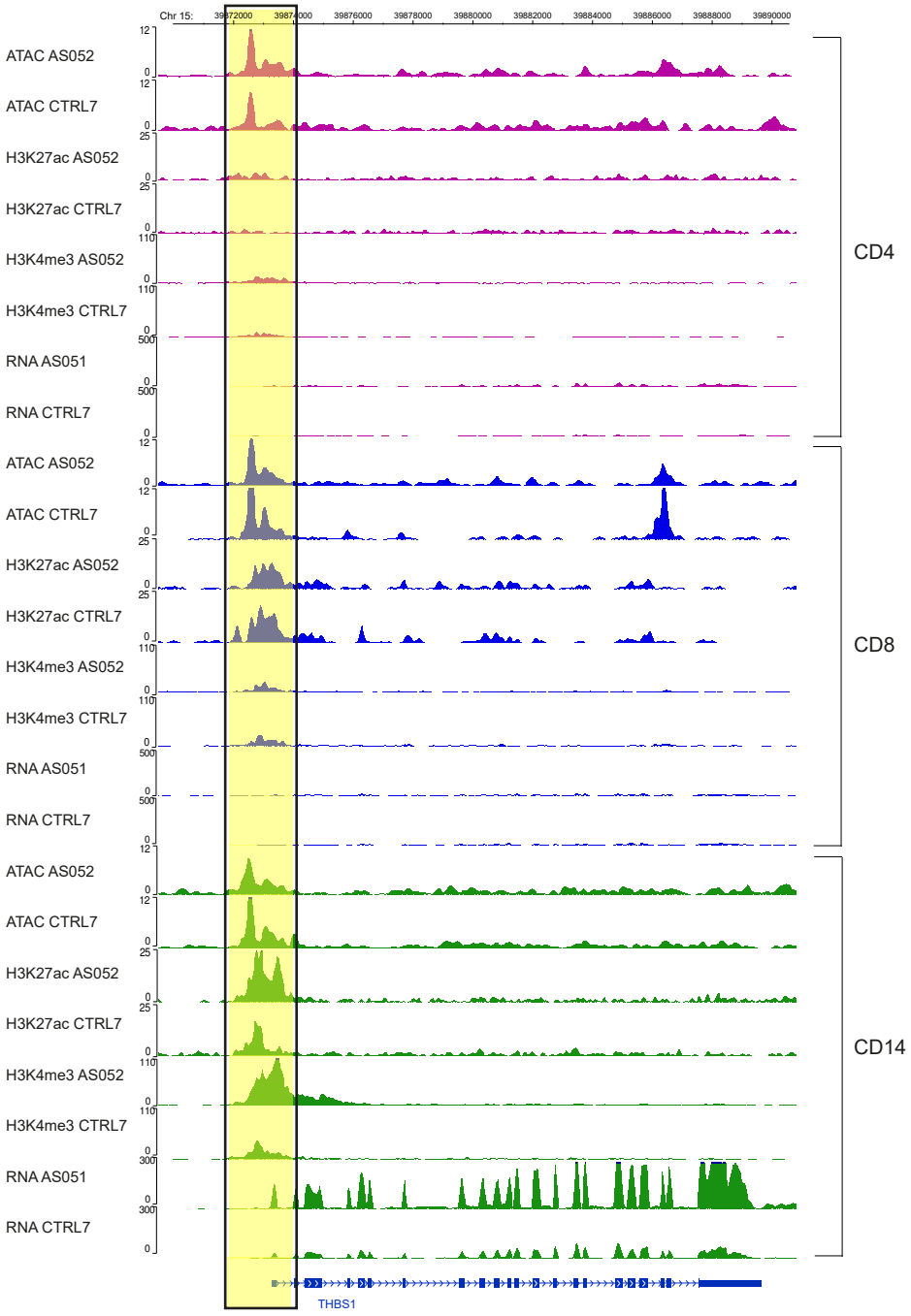


Figure 5.27: Chromatin accessibility and histone modifications over *THBS1*. WashU Epigenome Browser view of the results of ATAC, H3K27ac ChIP, H3K4me3 ChIP, and RNA-Seq across $CD4^+$, $CD8^+$, and $CD14^+$ cells over *THBS1*. One AS patient and one control sample are shown for each technique and cell type. The $CD14^+$ differential H3K27ac peak over the *THBS1* promoter is highlighted in yellow.

5.3 Discussion

5.3.1 Important Cell Types in AS Pathology

$CD4^+$ and $CD8^+$ T cells, and $CD14^+$ monocytes have all been implicated in AS and remain important cell types in understanding its pathology. Cell type specific peaks indicating chromatin accessibility, and enhancer and promoter locations overlapped multiple AS GWAS loci in each cell type, including *IL6R*, *IL1R1*, and *TGB2*, suggesting all these cell types may play a role in disease. $CD14^+$ cells had the highest number of cell type specific peaks for ATAC and ChIP and also overlapped the most GWAS regions. All three cell types are therefore important cell types to consider for functional studies of AS associated loci. $CD14^+$ cells were selected for the functional work carried out in Chapters 3 and 4, and remain a valid and useful model for understanding AS aetiology.

5.3.2 Cell Type Specific Pathway Enrichment

Pathway enrichment was performed on $CD4^+$, $CD8^+$, and $CD14^+$ cells using genes overlapping cell type specific ATAC and ChIP peaks to validate this method for defining genes and associated epigenetic trait peaks. All three cell types shared pathways related to immunity.

$CD4^+$ T Cells

The most significantly enriched pathways in the ATAC $CD4^+$ specific analysis were 'TCR signaling in naive $CD8^+$ T cells' and 'TCR signaling in naive $CD4^+$ T cells'. These pathway are expected come up as being significantly enriched as $CD4^+$ and $CD8^+$ share many genes and work together during an immune response. Naive $CD4^+$ T cells respond to MHC class II peptides and differentiate into effector subtypes to modulate the immune response, primarily the innate immune response through the secretion of cytokines to activate B cells and

cytotoxic $CD8^+$ T cells (Doyle and Strominger 1987; Luckheeram et al. 2012). The 'co-stimulatory signal during T-cell activation' was enriched in both the ATAC and H3K27ac analysis, which is expected of T cells. CD28 and CD247 are also T cell receptors that work with CD4 and CD8, and they appeared in pathways in both the ATAC and ChIP and analysis (Luckheeram et al. 2012). The most enriched pathway following the H3K27ac pathway was the 'role of ERBB2 in signal transduction and oncology'. *ERBB2* encodes a receptor tyrosine kinase has been associated with inflammation in a cancer context (Madson et al. 2006; Zeng et al. 2012). This pathway was also enriched in the ATAC analysis. Unfortunately, there were only 48 genes overlapping H3K4me3 $CD4^+$ peaks, which is not enough for a robust pathway analysis. Many pathways involved in immunity and inflammation were shared between the ATAC and H3K27ac analyses. These included the 'canonical NF-κB pathway', 'T cell signaling pathways', and the 'immune system'. All of these pathways are associated with $CD4^+$ T cell biology and would be expected to be enriched when examining genes associated with cell type specific open chromatin and histone modifications.

$CD8^+$ T Cells

The most significantly enriched pathway following the ATAC chromatin accessibility analysis was 'signaling by Rho GTPases'. Rho GTPases are involved in signal transduction in a variety of biological pathways, including the immune system. Rho GTPases are important in T cell biology as they play a role in their development, migration, signaling, and differentiation into T cell subtypes (Saoudi et al. 2014). Therefore, it is unsurprising that this is the top enriched pathway. Various NOTCH signaling pathways were also highly enriched in the ATAC analysis. NOTCH proteins are cell surface receptors found on a variety of cell types including $CD8^+$ and $CD4^+$ T cells (Amsen et al. 2015). They are believed to be involved in the differentiation of naive T cells into effector cell subsets so it is unsurprising that a number of NOTCH signaling

pathways are enriched in this analysis (Amsen et al. 2015). The most enriched pathway following the H3K27ac analysis was 'genes encoding the ECM and ECM-associated proteins including ECM-affilaited proteins, ECM regulators and secreted factors'. This and related pathways were also enriched in the $CD4^+$ analysis. The ECM is an important modulator of the inflammatory response. Many aspects of an immune response take place within the ECM and rely on ECM proteins (Morwood and Nicholson 2006; Tomlin and Piccinini 2018). Some ECM proteins have pro-inflammatory functions to propagate the immune response by activating key regulators (Jiang et al. 2005). In the H3K4me3 analysis, the 'B cell receptor signaling pathway' was the most enriched. Although this is a different immune cell type, T cells and B cells are both lymphoid cells and would share many genes. Some pathways shared between the chromatin accessibility analysis and the histone modification analyses include 'downstream signaling in naive $CD8^+$ T cells', 'Graft-versus-host disease', and 'T cell receptor signaling pathways'. Many of the enriched pathways are related to the immune response, which is expected for $CD8^+$ cells.

$CD14^+$ Monocytes

The most enriched pathway from the ATAC analysis in $CD14^+$ cells was 'cytokine activation of matrix metalloproteinases', which is unsurprising as matrix metalloproteinases are modulators of inflammation (Nissinen and Kähäri 2014). Another pathway that was significantly enriched was 'hematopoietic cell lineage' and was only enriched in $CD14^+$ cells. As $CD14^+$ cells are of the hematopoietic lineage and T cells are not, this is an expected finding. 'FcγR-mediated phagocytosis' was also enriched in the ATAC analysis. $CD14^+$ cells have the ability to phagocytose cells and bacteria, while T cells do not (Devitt et al. 1998; Schiff et al. 1997). Therefore, this pathway is expected to be enriched in $CD14^+$ cells and not in $CD4^+$ and $CD8^+$ T cells. Some of the most significantly enriched pathways were GCPR signaling pathways, following the

analysis with H3K27ac and H3K4me3 overlapped genes. GPCRs are essential receptors for cell signaling and activation in a wide variety of biological contexts. They have been found to be involved within the immune response as immune cells respond to stimulation from multiple types of GPCRs to initiate a pro-inflammatory response (Lämmermann and Kastenmüller 2019). Importantly chemokine receptors are GPCRs which can lead to activation, migration, and affinity for certain molecules related to the immune response in monocytes and other cell types (Lämmermann and Kastenmüller 2019; Wang 2018). Some other enriched pathways shared between the ATAC and ChIP analyses included ECM pathways, and signaling by interleukins, all of which are known to be involved in the immune response.

5.3.3 Uncovering Putative Enhancers in AS Patients

Around 98% of the genome is non-coding and over 90% of GWAS variants fall within these regions (Maurano et al. 2012). Therefore, understanding the role of the non-coding genome by combining transcriptomic and epigenetic datasets is essential in determining the genetic aetiology of disease. The approach in this chapter was to define AS associated regulatory regions by combining ATAC-Seq, ChIP-Seq, and RNA-Seq datasets from patients and controls. This approach has been used successfully to uncover non-coding drivers of disease. *PAX6* has been associated with diabetes through GWAS but the variants and regions modulating its expression remained unclear, until one study performed RNA-Seq, and 3C, and then combined this information with ChIP-Seq data (Buckle et al. 2018). They were able to define two regulatory elements in a gene desert that modulate *PAX6* in pancreatic cells. Unfortunately, the analysis presented in this chapter showed that none of the AS GWAS associated loci overlapped with differential regions of open chromatin, H3K27ac, or H3K4me3 between AS patients and controls. However, some loci were highlighted for future functional genomics work.

The integrated analysis across epigenetic marks and RNA-seq data uncovered two putative enhancers, and one differentially active promoter in $CD14^+$ cells between patients and controls. The most significant differential H3K27ac enhancer peak in $CD14^+$ cells is situated over the promoter *THBS1*, indicting that it is more active in AS patients, which is further evidenced by the gene being positively differentially expressed in AS patients. This peak is also situated in open chromatin and coincides with the H3K4me3 promoter mark.

THBS1 encodes an ECM glycoprotein involved in cell adhesion and interactions, and is a modulator of the inflammatory response in $CD14^+$ macrophages. It exhibits both anti- and pro-inflammatory functions through binding cell surface receptors to regulate cytokine secretion (Lopez-Dee et al. 2011). THBS1 interacts with CD47 receptors on $CD14^+$ cells to inhibit the CD47-CD14 interaction, and therefore inhibits the induction of LPS and reduces the activation of NF-κB and other pro-inflammatory genes, such as IL-1β (Stein et al. 2016). Binding of THBS1 can also induce production and secretion of the anti-inflammatory cytokine, IL-10, which is associated with injury clearance and recovery (Zhao et al. 2014). Furthermore, extracellular THBS1 plays a role in macrophage recognition and phagocytosis of apoptotic cells through binding of the cell surface receptor CD36 (Savill et al. 1992). *THBS1* and associated pathways are already implicated in tumor growth and cancer (Daubon et al. 2019). These results suggest it may play role a role in AS pathology, and presents an interesting target for further research in AS and autoimmunity. It would make an excellent candidate for CRISPR/Cas9 functional studies to determine the effects of gene knockout and also find variants or regulatory regions that may modulate its expression.

In addition to changes in *THBS1* gene regulation, two putative enhancers were uncovered by combining epigenetic and transcriptomic data. Differential H3K27ac peaks within *ANO10* and *CABIN1* were located within the introns of

these genes, which is consistent with them being enhancers. These peaks were larger in AS patients which indicates that these putative enhancers are stronger in patients than controls. Moreover, both genes were up-regulated in AS patients. The H3K27ac peaks also lie within open chromatin. Therefore, it is likely that these regions are enhancers for the genes they are situated within. Continued analysis including chromatin capture techniques would help further associate these putative enhancer regions with their target genes. These peaks are not currently annotated as being enhancers in publicly available data.

ANO10 encodes a family member of a transmembrane calcium-activated chloride channel. One of its family members, *ANO6*, has been implicated in innate immunity and found to play a role in macrophage phagocytosis (Ousingsawat et al. 2015). One study performed a GWAS for *Borrelia* infection, which is the bacteria that causes Lyme disease (Hammer et al. 2015). They found a coding mutation within *ANO10* to be associated with *Borrelia* infection and further implicated *ANO10* with normal macrophage migration and phagocytosis, and mutations within it decrease macrophage clearance of bacteria.

CABIN1 has been strongly associated with T cell receptor signaling. Calcineurin, a phosphatase, activates T cells during an immune response, and CABIN1 binds to calcineurin to inhibit this response (Sun et al. 1998). One study implicated calcineurin with RA and found its activity to be increased in patient synoviocytes and suggested this pathway as a target for treatment of rheumatic disease (Yoo et al. 2006).

Though *THBS1*, *ANO10*, and *CABIN1*, have never been explicitly associated with AS through GWAS, they could be new targets for AS research. GWAS findings can be important in understanding disease, but assaying the transcriptomic and epigenetic landscape in AS patients has led to uncovering new regulatory elements and genes that are involved in the immune response and could present promising targets for CRISPR/Cas9 mediated functional studies.

5.3.4 Limitations of Patient Vs Control Differential Analysis

As discussed in Chapter 1, AS is clinically heterogeneous with varying levels of disease activity. Although patients recruited for this study had active disease, it varied between individuals was partly assessed by BASDAI scores, which is a subjective measure of pain. Differences in the degree of disease activity may have impacted the results of the differential analysis between AS patients and controls. All patients recruited were biologic naive, but some were using different NSAIDS to relieve symptoms. NSAID usage could affect the inflammation occurring within a patient and alter the results shown here. Although none of the patients had taken an NSAID when their blood was drawn, continual usage of these drugs could alter immune cell counts and gene regulation. The results of this analysis could be improved and made more robust in the future by including more samples and being more stringent with patient recruitment, or stratify patients to make sure every patient meets the same criteria. The three different methods of ATAC-Seq used could have also affected the analysis, so a larger amount of samples processed using Omni-ATAC-Seq could help reduce noise in the samples. Additionally, Capture-C on AS patients could be integrated with the ATAC and ChIP data to provide another level of understanding to the epigenetic landscape in AS patients.

5.4 Conclusions

Understanding the epigenetic landscape between cell types and patients and controls is vital to understanding AS pathology. Integrating differential gene expression and differential epigentic marks between AS patients and controls uncovered three novel regulatory elements modulating immune relate genes, *THBS1*, *ANO10*, and *CABIN1*. None of these genes were associated with AS through GWAS, highlighting the importance of using functional genomics to

assay the epigenetic landscape. The work carried out here to define chromatin accessibility and active enhancer and promoter locations has identified new genes and putative enhancers for future functional studies with CRISPR/Cas9 to uncover the molecular mechanisms driving disease and provide a new understanding of AS pathology.

6

General Discussion

Science and everyday life cannot and should not be separated. -Rosalind Franklin

6.1 Model Systems for AS Biology 259
6.2 CRISPR/Cas9 as a Functional Genomics Tool 261
6.3 Integration of -Omics Data in AS Patients 266
6.4 Future Work 267
6.5 Conclusions 269

In this thesis, two novel putative regulatory elements at the *TNFRSF1A* locus were characterized using CRISPR/Cas9 in an iPSC derived macrophage model to understand genetic variation associated with AS. This thesis also explored the epigenetic landscape associated with AS to annotate putative causal variants and to define epigenetic changes in chromatin accessibility or histone modifications associated with regulatory elements that could be driving disease. The work presented here has been conducted to better understand the underlying genetic and molecular mechanisms involved in AS pathogenesis. An overview of the main findings of this thesis are discussed below, along with suggestions for future work to enhance these results and improve the understanding of AS aetiology.

6.1 Model Systems for AS Biology

The work in Chapter 3 has illustrated the difficulties in finding an appropriate cellular model system to study disease biology. The myeloid-like immortalized cell line, K562 cells, were easy to grow and transfect for CRISPR/Cas9

experiments, and commercial lines constitutively expressing Cas9 can be purchased. However, they were found to be primarily triploid with many chromosomal translocations and rearrangements. The CRISPR/Cas9 deletion experiments performed on them yielded no cells homozygous for the deletions of interest, and the bulk edited cell population showed no discernible change in gene expression. The K562 cell populations were also naturally genetically heterogeneous, making it hard to discern the proportion of edited cells. This cell line may be good for testing newly designed gRNAs, as evidence of large deletions are visible after a simple PCR and gel electrophoresis assay on a bulk population to determine that a gRNA worked, but the effects of the edits were too difficult to ascertain in this cell line.

iPSC derived macrophages presented a superior alternative to K562 cells, but they were not without their limitations. As noted by other scientists, the iPSC derived macrophages are a primitive macrophage and are not the same as $CD14^{+}$ macrophages derived from blood monocytes (Buchrieser et al. 2017). An advantage of this protocol is that cells can be harvested every week for experiments, but cells have been found to differ between harvests and cells taken from older factories may yield different results from newer ones (Buchrieser 2016). Therefore, generating replicates of these cell lines can be challenging. No replicates were performed on cells from different harvests for the RNA-Seq, ATAC-Seq, or ChIP-Seq analysis presented in Chapter 4, ensuring all cells used for a single sequencing experiment were taken from the same harvest to simplify the analysis. However, replicates would be required for future experiments to validate the results. Additionally, some experimental protocols call for a large number of cells, and sometimes iPSC derived macrophages cannot meet this demand. The number of cells harvested each week varied, typically with around 3 to 5 million cells available for an experiment, but decreased with factory age. The Capture-C experiment performed in Chapter 4 was

unsuccessful on iPSC derived macrophages due to cell numbers, as Capture-C requires around 10 million cells per cell line being investigated. ATAC-Seq experiments require around 50,000 cells and ChIPm requires around 100,000, which is much more easily obtained, especially when dividing the harvested cells for multiple stimulation treatments. Despite these limitations, iPSC derived macrophages are easier to work with than primary cells and are similar in their function, phenotype, and transcriptome to their blood derived counterparts and recapitulate macrophage biology reasonably well (Alasoo et al. 2015; Zhang et al. 2015; Buchrieser et al. 2017; Cao et al. 2019). They also have the advantage of being an isogenic model for CRISPR/Cas9 studies to investigate disease associated loci, which is not currently possible with primary cells.

6.2 CRISPR/Cas9 as a Functional Genomics Tool

GWAS associate many genetic variants with disease, but functional studies are required for distilling correlation and causation for genetic traits. Experiments to determine the impact of genetic variation on gene expression and their mechanism of action are essential in establishing disease causality. One paper, by Chakravarti et al., has established criteria for determining causality which they call "Koch's Postulates for Complex Human Diseases and Traits" and are as follows: 1) "Candidate gene variants are enriched in patients", 2) "Disruption of the gene in a model system gives rise to a phenotype that is accepted as relevant to the human phenotype", 3) "Phenotype in model system can be rescued with wild-type human alleles", and 4) "Phenotype in model system cannot be rescued with mutant human alleles" (Chakravarti et al. 2013). GWAS can be used to find genetic variants that are enriched in disease patients and therefore satisfy Chakravarti et al.'s first criterion, but no others. CRISPR/Cas9 is an invaluable tool that can be used to both generate and correct disease associated variants or genes to fully establish a causal relationship between genetic trait and disease.

6.2.1 CRISPR/Cas9 to Generate Disease Associated Alleles

Chapters 3 and 4 explored using CRISPR/Cas9 in a model system to characterize two putative enhancers within the *TNFRSF1A* AS GWAS associated locus. *TNFRSF1A* was prioritized using a new computational pipeline, Pi, fine mapping, and annotation from publicly available data. Two SNPs within this region, rs887477 and rs11064145, appeared to be in putative enhancers that had not been functionally characterized. These putative enhancers were selected for examination to establish the genes and pathways they modulate and provide a mechanism of action for how these SNPs and the *TNFRSF1A* locus may relate to disease.

The method employed was to design two gRNAs for each deletion region that would recognize and cut at the boundaries of this region. The DNA would then be repaired using NHEJ, and the desired deletion region would be excised. This method of using two gRNAs to excise regions of DNA was also used successfully in other studies (Liu et al. 2016). The two putative enhancers and four exons of *TNFRSF1A* were attempted to be removed in this way, but only the enhancer deletions were successful. Initially, the SNPs, rs887477 and rs11064145, were also going to be generated in iPSC for functional characterization, but this experiment never occurred due to time constraints.

Uncovering a Novel Enhancer

The Intronic Enhancer, which is situated within the intron of *TNFRSF1A* and encompasses rs887477, was successfully deleted in three clones of an iPSC cell line that were differentiated into macrophages. Its deletion resulted in a decrease in *TNFRSF1A* expression, consistent with the hypothesis that it is an enhancer for this gene. The Intronic Enhancer was also found to be situated within a region of open chromatin and a region of H3K27ac, which denotes active enhancers and further confirms this hypothesis. Therefore, CRISPR/Cas9 was used to

successfully validate the presence of an enhancer for *TNFRSF1A*. Unfortunately, the SNP rs887477 was not generated, so it is unclear whether it impacts enhancer function. However, because it is situated within this enhancer, it could possibly be the SNP driving the GWAS association at this locus. The A allele of rs887477 is situated within the predicted binding motif for the transcription factors, TFII-I and PPARα, while the C allele alters the binding motif to bind those two factors as well as MYC Associated Zinc Finger Protein (MAZ; AKA Serum Amyloid A Activating Factor 1). There is no explicit connection between MAZ and *TNFRSF1A* regulation, but it is thought to be involved in the regulation of inflammation induced serum amyloid A (*SAA*) family expression. SAA proteins are up-regulation during an inflammatory response and play a role in cytokine synthesis and is a chemotactic to recruit immune cells (Ye and Sun 2015). They have also been implicated in RA (O'Hara et al. 2000; Vallon et al. 2001). Therefore, rs887477 could have an immunomodulatory role through the generation of a binding motif for MAZ, but functional validation is required.

Uncovering a Novel Suppressor

The Intergenic Enhancer, which is situated approximately 3.5kb upstream of *TNFRSF1A* and encompasses rs11064145, was successfully deleted in two clones of an iPSC cell line that were differentiated into macrophages. Despite being predicted to act as an enhancer, the results from the functional analysis in Chapter 4 suggest that it may be a gene silencer or repressor instead. It was not found to be situated within strong peaks representing open chromatin and was not a region enriched for the enhancer mark, H3K27ac. When deleted, however, it did result in an up-regulation of *NCAPD2*, 150kb away, consistent with it being a repressor and not an enhancer. This type of functional annotation would not have been possible without CRISPR/Cas9 editing. Its use has defined a novel regulatory element that could be acting at a distance to modulate gene expression. The G allele for rs11064145 creates an additional binding motif for

GC-Rich Sequence DNA-Binding Factor 2 (GCF). In vitro studies have shown that GCF is a transcriptional repressor and has been found to repress the epidermal growth factor (EGF) receptor, β-actin, and the protease CANP (Beguinot et al. 1995). Though there is no obvious link to *NCAPD2*, this provides some evidence for how rs11064145 may modulate gene expression, although it would need to be generated in a cell line for conclusive results.

TNFRSF1A in Primary Cells

The *TNFRSF1A* locus was also examined in $CD4^+$, $CD8^+$, and $CD14^+$ primary cells derived from AS patient and healthy control blood in Chapter 5. ATAC-Seq and ChIP-seq performed on $CD14^+$ cells confirmed what was seen in the iPSC derived macrophages. In primary $CD14^+$ cells the Intronic Enhancer is situated within a region of open chromatin and is a strong enhancer based on H3K27ac and H3K4me3 peaks. The Intergenic Enhancer was situated within open chromatin, but not within an active enhancer. This locus also appeared promising for further investigation in $CD4^+$ and $CD8^+$ T cells.

6.2.2 CRISPR/Cas9 to Correct Disease Associated Alleles

Though not attempted in this thesis, another important way to study disease associated variants and determine causality is to correct variants or genes in a disease model. iPSCs derived from patients are a favourable model for this as they can be derived from patient cells, including PBMCs, to create an isogenic disease model for single variant investigation. Correcting alleles through CRISPR/Cas9 editing, or by adding back a non-functional gene or protein has been used successfully in a number of studies to uncover a causal relationship between genetic variants and diseases (Awad et al. 2015; Chung et al. 2013; Soldner et al. 2016; Boland et al. 2017; Murakami et al. 2017; Xu et al. 2017). One study used CRISPR/Cas9 in iPSC derived cortical neurons to associate SNPs with Alzheimer's disease (Paquet et al. 2016). They corrected SNPs in the genes

APP and *PSEN1*, which was able to rescue to the disease phenotype of amyloid-β protein secretion. This type of work could be an intriguing next step for AS research and help further validate associated genes and variants. However, with AS being such a heterogeneous disease with multiple cell types involved in its pathogenesis, picking an appropriate phenotype to measure is a challenge.

6.2.3 Challenges of CRISPR/Cas9 Mediated Genome Editing

Generating homozygous clones for the deletions of interest was a major challenge in this thesis. None were generated in K562 cells and mosaicism was a prominent issue in iPSCs, requiring much optimization to generate homozygous clones. The deletion of the *TNFRSF1A* exons was never successfully generated in a homozygous state. The RNP method used to generate homozygous clones was very new when this thesis work began, but is now often preferred over plasmids (Kim et al. 2014). It reduced the incidence of mosaicism and is highly efficient at editing iPSCs, so would be a promising approach for future experiments. Unfortunately, only two or three cell lines harbouring deletions of interest were developed, which made analysis difficult for experiments such as RNA-Seq. Statistically significant results are difficult to obtain from comparisons with so few samples. The lack of edited cell lines is partly due to the difficulty in generating them using current CRISPR/Cas9 technology, and also to practical restraints in differentiating a large amounts of macrophages from iPSCs.

In addition to mosaicism issues, CRISPR/Cas9 technology still suffers from off-target effects (Chen 2019). Although they were not specifically examined in this work, off target mutations may have been generated that could have impacted the results. Continual culture of iPSCs and sub-cloning can also lead to variant generation and cause the genotypes of the CRISPR/Cas9 edited cells to differ from the parental cell line, beyond the desired edit. Increasing the number of replicates of edited clones and controls could reduce the likelihood of unwanted

mutations altering the results of the analysis. Improved CRISPR/Cas9 techniques could also improve off-tagerts. New approaches for CRISPR/Cas9 mediated editing are being developed continuously. Recently, an approach entitled 'prime editing' was released which fuses Cas9 to a reverse transcriptase and a prime editing gRNA, which eliminates the need for a donor template (Anzalone et al. 2019). CRISPR/Cas9 could become much more efficient in the future and eliminate the issues of mosaicism and off-target effects, and generate many homozygous cell lines to improve functional studies.

6.3 Integration of -Omics Data in AS Patients

Chromatin modifications play an important role in modulating gene expression. Different epigenetic and transcriptomic datasets can be overlapped to generate hypotheses about how a variant or a region may be impacting gene expression and contribute to disease. One single dataset containing regions of open chromatin provides a limited amount of information, but when it is combined with histone marks, expression data, and chromatin interactions, then regulatory regions and target genes can be discerned. One study used this approach to great effect in osteoarthritis (Liu et al. 2018). They performed ATAC-Seq on patient knee cartilage samples to find differential regions of open chromatin. They then combined these results with histone modification, DNA-methylation, and gene expression data to uncover disease specific enhancers and target genes. Another recent study was able to use these types of data to define B cell subtypes and identify a new B cell mechanism related to the pathogenesis of systemic lupus erythematosus by combining chromatin accessibility and histone mark locations with transcriotomic data for five B cell subtypes (Scharer et al. 2019).

The analysis in Chapter 5 revealed differential regions of open chromatin as well as the locations of differential active enhancers (H3K27ac) and promoters (H3K4me3) across three immune cell types. When combined with transcriptomic

data, two differential enhancers and one differentially active promoter were found in $CD14^+$ cells. This combined analysis can lead to the identification of new targets for downstream functional studies involving CRISPR/Cas9.

One limitation is that enhancer activity and gene expression exhibit context specificity. The analysis in Chapter 4 illustrated that many enhancers and promoters change their activity upon LPS stimulation. This is in important consideration when studying immune mediated disorders. The analysis performed here was carried out on primary $CD4^+$, $CD8^+$, and $CD14^+$ cells to find any differential regions exhibiting cell type context specificity, but they were all in an unstimulated state. Stimulation with LPS or pro-inflammatory cytokines may uncover novel regulatory elements that are active only during an immune response. One recent study assayed the chromatin accessibility landscape in immune cell types in both unstimulated and stimulated states (Calderon et al. 2019). They found that stimulation greatly impacted the chromatin landscape, and identified genetic variants that influenced chromatin accessibility, but only under stimulated conditions, thus highlighting the importance of examining various cellular states when establishing causal disease variants.

6.4 Future Work

6.4.1 CRISPR/Cas9 Experiments

The methods employed in thesis of prioritizing AS associated variants, generating enhancer deletions with CRISPR/Cas9, and performing functional genomic analysis on edited cell lines was successful, but incredibly low-throughput. Multiple studies have used high-throughput techniques for assaying enhancers by using CRISPR/Cas9 screens (Korkmaz et al. 2016; Simeonov et al. 2017). This approach could be considered for the future to identify important enhancers or other regulatory regions. This would improve the throughput of the experiment

and allow for the interrogation of many genomic regions. Regions identified in the screen could then be examined further using the approach applied in Chapters 3 and 4 to fully characterize them.

6.4.2 Additional Cell Types

Multiple immune cell types have been associated with AS and the analysis in Chapter 5 illustrated that $CD4^+$, $CD8^+$, and $CD14^+$ cells have specific epigenetic landscapes that coincide with AS GWAS regions. $CD14^+$ macrophages were the selected cell type for the functional CRISPR/Cas9 work carried out in Chapters 3 and 4 due to their disease relevance and feasibility to differentiate. However, the CRISPR/Cas9 edited iPSCs produced in Chapter 3 could be differentiated into other important immune cell types. Protocols have been published for the differentiation of iPSCs into B cells (Carpenter et al. 2011) and T cells (Timmermans et al. 2009; Nishimura et al. 2013; Iriguchi and Kaneko 2019) and would help explore the cell type specificity of the enhancers examined here, or any other future loci edited in iPSCs.

6.4.3 Continued Patient Recruitment

To improve upon the differential analysis between patients and controls performed in Chapter 5, patients are still being recruited. Patient samples included in further analysis are also under investigation to make sure all samples are from individuals with current active disease. The difficulty in assessing disease activity may be impacting the results of the differential analysis. In addition to RNA-Seq, ATAC-Seq and ChIP-Seq experiments, Capture-C, to assay physical chromatin interactions, is now being performed on patient samples to add another layer of data and help identify target genes of regulatory regions. Genotyping of patients and controls is also underway to perform an eQTL analysis, using RNA-Seq data, to further annotate and uncover disease

causing variants. Enhancer RNAs (eRNAs) are also being investigated between patients and controls to better understand differential enhancer activity. Other analytical methods, such as chromHMM, could be used to further investigate differences between patients and controls. chromHMM is a software tool to summate chromatin states by using different histone modifications, and could help uncover disease relevant chromatin states (Ernst and Kellis 2012). Only $CD4^+$, $CD8^+$, and $CD14^+$ cells were examined in this thesis, but $CD19^+$ B cells are also being collected as they may be implicated in disease (Chen et al. 2015; Miossec and Kolls 2012). Further ChIP-Seq experiments assaying other histone marks and CTCF regions are also being considered to be performed on patient samples. Increased sample numbers and integrated -omics datasets should help establish the underlying molecular mechanisms of genetic and epigenetic variation and improve the current understanding of AS pathology.

6.5 Conclusions

The overall goal of this thesis was to utilize functional genomic approaches to advance the understanding of the genetic aetiology of AS. The underlying genetic aetiology and pathology of complex diseases are difficult to elucidate. Functional genomic approaches were utilized in this thesis to examine AS associated genetic loci in both AS patients and within a model system. CRISPR/Cas9 genome editing was used to successfully annotate two regulatory regions within the *TNFRSF1A* GWAS locus in $CD14^+$ cells. Moreover, these same edited iPSCs can be explored in other immune cell types. Epigenetic and gene expression data were successfully overlaid in AS patients to identify novel genes and regulatory regions that may be associated with AS pathology. These results have highlighted the importance of functional experiments in a biological system to annotate SNPs and regulatory regions, and with improved CRISPR/Cas9 techniques, hopefully causal relationships between genetic variation and AS can be established.

A

Supplementary Data from Chapter 3

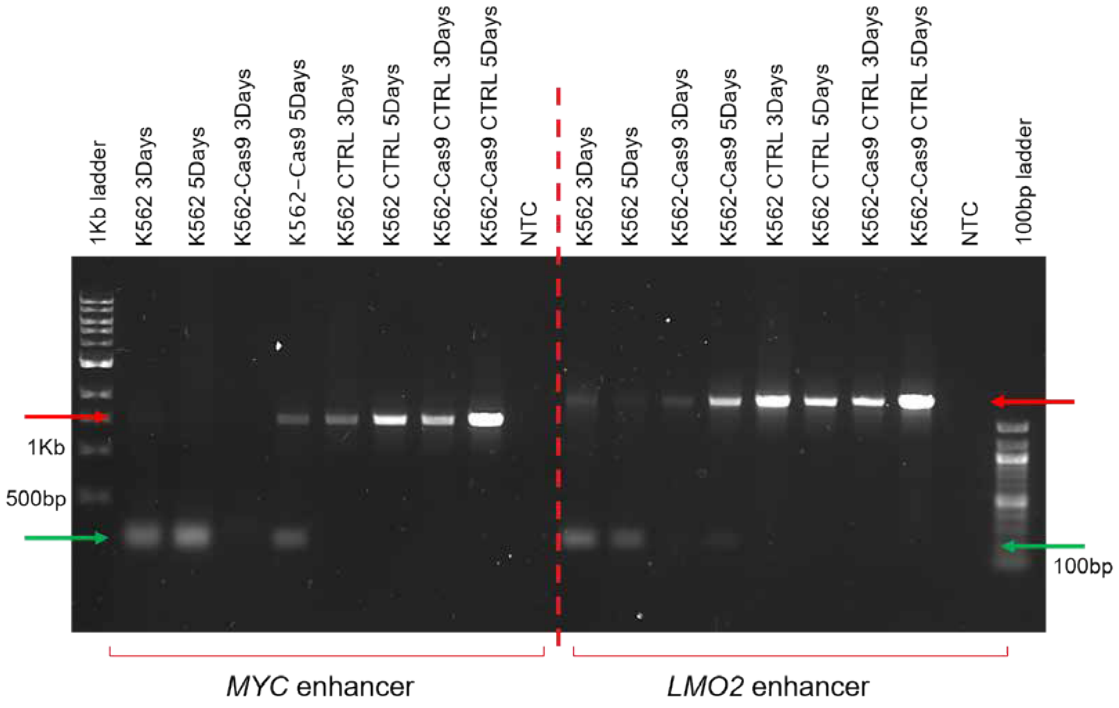


Figure A.1: Agarose gel indicating CRISPR/Cas9 deletion success in K562 cells (*MYC* and *LMO2*).
PCR primers were designed to amplify the deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. Two of the five enhancer deletions are shown here, enhancers for *MYC* and *LMO2*. DNA was extracted from a bulk cell population that had not been single cell sorted or clonally expanded. Transfections were done in both K562 and K562-Cas9 cells and were harvested at three days and five days post transfection. Unedited control cells for both cell lines were harvested and process at the same time as the transfected cells, labeled here as CTRL. If the *MYC* enhancer was successfully deleted it would show a band at approximately 215bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 1557bp (red arrow). If the *LMO2* enhancer was successfully deleted it would show a band at approximately 170bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 2043bp (red arrow). Two bands in a lane showing both the deleted and undeleted product suggest multiple genotypes exist in the cell population.

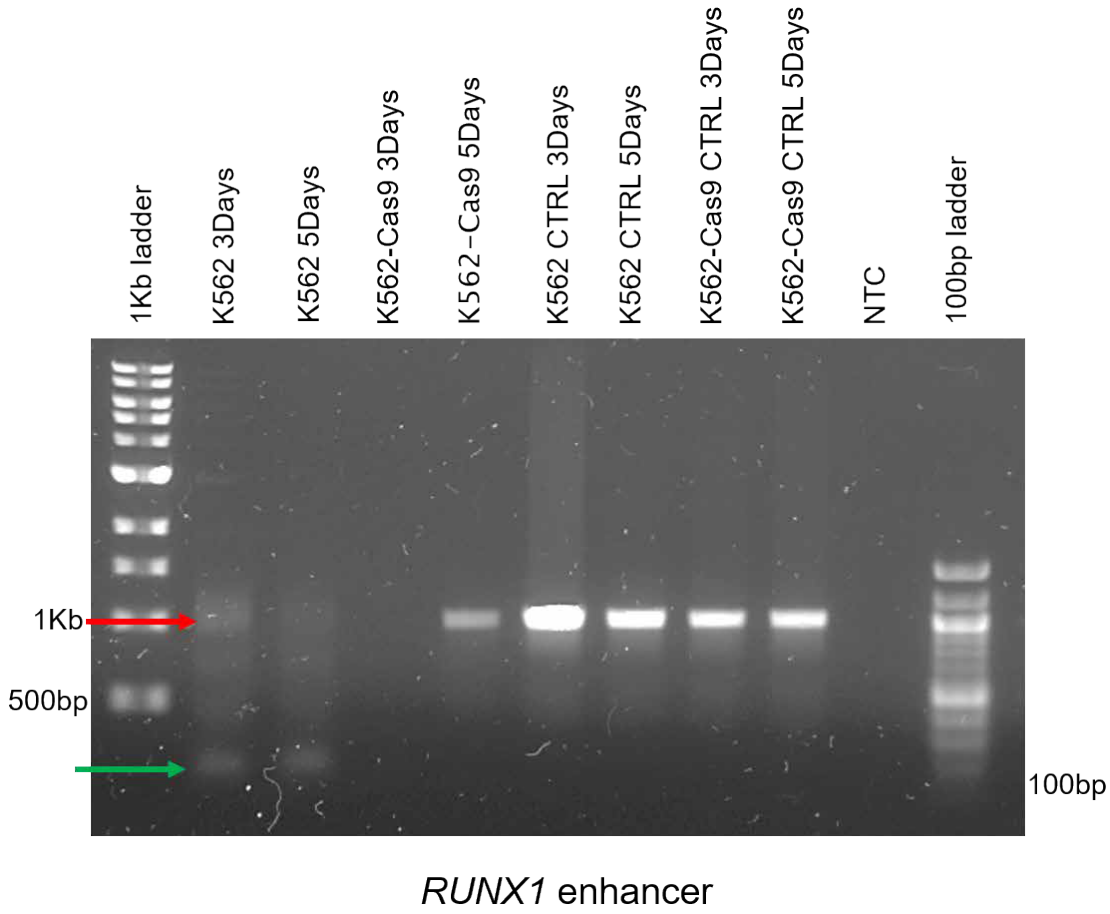


Figure A.2: Agarose gel indicating CRISPR/Cas9 deletion success in K562 cells (*RUNX1*)
PCR primers were designed to amplify the deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. One of the five enhancer deletions are shown here, enhancers for *RUNX1*. DNA was extracted from a bulk cell population that had not been single cell sorted or clonally expanded. Transfections were done in both K562 and K562-Cas9 cells and were harvested at three days and five days post transfection. Unedited control cells for both cell lines were harvested and process at the same time as the transfected cells, labeled here as CTRL. If the *RUNX1* enhancer was successfully deleted it would show a band at approximately 200bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 1023bp (red arrow). Two bands in a lane showing both the deleted and undeleted product suggest multiple genotypes exist in the cell population.

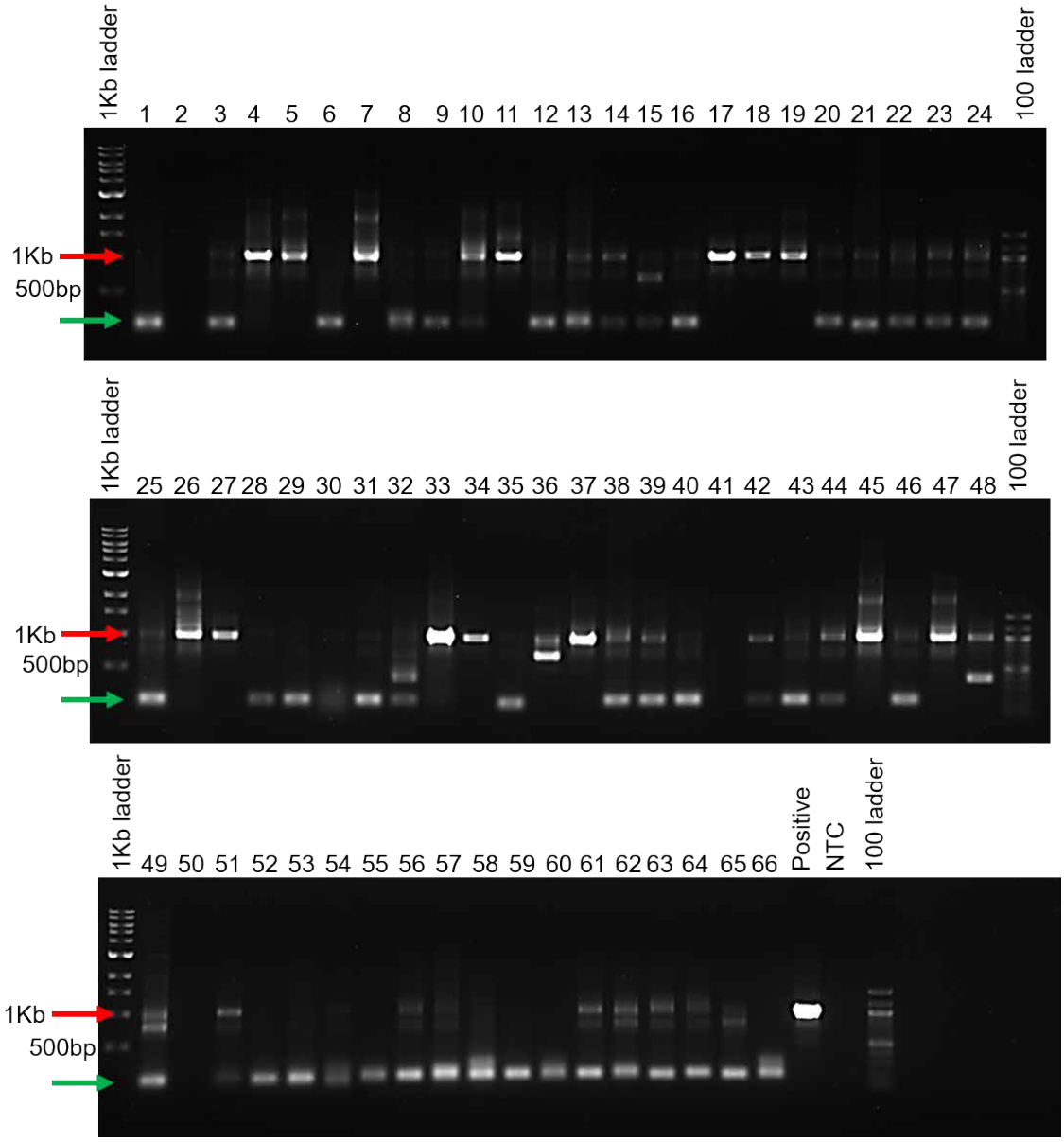


Figure A.3: Agarose gel indicating CRISPR/Cas9 *RUNX1* enhancer deletion success in K562 edited cell clones.
PCR primers were designed to amplify the deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. If the *RUNX1* enhancer was successfully deleted it would show a band at approximately 200bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 1023bp (red arrow). All 66 *RUNX1* Enhancer Deletion clones were screened, with 46 clones showing signs of a deletion (70%). Unedited K562 cells were used as a positive control. Two bands in a lane showing both the deleted and undeleted product suggest multiple genotypes exist in the cell population.

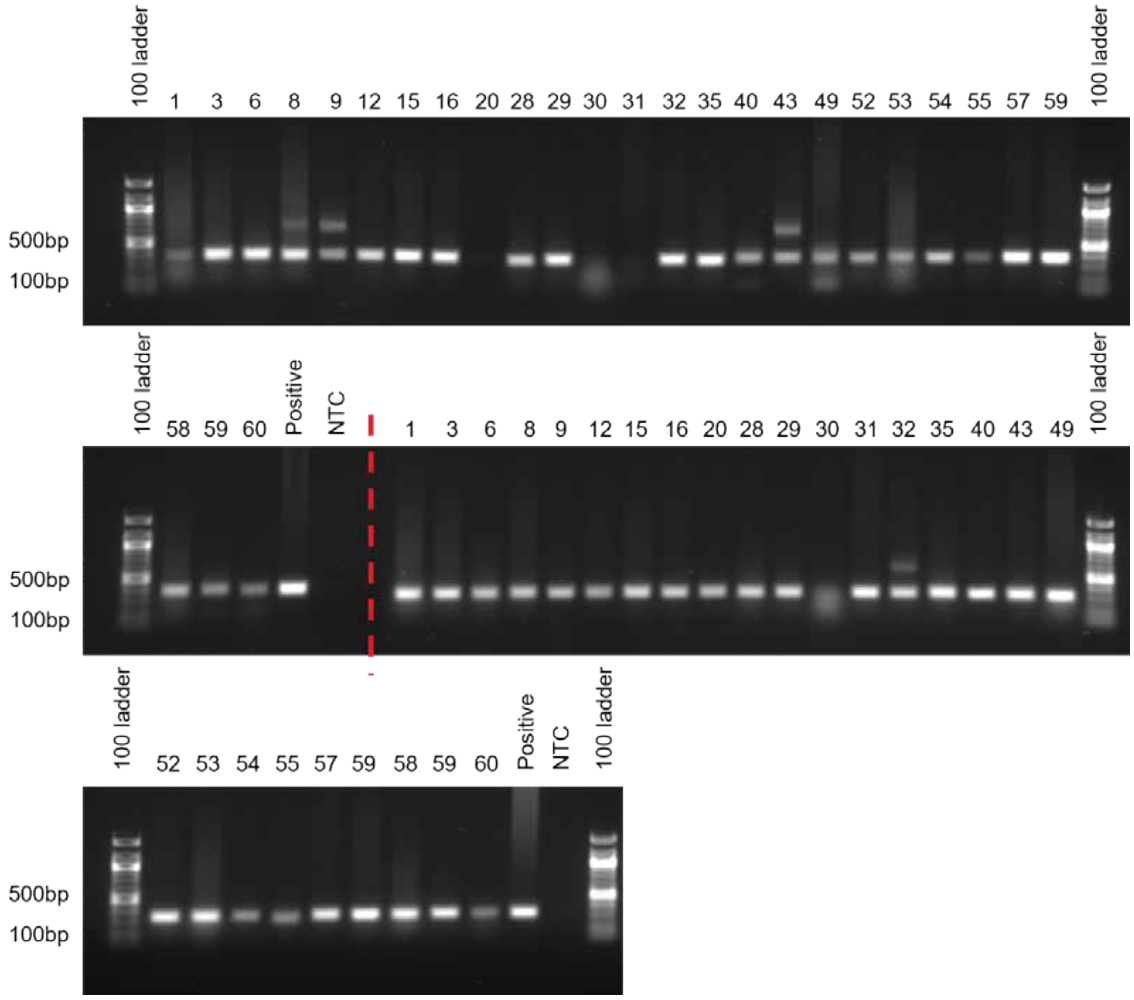


Figure A.4: Agarose gel indicating zygosity of K562 edited cell clones.
To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region, and one binding just outside. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The 5' end of the deletion region is shown first. If the clone is heterozygous it would amplify and show a band at 374bp. The 3' end of the deletion region is shown after the red dotted line. If the clone is heterozygous it would amplify and show a band at 310bp. Unedited K562 cells were used as a positive control. None of the clones appear homozygous.

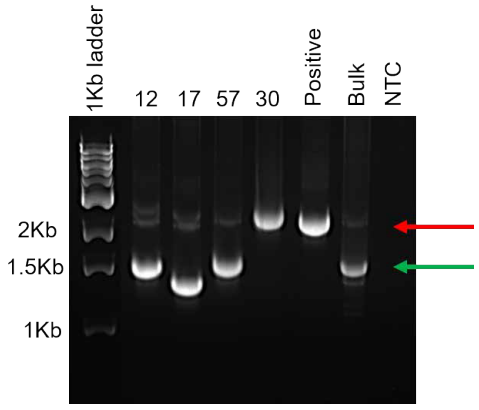


Figure A.5: Agarose gel indicating showing deletion region of K562 edited cell clones using long range primers.
Long range primers binding further away from the deletion region were used to amplify the deletion region. If the clones had no deletion a band would be present at 2347bp, and if the clone did have the deletion then a band would visible at 1644bp. Unedited K562 cells were used as a positive control. The bulk edited cell population is also shown.

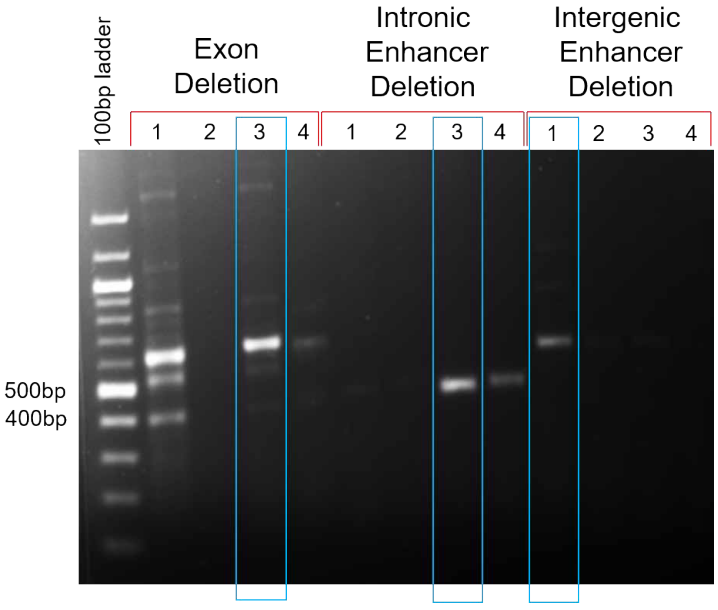


Figure A.6: Agarose gel showing gRNA test in HEK-293 cells.
Several gRNA constructs were tested in HEK-293 cells for editing efficiency. Four combinations were tested for each deletion and the bulk edited cell populations of HEK-293 cells were screened for the presence of a deletion using PCR and gel electrophoresis. The amplified fragments from the Exon Deletion cells would appear at 2005bp if the deletion had not occurred and at 592bp if it had. The amplified fragments from the Intronic Enhancer Deletion cells would appear at 4005bp if the deletion had not occurred and at 414bp if it had. The amplified fragments from the Intergenic Enhancer Deletion cells would appear at 1065bp if the deletion had not occurred and at 552bp if it had. The blue boxes indicate the gRNAs that were selected.

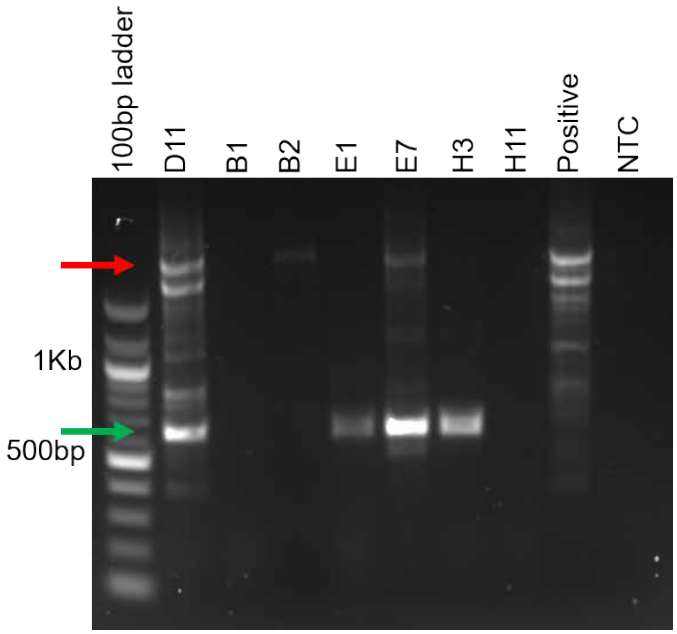


Figure A.7: Agarose gel indicating Exon Deletion success in edited iPSC clones. PCR primers were designed to amplify the entire deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. If the Exon Deletion was successful it would show a band at approximately 592bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 2005bp (red arrow). The clones shown here were the clones from the initial 192 that showed signs of the deletion.

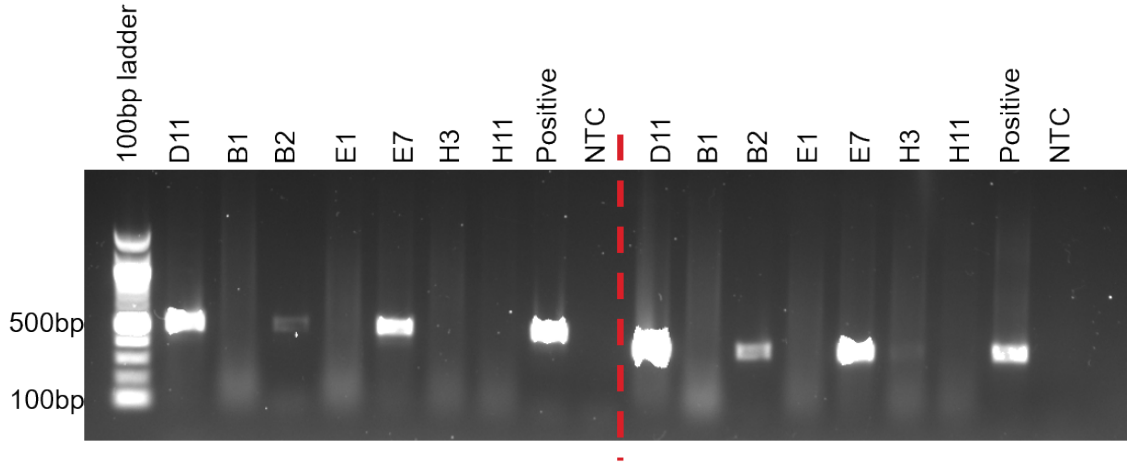


Figure A.8: Agarose gel indicating zygosity of the Exon Deletion in edited iPSC clones. To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region, and one binding just outside. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The 5' end of the deletion region is shown first. If the clone is heterozygous it would amplify and show a band at 509bp. The 3' end of the deletion region is shown after the red dotted line. If the clone is heterozygous it would amplify and show a band at 424bp. Unedited iPSCs were used as a positive control.

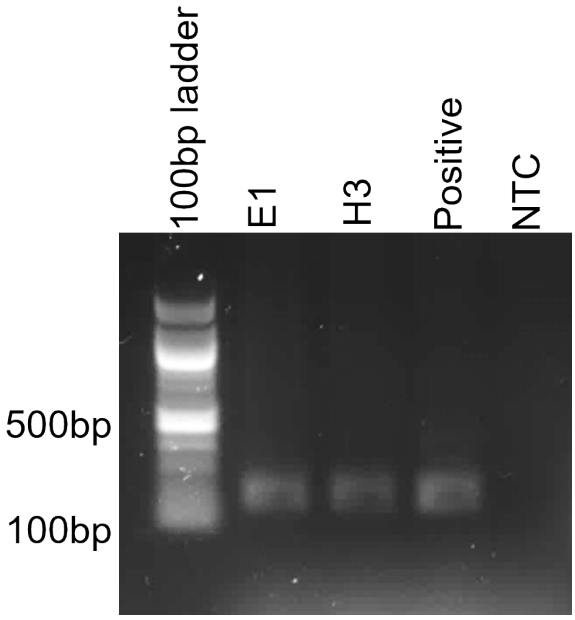


Figure A.9: Agarose gel indicating zygosity of the Exon Deletion in edited iPSC clones using internal primers.
To confirm heterozygosity, alternate primer sets were used with both primers lying within the deletion region. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The expected amplicon size is 163bp. Unedited iPSCs were used as a positive control. None of the clones appear homozygous.

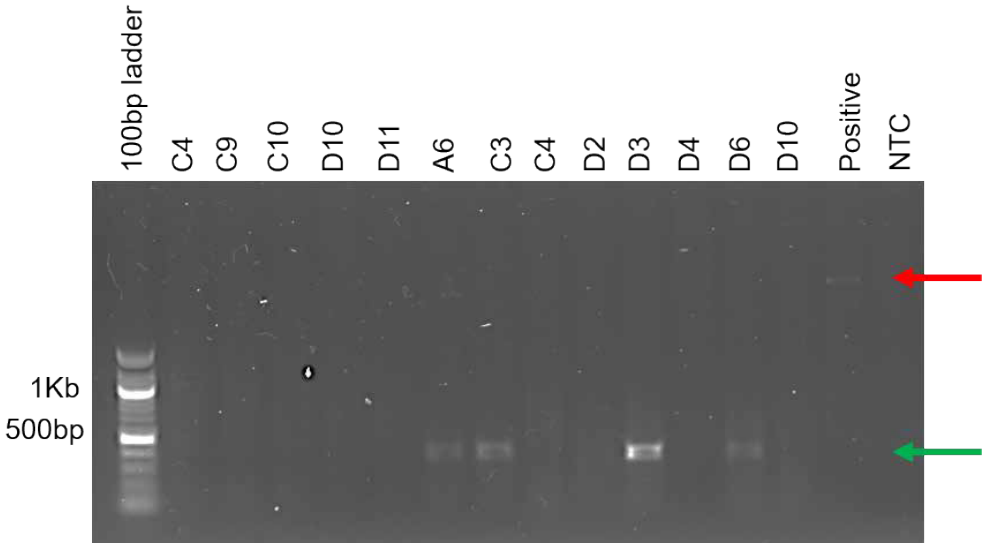


Figure A.10: Agarose gel indicating Intronic Enhancer Deletion success in edited iPSC clones.
PCR primers were designed to amplify the entire deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. If the Exon Deletion was successful it would show a band at approximately 414bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 4005bp (red arrow). The clones shown here were the clones from the initial 192 that showed signs of the deletion.

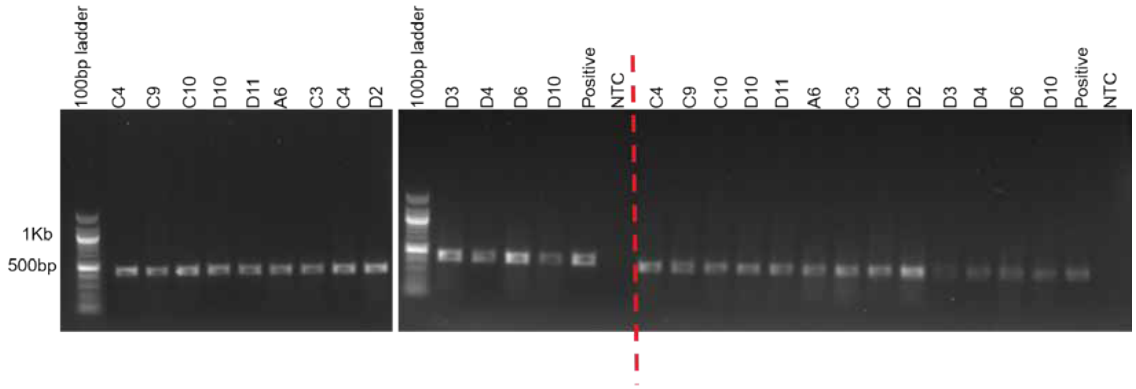


Figure A.11: Agarose gel indicating zygosity of the Intronic Enhancer Deletion in edited iPSC clones.
To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region, and one binding just outside. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The 5' end of the deletion region is shown first. If the clone is heterozygous it would amplify and show a band at 438bp. The 3' end of the deletion region is shown after the red dotted line. If the clone is heterozygous it would amplify and show a band at 377bp. Unedited iPSCs were used as a positive control.

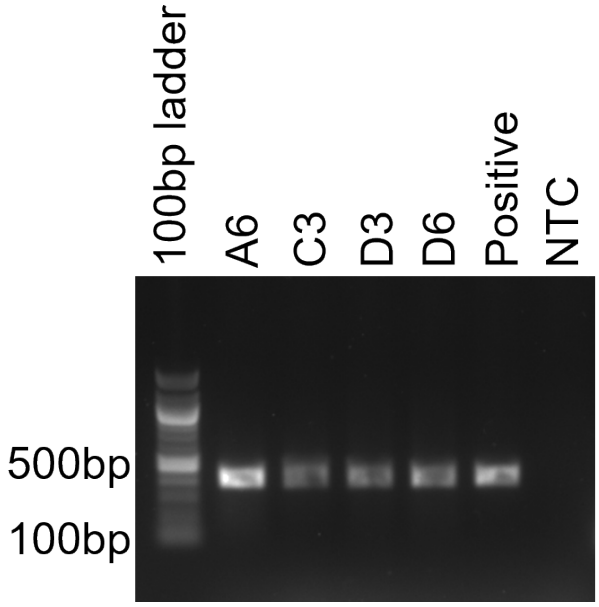


Figure A.12: Agarose gel indicating zygosity of the Intronic Enhancer Deletion in edited iPSC clones using internal primers.
To confirm heterozygosity, alternate primer sets were used with both primers lying within the deletion region. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The expected amplicon size is 394bp. Unedited iPSCs were used as a positive control. None of the clones appear homozygous.

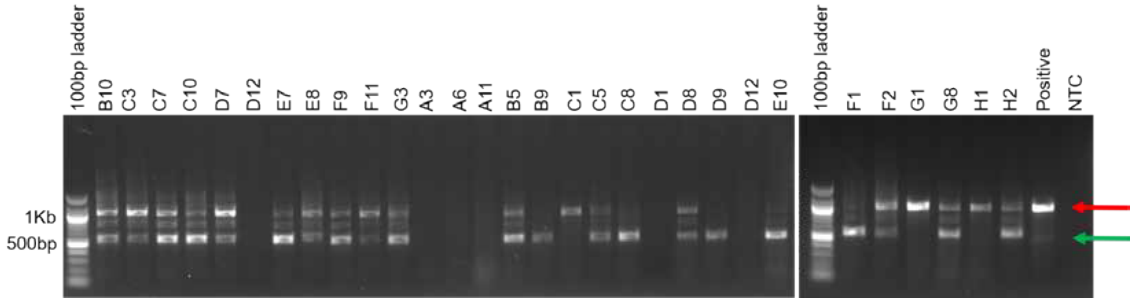


Figure A.13: Agarose gel indicating Intergenic Enhancer Deletion success in edited iPSC clones.
PCR primers were designed to amplify the entire deleted region using PCR. If the correct deletion occurred, the amplified region of DNA would appear as a smaller fragment when run on an agarose gel. If the Exon Deletion was successful it would show a band at approximately 552bp (green arrow), but if the enhancer was not deleted the product would be seen at approximately 1065bp (red arrow). The clones shown here were the clones from the initial 192 that showed signs of the deletion.

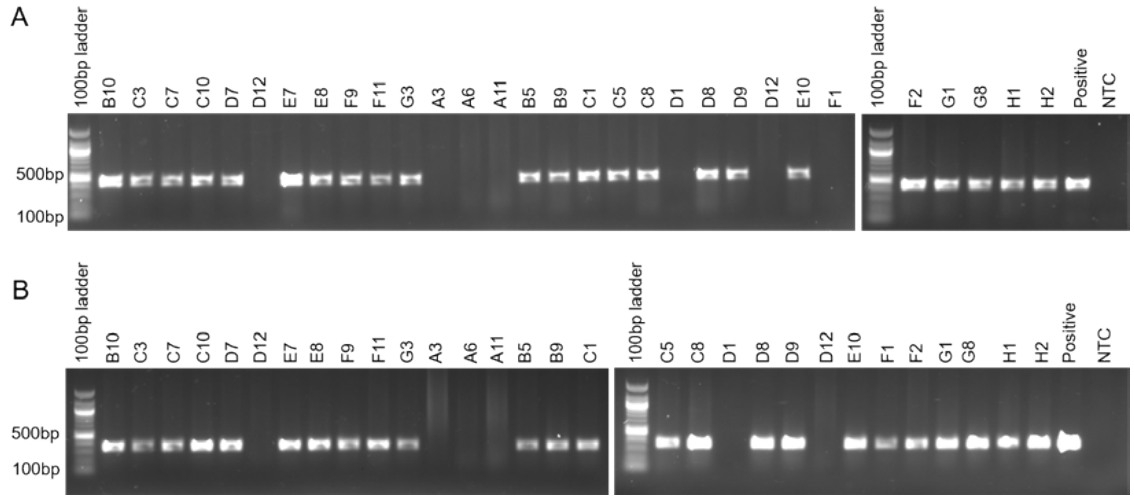


Figure A.14: Agarose gel indicating zygosity of the Intergenic Enhancer Deletion in edited iPSC clones.
To test for heterozygosity, alternate primer sets were used with one primer from each set lying within the deletion region, and one binding just outside. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The 5' end of the deletion region is shown first. If the clone is heterozygous it would amplify and show a band at 445bp. The 3' end of the deletion region is shown after the red dotted line. If the clone is heterozygous it would amplify and show a band at 349bp. Unedited iPSCs were used as a positive control.

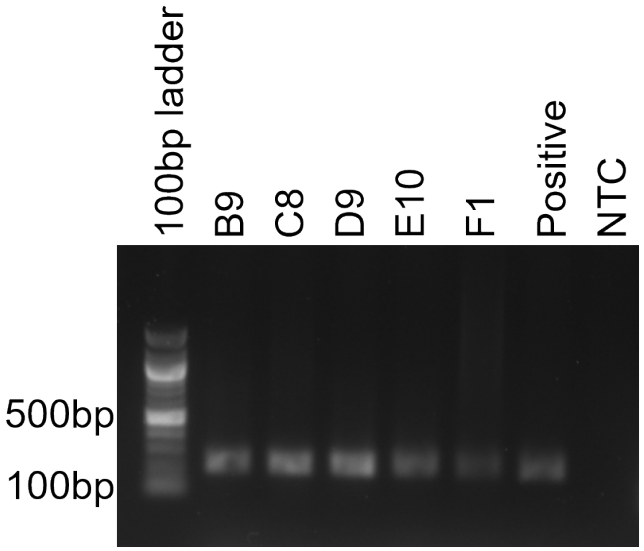


Figure A.15: Agarose gel indicating zygosity of the Intergenic Enhancer Deletion in edited iPSC clones using internal primers.
To confirm heterozygosity, alternate primer sets were used with both primers lying within the deletion region. The presence of a band on the gel means the deletion region is present in a heterozygous state, as amplification would not occur if the enhancer region has been removed. The expected amplicon size is 202bp. Unedited iPSCs were used as a positive control. None of the clones appear homozygous.

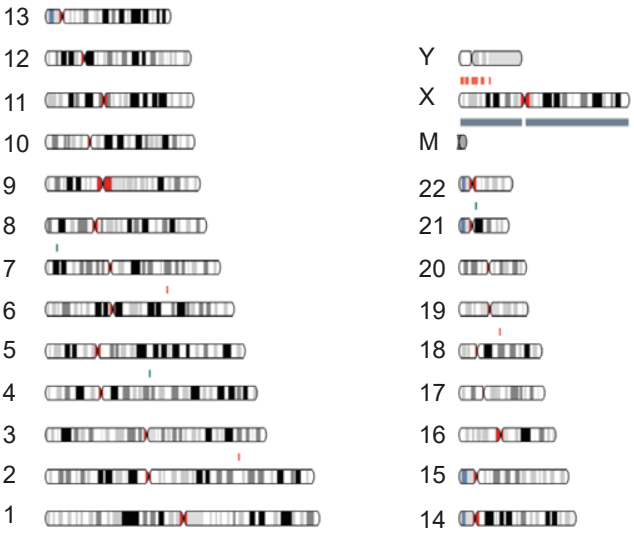


Figure A.16: Karyogram for Intronic Enhancer CRISPR Control 1.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intronic Enhancer A1 (CRISPR Control 1). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

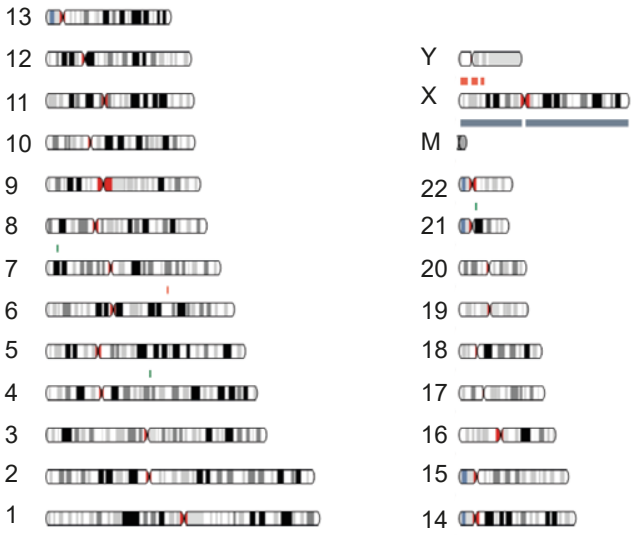


Figure A.17: Karyogram for Intronic Enhancer CRISPR Control 2.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intronic Enhancer B1 (CRISPR Control 2). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

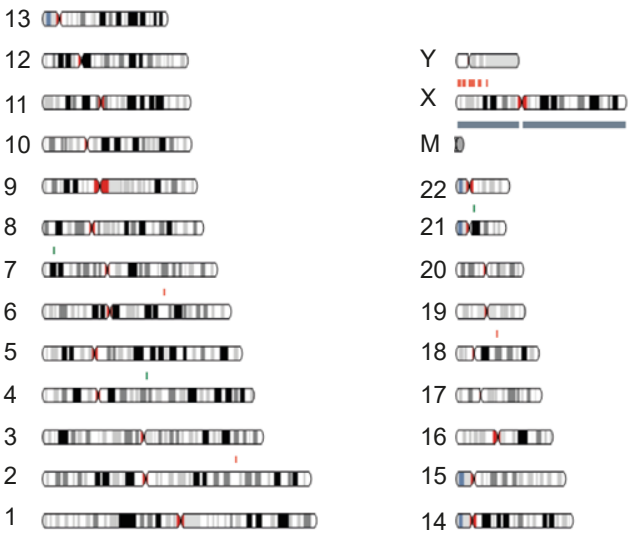


Figure A.18: Karyogram for Intronic Enhancer Deletion clone 1.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intronic Enhancer Deletion B6 (Intronic Enhancer Deletion 1). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate..



Figure A.19: Karyogram for Intronic Enhancer Deletion clone 2.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intronic Enhancer Deletion C5 (Intronic Enhancer Deletion 2). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

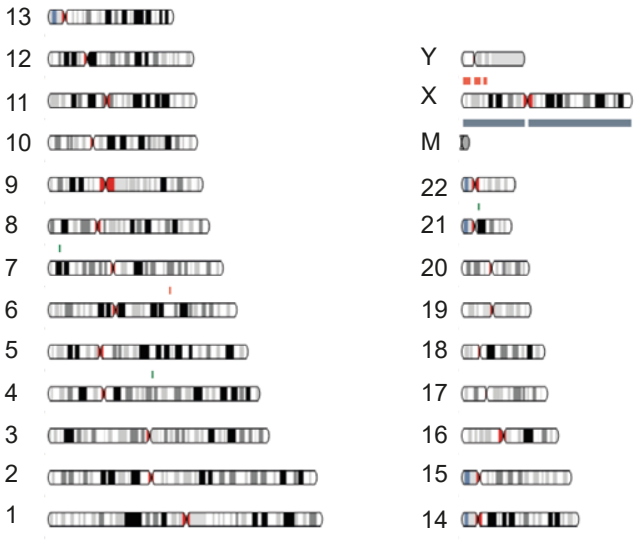


Figure A.20: Karyogram for Intronic Enhancer Deletion clone 3.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intronic Enhancer Deletion F7 (Intronic Enhancer Deletion 3). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

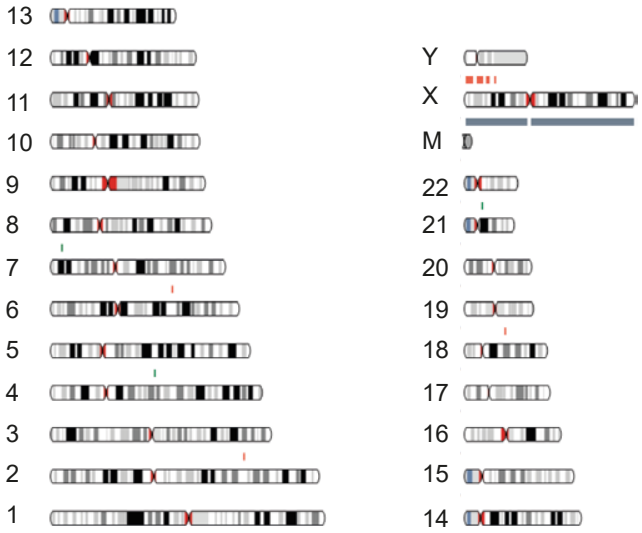


Figure A.21: Karyogram for Intergenic Enhancer CRISPR Control 1.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intergenic Enhancer C2 (CRISPR Control 1). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

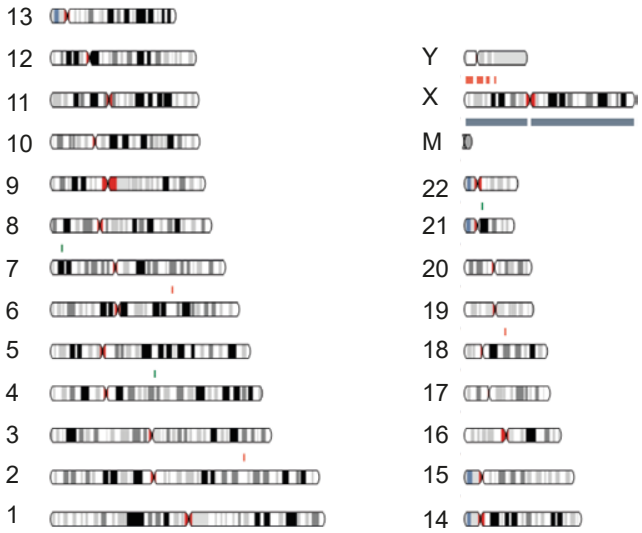


Figure A.22: Karyogram for Intergenic Enhancer CRISPR Control 2.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intergenic Enhancer G1 (CRISPR Control 2). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

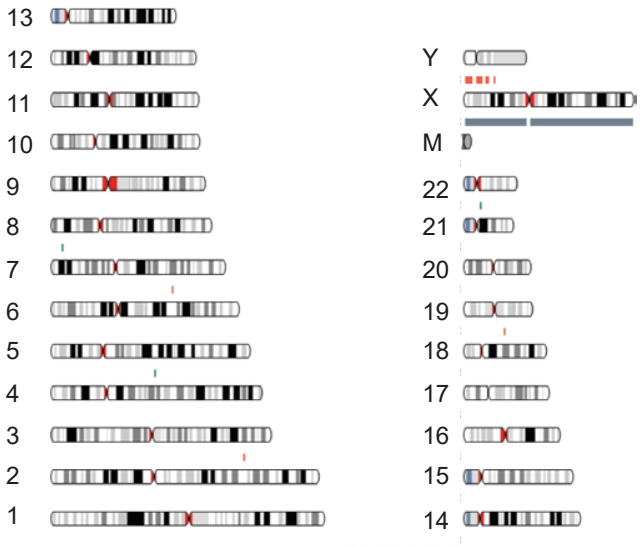


Figure A.23: Karyogram for Intergenic Enhancer Deletion clone 1.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intergenic Enhancer Deletion C8 sunclone 3 (Intergenic Enhancer Deletion 1). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

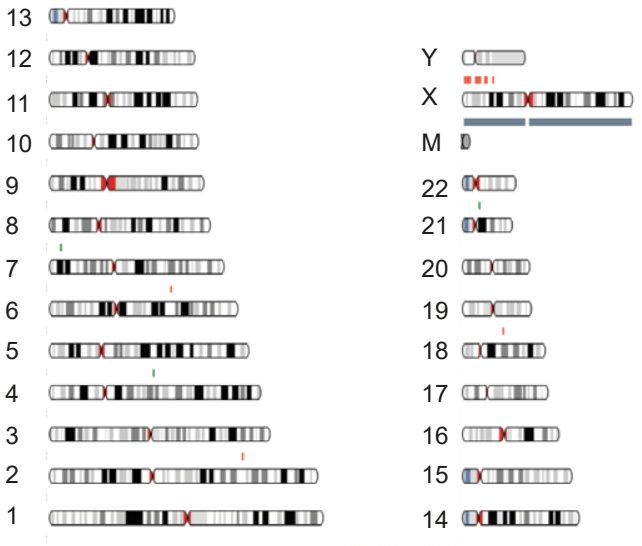


Figure A.24: Karyogram for Intergenic Enhancer Deletion clone 2.
Genome integrity was assessed by Illumina BeadChip array to look for gross abnormalities following the CRISPR/Cas9 editing process. This is the karyogram for sample Intergenic Enhancer Deletion C8 sunclone 4 (Intergenic Enhancer Deletion 2). A red line alongside a chromosome indicates a loss or single copy, green indicates a gain of copy, and grey indicates a loss of heterozygosity on autosomes, or two copies of X chromosome (female lines). This sample was compared to the parental control and showed no abnormalities indicative of affecting its growth or ability to differentiate.

B

Supplementary Data from Chapter 4

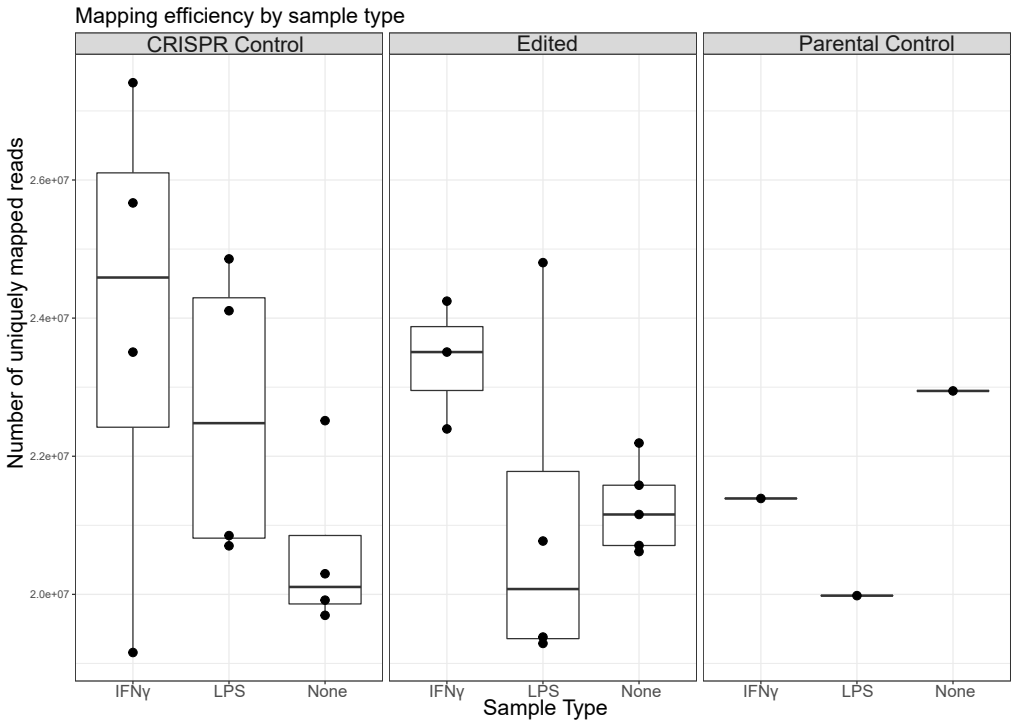


Figure B.1: RNA-Seq mapping efficiency by sample type.
RNA-Seq was carried out on edited and undedited iPSC derived macrophages under different stimulation conditions. All samples had initial read counts exceeding 20 million and all but five samples had at least 20 million uniquely mapped reads.

Gene Name	Log_2 fold Change	FDR
TRIM4	0.824	$1.40x10^{-4}$
FCGBP	-0.642	$6.87x10^{-3}$

Table B.1: Summary of differentially expressed genes following a genome-wide analysis between the Intronic Enhancer Deletion clones and CRISPR Controls after 48 hours of stimulation with IFNγ.
This table lists the differentially expressed genes following a genome-wide gene comparison and was not restricted to TADs. Only those genes with an FDR of <0.05 are shown.

Table B.2: Summary of differentially expressed genes following a genome-wide analysis between two Intergenic Enhancer Deletion edited clones, two CRISPR Controls, and one Parental Control cell line in an unstimulated state.
This table lists the differentially expressed genes following a genome-wide gene comparison and was not restricted a TAD. Only those genes with an FDR of <0.05 are shown.

Gene Name	Log_2 fold change	FDR
LY6E	-1.41	$1.33x10^{-31}$
LPL	-0.688	$7.76x10^{-8}$
CD93	0.709	$7.76x10^{-8}$
ADAMDEC1	-0.754	$7.26x10^{-7}$
ZFP41	-0.764	$9.06x10^{-6}$
EGR2	-0.759	$1.47x10^{-5}$
GALNT12	-0.765	$1.70x10^{-5}$
RP11-455F5.3	0.581	$4.56x10^{-5}$
ITGA3	-0.743	$4.56x10^{-5}$
CD9	-0.607	$6.13x10^{-5}$
EMP1	-0.569	$1.49x10^{-4}$
MSR1	-0.571	$3.59x10^{-4}$
PGBD5	-0.644	$3.62x10^{-4}$
TGM2	-0.665	$6.82x10^{-4}$
ANPEP	-0.642	$6.87x10^{-4}$
APOE	-0.609	$6.87x10^{-4}$
HTR2B	-0.552	$9.42x10^{-4}$
GPNMB	-0.564	$1.76x10^{-3}$
DCSTAMP	-0.657	$2.49x10^{-3}$
LINC01010	-0.655	$2.49x10^{-3}$
AFAP1L1	-0.559	$2.82x10^{-3}$
ALDH5A1	-0.647	$2.82x10^{-3}$
CDKN2B	-0.638	$3.09x10^{-3}$
FGR	-0.613	$5.35x10^{-3}$
FABP5P7	-0.542	$5.35x10^{-3}$
ITPKB	-0.493	$6.46x10^{-3}$
DTX1	-0.519	$1.07x10^{-2}$
FAIM3	-0.540	$1.07x10^{-2}$
GATM	0.472	$1.07x10^{-2}$
RAB11FIP4	-0.598	$1.07x10^{-2}$
USP2	-0.548	$1.07x10^{-2}$
CYTL1	-0.593	$1.07x10^{-2}$
NES	-0.596	$1.07x10^{-2}$
COL6A1	-0.572	$1.38x10^{-2}$
PLXDC2	-0.457	$1.46x10^{-2}$
VCAM1	-0.580	$1.46x10^{-2}$
FAM211A	-0.572	$1.52x10^{-2}$

Continued on following page

Table B.2 - *continued from previous page*

Gene Name	Log_2 fold change	FDR
ZNF562	-0.548	$1.62x10^{-2}$
ALCAM	-0.506	$1.62x10^{-2}$
LIMS1	-0.409	$1.65x10^{-2}$
CXXC5	-0.5239	$1.65x10^{-2}$
FABP4	-0.5449	$1.65x10^{-2}$
RP11	0.3329	$2.04x10^{-2}$
ITGAX	-0.4589	$2.04x10^{-2}$
NR4A3	-0.563	$2.19x10^{-2}$
ZMYND15	-0.555	$2.19x10^{-2}$
MCOLN3	-0.558	$2.24x10^{-2}$
TREM2	-0.449	$2.57x10^{-2}$
MMP2	-0.455	$2.60x10^{-2}$
EMR2	-0.546	$2.66x10^{-2}$
CRABP2	-0.546	$3.11x10^{-2}$
MNDA	-0.472	$3.11x10^{-2}$
BCAM	-0.500	$3.11x10^{-2}$
PDGFA	-0.545	$3.11x10^{-2}$
DOCK3	-0.545	$3.15x10^{-2}$
SPSB1	-0.543	$3.27x10^{-2}$
ITPR1	-0.500	$3.29x10^{-2}$
LPCAT1	0.373	$3.37x10^{-2}$
SESN3	-0.521	$3.40x10^{-2}$
GABARAPL1	-0.439	$3.57x10^{-2}$
BCL2A1	-0.535	$3.57x10^{-2}$
SLAMF7	-0.505	$3.57x10^{-2}$
CAPG	-0.391	$3.57x10^{-2}$
NR1H3	-0.470	$3.57x10^{-2}$
NRIP3	-0.531	$3.57x10^{-2}$
GLIPR2	-0.514	$3.57x10^{-2}$
DCLK2	-0.533	$3.57x10^{-2}$
PSEN2	-0.491	$3.60x10^{-2}$
ADARB1	-0.509	$4.63x10^{-2}$
CD84	-0.403	$4.68x10^{-2}$
SEMA7A	-0.492	$4.71x10^{-2}$
GAL3ST4	-0.468	$4.86x10^{-2}$

Gene Name	Log_2 fold Change	FDR
S100A6	-2.98	$9.63x10^{-8}$
PTCHD3P1	-2.64	$1.99x10^{-2}$
XIST	-0.33	$3.50x10^{-2}$

Table B.3: Summary of differentially expressed genes following a genome-wide analysis between the Intergenic Enhancer Deletion clones and CRISPR Controls after 48 hours of stimulation with LPS.
This table lists the differentially expressed genes following a genome-wide gene comparison and was not restricted to a TAD. Only those genes with an FDR of <0.05 are shown.

Table B.4: Summary of differentially expressed genes following a genome-wide analysis between Intergenic Enhancer deletion clones and CRISPR Controls after 48 hours of stimulation with IFNγ.
This table lists the differentially expressed genes following a genome-wide gene comparison and was not restricted a TAD. Only those genes with an FDR of <0.05 are shown.

Gene Name	Log_2 fold change	FDR
EMC10	-1.13	$5.60x10^{-16}$
S100A6	-1.66	$9.76x10^{-15}$
XIST	-3.64	$7.51x10^{-8}$
RBBP7	0.666	$1.03x10^{-5}$
FAIM3	-0.656	$1.05x10^{-4}$
CD93	0.669	$1.06x10^{-4}$
PTCHD3P1	-1.44	$1.06x10^{-4}$
CXCL9	-0.708	$1.06x10^{-4}$
TIMP3	-0.643	$1.29x10^{-4}$
SPP1	-0.611	$1.29x10^{-4}$
EGR2	-0.669	$2.32x10^{-4}$
HTR2B	-0.629	$4.20x10^{-4}$
RGCC	-0.645	$5.33x10^{-4}$
ADAMDEC1	-0.582	$5.90x10^{-4}$
GPNMB	-0.509	$5.94x10^{-4}$
PLD4	-0.633	$6.70x10^{-4}$
CD83	-0.556	$7.47x10^{-4}$
PLXNC1	-0.599	$1.29x10^{-3}$
SESN3	-0.631	$1.29x10^{-3}$
TMEM63C	-0.563	$1.58x10^{-3}$
MSR1	-0.573	$1.66x10^{-3}$
SCIN	-0.621	$1.75x10^{-3}$
DHRS9	-0.566	$2.34x10^{-3}$
ALCAM	-0.516	$2.58x10^{-3}$

Continued on following page

Table B.4 - *continued from previous page*

Gene Name	Log_2 fold change	FDR
CDKN2B	-0.598	$3.66x10^{-3}$
ABCG1	0.577	$3.66x10^{-3}$
CD209	-0.580	$4.02x10^{-3}$
MALAT1	-0.495	$4.69x10^{-3}$
LRP5	0.612	$5.17x10^{-3}$
ZFP41	-0.433	$5.63x10^{-3}$
SCD	0.449	$5.97x10^{-3}$
ETS1	0.565	$6.64x10^{-3}$
LGI2	0.567	$7.11x10^{-3}$
TNFSF18	-0.477	$7.47x10^{-3}$
FADS2	0.438	$9.54x10^{-3}$
FASN	0.442	$9.54x10^{-3}$
CD9	-0.471	$9.90x10^{-3}$
SLAMF7	-0.526	$1.18x10^{-2}$
MRAS	-0.548	$1.35x10^{-2}$
CTPS2	0.551	$1.35x10^{-2}$
PAQR6	0.474	$1.35x10^{-2}$
DHCR7	0.475	$1.46x10^{-2}$
KCNAB1	-0.379	$2.07x10^{-2}$
OLR1	-0.515	$2.12x10^{-2}$
LSS	0.450	$2.13x10^{-2}$
ACSM5	-0.501	$2.19x10^{-2}$
FADS1	0.421	$2.19x10^{-2}$
PCDH12	0.445	$2.19x10^{-2}$
METTL7A	-0.456	$2.23x10^{-2}$
APOE	-0.403	$2.24x10^{-2}$
CLEC11A	-0.519	$2.35x10^{-2}$
WIPF3	-0.460	$2.39x10^{-2}$
SERPINI1	-0.510	$2.48x10^{-2}$
PIGR	-0.499	$2.68x10^{-2}$
SYAP1	0.482	$2.93x10^{-2}$
CDKN2A	-0.504	$2.95x10^{-2}$
MYO1G	0.522	$3.97x10^{-2}$
ALDH5A1	-0.491	$4.15x10^{-2}$
ARRDC4	-0.469	$4.15x10^{-2}$

Name	Capture Fragments	Cis Reporter Fragments
PLEKHG6 Promoter	51929	38929
TNFRSF1A Promoter	62463	44012
TNFRSF1A rs887477	33986	27322
TNFRSF1A rs11064145	76894	52434
SCNN1A Promoter	62712	45723
LTBR Promoter	50408	36997
CD27 Promoter	75365	59885
TAPBPL Promoter	68532	51010
VAMP1 Promoter	71079	43622
NCAPD2 Promoter	68657	49075

Table B.5: Capture-C sequenced fragments for $CD14^+$ cells.
Capture fragments are the DpnII fragments targeted by the biotinylated oligo. Reporter fragments are any other fragment, sequenced in the same read as a capture fragment, indicating interaction. A minimum of 10000 fragments per capture site is required for analysis. Cis interactions refer to interactions occurring within the same chromosome as the capture fragment.

Name	Capture Fragments	Cis Reporter Fragments
PLEKHG6 Promoter	1078	439
TNFRSF1A Promoter	1134	456
TNFRSF1A rs887477	1113	390
TNFRSF1A rs11064145	1826	824
SCNN1A Promoter	1333	542
LTBR Promoter	1165	458
CD27 Promoter	1837	868
TAPBPL Promoter	1887	842
VAMP1 Promoter	2353	1086
NCAPD2 Promoter	2262	977

Table B.6: Capture-C sequenced fragments for Intronic Enhancer Deletion CRISPR/Cas9 edited clones.
Capture fragments are the DpnII fragments targeted by the biotinylated oligo. Reporter fragments are any other fragment, sequenced in the same read as a capture fragment, indicating interaction. A minimum of 10000 fragments per capture site is required for analysis, so this cell line was unusable. Cis interactions refer to interactions occurring within the same chromosome as the capture fragment.

Name	Capture Fragments	Cis Reporter Fragments
PLEKHG6 Promoter	670	226
TNFRSF1A Promoter	726	301
TNFRSF1A rs887477	695	217
TNFRSF1A rs11064145	1305	548
SCNN1A Promoter	780	297
LTBR Promoter	825	240
CD27 Promoter	1179	504
TAPBPL Promoter	1264	523
VAMP1 Promoter	1394	599
NCAPD2 Promoter	1150	535

Table B.7: Capture-C sequenced fragments for Intergenic Enhancer Deletion CRISPR/Cas9 edited clones.
Capture fragments are the DpnII fragments targeted by the biotinylated oligo. Reporter fragments are any other fragment, sequenced in the same read as a capture fragment, indicating interaction. A minimum of 10000 fragments per capture site is required for analysis, so this cell line was unusable. Cis interactions refer to interactions occurring within the same chromosome as the capture fragment.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6444927-6446228	*TNFRSF1A*	9.351	$1.96x10^{-9}$
chr12:6602133-6603649	*NCAPD2*	-0.391	0.728
chr12:6539162-6540634	NA	0.466	0.771
chr12:6570317-6572085	*TAPBPL*	0.317	0.771
chr12:6579176-6580574	*VAMP1*	-0.265	0.772
chr12:6532164-6533793	NA	0.388	0.806
chr12:6421670-6422903	*PLEKHG6*	-0.117	0.835
chr12:6432144-6433445	*PLEKHG6*	-0.130	0.835
chr12:6492518-6493983	*SCNN1A*	-0.123	0.835
chr12:6502999-6504310	NA	0.173	0.835
chr12:6525452-6526822	NA	-0.073	0.835
chr12:6560238-6561919	*TAPBPL*	-0.081	0.835

Table B.8: Regions of differential open chromatin between Intronic Enhancer Deletion clones and unedited controls (LPS stimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. Only one peak was significantly different (FDR of <0.05), which corresponds to the region that was deleted. The coordinates stated are in UCSC notation based on genome build hg19.

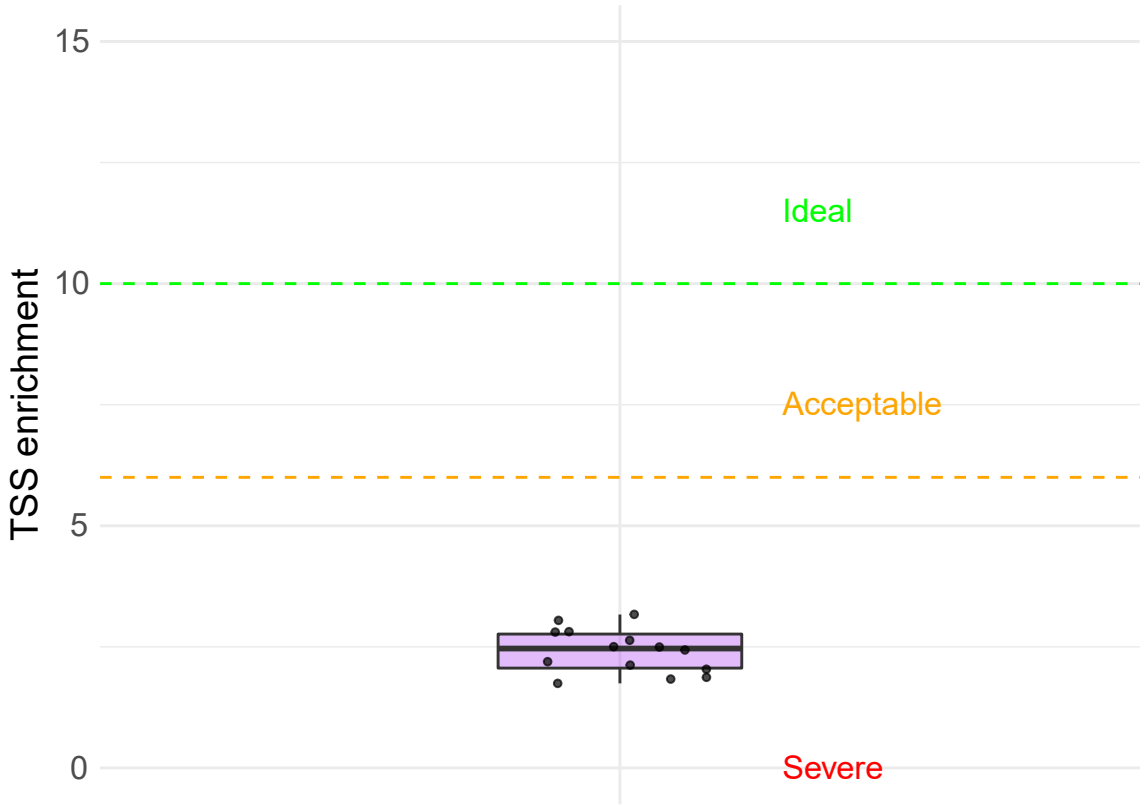


Figure B.2: TSS fold enrichment following Omni-ATAC on iPSC derived macrophages.
Samples containing an enhancer deletion (Enh Del) and control samples were sequenced in both an unstimulated and LPS stimulated state. ENCODE recommends samples having a TSS fold enrichment of >5. All samples demonstrated a fold enrichment of below 5, suggesting a low signal to noise ratio.

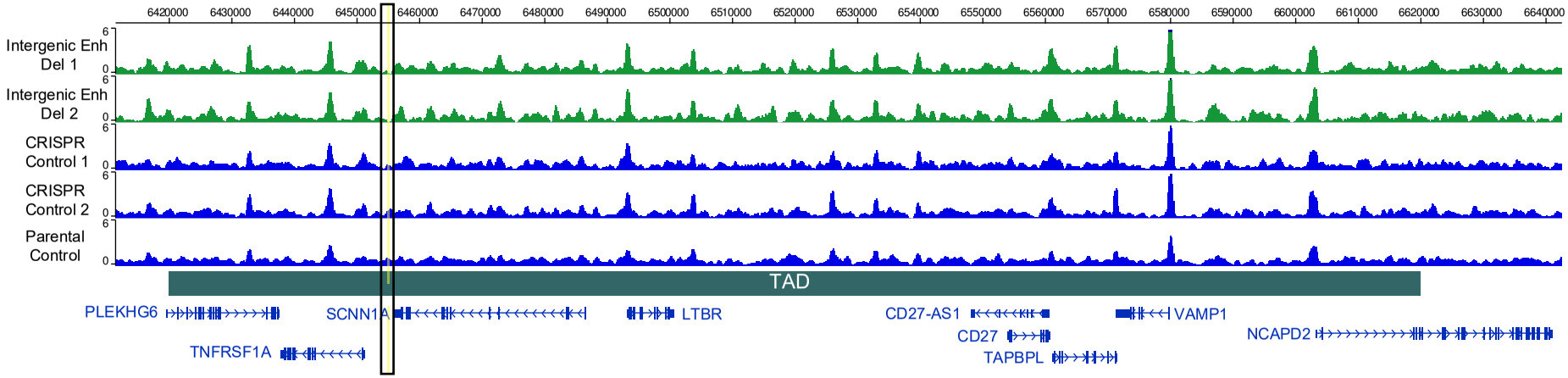


Figure B.3: Open chromatin peaks in Intergenic Enhancer Deletion clones and unedited controls in an unstimulated state. WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The Intergenic Enhancer is highlighted in yellow. None of the peaks were differential.

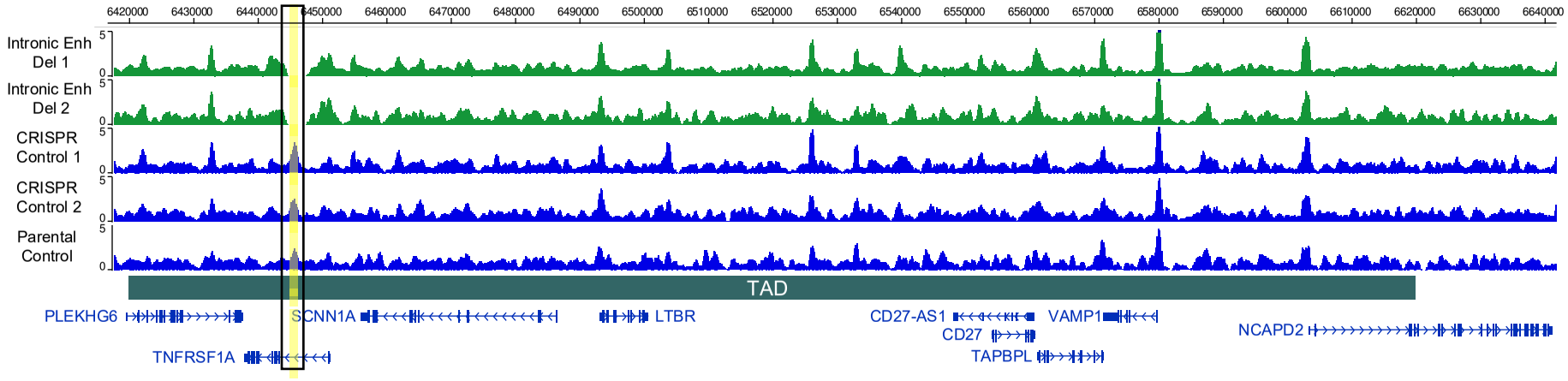


Figure B.4: Differentially open chromatin peaks in Intronic Enhancer Deletion clones and unedited controls after LPS stimulation. WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 1.96×10^{-9}).

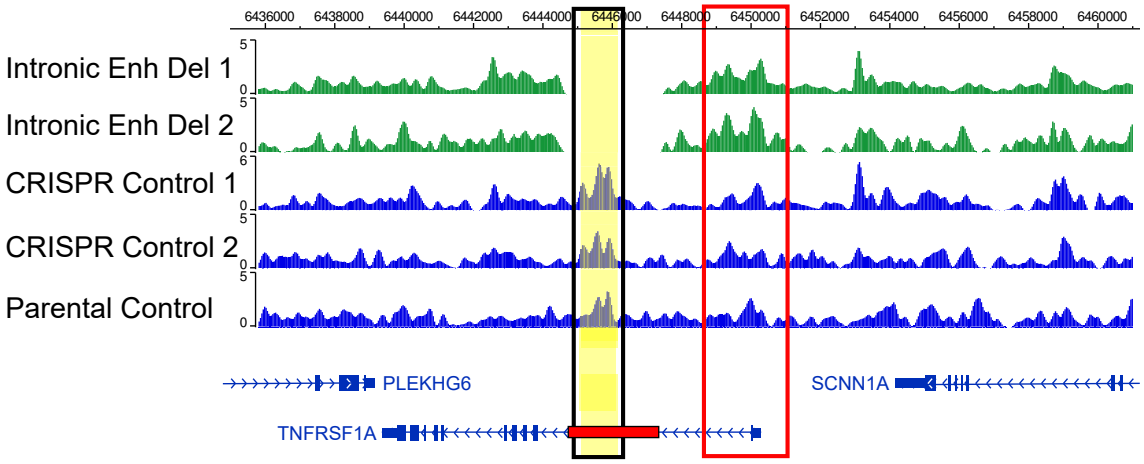


Figure B.5: Close up of the differentially open chromatin peak over the Intronic Enhancer Deletion in edited cells and unedited controls after LPS stimulation.
WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 1.96×10^{-9}). The deletion region is represented by the red box. The peak over the *TNFRSF1A* promoter is outlined in red. This peaks is not differential, but visually appears larger in edited cells.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6421670-6422903	*PLEKHG6*	-0.254	0.841
chr12:6432144-6433445	*PLEKHG6*	-0.329	0.841
chr12:6444927-6446228	*TNFRSF1A*	-0.454	0.841
chr12:6492518-6493983	*SCNN1A*	-0.202	0.841
chr12:6502999-6504310	NA	0.096	0.841
chr12:6525452-6526822	NA	0.123	0.841
chr12:6532164-6533793	NA	0.634	0.841
chr12:6539162-6540634	NA	0.481	0.841
chr12:6560238-6561919	*TAPBPL*	0.205	0.841
chr12:6570317-6572085	*TAPBPL*	0.231	0.841
chr12:6579176-6580574	*VAMP1*	-0.243	0.841
chr12:6602133-6603649	*NCAPD2*	-0.144	0.841

Table B.9: Regions of differential open chromatin between Intergenic Enhancer Deletion clones and unedited controls (LPS stimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks one were significantly different (FDR of <0.05) between samples. The coordinates stated are in UCSC notation based on genome build hg19.

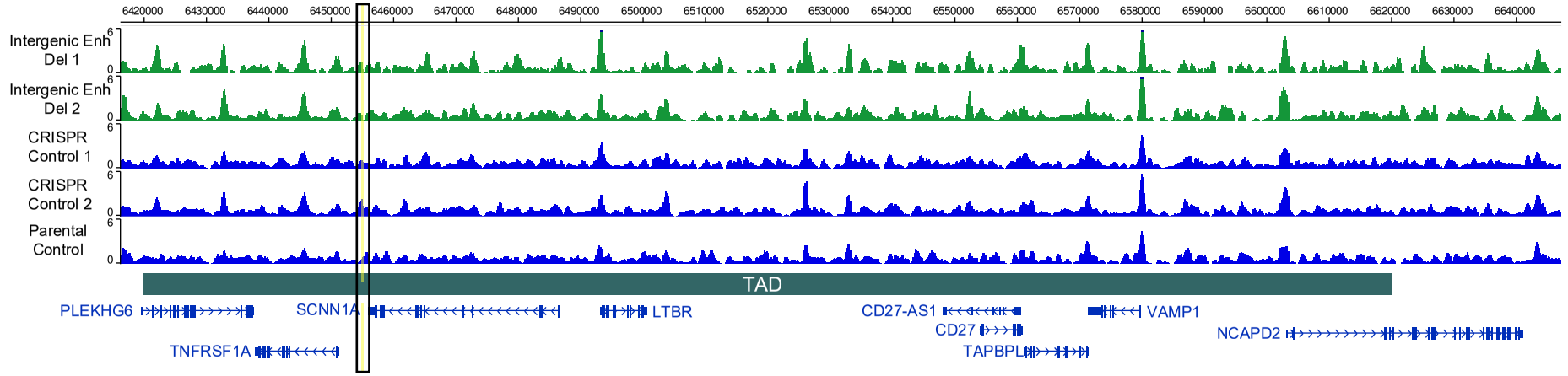


Figure B.6: Open chromatin peaks in Intergenic Enhancer Deletion clones and unedited controls after stimulation with LPS. WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, two CRISPR Controls, and the Parental Control are shown. The Intergenic Enhancer is highlighted in yellow. None of the peaks were differential.

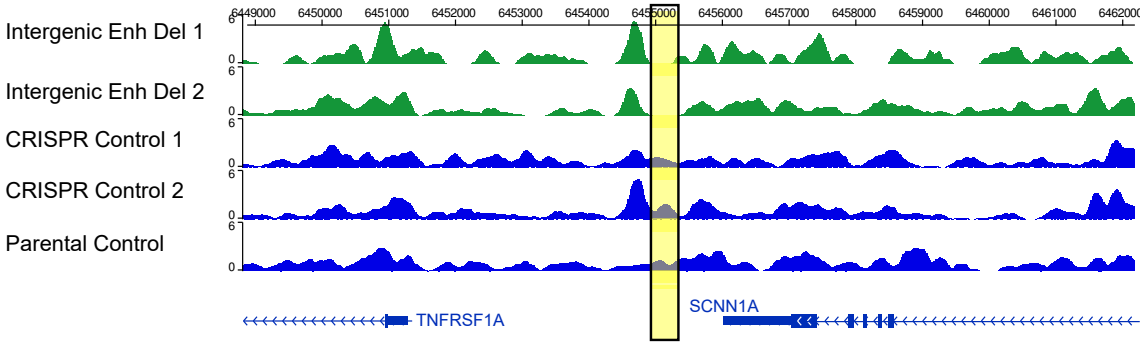


Figure B.7: Close up of the Intergenic Enhancer Deletion region in edited clones and unedited controls after stimulation with LPS, showing chromatin accessibility.
WashU Epigenome Browser view of the Omni-ATAC read density distribution for edited and unedited iPSC derived macrophages. Two Intronic Enhancer Deletion clones, two CRISPR Control clones, and the Parental Control are shown. The Intergenic Enhancer is highlighted in yellow, but no open chromatin peak was called in this region. The small peak next to this region does not appear in the unstimulated samples. None of the peaks were differential.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6421670-6422903	*PLEKHG6*	-0.710	0.121
chr12:6602133-6603649	*NCAPD2*	-0.198	0.516
chr12:6492518-6493983	*SCNN1A*	0.102	0.901
chr12:6525452-6526822	NA	-0.169	0.901
chr12:6539162-6540634	NA	-0.100	0.901
chr12:6560238-6561919	*TAPBPL*	0.080	0.901
chr12:6579176-6580574	*VAMP1*	0.152	0.901
chr12:6432144-6433445	*PLEKHG6*	-0.068	0.931
chr12:6444927-6446228	*TNFRSF1A*	0.190	0.931
chr12:6502999-6504310	NA	0.048	0.931
chr12:6532164-6533793	NA	0.038	0.931
chr12:6570317-6572085	*TAPBPL*	0.015	0.931

Table B.10: Regions of differential open chromatin between unstimulated and LPS stimulated Intronic Enhancer Deletion clones and controls.
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significant different in cells treated with LPS. The coordinates stated are in UCSC notation based on genome build hg19.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6421670-6422903	*PLEKHG6*	-0.702	0.168
chr12:6560238-6561919	*TAPBPL*	0.269	0.576
chr12:6432144-6433445	*PLEKHG6*	-0.115	0.983
chr12:6444927-6446228	*TNFRSF1A*	0.015	0.983
chr12:6492518-6493983	*SCNN1A*	0.106	0.983
chr12:6502999-6504310	NA	-0.037	0.983
chr12:6525452-6526822	NA	-0.102	0.983
chr12:6532164-6533793	NA	0.120	0.983
chr12:6539162-6540634	NA	-0.004	0.983
chr12:6570317-6572085	*TAPBPL*	-0.022	0.983
chr12:6579176-6580574	*VAMP1*	0.128	0.983
chr12:6602133-6603649	*NCAPD2*	-0.055	0.983

Table B.11: Regions of differential open chromatin between unstimulated and LPS stimulated Intergenic Enhancer Deletion clones and controls.
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks were significant different in cells treated with LPS. The coordinates stated are in UCSC notation based on genome build hg19.

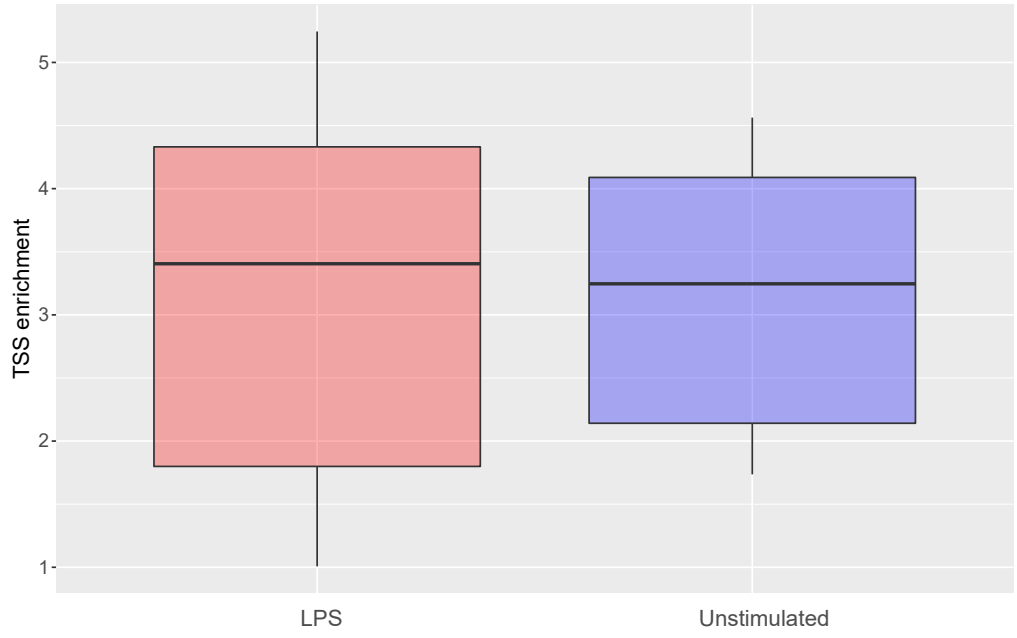


Figure B.8: TSS fold enrichment following ChIPm on iPSC derived macrophages.
Samples containing an enhancer deletion (Enh Del) and control samples were sequenced in both an unstimulated and LPS stimulated state.

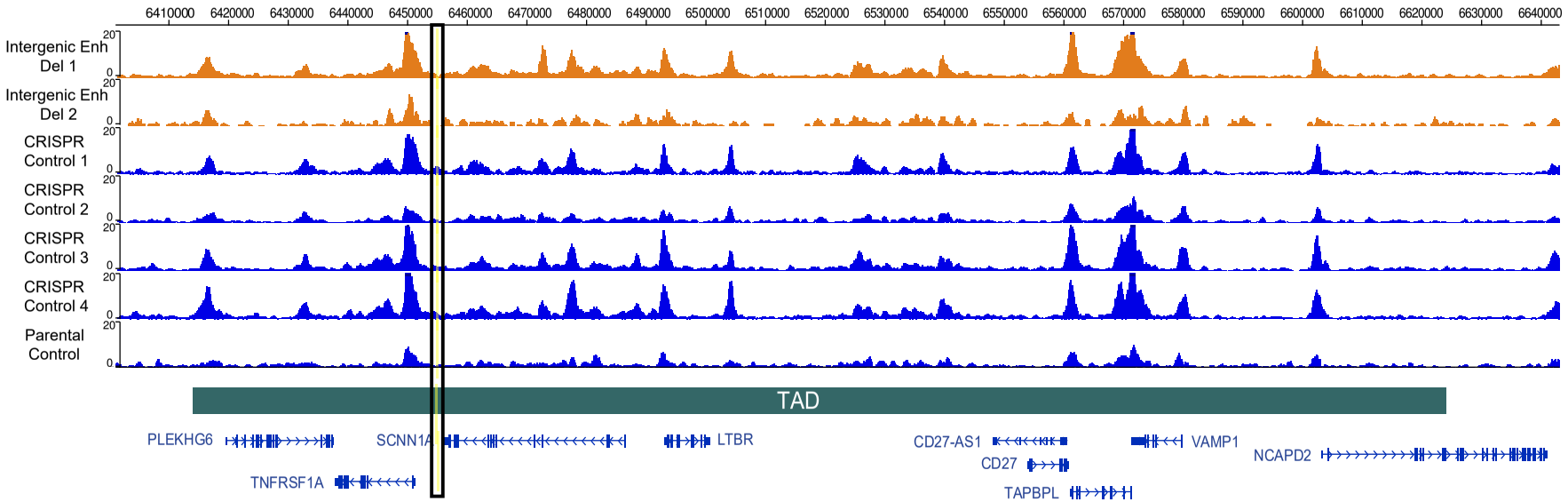


Figure B.9: Regions of H3K27ac enhancer histone mark modifications in Intergenic Enhancer Deletion clones and unedited controls in an unstimulated state.

WashU Epigenome Browser view of the ChIPmentation read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intergenic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. None of the peaks were significantly differential between edited and unedited samples. The deleted region is highlighted in yellow.

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6443654-6447439	*TNFRSF1A*	4.798	$1.12x10^{-26}$
chr12:6441496-6442515	*TNFRSF1A*	0.914	0.112
chr12:6501613-6502113	*LTBR*	-0.544	0.885
chr12:6568210-6573444	*TAPBPL/VAMP1*	-0.277	0.885
chr12:6603193-6604157	*NCAPD2*	0.489	0.885
chr12:6460247-6460887	*SCNN1A*	-0.222	0.885
chr12:6460934-6462955	*SCNN1A*	0.274	0.885
chr12:6467539-6468401	*SCNN1A*	0.224	0.885
chr12:6471971-6473762	*SCNN1A*	0.175	0.885
chr12:6476454-6478580	*SCNN1A*	0.137	0.885
chr12:6481586-6482482	*SCNN1A*	0.176	0.885
chr12:6487632-6489030	*LTBR*	0.163	0.885
chr12:6492222-6494533	*LTBR*	-0.300	0.885
chr12:6495594-6496638	*LTBR*	0.149	0.885
chr12:6538819-6540518	NA	-0.201	0.885
chr12:6559861-6562469	*CD27/TAPBPL*	-0.239	0.885
chr12:6578918-6580886	*VAMP1*	-0.202	0.885
chr12:6448993-6452438	*TNFRSF1A*	0.056	0.942
chr12:6453927-6455189	*TNFRSF1A*	0.078	0.942
chr12:6503005-6504920	*LTBR*	0.077	0.943
chr12:6432218-6433603	*PLEKHG6*	-0.098	0.972
chr12:6524686-6527917	NA	-0.018	0.972
chr12:6532784-6533655	NA	0.054	0.972
chr12:6601568-6603078	*NCAPD2*	0.023	0.972

Table B.12: Regions of differential chromatin with the H3K27ac histone mark between Intronic Enhancer Deletion clones and unedited controls (LPS stimulated). This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. Only one peak was significantly different (FDR of <0.05), which corresponds to the region that was deleted. The coordinates stated are in UCSC notation based on genome build hg19.

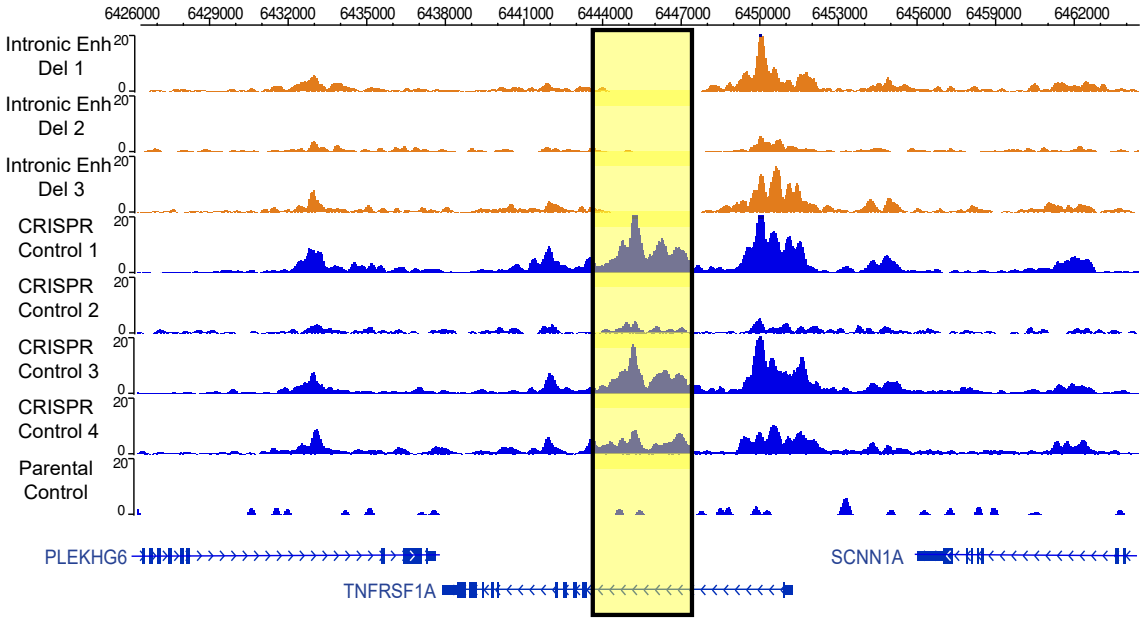


Figure B.10: Close up of the differential H3K27ac modification peak over the Intronic Enhancer Deletion in edited clones and unedited controls after stimulation with LPS. WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages. Three Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of $1.12x10^{-26}$).

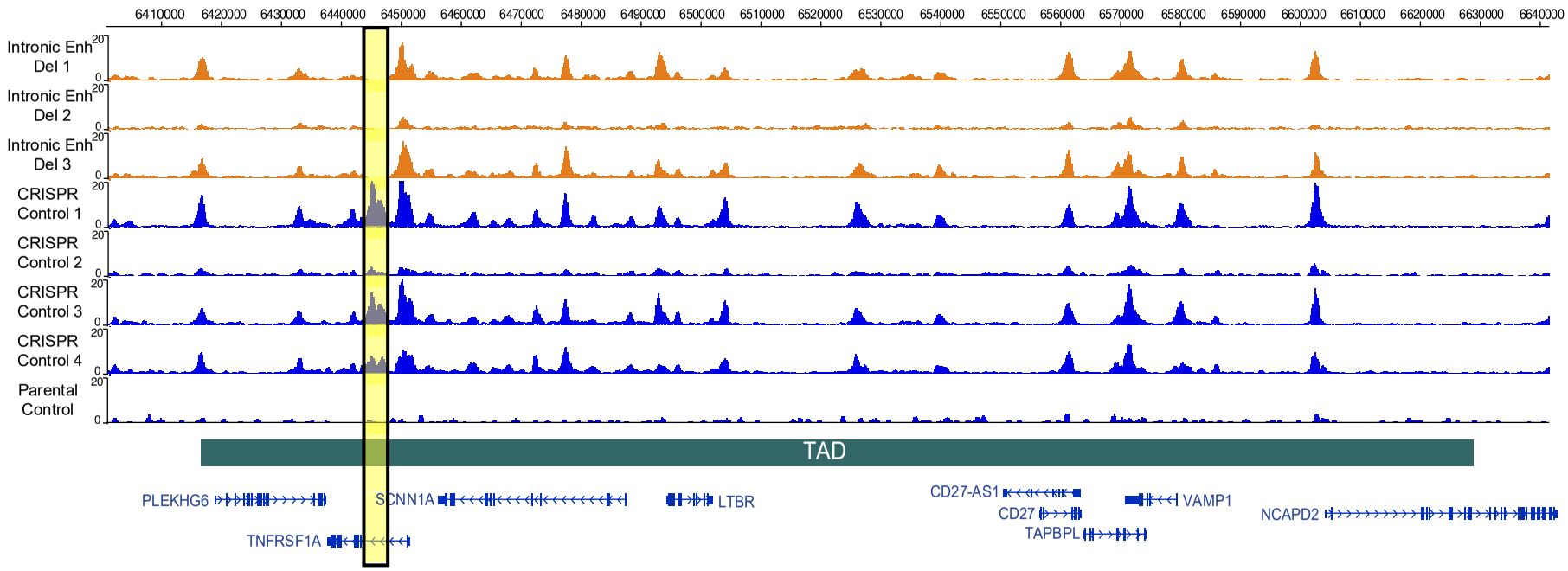


Figure B.11: Regions of H3K27ac enhancer histone mark modifications in Intronic Enhancer Deletion clones and unedited controls after stimulation with LPS.

WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Three Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. The significant differential peak over the Intronic Enhancer is highlighted in yellow (FDR of 1.12×10^{-26}).

Peak Coordinates	Proximal Gene	Log_2 fold Change	FDR
chr12:6453927-6455189	*TNFRSF1A*	0.820	0.450
chr12:6568210-6573444	*TAPBPL/VAMP1*	-0.465	0.450
chr12:6467539-6468401	*SCNN1A*	0.682	0.915
chr12:6476454-6478580	*SCNN1A*	0.374	0.937
chr12:6448993-6452438	*TNFRSF1A*	-0.182	0.989
chr12:6492222-6494533	*LTBR*	-0.226	0.989
chr12:6538819-6540518	NA	-0.355	0.989
chr12:6432218-6433603	*PLEKHG6*	0.104	0.999
chr12:6441496-6442515	*TNFRSF1A*	0.025	0.999
chr12:6443654-6447439	*TNFRSF1A*	-0.104	0.999
chr12:6460247-6460887	*SCNN1A*	-0.210	0.999
chr12:6460934-6462955	*SCNN1A*	-0.058	0.999
chr12:6471971-6473762	*SCNN1A*	-0.068	0.999
chr12:6481586-6482482	*SCNN1A*	-0.270	0.999
chr12:6487632-6489030	*LTBR*	0.077	0.999
chr12:6495594-6496638	*LTBR*	0.004	0.9699
chr12:6501613-6502113	*LTBR*	-0.236	0.999
chr12:6503005-6504920	*LTBR*	0.003	0.999
chr12:6524686-6527917	NA	0.096	0.999
chr12:6532784-6533655	NA	-0.210	0.999
chr12:6559861-6562469	*CD27/TAPBPL*	-0.143	0.999
chr12:6578918-6580886	*VAMP1*	-0.149	0.999
chr12:6601568-6603078	*NCAPD2*	0.119	0.999
chr12:6603193-6604157	*NCAPD2*	-0.013	0.999

Table B.13: Regions of chromatin with differential H3K27ac between Intergenic Enhancer Deletion clones and unedited controls (LPS stimulated).
This table lists the peaks called during the Omni-ATAC differential analysis, restricted to the *TNFRSF1A* TAD. None of the peaks one were significantly different (FDR of <0.05) between samples. The coordinates stated are in UCSC notation based on genome build hg19.

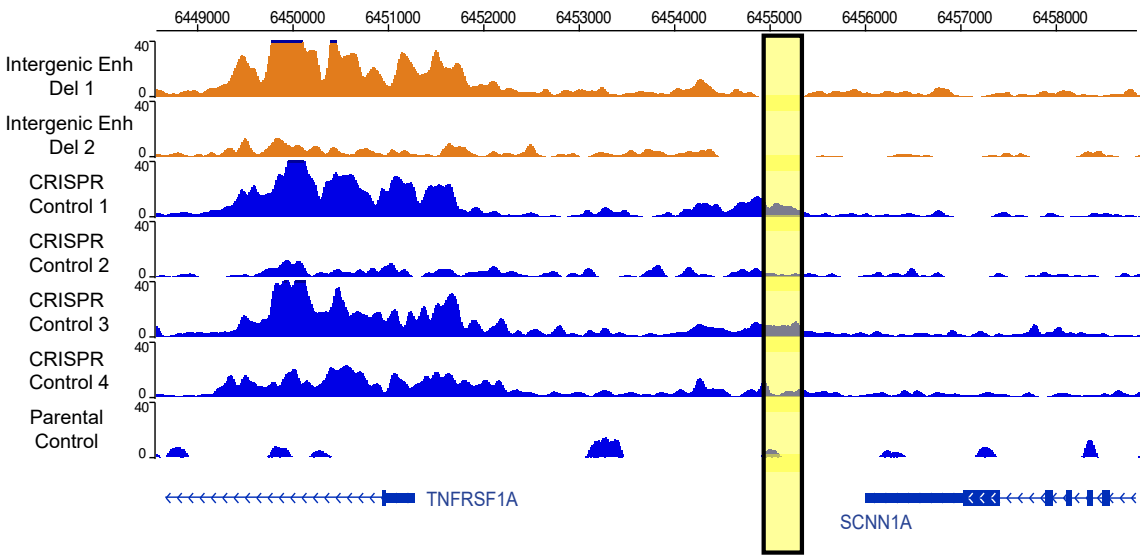


Figure B.12: Close up of the differential H3K27ac modification peak over the Intergenic Enhancer Deletion in edited clones and unedited controls after stimulation with LPS.
WashU Epigenome Browser view of the ChIPm read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. None of the peaks were significantly differential between edited and unedited samples. The deleted region is highlighted in yellow.

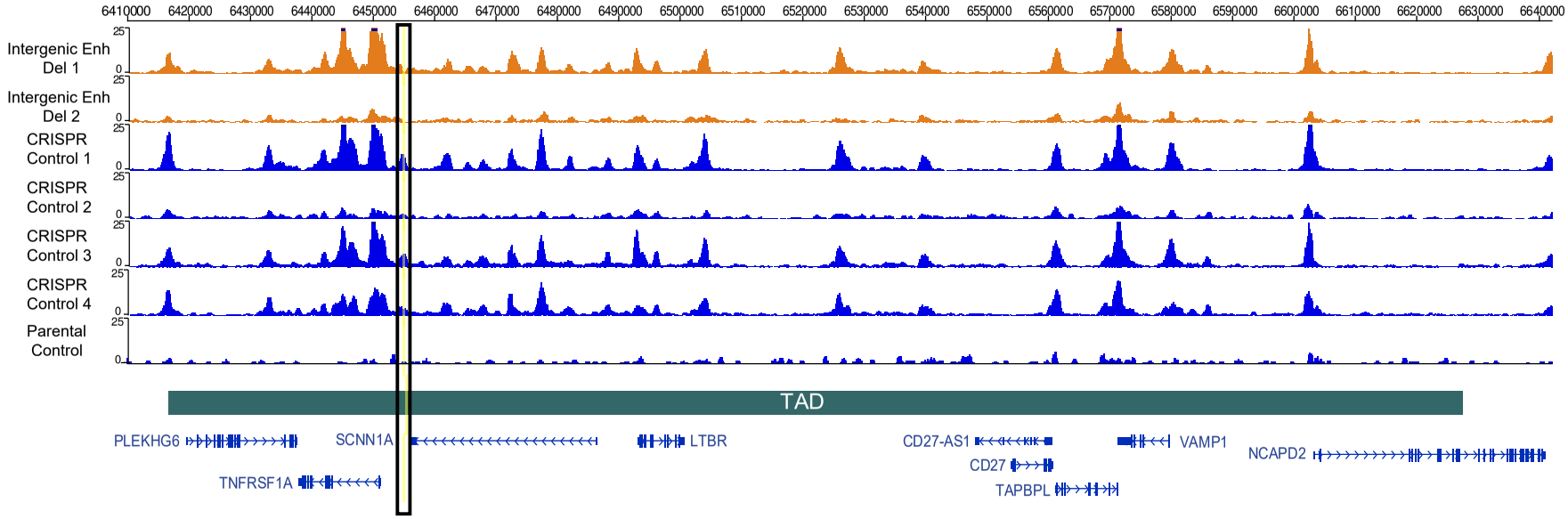


Figure B.13: Regions of H3K27ac enhancer histone mark modifications in Intergenic Enhancer Deletion clones and unedited controls after stimulation with LPS.

WashU Epigenome Browser view of the ChIPmentation read density distribution for edited and unedited iPSC derived macrophages across the entire TAD. Two Intronic Enhancer Deletion clones, four CRISPR Controls, and the Parental Control are shown. None of the peaks were significantly differential between edited and unedited samples. The deleted region is highlighted in yellow.

C

Supplementary Data from Chapter 5

Table C.1: Summary of the cell types used for ChIP and ATAC analysis from each patient. Three types of ATAC were used over the course of this study. Historic samples performed using Omni-ATAC on formaldehyde fixed samples are denoted by His-ATAC.

Patient ID	H3K27ac Cell Types	H3K4me3 Cell Types	ATAC Cell Types	ATAC Type	Batch
AS004	$CD4^+$ $CD14^+$	$CD4^+$ $CD14^+$	$CD4^+$	His-ATAC	3
AS014	$CD4^+$ $CD8^+$	$CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$	His-ATAC	3
AS018	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	His-ATAC	3
AS037	$CD8^+$	$CD8^+$	$CD8^+$	His-ATAC	3
AS038	$CD4^+$ $CD14^+$	$CD4^+$ $CD14^+$	$CD4^+$	His-ATAC	3
AS043	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
AS044	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
AS045	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
AS046	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
AS047	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1

Continued on following page

Table C.1 - *continued from previous page*

Patient ID	H3K27ac Cell Types	H3K4me3 Cell Types	ATAC Cell Types	ATAC Type	Batch
AS049	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
AS050	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
AS051	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
AS052	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Omni-ATAC	2

Table C.2: Summary of the cell types used for ChIP and ATAC analysis from healthy controls. Three types of ATAC were used over the course of this study. Historic samples performed using Omni-ATAC on formaldehyde fixed samples are denoted by His-ATAC.

Sample ID	H3K27ac Cell Types	H3K4me3 Cell Types	ATAC Cell Types	ATAC Type	Batch
CTRL1	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL2	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL3	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL4	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL5	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL6	None	None	$CD4^+$ $CD8^+$ $CD14^+$	ATAC-Seq	1
CTRL7	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Omni-ATAC	2
CTRL8	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
CTRL9	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
CTRL10	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	$CD4^+$ $CD8^+$ $CD14^+$	Fast-ATAC	2
CTRL11	$CD4^+$ $CD8^+$	$CD4^+$ $CD8^+$	$CD4^+$ $CD8^+$	His-ATAC	3
CTRL12	$CD4^+$	$CD4^+$	$CD4^+$	His-ATAC	3
CTRL14	$CD4^+$ $CD8^+$	$CD4^+$ $CD8^+$	$CD4^+$ $CD8^+$	His-ATAC	3

Continued on following page

Table C.2 - *continued from previous page*

Sample ID	H3K27ac Cell Types	H3K4me3 Cell Types	ATAC Cell Types	ATAC Type	Batch
CTRL16	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL17	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL18	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL19	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL20	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL21	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL22	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL23	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL24	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4
CTRL25	None	None	$CD4^+$ $CD14^+$	Omni-ATAC	4

ATAC Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr13:28020751-28021889	$3.30\text{x}10^{-109}$	$1.10\text{x}10^{-113}$	$3.20\text{x}10^{-6}$	*MTIF3*
chr4:84140080-84141555	$1.10\text{x}10^{-105}$	$1.00\text{x}10^{-108}$	$2.60\text{x}10^{-14}$	*COQ2*
chr6:143859908-143861279	$1.30\text{x}10^{-96}$	$4.80\text{x}10^{-105}$	$2.70\text{x}10^{-5}$	*FUCA2*
chr2:197124530-197125961	$2.10\text{x}10^{-95}$	$1.10\text{x}10^{-105}$	$2.30\text{x}10^{-15}$	*HECW2*
chr2:172019509-172020531	$1.10\text{x}10^{-93}$	$5.30\text{x}10^{-92}$	$1.30\text{x}10^{-18}$	*TLK1*
chr3:107811536-107812852	$3.20\text{x}10^{-92}$	$7.70\text{x}10^{-103}$	$7.20\text{x}10^{-8}$	*CD47*
chr8:82486790-82488091	$4.00\text{x}10^{-87}$	$4.30\text{x}10^{-76}$	$8.30\text{x}10^{-32}$	*FABP12*
chr2:192493319-192494478	$1.10\text{x}10^{-84}$	$1.40\text{x}10^{-82}$	$3.40\text{x}10^{-21}$	*NABP1*
chr1:101395502-101396991	$3.20\text{x}10^{-75}$	$9.70\text{x}10^{-82}$	$8.40\text{x}10^{-5}$	*SLC30A7*
chr2:109234983-109236032	$5.40\text{x}10^{-75}$	$3.60\text{x}10^{-75}$	$2.90\text{x}10^{-18}$	*LIMS1*
chr2:202312878-202314285	$2.20\text{x}10^{-73}$	$5.40\text{x}10^{-77}$	$1.80\text{x}10^{-12}$	*TRAK2*
chr3:97747343-97748759	$3.70\text{x}10^{-73}$	$2.00\text{x}10^{-69}$	$1.20\text{x}10^{-20}$	*GABRR3*
chr6:11722171-11723738	$1.90\text{x}10^{-72}$	$3.10\text{x}10^{-63}$	$1.70\text{x}10^{-32}$	*ADTRP*
chr4:40192082-40193553	$1.20\text{x}10^{-71}$	$6.60\text{x}10^{-77}$	0.0046	*RHOH*
chr5:56790170-56791924	$2.20\text{x}10^{-71}$	$5.50\text{x}10^{-56}$	$1.00\text{x}10^{-39}$	*ACTBL2*
chr4:143621219-143622629	$2.40\text{x}10^{-70}$	$2.70\text{x}10^{-74}$	$1.10\text{x}10^{-14}$	*INPP4B*
chr5:55772986-55774148	$2.80\text{x}10^{-67}$	$8.00\text{x}10^{-64}$	$8.20\text{x}10^{-23}$	NA
chr10:121414657-121415905	$4.40\text{x}10^{-67}$	$2.40\text{x}10^{-70}$	$7.90\text{x}10^{-17}$	*BAG3*
chr6:143094233-143095960	$2.20\text{x}10^{-66}$	$7.00\text{x}10^{-63}$	$2.70\text{x}10^{-25}$	*HIVEP2*
chr5:43109812-43111584	$7.40\text{x}10^{-66}$	$2.70\text{x}10^{-72}$	$2.20\text{x}10^{-8}$	*ZNF131*
chr11:3575396-3576572	$9.90\text{x}10^{-66}$	$1.70\text{x}10^{-64}$	$8.70\text{x}10^{-22}$	NA
chr12:102452772-102453798	$1.30\text{x}10^{-65}$	$8.70\text{x}10^{-67}$	$3.80\text{x}10^{-11}$	*WASHC3*
chr6:14784693-14786009	$2.30\text{x}10^{-64}$	$2.70\text{x}10^{-60}$	$5.40\text{x}10^{-20}$	NA
chr11:89972864-89974548	$3.30\text{x}10^{-63}$	$1.00\text{x}10^{-57}$	$1.00\text{x}10^{-22}$	*CHORDC1*
chr1:167483235-167484297	$5.30\text{x}10^{-63}$	$4.50\text{x}10^{-69}$	$7.80\text{x}10^{-5}$	*CD247*

Table C.3: Top 25 $CD4^+$ cell type specific ATAC peaks. The top 25 cell type specific ATAC peaks based on ANOVA global p-value.

ATAC Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr7:55637060-55638340	$7.50\text{x}10^{-115}$	$1.90\text{x}10^{-114}$	$1.30\text{x}10^{-30}$	*VOPP1*
chr12:10706484-10707986	$1.00\text{x}10^{-114}$	$2.60\text{x}10^{-86}$	$2.00\text{x}10^{-78}$	*KLRA1P*
chr7:151724341-151725631	$5.40\text{x}10^{-110}$	$9.40\text{x}10^{-76}$	$5.80\text{x}10^{-78}$	*GALNT11*
chr18:67562909-67564002	$1.30\text{x}10^{-100}$	$2.30\text{x}10^{-94}$	$5.80\text{x}10^{-48}$	*CD226*
chr5:172345340-172346567	$1.50\text{x}10^{-96}$	$1.70\text{x}10^{-62}$	$9.40\text{x}10^{-77}$	*ERGIC1*
chr2:87017419-87018892	$5.60\text{x}10^{-96}$	$2.70\text{x}10^{-68}$	$6.10\text{x}10^{-67}$	*CD8A*
chr12:10711437-10712957	$2.20\text{x}10^{-95}$	$2.40\text{x}10^{-65}$	$1.90\text{x}10^{-63}$	*KLRA1P*
chr16:75653883-75655332	$1.10\text{x}10^{-91}$	$1.30\text{x}10^{-84}$	$6.40\text{x}10^{-41}$	*ADAT1*
chr11:134098201-134099362	$3.10\text{x}10^{-91}$	$2.20\text{x}10^{-78}$	$9.30\text{x}10^{-42}$	*VPS26B*
chr18:808136-809186	$1.80\text{x}10^{-90}$	$6.20\text{x}10^{-92}$	$8.70\text{x}10^{-15}$	*YES1*
chr4:42648219-42649747	$3.20\text{x}10^{-90}$	$3.00\text{x}10^{-63}$	$3.20\text{x}10^{-58}$	*ATP8A1*
chr2:170539259-170540696	$1.20\text{x}10^{-88}$	$5.40\text{x}10^{-66}$	$1.70\text{x}10^{-59}$	*CCDC173*
chr10:98430196-98431576	$1.80\text{x}10^{-87}$	$2.10\text{x}10^{-53}$	$1.70\text{x}10^{-70}$	*PIK3AP1*
chr9:80852809-80854009	$7.30\text{x}10^{-84}$	$2.30\text{x}10^{-60}$	$1.40\text{x}10^{-53}$	*CEP78*
chr2:182010364-182011802	$1.10\text{x}10^{-83}$	$7.20\text{x}10^{-72}$	$1.20\text{x}10^{-44}$	*NA*
chr2:191559829-191561264	$1.90\text{x}10^{-79}$	$2.70\text{x}10^{-55}$	$4.80\text{x}10^{-49}$	*NAB1*
chr14:53615051-53616123	$2.60\text{x}10^{-78}$	$3.10\text{x}10^{-69}$	$3.10\text{x}10^{-32}$	*DDHD1*
chr11:85461324-85462370	$1.10\text{x}10^{-77}$	$1.70\text{x}10^{-82}$	$2.50\text{x}10^{-14}$	*SYTL2*
chr7:134325431-134326578	$9.80\text{x}10^{-77}$	$4.90\text{x}10^{-51}$	$5.80\text{x}10^{-51}$	*BPGM*
chr13:32963734-32965507	$3.80\text{x}10^{-74}$	$8.00\text{x}10^{-51}$	$1.40\text{x}10^{-51}$	*BRCA2*
chr13:81051502-81052960	$1.90\text{x}10^{-73}$	$3.00\text{x}10^{-47}$	$1.10\text{x}10^{-56}$	NA
chr6:106541826-106542870	$2.20\text{x}10^{-73}$	$8.80\text{x}10^{-74}$	$3.10\text{x}10^{-12}$	*PRDM1*
chr3:114013408-114014489	$3.20\text{x}10^{-72}$	$8.70\text{x}10^{-71}$	$8.60\text{x}10^{-20}$	*TIGIT*
chr12:122711488-122712676	$4.30\text{x}10^{-72}$	$3.90\text{x}10^{-67}$	$1.10\text{x}10^{-31}$	*DIABLO*
chr12:10649883-10651371	$2.70\text{x}10^{-71}$	$6.10\text{x}10^{-45}$	$1.70\text{x}10^{-52}$	*KLRC1*

Table C.4: Top 25 $CD8^+$ cell type specific ATAC peaks. The top 25 cell type specific ATAC peaks based on ANOVA global p-value.

ATAC Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr3:46013555-46014829	$1.40\text{x}10^{-252}$	$2.40\text{x}10^{-166}$	$8.20\text{x}10^{-148}$	*FYCO1*
chr7:50423737-50425009	$5.60\text{x}10^{-242}$	$5.30\text{x}10^{-146}$	$1.60\text{x}10^{-175}$	*IKZF1*
chr1:92561315-92562797	$9.40\text{x}10^{-239}$	$5.90\text{x}10^{-155}$	$2.00\text{x}10^{-153}$	*BTBD8*
chr3:15335878-15337138	$4.20\text{x}10^{-238}$	$8.10\text{x}10^{-166}$	$2.20\text{x}10^{-159}$	*SH3BP5*
chr3:126190796-126191849	$6.70\text{x}10^{-237}$	$6.20\text{x}10^{-153}$	$2.00\text{x}10^{-158}$	*ZXDC*
chr1:121253010-121254245	$3.80\text{x}10^{-236}$	$5.00\text{x}10^{-166}$	$8.50\text{x}10^{-150}$	*EMBP1*
chr15:93778951-93780466	$3.70\text{x}10^{-224}$	$5.60\text{x}10^{-146}$	$5.10\text{x}10^{-151}$	NA
chr2:228234616-228236191	$1.30\text{x}10^{-220}$	$1.00\text{x}10^{-149}$	$1.00\text{x}10^{-143}$	*TM4SF20*
chr21:42738243-42739307	$1.10\text{x}10^{-219}$	$1.30\text{x}10^{-145}$	$4.20\text{x}10^{-140}$	*MX2*
chr8:128771630-128773174	$2.60\text{x}10^{-218}$	$2.00\text{x}10^{-149}$	$9.60\text{x}10^{-141}$	*MYC*
chr8:42122716-42124099	$1.40\text{x}10^{-217}$	$3.70\text{x}10^{-141}$	$3.20\text{x}10^{-127}$	*IKBKB*
chr14:69065024-69066320	$2.30\text{x}10^{-217}$	$3.50\text{x}10^{-144}$	$2.30\text{x}10^{-153}$	*RAD51B*
chr15:92404112-92405162	$2.30\text{x}10^{-211}$	$2.00\text{x}10^{-127}$	$2.60\text{x}10^{-135}$	*SLCO3A1*
chr16:17352598-17353895	$2.80\text{x}10^{-211}$	$8.80\text{x}10^{-138}$	$1.80\text{x}10^{-124}$	*XYLT1*
chr3:122487971-122489512	$3.40\text{x}10^{-211}$	$1.40\text{x}10^{-126}$	$5.00\text{x}10^{-137}$	*HSPBAP1*
chr14:90055574-90056772	$1.30\text{x}10^{-210}$	$2.00\text{x}10^{-136}$	$3.60\text{x}10^{-122}$	*FOXN3*
chr2:61153559-61155046	$2.20\text{x}10^{-210}$	$4.70\text{x}10^{-144}$	$3.30\text{x}10^{-138}$	*REL*
chr3:58320389-58321750	$2.10\text{x}10^{-208}$	$1.20\text{x}10^{-148}$	$3.60\text{x}10^{-128}$	*PXK*
chr2:46774635-46775952	$3.00\text{x}10^{-208}$	$2.60\text{x}10^{-155}$	$6.80\text{x}10^{-127}$	*RHOQ*
chr4:71757567-71758761	$4.20\text{x}10^{-208}$	$8.20\text{x}10^{-135}$	$2.20\text{x}10^{-123}$	*MOB1B*
chr22:30666528-30667594	$5.80\text{x}10^{-208}$	$5.20\text{x}10^{-128}$	$8.90\text{x}10^{-140}$	*OSM*
chr2:96925690-96927167	$2.00\text{x}10^{-207}$	$3.10\text{x}10^{-138}$	$3.60\text{x}10^{-133}$	*TMEM127*
chr13:99964416-99965912	$7.10\text{x}10^{-207}$	$5.10\text{x}10^{-126}$	$1.90\text{x}10^{-142}$	*MIR548AN*
chr2:219234890-219236238	$4.20\text{x}10^{-205}$	$7.40\text{x}10^{-136}$	$1.80\text{x}10^{-140}$	*CATIP*
chr5:74615480-74616859	$2.00\text{x}10^{-204}$	$2.80\text{x}10^{-134}$	$2.30\text{x}10^{-144}$	*HMGCR*

Table C.5: Top 25 CD14$^+$ cell type specific ATAC peaks. The top 25 cell type specific ATAC peaks based on ANOVA global p-value.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr21:46044091-46052269	$3.51x10^{-78}$	$1.49x10^{-48}$	$2.97x10^{-09}$	*KRTAP10-9*
chr5:56787758-56793786	$6.52x10^{-72}$	$4.68x10^{-55}$	$1.96x10^{-49}$	*ACTBL2*
chr17:78783330-78789130	$8.46x10^{-62}$	$9.06x10^{-66}$	$1.77x10^{-21}$	*RPTOR*
chr12:122808955-122814469	$5.61x10^{-60}$	$8.73x10^{-61}$	$6.76x10^{-23}$	*CLIP1*
chr12:123955319-123970475	$2.23x10^{-59}$	$2.95x10^{-71}$	$1.80x10^{-09}$	*RILPL1*
chr17:63532423-63537862	$8.66x10^{-59}$	$7.30x10^{-70}$	$2.53x10^{-07}$	*AXIN2*
chr7:5869158-5881357	$1.79x10^{-57}$	$1.89x10^{-69}$	$4.64x10^{-09}$	*ZNF815P*
chr10:35298790-35303920	$7.59x10^{-55}$	$2.83x10^{-36}$	$3.06x10^{-42}$	*CUL2*
chr12:12149423-12153935	$3.60x10^{-53}$	$1.07x10^{-59}$	$3.09x10^{-15}$	*ETV6*
chr11:3530917-3538383	$5.49x10^{-53}$	$3.74x10^{-57}$	$6.27x10^{-21}$	NA
chr10:87368634-87375302	$6.20x10^{-52}$	$4.01x10^{-63}$	$1.19x10^{-07}$	*GRID1*
chr20:42780135-42784435	$6.67x10^{-50}$	$3.93x10^{-56}$	$5.13x10^{-17}$	*JPH2*
chr2:54317755-54323122	$9.79x10^{-50}$	$5.38x10^{-58}$	$3.78x10^{-08}$	*ACYP2*
chr8:100707749-100718441	$2.22x10^{-49}$	$9.21x10^{-49}$	$4.54x10^{-25}$	*VPS13B*
chr14:97872681-97877735	$1.01x10^{-47}$	$5.88x10^{-52}$	$1.16x10^{-15}$	*NA*
chr15:91859793-91865920	$3.20x10^{-47}$	$9.72x10^{-52}$	$6.75x10^{-19}$	*SV2B*
chr11:117797883-117806546	$3.80x10^{-47}$	$3.06x10^{-23}$	$2.70x10^{-08}$	*TMPRSS13*
chr6:143089894-143098204	$4.77x10^{-47}$	$4.09x10^{-54}$	$9.01x10^{-11}$	*HIVEP2*
chr8:37129611-37136581	$5.27x10^{-47}$	$5.83x10^{-58}$	$9.83x10^{-12}$	NA
chr17:8980276-8985505	$5.59x10^{-47}$	$1.75x10^{-27}$	$1.06x10^{-06}$	*NTN1*
chr3:178750332-178759107	$4.11x10^{-46}$	$4.31x10^{-54}$	$8.33x10^{-14}$	*ZMAT3*
chr1:213841335-213845942	$4.55x10^{-46}$	$9.84x10^{-46}$	$1.05x10^{-21}$	NA
chr21:37512156-37520010	$2.07x10^{-45}$	$4.70x10^{-52}$	$4.71x10^{-10}$	*CBR3*
chr9:35936015-35946913	$2.34x10^{-45}$	$5.57x10^{-53}$	$8.99x10^{-16}$	*OR2S2*
chr5:55304975-55311637	$5.66x10^{-45}$	$1.08x10^{-47}$	$4.40x10^{-19}$	*IL6ST*

Table C.6: Top 25 $CD4^+$ cell type specific H3K27ac ChIPm peaks. The top 25 cell type specific H3K27ac ChIPm peaks based on global p-value from the ANOVA.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr2:87007226-87024346	$2.91x10^{-114}$	$1.09x10^{-84}$	$1.21x10^{-89}$	*CD8A*
chr19:10622857-10637235	$1.04x10^{-81}$	$9.85x10^{-67}$	$3.11x10^{-55}$	*S1PR5*
chr2:232254825-232263061	$2.29x10^{-80}$	$1.26x10^{-50}$	$7.33x10^{-64}$	*B3GNT7*
chr2:232416053-232423684	$1.74x10^{-77}$	$5.37x10^{-71}$	$7.46x10^{-40}$	*NMUR1*
chr19:13152540-13158370	$1.46x10^{-70}$	$8.56x10^{-64}$	$3.46x10^{-39}$	*NFIX*
chr19:46269910-46275660	$4.25x10^{-58}$	$9.10x10^{-66}$	$4.09x10^{-12}$	*DMPK*
chr2:232385481-232400404	$9.41x10^{-58}$	$2.09x10^{-65}$	$6.34x10^{-19}$	*NMUR1*
chr19:3782465-3792410	$1.16x10^{-56}$	$1.65x10^{-57}$	$7.77x10^{-26}$	*MATK*
chr5:1149498-1157766	$5.59x10^{-56}$	$1.81x10^{-53}$	$1.10x10^{-28}$	*SLC12A7*
chr19:51668454-51673854	$4.60x10^{-54}$	$5.50x10^{-58}$	$9.23x10^{-17}$	*SIGLEC17P*
chr1:202161485-202174897	$4.82x10^{-54}$	$5.00x10^{-56}$	$1.85x10^{-25}$	*LGR6*
chr17:45294738-45303718	$6.14x10^{-53}$	$1.10x10^{-51}$	$6.61x10^{-27}$	*MYL4*
chr12:10632482-10636733	$3.01x10^{-52}$	$1.13x10^{-36}$	$3.59x10^{-41}$	*KLRC1*
chr1:11749397-11756319	$1.94x10^{-51}$	$1.99x10^{-46}$	$1.61x10^{-28}$	*DRAXIN*
chr17:47477554-47485794	$8.74x10^{-51}$	$7.47x10^{-47}$	$1.98x10^{-29}$	*PHB*
chr17:80730591-80736331	$1.12x10^{-50}$	$3.80x10^{-51}$	$2.86x10^{-22}$	*TBCD*
chr19:4706668-4712835	$2.81x10^{-50}$	$5.20x10^{-38}$	$3.33x10^{-36}$	*DPP9*
chr1:204040421-204046343	$2.01x10^{-49}$	$1.43x10^{-44}$	$3.72x10^{-28}$	*SOX13*
chr4:6196338-6205181	$9.40x10^{-49}$	$1.98x10^{-55}$	$4.00x10^{-18}$	*JAKMIP1*
chr5:1592008-1596685	$9.66x10^{-49}$	$1.93x10^{-51}$	$1.70x10^{-18}$	*SDHAP3*
chr1:202611847-202614847	$1.52x10^{-47}$	$4.82x10^{-49}$	$4.77x10^{-08}$	*SYT2*
chr5:172342647-172351923	$1.17x10^{-46}$	$3.26x10^{-45}$	$4.32x10^{-25}$	*ERGIC1*
chr11:134278363-134284609	$8.15x10^{-45}$	$6.87x10^{-48}$	$2.24x10^{-17}$	*B3GAT1*
chr4:15960899-15967700	$3.42x10^{-44}$	$7.96x10^{-37}$	$1.11x10^{-31}$	*FGFBP2*
chr12:54409641-54413992	$4.95x10^{-44}$	$3.12x10^{-42}$	$6.58x10^{-22}$	*HOXC4*

Table C.7: Top 25 $CD8^+$ cell type specific H3K27ac ChIPm peaks. The top 25 cell type specific H3K27ac ChIPm peaks based on global p-value from the ANOVA.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr6:146862661-146876404	$3.27\text{x}10^{-257}$	$1.76\text{x}10^{-121}$	$5.33\text{x}10^{-215}$	*RAB32*
chr7:81393373-81401798	$5.23\text{x}10^{-194}$	$1.16\text{x}10^{-97}$	$1.99\text{x}10^{-148}$	*HGF*
chr8:16216360-16222492	$3.42\text{x}10^{-189}$	$2.85\text{x}10^{-89}$	$1.05\text{x}10^{-145}$	NA
chr12:16498755-16508479	$4.73\text{x}10^{-181}$	$2.13\text{x}10^{-93}$	$5.81\text{x}10^{-140}$	*MGST1*
chr13:41749745-41754883	$6.54\text{x}10^{-177}$	$2.64\text{x}10^{-88}$	$5.17\text{x}10^{-138}$	*KBTBD7*
chr11:70167370-70176639	$8.86\text{x}10^{-172}$	$4.95\text{x}10^{-93}$	$1.96\text{x}10^{-134}$	*PPFIA1*
chr19:10454068-10460782	$8.83\text{x}10^{-168}$	$2.05\text{x}10^{-98}$	$3.78\text{x}10^{-127}$	*TYK2*
chr1:35808213-35815676	$1.86\text{x}10^{-163}$	$7.42\text{x}10^{-81}$	$2.43\text{x}10^{-130}$	*ZMYM4*
chr1:186647436-186651954	$1.27\text{x}10^{-158}$	$1.42\text{x}10^{-77}$	$6.07\text{x}10^{-125}$	*PTGS2*
chr14:89943988-89956437	$3.84\text{x}10^{-155}$	$1.38\text{x}10^{-92}$	$7.32\text{x}10^{-118}$	*FOXN3*
chr2:29171985-29178305	$4.64\text{x}10^{-155}$	$3.13\text{x}10^{-88}$	$8.49\text{x}10^{-118}$	*WDR43*
chr9:110308230-110313248	$1.27\text{x}10^{-154}$	$4.90\text{x}10^{-86}$	$2.33\text{x}10^{-116}$	NA
chr10:28686134-28695969	$5.21\text{x}10^{-153}$	$1.87\text{x}10^{-81}$	$3.78\text{x}10^{-127}$	NA
chr9:138955858-138963386	$2.40\text{x}10^{-152}$	$7.37\text{x}10^{-86}$	$3.40\text{x}10^{-116}$	*NACC2*
chr10:102011379-102019869	$2.68\text{x}10^{-150}$	$5.06\text{x}10^{-83}$	$5.41\text{x}10^{-117}$	*CWF19L1*
chr10:90517098-90522413	$1.72\text{x}10^{-149}$	$1.14\text{x}10^{-69}$	$9.01\text{x}10^{-131}$	*LIPN*
chr1:73098962-73104375	$4.27\text{x}10^{-149}$	$7.38\text{x}10^{-76}$	$9.97\text{x}10^{-106}$	NA
chr4:130245485-130252325	$5.73\text{x}10^{-149}$	$4.74\text{x}10^{-77}$	$1.25\text{x}10^{-118}$	NA
chr8:97770228-97776807	$9.91\text{x}10^{-149}$	$4.30\text{x}10^{-73}$	$2.99\text{x}10^{-132}$	*CPQ*
chr3:87136297-87140821	$1.12\text{x}10^{-148}$	$4.69\text{x}10^{-74}$	$1.79\text{x}10^{-111}$	NA
chr2:185628448-185633723	$9.45\text{x}10^{-148}$	$3.03\text{x}10^{-66}$	$3.81\text{x}10^{-102}$	*ZNF804A*
chr4:142247019-142257261	$3.18\text{x}10^{-146}$	$8.74\text{x}10^{-68}$	$2.45\text{x}10^{-128}$	NA
chr1:71877495-71881740	$4.18\text{x}10^{-145}$	$3.21\text{x}10^{-72}$	$4.82\text{x}10^{-94}$	*NEGR1*
chr2:165576950-165581978	$6.75\text{x}10^{-145}$	$4.92\text{x}10^{-67}$	$3.58\text{x}10^{-107}$	*COBLL1*
chr11:15819030-15824611	$3.02\text{x}10^{-144}$	$4.66\text{x}10^{-69}$	$7.37\text{x}10^{-121}$	NA

Table C.8: Top 25 CD14$^+$ cell type specific H3K27ac ChIPm peaks. The top 25 cell type specific H3K27ac ChIPm peaks based on global p-value from the ANOVA.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr10:5604232-5608180	$7.57x10^{-94}$	$1.39x10^{-98}$	$1.95x10^{-17}$	*CALML3*
chr11:11174958-11180173	$7.16x10^{-82}$	$3.18x10^{-91}$	$2.17x10^{-13}$	NA
chr4:143475682-143483284	$9.20x10^{-82}$	$2.22x10^{-85}$	$6.50x10^{-30}$	*INPP4B*
chr2:198755065-198760787	$1.15x10^{-81}$	$7.65x10^{-86}$	$8.75x10^{-22}$	*PLCL1*
chr11:3531691-3535795	$4.32x10^{-75}$	$9.39x10^{-57}$	$4.51x10^{-21}$	NA
chr6:149814335-149819034	$1.40x10^{-73}$	$1.78x10^{-80}$	$6.74x10^{-09}$	*PPIL4*
chr13:52514670-52519127	$7.50x10^{-73}$	$1.48x10^{-76}$	$2.61x10^{-17}$	*ATP7B*
chr11:95417582-95423676	$2.40x10^{-71}$	$2.18x10^{-82}$	$2.03x10^{-11}$	NA
chr2:204715859-204722246	$1.54x10^{-68}$	$1.62x10^{-74}$	$3.53x10^{-22}$	*CTLA4*
chr4:143306432-143310732	$3.42x10^{-66}$	$7.16x10^{-70}$	$1.65x10^{-07}$	*INPP4B*
chr5:55394801-55400291	$5.89x10^{-66}$	$9.07x10^{-74}$	$1.06x10^{-14}$	*ANKRD55*
chr20:21536123-21542015	$5.10x10^{-63}$	$2.41x10^{-75}$	$3.22x10^{-18}$	*NKX2-2*
chr2:161339998-161343523	$2.62x10^{-61}$	$3.01x10^{-63}$	$1.59x10^{-07}$	*RBMS1*
chr17:74635996-74641589	$4.94x10^{-60}$	$2.02x10^{-63}$	$3.36x10^{-26}$	*ST6GALNAC1*
chr5:1253706-1257358	$1.10x10^{-58}$	$1.58x10^{-66}$	$5.22x10^{-08}$	*TERT*
chr19:40168114-40172320	$2.94x10^{-55}$	$3.19x10^{-65}$	$2.00x10^{-09}$	*LGALS16*
chr1:183855621-183860059	$1.38x10^{-54}$	$3.20x10^{-62}$	$1.06x10^{-07}$	*RGL1*
chr2:207393441-207397192	$1.04x10^{-53}$	$9.47x10^{-57}$	$1.60x10^{-10}$	*ADAM23*
chr7:142031330-142036363	$3.79x10^{-53}$	$6.94x10^{-60}$	$1.00x10^{-12}$	*TRBV2*
chr1:179799106-179803187	$7.68x10^{-52}$	$3.40x10^{-58}$	$2.75x10^{-08}$	*TOR1AIP2*
chr11:60831704-60835544	$2.96x10^{-51}$	$5.72x10^{-55}$	$5.32x10^{-07}$	*CD5*
chr8:41622462-41627349	$8.86x10^{-51}$	$1.34x10^{-59}$	$5.80x10^{-10}$	*ANK1*
chr17:63532561-63537740	$7.76x10^{-50}$	$2.53x10^{-62}$	$1.39x10^{-06}$	*AXIN2*
chr3:185783695-185787457	$2.51x10^{-47}$	$3.47x10^{-52}$	$2.18x10^{-10}$	*ETV5*
chr1:234542869-234546867	$8.72x10^{-47}$	$1.68x10^{-48}$	$8.93x10^{-20}$	*TARBP1*

Table C.9: Top 25 $CD4^+$ cell type specific H3K4me3 ChIPm peaks. The top 25 cell type specific H3K4me3 ChIPm peaks based on global p-value from the ANOVA.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr9:139852648-139857181	$2.24x10^{-92}$	$2.25x10^{-90}$	$2.42x10^{-31}$	*LCN12*
chr10:80783496-80788652	$1.33x10^{-89}$	$2.84x10^{-91}$	$2.41x10^{-26}$	*ZMIZ1*
chr22:45102898-45108871	$2.80x10^{-81}$	$2.11x10^{-77}$	$4.01x10^{-43}$	*PRR5*
chr1:25420688-25428629	$8.30x10^{-77}$	$5.92x10^{-74}$	$3.44x10^{-40}$	NA
chr14:35883634-35887888	$2.77x10^{-73}$	$9.89x10^{-78}$	$4.21x10^{-11}$	*NFKBIA*
chr12:10704080-10714569	$2.84x10^{-72}$	$1.36x10^{-50}$	$1.78x10^{-64}$	*KLRA1P*
chr4:42671931-42676048	$2.97x10^{-69}$	$1.57x10^{-69}$	$8.73x10^{-18}$	*ATP8A1*
chr15:74690715-74694368	$1.14x10^{-68}$	$4.85x10^{-69}$	$3.23x10^{-22}$	*SEMA7A*
chr2:87068164-87072324	$2.02x10^{-64}$	$4.23x10^{-48}$	$1.31x10^{-39}$	*CD8B*
chr12:10650223-10653885	$6.32x10^{-64}$	$8.54x10^{-42}$	$3.55x10^{-48}$	*KLRC1*
chr10:98854025-98857849	$1.48x10^{-63}$	$3.36x10^{-61}$	$3.67x10^{-23}$	*SLIT1*
chr5:146130210-146134386	$1.24x10^{-61}$	$1.64x10^{-69}$	$1.58x10^{-09}$	*PPP2R2B*
chr17:74672656-74676507	$5.98x10^{-60}$	$1.22x10^{-62}$	$2.22x10^{-12}$	*MXRA7*
chr17:80731203-80735898	$8.79x10^{-60}$	$2.87x10^{-54}$	$6.20x10^{-32}$	*TBCD*
chr10:79022374-79026083	$9.59x10^{-60}$	$6.59x10^{-48}$	$4.92x10^{-34}$	*KCNMA1*
chr11:2404454-2411989	$1.01x10^{-59}$	$3.19x10^{-64}$	$1.92x10^{-24}$	*CD81*
chr12:10746352-10757313	$2.20x10^{-59}$	$6.58x10^{-41}$	$5.92x10^{-48}$	*KLRA1P*
chr16:53100060-53104289	$1.33x10^{-57}$	$3.60x10^{-48}$	$2.48x10^{-36}$	*CHD9*
chr16:3866906-3871787	$1.53x10^{-57}$	$2.49x10^{-61}$	$2.38x10^{-22}$	*CREBBP*
chr2:234292121-234298435	$2.14x10^{-57}$	$2.83x10^{-54}$	$2.39x10^{-30}$	*DGKD*
chr19:13995948-13999561	$2.05x10^{-54}$	$6.05x10^{-41}$	$1.16x10^{-37}$	*C19orf57*
chr7:129173791-129177991	$8.72x10^{-54}$	$1.38x10^{-44}$	$1.94x10^{-33}$	*SMKR1*
chr6:41299926-41304027	$4.19x10^{-53}$	$6.16x10^{-57}$	$5.58x10^{-19}$	*NCR2*
chr9:82496674-82499948	$8.92x10^{-53}$	$5.38x10^{-57}$	$4.21x10^{-09}$	NA
chr1:205448158-205452788	$1.46x10^{-52}$	$3.24x10^{-53}$	$9.80x10^{-26}$	*CDK18*

Table C.10: Top 25 $CD8^+$ cell type specific H3K4me3 ChIPm peaks. The top 25 cell type specific H3K4me3 ChIPm peaks based on global p-value from the ANOVA.

ChIPm Peak Coordinates	Global p-value	CD14 vs CD4 p-value	CD14 vs CD8 p-value	Proximal Gene
chr11:67774940-67782105	$3.25\text{x}10^{-269}$	$6.37\text{x}10^{-183}$	$1.85\text{x}10^{-171}$	*ALDH3B1*
chr17:56353363-56358971	$1.88\text{x}10^{-264}$	$6.83\text{x}10^{-176}$	$3.10\text{x}10^{-159}$	*MPO*
chr7:99967906-99983319	$8.54\text{x}10^{-258}$	$1.87\text{x}10^{-174}$	$1.73\text{x}10^{-179}$	*PILRA*
chr19:849404-855075	$2.54\text{x}10^{-253}$	$5.21\text{x}10^{-156}$	$9.30\text{x}10^{-139}$	*ELANE*
chr2:113879621-113891154	$9.43\text{x}10^{-252}$	$2.75\text{x}10^{-165}$	$6.73\text{x}10^{-178}$	*IL1RN*
chr9:139263567-139270258	$1.02\text{x}10^{-251}$	$6.43\text{x}10^{-163}$	$5.64\text{x}10^{-174}$	*CARD9*
chr19:47810971-47821386	$1.66\text{x}10^{-250}$	$3.56\text{x}10^{-160}$	$2.40\text{x}10^{-179}$	*C5AR1*
chr19:55383732-55395008	$1.47\text{x}10^{-248}$	$2.41\text{x}10^{-171}$	$6.38\text{x}10^{-163}$	*FCAR*
chr11:82598418-82604754	$5.13\text{x}10^{-247}$	$2.69\text{x}10^{-165}$	$1.31\text{x}10^{-161}$	*PRCP*
chr1:244977020-244982941	$3.57\text{x}10^{-246}$	$2.30\text{x}10^{-157}$	$7.24\text{x}10^{-165}$	*COX20*
chr3:157152682-157158021	$1.10\text{x}10^{-243}$	$3.00\text{x}10^{-145}$	$4.73\text{x}10^{-177}$	*PTX3*
chr5:176312632-176330688	$1.58\text{x}10^{-240}$	$7.09\text{x}10^{-167}$	$2.46\text{x}10^{-162}$	*HK3*
chr12:50424581-50429172	$2.40\text{x}10^{-239}$	$3.86\text{x}10^{-150}$	$2.50\text{x}10^{-159}$	*RACGAP1*
chr20:23026257-23034036	$5.78\text{x}10^{-239}$	$1.52\text{x}10^{-150}$	$2.46\text{x}10^{-172}$	*THBD*
chr19:42125118-42135609	$3.75\text{x}10^{-226}$	$3.92\text{x}10^{-133}$	$5.93\text{x}10^{-170}$	*CEACAM4*
chr3:98478525-98486581	$1.30\text{x}10^{-225}$	$2.58\text{x}10^{-139}$	$1.30\text{x}10^{-159}$	*ST3GAL6*
chr22:39346118-39357953	$1.04\text{x}10^{-224}$	$1.45\text{x}10^{-155}$	$4.85\text{x}10^{-142}$	*APOBEC3A*
chr13:33921979-33926657	$5.67\text{x}10^{-224}$	$1.04\text{x}10^{-146}$	$1.55\text{x}10^{-139}$	*STARD13*
chr19:55082024-55091395	$9.71\text{x}10^{-221}$	$7.75\text{x}10^{-143}$	$1.65\text{x}10^{-157}$	*LILRA2*
chr15:68494617-68500589	$1.63\text{x}10^{-218}$	$4.61\text{x}10^{-126}$	$6.37\text{x}10^{-159}$	*CALML4*
chr8:19795044-19799667	$3.32\text{x}10^{-218}$	$3.63\text{x}10^{-118}$	$2.21\text{x}10^{-160}$	*LPL*
chr3:122042013-122047764	$9.49\text{x}10^{-216}$	$2.05\text{x}10^{-132}$	$4.20\text{x}10^{-154}$	*CSTA*
chr13:50214664-50219396	$1.02\text{x}10^{-214}$	$3.17\text{x}10^{-143}$	$2.85\text{x}10^{-132}$	*ARL11*
chr20:48884834-48902242	$3.91\text{x}10^{-214}$	$1.34\text{x}10^{-147}$	$1.13\text{x}10^{-143}$	NA
chr1:153668841-153672665	$9.39\text{x}10^{-214}$	$7.43\text{x}10^{-131}$	$5.20\text{x}10^{-117}$	*NPR1*

Table C.11: Top 25 CD14^{+} cell type specific H3K4me3 ChIPm peaks. The top 25 cell type specific H3K4me3 ChIPm peaks based on global p-value from the ANOVA.

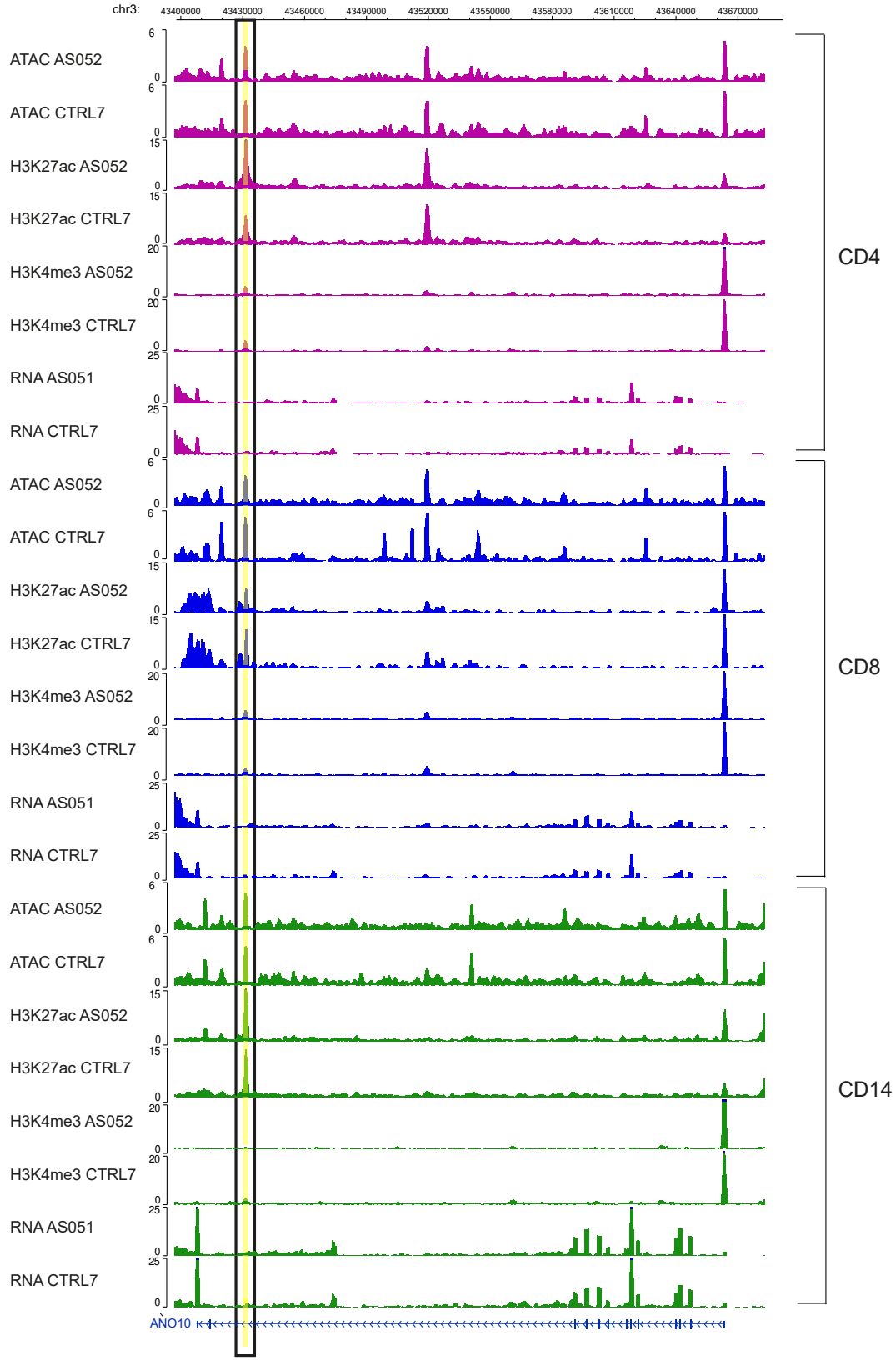


Figure C.1: Chromatin accessibility and histone modifications over *ANO10*. WashU Epigenome Browser view of the results of ATAC, H3K27ac ChIP, H3K4me3 ChIP, and RNA-Seq across $CD4^+$, $CD8^+$, and $CD14^+$ cells over *ANO10*. One AS patient and one control sample are shown for each technique and cell type. The $CD14^+$ differential H3K27ac peak over the intron of *ANO10* is highlighted in yellow.

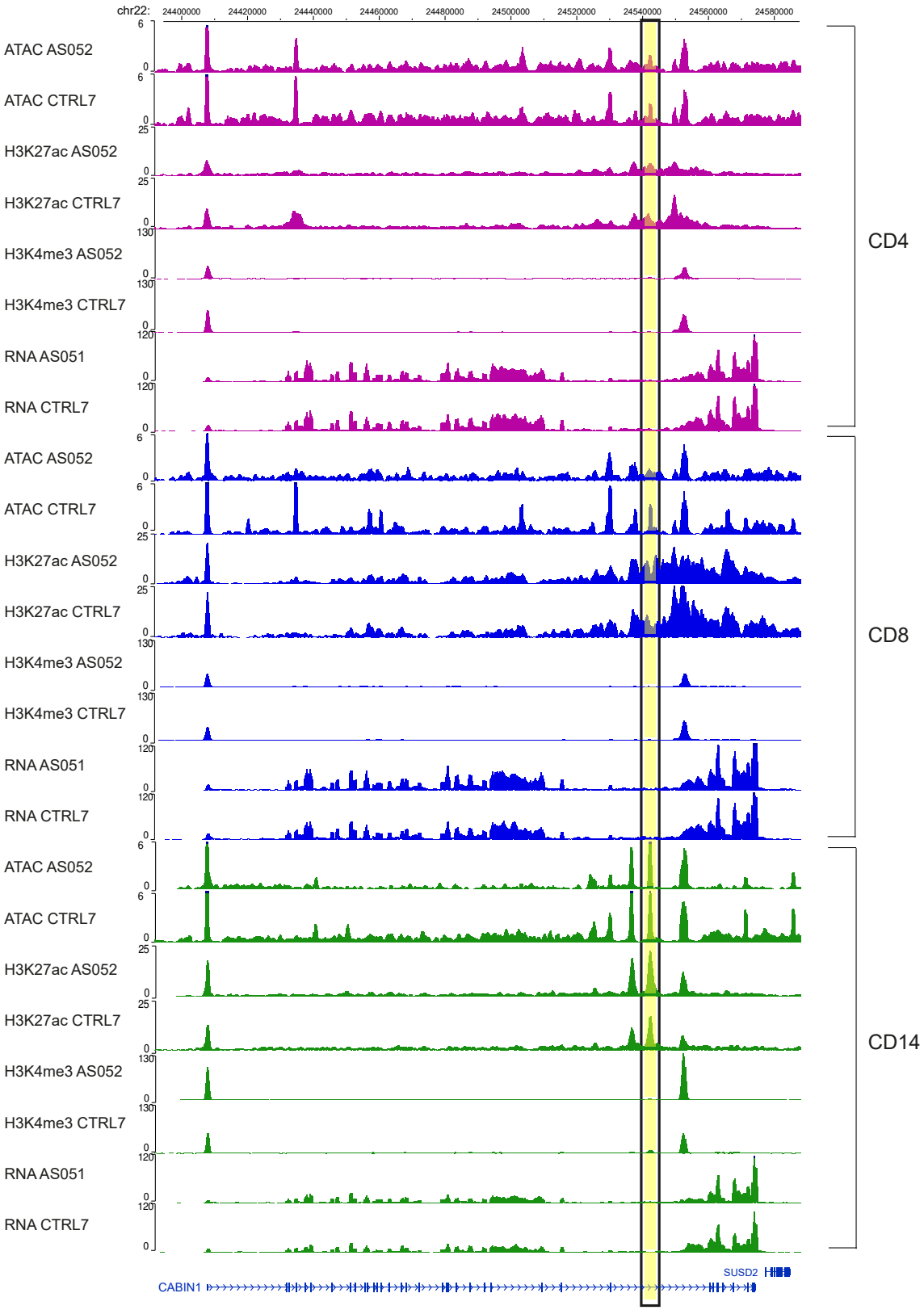


Figure C.2: Chromatin accessibility and histone modifications over *CABIN1*. WashU Epigenome Browser view of the results of ATAC, H3K27ac ChIP, H3K4me3 ChIP, and RNA-Seq across $CD4^+$, $CD8^+$, and $CD14^+$ cells over *CABIN1*. One AS patient and one control sample are shown for each technique and cell type. The $CD14^+$ differential H3K27ac peak over the intron of *CABIN1* is highlighted in yellow.

Bibliography

Adamopoulos, I. E., Tessmer, M., Chao, C.-C., et al. (2011). "IL-23 Is Critical for Induction of Arthritis, Osteoclast Formation, and Maintenance of Bone Mass". *Journal of Immunology* 187: pp. 951–959.

Adams, D., Altucci, L., Antonarakis, S. E., et al. (2012). "BLUEPRINT to decode the epigenetic signature written in blood". *Nature biotechnology* 30: p. 224.

Adams, M. D., Kelley, J. M., Gocayne, J. D., et al. (1991). "Complementary DNA sequencing: expressed sequence tags and human genome project". *Science* 252: pp. 1651–1656.

Agelopoulos, M., McKay, D., and Mann, R. (2012). "Developmental Regulation of Chromatin Conformation by Hox Proteins in Drosophila". *Cell Reports* 1: pp. 350–359.

Akiyama, Y., Maesawa, C., Ogasawara, S., et al. (2003). "Cell-type-specific repression of the maspin gene is disrupted frequently by demethylation at the promoter region in gastric intestinal metaplasia and cancer cells." *The American journal of pathology* 163: pp. 1911–9.

Alasoo, K., Martinez, F. O., Hale, C., et al. (2015). "Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription". *Scientific reports* 5: p. 12524.

Alge, C. S., Hauck, S. M., Priglinger, S. G., et al. (2006). "Differential protein profiling of primary versus immortalized human RPE cells identifies expression patterns associated with cytoskeletal remodeling and cell survival". *Journal of proteome research* 5: pp. 862–878.

Allen, R. L., O'Callaghan, C. A., McMichael, A. J., et al. (1999). "Cutting edge: HLA-B27 can form a novel β2-microglobulin-free heavy chain homodimer structure". *The Journal of Immunology* 162: pp. 5045–5048.

Altshuler, D. M., Gibbs, R. A., Peltonen, L., et al. (2010). "Integrating common and rare genetic variation in diverse human populations". *Nature* 467: pp. 52–58.

Amrolia, P. J., Gabbard, W., Cunningham, J. M., et al. (1998). "Maximal activity of an erythroid-specific enhancer requires the presence of specific protein binding sites in linked promoters". *Journal of Biological Chemistry* 273: pp. 13593–13598.

Amsen, D., Helbig, C., and Backer, R. A. (2015). "Notch in T cell differentiation: all things considered". *Trends in immunology* 36: pp. 802–814.

Anders, S., Pyl, P. T., and Huber, W. (2015). "HTSeq—a Python framework to work with high-throughput sequencing data". *Bioinformatics* 31: pp. 166–169.

Andersson, R., Gebhard, C., Miguel-Escalada, I., et al. (2014). "An atlas of active enhancers across human cell types and tissues". *Nature* 507: pp. 455–461.

Antonelli, M., Khan, M. A., and Magrey, M. N. (2015). "Differential adverse events between TNF-α inhibitors and IL-17 axis inhibitors for the treatment of spondyloarthritis". *Current Treatment Options in Rheumatology* 1: pp. 239–254.

Anzalone, A. V., Randolph, P. B., Davis, J. R., et al. (2019). "Search-and-replace genome editing without double-strand breaks or donor DNA". *Nature* 576: pp. 149–157.

Aoki, M., Aoki, H., Ramanathan, R., et al. (2016). "Sphingosine-1-phosphate signaling in immune cells and inflammation: roles and therapeutic potential". *Mediators of inflammation* 2016: p. 11.

Appel, H., Ruiz-Heiland, G., Listing, J., et al. (2009). "Altered skeletal expression of sclerostin and its link to radiographic progression in ankylosing spondylitis". *Arthritis & Rheumatism* 60: pp. 3257–3262.

Arévalo, M., Masmitjà, J. G., Moreno, M., et al. (2018). "Influence of HLA-B27 on the Ankylosing Spondylitis phenotype: results from the REGISPONSER database". *Arthritis research & therapy* 20: p. 221.

Arnason, B., Jacobs, G, Hanlon, M, et al. (1999). "TNF neutralization in MS-Results of a randomized, placebo-controlled multicenter study". *Neurology* 53.

Awad, O., Sarkar, C., Panicker, L. M., et al. (2015). "Altered TFEB-mediated lysosomal biogenesis in Gaucher disease iPSC-derived neuronal cells". *Human molecular genetics* 24: pp. 5775–5788.

Bannister, A. J. and Kouzarides, T. (2011). "Regulation of chromatin by histone modifications". *Cell research* 21: pp. 381–395.

Bannister, A. J., Zegerman, P., Partridge, J. F., et al. (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain". *Nature* 410: pp. 120–124.

Baraliakos, X., Kiltz, U., Peters, S., et al. (2016). "Efficiency of treatment with non-steroidal anti-inflammatory drugs according to current recommendations in patients with radiographic and non-radiographic axial spondyloarthritis". *Rheumatology* 56: pp. 95–102.

Baron, R. and Rawadi, G. (2007). "Targeting the Wnt/β-catenin pathway to regulate bone formation in the adult skeleton". *Endocrinology* 148: pp. 2635–2643.

Barrett, J. C., Hansoul, S., Nicolae, D. L., et al. (2008). "Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease". *Nature Genetics* 40: pp. 955–962.

Barski, A., Cuddapah, S., Cui, K., et al. (2007). "High-resolution profiling of histone methylations in the human genome". *Cell* 129: pp. 823–837.

Battle, A. and Montgomery, S. B. (2014). "Determining causality and consequence of expression quantitative trait loci". *Human Genetics* 133: pp. 727–735.

Baxter, J, Sen, N, Martínez, V. L., et al. (2011). "Positive supercoiling of mitotic DNA drives decatenation by topoisomerase II in eukaryotes". *Science* 331: pp. 1328–1332.

Bay-Jensen, A. C., Karsdal, M. A., Vassiliadis, E., et al. (2013). "Circulating citrullinated vimentin fragments reflect disease burden in ankylosing spondylitis and have prognostic capacity for radiographic progression". *Arthritis & Rheumatism* 65: pp. 972–980.

Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., et al. (2013). "Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis". *Nature genetics* 45: pp. 1353–1360.

Beguinot, L., Yamazaki, H., Pastan, I., et al. (1995). "Biochemical characterization of human GCF transcription factor in tumor cells." *Cell Growth Differ.* 6: pp. 699–706.

Belton, J.-M., McCord, R. P., Gibcus, J. H., et al. (2012). "Hi-C: a comprehensive technique to capture the conformation of genomes." *Methods* 58: pp. 268–276.

Benjamin, R and Parham, P (1992). "HLA-B27 and disease: a consequence of inadvertent antigen presentation?" *Rheumatic diseases clinics of North America* 18: pp. 11–21.

Bernier, F. P., Caluseriu, O., Ng, S., et al. (2012). "Haploinsufficiency of SF3B4, a component of the pre-mRNA spliceosomal complex, causes Nager syndrome." *American journal of human genetics* 90: pp. 925–933.

Bernstein, B. E., Stamatoyannopoulos, J. A., Costello, J. F., et al. (2010). "The NIH roadmap epigenomics mapping consortium". *Nature biotechnology* 28: p. 1045.

Blair, H. A. (2019). "Secukinumab: A Review in Ankylosing Spondylitis". *Drugs* 79: pp. 433–443.

Blair, H. A. and Dhillon, S. (2016). "Secukinumab: a review in ankylosing spondylitis". *Drugs* 76: pp. 1023–1030.

Blanco-Gómez, A., Castillo-Lluva, S., Mar Sáez-Freire, M. del, et al. (2016). "Missing heritability of complex diseases: Enlightenment by genetic variants from intermediate phenotypes". *Bioessays* 38: pp. 664–673.

Blaschek, A., Kries, R. v., Lohse, P., et al. (2018). "TNFRSF1A and MEFV mutations in childhood onset multiple sclerosis". *European Journal of Paediatric Neurology* 22: pp. 72–81.

Blomen, V. A., Májek, P., Jae, L. T., et al. (2015). "Gene essentiality and synthetic lethality in haploid human cells". *Science* 350: pp. 1092–1096.

Boland, M. J., Nazor, K. L., Tran, H. T., et al. (2017). "Molecular analyses of neurogenic defects in a human pluripotent stem cell model of fragile X syndrome". *Brain* 140: pp. 582–598.

Bonadies, N., Göttgens, B., and Calero-Nieto, F. J. (2015). "The LMO2-25 Region Harbours GATA2-Dependent Myeloid Enhancer and RUNX-Dependent T-Lymphoid Repressor Activity". *PloS one* 10: e0131577.

Bossone, S. A., Asselin, C., Patel, A. J., et al. (1992). "MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcriptional initiation and termination." *Proceedings of the National Academy of Sciences* 89: pp. 7452–7456.

Bowness, P. (2015). "HLA-B27". *Annual review of immunology* 33: pp. 29–48.

Bowness, P., Ridley, A., Shaw, J., et al. (2011). "Th17 cells expressing KIR3DL2+ and responsive to HLA-B27 homodimers are increased in ankylosing spondylitis". *The Journal of Immunology* 186: pp. 2672–2680.

Braun, J., Baraliakos, X., Deodhar, A., et al. (2019). "Secukinumab shows sustained efficacy and low structural progression in ankylosing spondylitis: 4-year results from the MEASURE 1 study". *Rheumatology* 58: pp. 859–868.

Brewerton, D. A., Hart, F. D., Nicholls, A, et al. (1973). "Ankylosing spondylitis and HL-A 27." *Lancet* 1: pp. 904–7.

Brody, J. A., Morrison, A. C., Bis, J. C., et al. (2017). "Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology". *Nature genetics* 49: pp. 1560–1563.

Brown, M. A., Pile, K. D., Kennedy, L. G., et al. (1996). "HLA class I associations of ankylosing spondylitis in the white population in the United Kingdom." *Annals of the rheumatic diseases* 55: pp. 268–70.

Brown, M. A., Laval, S. H., Brophy, S, et al. (2000). "Recurrence risk modelling of the genetic susceptibility to ankylosing spondylitis." *Annals of the rheumatic diseases* 59: pp. 883–6.

Brown, M. and Bradbury, L. A. (2017). "New approaches in ankylosing spondylitis". *Medical Journal of Australia* 206: pp. 192–194.

Brown, M. A., Kennedy, L. G., Macgregor, A. J., et al. (1997). "Susceptibility to ankylosing spondylitis in twins the role of genes, HLA, and the environment". *Arthritis & Rheumatism* 40: pp. 1823–1828.

Brown, M. A., Kenna, T., and Wordsworth, B. P. (2016). "Genetics of ankylosing spondylitis—insights into pathogenesis". *Nature Reviews Rheumatology* 12: pp. 81–91.

Buchrieser, J. (2016). "Understanding human mononuclear phagocyte ontogeny using human induced pluripotent stem cells (iPSCs)". PhD thesis. University of Oxford.

Buchrieser, J., James, W., and Moore, M. D. (2017). "Human induced pluripotent stem cell-derived macrophages share ontogeny with MYB-independent tissue-resident macrophages". *Stem cell reports* 8: pp. 334–345.

Buckle, A., Nozawa, R.-s., Kleinjan, D. A., et al. (2018). "Functional characteristics of novel pancreatic Pax6 regulatory elements". *Human molecular genetics* 27: pp. 3434–3448.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., et al. (2013). "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." *Nature methods* 10: pp. 1213–8.

Bunt, M. Van de, Cortes, A., Brown, M. A., et al. (2015). "Evaluating the Performance of Fine-Mapping Strategies at Common Variant GWAS Loci." *PLoS genetics* 11: e1005535.

Burton, P. R., Clayton, D. G., Cardon, L. R., et al. (2007). "Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants." *Nature genetics* 39: pp. 1329–37.

Buske, B., Szydlowski, M., Verkenne, C., et al. (2011). "Estimating myostatin gene effect on milk performance traits using estimated gene content for a large number of non-genotyped cows". *animal* 5: pp. 43–47.

Calderon, D., Nguyen, M. L., Mezger, A., et al. (2019). "Landscape of stimulation-responsive chromatin across diverse human immune cells". *Nature Genetics*: pp. 1–12.

Calin, A, Marder, A, Becks, E, et al. (1983). "Genetic differences between B27 positive patients with ankylosing spondylitis and B27 positive healthy controls". *Arthritis & Rheumatology* 26: pp. 1460–1464.

Caminero, A., Comabella, M., and Montalban, X. (2011). "Tumor necrosis factor alpha (TNF-α), anti-TNF-α and demyelination revisited: an ongoing story". *Journal of neuroimmunology* 234: pp. 1–6.

Campenhout, C. V., Cabochette, P., Veillard, A.-C., et al. (2019). "Guidelines for optimized gene knockout using CRISPR/Cas9". *BioTechniques* 66: pp. 295–302.

Canver, M. C., Smith, E. C., Sher, F., et al. (2015). "BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis". *Nature* 527: pp. 192–197.

Cao, X., Yakala, G. K., Hil, F. E. van den, et al. (2019). "Differentiation and Functional Comparison of Monocytes and Macrophages from hiPSCs with Peripheral Blood Derivatives". *Stem cell reports* 12: pp. 1282–1297.

Capon, F., Allen, M. H., Ameen, M., et al. (2004). "A synonymous SNP of the corneodesmosin gene leads to increased mRNA stability and demonstrates association with psoriasis across diverse ethnic groups". *Human Molecular Genetics* 13: pp. 2361–2368.

Carpenter, L., Malladi, R., Yang, C.-T., et al. (2011). "Human induced pluripotent stem cells are capable of B-cell lymphopoiesis." *Blood* 117: pp. 4008–11.

Carroll, D. (2011). "Genome engineering with zinc-finger nucleases". *Genetics* 188: pp. 773–782.

Casneuf, T., Peer, Y. Van de, and Huber, W. (2007). "In situ analysis of cross-hybridisation on microarrays and the inference of expression correlation". *BMC bioinformatics* 8: p. 461.

Chabaud, M. and Miossec, P. (2001). "The combination of tumor necrosis factor α blockade with interleukin-1 and interleukin-17 blockade is more effective for controlling synovial inflammation and bone resorption in an ex vivo model". *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 44: pp. 1293–1303.

Chakravarti, A., Clark, A. G., and Mootha, V. K. (2013). "Distilling pathophysiology from complex disease genetics". *Cell* 155: pp. 21–26.

Chang, C. C., Chow, C. C., Tellier, L. C., et al. (2015). "Second-generation PLINK: rising to the challenge of larger and richer datasets". *Gigascience* 4: s13742–015.

Chang, S.-C., Momburg, F., Bhutani, N., et al. (2005). "The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism." *Proceedings of the National Academy of Sciences of the United States of America* 102: pp. 17107–12.

Charbin, A., Bouchoux, C., and Uhlmann, F. (2013). "Condensin aids sister chromatid decatenation by topoisomerase II". *Nucleic acids research* 42: pp. 340–348.

Chen, B., Li, J., He, C., et al. (2017). "Role of HLA-B27 in the pathogenesis of ankylosing spondylitis". *Molecular medicine reports* 15: pp. 1943–1951.

Chen, J., Namiki, S., Toma-Hirano, M., et al. (2008). "The role of CD11b in phagocytosis and dendritic cell development". *Immunology letters* 120: pp. 42–48.

Chen, L., Fischer, R., Peng, Y., et al. (2014). "Critical Role of Endoplasmic Reticulum Aminopeptidase 1 in Determining the Length and Sequence of Peptides Bound and Presented by HLA–B27". *Arthritis and Rheumatology* 66: pp. 284–294.

Chen, L., Ridley, A., Hammitzsch, A., et al. (2016a). "Silencing or inhibition of endoplasmic reticulum aminopeptidase 1 (ERAP1) suppresses free heavy chain expression and Th17 responses in ankylosing spondylitis." *Annals of the rheumatic diseases* 75: pp. 916–23.

Chen, M., Zhang, L., Ren, Y., et al. (2015). "Defective Function of CD24(+)CD38(+) Regulatory B Cells in Ankylosing Spondylitis." *DNA and cell biology* 35: pp. 88–95.

Chen, M. and Manley, J. L. (2009). "Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches". *Nature reviews Molecular cell biology* 10: pp. 741–754.

Chen, S.-J. (2019). "Minimizing off-target effects in CRISPR-Cas9 genome editing". *Cell Biology and Toxicology* 35: 399–401.

Chen, X., Shen, Y., Draper, W., et al. (2016b). "ATAC-see reveals the accessible genome by transposase-mediated imaging and sequencing". *Nature methods* 13: pp. 1013–1020.

Chung, C. Y., Khurana, V., Auluck, P. K., et al. (2013). "Identification and rescue of α-synuclein toxicity in Parkinson patient–derived neurons". *Science* 342: pp. 983–987.

Ciccia, F., Rizzo, A., and Triolo, G. (2016). "Subclinical gut inflammation in ankylosing spondylitis". *Current opinion in rheumatology* 28: pp. 89–96.

Comabella, M., Caminero, A. B., Malhotra, S., et al. (2013). "TNFRSF1A polymorphisms rs1800693 and rs4149584 in patients with multiple sclerosis". *Neurology* 80: pp. 2010–2016.

Cong, L., Ran, F. A., Cox, D., et al. (2013). "Multiplex Genome Engineering Using CRISPR/Cas Systems". *Science* 339: pp. 819–823.

Conigliaro, P., Scrivo, R., Valesini, G., et al. (2011). "Emerging role for NK cells in the pathogenesis of inflammatory arthropathies". *Autoimmunity reviews* 10: pp. 577–581.

Consortium, G. P. (2015). "A global reference for human genetic variation". *Nature* 526: pp. 68–74.

Corces, M. R., Buenrostro, J. D., Wu, B., et al. (2016). "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution". *Nature genetics* 48: pp. 1193–1203.

Corces, M. R., Trevino, A. E., Hamilton, E. G., et al. (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues". *Nature methods* 14: pp. 959–962.

Corona-Sanchez, E. G., Muñoz-Valle, J. F., Gonzalez-Lopez, L., et al. (2012). "-383 A/C tumor necrosis factor receptor 1 polymorphism and ankylosing spondylitis in Mexicans: a preliminary study". *Rheumatology International* 32: pp. 2565–2568.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., et al. (2014). "Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits." *Genome research* 24: pp. 1–13.

Cortes, A., Hadler, J., Pointon, J. P., et al. (2013). "Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci International Genetics of Ankylosing Spondylitis Consortium (IGAS)". *Nature Genetics* 45: pp. 730–738.

Costello, M.-E., Ciccia, F., Willner, D., et al. (2015). "Brief report: intestinal dysbiosis in ankylosing spondylitis". *Arthritis & rheumatology* 67: pp. 686–691.

Cradick, T. J., Fine, E. J., Antico, C. J., et al. (2013). "CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity". *Nucleic acids research* 41: pp. 9584–9592.

Crawford, G. E., Holt, I. E., Whittle, J., et al. (2005). "Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS)". *Genome Research* 16: pp. 123–131.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., et al. (2010). "Histone H3K27ac separates active from poised enhancers and predicts developmental state". *Proceedings of the National Academy of Sciences* 107: pp. 21931–21936.

Cubeñas-Potts, C. and Corces, V. G. (2015). "Architectural proteins, transcription, and the three-dimensional organization of the genome". *FEBS Letters* 589: pp. 2923–2930.

Dailey, L. (2015). "High throughput technologies for the functional discovery of mammalian enhancers: new approaches for understanding transcriptional regulatory network dynamics". *Genomics* 106: pp. 151–158.

Danoy, P., Pryce, K., Hadler, J., et al. (2010). "Association of variants at 1q32 and STAT3 with ankylosing spondylitis suggests genetic overlap with Crohn's disease". *PLoS Genetics* 6: pp. 1–5.

Daubon, T., Léon, C., Clarke, K., et al. (2019). "Deciphering the complex role of thrombospondin-1 in glioblastoma development". *Nature communications* 10: p. 1146.

Davidson, S. I., Liu, Y., Danoy, P. A., et al. (2011). "Association of STAT3 and TNFRSF1A with ankylosing spondylitis in Han Chinese." *Annals of the rheumatic diseases* 70: pp. 289–92.

Davie, K., Jacobs, J., Atkins, M., et al. (2015). "Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling". *PLoS genetics* 11: e1004994.

Davies, J. O., Telenius, J. M., McGowan, S. J., et al. (2016). "Multiplexed analysis of chromosome conformation at vastly improved sensitivity". *Nature methods* 13: pp. 74–80.

De Jager, P. L., Jia, X., Wang, J., et al. (2009). "Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci". *Nature genetics* 41: pp. 776–782.

De Sandre-Giovannoli, A. and Lévy, N. (2006). "Altered splicing in prelamin A-associated premature aging phenotypes." *Progress in molecular and subcellular biology* 44: pp. 199–232.

Dean, L. E., Jones, G. T., MacDonald, A. G., et al. (2013). "Global prevalence of ankylosing spondylitis". *Rheumatology* 53: pp. 650–657.

Dekker, J., Rippe, K., Dekker, M., et al. (2002). "Capturing Chromosome Conformation". *Science* 295: pp. 1306–1311.

Delaneau, O., Marchini, J., and Zagury, J.-F. (2012). "A linear complexity phasing method for thousands of genomes". *Nature methods* 9: pp. 179–181.

Deleuze, V., Chalhoub, E., El-Hajj, R., et al. (2007). "TAL-1/SCL and its partners E47 and LMO2 up-regulate VE-cadherin expression in endothelial cells". *Molecular and cellular biology* 27: pp. 2687–2697.

Devitt, A., Moffatt, O. D., Raykundalia, C., et al. (1998). "Human CD14 mediates recognition and phagocytosis of apoptotic cells". *Nature* 392: pp. 505–509.

Di Meglio, P., Di Cesare, A., Laggner, U., et al. (2011). "The IL23R R381Q Gene Variant Protects against Immune-Mediated Diseases by Impairing IL-

23-Induced Th17 Effector Response in Humans". *PLoS ONE* 6. Ed. by M. von Herrath: e17160.

Dimos, J. T., Rodolfa, K. T., Niakan, K. K., et al. (2008). "Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons". *science* 321: pp. 1218–1221.

Dixon, J. R., Selvaraj, S., Yue, F., et al. (2012). "Topological domains in mammalian genomes identified by analysis of chromatin interactions". *Nature* 485: pp. 376–380.

Dobin, A., Davis, C. A., Schlesinger, F., et al. (2013). "STAR: ultrafast universal RNA-seq aligner". *Bioinformatics* 29: pp. 15–21.

Dodé, C., André, M., Bienvenu, T., et al. (2002). "The enlarging clinical, genetic, and population spectrum of tumor necrosis factor receptor–associated periodic syndrome". *Arthritis & Rheumatism* 46: pp. 2181–2188.

Doench, J. G., Fusi, N., Sullender, M., et al. (2016). "Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9". *Nature biotechnology* 34: pp. 184–191.

Doerner, A. M. and Zuraw, B. L. (2009). "TGF-β 1 induced epithelial to mesenchymal transition (EMT) in human bronchial epithelial cells is enhanced by IL-1β but not abrogated by corticosteroids". *Respiratory research* 10: p. 100.

Dong, C. (2008). "Th17 cells in development: an updated view of their molecular identity and genetic programming". *Nature Reviews Immunology* 8: pp. 337–348.

Dostie, J., Richmond, T. A., Arnaout, R. A., et al. (2006). "Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements". *Genome Research* 16: pp. 1299–1309.

Doyle, C. and Strominger, J. L. (1987). "Interaction between CD4 and class II MHC molecules mediates cell adhesion". *Nature* 330: pp. 256–259.

Duncan, D. D., Stupakoff, A., Hedrick, S. M., et al. (1995). "A Myc-associated zinc finger protein binding site is one of four important functional regions in the CD4 promoter." *Molecular and Cellular Biology* 15: pp. 3179–3186.

Dunham, I., Kundaje, A., Aldred, S. F., et al. (2012). "An integrated encyclopedia of DNA elements in the human genome". *Nature* 489: pp. 57–74.

Duran, A., Kobak, S., Sen, N., et al. (2016). "Fecal calprotectin is associated with disease activity in patients with ankylosing spondylitis". *Bosnian journal of basic medical sciences* 16: p. 71.

Durbin, R. M., Altshuler, D. L., Durbin, R. M., et al. (2010). "A map of human genome variation from population-scale sequencing". *Nature* 467: pp. 1061–1073.

Ebringer, A. (1989). "The relationship between Klebsiella infection and ankylosing spondylitis." *Bailliere's clinical rheumatology* 3: pp. 321–338.

Egusquiaguirre, S. P., Yeh, J. E., Walker, S. R., et al. (2018). "The STAT3 target gene TNFRSF1A modulates the NF-κB pathway in breast Cancer cells". *Neoplasia* 20: pp. 489–498.

Eichler, E. E., Flint, J., Gibson, G., et al. (2010). "Missing heritability and strategies for finding the underlying causes of complex disease". *Nature Reviews Genetics* 11: pp. 446–450.

Eijsbouts, C. Q., Burren, O. S., Newcombe, P. J., et al. (2019). "Fine mapping chromatin contacts in capture Hi-C data". *BMC genomics* 20: p. 77.

El-Maarri, O. (2003). "DNA methylation and human disease." *Nature Reviews Genetics* 544: pp. 135–144.

Ellinghaus, D., Jostins, L., Spain, S. L., et al. (2016). "Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci". *Nature Genetics* 48: pp. 510–8.

Erlandsen, H. and Stevens, R. C. (1999). "The Structural Basis of Phenylketonuria". *Molecular Genetics and Metabolism* 68: pp. 103–125.

Ermert, M, Pantazis, C, Duncker, H.-R., et al. (2003). "In situ localization of TNFα/β, TACE AND TNF receptors TNF-R1 and TNF-R2 in control and LPS-treated lung tissue". *Cytokine* 22: pp. 89–100.

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

Evans, D. M., Spencer, C. C. A., Pointon, J. J., et al. (2011). "Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility." *Nature genetics* 43: pp. 761–7.

Evans, R., Patzak, I., Svensson, L., et al. (2009). "Integrins in immunity". *Journal of cell science* 122: pp. 215–225.

Evnouchidou, I., Weimershaus, M., Saveanu, L., et al. (2014). "ERAP1–ERAP2 Dimerization Increases Peptide-Trimming Efficiency". *The Journal of Immunology* 193: pp. 901–908.

Fairfax, B. P. and Knight, J. C. (2014). "Genetics of gene expression in immunity to infection". *Current Opinion in Immunology* 30: pp. 63–71.

Fairfax, B. P., Makino, S., Radhakrishnan, J., et al. (2012). "Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles". *Nature Genetics* 44: pp. 502–510.

Fairfax, B. P., Humburg, P., Makino, S., et al. (2014). "Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression." *Science* 343: p. 1246949.

Fang, H., Knezevic, B., Burnham, K. L., et al. (2016). "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits". *Genome medicine* 8: p. 129.

Fang, H., De Wolf, H., Knezevic, B., et al. (2019). "A genetics-led approach defines the drug target landscape of 30 immune-related traits". *Nature genetics* 51: pp. 1082–1091.

Farh, K. K.-H., Marson, A., Zhu, J., et al. (2015). "Genetic and epigenetic fine mapping of causal autoimmune disease variants." *Nature* 518: pp. 337–43.

Fattahi, M. J., Jamshidi, A. R., Mahmoudi, M., et al. (2018). "Evaluation of the efficacy and safety of β-d-mannuronic acid in patients with ankylosing spondylitis: a 12-week randomized, placebo-controlled, phase I/II clinical trial". *International immunopharmacology* 54: pp. 112–117.

Feldtkeller, E., Lemmel, E.-M., and Russell, A. S. (2003). "Ankylosing spondylitis in the pharaohs of ancient Egypt". *Rheumatology international* 23: pp. 1–5.

Fernandes, H. J. R., Hartfield, E. M., Christian, H. C., et al. (2016). "ER Stress and Autophagic Perturbations Lead to Elevated Extracellular α-Synuclein in

GBA-N370S Parkinson's iPSC-Derived Dopamine Neurons". *Stem Cell Reports* 6: pp. 342–356.

Fishilevich, S., Nudel, R., Rappaport, N., et al. (2017). "GeneHancer: genome-wide integration of enhancers and target genes in GeneCards". *Database* 2017.

Follenzi, A., Santambrogio, L., and Annoni, A. (2007). "Immune responses to lentiviral vectors". *Current gene therapy* 7: pp. 306–315.

Forrest, M. P., Zhang, H., Moy, W., et al. (2017). "Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci". *Cell Stem Cell* 21: pp. 305–318.

Fulco, C. P., Munschauer, M., Anyoha, R., et al. (2016). "Systematic mapping of functional enhancer–promoter connections with CRISPR interference". *Science* 354: pp. 769–773.

Gagliani, N., Vesely, M. C. A., Iseppon, A., et al. (2015). "Th17 cells transdifferentiate into regulatory T cells during resolution of inflammation". *Nature* 523: pp. 221–225.

Gaj, T., Guo, J., Kato, Y., et al. (2012). "Targeted gene knockout by direct delivery of zinc-finger nuclease proteins". *Nature methods* 9: pp. 805–807.

Gaj, T., Gersbach, C. A., Barbas, C. F., et al. (2013). "ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering." *Trends in biotechnology* 31: pp. 397–405.

Gallagher, M. D. and Chen-Plotkin, A. S. (2018). "The post-GWAS era: from association to function". *The American Journal of Human Genetics* 102: pp. 717–730.

Gameiro, S. R., Caballero, J. A., and Hodge, J. W. (2012). "Defining the molecular signature of chemotherapy-mediated lung tumor phenotype modulation and increased susceptibility to T-cell killing". *Cancer Biotherapy and Radiopharmaceuticals* 21: pp. 23–35.

García-Medel, N., Sanz-Bravo, A., Van Nguyen, D., et al. (2012). "Functional Interaction of the Ankylosing Spondylitis-associated Endoplasmic Reticulum Aminopeptidase 1 Polymorphism and HLA-B27 in Vivo". *Molecular & Cellular Proteomics* 11: pp. 1416–1429.

Garcia-Montoya, L., Gul, H., and Emery, P. (2018). "Recent advances in ankylosing spondylitis: understanding the disease and management". *F1000Research* 7.

Garrett, S., Jenkinson, T., Kennedy, L. G., et al. (1994). "A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index." *The Journal of rheumatology* 21: pp. 2286–2291.

Gate, R. E., Cheng, C. S., Aiden, A. P., et al. (2018). "Genetic determinants of co-accessible chromatin regions in activated T cells across humans". *Nature genetics* 50: pp. 1140–1150.

Gentileschi, S., Vitale, A., Rigante, D., et al. (2018). "Prompt Clinical Response to Secukinumab in Patients with Axial Spondyloarthritis: Real Life Observational Data from Three Italian Referral Centers." *The Israel Medical Association journal: IMAJ* 20: pp. 438–441.

Ghoreschi, K., Laurence, A., Yang, X.-P., et al. (2010). "Generation of pathogenic Th17 cells in the absence of TGF-β signalling". *Nature* 467: p. 967.

Gilly, A., Southam, L., Suveges, D., et al. (2018). "Very low-depth whole-genome sequencing in complex trait association studies". *Bioinformatics* 35: pp. 2555–61.

Glazier, A. M., Nadeau, J. H., and Aitman, T. J. (2002). "Finding genes that underlie complex traits". *science* 298: pp. 2345–2349.

Gong, Y., Lazaris, C., Sakellaropoulos, T., et al. (2018). "Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries". *Nature communications* 9: p. 542.

Gopalappa, R., Suresh, B., Ramakrishna, S., et al. (2018). "Paired D10A Cas9 nickases are sometimes more efficient than individual nucleases for gene disruption". *Nucleic acids research* 46: e71.

Gracey, E., Yao, Y., Green, B., et al. (2016). "Sexual dimorphism in the Th17 signature of ankylosing spondylitis". *Arthritis & rheumatology* 68: pp. 679–689.

Gratacos, J, Collado, A, Filella, X, et al. (1994). "Serum cytokines (IL-6, TNF-α, IL-1β and IFN-γ) in ankylosing spondylitis: a close correlation between serum IL-6 and disease activity and severity". *Rheumatology* 33: pp. 927–931.

Greenwald, W. W., Chiou, J., Yan, J., et al. (2019). "Pancreatic islet chromatin accessibility and conformation reveals distal enhancer networks of type 2 diabetes risk". *Nature communications* 10: p. 2078.

Gregory, A. P., Dendrou, C. A., Attfield, K. E., et al. (2012). "TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis". *Nature* 488: pp. 508–511.

GTEx Consortium (2015). "The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans". *Science* 348: pp. 648–60.

Guenther, M. G., Frampton, G. M., Soldner, F., et al. (2010). "Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells". *Cell stem cell* 7: pp. 249–257.

Gundry, M. C., Brunetti, L., Lin, A., et al. (2016). "Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9". *Cell reports* 17: pp. 1453–1461.

Guo, X. and Li, X.-J. (2015). "Targeted genome editing in primate embryos". *Cell research* 25: p. 767.

Gupta, A., Godwin, A. K., Vanderveer, L., et al. (2003). "Hypomethylation of the Synuclein γ Gene CpG Island Promotes Its Aberrant Expression in Breast Carcinoma and Ovarian Carcinoma". *Cancer Research* 63: pp. 664–673.

Haibel, H, Brandt, H., Song, I., et al. (2007). "No efficacy of subcutaneous methotrexate in active ankylosing spondylitis: a 16-week open-label trial". *Annals of the rheumatic diseases* 66: pp. 419–421.

Hammer, C., Wanitchakool, P., Sirianant, L., et al. (2015). "A coding variant of ANO10, affecting volume regulation of macrophages, is associated with Borrelia seropositivity". *Molecular medicine* 21: pp. 26–37.

Hammer, R. E., Maika, S. D., Richardson, J. A., et al. (1990). "Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human β2m: an animal model of HLA-B27-associated human disorders". *Cell* 63: pp. 1099–1112.

Hansen, K. H., Bracken, A. P., Pasini, D., et al. (2008). "A model for transmission of the H3K27me3 epigenetic mark". *Nature cell biology* 10: p. 1291.

Haraksingh, R. R. and Snyder, M. P. (2013). "Impacts of variation in the human genome on gene regulation". *Journal of molecular biology* 425: pp. 3970–3977.

Harismendy, O., Notani, D., Song, X., et al. (2011). "9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response". *Nature* 470: p. 264.

Heiland, G. R., Appel, H., Poddubnyy, D., et al. (2012). "High level of functional dickkopf-1 predicts protection from syndesmophyte formation in patients with ankylosing spondylitis". *Annals of the rheumatic diseases* 71: pp. 572–574.

Heintzman, N. D., Stuart, R. K., Hon, G., et al. (2007). "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome". *Nature genetics* 39: p. 311.

Heinz, S., Romanoski, C. E., Benner, C., et al. (2015). "The selection and function of cell type-specific enhancers". *Nature reviews Molecular cell biology* 16: p. 144.

Hendel, A., Bak, R. O., Clark, J. T., et al. (2015). "Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells". *Nature biotechnology* 33: pp. 985–989.

Hendrich, B. and Bickmore, W. (2001). "Human diseases with underlying defects in chromatin structure and modification". *Human Molecular Genetics* 10: pp. 2233–2242.

Henriksson, J., Chen, X., Gomes, T., et al. (2019). "Genome-wide CRISPR screens in T helper cells reveal pervasive crosstalk between activation and differentiation". *Cell* 176: pp. 882–896.

Hilton, I. B. and Gersbach, C. A. (2015). "Enabling functional genomics with genome engineering". *Genome research* 25: pp. 1442–1455.

Hnisz, D., Weintraub, A. S., Day, D. S., et al. (2016). "Activation of proto-oncogenes by disruption of chromosome neighborhoods". *Science* 351: pp. 1454–1458.

Hon, G., Wang, W., and Ren, B. (2009). "Discovery and annotation of functional chromatin signatures in the human genome". *PLoS computational biology* 5: e1000566.

Honold, L. and Nahrendorf, M. (2018). "Resident and monocyte-derived macrophages in cardiovascular disease". *Circulation research* 122: pp. 113–127.

Howie, B., Fuchsberger, C., Stephens, M., et al. (2012). "Fast and accurate genotype imputation in genome-wide association studies through pre-phasing." *Nature genetics* 44: pp. 955–9.

Hu, P., Zhang, W., Xin, H., et al. (2016). "Single cell isolation and analysis". *Frontiers in cell and developmental biology* 4: p. 116.

Huang, W.-N., Tso, T. K., Kuo, Y.-C., et al. (2012). "Distinct impacts of syndesmophyte formation on male and female patients with ankylosing spondylitis". *International journal of rheumatic diseases* 15: pp. 163–168.

Hudnall, S. D., Meng, H., Lozovatsky, L., et al. (2016). "Recurrent genetic defects in classical Hodgkin lymphoma cell lines". *Leukemia & lymphoma* 57: pp. 2890–2900.

Hughes, J. R., Roberts, N., McGowan, S., et al. (2014). "Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment". *Nature Genetics* 46: pp. 205–212.

Hull, K. M., Drewe, E., Aksentijevich, I., et al. (2002). "The TNF receptor-associated periodic syndrome (TRAPS): emerging concepts of an autoinflammatory disorder". *Medicine* 81: pp. 349–368.

Hunt, S. E., McLaren, W., Gil, L., et al. (2018). "Ensembl variation resources". *Database* 2018.

Idriss, H. T. and Naismith, J. H. (2000). "TNFα and the TNF receptor superfamily: Structure-function relationship (s)". *Microscopy research and technique* 50: pp. 184–195.

International HapMap Consortium (2003). "The International HapMap Project". *Nature* 426: pp. 789–796.

Iriguchi, S. and Kaneko, S. (2019). "Toward the development of true "off-the-shelf" synthetic T-cell immunotherapy". *Cancer science* 110: p. 16.

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

Jansen, R., Hottenga, J.-J., Nivard, M. G., et al. (2017). "Conditional eQTL analysis reveals allelic heterogeneity of gene expression". *Human molecular genetics* 26: pp. 1444–1451.

Javed, M. R., Sadaf, M., Ahmed, T., et al. (2018). "CRISPR-Cas system: history and prospects as a genome editing tool in microorganisms". *Current microbiology* 75: pp. 1675–1683.

Javierre, B. M., Burren, O. S., Wilder, S. P., et al. (2016). "Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters". *Cell* 167: pp. 1369–1384.

Jehuda, R. B., Shemer, Y., and Binah, O. (2018). "Genome editing in induced pluripotent stem cells using CRISPR/Cas9". *Stem Cell Reviews and Reports* 14: pp. 323–336.

Jiang, D., Liang, J., Fan, J., et al. (2005). "Regulation of lung injury and repair by Toll-like receptors and hyaluronan". *Nature medicine* 11: p. 1173.

Jiang, Y., Chen, B., Duan, C., et al. (2015). "Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system". *Applied and Environmental Microbiology* 81: pp. 2506–2514.

John, S., Sabo, P. J., Canfield, T. K., et al. (2013). "Genome-scale mapping of DNase I hypersensitivity". *Current protocols in molecular biology* 103: pp. 21–27.

Karaderi, T, Pointon, J., Wordsworth, T., et al. (2012). "Evidence of genetic association between TNFRSF1A encoding the p55 tumour necrosis factor receptor, and ankylosing spondylitis in UK Caucasians." *Clinical experimental rheumatology* 30: pp. 110–113.

Karlsson, K. R., Cowley, S., Martinez, F. O., et al. (2008). "Homogeneous monocytes and macrophages from human embryonic stem cells following coculture-free differentiation in M-CSF and IL-3". *Experimental hematology* 36: pp. 1167–1175.

Kasela, S., Kisand, K., Tserel, L., et al. (2017). "Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells". *PLoS genetics* 13: e1006643.

Keene, M. A., Corces, V, Lowenhaupt, K, et al. (1981). "DNase I hypersensitive sites in Drosophila chromatin occur at the 5' ends of regions of transcription." *Proceedings of the National Academy of Sciences of the United States of America* 78: pp. 143–6.

Kemanetzoglou, E. and Andreadou, E. (2017). "CNS demyelination with TNF-α blockers". *Current neurology and neuroscience reports* 17: p. 36.

Kent, W. J., Sugnet, C. W., Furey, T. S., et al. (2002). "The human genome browser at UCSC." *Genome research* 12: pp. 996–1006.

Kilpinen, H., Waszak, S. M., Gschwind, A. R., et al. (2013). "Coordinated Effects of Sequence Variation on DNA Binding, Chromatin Structure, and Transcription". *Science* 342: pp. 744–747.

Kim, S., Kim, D., Cho, S. W., et al. (2014). "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins". *Genome research* 24: pp. 1012–1019.

Kim, Y.-G. and Chandrasegaran, S. (1994). "Chimeric restriction endonuclease." *Proceedings of the National Academy of Sciences* 91: pp. 883–887.

Kim, Y.-G., Cha, J., and Chandrasegaran, S. (1996). "Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain". *Proceedings of the National Academy of Sciences* 93: pp. 1156–1160.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., et al. (2015). "Engineered CRISPR-Cas9 nucleases with altered PAM specificities". *Nature* 523: p. 481.

Klingberg, E, Oleröd, G, Hammarsten, O, et al. (2016). "The vitamin D status in ankylosing spondylitis in relation to intestinal inflammation, disease activity, and bone health: a cross-sectional study". *Osteoporosis International* 27: pp. 2027–2033.

Knight, J. C. (2012). "Resolving the variable genome and epigenome in human disease". *Journal of Internal Medicine* 271: pp. 379–391.

Kochan, G., Krojer, T., Harvey, D., et al. (2011). "Crystal structures of the endoplasmic reticulum aminopeptidase-1 (ERAP1) reveal the molecular basis for N-terminal peptide trimming". *Proceedings of the National Academy of Sciences* 108: pp. 7745–7750.

Kollnberger, S., Bird, L., Sun, M.-Y., et al. (2002). "Cell-surface expression and immune receptor recognition of HLA–B27 homodimers". *Arthritis & Rheumatism* 46: pp. 2972–2982.

Kolovos, P., Knoch, T. A., Grosveld, F. G., et al. (2012). "Enhancers and silencers: an integrated and simple model for their function". *Epigenetics & chromatin* 5: p. 1.

Konermann, S., Brigham, M. D., Trevino, A. E., et al. (2015). "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex". *Nature* 517: p. 583.

Korkmaz, G., Lopes, R., Ugalde, A. P., et al. (2016). "Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9". *Nature biotechnology* 34: p. 192.

Kruger, P., Saffarzadeh, M., Weber, A. N., et al. (2015). "Neutrophils: between host defence, immune modulation, and tissue injury". *PLoS pathogens* 11: e1004651.

Ksontini, R., MacKay, S. L., and Moldawer, L. L. (1998). "Revisiting the role of tumor necrosis factor α and the response to surgical injury and inflammation". *Archives of surgery* 133: pp. 558–567.

Ku, C. S., Loy, E. Y., Salim, A., et al. (2010). "The discovery of human genetic variations and their use as disease markers: past, present and future". *Journal of human genetics* 55: p. 403.

Kukurba, K. R. and Montgomery, S. B. (2015). "RNA sequencing and analysis". *Cold Spring Harbor Protocols* 2015: pp. 671–683.

Kümpfel, T, Hoffmann, L., Pellkofer, H, et al. (2008). "Multiple sclerosis and the TNFRSF1A R92Q mutation: clinical characteristics of 21 cases". *Neurology* 71: pp. 1812–1820.

Kümpfel, T., Hoffmann, L.-A., Rübsamen, H., et al. (2007). "Late-onset tumor necrosis factor receptor–associated periodic syndrome in multiple sclerosis patients carrying the TNFRSF1A R92Q mutation". *Arthritis & Rheumatism* 56: pp. 2774–2783.

Kwon, D. Y., Zhao, Y.-T., Lamonica, J. M., et al. (2017). "Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC". *Nature Communications* 8: p. 15315.

LaFleur, M. W., Nguyen, T. H., Coxe, M. A., et al. (2019). "A CRISPR-Cas9 delivery system for in vivo screening of genes in the immune system". *Nature communications* 10: p. 1668.

Lämmermann, T. and Kastenmüller, W. (2019). "Concepts of GPCR-controlled navigation in the immune system". *Immunological reviews* 289: pp. 205–231.

Landi, M., Maldonado-Ficco, H., Perez-Alamino, R., et al. (2017). "Gender differences among patients with primary ankylosing spondylitis and spondylitis associated with psoriasis and inflammatory bowel disease in an iberoamerican spondyloarthritis cohort: Erratum". *Medicine* 96: e6485.

Langmead, B., Trapnell, C., Pop, M., et al. (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome". *Genome biology* 10: R25.

Larson, M. H., Gilbert, L. A., Wang, X., et al. (2013). "CRISPR interference (CRISPRi) for sequence-specific control of gene expression". *Nature protocols* 8: p. 2180.

Lee, Y. W., Lee, K.-M., Lee, J. M., et al. (2019). "The usefulness of fecal calprotectin in assessing inflammatory bowel disease activity". *The Korean journal of internal medicine* 34: p. 72.

Lei, Y., Zhang, X., Su, J., et al. (2017). "Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein". *Nature Communications* 8: p. 16026.

Lelli, K. M., Slattery, M., and Mann, R. S. (2012). "Disentangling the many layers of eukaryotic transcriptional regulation." *Annual review of genetics* 46: pp. 43–68.

Lesurf, R., Cotto, K. C., Wang, G., et al. (2015). "ORegAnno 3.0: a community-driven resource for curated regulatory annotation". *Nucleic acids research* 44: pp. D126–D132.

Li, D., Hsu, S., Purushotham, D., et al. (2019). "WashU Epigenome Browser update 2019". *Nucleic acids research* 47: pp. 158–165.

Li, H., Handsaker, B., Wysoker, A., et al. (2009). "The sequence alignment/map format and SAMtools". *Bioinformatics* 25: pp. 2078–2079.

Li, L., Wu, L. P., and Chandrasegaran, S. (1992). "Functional domains in Fok I restriction endonuclease." *Proceedings of the National Academy of Sciences* 89: pp. 4275–4279.

Liang, X., Potter, J., Kumar, S., et al. (2015). "Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection". *Journal of biotechnology* 208: pp. 44–53.

Liao, B.-Y. and Zhang, J. (2008). "Null mutations in human and mouse orthologs frequently result in different phenotypes". *Proceedings of the National Academy of Sciences* 105: pp. 6987–6992.

Lieberman-Aiden, E., Berkum, N. L. van, Williams, L., et al. (2009). "Comprehensive mapping of long-range interactions reveals folding principles of the human genome." *Science* 326: pp. 289–93.

Limón-Camacho, L., Vargas-Rojas, M. I., Vázquez-Mellado, J., et al. (2012). "In vivo peripheral blood proinflammatory T cells in patients with ankylosing spondylitis". *The Journal of rheumatology* 39: pp. 830–835.

Lin, P., Bach, M., Asquith, M., et al. (2014). "HLA-B27 and human β2-microglobulin affect the gut microbiota of transgenic rats". *PloS one* 9: e105684.

Lin, Y., Zeng, C., Lu, Z., et al. (2019). "A novel homozygous splice-site variant of NCAPD2 gene identified in two siblings with primary microcephaly: The second case report". *Clinical genetics* 96: pp. 98–101.

Linden, S. M. van der, Valkenburg, H. A., Jongh, B. M. de, et al. (1984). "The risk of developing ankylosing spondylitis in HLA-B27 positive individuals. A comparison of relatives of spondylitis patients with the general population." *Arthritis and rheumatism* 27: pp. 241–249.

Ling, G. S., Bennett, J., Woollard, K. J., et al. (2014). "Integrin CD11b positively regulates TLR4-induced signalling pathways in dendritic cells but not in macrophages". *Nature communications* 5: p. 3039.

Lino, C. A., Harper, J. C., Carney, J. P., et al. (2018). "Delivering CRISPR: a review of the challenges and approaches". *Drug delivery* 25: pp. 1234–1257.

Liu, J. Z., Van Sommeren, S., Huang, H., et al. (2015). "Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations". *Nature genetics* 47: p. 979.

Liu, Y., Chang, J.-C., Hon, C.-C., et al. (2018). "Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis". *Scientific reports* 8: p. 15499.

Liu, Z., Hui, Y., Shi, L., et al. (2016). "Efficient CRISPR/Cas9-mediated versatile, predictable, and donor-free gene knockout in human pluripotent stem cells". *Stem cell reports* 7: pp. 496–507.

Lonsdale, J., Thomas, J., Salvatore, M., et al. (2013). "The Genotype-Tissue Expression (GTEx) project". *Nature Genetics* 45: pp. 580–585.

Lopez-Dee, Z., Pidcock, K., and Gutierrez, L. S. (2011). "Thrombospondin-1: multiple paths to inflammation". *Mediators of inflammation* 2011: p. 10.

Love, M. I., Huber, W., and Anders, S. (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". *Genome biology* 15: p. 550.

Lozzio, C. B. and Lozzio, B. B. (1975). "Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome". *Blood* 45: pp. 321–334.

Lubberts, E., Joosten, L. A., Oppers, B., et al. (2001). "IL-1-independent role of IL-17 in synovial inflammation and joint destruction during collagen-induced arthritis". *The Journal of Immunology* 167: pp. 1004–1013.

Luckheeram, R. V., Zhou, R., Verma, A. D., et al. (2012). "CD4+ T cells: differentiation and functions". *Clinical and developmental immunology* 2012: p. 12.

Mack, D. G., Lanham, A. M., Palmer, B. E., et al. (2009). "CD27 expression on CD4+ T cells differentiates effector from regulatory T cell subsets in the lung". *The Journal of Immunology* 182: pp. 7317–7324.

Madson, J. G., Lynch, D. T., Tinkum, K. L., et al. (2006). "Erbb2 regulates inflammation and proliferation in the skin after ultraviolet irradiation". *The American journal of pathology* 169: pp. 1402–1414.

Magoč, T. and Salzberg, S. L. (2011). "FLASH: fast length adjustment of short reads to improve genome assemblies". *Bioinformatics* 27: pp. 2957–2963.

Mahmood, F. and Helliwell, P. (2017). "Ankylosing spondylitis: a review". *European medical journal* 2: pp. 134–139.

Maksymowych, W. P., Landewé, R., Conner-Spady, B., et al. (2007). "Serum matrix metalloproteinase 3 is an independent predictor of structural damage progression in patients with ankylosing spondylitis". *Arthritis & Rheumatism* 56: pp. 1846–1853.

Mans, R., Rossum, H. M. van, Wijsman, M., et al. (2015). "CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae". *FEMS yeast research* 15: p. 4.

Mantovani, A., Sica, A., Sozzani, S., et al. (2004). "The chemokine system in diverse forms of macrophage activation and polarization". *Trends in immunology* 25: pp. 677–686.

Markova, E. N., Kantidze, O. L., and Razin, S. V. (2011). "Transcriptional regulation and spatial organisation of the human AML1/RUNX1 gene". *Journal of cellular biochemistry* 112: pp. 1997–2005.

Martin, C.-A., Murray, J. E., Carroll, P., et al. (2016). "Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis". *Genes & development* 30: pp. 2158–2172.

Martín-Esteban, A., Guasp, P., Barnea, E., et al. (2016). "Functional Interaction of the Ankylosing Spondylitis-Associated Endoplasmic Reticulum Aminopeptidase 2 With the HLA-B*27 Peptidome in Human Cells". *Arthritis & Rheumatology* 68: pp. 2466–2475.

Martín-Mola, E, Sieper, J, Leirisalo-Repo, M, et al. (2010). "Sustained efficacy and safety, including patient-reported outcomes, with etanercept treatment over 5 years in patients with ankylosing spondylitis". *Clinical & Experimental Rheumatology* 28: p. 238.

Martinez, F. O. and Gordon, S. (2014). "The M1 and M2 paradigm of macrophage activation: time for reassessment". *F1000prime reports* 6.

Martinez-Gonzalez, O, Cantero-Hinojosa, J, Paule-Sastre, P, et al. (1994). "Intestinal permeability in patients with ankylosing spondylitis and their healthy relatives". *Rheumatology* 33: pp. 644–647.

Matsuda, M, Sakamoto, N, and Fukumaki, Y (1992). "Delta-thalassemia caused by disruption of the site for an erythroid-specific transcription factor, GATA-1, in the delta-globin gene promoter." *Blood* 80: pp. 1347–51.

Mau, W., Zeidler, H., Mau, R, et al. (1988). "Clinical features and prognosis of patients with possible ankylosing spondylitis. Results of a 10-year followup." *The Journal of Rheumatology* 15: pp. 1109–1114.

Maurano, M. T., Humbert, R., Rynes, E., et al. (2012). "Systematic Localization of Common Disease-Associated Variation in Regulatory DNA". *Science* 337: pp. 1190–1195.

Mauri, C., Gray, D., Mushtaq, N., et al. (2003). "Prevention of arthritis by interleukin 10-producing B cells." *The Journal of experimental medicine* 197: pp. 489–501.

Maxwell, J. R., Zhang, Y., Brown, W. A., et al. (2015). "Differential roles for interleukin-23 and interleukin-17 in intestinal immunoregulation". *Immunity* 43: pp. 739–750.

May, E., Dorris, M. L., Satumtira, N., et al. (2003). "CD8$\alpha\beta$ T cells are not essential to the pathogenesis of arthritis or colitis in HLA-B27 transgenic rats". *The Journal of Immunology* 170: pp. 1099–1105.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., et al. (2008). "Genome-wide association studies for complex traits: consensus, uncertainty and challenges". *Nature reviews genetics* 9: p. 356.

McDermott, M. F., Aksentijevich, I., Galon, J., et al. (1999). "Germline mutations in the extracellular domains of the 55 kDa TNF receptor, TNFR1, define a family of dominantly inherited autoinflammatory syndromes". *Cell* 97: pp. 133–144.

McGovern, D. P. B., Gardet, A., Törkvist, L., et al. (2010). "Genome-wide association identifies multiple ulcerative colitis susceptibility loci". *Nature Genetics* 42: pp. 332–337.

McHugh, K. and Bowness, P. (2012). "The link between HLA-B27 and SpA—new ideas on an old problem". *Rheumatology* 51: pp. 1529–1539.

McInnes, I. B., Kavanaugh, A., Gottlieb, A. B., et al. (2013). "Efficacy and safety of ustekinumab in patients with active psoriatic arthritis: 1 year results of the phase 3, multicentre, double-blind, placebo-controlled PSUMMIT 1 trial". *The Lancet* 382: pp. 780–789.

McVeigh, C. M. and Cairns, A. P. (2006). "Diagnosis and management of ankylosing spondylitis". *Bmj* 333: pp. 581–585.

McVicker, G., Geijn, B. van de, Degner, J. F., et al. (2013). "Identification of genetic variants that affect histone modifications in human cells." *Science* 342: pp. 747–9.

Mear, J. P., Schreiber, K. L., Münz, C., et al. (1999). "Misfolding of HLA-B27 as a result of its B pocket suggests a novel mechanism for its role in susceptibility to spondyloarthropathies". *The Journal of Immunology* 163: pp. 6665–6670.

Mells, G. F., Floyd, J. A., Morley, K. I., et al. (2011). "Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis". *Nature genetics* 43: p. 329.

Menegatti, S., Bianchi, E., and Rogge, L. (2019). "Anti-TNF Therapy in Spondyloarthritis and Related Diseases, Impact on the Immune System and Prediction of Treatment Responses". *Frontiers in immunology* 10: p. 382.

Mercan, R., Bitik, B., Tufan, A., et al. (2016). "The association between neutrophil/lymphocyte ratio and disease activity in rheumatoid arthritis and ankylosing spondylitis". *Journal of clinical laboratory analysis* 30: pp. 597–601.

Messeguer, X., Escudero, R., Farré, D., et al. (2002). "PROMO: detection of known transcription regulatory elements using species-tailored searches". *Bioinformatics* 18: pp. 333–334.

Miller, C. L., Anderson, D. R., Kundu, R. K., et al. (2013). "Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus." *PLoS genetics* 9: e1003652.

Miossec, P. and Kolls, J. K. (2012). "Targeting IL-17 and TH17 cells in chronic inflammation". *Nature Reviews Drug Discovery* 11: pp. 763–776.

Mizuki, N., Meguro, A., Ota, M., et al. (2010). "Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behçet's disease susceptibility loci". *Nature Genetics* 42: pp. 703–706.

Morwood, S. R. and Nicholson, L. B. (2006). "Modulation of the immune response by extracellular matrix proteins". *Archivum immunologiae et therapiae experimentalis* 54: pp. 367–374.

Murakami, N., Imamura, K., Izumi, Y., et al. (2017). "Proteasome impairment in neural cells derived from HMSN-P patient iPSCs". *Molecular brain* 10: p. 7.

Nair, R. P., Duffin, K. C., Helms, C., et al. (2009). "Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways". *Nature Genetics* 41: pp. 199–204.

Nakade, S., Tsubota, T., Sakane, Y., et al. (2014). "Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9". *Nature communications* 5: p. 5560.

Nakamura, N. and Takenaga, K. (1997). "Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines". *Clinical and Experimental Metastasis* 16: pp. 471–479.

Nishimura, T., Kaneko, S., Kawana-Tachikawa, A., et al. (2013). "Generation of rejuvenated antigen-specific T cells by reprogramming to pluripotency and redifferentiation". *Cell stem cell* 12: pp. 114–126.

Nissinen, L. and Kähäri, V.-M. (2014). "Matrix metalloproteinases in inflammation". *BBA-General Subjects* 1840: pp. 2571–2580.

Noordenbos, T., Yeremenko, N., Gofita, I., et al. (2012). "Interleukin-17-positive mast cells contribute to synovial inflammation in spondylarthritis". *Arthritis & Rheumatism* 64: pp. 99–109.

Nora, E. P., Lajoie, B. R., Schulz, E. G., et al. (2012). "Spatial partitioning of the regulatory landscape of the X-inactivation centre". *Nature* 485: pp. 381–385.

O'Hara, R., Murphy, E. P., Whitehead, A. S., et al. (2000). "Acute-phase serum amyloid A production by rheumatoid arthritis synovial tissue". *Arthritis Research & Therapy* 2: p. 142.

Ombrello, M. J., Kastner, D. L., and Remmers, E. F. (2015). "Endoplasmic reticulum-associated amino-peptidase 1 and rheumatic disease: genetics". *Current opinion in rheumatology* 27: p. 349.

Osgood, J. A. and Knight, J. C. (2018). "Translating GWAS in rheumatic disease: approaches to establishing mechanism and function for genetic associations with ankylosing spondylitis". *Briefings in functional genomics* 17: pp. 308–318.

Ousingsawat, J., Wanitchakool, P., Kmit, A., et al. (2015). "Anoctamin 6 mediates effects essential for innate immunity downstream of P2X 7 receptors in macrophages". *Nature communications* 6: p. 6245.

Pahl, H. L. (1999). "Signal transduction from the endoplasmic reticulum to the cell nucleus". *Physiological reviews* 79: pp. 683–701.

Pahl, H. L. and Baeuerle, P. A. (1997). "The ER-overload response: activation of NF-κB". *Trends in biochemical sciences* 22: pp. 63–67.

Palla, I, Trieste, L, Tani, C, et al. (2012). "A systematic literature review of the economic impact of ankylosing spondylitis". *Clinical and Experimental Rheumatology* 30: S136.

Pan, C., Kumar, C., Bohl, S., et al. (2009). "Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions". *Molecular & Cellular Proteomics* 8: pp. 443–450.

Pan, G., Tian, S., Nie, J., et al. (2007). "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells". *Cell stem cell* 1: pp. 299–312.

Papapetrou, E. P. and Schambach, A. (2016). "Gene insertion into genomic safe harbors for human gene therapy". *Molecular Therapy* 24: pp. 678–684.

Paquet, D., Kwart, D., Chen, A., et al. (2016). "Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9". *Nature* 533: p. 125.

Parvaneh, N., Mamishi, S., Rezaei, A., et al. (2010). "Characterization of 11 new cases of leukocyte adhesion deficiency type 1 with seven novel mutations in the ITGB2 gene". *Journal of clinical immunology* 30: pp. 756–760.

Pedersen, O. B., Svendsen, A. J., Ejstrup, L, et al. (2008). "Ankylosing spondylitis in Danish and Norwegian twins: occurrence and the relative importance of genetic vs. environmental effectors in disease causation." *Scandinavian journal of rheumatology* 37: pp. 120–6.

Pelletier, M., Maggi, L., Micheletti, A., et al. (2010). "Evidence for a cross-talk between human neutrophils and Th17 cells". *Blood* 115: pp. 335–343.

Peng, R., Lin, G., and Li, J. (2016). "Potential pitfalls of CRISPR/Cas9-mediated genome editing". *The FEBS journal* 283: pp. 1218–1231.

Peschon, J. J., Torrance, D. S., Stocking, K. L., et al. (1998). "TNF receptor-deficient mice reveal divergent roles for p55 and p75 in several models of inflammation". *The Journal of Immunology* 160: pp. 943–952.

Petrif, F., Giles, R., H., et al. (1995). "Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP". *Nature* 376: pp. 348–351.

Pfeffer, K., Matsuyama, T., Kündig, T. M., et al. (1993). "Mice deficient for the 55 kd tumor necrosis factor receptor are resistant to endotoxic shock, yet succumb to L. monocytogenes infection". *Cell* 73: pp. 457–467.

Phillips-Cremins, J., Sauria, M., Sanyal, A., et al. (2013). "Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment". *Cell* 153: pp. 1281–1295.

Poddubnyy, D., Conrad, K., Haibel, H., et al. (2014a). "Elevated serum level of the vascular endothelial growth factor predicts radiographic spinal progression in patients with axial spondyloarthritis". *Annals of the rheumatic diseases* 73: pp. 2137–2143.

Poddubnyy, D., Hermann, K.-G. A., Callhoff, J., et al. (2014b). "Ustekinumab for the treatment of patients with active ankylosing spondylitis: results of a 28-week, prospective, open-label, proof-of-concept study (TOPAS)". *Annals of the rheumatic diseases* 73: pp. 817–823.

Pointon, J. J., Harvey, D., Karaderi, T., et al. (2010). "The chromosome 16q region associated with ankylosing spondylitis includes the candidate gene tumour necrosis factor receptor type 1-associated death domain (TRADD)". *Annals of the Rheumatic Diseases* 69: pp. 1243–1246.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., et al. (2010). "Detection of nonneutral substitution rates on mammalian phylogenies". *Genome research* 20: pp. 110–121.

Probert, L., Eugster, H.-P., Akassoglou, K., et al. (2000). "TNFR1 signalling is critical for the development of demyelination and the limitation of T-cell responses during immune-mediated CNS disease". *Brain* 123: pp. 2005–2019.

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

Raj, T., Rothamel, K., Mostafavi, S., et al. (2014). "Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes." *Science* 344: pp. 519–23.

Ramagopalan, S. V., Knight, M., Ebers, G. C., et al. (2007). "Origins of magic: review of genetic and epigenetic effects". *Bmj* 335: pp. 1299–1301.

Ramakrishnan, V. and Pace, B. S. (2011). "Regulation of γ-globin gene expression involves signaling through the p38 MAPK/CREB1 pathway". *Blood Cells, Molecules, and Diseases* 47: pp. 12–22.

Rao, S. S., Huntley, M. H., Durand, N. C., et al. (2014). "A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping". *Cell* 159: pp. 1665–1680.

Ray, A., Dhar, S., and Ray, B. K. (2011). "Control of VEGF expression in triple-negative breast carcinoma cells by suppression of SAF-1 transcription factor activity". *Molecular Cancer Research* 9: pp. 1030–1041.

Raychaudhuri, S. P. and Deodhar, A. (2014). "The classification and diagnostic criteria of ankylosing spondylitis". *Journal of Autoimmunity* 48-49: pp. 128–133.

Reed, M., Dharmage, S, Boers, A, et al. (2008). "Ankylosing spondylitis: an Australian experience". *Internal medicine journal* 38: pp. 321–327.

Reveille, J. D., Sims, A.-M., Danoy, P., et al. (2010). "Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci." *Nature genetics* 42: pp. 123–7.

Revez, J. A., Bain, L., Chapman, B., et al. (2013). "A new regulatory variant in the interleukin-6 receptor gene associates with asthma risk". *Genes and immunity* 14: p. 441.

Ridley, A., Hatano, H., Wong-Baeza, I., et al. (2016). "Activation-Induced Killer Cell Immunoglobulin-like Receptor 3DL2 Binding to HLA–B27 Licenses Pathogenic T Cell Differentiation in Spondyloarthritis". *Arthritis & Rheumatology* 68: pp. 901–914.

Rivera, C. and Ren, B. (2013). "Mapping Human Epigenomes". *Cell* 155: pp. 39–55.

Robertson, J., Tindell, A., Crosbie, D., et al. (2018). "190 Real-world experience of secukinumab for psoriatic arthritis and axial spondyloarthritis". *Rheumatology* 57: key075–414.

Robinson, P. C. and Brown, M. A. (2014). "Genetics of ankylosing spondylitis". *Molecular immunology* 57: pp. 2–11.

Rockman, M. V. and Kruglyak, L. (2006). "Genetics of global gene expression". *Nature Reviews Genetics* 7: pp. 862–872.

Romero-Sanchez, C, Jaimes, D., Londoño, J, et al. (2011). "Association between Th-17 cytokine profile and clinical features in patients with spondyloarthritis". *Clinical and Experimental Rheumatology* 29: p. 828.

Rothbart, S. B. and Strahl, B. D. (2014). "Interpreting the language of histone and DNA modifications". *BBA-Gene Regulatory Mechanisms* 1839: pp. 627–643.

Roux-Lombard, P., Punzi, L., Hasler, F., et al. (1993). "Soluble tumor necrosis factor receptors in human inflammatory synovial fluids". *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology* 36: pp. 485–489.

Ruan, W.-F., Xie, J.-T., Jin, Q., et al. (2018). "The diagnostic and prognostic role of interleukin 12B and Interleukin 6R gene polymorphism in patients with ankylosing spondylitis". *Journal of Clinical Rheumatology* 24: pp. 18–24.

Rudwaleit, M, Van Der Heijde, D, Landewé, R, et al. (2009). "The development of Assessment of SpondyloArthritis international Society classification criteria for axial spondyloarthritis (part II): validation and final selection". *Annals of the rheumatic diseases* 68: pp. 777–783.

Rusman, T, Vollenhoven, R. van, and Horst-Bruinsma, I. van der (2018). "Gender differences in axial spondyloarthritis: women are not so lucky". *Current rheumatology reports* 20: p. 35.

Ryall, N. H. and Helliwell, P (1998). "A critical review of ankylosing spondylitis". *Critical Reviews in Physical and Rehabilitation Medicine* 10: pp. 265–301.

Safa, A. (2012). "c-FLIP, a master anti-apoptotic regulator". *Experimental oncology* 34: p. 176.

Salomon, M. P., Li, W. L. S., Edlund, C. K., et al. (2016). "GWASeq: targeted re-sequencing follow up to GWAS." *BMC genomics* 17: p. 176.

Sanjana, N. E., Wright, J., Zheng, K., et al. (2016). "High-resolution interrogation of functional elements in the noncoding genome". *Science* 353: pp. 1545–1549.

Sanz-Bravo, A., Alvarez-Navarro, C., Martín-Esteban, A., et al. (2018). "Ranking the Contribution of Ankylosing Spondylitis-associated Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) Polymorphisms to Shaping the HLA-B* 27 Peptidome". *Molecular & Cellular Proteomics* 17: pp. 1308–1323.

Saoudi, A., Kassem, S., Dejean, A. S., et al. (2014). "Rho-GTPases as key regulators of T lymphocyte biology". *Small GTPases* 5: e983862.

Saric, T., Chang, S.-C., Hattori, A., et al. (2002). "An IFN-γ–induced aminopeptidase in the ER, ERAP1, trims precursors to MHC class I–presented peptides". *Nature immunology* 3: p. 1169.

Sarin, R., Wu, X., and Abraham, C. (2011). "Inflammatory disease protective R381Q IL23 receptor polymorphism results in decreased primary CD4+ and CD8+ human T-cell functional responses". *Proceedings of the National Academy of Sciences* 108: pp. 9560–9565.

Sato, K., Suematsu, A., Okamoto, K., et al. (2006). "Th17 functions as an osteoclastogenic helper T cell subset that links T cell activation and bone destruction". *Journal of Experimental Medicine* 203: pp. 2673–2682.

Sato, M., Koriyama, M., Watanabe, S., et al. (2015). "Direct injection of CRISPR/Cas9-related mRNA into cytoplasm of parthenogenetically activated porcine oocytes causes frequent mosaicism for indel mutations". *International journal of molecular sciences* 16: pp. 17838–17856.

Savill, J., Hogg, N., Ren, Y., et al. (1992). "Thrombospondin cooperates with CD36 and the vitronectin receptor in macrophage recognition of neutrophils undergoing apoptosis." *The Journal of clinical investigation* 90: pp. 1513–1522.

Sawcer, S., Hellenthal, G., Pirinen, M., et al. (2011). "Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis". *Nature* 476: p. 214.

Scharer, C. D., Blalock, E. L., Mi, T., et al. (2019). "Epigenetic programming underpins B cell dysfunction in human SLE". *Nature immunology* 20: pp. 1071–1082.

Schaub, M. A., Boyle, A. P., Kundaje, A., et al. (2012). "Linking disease associations with regulatory information in the human genome." *Genome research* 22: pp. 1748–59.

Schena, M., Shalon, D., Davis, R. W., et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray". *Science* 270: pp. 467–470.

Schiff, D. E., Kline, L., Soldau, K., et al. (1997). "Phagocytosis of gram-negative bacteria by a unique CD14-dependent mechanism". *Journal of leukocyte biology* 62: pp. 786–794.

Schittenhelm, R. B., Sian, T. C. L. K., Wilmann, P. G., et al. (2015). "Revisiting the arthritogenic peptide theory: quantitative not qualitative changes in the peptide repertoire of HLA–B27 allotypes". *Arthritis & rheumatology* 67: pp. 702–713.

Schmidl, C., Rendeiro, A. F., Sheffield, N. C., et al. (2015). "ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors". *Nature methods* 12: p. 963.

Schmidt, D., Wilson, M. D., Spyrou, C., et al. (2009). "ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions". *Methods* 48: pp. 240–248.

Schmidt, S., Moser, M., and Sperandio, M. (2013). "The molecular basis of leukocyte recruitment and its deficiencies". *Molecular immunology* 55: pp. 49–58.

Schmitt, A. D., Hu, M., Jung, I., et al. (2016). "A compendium of chromatin contact maps reveals spatially active regions in the human genome". *Cell reports* 17: pp. 2042–2059.

Schneider-Brachert, W., Tchikov, V., Neumeyer, J., et al. (2004). "Compartmentalization of TNF receptor 1 signaling: internalized TNF receptosomes as death signaling vesicles". *Immunity* 21: pp. 415–428.

Schoenborn, J. R. and Wilson, C. B. (2007). "Regulation of interferon-γ during innate and adaptive immune responses". *Advances in immunology* 96: pp. 41–101.

Schofield, E., Carver, T., Achuthan, P., et al. (2016). "CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets". *Bioinformatics* 32: pp. 2511–2513.

Schumann, K., Lin, S., Boyer, E., et al. (2015). "Generation of knock-in primary human T cells using Cas9 ribonucleoproteins". *Proceedings of the National Academy of Sciences* 112: pp. 10437–10442.

Schütze, S., Tchikov, V., and Schneider-Brachert, W. (2008). "Regulation of TNFR1 and CD95 signalling by receptor compartmentalization". *Nature reviews Molecular cell biology* 9: p. 655.

Serwold, T., Gaw, S., and Shastri, N. (2001). "ER aminopeptidases generate a unique pool of peptides for MHC class I molecules". *Nature immunology* 2: p. 644.

Serwold, T., Gonzalez, F., Kim, J., et al. (2002). "ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum". *Nature* 419: p. 480.

Shalem, O., Sanjana, N. E., Hartenian, E., et al. (2014). "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells". *Science* 343: pp. 84–87.

Sherlock, J. P., Joyce-Shaikh, B., Turner, S. P., et al. (2012). "IL-23 induces spondyloarthropathy by acting on ROR-γt+ CD3+CD4-CD8- entheseal resident T cells". *Nature Medicine* 18: pp. 1069–1076.

Sieper, J., Listing, J., Poddubnyy, D., et al. (2016). "Effect of continuous versus on-demand treatment of ankylosing spondylitis with diclofenac over 2 years on radiographic progression of the spine: results from a randomised multicentre trial (ENRADAS)". *Annals of the rheumatic diseases* 75: pp. 1438–1443.

Silva Correia, J. da, Soldau, K., Christen, U., et al. (2001). "Lipopolysaccharide is in close proximity to each of the proteins in its membrane receptor complex transfer from CD14 to TLR4 and MD-2". *Journal of Biological Chemistry* 276: pp. 21129–21135.

Simeonov, D. R., Gowen, B. G., Boontanrart, M., et al. (2017). "Discovery of stimulation-responsive immune enhancers with CRISPR activation". *Nature* 549: pp. 111–115.

Smemo, S., Tena, J. J., Kim, K.-H., et al. (2014). "Obesity-associated variants within FTO form long-range functional connections with IRX3". *Nature* 507: pp. 371–375.

Smith, E. N., D'Antonio-Chronowska, A., Greenwald, W. W., et al. (2019). "Human iPSC-Derived Retinal Pigment Epithelium: A Model System for Prioritizing and Functionally Characterizing Causal Variants at AMD Risk Loci". *Stem cell reports* 12: pp. 1342–1353.

Smith, J. A. and Colbert, R. A. (2014). "The interleukin-23/interleukin-17 axis in spondyloarthritis pathogenesis: Th17 and beyond". *Arthritis & rheumatology* 66: pp. 231–241.

Smith, J. A., Märker-Hermann, E., and Colbert, R. A. (2006). "Pathogenesis of ankylosing spondylitis: Current concepts". *Best Practice & Research Clinical Rheumatology* 20: pp. 571–591.

Soldner, F., Stelzer, Y., Shivalila, C. S., et al. (2016). "Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression". *Nature* 533: pp. 95–99.

Spain, S. L. and Barrett, J. C. (2015). "Strategies for fine-mapping complex traits." *Human molecular genetics* 24: R111–9.

Spoorenberg, A, Dougados, M, Mielants, H, et al. (1999). "Relative value of erythrocyte sedimentation rate and C-reactive protein in assessment of disease activity in ankylosing spondylitis." *The Journal of rheumatology* 26: pp. 980–984.

Spreafico, M., Grillo, B., Rusconi, F., et al. (2018). "Multiple layers of CDK5R1 regulation in Alzheimer's disease implicate long non-coding RNAs". *International journal of molecular sciences* 19: p. 2022.

Stein, E. V., Miller, T. W., Ivins-O'Keefe, K., et al. (2016). "Secreted thrombospondin-1 regulates macrophage interleukin-1β production and activation through CD47". *Scientific reports* 6: p. 19684.

Stuart, P., Nair, R., Tsoi, L., et al. (2015). "Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture". *The American Journal of Human Genetics* 97: pp. 816–836.

Subramanian, A., Tamayo, P., Mootha, V. K., et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". *Proceedings of the National Academy of Sciences* 102: pp. 15545–15550.

Sun, L., Youn, H.-D., Loh, C., et al. (1998). "Cabin 1, a negative regulator for calcineurin signaling in T lymphocytes". *Immunity* 8: pp. 703–711.

Sveaas, S., Berg, I., Provan, S., et al. (2015). "Circulating levels of inflammatory cytokines and cytokine receptors in patients with ankylosing spondylitis: a cross-sectional comparative study". *Scandinavian journal of rheumatology* 44: pp. 118–124.

Syrbe, U., Callhoff, J., Conrad, K., et al. (2015). "Serum adipokine levels in patients with ankylosing spondylitis and their relationship to clinical parameters and radiographic spinal progression". *Arthritis & Rheumatology* 67: pp. 678–685.

Takahashi, K. and Yamanaka, S. (2006). "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors". *cell* 126: pp. 663–676.

Tan, S., Wang, R., and Ward, M. M. (2015). "Syndesmophyte growth in ankylosing spondylitis". *Current opinion in rheumatology* 27: p. 326.

Tau, G and Rothman, P (1999). "Biologic functions of the IFN-γ receptors". *Allergy* 54: p. 1233.

Taurog, J. D., Richardson, J. A., Croft, J., et al. (1994). "The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats." *Journal of Experimental Medicine* 180: pp. 2359–2364.

Taurog, J. D., Chhabra, A., and Colbert, R. A. (2016). "Ankylosing spondylitis and axial spondyloarthritis". *New England Journal of Medicine* 374: pp. 2563–2574.

Tazi, J., Bakkour, N., and Stamm, S. (2009). "Alternative splicing and disease". *BBA - Molecular Basis of Disease* 1792: pp. 14–26.

Timmermans, F., Velghe, I., Vanwalleghem, L., et al. (2009). "Generation of T cells from human embryonic stem cell-derived hematopoietic zones". *The Journal of Immunology* 182: pp. 6879–6888.

Ting, P. Y., Parker, A. E., Lee, J. S., et al. (2018). "Guide Swap enables genome-scale pooled CRISPR–Cas9 screening in human primary cells". *Nature methods* 15: p. 941.

Tomlin, H. and Piccinini, A. M. (2018). "A complex interplay between the extracellular matrix and the innate immune response to microbial pathogens". *Immunology* 155: pp. 186–201.

Tracey, D., Klareskog, L., Sasso, E. H., et al. (2008). "Tumor necrosis factor antagonist mechanisms of action: a comprehensive review". *Pharmacology & therapeutics* 117: pp. 244–279.

Triantafilou, M. and Triantafilou, K. (2002). "Lipopolysaccharide recognition: CD14, TLRs and the LPS-activation cluster". *Trends in immunology* 23: pp. 301–304.

Tsompana, M. and Buck, M. J. (2014). "Chromatin accessibility: a window into the genome". *Epigenetics & Chromatin* 7: p. 33.

Tsui, H. W., Inman, R. D., Paterson, A. D., et al. (2005). "ANKH variants associated with ankylosing spondylitis: gender differences". *Arthritis research & therapy* 7: R513.

Tu, Z., Yang, W., Yan, S., et al. (2017). "Promoting Cas9 degradation reduces mosaic mutations in non-human primate embryos". *Scientific reports* 7: p. 42081.

Ueda, H., Howson, J. M. M., Esposito, L., et al. (2003). "Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease". *Nature* 423: pp. 506–511.

Vallon, R., Freuler, F., Desta-Tsedu, N., et al. (2001). "Serum amyloid A (apoSAA) expression is up-regulated in rheumatoid arthritis and induces transcription of matrix metalloproteinases". *The Journal of Immunology* 166: pp. 2801–2807.

Van Der Heijde, D., Ramiro, S., Landewé, R., et al. (2017a). "2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis". *Annals of the rheumatic diseases* 76: pp. 978–991.

Van Der Heijde, D., Deodhar, A., Wei, J. C., et al. (2017b). "Tofacitinib in patients with ankylosing spondylitis: a phase II, 16-week, randomised, placebo-controlled, dose-ranging study". *Annals of the rheumatic diseases* 76: pp. 1340–1347.

Van Der Linden, S., Valkenburg, H. A., and Cats, A. (1984). "Evaluation of diagnostic criteria for ankylosing spondylitis". *Arthritis & Rheumatism* 27: pp. 361–368.

Van Oosten, B., Barkhof, F, Truyen, L, et al. (1996). "Increased MRI activity and immune activation in two multiple sclerosis patients treated with the monoclonal anti-tumor necrosis factor antibody cA2". *Neurology* 47: pp. 1531–1534.

Visscher, P. M. and Hill, W. G. (2009). "The limits of individual identification from sample allele frequencies: theory and statistical analysis". *PLoS genetics* 5: e1000628.

Visscher, P. M., Wray, N. R., Zhang, Q., et al. (2017). "10 years of GWAS discovery: biology, function, and translation". *The American Journal of Human Genetics* 101: pp. 5–22.

Wallis, D., Haroon, N., Ayearst, R., et al. (2013). "Ankylosing spondylitis and nonradiographic axial spondyloarthritis: part of a common spectrum or distinct diseases?" *The Journal of rheumatology* 40: pp. 2038–2041.

Wanders, A., Heijde, D. v. d., Landewé, R., et al. (2005). "Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial". *Arthritis & Rheumatism* 52: pp. 1756–1765.

Wang, D. (2018). "The essential role of G protein-coupled receptor (GPCR) signaling in regulating T cell immunity". *Immunopharmacology and immunotoxicology* 40: pp. 187–192.

Wang, J., Zibetti, C., Shang, P., et al. (2018a). "ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration". *Nature communications* 9: p. 1364.

Wang, T., Wei, J. J., Sabatini, D. M., et al. (2014). "Genetic screens in human cells using the CRISPR/Cas9 system". *Science* 343: pp. 80–84.

Wang, T., Birsoy, K., Hughes, N. W., et al. (2015). "Identification and characterization of essential genes in the human genome". *Science* 350: pp. 1096–1101.

Wang, X., Hayes, J., Xu, D., et al. (2018b). "Validation of prostate cancer risk variants by CRISPR/Cas9 mediated genome editing". *BioRxiv*: p. 337022.

Wang, Z., Gerstein, M., and Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics". *Nature reviews genetics* 10: p. 57.

Ward, L. D. and Kellis, M. (2012). "Interpreting noncoding genetic variation in complex traits and human disease". *Nature Biotechnology* 30: pp. 1095–1106.

Weintraub, H and Groudine, M (1976). "Chromosomal subunits in active genes have an altered conformation". *Science* 193: pp. 848–856.

Werken, H. J. G. van de, Landan, G., Holwerda, S. J. B., et al. (2012). "Robust 4C-seq data analysis to screen for regulatory DNA interactions". *Nature Methods* 9: pp. 969–972.

Westra, H.-J. and Franke, L. (2014). "From genome to function by studying eQTLs". *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* 1842: pp. 1896–1902.

Westra, H.-J., Peters, M. J., Esko, T., et al. (2013). "Systematic identification of trans eQTLs as putative drivers of known disease associations". *Nature genetics* 45: p. 1238.

Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012). "RNA-guided genetic silencing systems in bacteria and archaea". *Nature* 482: p. 331.

Wilgenburg, B. van, Browne, C., Vowles, J., et al. (2013). "Efficient, Long Term Production of Monocyte-Derived Macrophages from Human Pluripotent Stem Cells under Partly-Defined and Fully-Defined Conditions". *PLoS ONE* 8. Ed. by D. T. Covas: e71098.

Wright, C, Edelmann, M, DiGleria, K, et al. (2009). "Ankylosing spondylitis monocytes show upregulation of proteins involved in inflammation and the ubiquitin proteasome pathway." *Annals of the rheumatic diseases* 68: pp. 1626–32.

Wright, S. D., Ramos, R. A., Tobias, P. S., et al. (1990). "CD14, a receptor for complexes of lipopolysaccharide (LPS) and LPS binding protein". *Science* 249: pp. 1431–1433.

Wroblewska, A., Dhainaut, M., Ben-Zvi, B., et al. (2018). "Protein barcodes enable high-dimensional single-cell CRISPR screens". *Cell* 175: pp. 1141–1155.

Wu, Y., Zheng, Z., Visscher, P. M., et al. (2017). "Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data". *Genome biology* 18: p. 86.

Wyman, C. and Kanaar, R. (2006). "DNA double-strand break repair: all's well that ends well". *Annual Review of Genetics* 40: pp. 363–383.

Xanthoulea, S., Pasparakis, M., Kousteni, S., et al. (2004). "Tumor necrosis factor (TNF) receptor shedding controls thresholds of innate immune activation that balance opposing TNF functions in infectious and inflammatory diseases". *Journal of Experimental Medicine* 200: pp. 367–376.

Xu, X., Tay, Y., Sim, B., et al. (2017). "Reversal of phenotypic abnormalities by CRISPR/Cas9-mediated gene correction in Huntington disease patient-derived induced pluripotent stem cells". *Stem cell reports* 8: pp. 619–633.

Xueyi, L., Lina, C., Zhenbiao, W., et al. (2013). "Levels of Circulating Th17 Cells and Regulatory T Cells in Ankylosing Spondylitis Patients with an Inadequate Response to Anti-TNF-α Therapy". *Journal of Clinical Immunology* 33: pp. 151–161.

Yago, T., Nanke, Y., Kawamoto, M., et al. (2017). "IL-23 and Th17 disease in inflammatory arthritis". *Journal of clinical medicine* 6: p. 81.

Yang, J., Rajan, S. S., Friedrich, M. J., et al. (2019). "Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming". *Stem cell reports* 12: pp. 757–771.

Ye, R. D. and Sun, L. (2015). "Emerging functions of serum amyloid A in inflammation". *Journal of leukocyte biology* 98: pp. 923–929.

Yen, D., Cheung, J., Scheerens, H., et al. (2006). "IL-23 is essential for T cell–mediated colitis and promotes inflammation via IL-17 and IL-6". *The Journal of clinical investigation* 116: pp. 1310–1316.

Yona, S., Kim, K.-W., Wolf, Y., et al. (2013). "Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis". *Immunity* 38: pp. 79–91.

Yoo, S.-A., Park, B.-H., Park, G.-S., et al. (2006). "Calcineurin is expressed and plays a critical role in inflammatory arthritis". *The Journal of Immunology* 177: pp. 2681–2690.

Yu, Z.-H., Lun, S.-M., He, R., et al. (2017). "Dual function of MAZ mediated by FOXF2 in basal-like breast cancer: Promotion of proliferation and suppression of progression". *Cancer letters* 402: pp. 142–152.

Zambrano-Zaragoza, J. F., Agraz-Cibrian, J. M., González-Reyes, C., et al. (2013). "Ankylosing spondylitis: From cells to genes". *International Journal of Inflammation* 2013: p. 16.

Zanoni, I., Ostuni, R., Marek, L. R., et al. (2011). "CD14 controls the LPS-induced endocytosis of Toll-like receptor 4". *Cell* 147: pp. 868–880.

Zeng, S., Yang, Y., Tan, Y., et al. (2012). "ERBB2-induced inflammation in lung carcinogenesis". *Molecular biology reports* 39: pp. 7911–7917.

Zhang, H., Xue, C., Shah, R., et al. (2015). "Functional analysis and transcriptomic profiling of iPSC-derived macrophages and their application in modeling Mendelian disease". *Circulation research* 117: pp. 17–28.

Zhang, Y., Liu, F., Zhang, C., et al. (2019). "Non-SMC condensin I complex subunit D2 is a prognostic factor in triple-negative breast cancer for the ability to promote cell cycle and enhance invasion". *The American journal of pathology* 190: pp. 37–47.

Zhang, Y., An, L., Xu, J., et al. (2018). "Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus". *Nature communications* 9: p. 750.

Zhang, Y., Shi, S., Ciurli, C., et al. (2002). "Animal models of ankylosing spondylitis". *Current rheumatology reports* 4: pp. 507–512.

Zhang, Y., Liu, T., Meyer, C. A., et al. (2008). "Model-based analysis of ChIP-Seq (MACS)". *Genome biology* 9: R137.

Zhao, Y., Xiong, Z., Lechner, E. J., et al. (2014). "Thrombospondin-1 triggers macrophage IL-10 production and promotes resolution of experimental lung injury". *Mucosal immunology* 7: p. 440.

Zhou, X. and Wang, T. (2012). "Using the Wash U Epigenome Browser to examine genome-wide sequencing data". *Current protocols in bioinformatics* 40: pp. 10–10.

Zinovieva, E., Bourgain, C., Kadi, A., et al. (2009). "Comprehensive Linkage and Association Analyses Identify Haplotype, Near to the TNFSF15 Gene, Significantly Associated with Spondyloarthritis". *PLoS Genetics* 5: e1000528.