Exploring the Role of Helios in Systemic Lupus Erythematosus Regulatory Mechanisms in T Cells

Suriguga, Suri

Awarding institution:
King's College London

Exploring the Role of Helios in Systemic Lupus Erythematosus: Regulatory Mechanisms in T Cells

XXX Suriguga

Department of Medical & Molecular Genetics
King's College London

This thesis is submitted for the degree of
Doctor of Philosophy (PhD)

Table of Contents

List of Tables 7

List of Figures 8

Abstract 11

Publications 15

Acknowledgements 16

COVID-19 Impact Statement 18

Contribution Statement 18

Chapter 1. Introduction 19

1.1 Overview of Systemic Lupus Erythematosus (SLE) 19

1.1.1 Epidemiology 19

1.1.2 Aetiology and Risk Factors 20

1.1.3 Pathophysiology 24

1.1.4 Clinical aspects: diagnosis and treatment of SLE 31

1.2 Regulation of Transcription and Gene Expression in Eukaryotes 32

1.2.1 Overview of transcription factors 33

1.2.2 Methodologies for Transcription Factor Studies 35

1.3 Overview of *IKZF2* 41

1.3.1 The Ikaros family of transcription factors 41

1.3.2 Current Understanding of the function of Helios in human cells 43

1.3.2 Current understanding of the Role of Helios in SLE 45

1.4 Thesis aims 47

Chapter 2. Chip-seq binding sites of Helios in Jurkat T cells 48

2.1 Introduction 48

2.2 Methods 51

2.2.1 Helios ChIP-Seq data in Jurkat T cells 51

2.2.2 Antibody suitability assessment 54

2.2.2 Nearest gene annotation of the binding sites of Helios in Jurkat T cell 56

2.2.3 Chromatin States annotation of the Helios binding sites in Jurkat T cells..... 57

2.2.4 Enrichment of Helios binding sites in disease-associated genetic variations 58

2.2.5 Overlap of Helios binding sites with the putative causal variants of SLE risk loci in Jurkat T cells..... 59

2.3 Results 59

2.3.1 Validation of the Helios antibody used for Chip-seq..... 59

2.3.2 Genome-wide discovery of Helios binding sites on Jurkat T cell line 60

2.3.2 ChromHMM states of the binding sites..... 62

2.3.3 Enrichment of Helios binding sites in autoimmune diseases associated genetic variations 65

2.3.4 Overlap of Helios binding sites with the putative causal variants of SLE risk loci in Jurkat T cells..... 69

2.4 Discussion 70

Chapter 3. Differential Expression Induced by Helios Knockdown in Jurkat T Cells 72

3.1 Introduction 72

3.2 Methods 74

3.2.1 Data access of differentially expressed genes between Helios knockdown vs.WT-Helios jurkat T cell 74

3.2.2 Differential expression analysis between Helios wild-type and Helios knockdown in Jurkat T cells 75

3.2.3 Functional annotation of the DEGs between Helios knockdown and wild type in Jurkat T cells..... 75

3.2.4 Network analyis of the DEGs between wild-type and knockdown of Helios in Jurkat T cells 76

3.3 Results 77

3.3.1 DEGs between Helios wild-type and Helios knockdown in Jurkat T cells 77

3.3.2 Change of expression of the genes *ELF1* and *IKZF1* between Helios knockdown and Helios wild type in Jurkat T cells..... 78

3.3.3 Functional annotation of the DEGs 79

3.3.4 PPI network analysis and modules identification for the DEGs 80

3.4 Discussion 102

Chapter 4. prioritizing the target genes of Helios in Jurkat T cells 103

4.1 Introduction 103

4.2 Methods 106

4.2.1 Annotating chip-seq binding sites of Helios in Jurkat T cells using FANTOM5 promoter annotation data in Jurkat T cells and Enhancers-genes interaction data in Jurkat T cells from EnhancerAtlas 2.0 106

4.2.3 Integration of chip-seq data and Microarray data 107

4.2.4 Identifying hub genes of prioritized target genes of Helios in Jurkat T cells 107

4.3 Results 107

4.3.1 The target genes of Helios in Jurkat T cells annotated from binding sites 108

4.3.2 Prioritized target genes of Helios in Jurkat T cells 109

4.3.3 Overlapped pathways enriched in Helios target genes annotatated from binding sites and DEGs caused by Helios knockdown in Jurkat T cells 112

4.3.4 PPI network and hub genes of prioritized target genes of Helios in Jurkat T cells 116

4.4 Discussion 117

Chapter 5. Trans-cell type Helios binding sites prediction 120

5.1 Introduction 120

5.2 Methods 123

5.2.1 DNase I hypersensitivity sites hotspots in GM12878 cells 123

5.2.2 Motif-based Prediction of Trans-cell type Binding of Helios in GM12878 Cells 124

5.2.3 Predicting Trans-cell type Bindings Through Shared Genomic Regions with Known Helios Binding Regions 124

5.2.4 Validation of the efficiency of the different methods 125

5.2.5 Predicting binding sites and target genes of Helios in various T cell subtypes 126

5.2.5.1 predicting binding sites and target genes of Helios in CD4+ naïve T cells 127

5.2.5.2 Predicting binding sites and target genes of Helios in Th1 cells 127

5.2.5.3 predicting binding sites and target genes of Helios in Th17 cells 128

5.3 Results 129

5.3.1 Validation of two trans cell type binding sites prediction methods 129
5.3.2 Predicted binding sites and target genes of Helios in CD4+ naïve T cells 131
5.3.3 Predicted binding sites and target genes of Helios in Th1 cells 133
5.3.4 Predicted binding sites and target genes of Helios in Th17 cells 134
5.3.5 Shared predicted target genes of Helios and the pathways over different cell types 136
5.4 Discussion 138
Chapter 6. Regulatory effects of *IKZF2* in different cell types in SLE patients 140
6.1 Introduction 140
6.2 Methods 142
6.2.1 IMMUNEUXT dataset introduction 142
6.2.2 Analysis of Differentially Expressed Genes (DEGs) in Various T Cell Subtypes between Healthy Controls and SLE Patients 143
6.2.3 Gene set enrichment analysis of target genes of Helios in DEGs between controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells 143
6.2.4 PPI network and functional annotation of core enrichment genes in Th1 cells 144
6.2.5 PPI network and functional annotation of core enrichment genes in Th17 cells 144
6.2.6 Correlation of expression of *IKZF2* with the target genes of Helios in Th1 cells cells 145
6.2.7 Correlation of expression of *IKZF2* with predicted target genes of Helios in Th17 cells cells 146
6.3 Results 146
6.3.1 Change of Expression of *IKZF2* in different T cell subtypes in SLE compared to Healthy controls 146
6.3.2 Change of Expression of Helios's target genes in different T cell subtypes in SLE compared to Healthy controls 148
6.3.3 Gene set enrichment analysis of predicted target genes of Helios in DEGs between Healthy controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells 149
6.3.4 Functional annotation and Hub genes identification of core enrichment genes in Th1 cells 151
6.3.5 Hub genes identification of core enrichment genes in Th17 cells 155
6.3.6 Correlation between *IKZF2* and Helios' core predicted target genes related to SLE in Th1 and Th17 cells 157

6.3.7 Shared Th1 and Th17 cells core enriched target genes of Helios in DEGs between controls and SLE 160

6.3.8 Expression Changes and Correlation with *IKZF2* of Recognized Key Predicted Target Genes of Helios related to SLE 161

6.4 Discussion 162

7. Conclusion 167

8. Reference 169

List of Tables

Table 1. 1 Association of *IKZF2* Gene Variants with Systemic Lupus Erythematosus Risk as Identified in Genome-Wide Association Studies [9]. 47

Table 2. 1 Sequencing Metrics for Helios ChIP-Seq in Jurkat Cells 53

Table 2. 2 Antibodies used for Immunoprecipitations followed by Western Blot 56

Table 2. 3 18-expanded ChromHMM states from Roadmap Epigenomic Program [210] (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html) 57

Table 4. 1 Forty-three shared target genes both annotated from promoter and enhancer data ... 110

Table 4. 2 Fifty-six prioritized target genes of Helios in Jurkat T cells 110

Table 4. 3 Eighty-three pathways enriched in both Helios binding targets and DEGs followed by Helios knockdown 112

Table 5. 1 Thirty-nine shared predicted target genes across Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells. 137

Table 5. 2 Proportion of unique predicted target genes in different cell types 137

Table 5. 3 Shared pathways enriched in predicted target genes over four cell types 138

Table 6. 1 Expression Changes and Correlation with *IKZF2* of Recognized Key Target Genes of Helios in Th1 and Th17 Cells 161

List of Figures

Figure 1. 1 The analogy between Mendelian randomization and randomized controlled trial..... 24

Figure 1. 2 Five different levels of regulation of gene expression in eukaryotes..... 33

Figure 1. 3 Transcription factors bind to promoters/enhancers to initiate transcription 35

Figure 1. 4 Chromatin immunoprecipitation assays with sequencing (ChIP-Seq) workflow to identifying binding sites of a transcription factor. 39

Figure 2. 1 Alignment of Helios Protein Isoforms and Antibody Targeting Sequence..... 54

Figure 2. 2 Assessment of Helios antibody suitability 59

Figure 2. 3 Genome-wide characterization of Helios Binding Sites in Jurkat T cells 61

Figure 2. 4 Feature distribution of the expanded ChromHMM 18 states annotated for the binding sites of Helios in Jurkat T cells 63

Figure 2. 5 Enrichment of Helios binding sites for the GWAS risk variants of Autoimmune Diseases 66

Figure 3. 1 Volcano plot illustrating the differential expression between wild-type and Helios knockdown in Jurkat T cells 77

Figure 3. 2 Functional annotation of the DEGs between wild-type and Helios-knockdown Jurkat T cells..... 79

Figure 3. 3 Gene Interaction Network and Pathway Enrichment in Module 1 of DEGs between wild type and Helios knockdown in Jurkat T cells 81

Figure 3. 4 Gene Interaction Network and Pathway Enrichment in Module 2 of DEGs between wild type and Helios knockdown in Jurkat T cells 83

Figure 3. 5 Gene Interaction Network and Pathway Enrichment in Module 3 of DEGs between wild type and Helios knockdown in Jurkat T cells 85

Figure 3. 6 Gene Interaction Network and Pathway Enrichment in Module 4 of DEGs between wild type and Helios knockdown in Jurkat T cells 87

Figure 3. 7 Gene Interaction Network and Pathway Enrichment in Module 5 of DEGs between wild type and Helios knockdown in Jurkat T cells 89

Figure 3. 8 Gene Interaction Network and Pathway Enrichment in Module 6 of DEGs between wild type and Helios knockdown in Jurkat T cells 92

Figure 3. 9 Gene Interaction Network and Pathway Enrichment in Module 7 of DEGs between wild type and Helios knockdown in Jurkat T cells 94

Figure 3. 10 Gene Interaction Network and Pathway Enrichment in Module 8 of DEGs between wild type and Helios knockdown in Jurkat T cells 96

Figure 3. 11 Gene Interaction Network and Pathway Enrichment in Module 9 of DEGs between wild type and Helios knockdown in Jurkat T cells 100

Figure 4. 1 Pathways enriched for the target genes of Helios annotated from the binding sites in Jurkat T cells..... 109

Figure 4. 2 Integration of Helios Binding Sites with Promoter and Enhancer Annotations in Jurkat T Cells 110

Figure 4. 3 Target genes and pathways overlap between bindings site annotation and DEGs of Helios in Jurkat T cells 111

Figure 4. 4 Protein-Protein Interaction Network of Prioritized Helios Target Genes in Jurkat T Cells 116

Figure 5. 1 Comparison of True Positive Rate and Precision in Trans-Cell Type Helios Binding Site Prediction Methods..... 130

Figure 5. 2 Integration of Helios Binding Sites with Promoter and Enhancer Annotations in CD+ naïve T Cells 132

Figure 5. 3 Pathway Enrichment for the predicted target genes of Helios in CD4+ naïve T cells 132

Figure 5. 4 Pathway Enrichment for the predicted target genes of Helios in Th1 cells..... 134

Figure 5. 5 Pathway Enrichment for the predicted target genes of Helios in Th17 cells..... 135

Figure 5. 6 Overlap of predicted target genes and pathways enriched in the predicted target genes across Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells 136

Figure 6. 1 Differential Expression of *IKZF2* Across Various T Cell Subtypes between Healthy controls and SLE patients..... 147

Figure 6. 2 Results of gene set enrichment analysis of predicted target genes of Helios in DEGs between Healthy controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells 149

Figure 6. 3 Pathway enrichment of core enriched Helios target genes in DEGs between HC and SLE in Th1 cells 152

Figure 6. 4 Modules identified from the core enriched genes of predicted Helios target genes to DEGs between healthy controls and SLE patients in Th1 cells 153

Figure 6. 5 Pathway enrichment of core enriched Helios target genes in DEGs between HC and SLE in Th1 cells 155

Figure 6. 6 Modules identified from the core enriched genes of predicted Helios target genes to DEGs between healthy controls and SLE patients in Th17 cells 156

Figure 6. 7 Top 20 predicted target genes of Helios in Th1 T cells based on the correlation with *IKZF2* in Th1 cells of SLE patients 158

Figure 6. 8 Top 20 predicted target genes of Helios in Th17 T cells based on the correlation with *IKZF2* in Th17 cells of SLE patients 158

Figure 6. 9 Venn Diagram Illustrating Shared Target Genes of Helios in Th1 and Th17 Cells Related to SLE 160

Abstract

Backgrounds: Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease associated with a wide range of clinical features involving different organs and the prognosis is also highly variable. *IKZF2*, which encodes Helios, is a Krüppel-like zinc finger transcription factor belonging to the Ikaros family. This group of transcription factors is integral to the regulation of the immune system. Furthermore, *IKZF2* has been identified as a risk locus for SLE. Given that the expression of Helios is largely restricted to T-cells, and dysregulation of T cell function is an important factor for SLE pathogenesis, I sought to unravel the biological mechanism for the association by leveraging ChIP-Seq data for Helios in Jurkat T-cells. This study employed a multi-dimensional approach to delve into the intricacies of Helios in Jurkat T cells. The research integrated genome-wide identification of Helios binding sites, functional annotation of these sites, and their intersection with gene expression data, offering a comprehensive view of its regulatory network. Furthermore, the investigation extended beyond Jurkat T cells, exploring Helios binding sites and target genes across various T cell subtypes and their potential implications in SLE.

Methods: Using a multi-faceted bioinformatics approach, I aimed to investigate the regulatory role of the transcription factor Helios in Jurkat T cells. The research encompassed the comprehensive identification of Helios binding sites and functional annotation of these binding sites using multi-omic datasets from either publicly available sources or accessed through collaboration. Using these resources I also went on to explore the binding sites and target genes of the transcription factor Helios in various T cell subtypes. I predicted Helios binding sites across CD4+ naïve T cells, Th1 cells, and Th17 cells by combing its binding sites in Jurkat T cells and DNase I hypersensitivity data in these cell types. Subsequently, I identified target genes associated with these binding sites using pre-established promoter annotation and

enhancer annotation data and conducted gene set enrichment analysis to reveal their role in SLE.

Results: The extensive analysis of Helios in T cell regulation and its impact on SLE has yielded significant insights. A genome-wide chip-seq analysis in Jurkat T cells identified 5,068 binding sites of Helios, indicating its extensive role in gene regulatory mechanisms within T cells. Notably, a substantial proportion of these sites are located in promoter regions and distal intergenic areas, suggesting Helios's involvement in both direct transcription initiation and long-range gene regulation. ChromHMM annotation further revealed the majority of these sites are associated with active promoters and enhancers. The study also observed variable overlaps of Helios binding sites with SNPs in different diseases, with significant associations in Crohn's Disease, Multiple Sclerosis, and Rheumatoid Arthritis, highlighting its potential role in these conditions. Differential gene expression analysis following Helios knockdown in Jurkat T cells showed significant changes in 1,072 genes, with pathway analysis revealing enrichment in crucial processes like cholesterol biosynthesis, apoptosis, and T cell receptor regulation. Utilizing the MCODE plugin in Cytoscape, ten distinct biological modules were identified, encompassing a range of functions from epigenetic modulation to immune responses. A total of 56 priority target genes of Helios in Jurkat T cells were identified by intersecting the genes annotated from Helios binding sites in Jurkat T cells with the differentially expressed genes (DEGs) observed following Helios knockdown in the same cell type. This list includes several key genes that have been previously implicated in the development of SLE, such as *IRF4*, *PRKCB*, and *CD9*.

Significant changes (adjusted p value <0.05) in the expression of Helios in specific T-cell subtypes in SLE patients compared to healthy controls are observed, with Th1 cells showing the most substantial difference (log (FC) =1.96 using with our analysis based on a Japanese bulk RNA-seq dataset that encompasses data from various immune cell types across both healthy controls and SLE patients. In order to observe the function of Helios in different T cell subtypes, we employed a comprehensive bioinformatics approach to predict Helios transcription factor binding sites and their target genes in various T cell subtypes. Our focus was primarily on Th1,

naïve CD4+, and Th17 cell types due to the availability of relevant datasets for DNase I hypersensitivity hotspots and pre-established annotation data in these specific subtypes.

Gene Set Enrichment Analysis revealed that the predicted target genes were enriched at the "top" of the list of differentially expressed genes when comparing healthy controls and SLE patients in both Th1 and Th17 cells (with a significance threshold set at p-value <0.05). I identified 190 Helios' predicted target genes in Th1 cells and 70 Helios' predicted target genes in Th17 cells that were found to be centrally contributed to differentially expressed genes (DEGs) between healthy controls and SLE patients. Utilizing protein-protein interaction (PPI) networks and plugins like MCODE and Cytohubba in Cytoscape, I pinpointed key hub genes associated with SLE in these cell types. Specifically, *CCNA2, MRPL58*, and *CXCR6* emerged as hub genes in Th1 cells, while *IRF4* and *CBFB* were identified in Th17 cells. Additionally, seven genes – *IFI6, FKBP5, TRIB1, PRDM1, TRAT1, LINC00426*, and *CCND3* – were found to be core predicted target genes of Helios in both Th1 and Th17 cells within the DEGs between healthy controls and SLE patients. Further analysis of the expression changes and correlation with IKZF2 expression in these 12 genes revealed that *IFI6* in both Th1 and Th17 cells, as well as *CCNA2* in Th1 cells, are particularly significant. These genes not only exhibit considerable changes in expression but also show a notable correlation with *IKZF2*, underscoring their importance as target genes of Helios and their potential role in SLE pathogenesis in these T cell subsets.

Conclusion: This study elucidates the role of the transcription factor Helios in T cell regulation and its implications in SLE. The study revealed that Helios exhibits a predilection for binding sites near transcription start sites and active promoter regions, indicating its pivotal role in the regulation of gene expression. Functional annotation disclosed Helios' involvement in biological processes such as sterol biosynthesis and cholesterol metabolism, epigenetic modification and immune responses. the study explored changes in Helios expression across various T cell subtypes, particularly highlighting Th1 cells as the most affected. Additionally, gene Set Enrichment Analysis showed that Helios' predicted target genes were significantly enriched

among differentially expressed genes in Th1 and Th17 cells in SLE patients, indicating its potential role in the development of this autoimmune disease within these cell types. Through protein-protein interaction network analysis, hub target genes of Helios related to SLE in each cell type were identified.

Publications

Wang Y, **Guga S**, Wu K, et al. COVID-19 and systemic lupus erythematosus genetics: A balance between autoimmune disease risk and protection against infection. PLoS Genet. 2022;18(11):e1010253.

Guga S, Wang Y, Graham DC, Vyse TJ. A review of genetic risk in systemic lupus erythematosus. Expert Rev Clin Immunol. 2023;19(10):1247-1258

Acknowledgements

I would firstly like to express my utmost gratitude to my supervisors Dr. Deborah Cunninghame Graham and Professor Tim Vyse, for the opportunity to pursue this PhD. Their invaluable guidance, instruction, and encouragement have been the cornerstones of my journey through the last four years. Their unwavering support has been instrumental to my development and research, and for this, I am profoundly thankful. I am also grateful to the China Scholarship Council (CSC) and King's College London for their generous financial support through the PhD Scholarship, which has been essential for my academic endeavours.

I am indebted to my thesis committee members, Professor Peter Zammit, Dr. Jordana Bell, Dr Mansoor Saqi, and Dr Cameron Osborne, for their insightful suggestions and constructive critiques that significantly shaped my doctoral research.

This thesis would not have been achievable without the help and scientific input of my colleges both past and present in the Immunogenetics group. I owe a debt of gratitude for their collaborative spirit and assistance. A special mention to Nick, whose professional expertise and genuine support have been particularly noteworthy.

I extend my heartfelt thanks to Kejia, Susan, Lu, Yu Xuan, Chen, and Siwen for their kindness and camaraderie during my PhD. I wish you all every success in your future endeavours.

To my husband and my parents-in-law, I am grateful for your endless support and encouragement throughout this challenging and rewarding process.

To my dear sister, whose infectious laughter has been my greatest source of joy—thank you for the endless energy and the light-hearted moments that have brightened my days and sustained me through this journey. Your support means the world to me.

Most importantly, I dedicate this thesis to my exceptional parents, whose unwavering belief in me and support for my decisions have been my constant source of strength and inspiration. You are truly special to me, and this achievement is as much yours as it is mine.

COVID-19 Impact Statement

The first year of this project was substantially disrupted by the COVID-19 pandemic. The restrictions and safety protocols in place required all meetings to be conducted virtually over platforms like Zoom, significantly impacting face-to-face collaboration and communication. Additionally, the pandemic caused considerable delays in obtaining Data Transfer Agreements for essential data from the source, which further impeded the progress of our research activities. Despite these challenges, the team demonstrated resilience and adaptability, ensuring continued progress towards our research objectives.

Contribution Statement

Throughout this thesis, the contributions of various team members are acknowledged, but for clarity, the division of work is outlined here. The ChIP-Seq experiments, quality control of ChIP-Seq data, and validation of antibody suitability were conducted by previous lab members Chris Odhams and Andrea Cortini. All subsequent aspects of the project, including data analysis, interpretation of results, preparation of figures and tables, were carried out by myself.

Chapter 1. Introduction

1.1 Overview of Systemic Lupus Erythematosus (SLE)

The immune system is an interactive network of cells and molecules with specialised roles in host defense against infection. The ability to discriminate between self and non-self-antigens is crucial to the function of the immune system as a specific defense against invading microorganisms. Failure of self-tolerance can result in pathological states causing autoimmune disease. Autoimmune diseases have an overall prevalence of 3–5% in the general population [1]. SLE is a complex multisystem autoimmune disease with a wide range of signs and symptoms in affected individuals.

1.1.1 Epidemiology

The incidence of SLE is affected by genetic ancestry and sex. The female: male incidence ratio is 10:1 in adults. The disease burden of SLE has been shown to be highest those with African ancestry, followed by Asian (south or east) and those with European ancestry have the lowest prevalence (and severity) of SLE. The incidence of SLE varies geographically: according to a recent systematic review of the worldwide epidemiology of SLE [2], The overall global incidence of SLE ranges from 1.5 to 11 per 100,000 person-years, and the global prevalence from 13 to 7,713.5 per 100,000 person-years. Racial disparity is seen in the mortality of SLE: compared with European ancestry SLE patients, cumulative SLE mortality was significantly higher among those with African ancestry [3]. This likely reflects both disease severity and socio-economic factors in some settings.

1.1.2 Aetiology and Risk Factors

Despite extensive research over the years, the exact cause of SLE remains elusive. While some genetic predisposition is involved, environmental triggers such as infections, sun exposure, and certain medications are thought to play a role in the development of the disease [4]. Additionally, hormonal factors, particularly in women, are believed to contribute to the higher prevalence of SLE in females [5].

1.1.2.1 Update on findings from genetic studies

Although a few autoimmune diseases, such as autoimmune polyendocrinopathy syndrome type 1, which can be explained by a mutation in the autoimmune regulator (AIRE) gene, are monogenic; most autoimmune diseases are polygenic (complex genetic traits), with each susceptibility gene conveying a modest and non-exclusive elevation of risk.

Genome-wide association studies (GWAS) has led to radical increase in the identification of genetic risk factors that underpin complex diseases. to date, these studies have revealed around 200 risk loci associated with SLE [6]. Although many genetic risk loci for SLE have been identified, the explained heritability by these risk loci is still low. Morris DL et al. [7] used all genotyped SNPs in a Chinese cohort and a European cohort to calculate heritability in both populations. The results revealed 28% explained heritability in Chinese subjects and 27% in Europeans. López-Cortegano E et al. [8] reported the 17% explained heritability in SLE calculated using a selection of the most informative SNPs (P-value ≤5×10−8) available from the NHGRI-EBI GWAS Catalog [9]. One speculation surrounding the missing heritability in GWAS studies pertains to the potential contributions of rare variants that may not be adequately captured by SNP-chips. Such variants can be captured by exome sequencing or even more reliably by whole-genome sequencing (WGS). By performing whole-genome sequencing (WGS) of samples from 71 Swedish SLE trio families with two healthy parents and one child affected by

SLE, by Almlöf JC et al. [10] identified one previously reported homozygous nonsense mutation in the *C1QC* (Complement C1q C Chain) gene and seven ultra-rare, coding heterozygous variants in five genes (*C1S*, *DNASE1L3*, *DNASE1*, *IFIH1*, and *RNASEH2A*) involved in monogenic SLE. As the utilization of WGS becomes more widespread in future studies, it offers the potential to provide a more comprehensive understanding of the genetic landscape of SLE and unveil additional genetic factors that contribute to its heritability.

In the context of complex diseases, individual genetic variants often have a modest impact on disease risk. Nevertheless, their collective effect can yield meaningful risk predictions and enhance risk stratification. To this end, a polygenic risk score (PRS) amalgamates genetic risks identified through GWAS to generate a score that predicts an individual's disease risk based on their genotype. Chen L et al. [11]conducted a genetic risk score analysis for SLE across Chinese and European populations. Utilizing three European and two Chinese GWAS datasets and training on a dataset for one population, they found that the best performing GRS results in good predictive power with an area under the Receiver Operator Curve ROC curve (AUC) for SLE equal to 0.72 and 0.67 in two different cohorts. Wang YF et al. [12] found that an ancestry-matched risk score demonstrates 73.4% sensitivity and 65.4% specificity in individuals of Guangzhou ancestry and disease risk increased with higher PRS, with individuals in the highest PRS decile having a much higher disease risk than those in the lowest decile. In addition to providing value in the diagnosis of SLE, PGR is also shown to be useful in the differential diagnosis of SLE from other autoimmune diseases: Knevel et al. [13] developed a genetic probability tool (G-PROB) to calculate the probability of different inflammatory arthritis causing conditions (rheumatoid arthritis, systemic lupus erythematosus, spondyloarthropathy, psoriatic arthritis, and gout) for a patient using genetic risk scores. Calibration of G-probabilities with disease status was high, with regression coefficients from 0.90 to 1.08 (1.00 is ideal). G-probabilities discriminated true diagnoses across the three cohorts with pooled areas under the curve (95% CI) of 0.69 (0.67 to 0.71), 0.81 (0.76 to 0.84), and 0.84 (0.81 to 0.86), respectively. They further observed that 35% of the patients were misdiagnosed at the initial visit. In 77% of

patients, the final diagnosis was within the top two diseases with highest G-probabilities. This demonstrated that converting genotype information before a clinical visit into an interpretable probability value could significantly improve differential diagnosis.

1.1.2.2 Findings from Mendelian randomization (MR) studies

Numerous studies have established associations between SLE incidence and various environmental factors, including exposure to silica, smoking, alcohol consumption, infections, and vaccinations and other diseases such as endometriosis, allergic rhinitis, atopic dermatitis [14-17].

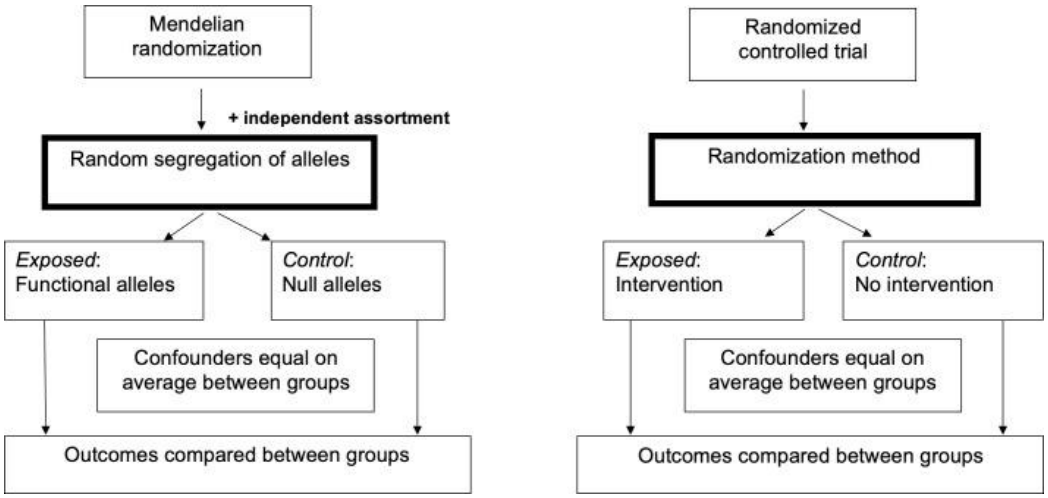
While observational studies demonstrate these associations, they fall short of establishing causation. The observed correlations are susceptible to potential "false correlations" influenced by numerous confounding factors, and the presence of reverse causality cannot be dismissed. Mendelian randomization (MR) study is a robust analytical method that can be applied to GWAS summary statistics data to estimate the causal relationship between traits and diseases. MR requires an exposure and an outcome and uses genetic variation to provide evidence that supports or rejects the hypothesis that the exposure has a causal effect on the outcome. In MR, it is assumed that specifically selected genetic variants behave similarly to treatment assignment, that the population is divided into subgroups in a way that mimics randomization (**Figure 1.1**) [18], and that the instrumental variables selected (usually SNP) should satisfy the relevance, exchangeability and exclusion restrictions.

In a study conducted by Bae, SC et al. [19] , the inverse variance-weighted (IVW) method suggests a causal link between periodontitis and SLE ($\beta < 0.01$, $SE < 0.01$, $P = 0.046$), while the weighted median approach and MR-Egger regression do not support the finding. In an analysis

conducted by Inamo J [20], using multiple Mendelian Randomization methods, a consistent and statistically significant causal association between Celiac disease and and SLE was demonstrated. Xiang K et al.'s MR study on the gut microbiome and SLE [21] revealed that the gut microbiota exerts a causal effect on the risk of SLE. Specific microbiota levels such as Bacillales, Coprobacter, Lachnospira, and Actinobacteria were found to have a negative association with SLE risk, while Bacilli, Lactobacillales, and Eggerthella may serve as risk factors for the development of SLE. Certain lifestyle factors were assessed in MR studies to discern their causal association with SLE. These investigations, including studies on coffee consumption [22] , alcohol intake [23, 24], smoking [23], and sleep disorders [25], notably did not reveal a causal link with the development of SLE.

MR methods have also been applied to data from methylation and gene expression quantitative trait loci (meQTL and eQTL) and plasma protein level QTL (pQTL) studies to explore potential causal factors for SLE. Mo X et al. [26] found that DNA methylation of 15 loci and mRNA expression of 21 genes were causally associated with SLE, particularly methylation and mRNA expression of known SLE risk genes, *UBE2L3* and *BLK*.

Figure 1. 1 The analogy between Mendelian randomization and randomized controlled trial.



Mendelian randomization (left) uses genetic variation to link risk factors with outcomes, circumventing confounders through nature's random assignment of genes. Randomized controlled trials (right) achieve similar confounder balance by randomly allocating subjects to intervention or control groups, allowing direct comparison of outcomes.

1.1.3 Pathophysiology

Although the exact mechanism is not fully understood, SLE is widely thought to be an autoimmune disease characterized by the excessive production of nuclear autoantibodies and

proteins. With the underlying genetic propensity and in response to environmental triggers, the balance of the immune system shifts to turn against itself, which primarily leads to activation of both innate and adaptive immunity. The ensuing activation of autoreactive B cells by T cells and leads to immune complexes deposition in tissues leading to an autoimmune cascade that may be limited to the single organ or cause a widespread systemic involvement.

1.1.3.1 Role of innate immune system in the development of SLE

The innate immune system consists of immune cells, including macrophages, neutrophils, dendritic cells, Natural killer cells, basophils, and innate lymphoid cells (ILCs), that circulate in blood or reside in tissues and are poised to respond to pathogens or inflammatory stimuli. The innate immune system has been indicated as a key player in the pathogenesis of SLE.

A role for macrophages in the pathogenesis of SLE is supported by the discovery that SLE macrophages were defective in their ability to clear apoptotic cell debris [27]. It is proposed the defect in the phagocytosis and clearance of apoptotic cell prolonged exposure of autoantigens to the adaptive immune cells provides survival signals for autoreactive B cells and consequently loss of tolerance to nuclear antigens released from apoptotic cells [28]. Notably, research suggests a different role of M1 and M2 macrophages in the progression of SLE. M1 macrophages increase the severity of the condition, while M2 macrophages reduce it [29]. Furthermore, cytokines that encourage M2 polarization, such as IL-4, have been reported to potentially offer therapeutic benefits in alleviating SLE symptoms [29].

Neutrophils are proposed to play a pathogenic role in SLE. Activated neutrophils in SLE do release proteases, factors that cause tissue damage, and reactive oxygen species (ROS), which collectively contribute to tissue destruction in SLE. Furthermore, the release of large amounts

of cytokines and chemokines by these activated neutrophils can lead to dysregulation of the immune system. In patients with SLE, a significant presence of neutrophil extracellular traps (NETs) is observed in the kidneys, skin, and blood [30, 31]. NETs are net-like structures composed of DNA-histone complexes and proteins, which are released by activated neutrophils into the extracellular space as a defense mechanism against invading pathogens. Suboptimal clearance and/or excessive formation of NETs (NETosis) are considered to result in autoantigen externalization and induce type I IFN synthesis and endothelial damage [32]. The abundance of these NETs is directly associated with the level of disease activity in SLE [33].

Studies have documented dysregulated dendritic cells (DCs) play a critical role in the initiation and development of SLE [34]. Several reports indicate that the frequency, composition, and phenotype of DCs in SLE patients differ from those of healthy individuals [34]. Decreased frequencies of Plasmacytoid DCs (pDCs) or Myeloid DCs were most often associated with active disease and to a lesser degree with nonactive disease [35]. Interestingly, studies showing peripheral pDCs decreases observed a concomitant infiltration of pDCs in the skin and kidney of patients with lupus, suggesting that active pDCs may have migrated to the sites of inflammation [36-38]. Plasmacytoid DCs (pDCs) in patients with SLE produce high levels of IFNα that causes a positive-feedback loop in the activation of innate and adaptive immunity [39]. Type I interferons, including IFNα, have recently been identified for their significant relationship with SLE in current research [40]. Type I IFNs not only activate dendritic cells (DCs) and nature killer (NK) cells involved in innate immunity, but also activate B cells and T cells to trigger the adaptive immune system. Evidence suggests that type I IFN may participate in the initial breaks in tolerance of the disease [40]. Increased IFNα levels in SLE patients correlate with increased disease activity and severity [41, 42]. Furthermore, IFN-α regulated gene transcripts have been found significantly upregulated in the peripheral blood of both pediatric and adult SLE patients, as shown by gene expression profiling [43, 44]. The findings make IFN-α a promising therapeutic target for SLE [45]. Moreover, the IFN-α signature is being evaluated as a potential new biomarker for monitoring SLE disease activity [46].

1.1.3.2 Role of B cells in the development of SLE

Abnormalities in B cells play pivotal roles in the pathogenesis of SLE. These abnormalities encompass several key aspects[47]. First, there are disturbances in peripheral B cell subsets, characterized by an imbalance in the numbers of different B cell subtypes, with an increase in class-switched memory B cells relative to naïve B cells[48]. Moreover, proportion of activated Memory B cell subsets (CD27+IgD+CD95+ B cells) are shown to be associated with disease activity in SLE [49]. Second, the abnormal activation of B cells is another central feature. Dysregulations in crucial pathways, including the B-cell receptor (BCR), toll-like receptor (TLR), and B-cell activating factor receptor (BAFF-R) pathways, are contributors [50]. Third, there is a notable defect in three immune checkpoints of B cell development in SLE patients [51]. SLE signalling pathway disorder, abnormal ubiquitination, TLR7 and TLR9 effects, and lack of CD72, result in loss of B cell immune tolerance to self-antigens, allowing autoreactive B cells to survive and produce autoantibodies[47]. Finally, immune complexes of autoantibodies and self-antigens deposit in target organs, causing inflammation and tissue damage[52].

1.1.3.3 Role of T cells in the development of SLE

T cell dysregulation has been increasingly acknowledged as central to SLE pathogenesis and is marked by an imbalance between populations with immunosuppressive functions and pathogenic T cell subsets, which contribute to the break in immune tolerance and ongoing inflammation (**Figure 1. 2**). T cells exhibit significant heterogeneity, with CD4+ helper T cells and CD8+ cytotoxic T cells comprising the majority. CD4+ helper T cells play a pivotal role in coordinating immune responses through the provision of co-stimulatory signals and cytokines.

Five main CD4+ T helper cell subsets have been identified: Th1, Th2, Th17, Treg (T regulatory) and Tfh (follicular helper) cells.

T follicular helper (Tfh) cells play a pivotal role in various aspects of the immune response, such as germinal center induction, isotype-switching, and somatic hypermutation [53]. They express the chemokine receptor CXCR5 and produce the cytokine IL-21, both of which are required for their contribution to germinal center formation. Inducible co-stimulator (ICOS), programmed cell death protein 1 (PD-1), and CD40L are also expressed on Tfh cells and these molecules are required for activation of B cells. Patients with SLE often produce somatically mutated IgG+ autoantibodies, indicating the involvement of germinal centers, where Tfh cells exert their influence. Several mouse models and clinical evidence demonstrate that dysregulated Tfh-cell and germinal center responses are linked to SLE [54]. In lupus nephritis lesions, Tfh-like cells expressing ICOS, PD-1, BCL-6, and IL-21 form ectopic germinal centers [55]. Additionally, a subset of SLE patients has an increased population of circulating CXCR5+ICOS+PD-1+ Tfh cells, which correlates with disease activity[56]. Of particular interest are the observations made on alterations in the composition of Tfh cells subsets in SLE, associated with disease activity. Ratio of Tfh2 (CXCR3(-) CCR6(-))over Tfh1(CXCR3 (+) CCR6(-)) are increased in SLE patients as compared to controls and disease activity correlates with the frequency of Tfh2 cells [57].

Regulatory T (Treg) cells are a unique T cell subset population that suppresses the immune response and maintains self-tolerance, suppressing autoreactive lymphocytes in healthy individuals. Controversial results were reported whether decreased and increased Tregs cells frequencies were reported in SLE[58]. Nevertheless, a disturbed balance between effector T cells and Treg cells is evident in SLE [59], which is due to an imbalanced T cell cytokine profile characterized by decreased Interleukin-2 (IL-2) [60]. It has been well-documented that SLE is associated with impaired IL-2 production by T cells [61]. Restoring the balance through IL-2

treatment in lupus-prone mice slows disease progression. Conversely, reducing Treg cell numbers in healthy mice accelerates the disease onset [58].

Th1 cells are recognized for their production of the pro-inflammatory cytokine interferon-gamma (IFN-γ) and play a crucial role in cell-mediated inflammatory responses as well as defense against intracellular pathogens. There is accumulating evidence suggesting that Th1 cells may contribute to tissue injury in murine lupus models [62]. In addition, Th1 cells are found to be enriched in inflamed kidneys and the number of Th1 cells in urine correlates with disease activity in patients with lupus nephritis [63]. Notably, patients with SLE exhibit elevated levels of IFN-γ compared to controls, and these levels positively correlate with SLE disease activity index (SLEDAI) scores [64, 65]. However, administering anti-interferon-γ monoclonal antibody to lupus patients didn't yield significant results in terms of reducing disease activity, organ damage, or autoantibody titer[66], which raises questions about the role of Th1 cells in lupus pathogenesis.

Th2 cells play crucial roles in maintaining humoral immunity by secreting key cytokines such as IL-4, IL-5, and IL-13, and their increased frequency in both SLE mouse models and SLE patients [67-69], suggests potential involvement in SLE development. IL-4 activates Janus-family tyrosine kinases Jak1 and Jak3 that lead to phosphorylation of the transcriptional factor Stat6. Studies have found that *IL-4* gene polymorphism is significantly associated with SLE susceptibility [70, 71]. SLE patients have significantly lower levels of IL-4 in plasma or serum compared to healthy controls [72, 73], implicating a critical role of IL-4 in SLE. In lupus prone mice, blocking IL-4 decreases IgG anti-double-stranded DNA antibodies (anti-dsDNA), while administration of IL-4 enhances the production of this autoantibody [74]. Traditionally, IL-5 has been recognized as a cytokine that stimulates activated B cells to produce antibodies and promotes the proliferation and differentiation of eosinophils from precursor cells into mature forms. IL-5 was reported to be overexpressed in skin inflammation[75]. IL-13 promotes the proliferation and differentiation

of B cells, and induces the expression of CD23 and IgE [76]. Multiple studies have documented elevated circulating levels of IL-13 in serum samples from SLE patients, significantly correlating with disease activity [77, 78]. Moreover, IL-13 levels are notably higher in lupus nephritis patients compared to those with SLE without renal involvement and healthy controls. Additionally, the association between IL-13 polymorphism and SLE risk in the Chinese population has been explored [77].

Th17 cells play a crucial role in the immune system by promoting inflammation and defending against bacterial and fungal infections. They are characterized by their production of the cytokine interleukin-17 (IL-17) and are involved in various autoimmune and inflammatory diseases due to their ability to trigger immune responses. The role of Th17 cells in SLE has been supported by elevated serum IL-17 levels [79, 80] and increased circulating Th17 cell frequencies [81, 82], both of which align with disease activity [83-85]. In BXD mice with autoimmune tendencies, blocking IL-17 signaling led to reduced development of germinal center B cells and decreased production of autoantibodies, including anti-dsDNA antibodies[86]. Additionally, In SLE patients, an elevated Th17 level, a lower Treg level, and an increased Th17/Treg ratio are commonly observed, which is often associated with the disease activity [87-89].

CD8+ T cells are cytotoxic immune cells that kill infected or damaged cells by releasing cytotoxins, such as granzymes and perforins, enabling them to regulate infection, malignancy, and autoreactive immunity. The role of CD8+ T cells in SLE is not well determined. Patients with active SLE have increased numbers of activated CD8+ T cells [90] and CD8+ T cells are present in the kidneys of patients with lupus nephritis [91]. On the other hand, CD8+ T cells from the peripheral blood of SLE patients frequently display a reduction in effector function [92], including attenuated granzyme B and perforin production [93]. This impaired cytotoxicity allows autoreactive B cells to persist, leading to the production of autoantibodies and the

characteristic autoimmune response in SLE. Moreover, SLE patients with compromised CD8+ T cell function are at higher risk of infections [94], including challenges in controlling latent viruses such as Epstein–Barr virus [95, 96]. Double negative T cells, a subset of αβ T cells characterized by the absence of CD4 or CD8 surface markers, are significantly elevated in SLE patients [97] and in murine disease models [98]. While their exact origin is not fully understood, evidence suggests they may derive from self-reactive CD8+ T cells [99, 100]. These double negative T cells are believed to play a prominent role in SLE pathogenesis by infiltrating target organs (such as the kidneys [101]) and being a major source of IL-17 production [102, 103].

1.1.4 Clinical aspects: diagnosis and treatment of SLE

SLE is a chronic autoimmune disease that causes inflammation and can affect multiple organs like the skin, joints, kidneys, and lungs, with varying severity. Diagnosing and treating SLE is challenging due to its diverse symptoms, different patient subsets, comorbidities, and medication effects. The diagnosis is clinical and supported by lab tests showing immune reactivity or inflammation. The diagnosis typically follows the 1997 American College of Rheumatology (ACR) criteria, which require four of eleven specific parameters for a diagnosis. Combining it with newer criteria (SLICC-2012 and EULAR/ACR-2019) could be helpful in early classification of SLE [104]. Assessing disease activity using tools such as the SLE Disease Activity Index (SLEDAI) aids in effectively managing SLE during both remission and flare-up periods [105]. The management of SLE is complex and require a multidisciplinary approach. Treatment goals include long-term patient survival, prevention of organ damage and improvement of health-related quality of life. The therapeutic focus is on achieving remission or at least low disease activity and prevention of flares. According to the recommendations from EULAR, updated in 2019, based on patient preference and the heterogeneity and severity of the condition, several therapeutics such as antimalarials (hydroxychloroquine), immunosuppressives (azathioprine, methotrexate, mycophenolate), and glucocorticoids may have to be considered simultaneously.

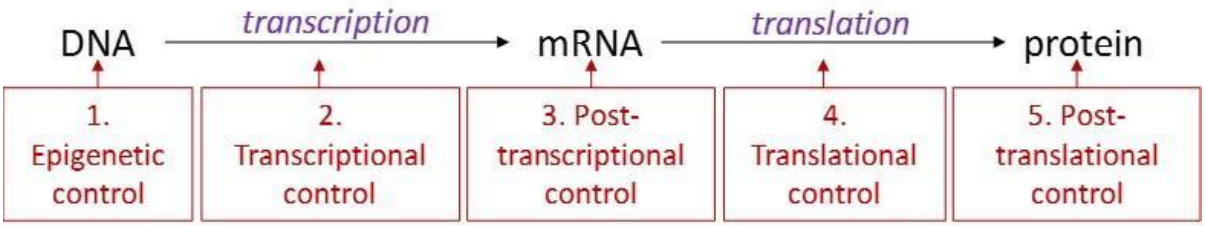
Special considerations are given to specific manifestations like cutaneous, neuropsychiatric, and renal disease. For lupus nephritis (LN), treatment targets include reducing proteinuria and maintaining stable kidney function, with specific protocols for proliferative LN involving medications like mycophenolate mofetil or cyclophosphamide. Unfortunately, standard therapies don't always achieve remission, leading to interest in biologics like belimumab and rituximab, especially in refractory cases [106]. Research is actively ongoing to discover and evaluate new treatment alternatives[107].

1.2 Regulation of Transcription and Gene Expression in Eukaryotes

In cellular biology, the synthesis of essential proteins in a timely manner is fundamental for cellular functionality. This process, known as gene expression, involves the activation of genes to produce mRNA and proteins, a mechanism that is meticulously regulated in both unicellular and multicellular organisms. Cells exert control over the timing, quantity, and discontinuation of protein production as required. In prokaryotes, DNA organization is typically in a circular chromosome within the cytoplasm. Genes related to specific functions or biochemical pathways are grouped in operons, allowing coordinated control of gene expression. Operons are transcribed into a single mRNA molecule, enabling unified regulation. Prokaryotic gene expression is modulated by activators, repressors, and inducers, which respectively increase, suppress, or deactivate transcription. Eukaryotic gene regulation is more intricate, occurring independently for each gene and across various stages from DNA to protein synthesis. This regulation encompasses multiple levels: epigenetic, transcriptional, post-transcriptional, translational, and post-translational (**Figure 1. 2**)**.** Epigenetic mechanisms modify DNA accessibility. Transcriptional control determines gene activation status. Post-transcriptional modifications impact mRNA processing and stability. Translational controls dictate protein synthesis efficiency. Post-translational modifications alter protein function. Among these, the transcriptional level is often considered the most critical point of regulation, as it is the

primary gateway to gene expression, determining the initial and fundamental output of the genetic information.

Figure 1. 2 Five different levels of regulation of gene expression in eukaryotes



The Central Dogma is illustrated with arrows indicating the points at which different types of eukaryotic gene expression regulation can occur.

1.2.1 Overview of transcription factors

Transcriptional regulation is control of whether or not an mRNA is transcribed from a gene in a particular cell. Like prokaryotic cells, the transcription of genes in eukaryotes requires an RNA polymerase to bind to a promoter to initiate transcription. In eukaryotes, RNA polymerase requires transcription factors to facilitate transcription initiation. Transcription factors (TFs) are regulatory proteins whose function is to activate (or more rarely, to inhibit) transcription of DNA by binding to specific DNA sequences. TFs have defined DNA-binding domains with up to 106-fold higher affinity for their target sequences than for the remainder of the DNA strand. Based on broad structural similarities of their DBDs, transcription factors (TFs) can be divided

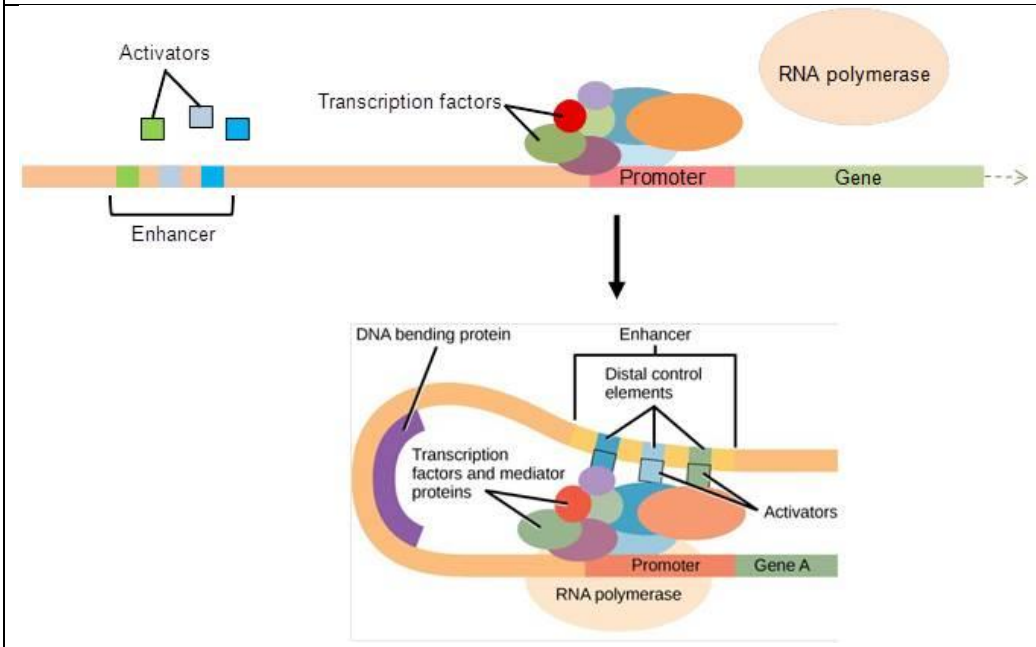
into several categories: Zinc Finger TFs, Helix-Turn-Helix (HTH) TFs, Leucine Zipper TFs, Helix-Loop-Helix (HLH) TFs etc. [108].

TFs control the expression of most genes by binding to promoter and/or enhancer regions of DNA. In eukaryotic genes, the promoter region is located immediately upstream of the coding sequence, where transcription by RNA polymerase II (RNA Pol II) is initiated. Alternatively, when TFs bind to enhancers, which act independent of orientation, distance, and location with respect to the genes they regulate[109], they enhance transcriptional activity by facilitating the formation of a more favorable chromatin structure. This is often achieved through the DNA looping that brings enhancers in proximity to target promoters (**Figure 1. 3** [110]). This reliance on three-dimensional proximity for enhancer function allows enhancers to impact genes that are far away on the DNA sequence (>1 kb) rather than those nearby. Although mechanisms of transcriptional activation at promoters have been extensively characterized, understanding the way genes are controlled by enhancer regions remains an area of intense study. A breakthrough in in this area has been the identification of topologically associated domains (TADs)[111]. TADs are essentially extruded chromatin loops there are minimal interdomain interactions and maximal intradomain interactions. TADs are usually consistent within specific cell types and help ensure that critical genes for cell identity are regulated by multiple enhancers, maintaining stable gene expression[112]. They can be reorganized when progenitor cells differentiate, altering the cells' gene expression patterns. Interestingly, not every gene within a TAD relies on it for functionality, indicating more complexity in the genomic structure. This complexity could be significant for developing targeted therapies for specific genes within TADs [113].

Transcription factors (TFs), which account for about 8% of all human genes, are linked to a diverse range of diseases and physical traits [114]. Research by the Human Phenotype Ontology project reveals that out of all human TFs, 313 (or 19.1%) are connected to at least one

phenotype which is a significantly larger proportion compared to the general gene population, where 16.2% show such connections ($p = 0.002$, proportions test)[115]. For example, mutations in TF genes are notably prevalent in conditions like anterior pituitary hypoplasia, underscoring the critical role of TFs in developmental disorders[115]. Mutations in TF genes can also lead to aberrant cell growth and cancer, as they regulate critical genes involved in cell proliferation and death[116]. Moreover, TFs are central in metabolic diseases like diabetes, where mutations in pancreatic TFs disrupt insulin regulation[117]. Additionally, genome-wide association studies have found TF loci associated with polygenic autoimmune diseases[114].

Figure 1. 3 Transcription factors bind to promoters/enhancers to initiate transcription



The promoter binds to transcription factors and helps RNA polymerase to bind and start transcription. Many genes also have upstream enhancers. Enhancers attract activator proteins, loop towards the promoter, and facilitate the commencement of transcription by RNA polymerase.

1.2.2 Methodologies for Transcription Factor Studies

Studying transcription factor binding sites and their target genes is essential for understanding gene regulation mechanisms. Various techniques are used in this area of research, each with its unique applications and insights.

1.2.2.1 Electrophoretic Mobility Shift Assay (EMSA)

The Electrophoretic Mobility Shift Assay (EMSA), also known as the gel shift assay, has been a staple in molecular biology since its development in the 1980s for the investigation of protein-DNA interactions, especially the binding of transcription factors to DNA. This technique, described by Fried and Crothers in 1981, exploits the differential electrophoretic mobility of molecules in a nondenaturing gel, a property dependent on their size, shape, or charge. When a protein binds to DNA, it forms a complex that exhibits slower migration through the gel than unbound DNA. This mobility shift, which signals binding, can be detected using autoradiography or fluorescence. To facilitate this process, transcription factors (TF) or DNA-binding domains (DBD) are mixed with labeled target DNA sequences in a buffer of low ionic strength and incubated to allow complex formation. Electrophoresis is then used to separate these bound complexes from any free DNA. EMSA is particularly useful in pinpointing the specific DNA sequences that transcription factors bind to, thereby identifying their target genes and regulatory regions. This is achieved by using a variety of DNA sequences as probes and assessing the impact of nucleotide changes on TF binding, which elucidates consensus binding motifs [118]. Moreover, EMSA can quantify the binding affinity and specificity of a transcription factor, offering insights into the dynamics of gene regulation [119]. Additionally, EMSA can be employed to study the formation of transcription factor complexes by adding other proteins or cofactors to the binding reaction, highlighting the multi-protein nature of transcriptional regulation [120]. Collectively, EMSA's ability to characterize transcription factor-DNA interactions renders it invaluable for advancing our understanding of the intricate mechanisms underlying gene expression and its associated biological processes and diseases.

1.2.2.2 DNA Footprinting

DNA footprinting, a technique integral to molecular biology, enables the precise identification of transcription factor (TF) binding sites on DNA. This method, first introduced in the 1970s, starts with labeling one end of the DNA and incubating it with DNA-binding proteins. then the DNA is degraded using agents such as DNase1 [121] or hydroxyl radicals [122], leading to a collection of fragments of different size. When analyzed using gel electrophoresis, these fragments create a characteristic ladder pattern ("footprint"). A DNA-bound protein shields the region near its binding site from the degrading agent, and the protected region can be determined by comparing sample and control footprints. Sequences corresponding to the protected areas can then be aligned to identify binding sites for the TF analyzed. It's worth noting that this method is highly sensitive to the protein concentration, with higher TF concentrations resulting in the protection of weaker binding sites [123]. Despite its effectiveness, traditional footprinting is low-throughput and requires significant amounts of purified protein. Advancements like in vivo footprinting and high-throughput techniques such as DNase-seq have expanded its applicability, enabling genome-wide studies of protein-DNA interactions within their natural cellular context [124]. These developments have solidified DNA footprinting's role as a crucial tool for unraveling the complex interactions between TFs and DNA, providing deep insights into the regulatory networks governing cellular processes and disease states.

1.2.2.3 ChIP-seq

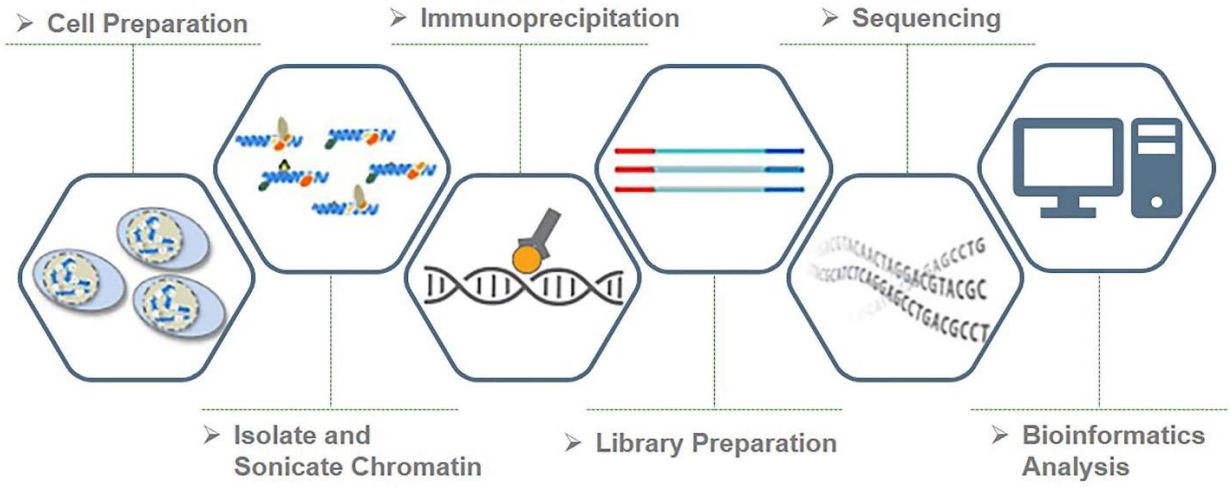
ChIP-seq [125] has revolutionized the study of TF-binding sites in vivo by enabling the genome-wide identification of region occupied by a TF of interest. Chromatin immunoprecipitation (ChIP) accompanied by NGS is one of the most successful NGS applications in identifying the binding sites of transcription factors and defining epigenetic modifications. Chromatin

immunoprecipitation followed by sequencing (ChIP-seq) first utilize the chromatin immunoprecipitation(ChIP) technique to pull down attached DNA fragments then identify the binding sites of DNA-associated proteins with massively parallel DNA sequencing. With the rapid development of sequencing technology, ChIP-seq has become the standard assay for whole-genome mapping of protein–DNA interactions in vitro and in vivo. Basically, cells in the ChIP assay are treated with a crosslinking agent to covalently bind any DNA-binding protein to the chromatin. Then, cell lysis is performed, and genomic DNA is extracted and sonicated to generate fragmented chromatin. An antibody specific to the protein of interest is introduced to the sonicated material, facilitating the isolation of the protein along with all associated DNA through immunoprecipitation. The DNA is released by reversing the crosslinking between DNA and protein to separate them and cleaning DNA with an extraction. Finally, after size selection, all the resulting ChIP-DNA fragments are sequenced simultaneously using a genome sequencer followed by a series of bioinformatic analysis including quality control, alignment of the reads to a reference genome and peak calling etc. (**Figure 1. 4**). ChIP-seq offers a distinct advantage over other methods like DNA footprinting, Electrophoretic Mobility Shift Assay (EMSA), and DNase-seq in terms of scope and specificity. While footprinting and EMSA provide detailed insight at specific loci, they do not possess the high-throughput capability to analyze the entire genome and lack the direct association of specific proteins with DNA sequences [125]. DNase-seq, though effective in identifying open chromatin regions, does not inherently indicate the binding of specific TFs at these sites [124]. ChIP-seq overcomes these limitations by not only indicating chromatin accessibility but also by directly connecting specific TFs to their binding sites across the genome. Additionally, Chromatin immunoprecipitation (ChIP) represents a valuable alternative to probing such interactions in vivo under physiological conditions [126]. However, this technique is not no limitation. A significant challenge lies in the dependency on the quality and specificity of the antibodies used; non-specific binding or subpar antibodies can lead to inaccurate mapping of binding sites and ChIP-grade antibodies are not available for many TFs. Additionally, ChIP-seq's resolution, while superior to previous methods, does not always achieve single-nucleotide precision, which can result in ambiguity in pinpointing exact binding motifs within identified regions of enrichment [127]. The interpretation of the vast data

generated by ChIP-seq also poses a challenge, necessitating sophisticated bioinformatics tools for analysis, and the risk of misinterpretation remains a concern [128]. Furthermore, the

technique may produce false positives due to non-specific antibody interactions or chromatin structure artifacts, and false negatives can occur, particularly with low-abundance TFs or transient interactions [129]. Despite these challenges, ongoing advancements in antibody development, sequencing technology, and data analysis continue to enhance the efficacy and reliability of ChIP-seq, solidifying its importance in exploring gene regulation and TF dynamics.

Figure 1. 4 Chromatin immunoprecipitation assays with sequencing (ChIP-Seq) workflow to identifying binding sites of a transcription factor.



The workflow typically starts with the process 'Cell Preparation' where cells are cultured and collected. Next is 'Isolate and Sonicate Chromatin,' where DNA-protein complexes are isolated, and the DNA is sheared into smaller fragments. The 'Immunoprecipitation' step follows, where specific antibodies are used to target and pull down the transcription factor-DNA complexes. 'Library Preparation' is where the DNA fragments are prepared for sequencing. 'Sequencing' itself is depicted as the generation of readable DNA sequences. Finally, 'Bioinformatics Analysis' involves using computational tools to analyze the sequence data to identify the binding sites of the transcription factors on the DNA.

1.2.3.4 RNA Interference (RNAi) and CRISPR/Cas9

Gene expression profile analysis before and after knockdown or knockout is one of the most important strategies for obtaining target genes of transcription factors and exploring their functions. A large number of studies show that gene expression profile analysis before and after knockdown or knockout effectively helps identify target genes of TFs and explore TF functions [130]. While these approaches do not directly reveal TF binding sites, they provide valuable insights into the genes and pathways regulated by these factors. Gene knockdown, typically achieved through techniques like RNA Interference (RNAi) [131], involves reducing the expression of a particular gene. This can lead to changes in the expression of downstream genes, providing insights into potential targets of the TF whose binding sites are being investigated. Similarly, gene knockout, commonly accomplished using CRISPR/Cas9 technology[132], eliminates the expression of a target gene. This perturbation can lead to significant alterations in the cellular transcriptome, with downstream genes potentially being regulated by the TF in question. Then the resulting alterations in gene expression profiles can be examined by RNA-sequencing or microarrays. Computational analysis of gene expression data from these experiments assists in identifying potential target genes and pathways influenced by the TFs. By integrating these approaches with ChIP-seq, which directly identifies

TF binding sites, we can gain a comprehensive understanding of the regulatory landscape. Genes showing significant expression changes along with evidence of transcription factor binding can be annotated as direct target genes. Indirect targets are identified as genes that display expression changes but show no indication of binding by the target transcription factor. These experimental approaches have been instrumental in advancing our knowledge of TF binding sites and gene regulation.

1.3 Overview of *IKZF2*

1.3.1 The Ikaros family of transcription factors

The Ikaros gene family encodes zinc finger transcription factors, comprising Ikaros (also called *IKZF1*), Helios (*IKZF2*), Aiolos (*IKZF3*), Eos (*IKZF4*) and Pegasus (*IKZF5*), which are involved in the development and differentiation of lymphoid cells [133]and play a critical role in regulating cell-fate decisions during haematopoiesis, particularly in the development of the adaptive immune system [134]. The defining feature of the Ikaros family of DNA binding transcription factors is the presence of two distinct and highly conserved C2H2-type zinc-finger domains of Krüppel ancestry. The first set of zinc fingers is found at the amino (N)-terminus of the protein and dictates the ability to bind sequence-specific DNA. The second set of two zinc fingers (ZF5-6) is located at the carboxy (C)-terminus of the protein and confers the ability to form dimers with itself or other members of the family, as well as to interact with other transcriptional regulators[134]. Despite the substantial sequence similarity among Ikaros family proteins, their expression is notably diverse across different cell types and tissues. Ikaros, Helios, and Aiolos predominantly localize to lymphoid cells and their precursors across various developmental stages. Ikaros can also be found in the brain, and both Ikaros and Helios are present in erythroid cells. On the other hand, Eos and Pegasus have a broader expression profile, being present in several tissues, including skeletal muscle, liver, brain, and heart [135]. The Ikaros family of transcription factors modulates gene expression through several mechanisms: firstly, by altering chromatin structure via their association with complexes like the nucleosome remodeling deacetylase (NuRD); secondly, by enhancing the activity of the RNA Polymerase II transcription initiation complex; and thirdly, by facilitating chromosome conformational changes, which involves mediating interactions between distant cis-regulatory regions [136].

The Ikaros gene family is increasingly recognized for its connection with a variety of diseases, from hematological cancers to autoimmune diseases. *IKZF1*, the foremost and extensively researched member, is a key tumor suppressor in B-cell acute lymphoblastic leukemia (ALL)[137]. Its high-frequency deletion in BCR-ABL1–positive ALL, a particularly high-risk subtype, correlates with poorer outcomes in patients with ALL[138]. Several GWASs, including those on SLE [139], inflammatory bowel disease (IBD) [140], Stevens-Johnson syndrome (SJS) [141] and Type I diabetes[142] , have revealed the link between *IKZF1* and susceptibility to autoimmune diseases, indicating the crucial role of *IKZF1* regulation in the maintenance of self-tolerance. Another two members from the Ikaros family, namely *IKZF2* and *IKZF3*, have also been linked to autoimmune diseases in large-scale European GWAS studies [143]. Foxp3(+) Helios (+) Treg, unlike Foxp3(+) Helios(-) Treg, were found to be significantly increased in SLE patients and expanded in active SLE [144]. This suggests that Helios plays a critical role in the suppressive function of Tregs and potentially contributes to the progression of SLE. Meanwhile, *IKZF3*, or Aiolos, predominantly found in mature peripheral B cells, is vital in B-cell differentiation and associated malignancies [133, 145]. It also has been observed that Aiolos controls cell death in T cells by regulating Bcl-2 expression and its cellular localization [146]. The mice lacking Aiolos exhibit symptoms akin to SLE [147]. Additionally, genetic polymorphisms in *IKZF3* are associated with a range of immune-related diseases, extending beyond SLE [148]to conditions like asthma [149], rheumatoid arthritis [150], and ankylosing spondylitis [150].

An additional aspect of disease manifestation related to the Ikaros gene family arises from the imbalance between the functional, wild-type isoforms and the dysfunctional, negative isoforms. Alternative splicing of *IKZF1*, *IKZF2*, and *IKZF3* generates a diverse array of protein isoforms. Canonical isoforms, able to bind DNA, have three or four zinc-fingers within this motif and are able to recognize and bind to the consensus sequence GGAAA as dimers[151]. Isoforms with two or less zinc-fingers act in a dominant negative fashion as they retain the ability to dimerize with canonical isoforms through the C-terminal domain but are unable to mediate DNA interactions [152]. These shorter isoforms function, therefore, in a classical dominant-negative

way, being able to substantially inhibit DNA binding of their dimeric partners [153]. Dysregulated expression of Ikaros family members, particularly splice variants missing the N-terminal DNA-binding zinc fingers, is linked to various hematological malignancies [153, 154]. For example, short isoforms of Ikaros are frequently overexpressed in human leukemia[155, 156]. Similarly, short isoform of Helios has been found to be overexpressed in patients with T-cell acute lymphoblastic leukemia [153, 157]. Moreover, retroviral expression of Helios isoform that lacked the N-terminal DNA-binding domain in hematopoietic progenitor cells of reconstituted mice led to the development of aggressive, transplantable T cell lymphoma[158].

1.3.2 Current Understanding of the function of Helios in human cells

Expression of Helios (*IKZF2*) is detected globally during embryonic hematopoiesis [159]. However, in adults, the expression was largely restricted to the T cell lineage and no expression was observed in mature B-cells, dendritic cells, and myeloid cells [160]. Helios is present at low levels in pro-B cells (CD45R+CD43+from Rag–/–bone marrow) and decreases as they progress to pre-B cells and not significantly expressed in mature B cells [159]. This silencing process of Helios is critical to B cell development. Forced wild-type Helios expression in B cell leads to metastatic lymphoma in Helios transgenic mice [161]. However, analysis of human tumor lines indicates that Helios acts as a tumor suppressor [162] , evidenced by adult T-cell malignancy patients expressing a defective Helios isoform, which leads to T-cell lymphomas when overexpressed in mice [153, 157, 163, 164]. Forced expression of this isoform of Helios in mice has been linked to the development of T-cell lymphomas [158]. It has been demonstrated that Helios is expressed in myeloid leukemia cells and its depletion in acute myeloid leukemia cells results in decreased colony growth and delayed oncogenesis [165]. These findings suggest that Helios appears to act as a tumor suppressor but can be tumorigenic when expressed in inappropriate cell types. Additionally, single cell RNA-seq data from the Human Protein Atlas (HPA) [166] suggest that *IKZF2* is expressed in a variety of cell types beyond blood cells, with higher expression in some neuronal cell types such as oligodendrocyte precursor cells and

astrocytes than immune cells. However, the role of Helios in these neural cells has not been extensively studied.

Studies have been more focusing on determining the role of Helios in T cells. Helios likely plays a significant role in T cell development. Helios' expression peaks in the early stages of T cell development and decreases as T cells mature in the thymus and are exported to the periphery [159]. Before the Helios knockout mouse was formed, over-expression studies found that T-cell differentiation was blocked at the CD4- CD8- stage in the thymus and overexpression of dominant negative isoforms of Helios were found to increase T-cell proliferation [133]. These studies, combined with evidence of up-regulated dominant-negative isoforms in T-cell leukaemias and lymphomas, suggest a role for Helios in restraining T-cell lineage progression [167]. However, Helios knockout mice did not exhibit an obvious T-cell phenotype, implying functional compensation from other Ikaros family members or its specific role in certain T-cell subsets [168]. Interestingly, in a separate investigation, Helios-deficient mice were observed to develop an auto-inflammatory phenotype later in life, resembling rheumatoid arthritis, with an increased number of activated CD4+ and CD8+ T-cells, T-follicular helper cells, germinal center B-cells, and the production of autoantibodies.

Helios may also have a role in T cell activation and signaling. Helios has also been identified as a marker of T-cell activation in Tregs and CD4+ and CD8+ T-cells [169-171]. Furthermore, Helios has been linked to the regulation of cytokine production, including the inhibition of IL-2, IFN-γ, IL-17 and TNFα cytokines in Tregs [172, 173]. Helios exhibits upregulation in CD8+ T cells upon activation, and its presence has been observed in specific CD8+ cell subsets during chronic HIV infection in vivo, indicating a potential role in regulating these populations [169, 174]. Recent studies have provided evidence that Helios may contribute to the promotion of preferential differentiation of human fetal CD4+ naïve T cells into regulatory T cells[175]. Some studies demonstrate that Helios is potentially a specific marker of thymic-derived Treg cells (tTreg)

from peripherally derived Treg(pTreg) [176]. However, this is controversial as many studies did not observe the clear distinction of Treg subsets on the basis of Helios expression. Recently conducted antigen-specific studies demonstrated Helios to be expressed in also iTregs induced in vivo [177].

Studies have unveiled pivotal role of Helios in regulating Treg cell identity and function. Research have revealed that Helios is highly expressed at the mRNA level in Treg cells compared with conventional CD4+ T cells (>10-fold) [168]. Foxp3 is regarded as the major transcription factor for T regulatory (Treg) cells and expression of Foxp3 is used to identify and quantitate Treg cells in mouse models. It was demonstrated that Helios is co-expressed with Foxp3 in 70-75% of murine Foxp3+ T cells and even 85-90% in human Foxp3+ T cells[178]. Helios has been found to bind to the Foxp3 promoter directly to regulate its expression and knocking-down Helios with siRNA oligonucleotides resulted in down-regulation of FoxP3 [179]. It is clear that Foxp3+ Helios+ Tregs possess greater suppressive potential than Foxp3+ Helios- Tregs. The research demonstrated that Helios+ Treg cells exhibited a more highly activated phenotype, greater suppressive activity [179-182], more stable Foxp3 expression[183] and reduced cytokine expression compared to FoxP3+Helios− cells [184, 185]. Furthermore, suppression of Helios message in human Tregs was found to decrease Foxp3 expression and impair its immunosuppressive functions [179]. Epigenetically, Helios+ total Tregs showed more significant demethylation of the Treg-specific demethylation region (TSDR) [181]. Treg-specific demethylation region (TSDR) was found to be restricted to natural Treg [179] and further associated with high and enduring FoxP3 expression upon Treg [184]. Furthermore, loss of Helios expression in nTregs also has been shown to be correlated with TSDR remethylation[186].

1.3.2 Current understanding of the Role of Helios in SLE

The role of Helios in SLE is gaining increasing attention due to its potential involvement in immune regulation. large genome-wide genetic association studies have identified *IKZF2* as a risk locus for SLE [7, 143, 187]. A summary of these associations is presented in the **Table 1. 1**, which is obtained from the GWAS catalog [9], detailing various risk alleles within the *IKZF2* gene that are statistically linked to SLE. Aberrant Helios expression in different T-cell subsets were reported from MPJ and LPR mouse models of lupus relative to C57BL/6 mice that are not prone to lupus[188]. These lupus-prone models show significant changes in their dendritic cell (DC) compartment and unusual Helios levels in several T-cell types, including CD4+ conventional T cells (Tconv), CD4+ regulatory T cells (Treg), CD8+ Treg, and double-negative (DN) T cells, as lupus progresses. Additionally, they showed a memory-like phenotype, increased baseline pSTAT5a levels, and a limited variety in their TCR repertoire. These findings point to Helios as a potentially significant factor in the immune dysregulation associated with lupus, and it may serve as an important marker or therapeutic target for understanding and treating this disease[188]. Remarkably, a high proportion of FoxP3+Helios+ cells have been detected in the circulating CD4+ T-cell population in cases of active SLE, with the prevalence of these cells being closely linked to the level of disease activity [144, 189]. They are thought to possess suppressive activity, as similar to the FoxP3+Helios+Treg cells in healthy individuals, FoxP3+Helios+ Treg cells in SLE do not express cytokines and show a hypomethylated TSDR [144, 189]. Furthermore, FoxP3+Helios+ Treg cells in SLE patients display typical levels of the chemokine receptors CXCR3 and CCR4. This expression facilitates their migration to areas of inflammation[144]. Additionally, they showed a memory-like phenotype, increased baseline levels of phosphorylated Signal Transducer and Activator of Transcription 5a (pSTAT5a, an activated form of the STAT5a protein crucial for cellular signaling), and a limited variety in their T-cell receptor (TCR) repertoire. This suggests that their proliferation in active SLE is likely induced by cytokine signaling and TCR activation, possibly as an immune system strategy to modulate or mitigate the characteristic autoreactive effector responses in SLE[144]. These studies collectively underscore the complex and pivotal role of Helios in modulating immune responses in SLE, marking it as a significant target for future research and potential therapeutic strategies.

Table 1. 1 Association of *IKZF2* Gene Variants with Systemic Lupus Erythematosus Risk as Identified in Genome-Wide Association Studies [9].

Risk Allele	Locations	P value	Mapped genes	AccessionID	PubmedID	Author
rs10048743-G	2:213025508	2.00E-10	*IKZF2*	GCST003156	26502338	Bentham J [143]
rs3768792-C	2:213006985	1.00E-13	*IKZF2*	GCST003155	26502338	Bentham J [143]
rs10048743-?	2:213025508	2.00E-08	*IKZF2*	GCST003622	27399966	Morris DL [7]
rs2371790-A	2:213019942	6.00E-09	*IKZF2*	GCST011956	33272962	Yin X [187]
rs12470231-A	2:213011526	2.00E-08	*IKZF2*	GCST90270940	36750564	Khunsriraksakul C [190]

Each entry in the table lists a unique single nucleotide polymorphism (SNP) identified by an rsID, the risk allele associated with SLE, its chromosomal location on chromosome 2, and the p-value indicating the strength of the association.

1.4 Thesis aims

The primary objective of this thesis is to investigate the biological mechanisms linking *IKZF2* (which encodes Helios) to SLE. Central to this investigation is a thorough functional annotation of Helios binding sites, utilizing ChIP-Seq data derived from Jurkat T-cells. The research expands to encompass a wide range of T-cell subtypes, employing sophisticated multi-omics bioinformatics techniques to gain deeper insights into the role of Helios in the context of SLE, in order to contribute significantly to the understanding of the molecular mechanisms driving SLE pathogenesis.

Chapter 2. Chip-seq binding sites of Helios in Jurkat T cells

2.1 Introduction

Studies have demonstrated that Helios likely plays a significant role in T cell development, activation, and signaling. This chapter focuses on the application of Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) to map the genome binding sites of Helios in Jurkat T cells. Jurkat T cells, being a widely used human T cell model, share many characteristics with primary human T cells, making them a relevant and informative system for studying T cell biology. The findings from the Helios binding sites in Jurkat T cells can offer significant implications for understanding Helios' transcriptional function in human T cells.

The chromatin state of binding sites for a transcription factor plays a crucial role in determining gene expression and cellular functions. Chromatin, the complex of DNA and proteins in the cell nucleus, is not a static structure but rather exhibits dynamic changes that influence gene regulation. Epigenetic modifications, such as DNA methylation and histone modifications, shape the chromatin landscape, either promoting or inhibiting access to DNA by transcription factors and other regulatory proteins. A histone modification is a covalent post-translational modification (PTM) to histone proteins which are often found in recurring combinations at promoters, enhancers and repressed regions. These combinations are referred to as ‘chromatin states’ and can be used to define transcriptionally active/silent regions in genomes. For example, H3K4me1 alone marks primed enhancers, while when present with H3K27ac mark active enhancers. H3K4me3 coupled with a high ratio of H3K4me3 to H3K4me1 are promoter-associated marks.H3K36me3 histone modifications and RNA polymerase (Pol) II ChIP signal correlate with transcribed regions, while H3K27me3 or H3K9me3 is associated with repressive chromatin states [191, 192].

ChromHMM has been recognized for its ability to provide an informative compression of multi-track epigenomic signals into chromatin state sequences, making it a valuable tool for characterizing epigenetic dynamics across multiple human cell types [193]. The algorithm has also been used to explain a large portion of the epigenome through chromatin segmentation, aiming to elucidate the observed epigenomic data as a sequence of hidden chromatin states [194]. This algorithm is based on a multivariate Hidden Markov Model and is used to capture significant combinatorial interactions between different chromatin marks in their spatial context across various epigenomes [195]. The Roadmap Epigenomics project has utilized ChromHMM to interrogate the epigenomes of 127 human tissues and cell types, integrating various techniques such as ChIP-seq, whole genome bisulfite sequencing, DNase hypersensitivity assays, and RNA sequencing [196]. Additionally, the project has included ChromHMM states corresponding to enhancer or promoter elements from the 15-state core model, 18-state model and 25-state model, along with histone modification ChIP-seq peaks and DNase hypersensitivity data peaks [197]. Furthermore, large consortium projects including ENCODE, Blueprint, CEEHRC, Mouse ENCODE, and Drosophila modENCODE have also employed ChromHMM as a basis for their analyses [195]. The use of ChromHMM in the Roadmap Epigenomics project has been instrumental in providing over 300 chromatin state maps for a diverse collection of human tissues, blood lineages, and stem cells [198]. The algorithm has also been applied to all BLUEPRINT samples with full reference epigenome histone modification alignment files, defining 25 epigenetic states [199]. The 15-state ChromHMM model offers a balance between simplicity and detail, making it suitable for studies where a moderate level of chromatin state detail is sufficient. It's particularly useful when broad categorizations are the primary goal, as it simplifies interpretation compared to models with more states. The 25-state ChromHMM model, offering the highest level of detail, is best suited for in-depth studies requiring the most nuanced chromatin state analysis. This model is optimal for research that necessitates very fine distinctions between different chromatin states. However, it is also the most complex to interpret and may require more advanced analysis capabilities. 18-State ChromHMM model provides more granularity than the 15-state model, making it suitable for studies that require a more nuanced understanding of chromatin states. It's a good middle

ground between the simpler 15-state and the more complex 25-state models. The 18-state ChromHMM model is often the preferred choice for many studies [200-202], as it strikes a balance between providing sufficient detail for a nuanced analysis and maintaining a level of simplicity that facilitates interpretation. Expanding beyond the initial scope of the Roadmap project, EpiMap [203] has utilized the Roadmap's 18-state ChromHMM model to provide chromatin state annotations for 833 biosamples. In this chapter, I exploit the capabilities of ChromHMM with 18 states in jurkat T cells from EpiMap [203], to systematically screen all Helios-binding sites in Jurkat T cells. By annotating the functional states of the identified Helios binding sites using ChromHMM, this chapter aims to unravel the complex regulatory landscape surrounding Helios interactions within the context of chromatin.

Exploring the enrichment of Helios binding sites in disease-associated genetic variations holds significant promise for shedding light on its potential involvement in disease pathogenesis. By correlating Helios binding with disease-associated genetic loci, I aim to uncover novel regulatory mechanisms and potential links between Helios and disease phenotypes. The GWAS Catalog [204], a comprehensive database maintained by the National Human Genome Research Institute (NHGRI), serves as a valuable resource for researchers investigating the genetic underpinnings of diseases. It compiles results from numerous GWAS studies and provides a curated list of disease-associated genetic loci. In this chapter, I aim to identify the overlap between Helios binding regions and autoimmune disease-associated genetic variants obtained from the GWAS Catalog and perform enrichment analysis to define the significant overlap, thus providing insights into whether Helios physically interacts with genetic variants that have been implicated in specific autoimmune diseases.

Identifying potential causal variants post-GWAS discovery of risk loci is pivotal in deepening our understanding of the genetic basis of complex diseases and traits, prioritizing potential therapeutic targets, guiding future functional studies, and ultimately translating genetic

discoveries into clinical applications. Traditional approaches like luciferase reporter assays and CRISPR-Cas9 can only assess one genomic region at a time, resulting in limited validation of genetic variants. Massively Parallel Reporter Assays (MPRAs) is a valuable tool that allows the testing of the potential regulatory role of thousands of sequences with unique variants in a single experiment [205]. Lu et al. [206] employed MPRA to screen 3073 GWAS-linked SLE variants at 91 loci in GM12878 cells and Jurkat T cells. 51 variants in the Epstein-Barr virus-transformed B cell line GM12878 and 92 in Jurkat T cells are nominated as putative causal variants due to their demonstration of cell-type-specific allelic enhancer activities in these respective cell lines. In this chapter, I leverage the findings of Lu et al.'s study to explore the intersection of Helios binding sites and these identified SLE putative causal variants in Jurkat T cells. This approach aims to further elucidate the potential involvement of Helios in the progression and development of SLE in T cells.

2.2 Methods

2.2.1 Helios ChIP-Seq data in Jurkat T cells

The *IKZF2* ChIP-Seq data in Jurkat T cells was generated by two previous members of the ImmunoGenetics group, Chris Odhams and Andrea Cortini. Briefly, The Jurkat E6-1 cell-line was purchased from the American Type Culture Collection (ATCC) and cultured at 5% CO2, 37°C in RPMI 1640 medium supplemented with 2mM L-glutamine, 10% FBS, 100units/ml penicillin, 100µg/ml streptomycin, 1.5g/L sodium bicarbonate, 4.5g/L glucose, 10mM HEPES, and 1.0mM sodium pyruvate. ChIP-Seq was were performed according to the ENCODE guidelines [Landt S 2012] using anti-Helios (rabbit polyclonal, 13554-1-AP, Proteintech) and compared with input controls in Jurkat cells. The ChIP was performed in biological duplicates (Replicate 1 and Replicate 2) using the ChIP-IT High Sensitivity Kit (Active Motif). Libraries were created from the purified fragments for both ChIP and input using the NEBNext ChIP-Seq Library Prep Master Mix Set for

Illumina (NEB). Purification and size selection was performed using the AMPure XP Beads (Beckman Coulter). Quality assessment of each library was carried out using the Agilent 2100 Bioanalyzer and the Agilent High Sensitivity DNA Kit. Libraries were sequenced on the Illumina HiSeq 2500 platform with single-end 50bp reads. The Uniform Processing Pipeline for transcription factor ChIP-Seq described by the ENCODE Phase 3 specifications was followed [Landt S 2012]. For all sequence data, reads were mapped to hg19 using BWA 0.7.10 with dynamic read-trimming, a seed length of 32, and a maximum of two mismatches in the seed 'aln -q 5 -l 32 -k 2'. Duplicate, unmapped, low quality, and non-uniquely aligned reads were removed using SAMTools v1.3.1 'view -F 1804'. The strand cross-correlation coefficients were generated using SPP v1.14. The irreproducible discovery rate (IDR) method was implemented to measure consistency between replicates and to determine an optimal cut-off for peak significance [Landt S 2012]. Relaxed peaks were called in narrowPeak format for all true biological replicates, self-pseudo-replicates, pooled replicates, and pooled-pseudo-replicates using SPP v1.14 against the merged input control '-npeak=300000 -savn'. A final IDR threshold of 2% was used taking the conservative set of peaks (derived from IDR of true biological replicates). Blacklisted regions were filtered using the hg19 Mapability Consensus Excludable bed file using bedops v6.1.1 (1bp overlap). Signal tracks depicting control-normalized tag density in the bigWig file format were created using the bdgcmp function of MACS2 v2.1.0 expressed as both fold-over control at each position and as P-value. Data were visualized using the Integrative Genomics Viewer. Narrow Peak files were exported from the MACS2 programme for this project. Sequencing Information is listed in **Table 2.1**.

Table 2. 1 Sequencing Metrics for Helios ChIP-Seq in Jurkat Cells

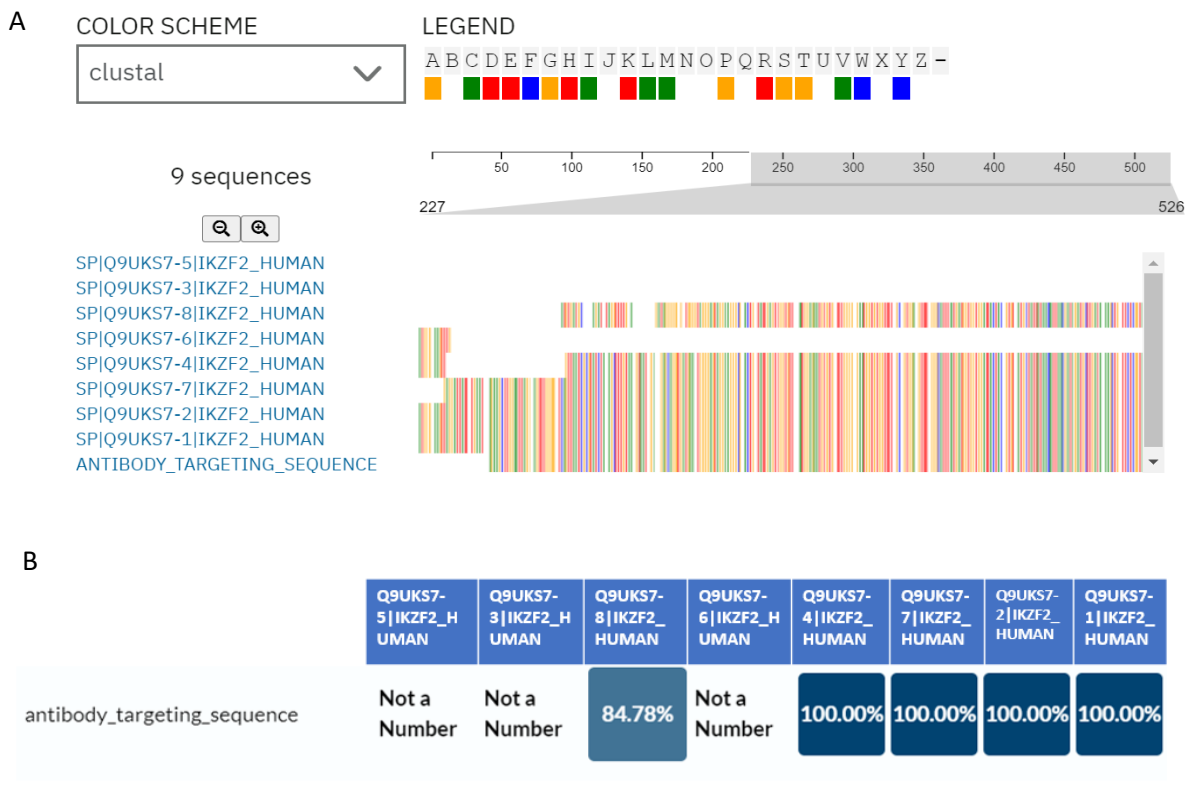
Transcription Factor	Cell Type	Replicate	Number of Mapped Reads	Number of peaks in the optimal set	Number of peaks in the conservative set	Self-consistency ratio	Result of the reproducibility test	Number of peaks from pooled pseudoreplicates	Number of peaks from replicate 1 self-pseudoreplicates	Number of peaks from true replicates	Number of peaks from replicate 2 self-pseudoreplicates
Helios	Jurkat	1	11646633	5237	5068	2.317	borderline	5310	1673	5130	722
Helios	Jurkat	2	8435307								
Control	Jurkat	Unary	60744855	-	-	-	-	-	-	-	-

53

2.2.2 Antibody suitability assessment

The rabbit polyclonal Helios (*IKZF2*) antibody (Catalog #13554-1-AP, Proteintech) is generated using a peptide sequence corresponding to amino acids 227-526 of the Helios protein. Multiple sequence alignment using Clustal Omega [207] demonstrates that the epitope is highly conserved across several isoforms (Q9UKS7-4, Q9UKS7-7, Q9UKS7-2, Q9UKS7-1). However, significant absence of alignment in isoforms Q9UKS7-5, Q9UKS7-3, and Q9UKS7-6 indicate potential differences in antibody binding affinity for these isoforms (**Figure 2. 1**).

Figure 2. 1 Alignment of Helios Protein Isoforms and Antibody Targeting Sequence



A) This figure shows the multiple sequence alignment of eight isoforms of the Helios protein (*IKZF2*) along with the peptide sequence used for generating the Helios antibody for ChIP-seq, using Clustal Omega. The alignment is visualized with the Clustal color scheme to highlight the conservation and differences among the sequences. Regions with consistent colors indicate conserved amino acids among the isoforms and the antibody targeting sequence. B) The percentage identity of the antibody targeting sequence with various isoforms of the Helios protein (*IKZF2*) in humans. Isoforms Q9UKS7-2, Q9UKS7-7, Q9UKS7-4, and Q9UKS7-1 have 100% identity with the antibody targeting sequence, indicating perfect matches. Isoform Q9UKS7-8 has an 84.78% identity, suggesting a high but not complete match. Isoforms Q9UKS7-5, Q9UKS7-3, and Q9UKS7-6 have results listed as "Not a Number," indicating insufficient alignment or data to calculate a meaningful percentage identity.

Immunoprecipitations (IPs) of Ikaros, Helios, and Aiolos from nuclear lysates of GM12878 cells were performed to test antibody specificity. The lysates were incubated overnight with specific antibodies (listed in **Table 2.1**) against Ikaros, Helios, and Aiolos, followed by the addition of Protein G agarose beads to capture the antibody-protein complexes. The beads were washed to remove non-specific binding, and the bound proteins were eluted by boiling in sodium dodecyl sulfate (SDS) sample buffer. The unbound fractions and IP fractions were then analyzed by Western blotting (WB). Proteins were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE), transferred to nitrocellulose membranes, blocked in 5% milk-phosphate buffered saline (PBS) solution, and incubated overnight at 4°C with primary antibodies (α-Ikaros, α-Helios, α-Aiolos). After washing with 0.1% Tween-20 in PBS (Tween-PBS), the membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 hour at room temperature, and the signal was developed using Luminata Crescendo Western HRP substrate and visualized on CL-Xposure film. The house-keeping gene, beta-actin (ACTB), was used as a loading control. Membranes were stripped by submersion in Restore™ Western Blot Stripping Buffer (Thermo Fisher) for 10 minutes at room temperature. The membrane was then re-probed with primary antibodies against Helios (α-Helios) and incubated with HRP-conjugated secondary antibodies,

followed by signal development using Luminata Crescendo Western HRP substrate. ImageJ was used to calculate the density of the bands relative to the loading control.

Table 2. 2 Antibodies used for Immunoprecipitations followed by Western Blot

Target	Source	Molecular Weight	Catalogue	Dilution for WB
Ikaros (*IKZF1*)	Goat Polyclonal (IgG)	50kDa (Calculated) 70kDa (Observed)	Santa Cruz E-20 sc-9861	1:500
Helios (*IKZF2*)	Rabbit Polyclonal (IgG)	57kDa (Calculated) 65kDa (Observed)	proteintech 13554-1-AP	1:500
Aiolos (*IKZF3*)	Rabbit Monoclonal (IgG)	58kDa (Calculated) 58kDa (Observed)	abcam ab139408	1:20,000

2.2.2 Nearest gene annotation of the binding sites of Helios in Jurkat T cell

The ChIPseeker [208] package in R, specifically its annotatePeak function, was employed to assign the nearest genes to Helios binding sites on the hg19 human genome. This function annotates peaks by comparing their midpoints to the transcription start sites (TSS) of genes, which are defined within a default range extending from 3 kilobases (kb) upstream to 3kb downstream of the TSS. To visually represent the characteristics of the binding site annotations, a distribution plot showing the location of transcription factor-binding sites in relation to transcription start sites (TSS), as well as a pie chart that provides an overview of the genomic annotation of these sites were produced.

2.2.3 Chromatin States annotation of the Helios binding sites in Jurkat T cells

The chip-seq peak regions of Helios in Jurkat T cells were intersected using R package IRanges [209] with the expanded 18-state ChromHMM annotation of Jurkat T cells under human genome build hg19 which was downloaded from EpiMap [203]. The ChromHMM 18 states are as follows: TssA (Active TSS), TssFlnkU (Flanking Active TSS Upstream), TssFlnkD (Flanking Active TSS Downstream), TssFlnk (Transcription Start Site Flanking), Tx (Strong transcription), TxWk (Weak transcription), EnhG1 (Genic enhancer 1), EnhG2 (Genic enhancer 2), EnhA1 (Active Enhancer 1), EnhA2 (Active Enhancer 2), EnhWk (Weak Enhancer), ZNF/Rpts (ZNF genes & repeats), Het (Heterochromatin), TssBiv (Bivalent/Poised TSS), EnhBiv (Bivalent Enhancer), ReprPC (Repressed PolyComb), ReprPCWk (Weak Repressed PolyComb), and Quies (Quiescent/Low). The results were visualized using the R package ggplot2.

Table 2. 3 18-expanded ChromHMM states from Roadmap Epigenomic Program [210] **(https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html)**

STATE NO.	MNEMONIC	DESCRIPTION
1	TssA	Active TSS
2	TssFlnk	Flanking TSS
3	TssFlnkU	Flanking TSS Upstream
4	TssFlnkD	Flanking TSS Downstream
5	Tx	Strong transcription
6	TxWk	Weak transcription
7	EnhG1	Genic enhancer1
8	EnhG2	Genic enhancer2
9	EnhA1	Active Enhancer 1
10	EnhA2	Active Enhancer 2
11	EnhWk	Weak Enhancer
12	ZNF/Rpts	ZNF genes & repeats

13	Het	Heterochromatin
14	TssBiv	Bivalent/Poised TSS
15	EnhBiv	Bivalent Enhancer
16	ReprPC	Repressed PolyComb
17	ReprPCWk	Weak Repressed PolyComb
18	Quies	Quiescent/Low

2.2.4 Enrichment of Helios binding sites in disease-associated genetic variations

All associations v1.0 file was downloaded from GWAS catalog [204] (https://www.ebi.ac.uk/gwas/docs/file-downloads). Specific variants linked to Ankylosing Spondylitis, Crohn's Disease, Multiple Sclerosis, Sjögren's Syndrome, SLE, and Rheumatoid Arthritis were extracted based on a significance threshold of $p < 5 \times 10^{-8}$, utilizing R programming. 1000 Genomes Project Phase 3 data was downloaded from 1000 Genome Project (https://www.internationalgenome.org/category/data-access/). SNPS with high LD ($r^2 > 0.8$) with the initial autoimmune disease-associated SNPs identified using 1000 Genomes Project Phase 3 data (GRCh37 build) [211] were also included for analysis, this was conducted using plink v1.9-beta6.10 [212]. The genomic coordinates of disease-associated SNPs (including high LD SNPs) were mapped to Helios binding site regions to detect their overlap using the findOverlapPairs () function of R package IRanges [209]. Overlapping was determined when SNPs fall within the binding site. Subsequently, an enrichment analysis was conducted to evaluate the statistical significance of Helios binding to the SNPs associated with each autoimmune disease using Chi-Square Test and Fisher's Exact Test. Chi-square test was applied to examine the enrichment of Helios binding sites among SNPs linked to autoimmune diseases, comparing this against the binding frequency observed in a reference set of SNPs from the 1000 Genome Project. This step aimed to determine if there was a notable deviation in the binding pattern of Helios to disease-associated SNPs as opposed to random distribution across the

genome. Secondly, recognizing the limitations posed by small sample sizes in some cases, Fisher's Exact Test was also employed.

2.2.5 Overlap of Helios binding sites with the putative causal variants of SLE risk loci in Jurkat T cells

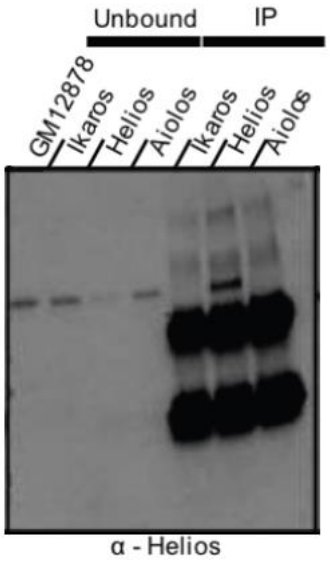
The list of putative causal variants in Jurkat T cells was downloaded from the supplementary data of Lu et al.'s study [206]. I employed the R package IRanges [209] to identify intersections between the genomic coordinates of these expanded list of variants and the Helios binding sites in Jurkat T cells.

2.3 Results

2.3.1 Validation of the Helios antibody used for Chip-seq

The Western blotting after immunoprecipitation shows that in the unbound fractions, bands corresponding to Helios are present, indicating some Helios protein remained unbound during IP. In the IP fractions, a strong band in the Helios IP Lane confirms successful immunoprecipitation of Helios by the α-Helios antibody, demonstrating its specificity and effectiveness. The absence or faint bands in the Ikaros and Aiolos IP lanes indicate that the α-Helios antibody does not significantly cross-react with Ikaros or Aiolos (**Figure 2. 1**). These results validate the α-Helios antibody's high specificity and suitability for ChIP-seq experiment.

Figure 2. 2 Assessment of Helios antibody suitability



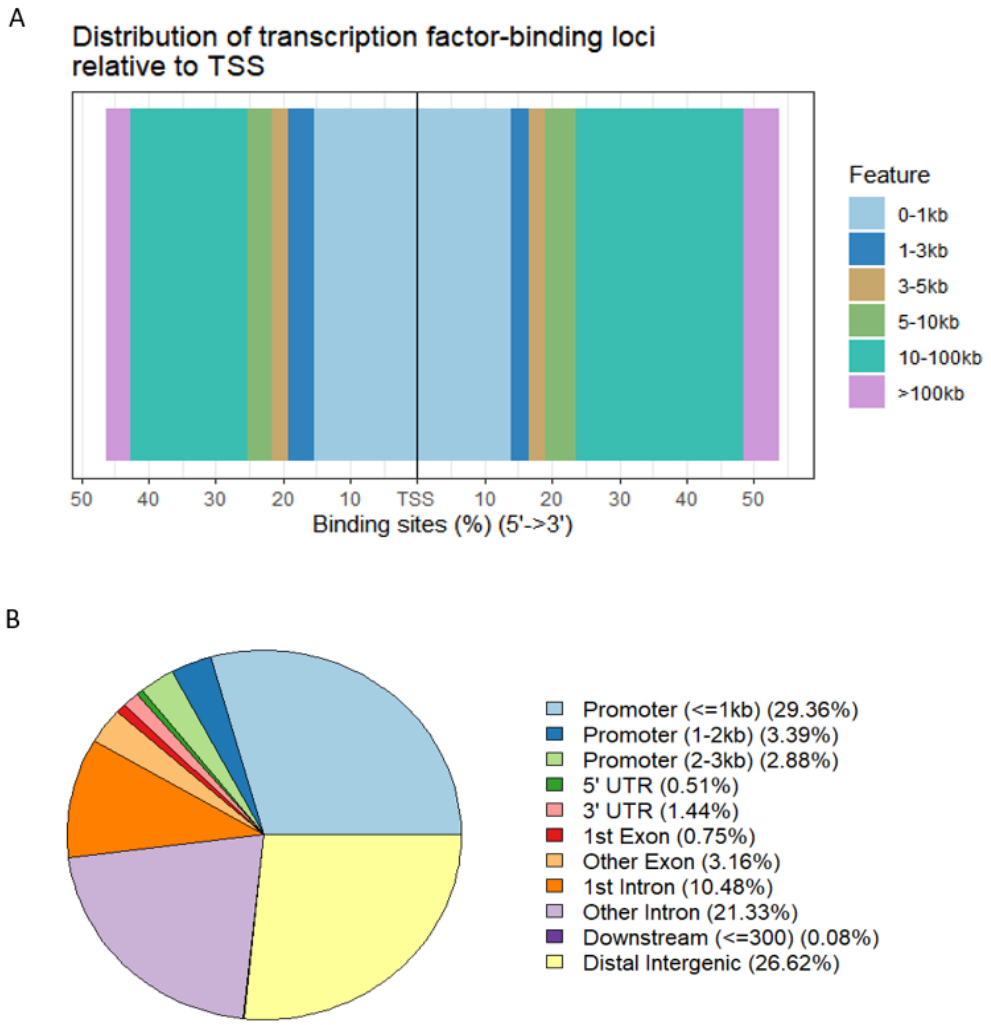
Immunoprecipitations (IPs) of Ikaros, Helios and Aiolos from nuclear lysates of GM12878 cells followed by Western blotting (WB) were performed to test antibodies specificity. The results demonstrate the depletion of Helios from the unbound fraction and its subsequent enrichment in the immunoprecipitated (IP) fraction, confirming the successful IP of Helios by the Helios antibody. There is no significant cross-reactivity observed in the Ikaros and Aiolos IP lanes, indicating the high specificity and efficiency of the Helios antibody used.

2.3.2 Genome-wide discovery of Helios binding sites on Jurkat T cell line

A total of 5,068 ChIP-Seq peaks were identified for Helios in Jurkat T cells. These peaks represent the conservative set of peak-calls, which were consistently found in both biological replicates and passed the 2% IDR (Irreproducible Discovery Rate) threshold. These peaks are mapped to 2937 nearest genes. Feature distribution of the distance from peaks to their nearest genes (**Figure 2.1** A) shows most peaks are centered on 10-100kb or 0-1kb region from the TSS and most peaks are located at the close promoter (defined as $\leqslant$1kb) (29.35%) or distal intergenic (31.69%) region of the annotated nearest genes(**Figure 2.1 B**). The full annotation results of the peaks to the nearest genes are detailed in **Supplementary Table 2.1**

(https://www.dropbox.com/scl/fi/gjkkbejwnc9524xmtjien/Supplementary-Table-2.1.peakanno_jurkatT_HeliosBStonearestgenes.csv?rlkey=weymm7vo1pm6aqbhlfg0unsu1&st=0u5o025v&dl=0).

Figure 2. 3 Genome-wide characterization of Helios Binding Sites in Jurkat T cells



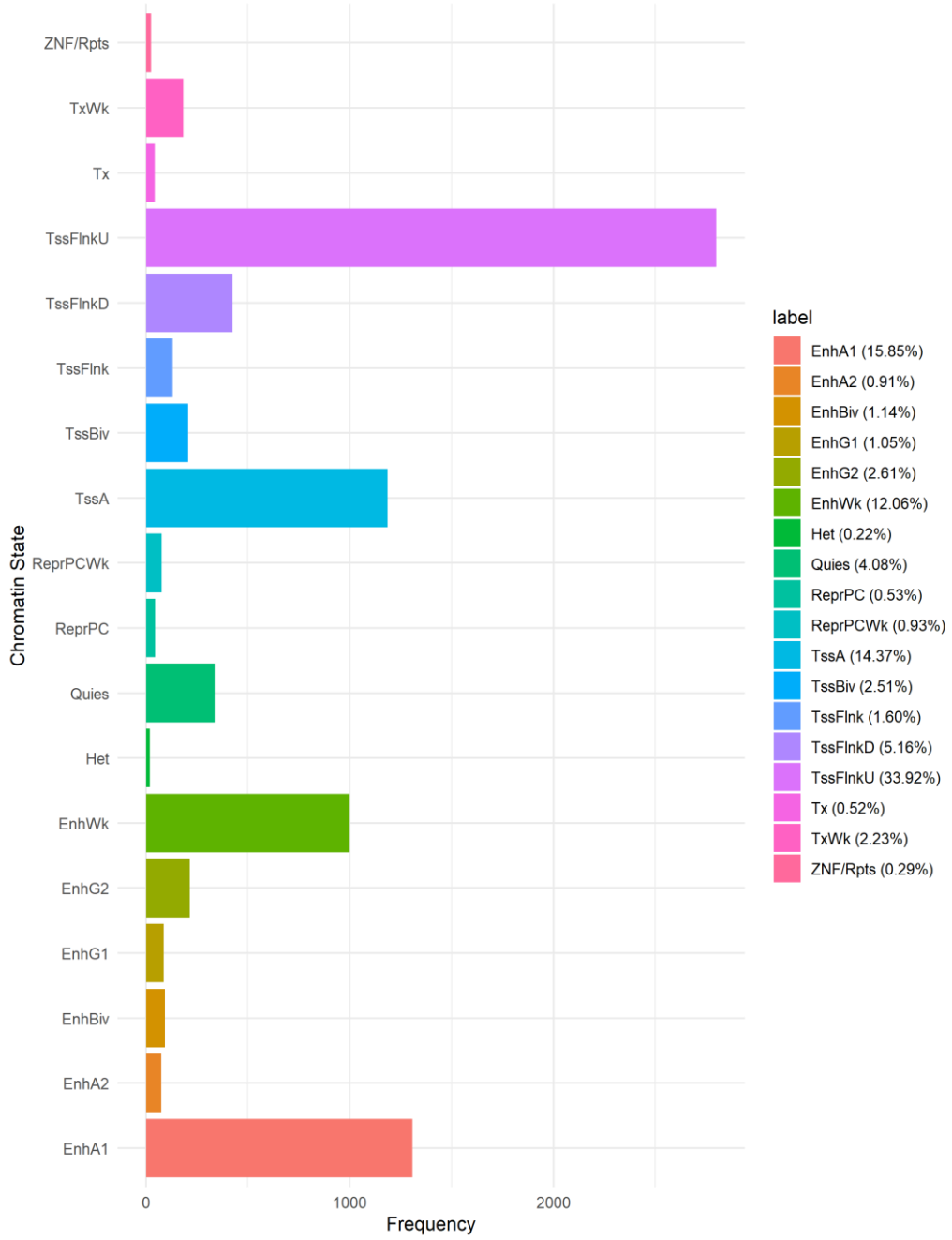
A) Feature distribution of the distance from peaks to TSS of their nearest genes shows most peaks are centered on 10-100kb or 0-1kb region from the TSS. B) Feature distribution of the distance from peaks to their nearest genes shows the most peaks are located at the close promoter (29.36%) or distal intergenic (26.62%) region of the annotated genes.

2.3.2 ChromHMM states of the binding sites

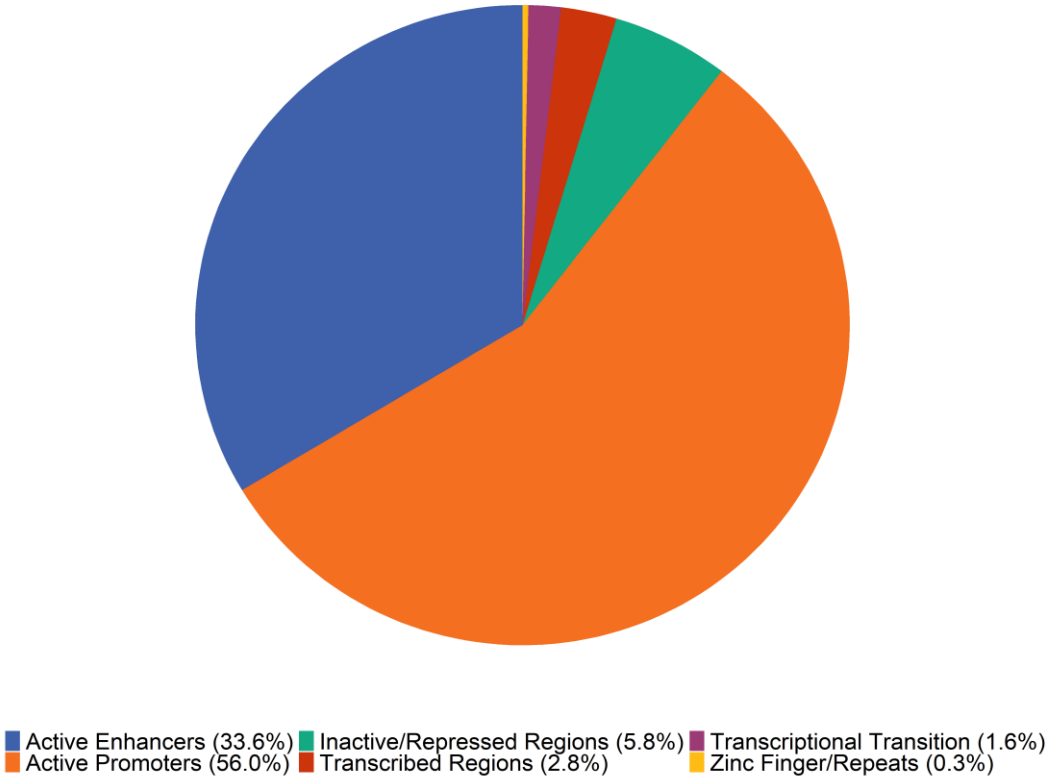
Feature distribution of the expanded ChromHMM 18 states annotated for the binding sites of Helios in Jurkat T cells is shown in **Figure 2.2 A**. Among the annotated chromatin states, the Transcription Start Site Flanking Upstream (TssFlnkU) state constituted the highest proportion (33.89%), followed by Active Enhancer 1 (EnhA1) (15.8%). To provide a more comprehensive overview, we have consolidated the original 18 states into six main categories: Active Promoters (TssA, TssFlnkU, TssFlnkD, TssBiv), Active Enhancers (EnhA1, EnhA2, EnhG1, EnhG2, EnhWk, EnhBiv), Inactive/Repressed Regions (Quies, ReprPC, ReprPCWk, Het), Transcriptional Transition (TssFlnk), Transcribed Regions (TxWk, Tx), and Zinc Finger/Repeats (ZNF/Rpts). The percentages of occurrence within each of these broader categories have been computed. As evident from **Figure 2. 2 B**, the highest proportion of the states falls within the category of Active Promoters (56 %), with Active Enhancers comprising the second largest segment (33.6%).

Figure 2. 4 Feature distribution of the expanded ChromHMM 18 states annotated for the binding sites of Helios in Jurkat T cells

A



B



A) Distribution of the 18 ChromHMM states annotated for the binding sites of Helios in Jurkat T cells. B) Six main categories of the chromHMM state annotation of binding sites of Helios in Jurkat T cells. The eighteen states have been classified into six principal categories: Active Promoters (TssA, TssFlnkU, TssFlnkD, TssBiv), Active Enhancers (EnhA1, EnhA2, EnhG1, EnhG2, EnhWk, EnhBiv), Inactive/Repressed Regions (Quies, ReprPC, ReprPCWk, Het), Transcriptional Transition (TssFlnk), Transcribed Regions (TxWk, Tx), and Zinc Finger/Repeats (ZNF/Rpts). The percentages of occurrence within each of these broader categories have been computed and visualized in this figure.

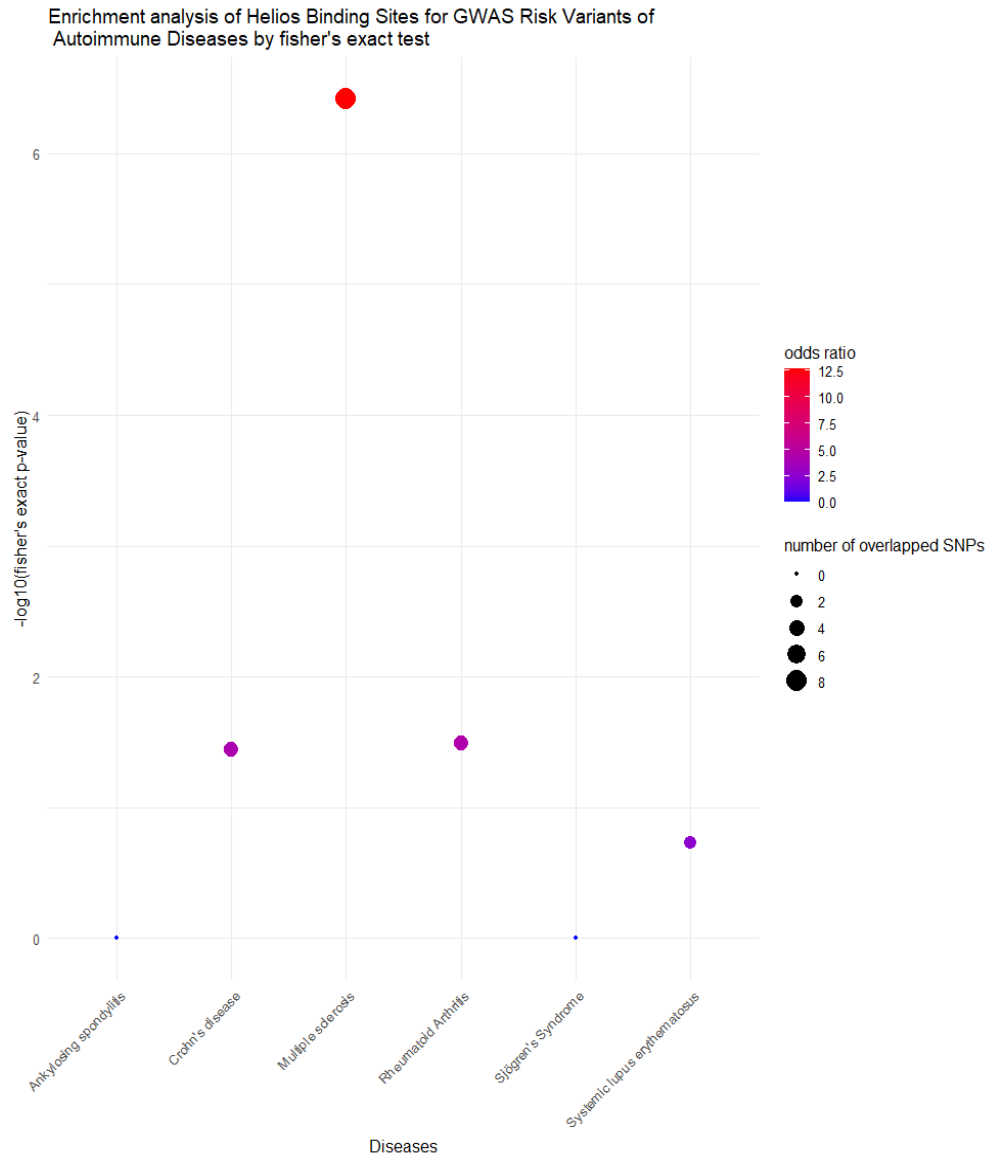
2.3.3 Enrichment of Helios binding sites in autoimmune diseases associated genetic variations

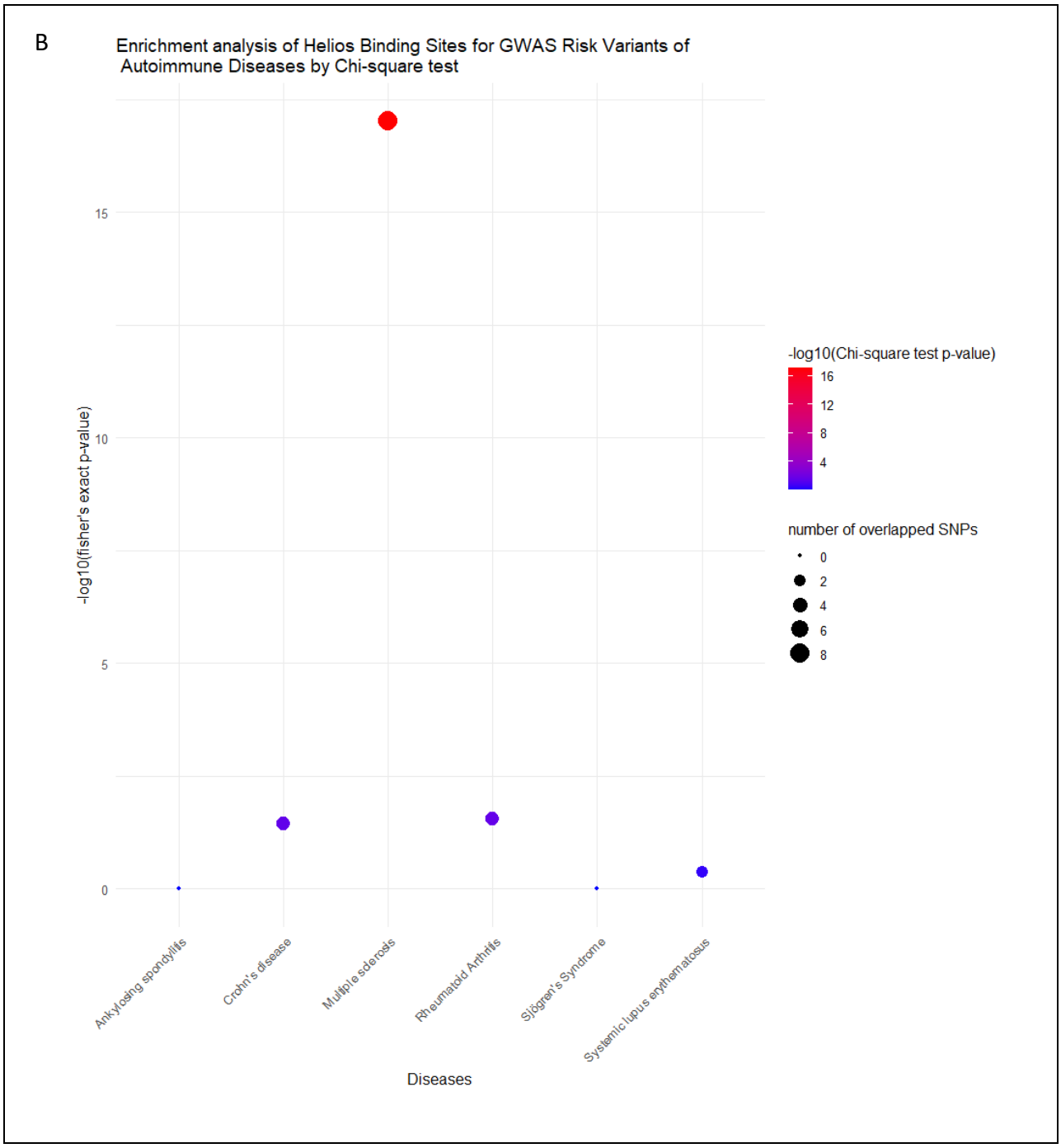
The total count of genetic variants, including both the initial SNPs and the supplementary SNPs exhibiting high linkage disequilibrium (LD) with the initial SNPs, associated with each respective disease are as follows: 94 in Ankylosing Spondylitis, 742 in Crohn's Disease, 655 in Multiple Sclerosis, 718 in Rheumatoid Arthritis, 33 in Sjögren's Syndrome, and 828 in Systemic Lupus Erythematosus (**Table 2.2**). The number of the SNPs that overlap with the binding sites of Helios in Jurkat T cells are: 0 in Ankylosing Spondylitis, 3 in Crohn's Disease, 8 in Multiple Sclerosis, 3 in Rheumatoid Arthritis, 0 in Sjögren's Syndrome, and 2 in Systemic Lupus Erythematosus (**Table 2.3**). The genomic coordinates for these SNPs were ascertained through the getBM() function from the BiomaRt R package, with some SNPs omitted due to unavailable positional data. The remaining SNPs (**Table 2.3**) were analyzed for their location within Helios binding sites using the IRanges package. The statistical evaluations, employing both Fisher's exact test and the Chi-square test, showed a high degree of agreement and identified significant overlaps for Crohn's Disease, Multiple Sclerosis, and Rheumatoid Arthritis (**Table 2.3, Figure 2.3)**. The risk variants that overlap with Helios binding sites for each disease are listed in **Table 2.4**. Multiple Sclerosis had the highest number of overlapping SNPs (8) with Helios binding sites, a highly significant Chi-square p-value (9.70E-18), a significant Fisher's exact test p-value (3.83E-07), and the largest odds ratio (12.6363), suggesting a strong association in this context.

For SLE, the two SNPs found to coincide with Helios binding sites are rs11185603 (on chromosome 7 at position 50306810) and rs7329174 (on chromosome 13 at position 41558110), based on the hg19 human genome reference (**Table 2.4**). While the Fisher's exact test presents an odds ratio of 2.5414, indicating a moderate association with Helios binding sites, the Chi-square test and Fisher's exact test p-values, being 0.4224 and 0.1869 respectively, did not reach the conventional threshold for statistical significance, typically set at $p < 0.05$. Therefore, despite the observed moderate association, the link between Helios binding sites and SLE-associated SNPs does not appear to be statistically significant in this analysis.

Figure 2. 5 Enrichment of Helios binding sites for the GWAS risk variants of Autoimmune Diseases

A




B
Enrichment analysis of Helios Binding Sites for GWAS Risk Variants of Autoimmune Diseases by Chi-square test
-log10(fisher's exact p-value)
15
10
5
0
Ankylosing spondylitis
Crohn's disease
Multiple sclerosis
Rheumatoid Arthritis
Sjögren's Syndrome
Systemic lupus erythematosus
Diseases
-log10(Chi-square test p-value)
16
12
8
4
number of overlapped SNPs
0
2
4
6
8

A) The figure represents an enrichment analysis of Helios Binding Sites for GWAS Risk Variants of Autoimmune Diseases using Fisher's exact test. The y-axis shows the negative logarithm (base 10) of the Fisher's exact p-value, indicating the significance of enrichment, with higher values indicating more significant enrichment. The x-axis lists various autoimmune diseases being analyzed. Dot size represents the number of overlapping SNPs, while the color gradient from blue to red indicates the odds ratio, with red representing higher odds ratios.

B) The figure represents an enrichment analysis of Helios Binding Sites for GWAS Risk Variants of Autoimmune Diseases using Chi-square test. The y-axis shows the negative logarithm (base 10) of the Fisher's exact p-value, indicating the significance of enrichment, with higher values indicating more significant enrichment. The x-axis lists various autoimmune diseases being analyzed. The color scale on the right side indicates the range of p-values, with the color intensity increasing with the significance of the enrichment (i.e., lower p-values).The size of the dots represents the odds ratio from the Fisher's exact test, with larger dots indicating a higher odds ratio, suggesting a stronger association between Helios Binding Sites and the risk variants for the diseases.

Table 2. 3 The number of SNPs associated with each of the autoimmune diseases

Autoimmune diseases	Initial diseases associated SNPs from GWAScatalog(p<5e-8)	Initial SNPs+ SNPswithhighLD with initial SNPs
Ankylosing spondylitis	44	94
Crohn's disease	308	742
Multiple sclerosis	337	655
Rheumatoid Arthritis	421	718
Sjögren's Syndrome	22	33
Systemic lupus erythematosus	489	828

Table 2. 4 The results of enrichment analysis of Helios binding sites for GWAS risk variants of autoimmune diseases

Diseases	Number of	Number of SNPs	Chi-square	Fisher’s	Fisher’s

	SNPs Tested for overlap	overlap with Helios binding sites	test Pvalue	exact test Pvalue	exact test odds ratio
Ankylosing Spondylitis	93	0	1	1	0
Crohn's disease	737	3	0.0348	0.0359	4.2053
Multiple Sclerosis	654	8	9.70E-18	3.83E-07	12.6363
Rheumatoid Arthritis	706	3	0.0283	0.0323	4.3904
Sjogren's Syndrome	30	0	1	1	0
Systemic lupus erythematosus	813	2	0.4224	0.1869	2.5414

Table 2. 5 The list of risk variants of each disease overlap with Helios binding sites

Disease	Resfsnp ID	Chromosome	Position_hg19
Multiple Sclerosis	rs9909593	chr17	37970149
Multiple Sclerosis	rs4812772	chr20	42579051
Multiple Sclerosis	rs4812773	chr20	42579148
Multiple Sclerosis	rs4245080	chr11	128421586
Multiple Sclerosis	rs4262739	chr11	128421175
Multiple Sclerosis	rs4939489	chr11	60793648
Multiple Sclerosis	rs4939490	chr11	60793651
Multiple Sclerosis	rs4939491	chr11	60793722
Crohn's disease	rs9656588	chr7	50306780
Crohn's disease	rs3792112	chr2	234176609
Crohn's disease	rs7329174	chr13	41558110
Rheumatoid Arthritis	rs10117059	chr9	123653477
Rheumatoid Arthritis	rs968567	chr11	61595564
Rheumatoid Arthritis	rs12889006	chr14	69260563
Systemic Lupus Erythematosus	rs11185603	chr7	50306810
Systemic Lupus Erythematosus	rs7329174	chr13	41558110

2.3.4 Overlap of Helios binding sites with the putative causal variants of SLE risk loci in Jurkat T cells

None of the set of 92 putative causal variants for SLE in Jurkat T cells was identified as being bound by the transcription factor Helios in Jurkat T cells.

2.4 Discussion

The genome-wide chip-seq analysis found 5,068 binding sites for Helios in Jurkat T cells, mapping to 2937 nearest genes, indicating the substantial impact of Helios on gene regulatory mechanisms within T cells. The 29.35% of peaks found in the close promoter region underline its direct regulatory role at the site of transcription initiation. Conversely, the 25.96% of peaks located in distal intergenic regions point towards a potential role in long-range gene regulation. These findings are consistent with the results of expanded 18-state ChromHMM annotation of the binding sites, which shows the highest proportion of the states falls within the category of Active Promoters (55.9%), with Active Enhancers comprising the second largest segment (33.7%). Collectively, these findings imply Helios' multifaceted role in gene regulation, encompassing both proximal transcription initiation and distal, enhancer-mediated regulatory functions.

The overlap of Single Nucleotide Polymorphisms (SNPs) with Helios binding sites in Jurkat T cells varies across different diseases: none in Ankylosing Spondylitis and Sjögren's Syndrome, three in Crohn's Disease, eight in Multiple Sclerosis, three in Rheumatoid Arthritis, and two in Systemic Lupus Erythematosus. Notably, statistical analysis using two-sided Fisher’s exact tests indicates a significant overlap ($p < 0.05$) of Helios binding sites with genetic variants in Crohn's Disease, Multiple Sclerosis, and Rheumatoid Arthritis. These findings suggest a potential association between Helios binding sites and genetic variants in these specific diseases, highlighting the relevance of Helios in the context of immune-related disorders and indicate disease-specific differences in the regulatory impact of Helios on T-cell function and immune responses.

While Helios is not found to bind to any of the putative causal variants of SLE in Jurkat T cells identified by Lu et al.[206], it is found to bind to two risk variants associated with SLE:

rs11185603 and rs7329174. The risk locus containing rs7329174 and rs57668933 are long been identified as SLE risk locus with rs7329174 showing the most significant association with the disease[213]. The nearest gene of this locus is the gene *ELF1*. Both of rs7329174 and rs57668933 are identified as eQTL for *ELF1* in multiple blood cells including B cells and T cells [214, 215]. Additionally, using CRISPR/Cas9, Fazel-Najafabadi M et al. validated rs57668933 as a functional variant regulating *ELF1*, in B-cells. The binding of Helios at SNP rs7329174 suggests its role in SLE development as an upstream regulator of the gene *ELF1*. The *ELF1* gene, also known as E74-like factor 1, has been implicated in various biological processes, particularly in the context of the immune system. *ELF1* is a member of the ETS family of transcription factors and is known to be involved in T cell development and function. It has been linked to the regulation of antibody heavy chain production in B-cells [216], suggesting its involvement in immune responses. Studies also suggest that decreased levels of DNA-binding *ELF1* found in SLE T cells could explain decreased expression of CD3ζ chain and increased expression of FcRg[217], which compromise the significant alterations in T cell signaling mechanisms in SLE.

Another risk variant of SLE that Helios binds to is rs11185603, of which the nearest gene is *IKZF1*. rs11185603 is also identified as eQTL for *IKZF1* in multiple blood cells including B cells and T cells [214, 215]. In addition, the SNP rs11185603 has been identified to interact with the promoter of the *IKZF1* gene and deletion of the region containing the rs4385425 proxy ($r2 = 0.99$) to rs11185603 induced expression of Ikaros (encoded by *IKZF1*) in nearly half of the Jurkat cells. These findings suggest the significant impact of rs11185603 on *IKZF1* expression and regulation. Also, the finding that Helios binds to this SNP is consistent with the knowledge that the members in Ikaros family interact with each other through homodimer and heterodimer formations. [218] *IKZF1* can form homodimers (e.g., *IKZF1*/*IKZF1*) as well as heterodimers with *IKZF2* (e.g., *IKZF1*/*IKZF2*). This ability to form both homodimers and heterodimers adds to the functional complexity of these transcription factors, significantly influencing their role in the regulation of immune responses and development of immune cells.

Chapter 3. Differential Expression Induced by Helios Knockdown in Jurkat T Cells

3.1 Introduction

The investigation of gene function in conjunction with gene knockdown or overexpression, represents a powerful approach to unraveling the molecular mechanisms underlying various biological processes. This integrated methodology allows for the comprehensive assessment of gene expression changes and their functional implications, providing valuable insights into the regulatory roles of specific genes in diverse cellular contexts. Over the past decade, gene expression profiling has evolved remarkably, transitioning from analyses focused on single genes or small gene clusters to comprehensive examinations at a global scale, thanks to advancements in high-throughput technologies. Microarray and RNA sequencing (RNA-seq) have emerged as the predominant methods for whole transcriptome profiling. DNA microarray technology, developed in the 1990s, involves hybridizing DNA fragments (such as oligonucleotides, genomic fragments, or cDNAs) converted from messenger RNAs of a sample to probes fixed on a solid matrix. The hybridization is quantified, typically using fluorescence-based detection systems. Although once the leading method for wide-scale gene expression analysis, the popularity of microarrays has waned in favor of RNA-seq, which offers greater resolution and sensitivity, particularly for low-abundance transcripts [219].

RNA-seq boasts several advantages over microarrays, such as not requiring a reference genome or pre-labeled probes, and its ability to generate data without prior knowledge of the genome sequence. However, it is important to recognize that microarrays have been a reliable source of valuable data and remain relevant in certain contexts. Given the high cost and time-intensive nature of RNA-seq and microarray experiments, we sought to leverage publicly available gene expression data to investigate the impact of Helios modulation in Jurkat T cells. By employing a targeted search strategy using keywords "RNA-seq OR microarray/gene expression" AND "knockdown/knockout/overexpression" AND "Helios/*IKZF2*" AND "Jurkat T cells," we identified a relevant microarray dataset [167]. This dataset specifically encompasses gene expression profiles associated with Helios knockdown using shRNA and Helios overexpression via an ATL-type expression plasmid, in comparison to wild-type Jurkat T cells. While the dataset encompasses both overexpression and knockdown profiles of Helios in Jurkat T cells, this study focuses on comparing the transcriptional profiles of cells with Helios knockdown to their wild-type equivalents. This decision is underpinned by the rationale that knockdown experiments more closely replicate loss-of-function mutations, offering a more physiologically pertinent insight into a gene's natural function.

Functional annotation and enrichment analysis play a pivotal role in gaining comprehensive insights into the genes driving pathophysiological mechanisms and providing a systemic view for interpreting data. Functional annotation involves assigning biological information to genomic elements, while functional enrichment analysis determines over-represented gene functional categories or associations with diseases in a large set of genes. In the context of investigating the impact of Helios knockdown in Jurkat T cells, functional annotation analysis, including Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, WikiPathways and Reactome pathways. The Gene Ontology (GO) [220] is a widely recognized bioinformatics initiative that unifies the representation of gene and gene product attributes across all species. It encompasses three main categories: molecular function (MF) of gene products, biological processes (BP) in which those actions occur, and the cellular

component (CC) where the genes are present. KEGG [221] is a database resource that provides insights into advanced functions and utilities of biological systems at the molecular level. WikiPathways [222] is a collaborative and open science biological pathway database known for its community-driven nature. Reactome [223] is a comprehensive bioinformatics database that provides curated information on reactions, pathways, and biological processes. Metascape [224] is used in this chapter to detect enrichment of these pathways in the DEGs between Helios wild-type and knockdown in Jurkat T cells. Metascape is a powerful web-based portal designed to provide a comprehensive gene list annotation. It combines functional enrichment, interactome analysis, gene annotation, and membership search to leverage over 40 independent knowledgebases within one integrated portal. Additionally, it facilitates comparative analyses of datasets across multiple independent and orthogonal experiments by merging clusters based on their membership similarities.

3.2 Methods

3.2.1 Data access of differentially expressed genes between Helios knockdown vs.WT-Helios jurkat T cell

Gene expression profile of Wild type Helios and Helios knockdown in Jurkat cells were downloaded from Gene Expression Omnibus (GEO) database of National Center for Biotechnology Information (NCBI) (accession number GSE41796). The raw microarray data in the format of “txt.gz” files of WT-Helios/Helios knockdown in Jurkat cells (GSM1024425,GSM1024426,GSM1024427,GSM1024428,GSM1024429,GSM1024430) were processed.

3.2.2 Differential expression analysis between Helios wild-type and Helios knockdown in Jurkat T cells

The R/Bioconductor software package limma [225] was used for background correction, normalization and differential expression analysis was done between Helios knockdown group and wild type group using limma package according to the users' guide. This returns an adjusted p-value and log2FC for differential expression between Helios knockdown and control groups using an empirical Bayes method. Considering the technical characteristics of microarray data, including its more limited dynamic range and sensitivity compared to RNA-seq, we adopted a less stringent criterion of log2FC > 0.585 (corresponding to a 1.5-fold change) and an adjusted p-value < 0.05 to identify differentially expressed genes. R package biomaRt 2.42.0 [226] were used to convert probeIDs to genesymbols. The expression level and distribution of DEGs were visualized using a volcano plot by the ggplot2 package (versions 3.4.0) [227], respectively.

3.2.3 Functional annotation of the DEGs between Helios knockdown and wild type in Jurkat T cells

To better understand the biological functions of the DEGs, Gene ontology Biological Process(GO_BP), Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway, Wikipathway and Reactome pathway analysis of differentially expressed genes were conducted using the web tool Metascape [224] (https://metascape.org/gp/index.html#/main/step1). Metascape provided pathway and process enrichment analysis utilizing KEGG Pathway, GO Biological Processes, Reactome Gene Sets, and WikiPathways, with the entire genome as the enrichment background. Terms meeting criteria of q-value < 0.05, a minimum of three overlapping genes, and an enrichment factor > 1.5 were clustered based on similarity, employing kappa scores for hierarchical clustering. Clusters with a similarity > 0.3 were recognized, with the most

significant term in each cluster representing it. P-values were calculated via the cumulative hypergeometric distribution, and q-values were adjusted using the Benjamini-Hochberg procedure for multiple testing [228]. The summarized pathways were visualizad using R package ggplot2.

3.2.4 Network analyis of the DEGs between wild-type and knockdown of Helios in Jurkat T cells

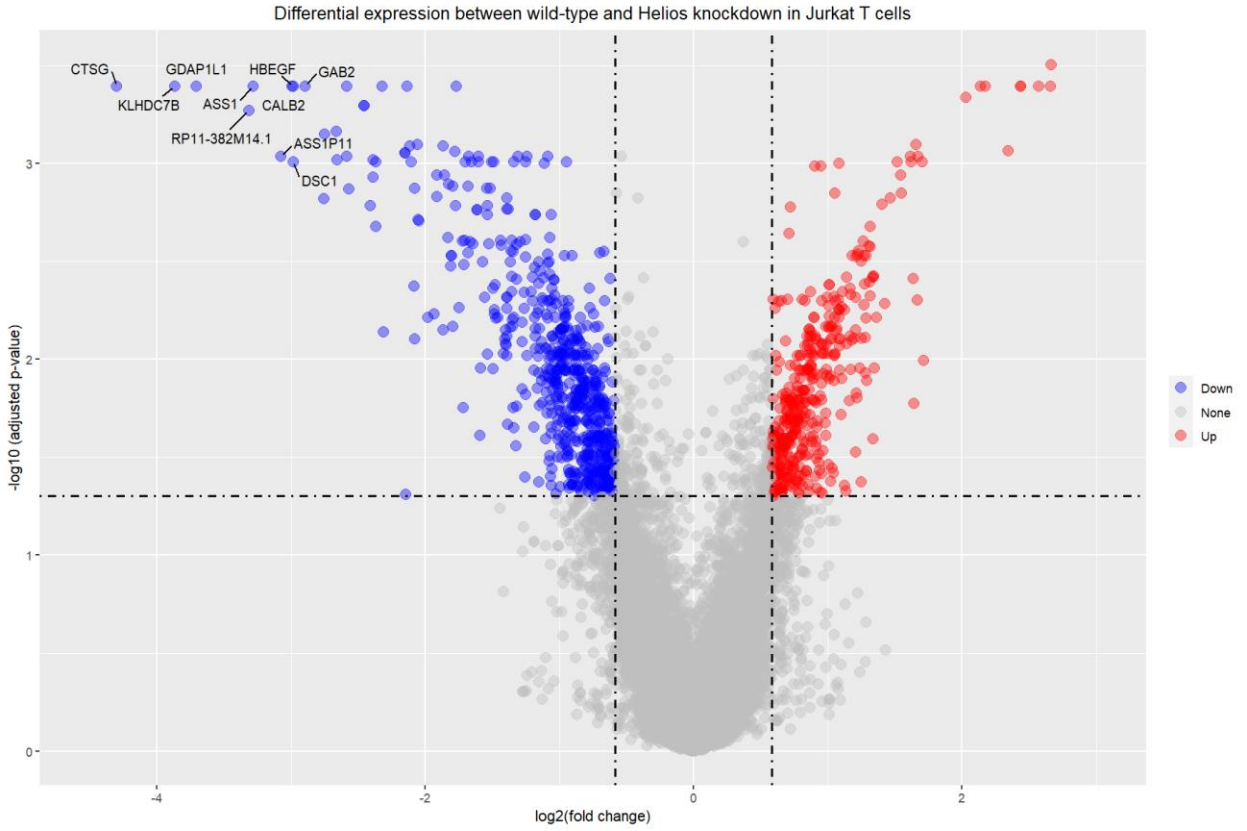
The DEGs were submitted to the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) 12.0 database to build the protein–protein interactive (PPI) network, employing a medium confidence threshold of 0.400 and a false discovery rate (FDR) of 0.05. The result was visualized using Cytoscape 3.7.2. Subsequently, we conducted module analysis using the MCODE plugin within Cytoscape 3.7.2, with a cutoff MCODE score of >2. MCODE plugin in Cytoscape was applied to identify modules with the criteria including degree cutoff of 2, node score cutoff of 0.2, k-core of 2, and maximum depth of 100. Additionally, we identified the top hub genes within each module using the CytoHubba plugin in Cytoscape 3.7.2, employing the Maximal Clique Centrality (MCC) algorithm. Metascape (https://metascape.org/gp/index.html#/main/step1) was used to conduct Functional annotation for the modules that have more than 5 nodes. The analyses included KEGG Pathway, GO Biological Processes, Reactome Gene Sets, and WikiPathways, with the entire homo sapien genome as the enrichment background. Terms meeting criteria of q-value < 0.05, a minimum of three overlapping genes, and an enrichment factor > 1.5 were clustered based on similarity, employing kappa scores for hierarchical clustering. Clusters with a similarity > 0.3 were recognized, with the most significant term in each cluster representing it. P-values were calculated via the cumulative hypergeometric distribution, and q-values were adjusted using the Benjamini-Hochberg procedure for multiple testing. The summarized pathways were visualized using R package ggplot2.

3.3 Results

3.3.1 DEGs between Helios wild-type and Helios knockdown in Jurkat T cells

A total of 1072 genes exhibited significant differential expression between Helios knockdown and WT-Helios Jurkat T cells, as determined by an adjusted p-value threshold of less than 0.05 and |log2FC|>0.585 (refer **Supplementary Table 3.1** https://www.dropbox.com/scl/fi/3whix1s9tz6wtosb2pq74/Supplementary-Table-3.1.1072DEGs.txt?rlkey=p2lt39mqg6eg8ipy4ycsxktb3&st=8exu9yk3&dl=0). Among these DEGs, 47 demonstrated a |log2FC| greater than 2.0, while 358 exhibited a |log2FC| greater than 1.0. The top 10 genes characterized by the highest |log2FC| values are as follows: *CTSG* (Cathepsin G), *KLHDC7B* (Kelch Domain-Containing Protein 7B), *GDAP1L1*(Ganglioside-Induced Differentiation-Associated Protein 1-Like 1), *RP11-382M14.1, ASS1* (argininosuccinate synthase 1), *ASS1P11* (Argininosuccinate Synthetase 1 Pseudogene 11), *CALB2* (Calbindin 2), *HBEGF* (Heparin Binding EGF Like Growth Factor), *DSC1* (Desmocollin 1) (**Figure 3.1**).

Figure 3. 1 Volcano plot illustrating the differential expression between wild-type and Helios knockdown in Jurkat T cells



Each point represents a gene, with red points indicating upregulated genes (log2FC > 0.585 & adjusted p value<0.05), blue points indicating downregulated genes (log2FC < -0.585 & adjusted p value<0.05), and gray points representing non-significant genes (adjusted P-value $\geqslant$ 0.05 or absolute log2 fold change $\leqslant$ 0.585).

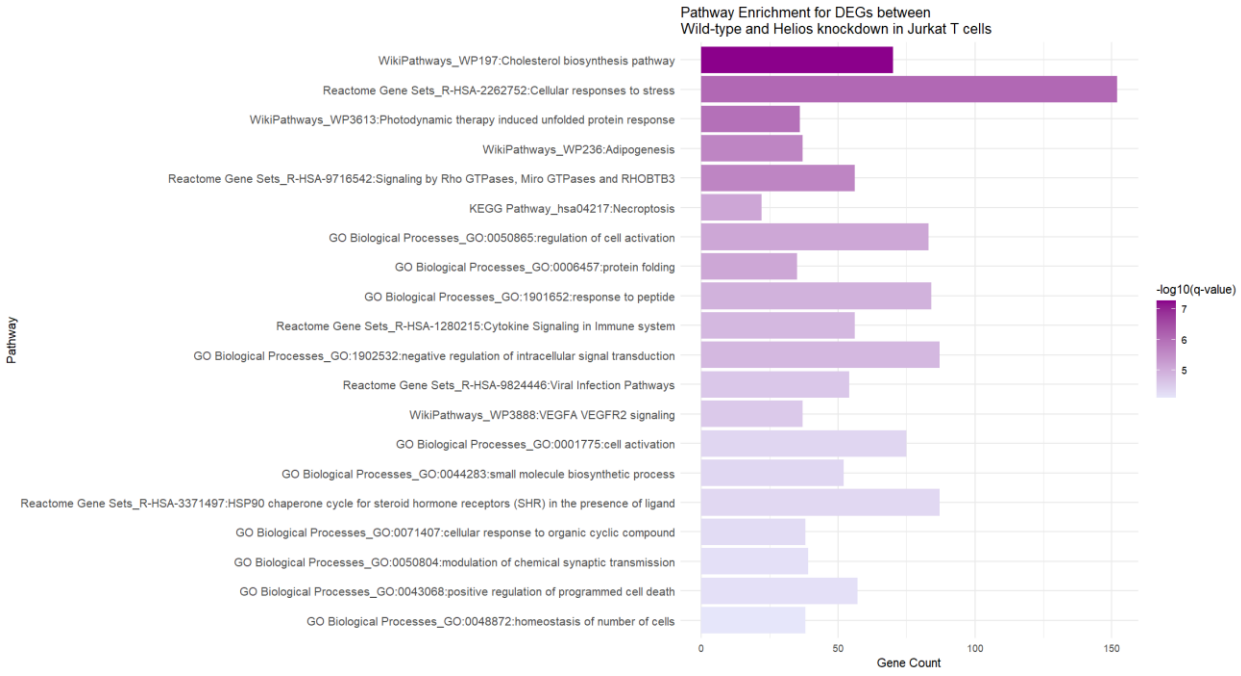
3.3.2 Change of expression of the genes *ELF1* and *IKZF1* between Helios knockdown and Helios wild type in Jurkat T cells

The gene *ELF1* didn't show significant change of expression between Helios knockdown and wild-type groups (log2FC = 0.24, adjusted p value > 0.05). *IKZF1* shows significant change of expression between two groups (log2FC =0.65, adjusted p value = 0.037) (**Supplementary Table 3.1**).

3.3.3 Functional annotation of the DEGs

The pathway enrichment analysis for the differential expressed genes between wild-type and Helios-knockdown Jurkat T cells reveals a significant interplay of metabolic and immune-regulatory pathways. These include the regulation of cholesterol biosynthesis, cellular stress responses, Rho GTPase-mediated signaling, and the unfolded protein response. Additionally, key immune system pathways such as Cytokine Signaling in Immune system and Viral Infection Pathways are highlighted (**Figure 3.2**).

Figure 3. 2 Functional annotation of the DEGs between wild-type and Helios-knockdown Jurkat T cells



The color gradient represents the -log10 transformed q-values from Metascape, with darker shades of red indicating higher significance levels. The bars depict the number of genes associated with each pathway, with pathways related to cholesterol biosynthesis, cellular responses to stress, and immune signaling pathways, including cytokine signaling and viral infection responses, being most prominently enriched.

3.3.4 PPI network analysis and modules identification for the DEGs

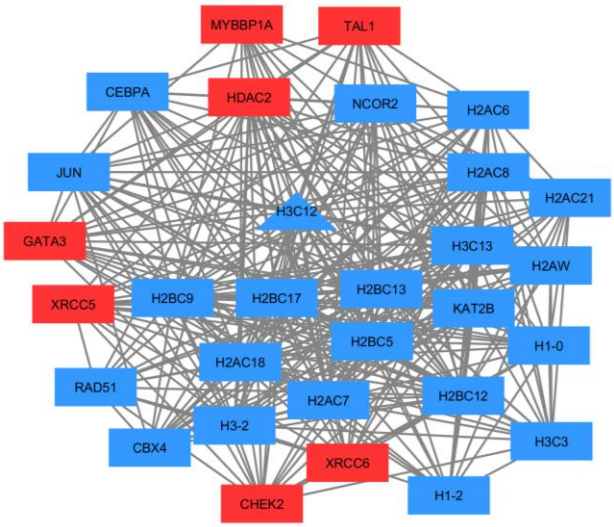
PPI analysis of the prioritized target genes of Helios in Jurkat T cells was based on the STRING database and the results were visualized using Cytoscape. After removing singleton nodes, a network with 698 nodes and 3683 edges. Ten modules were found by MCODE plugin in Cytoscape among the prioritized target genes with module identification criteria including degree cutoff of 2, node score cutoff of 0.2, k-core of 2, maximum depth of 100, and nodes>5. Metascape (https://metascape.org/gp/index.html#/main/step1) was used to conduct Functional annotation for the modules.

Module 1 has 30 nodes and 327 edges, of which the hub gene is *H3C12* (H3 Clustered Histone 12). The pathway enrichment analysis of this module points to significant activity in histone deacetylation, chromatin organization, histone arginine methylation, and chromosome organization. These processes are crucial for gene expression regulation and indicate Helios' role in epigenetic modulation. Additionally, pathways such as the Notch-HLH transcription pathway and ATM signaling are implicated, suggesting alterations in cell fate decisions and DNA damage responses. The results underscore the comprehensive impact of Helios on cellular mechanisms governing genome stability, cell cycle progression, and lineage commitment in T cells (**Figure 3. 3**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-

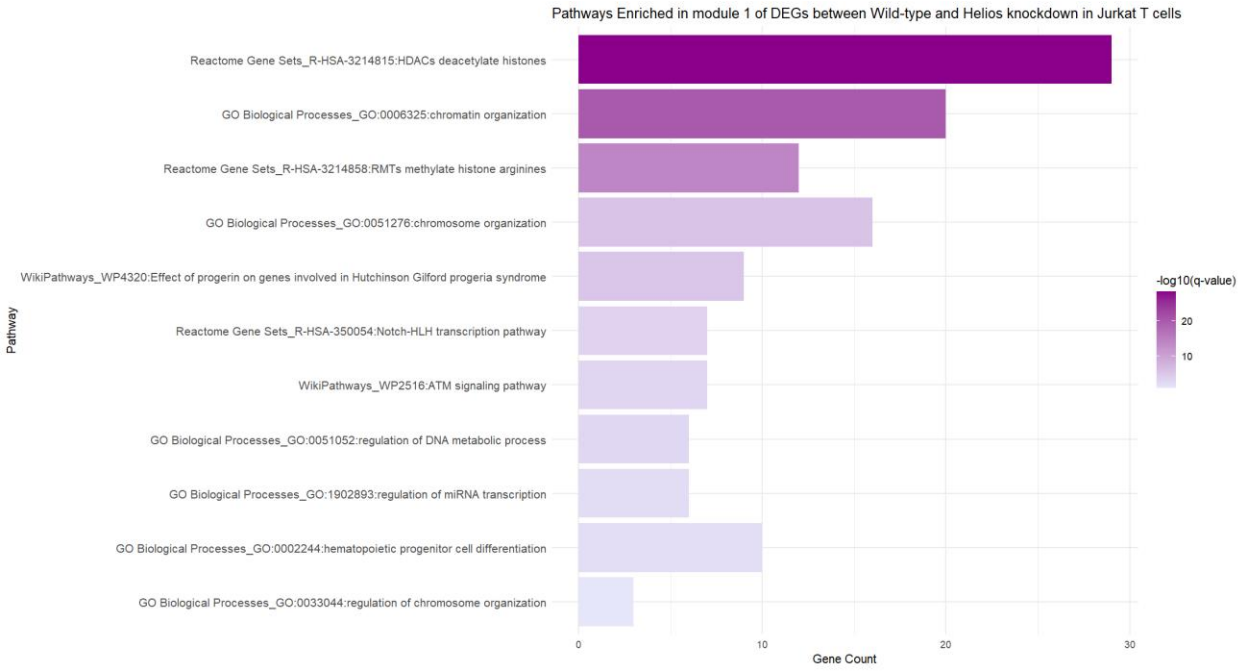
3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 3 Gene Interaction Network and Pathway Enrichment in Module 1 of DEGs between wild type and Helios knockdown in Jurkat T cells

A


MYBBP1A
TAL1
CEBPA
HDAC2
NCOR2
H2AC6
JUN
H2AC8
H2AC21
H3C12
GATA3
H3C13
H2AW
XRCC5
H2BC9
H2BC17
H2BC13
KAT2B
H2BC5
H1-0
H2AC18
RAD51
H2AC7
H2BC12
H3-2
CBX4
H3C3
XRCC6
H1-2
CHEK2

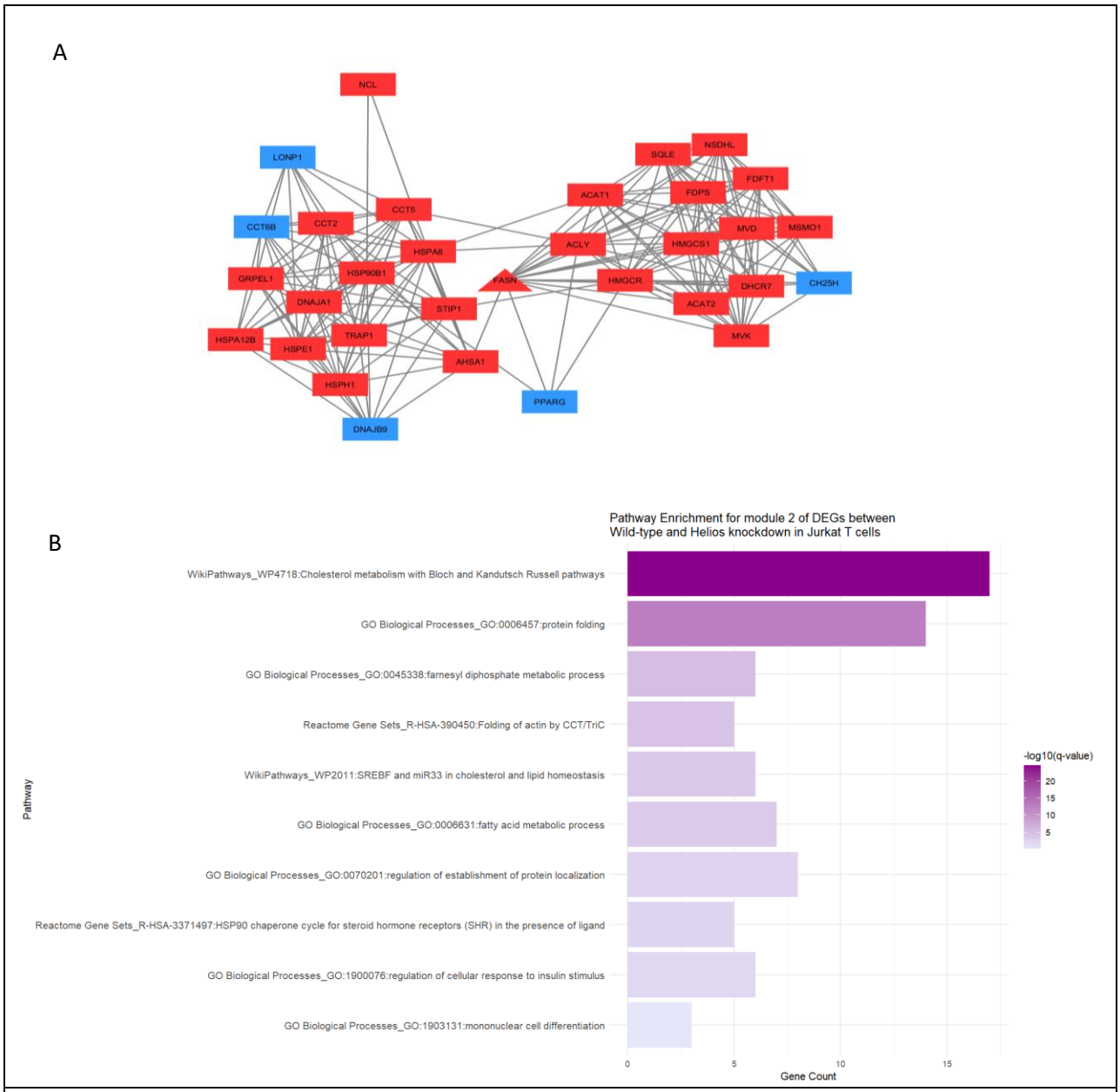
B


Pathways Enriched in module 1 of DEGs between Wild-type and Helios knockdown in Jurkat T cells
Reactome Gene Sets_R-HSA-3214815:HDACs deacetylate histones
GO Biological Processes_GO:0006325:chromatin organization
Reactome Gene Sets_R-HSA-3214858:RMTs methylate histone arginines
GO Biological Processes_GO:0051276:chromosome organization
WikiPathways_WP4320:Effect of progerin on genes involved in Hutchinson Gilford progeria syndrome
Reactome Gene Sets_R-HSA-350054:Notch-HLH transcription pathway
WikiPathways_WP2516:ATM signaling pathway
GO Biological Processes_GO:0051052:regulation of DNA metabolic process
GO Biological Processes_GO:1902893:regulation of miRNA transcription
GO Biological Processes_GO:0002244:hematopoietic progenitor cell differentiation
GO Biological Processes_GO:0033044:regulation of chromosome organization
Pathway
0
10
20
30
Gene Count
-log10(q-value)
20
10

A) The interaction network of genes in module 1, featuring key histone and transcriptional regulators. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *H3C12* identified as the hub gene of the module. B) The pathway enrichment analysis for module 1 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 2 contains 32 nodes and 202 edges. The hub gene is *FASN* (Fatty Acid Synthase). The pathway enrichment analysis of Module 2 DEGs highlights a pronounced influence on cholesterol metabolism. Protein folding processes are also markedly enriched, indicating a potential impact on protein homeostasis. Additionally, pathways related to farnesyl diphosphate metabolism and the HSP90 chaperone cycle reflect changes in lipid metabolism and molecular chaperone activities, which are essential for cellular function and stress response (**Figure 3. 4**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 4 Gene Interaction Network and Pathway Enrichment in Module 2 of DEGs between wild type and Helios knockdown in Jurkat T cells



A) The interaction network of genes in module 2. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *FASN* is the hub gene of this module. B) The pathway enrichment analysis for module 2 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 3 contains 10 nodes and 43 edges. The hub gene is *LYAR* (Ly1 Antibody Reactive). The

pathway enrichment analysis of Module 3 DEGs highlights a pronounced influence on cholesterol metabolism. Protein folding processes are also markedly enriched, indicating a potential impact on protein homeostasis. Additionally, pathways related to farnesyl diphosphate metabolism and the HSP90 chaperone cycle reflect changes in lipid metabolism and molecular chaperone activities, which are essential for cellular function and stress response (**Figure 3. 5**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 5 Gene Interaction Network and Pathway Enrichment in Module 3 of DEGs between wild type and Helios knockdown in Jurkat T cells

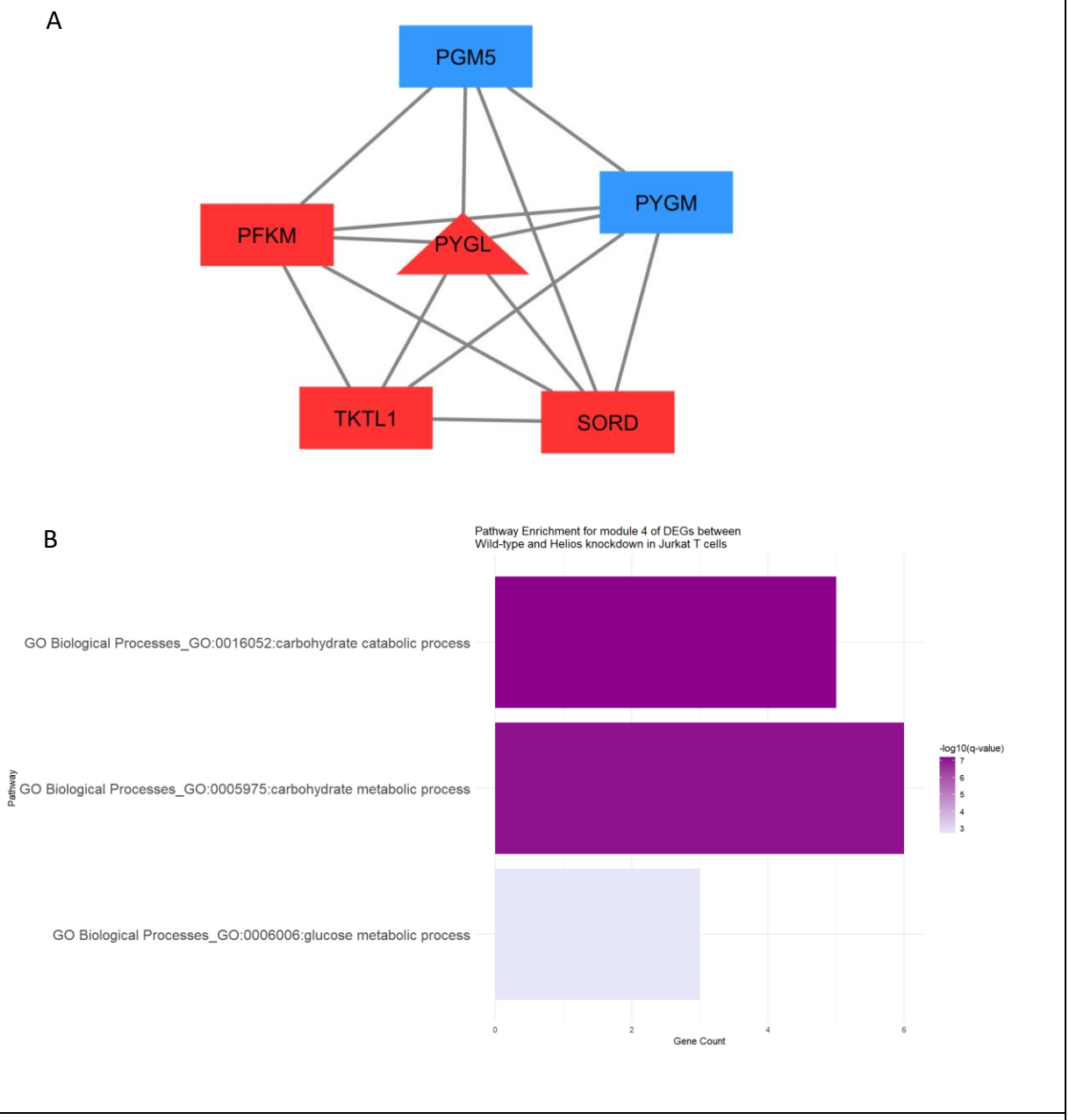
A) The interaction network of genes in module 3, featuring key histone and transcriptional regulators. The red color of the nodes mean they are upregulated following Helios knockdown in Jurkat T cells. The central triangle node *LYAR* is the hub gene of this network. B) The pathway enrichment analysis for module 3 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 4 contains 6 nodes and 14 edges. The hub gene is *PYGL* (Glycogen Phosphorylase L). The pathway enrichment analysis of Module 4 DEGs highlights the involvement in in

carbohydrate-related processes. This includes carbohydrate catabolic processes and carbohydrate metabolic processes, with a specific emphasis on glucose metabolism (**Figure 3. 6**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

.

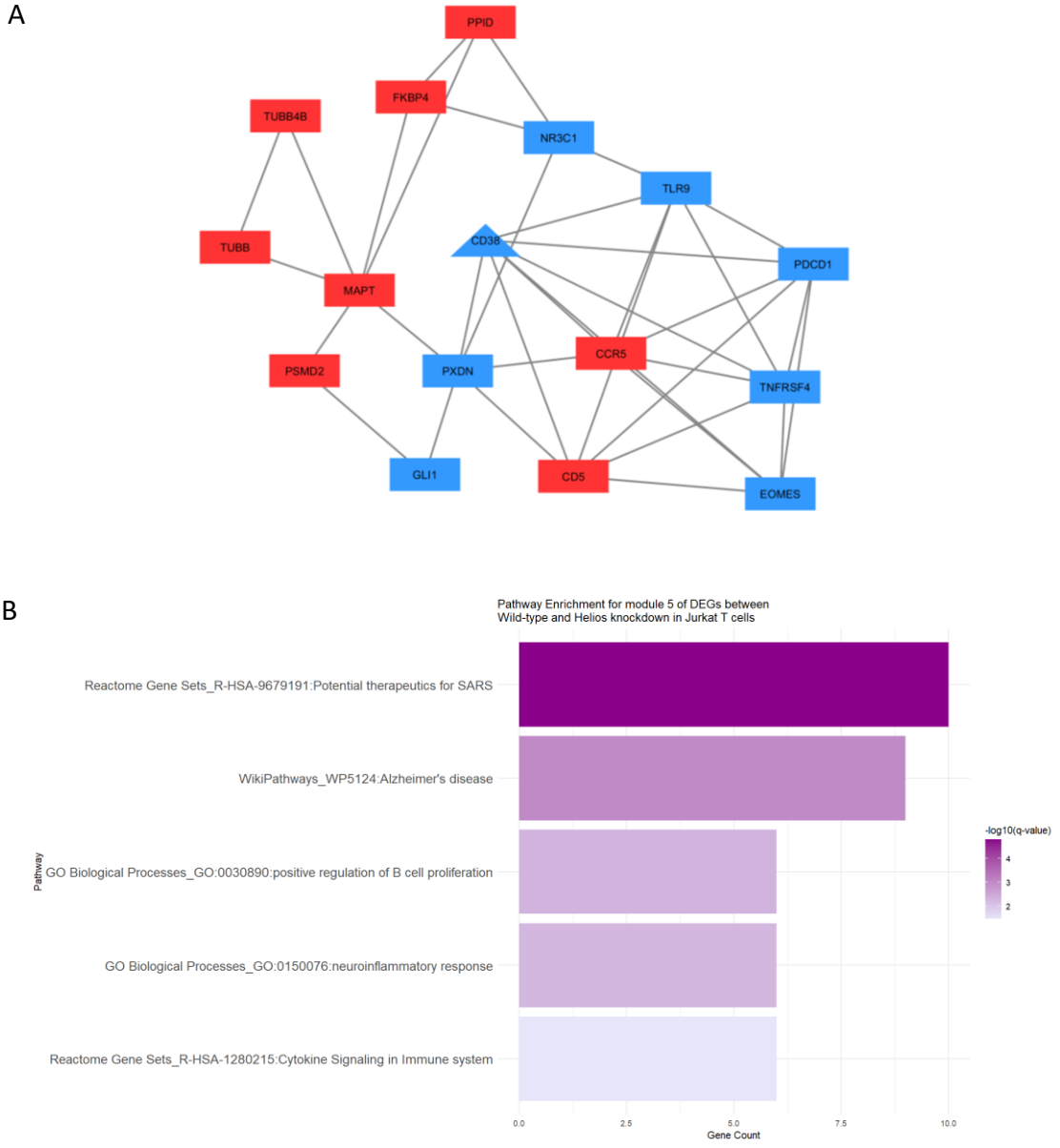
Figure 3. 6 Gene Interaction Network and Pathway Enrichment in Module 4 of DEGs between wild type and Helios knockdown in Jurkat T cells



A) The interaction network of genes in module 4. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *PYGL* identified as the hub gene of this module. B) The pathway enrichment analysis for module 4 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 5 contains 16 nodes and 36 edges. The hub gene is *CD38* (CD38 Molecule). The pathway enrichment analysis of Module 5 DEGs highlights the involvement of these genes in infection response and neuroinflammatory processes including pathways potential therapeutics for SARS, Alzheimer's disease, the positive regulation of B cell proliferation, neuroinflammatory response, and cytokine signaling in the immune system (**Figure 3. 7**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 7 Gene Interaction Network and Pathway Enrichment in Module 5 of DEGs between wild type and Helios knockdown in Jurkat T cells



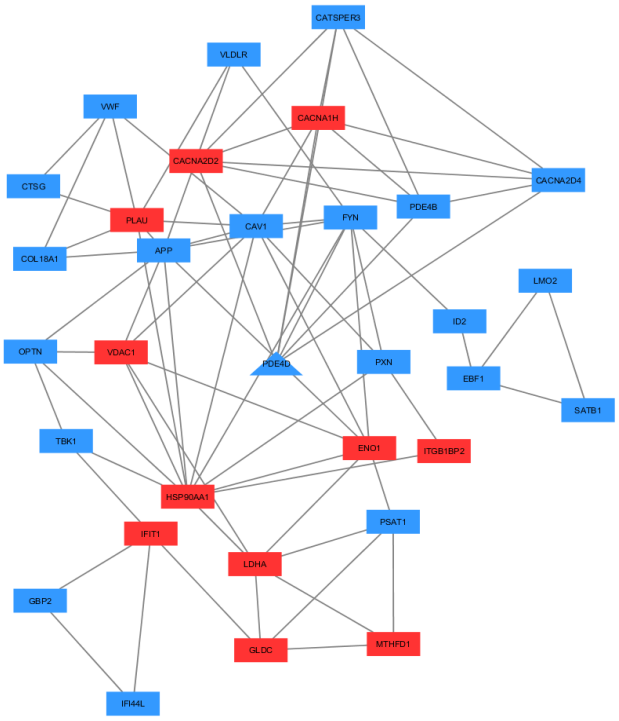
A) The interaction network of genes in module 5. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *CD38* identified as the hub gene of this module. B) The pathway enrichment analysis for module 5 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 6 contains 32 nodes and 74 edges. The hub gene is *PDE4D* (Phosphodiesterase 4D). The

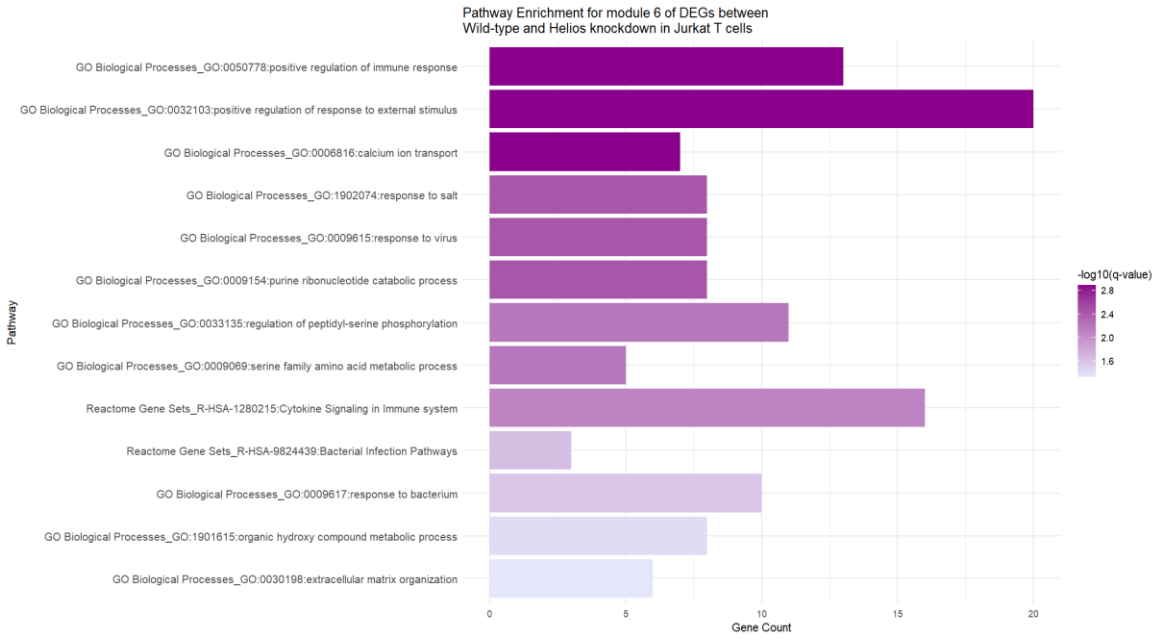
pathway enrichment analysis for module 6 genes identifies several key biological processes. These processes include the positive regulation of response to external stimulus, calcium ion transport, and regulation of muscle contraction, indicating a broad range of cellular responses. Additionally, the pathways involved in response to virus, salt, and bacterium, as well as serine family amino acid metabolism, are highlighted (**Figure 3. 8**). This suggests an activation of various stress and immune responses, reflecting the intricate interplay between metabolic processes and the immune system in this cellular model. For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 8 Gene Interaction Network and Pathway Enrichment in Module 6 of DEGs between wild type and Helios knockdown in Jurkat T cells

A



B

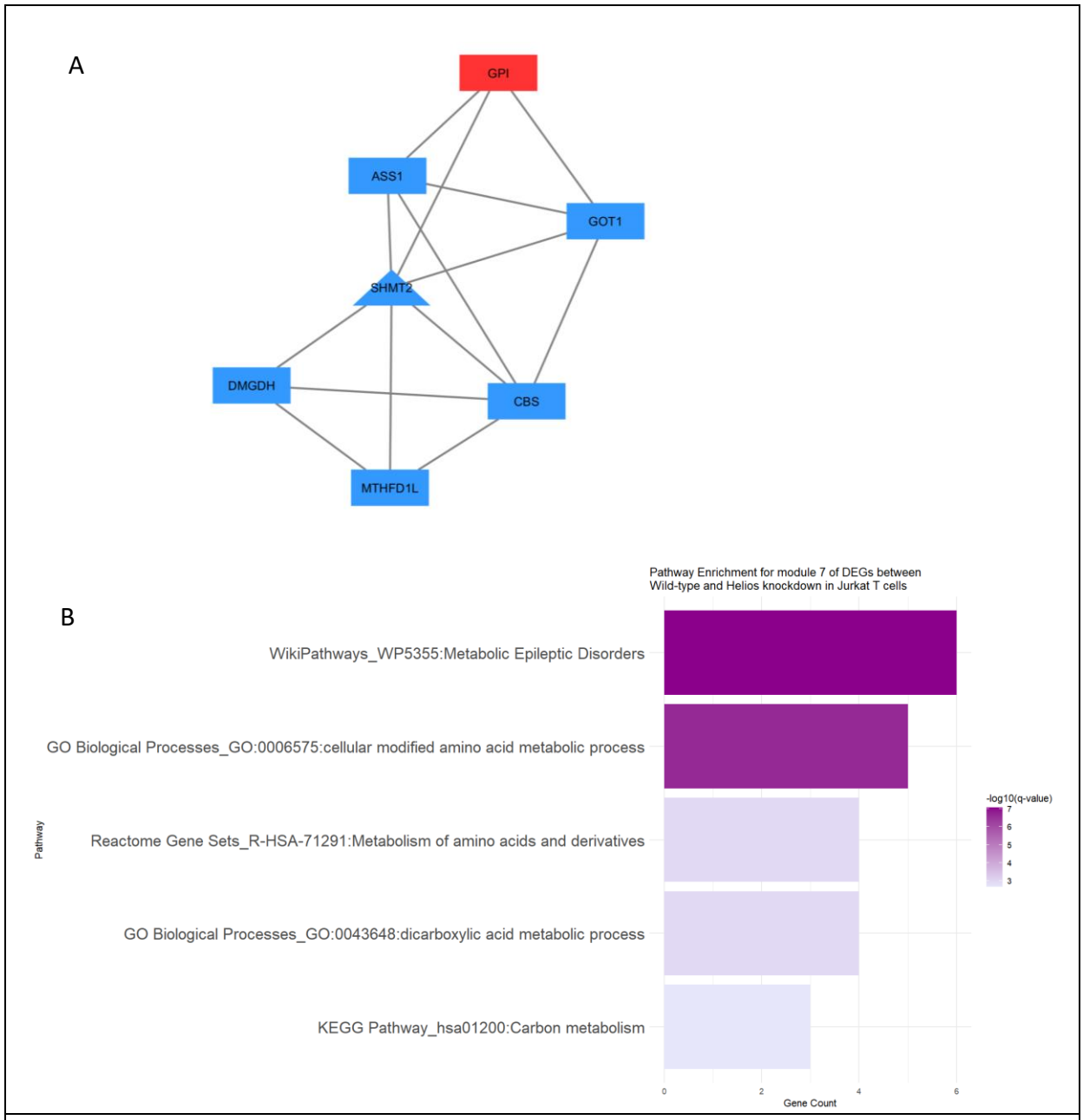


A) The interaction network of genes in module 6. Blue nodes represent downregulated

genes, red nodes represent upregulated genes, and the central triangle node *PDE4D* identified as the hub gene of this module. B) The pathway enrichment analysis for module 6 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 7 contains 7 nodes and 14 edges of which the hub gene is *SHMT2* (Serine Hydroxymethyltransferase 2). Module 7 genes are significantly associated with metabolic processes, particularly those related to epileptic disorders and the metabolism of amino acids and derivatives. There's a notable emphasis on cellular processes involving modified amino acids and dicarboxylic acid metabolism. The metabolic pathways implicated, such as carbon metabolism, suggest a potential disruption in metabolic homeostasis and energy production in response to alterations in Helios activity (**Figure 3.9**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 9 Gene Interaction Network and Pathway Enrichment in Module 7 of DEGs between wild type and Helios knockdown in Jurkat T cells

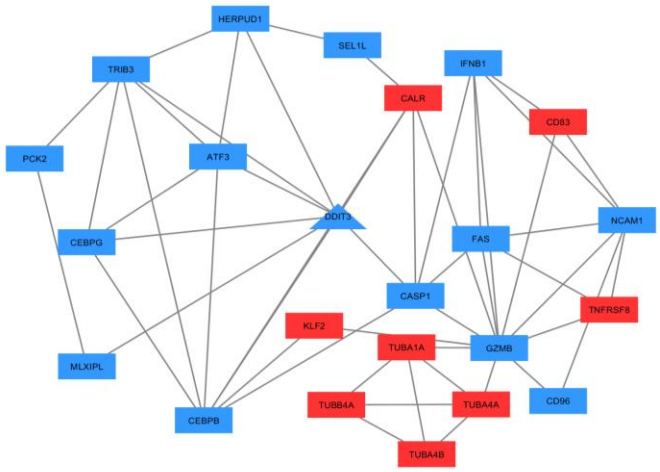


A) The interaction network of genes in module 7. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *SHTM2* identified as the hub gene of this module. B) The pathway enrichment analysis for module 7 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

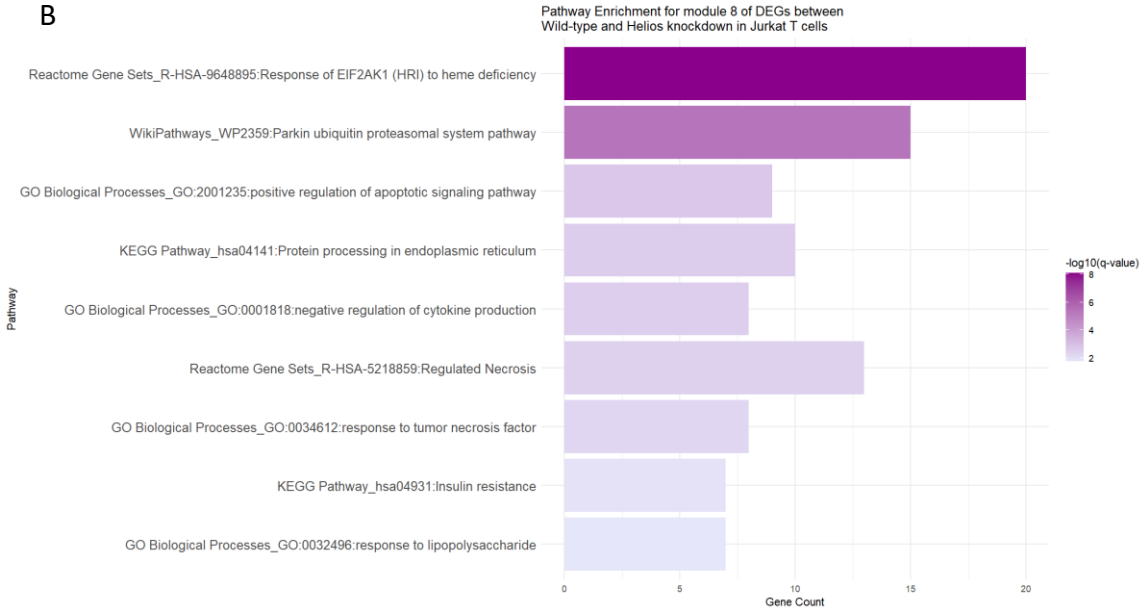
Module 8 contains 23 nodes and 51 edges of which the hub gene is *DDIT3* (DNA Damage Inducible Transcript 3). The pathway enrichment analysis for module 8 genes emphasizes several key biological processes and pathways including responses to heme deficiency, the Parkin ubiquitin proteasomal system, apoptotic signaling, protein processing in the endoplasmic reticulum, and cytokine production regulation. It also highlights involvement in regulated necrosis, response to tumor necrosis factor, insulin resistance, and response to lipopolysaccharide. These pathways collectively suggest a broad range of cellular responses, spanning metabolic regulation, stress response, and immune signaling of this module (**Figure 3.10**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 10 Gene Interaction Network and Pathway Enrichment in Module 8 of DEGs between wild type and Helios knockdown in Jurkat T cells

A



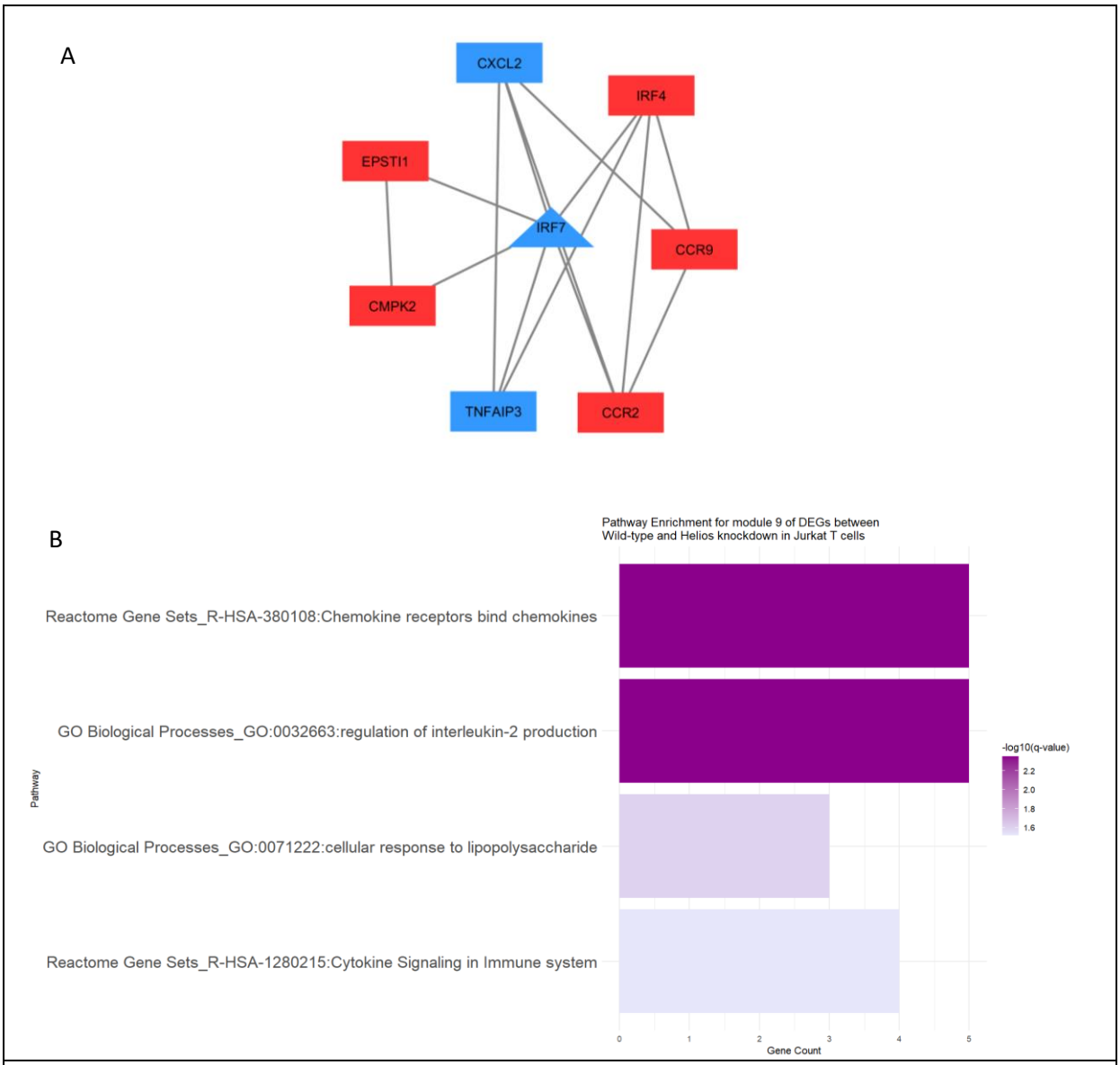
B



A) The interaction network of genes in module 8. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *DDIT3* identified as the hub gene of this module. B) The pathway enrichment analysis for module 8 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 9 contains 8 nodes and 14 edges, of which hub gene is *IRF7* (interferon regulatory factor 7). The enrichment analysis for module 9 genes primarily focuses on immune system-related pathways. Key pathways include chemokine receptor interactions with chemokines, the regulation of interleukin-2 production, cellular responses to lipopolysaccharide, and cytokine signaling within the immune system. These findings underscore the significant role of Helios in regulating immune response pathways, particularly those involving chemokine signaling and cytokine production, highlighting its potential impact on immune cell functions (**Figure 3. 10**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 10 Gene Interaction Network and Pathway Enrichment in Module 9 of DEGs between wild type and Helios knockdown in Jurkat T cells



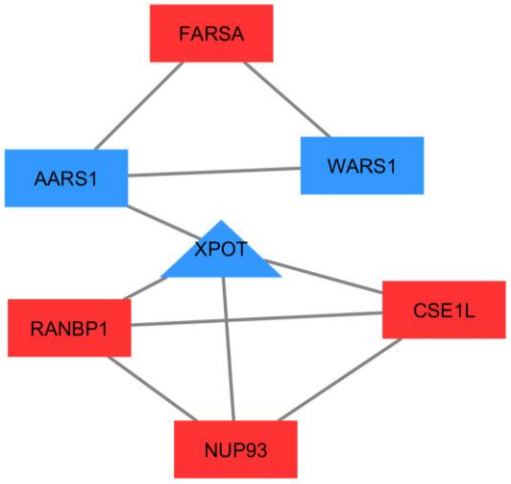
A) The interaction network of genes in module 9. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *IRF7* identified as the hub gene of this module. B) The pathway enrichment analysis for module 9 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

Module 10 has 7 nodes and 10 edges, of which hub gene is *XPOT*. These genes are enriched in

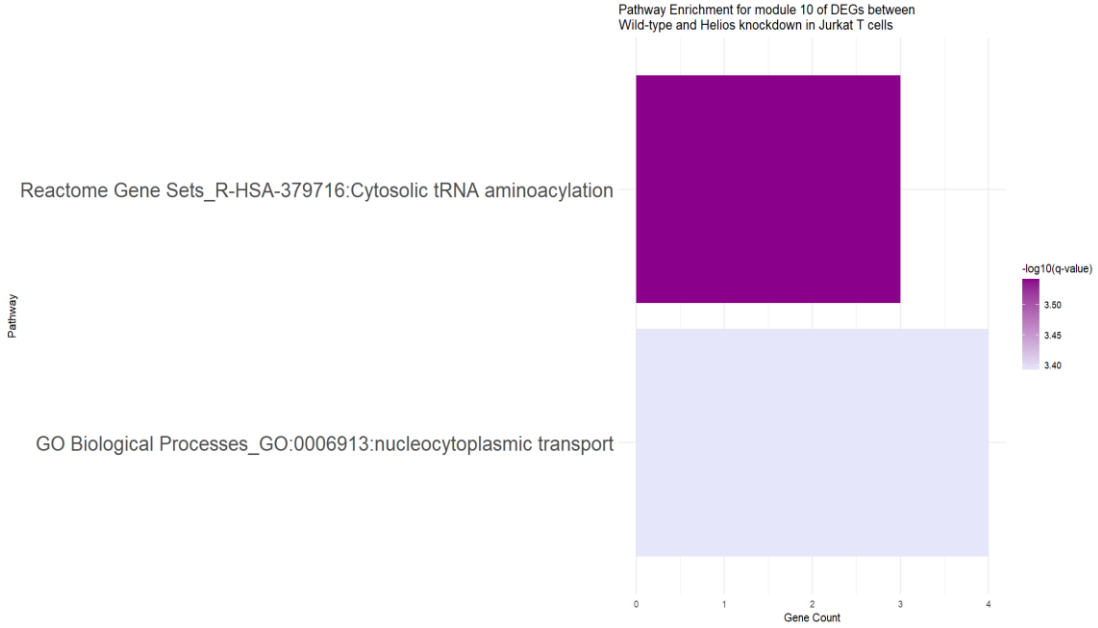
the pathways related to Cytosolic tRNA aminoacylation and nucleocytoplasmic transport (**Figure 3. 11**). For a comprehensive list of all enriched pathways, refer to **Supplementary Table 3.2** (https://www.dropbox.com/scl/fi/r0b8b3hakvupro5q7oeef/Supplementary-Table-3.2-Pathways_enriched_for_the_modules.xlsx?rlkey=xj2ehbde90zzr3u3y4jluynvk&st=ru82flqo&dl=0).

Figure 3. 11 Gene Interaction Network and Pathway Enrichment in Module 9 of DEGs between wild type and Helios knockdown in Jurkat T cells

A



B



A) The interaction network of genes in module 10. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *XPOT (Exportin For TRNA)* identified as the hub gene of this module. B) The pathway enrichment analysis for module 9 genes, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

3.4 Discussion

The comprehensive analysis of differential gene expression between Helios knockdown and wild-type Jurkat T cells, as determined from the Gene Expression Omnibus (GEO) database, reveals significant alterations in gene expression patterns. A total of 1072 genes displayed significant differential expression, with the most prominent changes observed in genes like *CTSG* (Cathepsin G), *KLHDC7B* (Kelch Domain Containing 7B), and *GDAP1L1* (Ganglioside Induced Differentiation Associated Protein 1 Like 1). The functional annotation and pathway analysis of the DEGs, utilizing tools such as Gene Ontology (GO), KEGG, WikiPathways, and BioPlanet, provide deeper insights into the biological processes and pathways affected by Helios knockdown. The enrichment of pathways related to cholesterol biosynthesis, apoptosis, and T cell receptor regulation in the DEGs suggests a multifaceted role of Helios in metabolic processes and immune signaling. The functional annotation of differentially expressed genes in Jurkat T cells, comparing wild-type with Helios knockdown, reveals diverse and intricate cellular pathways across ten modules. Ten modules were identified using MCODE plugin Modules 1 and 3 emphasize epigenetic modulation and lipid metabolism, respectively, highlighting Helios' role in gene expression regulation and stress response. Module 2 and 4 showcase significant impacts on cholesterol and carbohydrate metabolism, indicating effects on lipid and energy metabolism. Module 5 relates to neuroinflammatory and infection response processes, while Module 6 spans a variety of cellular responses including muscle contraction and immune responses. Module 7 focuses on metabolic processes linked to amino acids and epileptic disorders, suggesting alterations in metabolic homeostasis. Modules 8 and 9 emphasize responses to stress and immune signaling, highlighting roles in apoptotic signaling and cytokine production. Lastly, Module 10 is enriched in pathways related to genetic translation processes like cytosolic tRNA aminoacylation. Together, these findings underscore the comprehensive and multi-faceted role of Helios in regulating diverse cellular functions, including metabolism, immune response, and gene expression in T cells. While the role of Helios in immune regulation

is well-documented, research into its involvement in metabolic processes is notably sparse. Therefore, further research is needed to explore the potential links between Helios and metabolic pathways, particularly in the context of adipogenesis and cholesterol biosynthesis, shedding light on its broader physiological functions.

The lack of significant change in *ELF1* expression between the Helios knockdown and wild-type groups may suggest the regulatory role of Helios in *ELF1* is part of a broader, complex network and the presence of redundant or compensatory mechanisms among the network. Conversely, the significant alteration in *IKZF1* expression in Helios knockdown cells solidifies the regulatory relationship between Helios and *IKZF1*. Given the known involvements of both *IKZF2* and *IKZF1* in immune function and their association with SLE, the examination of expression changes in *IKZF1* and *IKZF2* within the context of SLE, compared to healthy controls, could provide invaluable insights into the molecular underpinnings of SLE, enhancing our understanding of its pathogenesis and potentially informing novel therapeutic approaches.

Chapter 4. prioritizing the target genes of Helios in Jurkat T cells

4.1 Introduction

The binding of transcription factors to regulatory regions, such as gene promoters or enhancers, plays a crucial role in the modulation of gene expression. Chip-seq experiments identify hundreds or thousands of binding sites for most factors. Not every identified binding site implies a functional role in gene regulation [229]. This is because transcription factors and other DNA-binding proteins may interact with DNA sequences due to their inherent affinity for these sequences, without necessarily triggering any alterations in gene expression. Therefore,

methods are needed to determine which of these sites are true targets and whether they are functional. Perturbing the transcription factor coding gene by overexpression or knockdown and measuring the effects on cellular gene expression provides useful information on the function of the factor. Integrating binding and gene expression data of the factor perturbation to prioritize the target genes and define direct target genes has been a valuable approach in this regard [230].

To effectively determine the gene regulatory roles of Helios binding sites, first step is to understand how these sites are connected to specific genes. A common approach is to annotate binding sites to their nearest gene, but this method can be oversimplified and may not always yield accurate results. This is because it overlooks the fact that a single regulatory element might influence multiple genes, and the associated target genes could be located at a considerable distance from the regulatory element, linked through intra- or intrachromosomal long-range chromatin interactions. To address these complexities, I employ a strategy that integrates multi-epigenomic data for more precise annotation of binding sites. This method takes into account the distinct roles of promoters and enhancers in gene regulation. By utilizing resources like the FANTOM5 [231] promoter annotations and Enhancer Atlas [232], we can identify overlaps between Helios binding sites and established promoter or enhancer regions. The genes associated with these overlapping promoters or enhancers are then considered as potential targets of the Helios binding sites, providing a more nuanced understanding of Helios' regulatory influence in the genome.

The Functional Annotation of the Mammalian Genome 5 (FANTOM5) [231] project represents a pivotal advancement in our understanding of mammalian genomes, providing comprehensive expression profiles and functional annotations of cell-type-specific transcriptomes with broad applications in biomedical research. Utilizing Cap Analysis of Gene Expression (CAGE) technology, FANTOM5 has systematically identified and cataloged Transcription Start Sites

(TSSs) across the mammalian genome. This method enables the precise determination of transcription initiation points, thereby accurately delineating promoter regions. The project measures the distance between individual peaks and the 5′ ends of known full-length transcripts, assigning peaks within 500 bases of the 5′ end of known transcript models to the corresponding gene. Expanding its scope, FANTOM5 has also developed an extensive enhancer atlas, crucial for understanding the non-coding regions of the genome that are instrumental in gene regulation. Using CAGE, FANTOM5 identifies enhancers by analyzing bidirectional CAGE transcripts, characteristic of active enhancer regions that often produce short, unstable RNA molecules known as eRNAs. This comprehensive mapping includes potential enhancer locations across various cell types and tissues, offering an unparalleled view of the dynamic regulation of gene expression. In addition to mapping, FANTOM5 provides functional annotations of these enhancers, linking them to potential target genes and elucidating their roles in gene regulation networks. This integration with epigenetic data, such as histone modification patterns and chromatin state, enhances the prediction and understanding of enhancer activities. Enhancers are annotated to promoters that exhibit close correlation within a 500kb radius. This annotation, when integrated with gene information from promoter data, facilitates the identification of target genes associated with these enhancers.

EnhancerAtlas 2.0 [232] is another useful database. It consolidates enhancer data from a wide array of genome-wide methodologies, encompassing H3K4me1/H3K27ac histone marks, Dnase-seq/ATAC-seq accessibility, P300 occupancy, POLR2A CAGE-seq transcription initiation, ChIA-PET chromatin interactions, GRO-seq transcription rates, STARR-seq enhancer function, and MPRA functional assays. The database now contains 13,494,603 enhancers across 586 tissue and cell types, including 6,031,402 enhancers in 277 different tissue and cell types specific to Homo sapiens. The authors employed ChIA-PET and Hi-C as gold standard references to establish training datasets for their algorithm which is developed to predict the target genes of enhancers based on six features, including correlation between enhancer activity and gene expression across cell types, gene expression level of target genes, genomic distance between

an enhancer and its target gene, enhancer signal, average gene activity in the region between the enhancer and target gene and enhancer–enhancer correlation from the same cell type. In the context of enhancers annotation, it gives a more comprehensive, integrated knowledge than FANTOM5 enhancer atlas.

4.2 Methods

4.2.1 Annotating chip-seq binding sites of Helios in Jurkat T cells using FANTOM5 promoter annotation data in Jurkat T cells and Enhancers-genes interaction data in Jurkat T cells from EnhancerAtlas 2.0

Promoter annotation data specific to Jurkat T cells under human genome assembly hg19 was retrieved from the FANTOM5 database (https://fantom.gsc.riken.jp/5/). Enhancer-gene interaction data in Jurkat T cells under human genome assembly hg19 was downloaded from EnhancerAtlas 2.0 (http://enhanceratlas.org/index.php). The annotation process involves comparing the genomic coordinates of Helios binding sites with the locations of location of promoters and enhancers in Jurkat T cells. Where these regions coincide, we infer that Helios may exert its regulatory influence on the associated genes within the context of promoter or enhancer activity. Overlapping regions between Helios binding sites in Jurkat T cells and promoter and enhancer regions in Jurkat T cells were identified utilizing the R package IRanges [209]. The genes associated with the identified overlapping promoter or enhancer regions with the binding sites were considered as potential binding targets of Helios in Jurkat T cells. This group of genes has been compiled into a list defined as "Helios binding targets," which is utilized for next step analysis. The functional annotation of these genes was conducted using Metascape.

4.2.3 Integration of chip-seq data and Microarray data

To prioritize the target genes identified by the above steps, we integrated them with the differentially expressed genes (DEGs) obtained from microarray data comparing Helios knockdown and wild-type Jurkat T cells. We define the prioritized target genes as those that appear both in the DEGs (the results discussed in **Chapter 3.3.1**) and as "Helios binding targets" in Jurkat T cells, identified from promoters and enhancers annotation data (the results discussed in **Chapter 4.2.1**). Subsequently, Metascape [228] was used to perform enrichment analysis on the prioritized target genes, employing databases such as KEGG, GO-BP, Reactome, and WikiPathway. The pathways with q-values < 0.05 (log (q value) <0.05) were considered significant.

Additionally, the overlaps between the pathways enriched for "Helios binding targets" and DEGs were also identified using intersect() function of R package dplyr [233]. The Venn diagram of the intersection was visualized using VennDiagram [234] package.

4.2.4 Identifying hub genes of prioritized target genes of Helios in Jurkat T cells

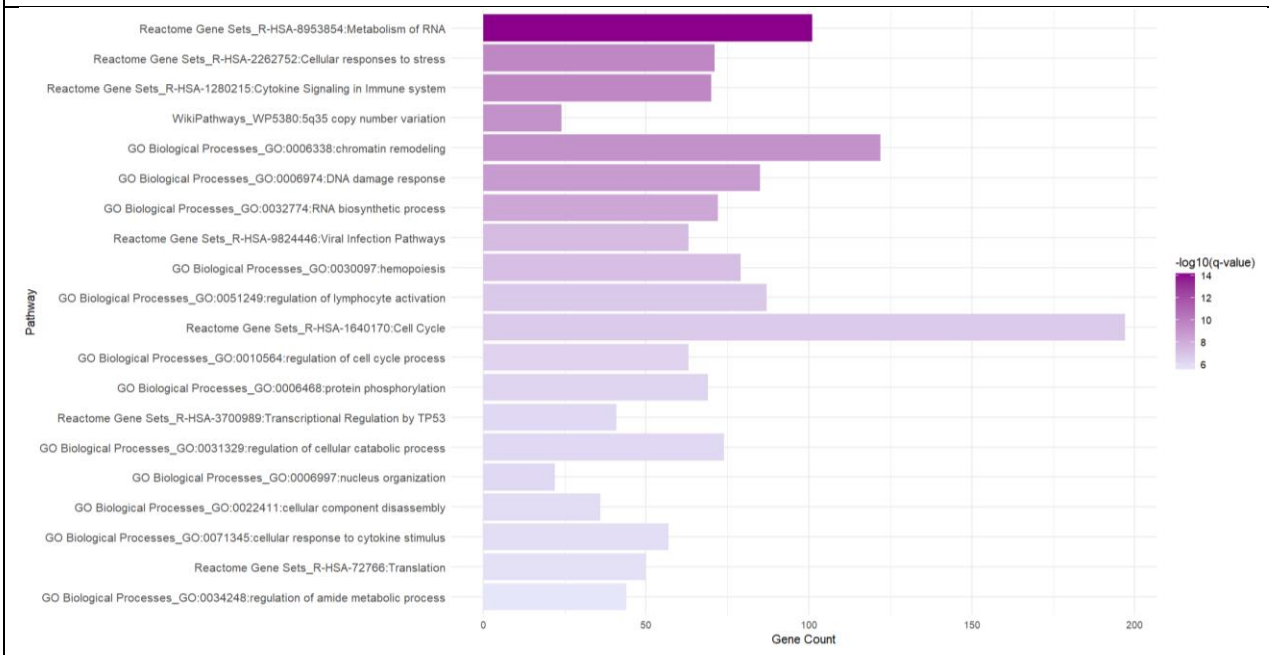
The prioritized target genes were submitted to the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) 12.0 database to build the protein–protein interactive (PPI) network, employing a medium confidence threshold of 0.400 and a false discovery rate (FDR) of 0.05. The result was visualized using Cytoscape 3.7.2. Ten top hub genes were selected using CytoHubba plugin of Cytoscape 3.7.2 through MMC (Maximum Margin Criterion) algorithm.

4.3 Results

4.3.1 The target genes of Helios in Jurkat T cells annotated from binding sites

The analysis identified 775 Helios binding sites in Jurkat T cells that coincide with 1,123 FANTOM5 annotated promoters, linking Helios to the regulation of 720 unique genes (refer to **Supplementary Table 4.1** https://www.dropbox.com/scl/fi/j3smrub1c3zfrladxlgkn/Supplementary-Table-4.1.Helios_genes_annotatedfrompromoters.txt?rlkey=7xau38m7vd6vlgmtrjj3m5w93&st=iqqsrwaj&dl=0 for details). Additionally, 36 binding sites overlapped with 30 EnhancerAtlas 2.0 enhancer regions, corresponding to 767 genes (as shown in **Supplementary Table 4.2** https://www.dropbox.com/scl/fi/d146j8o61rq2huu2j0b7k/Supplementary-Table-4.2.Helios_genes_annotatedfromenhancer.txt?rlkey=o4r8eezpq4td8vj2jrag9me1d&st=8q4sp61s&dl=0). 14 binding sites overlapped with both promoter and enhancer regions (**Supplementary Table 4.3** https://www.dropbox.com/scl/fi/nfx0maq36xllgvgu2qg34/Supplementary-Table-4.3.Annotaion-of-binding-sites-that-overlap-with-both-promoter-regions-and-enhancer-regions.xlsx?rlkey=6xma74xu04zb1k2s64bp893p9&st=ajg7es2j&dl=0), 43 genes (**Table 4.1**) are commonly annotated from both FANTOM5 promoter regions and EnhancerAtlas 2.0 enhancer regions (**Figure 4.2**). In overall, 1444 unique genes (**Supplementary Table 4.4** https://www.dropbox.com/scl/fi/vrhyz03bl5pzai697uvb5/Supplementary-Table-4.4.1444-targetsgenes_promoter_enhancer.txt?rlkey=se5e2fqimfketzwn1wgw5ucdg&st=73tw3t9b&dl=0) are identified from Helios binding sites by this method. The functional annotation of these genes shows they are enriched in various biological processes, predominantly in the metabolism of RNA, cellular responses to stress, and cytokine signaling in the immune system. These genes also align with pathways involved in chromatin remodeling, DNA damage response, RNA biosynthetic processes, and viral infection pathways. They are also involved in the critical cellular functions like hemopoiesis, regulation of lymphocyte activation, the cell cycle, and protein phosphorylation.

Figure 4. 1 Pathways enriched for the target genes of Helios annotated from the binding sites in Jurkat T cells

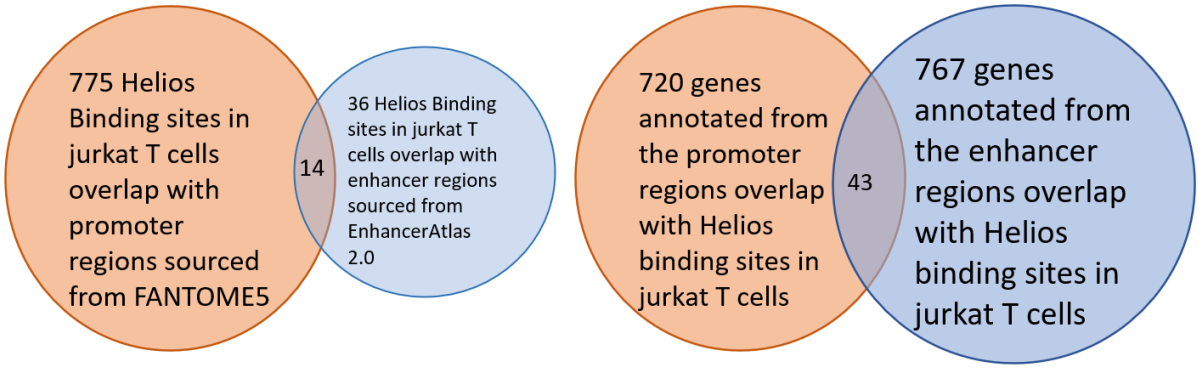


Each bar represents a different biological pathway or process, with the length of the bar corresponding to the number of genes associated with that pathway (Gene Count). The color gradient, ranging from light to dark purple, indicates the level of statistical significance, with darker shades representing higher significance as shown by the negative logarithm of the q-value (-log10(q-value)).

4.3.2 Prioritized target genes of Helios in Jurkat T cells

After finding overlap of this group of genes with the DEGs between wild type and Helios knockdown in Jurkat T cells, I identified a subset of 56 prioritized target genes (**Table 4.2 & Figure 4. 3 A**). The functional annotation of these prioritized genes did not yield any statistically significant enriched pathways with the criteria: terms q-values < 0.05, a minimum of three overlapping genes, and an enrichment factor > 1.5.

Figure 4. 2 Integration of Helios Binding Sites with Promoter and Enhancer Annotations in Jurkat T Cells



The paired Venn diagrams depict the interplay between Helios transcription factor binding sites and associated gene regulatory regions in Jurkat T cells. The diagram on the left indicates that out of the total Helios binding sites identified, 775 coincide with promoter regions cataloged in the FANTOM5 database, while a subset of 36 overlaps with enhancer regions as characterized by EnhancerAtlas 2.0. On the right, the diagram presents a breakdown of gene annotations, revealing that 720 genes associated with Helios binding sites correspond to promoter regions, and 767 genes correspond to enhancer regions.

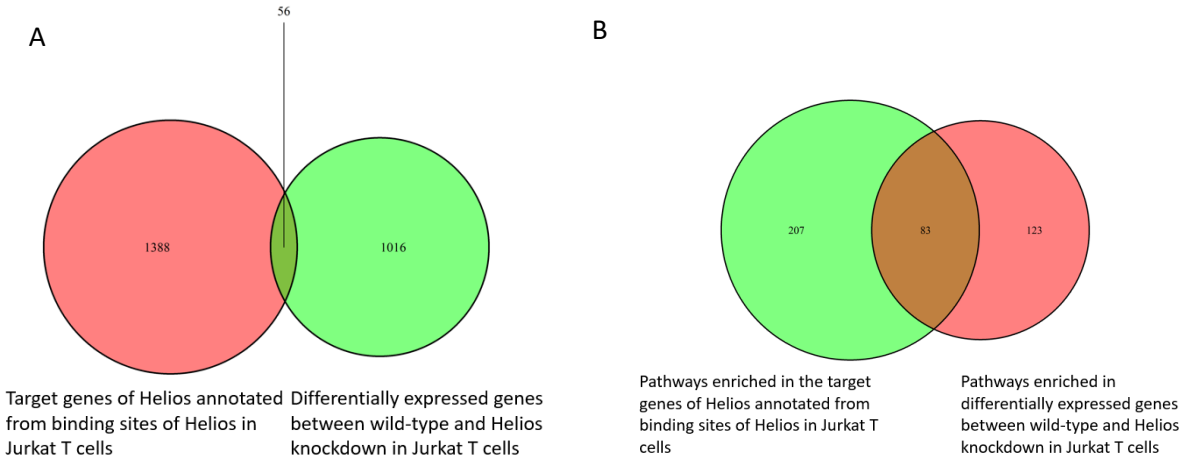
Table 4. 1 Forty-three shared target genes both annotated from promoter and enhancer data

ARL2	*C11orf2*	*COPS7A*	*THOC1*	*RUSC1*	*FBXW2*
SNX15	*MALAT1*	*C12orf57*	*TYMS*	*GGPS1*	*CREB3*
CALM3	*EHBP1L1*	*ELF1*	*CCDC94*	*CDC7*	*HINT2*
CALR3	*MUS81*	*WBP4*	*ZNF576*	*GLYCTK*	
MED26	*BBS1*	*TRIM69*	*PLAUR*	*BNIP1*	
COX8A	*AAAS*	*EARS2*	*TPGS1*	*ZNF184*	
OTUB1	*TNFRSF1A*	*MED11*	*GBA*	*GNB2*	
RPS6KA4	*ZNF384*	*RNF167*	*HCN3*	*FIS1*	

Table 4. 2 Fifty-six prioritized target genes of Helios in Jurkat T cells

FEN1P1	*STIP1*	*KRT1*	*MUM1*	*HIST1H2BL*	*PPM1H*	*PPID*
TPM3	*PYGM*	*ITGB7*	*CHERP*	*PVRIG*	*CCNB1IP1*	*KIF20A*
HAX1	*MALAT1*	*HOXC6*	*BBC3*	*SRRT*	*CHRNA5*	*NR3C1*
ZBTB7B	*CD248*	*AHSA1*	*SEPW1*	*HINT2*	*RFX8*	*NRN1*
HCN3	*VWF*	*PATL2*	*ID1*	*PDSS1*	*CTDSP1*	*RHBDD2*
FDPS	*CD9*	*PRKCB*	*MANF*	*AKIP1*	*SPTBN1*	*SLA*
GNG4	*ACRBP*	*SPNS3*	*ALAS1*	*VSIG10*	*LZTFL1*	*EPHX2*
FLRT1	*SPSB2*	*CFD*	*GNL3*	*PPFIBP1*	*C4orf33*	*S1PR3*

Figure 4. 3 Target genes and pathways overlap between bindings site annotation and DEGs of Helios in Jurkat T cells



A) This Venn diagram represents the overlap between genes targeted by Helios based on ChIP-seq binding site data in Jurkat T cells (red circle) and genes differentially expressed as a result of Helios knockdown in Jurkat T cells (green circle). The intersection (shaded area) shows the number of genes that are both bound by Helios and differentially expressed. B) The Venn diagram depicts the overlap between pathways enriched in target genes of Helios based on its binding sites in Jurkat T cells (green circle) and pathways enriched in genes differentially expressed following Helios knockdown in the same cell type (red circle). The overlap (brown area) represents pathways that are common to both sets.

4.3.3 Overlapped pathways enriched in Helios target genes annotatated from binding sites and DEGs caused by Helios knockdown in Jurkat T cells

There are 83 pathways which are both enriched in “Helios binding targets” based on its binding sites in Jurkat T cells and DEGs followed by Helios knockdown in Jurkat T cells (**Figure 4. 3 B & Table 4.3**). These overlapped pathways span a diverse range of biological processes, notably encompassing cellular stress responses, programmed cell death, and DNA repair mechanisms. This indicates Helios' involvement in maintaining genomic integrity and cellular homeostasis. Several pathways relate to transcriptional regulation, including chromatin modification, highlighting Helios' role in epigenetic regulation. Immune system-related pathways, such as cytokine signaling and T cell differentiation, are prominently featured, underlining the transcription factor's integral role in immune responses. The presence of disease-specific pathways, particularly systemic lupus erythematosus, underscores Helios' potential impact on the disease.

Table 4. 3 Eighty-three pathways enriched in both Helios binding targets and DEGs followed by Helios knockdown

	Term	Category	Description
1	R-HSA-2262752	Reactome Gene Sets	Cellular responses to stress

2	R-HSA-9609646	Reactome Gene Sets	HCMV Infection
3	R-HSA-9609690	Reactome Gene Sets	HCMV Early Events
4	R-HSA-9610379	Reactome Gene Sets	HCMV Late Events
5	R-HSA-9645723	Reactome Gene Sets	Diseases of programmed cell death
6	R-HSA-1912422	Reactome Gene Sets	Pre-NOTCH Expression and Processing
7	R-HSA-9018519	Reactome Gene Sets	Estrogen-dependent gene expression
8	R-HSA-8939211	Reactome Gene Sets	ESR-mediated signaling
9	R-HSA-9616222	Reactome Gene Sets	Transcriptional regulation of granulopoiesis
10	R-HSA-68875	Reactome Gene Sets	Mitotic Prophase
11	R-HSA-1912408	Reactome Gene Sets	Pre-NOTCH Transcription and Translation
12	R-HSA-9821002	Reactome Gene Sets	Chromatin modifications during the maternal to zygotic transition (MZT)
13	R-HSA-157118	Reactome Gene Sets	Signaling by NOTCH
14	R-HSA-3214815	Reactome Gene Sets	HDACs deacetylate histones
15	R-HSA-9710421	Reactome Gene Sets	Defective pyroptosis
16	R-HSA-73728	Reactome Gene Sets	RNA Polymerase I Promoter Opening
17	R-HSA-2559583	Reactome Gene Sets	Cellular Senescence
18	R-HSA-977225	Reactome Gene Sets	Amyloid fiber formation
19	R-HSA-2559580	Reactome Gene Sets	Oxidative Stress Induced Senescence
20	R-HSA-427389	Reactome Gene Sets	ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression
21	R-HSA-5334118	Reactome Gene Sets	DNA methylation
22	R-HSA-1474165	Reactome Gene Sets	Reproduction
23	R-HSA-5625886	Reactome Gene Sets	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3
24	R-HSA-5250924	Reactome Gene Sets	B-WICH complex positively regulates rRNA expression
25	R-HSA-427359	Reactome Gene Sets	SIRT1 negatively regulates rRNA expression
26	R-HSA-68616	Reactome Gene Sets	Assembly of the ORC complex at the origin of replication

27	hsa04613	KEGG Pathway	Neutrophil extracellular trap formation
28	R-HSA-5250913	Reactome Gene Sets	Positive epigenetic regulation of rRNA expression
29	R-HSA-9006931	Reactome Gene Sets	Signaling by Nuclear Receptors
30	R-HSA-212300	Reactome Gene Sets	PRC2 methylates histones and DNA
31	R-HSA-2559582	Reactome Gene Sets	Senescence-Associated Secretory Phenotype (SASP)
32	R-HSA-912446	Reactome Gene Sets	Meiotic recombination
33	R-HSA-2299718	Reactome Gene Sets	Condensation of Prophase Chromosomes
34	hsa05034	KEGG Pathway	Alcoholism
35	R-HSA-8939236	Reactome Gene Sets	RUNX1 regulates transcription of genes involved in differentiation of HSCs
36	R-HSA-1500620	Reactome Gene Sets	Meiosis
37	R-HSA-5578749	Reactome Gene Sets	Transcriptional regulation by small RNAs
38	R-HSA-5625740	Reactome Gene Sets	RHO GTPases activate PKNs
39	R-HSA-8936459	Reactome Gene Sets	RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function
40	R-HSA-68886	Reactome Gene Sets	M Phase
41	R-HSA-211000	Reactome Gene Sets	Gene Silencing by RNA
42	hsa05322	KEGG Pathway	Systemic lupus erythematosus
43	R-HSA-73854	Reactome Gene Sets	RNA Polymerase I Promoter Clearance
44	R-HSA-73864	Reactome Gene Sets	RNA Polymerase I Transcription
45	R-HSA-3214847	Reactome Gene Sets	HATs acetylate histones
46	R-HSA-69278	Reactome Gene Sets	Cell Cycle, Mitotic
47	R-HSA-201722	Reactome Gene Sets	Formation of the beta-catenin:TCF transactivating complex
48	R-HSA-73772	Reactome Gene Sets	RNA Polymerase I Promoter Escape
49	R-HSA-427413	Reactome Gene Sets	NoRC negatively regulates rRNA expression
50	R-HSA-5250941	Reactome Gene Sets	Negative epigenetic regulation of rRNA expression
51	R-HSA-110328	Reactome Gene Sets	Recognition and association of DNA glycosylase with site containing an affected pyrimidine

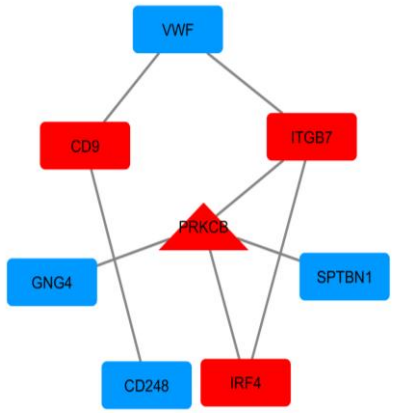
52	R-HSA-110329	Reactome Gene Sets	Cleavage of the damaged pyrimidine
53	R-HSA-73928	Reactome Gene Sets	Depyrimidination
54	R-HSA-9816359	Reactome Gene Sets	Maternal to zygotic transition (MZT)
55	R-HSA-73929	Reactome Gene Sets	Base-Excision Repair, AP Site Formation
56	R-HSA-171306	Reactome Gene Sets	Packaging Of Telomere Ends
57	R-HSA-5617472	Reactome Gene Sets	Activation of anterior HOX genes in hindbrain development during early embryogenesis
58	R-HSA-5619507	Reactome Gene Sets	Activation of HOX genes during differentiation
59	R-HSA-1221632	Reactome Gene Sets	Meiotic synapsis
60	R-HSA-2559586	Reactome Gene Sets	DNA Damage/Telomere Stress Induced Senescence
61	R-HSA-68867	Reactome Gene Sets	Assembly of the pre-replicative complex
62	R-HSA-69002	Reactome Gene Sets	DNA Replication Pre-Initiation
63	R-HSA-1640170	Reactome Gene Sets	Cell Cycle
64	R-HSA-110330	Reactome Gene Sets	Recognition and association of DNA glycosylase with site containing an affected purine
65	R-HSA-110331	Reactome Gene Sets	Cleavage of the damaged purine
66	R-HSA-73927	Reactome Gene Sets	Depurination
67	R-HSA-8878171	Reactome Gene Sets	Transcriptional regulation by RUNX1
68	R-HSA-606279	Reactome Gene Sets	Deposition of new CENPA-containing nucleosomes at the centromere
69	R-HSA-774815	Reactome Gene Sets	Nucleosome assembly
70	R-HSA-195258	Reactome Gene Sets	RHO GTPase Effectors
71	R-HSA-3214858	Reactome Gene Sets	RMTs methylate histone arginines
72	R-HSA-9670095	Reactome Gene Sets	Inhibition of DNA recombination at telomere
73	R-HSA-69306	Reactome Gene Sets	DNA Replication
74	R-HSA-212165	Reactome Gene Sets	Epigenetic regulation of gene expression
75	GO:0006338	GO Biological Processes	chromatin remodeling
76	R-HSA-73884	Reactome Gene Sets	Base Excision Repair
77	R-HSA-9716542	Reactome Gene Sets	Signaling by Rho GTPases, Miro GTPases and RHOBTB3

78	R-HSA-194315	Reactome Gene Sets	Signaling by Rho GTPases
79	R-HSA-9012999	Reactome Gene Sets	RHO GTPase cycle
80	GO:1903706	GO Biological Processes	regulation of hemopoiesis
81	GO:0043408	GO Biological Processes	regulation of MAPK cascade
82	GO:0080135	GO Biological Processes	regulation of cellular response to stress
83	R-HSA-9824446	Reactome Gene Sets	Viral Infection Pathways

4.3.4 PPI network and hub genes of prioritized target genes of Helios in Jurkat T cells

PPI analysis of the prioritized target genes of Helios in Jurkat T cells was based on the STRING database and the results were visualized using Cytoscape. After an initial refinement process that excluded singleton nodes and isolated smaller networks comprising fewer than five nodes, the remaining network consisted of 8 nodes interconnected by 8 edges (**Figure 4. 4**). Within this network, the Cytohubba plugin was employed, utilizing the Maximal Clique Centrality (MCC) algorithm, to pinpoint *PRKCB* as the central hub gene.

Figure 4. 4 Protein-Protein Interaction Network of Prioritized Helios Target Genes in Jurkat T Cells



The protein-protein interaction (PPI) analysis of prioritized Helios target genes in Jurkat T cells, based on the STRING database and visualized using Cytoscape. Blue nodes represent downregulated genes, red nodes represent upregulated genes, and the central triangle node *PRKCB* as the hub gene.

4.4 Discussion

The analysis of Helios binding sites in Jurkat T cells revealed significant associations with FANTOM5 annotated promoters and EnhancerAtlas 2.0 enhancer regions, indicating a potential role for Helios in the regulation of gene expression. Specifically, the study identified 775 Helios binding sites that coincide with 1,123 FANTOM5 annotated promoters, linking Helios to the regulation of 720 unique genes. Additionally, 36 binding sites overlapped with 30 enhancer regions, corresponding to 767 genes. Overall, the analysis identified 1444 unique genes associated with Helios binding sites, suggesting a broad impact on gene regulation. These genes are enriched in various biological processes, predominantly in the metabolism of RNA, cellular responses to stress, and cytokine signaling in the immune system. These genes also align with pathways involved in chromatin remodeling, DNA damage response, RNA biosynthetic processes, and viral infection pathways. The genes also intersect with critical cellular functions like hemopoiesis, regulation of lymphocyte activation, the cell cycle, and protein phosphorylation.

In Chapter 3, I identified 1072 DEGs between wild-type and Helios knockdown in Jurkat T cells. In the current Chapter, the annotation of the Helis binding sites in Jurkat T cells using established promoters and enhancers annotation data recognized 1444 target genes of Helios in Jurkat. However, only 56 genes are found to be shared in these two sets and they are not enriched in any pathway. This finding suggests that the differential expression of the majority of

genes following Helios knockdown may be the result of indirect regulatory mechanisms or compensatory responses within the cell. Despite Helios binding to numerous genomic sites, the limited overlap with differentially expressed genes (DEGs) implies that Helios might influence gene expression through secondary pathways, such as modulating the activity of other transcription factors, engaging in chromatin remodelling, or affecting mRNA processing. This assumption is supported by the observation that the “Helios binding targets” are involved in the pathways such as metabolism of RNA and chromatin remodelling. Additionally, the lack of significant expression change in many Helios-bound genes could be attributed to compensatory mechanisms within the cellular network. These mechanisms might include the activation of alternative pathways or other transcription factors that counterbalance the effects of Helios knockdown, maintaining gene expression levels despite the loss of Helios function. This is supported by the observation that 83 pathways are shared between the Helios binding target genes and DEGs following Helios knockdown in Jurkat T cells, despite a small overlap of genes, suggests that Helios may influence a common set of biological processes and pathways in these cells. The fact that these pathways are enriched in both Helios binding sites and DEGs implies that Helios could play a significant role in regulating these pathways, potentially impacting key cellular functions and the immune response. This also indicates that while direct gene regulation by Helios may be limited, its overall impact on cell function and disease-related pathways is more substantial. It is also essential to acknowledge how microarray technology might limit our insights, as its limitations could mask the full extent of Helios's direct regulatory effects on gene expression. Microarray technology, while useful for broad gene expression studies, has limitations in its sensitivity [235]. This means it may not detect small changes in gene expression levels, especially for genes expressed at low levels or those with subtle changes. As a result, the actual influence of Helios on gene expression might be underrepresented because subtle yet biologically significant changes could go undetected. Microarray provides a 'snapshot' of gene expression at a specific time point [236]. This static nature of data collection means that it might miss capturing dynamic changes in gene expression over time. For instance, if Helios affects gene expression in a temporal manner (i.e., only at certain stages or times after knockdown), these changes might not be detected in a

single microarray snapshot. Given these insights, future research into Helios's role should ideally incorporate a combination of various genomic and transcriptomic technologies. Techniques like RNA-Seq [237] offer higher sensitivity and a broader dynamic range, potentially revealing subtler changes in gene expression and capturing more comprehensive profiles of transcriptional activity.

Despite the absence of any enriched pathways for the prioritized target genes of Helios, which complicates the understanding of their collective function, individual examination of these genes reveals insightful information. Notably, a number of these genes have been identified as being associated with SLE. *PRKCB* (protein kinase C beta), identified as the hub gene among the prioritized Helios target genes, has been linked to susceptibility and pathogenesis of SLE. Research has marked *PRKCB* as a risk locus for SLE [238] and noted increased *PRKCB* mRNA expression in peripheral blood mononuclear cells of SLE patients [239], suggesting a significant role in the disease[239]. Similarly, *MUM1/IRF4* has been implicated in lupus pathogenesis, particularly through its involvement in dendritic cell dysfunction [240], and has been associated with specific gene expression signatures in SLE [241]. *IRF4* deficiency has been observed to reduce lupus nephritis, albeit with increased systemic cytokine production, indicating its complex influence in SLE [242]. The role of *IRF4* in SLE is further underscored by its association with differential microRNA expression patterns in CD4+ and CD19+ cells from asymptomatic SLE patients [243] and thalidomide has been found to exert anti-inflammatory effects in cutaneous lupus by inhibiting the IRF4/NF-κB and AMPK1/mTOR pathways [244]. The involvement of *CD9* in SLE has also been a subject of study, with observations of reduced *CD9* expression in dendritic cells of SLE patients, suggesting its role in the disease's pathophysiology [244]. Further evidence of *CD9*'s role in SLE includes its heightened expression on marginal zone B cells linked to B cell hyperactivity in autoimmune conditions [245], increased presence in CD9-positive exosomes in SLE patients [245], and its potential involvement in intercellular communication via exosomes in SLE[246]. However, the differential expression of these genes across various T cell types between SLE and healthy controls has not yet been addressed. This topic will be explored

in Chapter 6, where we will delve into the cell-type-specific change of expression patterns of the *IKZF2* and its target genes. Given the intriguing connections between several of these genes and SLE, their validation as targets of Helios in T cells within the context of SLE emerges as a pivotal area for future study. The individual roles of genes such as *PRKCB*, MUM1/IRF4, and *CD9* in SLE pathogenesis, as evidenced by existing research, underscore the necessity of further investigation into how Helios influences their expression and function in T cells. This focused validation could illuminate the specific mechanisms by which Helios contributes to the cellular and molecular pathology of SLE. As such, these genes stand out as key candidates for detailed exploration in future research endeavours. This includes the analysis of differential gene expression patterns through RNA-seq, particularly following Helios knockdown in various T cell subsets. Such studies will provide deeper insights into how Helios regulation affects these genes in the context of T cell functionality.

Chapter 5. Trans-cell type Helios binding sites prediction

5.1 Introduction

In the preceding chapters, an in-depth exploration of Helios function in Jurkat T cells has been conducted. However, to comprehensively understand the role of Helios in T cell biology, it is imperative to investigate its function across different T cell subtypes. The identification of Helios as a marker for specific T cell subsets and its association with distinct functional properties underscores the significance of elucidating its function in diverse cellular contexts. Nevertheless, conducting extensive ChIP-seq experiments to delineate binding profiles for each cell type is impractical due to constraints in time and resources. Therefore, there is a critical need for accurate computational approaches to predict transcription factor (TF) binding sites in different cell types. Given that transcription factors (TFs) recognize specific DNA sequences,

known as motifs, to regulate gene expression, leveraging this characteristic, prediction of transcription factor binding sites can be performed by scanning a DNA sequence of interest with known motifs of transcription factor of interest [247]. HOMER[248] (Hypergeometric Optimization of Motif EnRichment) is a widely used bioinformatics tool that specializes in identifying motifs from ChIP-Seq data. The tool extracts DNA sequences corresponding to these peaks and performs motif search, identifying short, recurring sequence patterns that are significantly enriched in the peak regions compared to the background. The output of HOMER can be further used to scan DNA sequence of interest to predict the binding sites using tools like FIMO which utilizes PWMs to scan DNA sequences, scoring each potential binding site for a given motif and identifying those that significantly match the motif pattern. A critical aspect of FIMO's analysis is setting a threshold score to determine significant matches. This threshold is usually based on statistical significance, often using p-values or q-values, to ensure that only the most probable binding sites are considered. When a position in the DNA sequence exceeds this threshold, it is flagged as a potential binding site.

A TF can recognize and preferentially bind to a specific set of DNA sequences referred to as a “binding motif” and these binding preferences can be inferred using sequence alone. However, predictions based on sequence alone lack specificity because they do not consider the alterations of chromatin accessibility between different cell types. Chromatin accessibility is the degree to which regions the genome is “open” to allow transcriptional factors to bind. These regions of chromatin accessibility are cell-type specific and are thereby one of the major contributors to the cell type specificity of TF binding [249]. Studies have shown that incorporating open chromatin information from DNase-seq or ATAC-seq with TF binding motif information can substantially improve predicting TF-bound sites [250-252]. These methods reveal regions of the genome that are actively involved in gene regulation due to their open chromatin structure. Furthermore, the use of chromatin accessibility data enables the distinction of cell type-specific regulatory patterns, as chromatin accessibility varies between different cell types, reflecting their unique regulatory environments [253]. DNase I

hypersensitivity sequencing, or DNase-seq, is a powerful method for assessing chromatin accessibility by digesting chromatin with DNaseI and subsequently sequencing the resulting fragments to identify regions of open chromatin. These regions, known as DNase I hypersensitive sites (DHSs), are typically associated with active gene regulatory elements like promoters and enhancers. In DNase-seq, cells are treated with DNase I, allowing the enzyme to cut the exposed, accessible DNA. The DNA is then purified and sequenced to identify the cleavage sites, effectively mapping chromatin accessibility across the genome[124]. This method hinges on the principle that certain regions of chromatin are more open, allowing the enzyme DNase I to cleave the DNA. These accessible regions, known as DNaseI hotspots, are indicative of active regulatory areas, such as promoters and enhancers, where transcription factors and other regulatory proteins can bind to influence gene expression. By treating cells with DNase I and sequencing the resultant DNA fragments, DNase-seq provides a high-resolution map of these hotspots across the genome, highlighting areas of potential regulatory activity. In this chapter, we utilized HOMER to identify DNA binding motifs of Helios in Jurkat T cells, then employed FIMO to scan DNase I hypersensitivity site hotspots in GM12787 cells to predicted the binding sites of Helios in GM12787 cells. To assess the accuracy and performance of our predictive method, we utilized actual Helios binding sites in GM12787 cells, as determined by ChIP-seq data, as a benchmark for validation.

In addition to the motif-based prediction approach that utilizes DNase I hypersensitivity data to identify potential DNA binding motifs of Helios, I also employed a distinct yet related method that can be defined as “genomic overlap analysis”. This method is based on the observation that for certain transcription factors, a significant proportion of binding sites—ranging from approximately 50% to 90%—are shared across various cell types[254]. This insight underpins the rationale for employing genomic overlap analysis that identifies shared genomic regions indicative of conserved transcription factor binding. By determining the extent of overlap, and setting a minimum threshold based on the size of Helios ChIP-seq peaks, this method seeks to identify regions where the chromatin structure in other cell types is conducive to Helios binding,

as observed in Jurkat T cells. Similar to the motif-based approach, we applied the genomic overlap analysis method to predict Helios binding sites in GM12787 cells. This involved identifying areas of overlap between DNase I hypersensitivity site hotspots in GM12787 cells and established Helios binding sites in Jurkat T cells. The rationale behind this approach was that regions of chromatin accessibility in GM12787 cells, as marked by DNase I hypersensitivity, which coincide with Helios binding sites in Jurkat T cells, could indicate potential binding sites for Helios in GM12787 cells. To validate the effectiveness of this method, we compared our predicted binding sites with the experimentally determined Helios binding sites in GM12787 cells, as obtained from ChIP-seq data.

After conducting a comparison of two distinct methodologies — motif-based prediction and genomic overlap analysis — we selected the method that demonstrated superior performance in accurately identifying Helios binding sites in GM12787 cells to predict the binding sites of Helios across various T cell subtypes. The predicted binding sites of Helios in various T cell subtypes then annotated to the genes by the method described in Chapter 4, which involves an integrated analysis using two pivotal datasets: the FANTOM5 promoter annotation data and the enhancer-gene interaction data from Enhancer Atlas. Furthermore, the use of chromatin accessibility data enables the distinction of cell type-specific regulatory patterns, as chromatin accessibility varies between different cell types, reflecting their unique regulatory environments. Within the same cells, chromatin accessibility changes upon stimulation, as demonstrated by studies showing dynamic alterations in chromatin landscapes in response to various stimuli [255, 256].

5.2 Methods

5.2.1 DNase I hypersensitivity sites hotspots in GM12878 cells

The DNase I hypersensitivity hotspots in GM12878 cells were sourced from the UCSC Genome Browser [257], specifically selecting files for replicate 1 and replicate 2 based on the hg19 genome build. To ensure the robustness and reliability of our data, we employed the IRanges package (version 2.36.0) [258] in R for intersection analysis. This process involved identifying common regions between the two replicates, thereby determining the overlaps. Regions found to be intersecting in both replicate 1 and replicate 2 were considered as definitive hotspots.

5.2.2 Motif-based Prediction of Trans-cell type Binding of Helios in GM12878 Cells

findMotifsGenome.pl function of HOMER [259] was applied to identify DNA binding motifs associated with Helios in Jurkat T cells with -len > 8. We specifically focused on motifs with a length greater than 8 base pairs to ensure a comprehensive search for biologically relevant and complex motifs. Following the motif identification with HOMER, FIMO (Find Individual Motif Occurrences) was used to scan the intersections between two DNase I hypersensitivity sites hotspots replicates in GM12787 cells. Our approach included a thorough examination across a range of statistical stringency levels, applying different p-value cutoffs (ranging from 1e-10 to 1e-300) to capture a broad spectrum of potential motif occurrences. The binding sites predicted based on the only top motif identified in Jurkat T cells are also analyzed. For the results obtained from FIMO, those predicted binding sites with a q-value of less than 0.05 are considered significant.

5.2.3 Predicting Trans-cell type Bindings Through Shared Genomic Regions with Known Helios Binding Regions

leveraging the assumption that shared open chromatin regions may indicate conserved transcription factor binding, I used the function IRanges package (version 2.36.0) [258] to perform a genomic overlap analysis between DNase I hypersensitivity site hotspots in GM12878 cells replicate and the binding sites of Helios in Jurkat T cells. In order to achieve a more detailed assessment of how varying degrees of minimal length of overlap between influence the identification of Helios binding sites, I evaluated multiple overlap lengths. Specifically, I examined minimal overlap lengths of 100, 200, 295, 300, 400, 500, and 590 base pairs (bp). For each length, the predicted binding sites of Helios in GM12878 cells are defined as those peaks that have a minimum overlap of that length with any of the DNase I hypersensitivity site hotspots in GM12878 cells.

5.2.4 Validation of the efficiency of the different methods

The IDR conservative narrow peak file (ENCFF337XDI) of chip-seq binding sites file of Helios in GM12878 cells is downloaded from ENCODE (The Encyclopedia of DNA Elements) [260] (https://www.encodeproject.org/). The comparison process involved aligning the predicted binding sites, derived from each of two methods, with the actual Helios binding sites in GM12878 cells as per the ENCODE data. The predicted binding sites that coincided with the ENCODE-verified binding sites were deemed True Positives (TP), signifying accurate predictions. Conversely, predicted sites that did not correspond to any actual binding sites in the ENCODE data were labeled as False Positives (FP), indicating erroneous predictions. Additionally, any real binding sites from the ENCODE data that were not captured by our predictions were classified as False Negatives (FN), pointing to missed detections.

To quantitatively assess the performance of each prediction method, I calculated two crucial metrics: True Positive Rate (TPR) and Precision. The True Positive Rate, defined as TPR = TP / (TP + FN), measures the sensitivity of our methods, indicating the proportion of actual binding sites that were correctly identified. Precision, calculated as Precision = TP / (TP + FP), evaluates

the accuracy of our predictions, reflecting the proportion of predicted sites that were indeed true binding sites.

5.2.5 Predicting binding sites and target genes of Helios in various T cell subtypes

I employed a comprehensive bioinformatics approach to predict binding sites and target genes of the transcription factor Helios across various T cell subtypes. Our approach involved the prediction of conserved trans-cell type binding events through the genomic overlap of DNase I hypersensitivity sites hotspots within the target cell type and Helios binding regions in Jurkat T cells, as this method exhibited superior predictive precision (**Figure 5. 1**). Subsequently, we utilized the FANTOM5 promoter annotation data and the enhancer-gene interaction data from Enhancer Atlas 2.0 to identify the target genes associated with the predicted trans-cell type binding sites of Helios. The super-enhancer-gene association data from the study conducted by Hnisz et al. [261] served as a supplementary source for enhancer annotation. The dataset is a catalogue of super-enhancers in 86 human cell and tissue types using H3K27ac ChIP-Seq data and super enhancers were linked to the nearest expressed transcripts based on their transcription start site (TSS) proximity, with expressed transcripts defined by a minimum mean H3K27ac ChIP-Seq density of 0.5 rpm/bp within a 500-base pair window around the TSS, and existing enhancer-gene assignments confirmed by prior experimental validation were retained. As DNase I hypersensitivity hotspots data is essential for the analysis. Our study's scope was naturally limited to T cell subtypes for which DNase I hypersensitivity hotspots data were available, including Th1, naïve CD4+, and Th17 cells. Data for these cell types were obtained from publicly accessible sources. Despite the availability of DNase I hypersensitivity hotspots for T regulatory (Treg) cells on the UCSC browser [257], our analysis had to exclude this cell subtype due to the lack of corresponding promoter and enhancer annotation data.

5.2.5.1 predicting binding sites and target genes of Helios in CD4+ naïve T cells

I downloaded DNase I hypersensitivity site hotspot files for CD4+ naïve T cells, specifically CD4+ naïve Wb11970640 from a 26-year-old Caucasian male and CD4+ naïve wb78495824 from a 35-year-old Caucasian female, via the UCSC browser. Using the R package IRanges, we identified intersections between these two datasets, defining them as the final DNase I hotspots in CD4+ naïve T cells. Additionally, we acquired promoter annotation data for CD4+ naïve T cells from the FANTOM5 project, and sourced enhancer-associated gene pairs specific to CD4+ naïve T cells from EnhancerAtlas 2.0.

To predict Helios binding sites within these cells, we employed IRanges again to pinpoint overlapping regions between the shared DNase I hypersensitivity site hotspots from the two replicates and Helios binding sites observed in Jurkat T cells, setting a minimum overlap threshold of 295 base pairs. Regions exhibiting overlaps in both datasets were earmarked as potential Helios binding sites in CD4+ naïve T cells.

Further, I cross-referenced these predicted Helios binding sites with CD4+ naïve T cell-specific promoter annotation data from the FANTOM5 database and enhancer regions from EnhancerAtlas 2.0. We then attributed genes associated with these promoters and enhancers to their corresponding overlapping binding sites.

5.2.5.2 Predicting binding sites and target genes of Helios in Th1 cells

DNase I hypersensitivity sites hotspots files for Th1 cells, specifically Th1 Wb33676984 (from an Asian, female 26-year-old) and Th1_Wb54553204 replicate1 and replicate2 (from a Caucasian,

male 33-year-old) were downloaded from the UCSC browser. Utilizing the R package IRanges, we determined the intersections between these datasets, thereby establishing them as the definitive DNase I hotspots in Th1 cells.

Due to the FANTOM5 project's lack of promoter annotation for Th1 cells, our analysis of predicted binding sites was based solely on enhancer-gene interaction data. We downloaded this specific data for Th1 cells from EnhancerAtlas 2.0. Continuing with IRanges, we identified overlapping regions between the DNase I hypersensitivity site hotspots common to both replicates and Helios binding sites observed in Jurkat T cells. We set a minimum overlap threshold of 295 base pairs for this analysis. The regions showing overlaps across both datasets were designated as potential Helios binding sites in Th1 cells. Subsequently, we employed the IRanges package to detect overlaps between these predicted Helios binding sites and Th1 cell-specific enhancers. Genes associated with these enhancers were then linked to their corresponding overlapping binding sites.

5.2.5.3 predicting binding sites and target genes of Helios in Th17 cells

DNase I hypersensitivity sites hotspots files for Th17 cells (T helper cells expressing IL-17, primary pheresis of single normal subject) was downloaded from UCSC browser. Notably, the FANTOM5 project does not provide promoter annotations for Th17 cells, and EnhancerAtlas 2.0 lacks enhancer-gene interaction data for these cells. Consequently, the annotation of predicted Helios binding sites in Th17 cells was based solely on enhancer-gene associations, using enhancer-associated gene pairs for Th17 cells as identified in the study by Hnisz et al [261].

To predict Helios binding sites within Th17 cells, we utilized the R package IRanges to pinpoint overlapping regions between the DNase I hypersensitivity sites hotspots and Helios binding

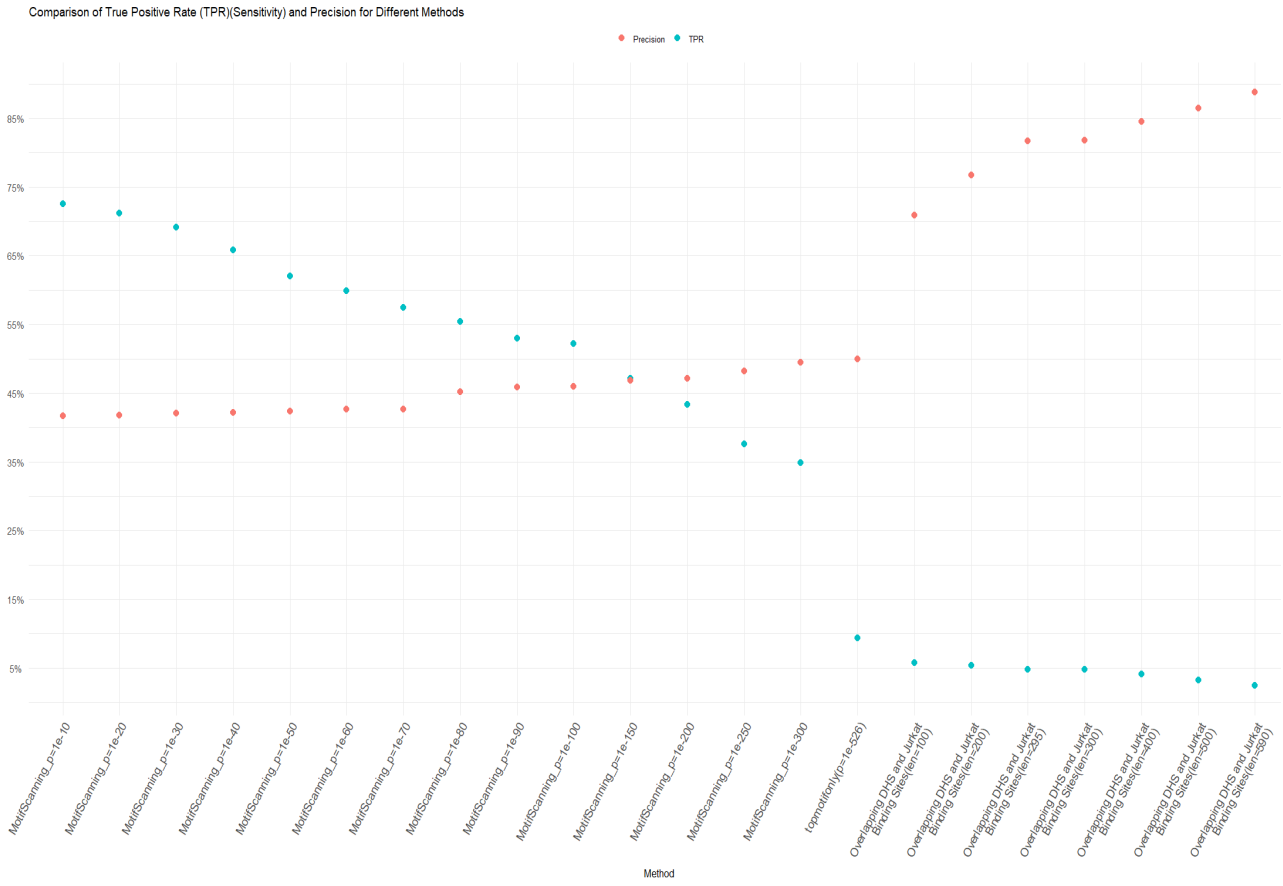
sites observed in Jurkat T cells, setting a minimum overlap threshold of 295 base pairs. These overlapping regions were designated as potential Helios binding sites in Th17 cells. In the subsequent stage, we identified overlaps between these predicted Helios binding sites and enhancers specific to CD4+ naïve T cells. Finally, genes associated with these enhancers were mapped to their respective overlapping binding sites.

5.3 Results

5.3.1 Validation of two trans cell type binding sites prediction methods

The results of validating the efficiency of the methods are summarized in **Figure 5. 1**, which presents the Sensitivity (True Positive Rate) and Precision for each method. As the cutoff p-value for motif-based prediction decreases, Precision demonstrates an increasing trend, while the True Positive Rate exhibits a declining pattern. Similarly, as the minimal overlap length for the method of intersecting DNase I hypersensitivity sites hotspots and binding sites of Helios in Jurkat T cells increases, Precision increases, and the Sensitivity declines. The method of intersecting DNase I hypersensitivity sites hotspots of GM12878 cells with the known binding sites of Helios in Jurkat T cells generally demonstrates higher Precision. Consequently, this method is prioritized for further prediction of Helios binding sites across various T cell subsets due to its ability to minimize false positives, which is crucial when downstream actions or decisions are based on the predicted outcomes. In terms of minimal overlap length between known binding sites of Helios in Jurkat T cells and DNase I hypersensitivity sites hotspots in the target cell types, 295 bp was chosen because it is half of the Helios ChIP-seq peaks in Jurkat T cells. Choosing this length ensures that the summits of the known ChIP-seq peaks are considered while still capturing a significant number of true binding sites. The prediction using this parameter shows Precision of 81.7 % and true positive rate of 4.85 %.

Figure 5. 1 Comparison of True Positive Rate and Precision in Trans-Cell Type Helios Binding Site Prediction Methods



The validation results for two computational methods predicting trans-cell type Helios binding sites in GM12878 cells were illustrated. The True Positive Rate (TPR) and Precision metrics are plotted for each method. The TPR, indicating sensitivity, is the proportion of actual ENCODE-verified binding sites correctly identified by the prediction methods. Precision reflects the accuracy, represented by the proportion of predicted binding sites that were confirmed as true binding sites. The graph shows a trend where Precision increases as the p-value cutoff for motif-based prediction becomes more stringent, while the TPR conversely decreases. The intersection method between DNase I hypersensitivity sites hotspots and known Helios binding sites in Jurkat T cells achieves the highest Precision at 81.7 % but has the lowest TPR at 4.85%.

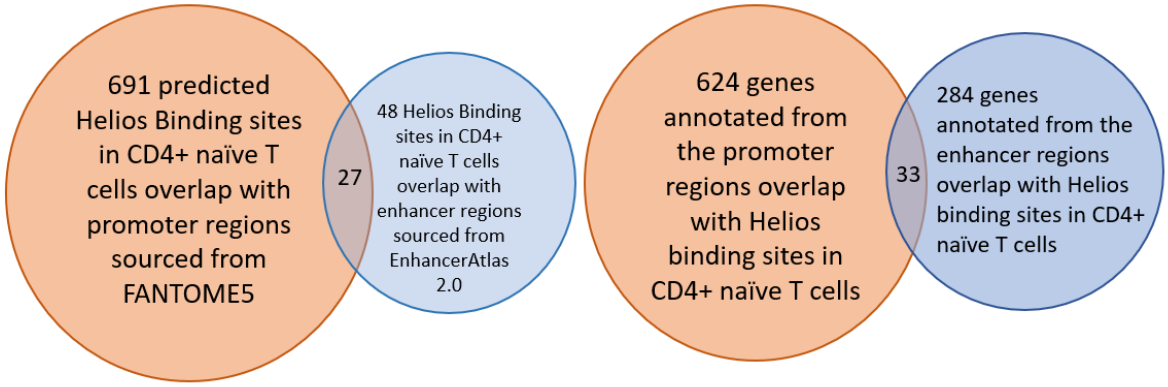
5.3.2 Predicted binding sites and target genes of Helios in CD4+ naïve T cells

2415 binding sites were predicted in CD4+ naïve T cells. Among these, 691 overlap with promoter regions from FANTOM5, corresponding to 624 genes, while 48 overlap with enhancer regions from EnhancerAtlas 2.0, linked to 284 genes (**Figure 5. 2**). In total, this analysis has led to the prediction of 875 target genes (**Supplementary Table 5.1** https://www.dropbox.com/scl/fi/9iffjv5kg4f29kb4xvs6j/Supplementary-Table-5.1.cd4_predicted_targetgenes.xlsx?rlkey=keza81xiin19vcmcuwwdsr6i4&st=8yzf6s08&dl=0).

The pathway enrichment analysis for predicted target genes of Helios in CD4+ naïve T cells reveals a spectrum of biological processes and pathways. Notably, there is a significant enrichment in pathways related to cellular responses to stress, non-coding RNA metabolic processes, and the metabolism of RNA, suggesting a role for Helios in managing cellular stress and RNA dynamics. Other enriched processes include protein catabolism, cell cycle regulation, and cellular responses to stress, highlighting the potential impact of Helios on cell proliferation and survival. Several pathways involved in DNA damage response and signaling pathways mediated by growth factors also feature prominently, indicating Helios's involvement in genomic stability and signal transduction. Immune system pathways, particularly cytokine signaling, innate immune response, and the negative regulation of immune processes, are significantly enriched, emphasizing Helios's regulatory role in immune function. Additionally, pathways related to RNA biosynthetic processes, HIV-1 infection, mitotic cell cycle, and nucleus organization are affected, which may reflect the broad regulatory network influenced by Helios in T cell biology. Disease-specific pathways, such as those involved in Shigellosis and VEGFA-VEGFR2 signaling in cancer, also appear to be modulated, potentially linking Helios function to disease mechanisms. These results collectively underscore the multifaceted role of Helios in orchestrating various cellular and immunological pathways in CD4+ naïve T cells. The summarized pathways are shown in **Figure 5. 3**, For a comprehensive list of all enriched

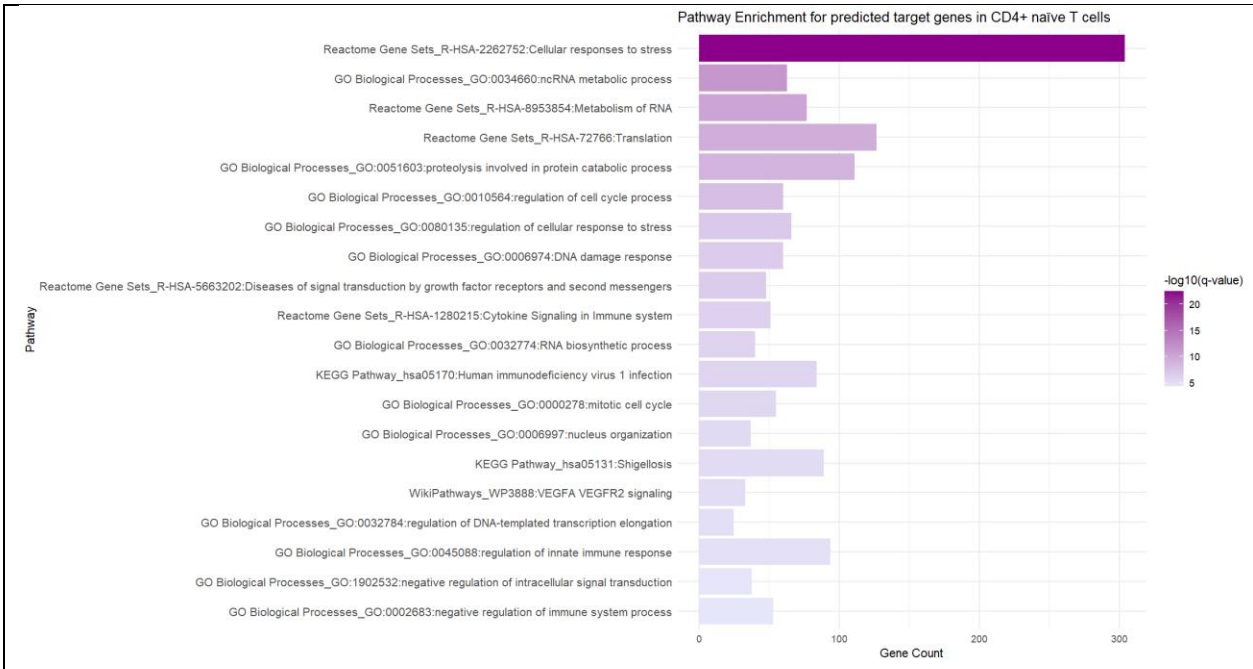
pathways, refer to **Supplementary Table 5.1** (https://www.dropbox.com/scl/fi/9iffjv5kg4f29kb4xvs6j/Supplementary-Table-5.1.cd4_predicted_targetgenes.xlsx?rlkey=keza81xiin19vcmcuwwdsr6i4&st=8yzf6s08&dl=0).

Figure 5. 2 Integration of Helios Binding Sites with Promoter and Enhancer Annotations in CD+ naïve T Cells



The paired Venn diagrams depict the interplay between Helios transcription factor binding sites and associated gene regulatory regions in CD4+ naïve T cells. The diagram on the left indicates that out of the total Helios binding sites identified, 691 coincide with promoter regions cataloged in the FANTOM5 database, while a subset of 27 overlaps with enhancer regions as characterized by EnhancerAtlas 2.0. On the right, the diagram presents a breakdown of gene annotations, revealing that 624 genes associated with Helios binding sites correspond to promoter regions, and 284 genes correspond to enhancer regions.

Figure 5. 3 Pathway Enrichment for the predicted target genes of Helios in CD4+ naïve T cells



The pathway enrichment analysis for predicted target genes of Helios in CD4+ T cells, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

5.3.3 Predicted binding sites and target genes of Helios in Th1 cells

2480 binding sites were predicted in Th1 cells, with 200 of these overlapping with enhancer regions identified in EnhancerAtlas2.0, corresponding to 1193 genes (**Supplementary Table 5.2** https://www.dropbox.com/scl/fi/yvmpl7p5d5hlqvoxcqf59/Supplementary-Table-5.2.Th1cells_predicted_targetgenes.xlsx?rlkey=bjnwxv8zv0mupbb9kp9gwzqju&st=8wg6nz8e&dl=0). The enriched pathways for these predicted target genes of Helios in Th1 cells encompass a range of biological processes including RNA metabolism, stress responses, and cytokine signaling in the immune system. Specific pathways related to diseases of signal transduction and cellular responses to external stimuli are highlighted, indicating Helios' influence on cell signaling and immune response. Additionally, processes such as RNA biosynthesis, cell cycle,

and DNA damage response are notable, underscoring the transcription factor's potential impact on gene expression regulation, cell proliferation, and genomic stability. The summarized terms are shown in **Figure 5. 4**

Figure 5. 4 Pathway Enrichment for the predicted target genes of Helios in Th1 cells

The pathway enrichment analysis for predicted target genes of Helios in Th1 cells, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

5.3.4 Predicted binding sites and target genes of Helios in Th17 cells

2283 binding sites were predicted in Th17 cells, with 1374 of these overlapping with enhancer regions identified in the study by Hnisz et al. [261], corresponding to 843 genes (**Supplementary Table 5.3** https://www.dropbox.com/scl/fi/7ykb3nvcswwozh91fxgu0/Supplementary-Table-5.3.Th17cells_predicted_targetgenes.xlsx?rlkey=0qfawwuuiqrk7sj3avqlhllok&st=sz44007x&dl=

0). Functional annotation of these genes revealed a strong enrichment in pathways related to T cell activation and differentiation, specifically those driving the Th17 immune response. Hemopoiesis and various facets of T cell and leukocyte activation and differentiation, including positive selection and commitment within the T cell lineage, are prominently featured. These results suggest that Helios may play a significant role in regulating the development and function of Th17 cells, which are critical in autoimmune conditions and inflammatory responses. The summarized terms are shown in **Figure 5. 5.**

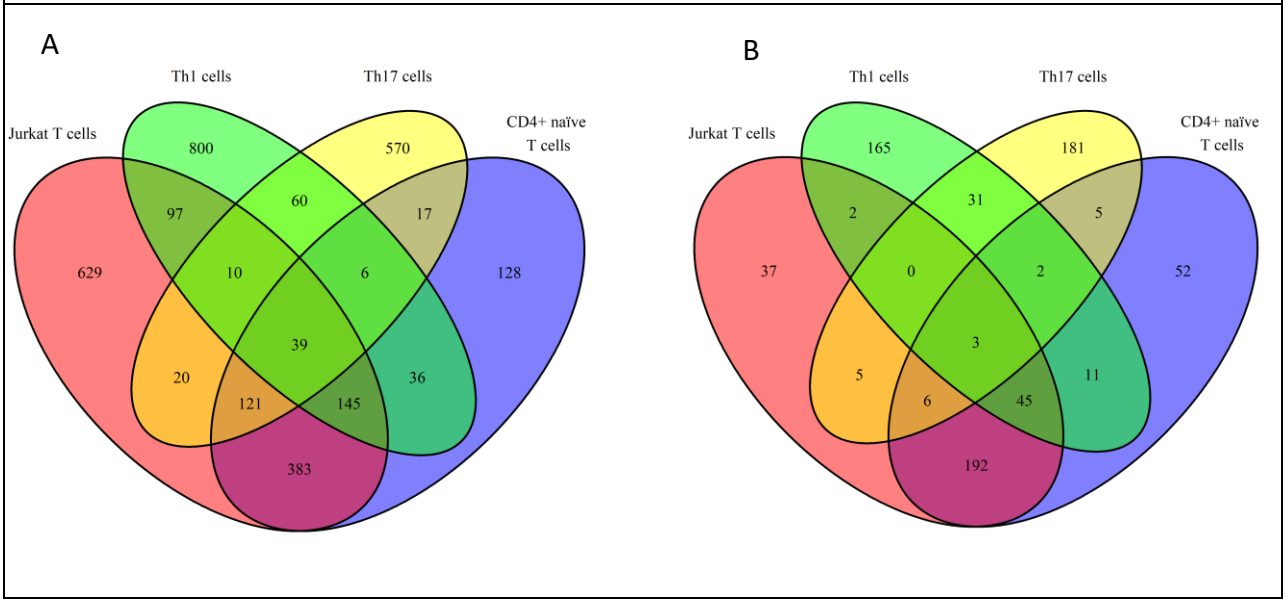
Figure 5. 5 Pathway Enrichment for the predicted target genes of Helios in Th17 cells

The pathway enrichment analysis for predicted target genes of Helios in Th17 cells, where bar length denotes gene count and color intensity reflects the statistical significance of the enrichment, quantified by the -log10(q-value).

5.3.5 Shared predicted target genes of Helios and the pathways over different cell types

Thirty-nine genes are found to be shared across four T cell subsets: Jurkat T cells, CD4+ naïve T cells, Th1 cells and Th17 cells (**Table 5. 1**). Among the four T cell subsets analyzed, Th17 cells stand out with the highest proportion of unique predicted target genes—67.62%—suggesting a distinct regulatory profile for Helios within this subset. Jurkat T cells have fewer unique targets in comparison, at 43.56%. CD4+ naïve T cells have the lowest uniqueness in their gene targets at 14.63%. Th1 cells closely match Th17 in terms of unique target gene proportion, with 67.06% (**Table 5. 2**). Three pathways are overlapped between enriched pathways within the predicted target genes of Helios among the four T cell types (**Table 5. 3**): R-HSA-2262752: Cellular responses to stress, GO:1903706: regulation of hemopoiesis, R-HSA-1280218: Adaptive Immune System.

Figure 5. 6 Overlap of predicted target genes and pathways enriched in the predicted target genes across Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells



A) The diagram illustrates the unique and shared predicted target genes of Helios across four T cell subsets: Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells. Each subset is represented by a colored shape, with overlapping areas indicating shared target genes. The numbers in each segment represent the count of target genes unique to or shared between the cell types, revealing the extent of Helios's potential regulatory influence across different T cell lineages. B) The Diagram showcases the distribution and intersection of pathways enriched within the predicted target genes of Helios among four T cell types: Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells. The numbers in each section denote the count of unique or shared pathways, providing insight into the common biological processes influenced by Helios across different T cell lineages.

Table 5. 1 Thirty-nine shared predicted target genes across Jurkat T cells, CD4+ naïve T cells, Th1 cells, and Th17 cells.

CDC42	*GAPDH*	*GNB2*	*C16orf80*	*IMMT*	*RPS14*
GPR137	*C12orf57*	*ETS1*	*KRI1*	*KDM3A*	*PTGER4*
RASGRP2	*PTPN6*	*FLI1*	*TNFSF9*	*TRAT1*	*PIK3R1*
MALAT1	*WBP4*	*RPS13*	*RHOC*	*NBEAL2*	*RNF19A*
SCYL1	*B2M*	*PSMC3*	*IFI6*	*SMARCA5*	
FAM89B	*MIDN*	*FCHSD2*	*RPS7*	*MATR3*	
TNFRSF1A	*TPM4*	*PCBP2*	*TMEM18*	*TMEM173*	

Table 5. 2 Proportion of unique predicted target genes in different cell types

	overall number of target genes	Number of Unique predicted target genes	Number of predicted target genes Shared with other cell types	proportion of unique predicted target genes
Jurkat T	1444	629	815	43.56%
CD4+naïve T	875	128	747	14.63%
Th1	1193	800	393	67.06%
Th17	843	570	273	67.62%

Table 5. 3 Shared pathways enriched in predicted target genes over four cell types

	Term	Category	Description
1	R-HSA-2262752	Reactome Gene Sets	Cellular responses to stress
2	GO:1903706	GO Biological Processes	regulation of hemopoiesis
3	R-HSA-1280218	Reactome Gene Sets	Adaptive Immune System

5.4 Discussion

The validation of two trans cell type binding sites prediction methods revealed interesting patterns in the efficiency of the methods. As depicted in **Figure 5. 1,** the True Positive Rate and Precision for each method were summarized. It was observed that as the cutoff p value for motif-based prediction decreases, Precision shows an increasing trend, while the True Positive Rate exhibits a declining pattern. Notably, the method of intersecting DNase I hypersensitivity site hotspots and binding sites of Helios in Jurkat T cells demonstrated higher Precision albeit with a lower True Positive Rate. This method has been chosen to further predict trans-cell type Helios binding sites due to its ability to minimize false positives, which is crucial when downstream actions or decisions are based on the predicted outcomes.

Furthermore, the prediction of binding sites and target genes of Helios in CD4+ naïve T cells resulted in the identification of 2415 binding sites and 875 target genes, with involvement in several key pathways involved in cellular stress and RNA dynamics, cell proliferation, survival, and genomic stability. Notably, Helios appears to play a critical regulatory role in the immune system, particularly in cytokine signaling, innate immune response, and the negative regulation of immune processes. Additionally, the study revealed Helios's involvement in broader

pathways such as RNA biosynthetic processes, HIV-1 infection, mitotic cell cycle, and nucleus organization, as well as in disease-specific pathways like Shigellosis and VEGFA-VEGFR2 signaling in cancer. These findings collectively highlight Helios's multifaceted role in orchestrating various cellular and immunological pathways, underlining its critical function in CD4+ T cell biology and potential implications in disease mechanisms.

In the analysis of Th17 cells, I identified 2283 binding sites and 843 target genes of Helios. Functional annotation of these genes indicated a pronounced enrichment in pathways related to T cell activation and differentiation, with a specific focus on the Th17 immune response. Key aspects such as hemopoiesis, T cell and leukocyte activation, and differentiation, including critical steps like positive selection and commitment within the T cell lineage, were notably prominent in the study's findings. These insights suggest that Helios could have a considerable impact on the development and functioning of Th17 cells, which are vital in the context of autoimmune diseases and inflammatory responses.

The discovery of thirty-nine genes shared across four T cell types indicates that Helios regulates a core set of genes essential to various T cell subsets. This finding indicates that Helios has some common regulatory functions in these cells. These shared targets could be involved in fundamental processes necessary for the basic functioning of all these T cell types. These shared genes are likely pivotal for understanding Helios's function across all these cell types. Th17 cells have the most unique target genes, suggesting a specialised regulatory role for Helios. Th1 cells show a similar pattern of Helios influence. Additionally, I have identified three key pathways that are common among the predicted target genes of Helios in these T cell types: cellular responses to stress, the regulation of hemopoiesis, and the adaptive immune system. The presence of these overlapping pathways underscores the fundamental roles Helios plays in T cell functionality across various subsets, highlighting its significance in both normal immune function and in the pathogenesis of immune-related diseases like SLE.

In summary, the validation of the efficiency of the prediction methods, along with the prediction of the binding sites and target genes of Helios in different T cell types, provides valuable insights into the regulatory mechanisms and pathways involved in T cell differentiation and function. The distinct and shared pathway enrichment observed in each T cell subtype underscores Helios's role in exerting both universal and specialised regulatory influences within T cells. The findings from this Chapter highlight Helios's multifaceted role in T cell regulation, affecting a variety of cellular and immunological pathways. This underlines its significant function in T cell biology and potential implications in various disease mechanisms.

Chapter 6. Regulatory effects of *IKZF2* in different cell types in SLE patients

6.1 Introduction

Understanding the expression changes of specific genes in different cell types between diseases and healthy controls is crucial for comprehending the cellular mechanisms underlying diseases and role of the genes of interest in diseases. The methods often involve analyzing gene count matrices across various cell types, employing tools like differential expression analysis. This can reveal how genes like Helios behave differently in immune cells under diseased conditions, providing valuable information for targeted therapeutic strategies. Analyzing the change in expression of specific genes between diseased and control groups across various cell types is crucial for understanding the molecular mechanisms underlying diseases. Several methods are employed for this purpose, including RNA Sequencing (RNA-Seq), Quantitative Real-Time PCR

(qRT-PCR), and Microarray Analysis. Additionally, bioinformatics tools such as edgeR[262] and DESeq2[263] are widely used for statistical analysis, identifying significant expression differences crucial for understanding cellular and molecular disease mechanisms. RNA-Seq [237] offers a comprehensive genome-wide expression analysis, providing insights into the differential expression of genes across various cell types in diseased and control samples[264]. It enables the identification of disease-specific alterations in gene expression and cellular composition. On the other hand, qRT-PCR[265] provides targeted, highly sensitive measurement of specific genes, allowing for precise quantification of gene expression changes in different cell types. While less sensitive than RNA-Seq, Microarray Analysis[266] can measure thousands of genes simultaneously, providing a broad view of gene expression changes across different cell types.

The Gene Set Enrichment Analysis (GSEA) method has been widely used in various studies to interpret gene expression data and identify statistically significant differences in gene expression across different biological conditions. GSEA assesses whether predefined sets of genes show statistically significant differences in expression across different biological states, providing a knowledge-based approach for interpreting gene expression data. It has been applied to identify potential biomarkers, understand immune system remodeling, and interpret omics data, making it a versatile tool for biological research [267-269]. Additionally, GSEA has been utilized to visualize enrichment plots and identify key genes and pathways in various biological contexts, demonstrating its utility in functional interpretation of large-scale data[269, 270]. GSEA has become a popular technique due to its ability to provide insights into the functional dynamics of genes and their roles in specific biological contexts, making it a valuable tool for understanding gene expression patterns and their implications in diverse biological processes[271]. In this study, Gene Set Enrichment Analysis (GSEA) is utilized to detect the enrichment of predicted target genes of Helios in CD4+ naïve T cells, Th1 cells, and Th17 cells, focusing on differences between SLE patients and healthy controls. By identifying Helio's predicted target gene sets with significant enrichment scores in DEGs of different T cells

between controls and SLE, GSEA can reveal key target genes, pathways, and biological processes where Helios might play a crucial role in the context of SLE. This approach helps to understand the potential impact of Helios in immune regulation and disease pathogenesis, specifically in relation to SLE.

Transcription factors (TFs) play a crucial role in regulating gene expression by binding to specific DNA sequences in the promoter region of target genes or by forming complexes with other DNA binding proteins. This binding can lead to the upregulation or downregulation of gene expression, and the levels of the target gene are expected to change in concert with fluctuations in the TF's levels.[272] This biological premise allows for the use of correlation analysis as a means to validate computational predictions of transcription factor targets. The statistical significance of the correlation serves as an indicator of a functional link, providing empirical support for the computational models. Such an approach is particularly valuable for sifting through potential false positives, refining the predictive output by highlighting those genes with a more probable direct or indirect regulatory connection to the transcription factor. Genes exhibiting strong correlations can be earmarked for further experimental validation, such as chromatin immunoprecipitation or functional reporter assays, to confirm the regulatory relationship. In this Chapter, a correlation analysis was conducted to explore the relationship between *IKZF2* expression and its recognized core target genes related to SLE. objective was to refine our understanding of the gene regulatory networks involving Helios (*IKZF2*) and to provide additional layers of information that could assist in prioritizing these genes for further research.

6.2 Methods

6.2.1 IMMUNEUXT dataset introduction

The ImmuNexUT [273] dataset is a comprehensive gene regulation atlas encompassing RNA-seq data on a diverse range of immune cell types within the context of immune-mediated diseases. This dataset has captured distinctive gene expression profiles across 28 different immune cell types, sourced from 337 patients diagnosed with 10 categories of immune-mediated diseases, as well as data from 79 healthy volunteers. Recently, after recent expansion, this dataset now encompassing data from a total of 416 donors. Combined with whole-genome sequencing genotype information, the study enables the identification of cell-type-specific and context-dependent expression quantitative trait loci (eQTLs). The dataset covers gene expression over 27 immune cell types and genotype from 159 SLE and 89 healthy donors.

6.2.2 Analysis of Differentially Expressed Genes (DEGs) in Various T Cell Subtypes between Healthy Controls and SLE Patients

The gene count matrices of 159 SLE and 89 healthy donors in 8 T cell subtypes (CD4+naïve T cells, CD4+ memory T cells, naïve regulatory T cells, effector regulatory T cells, T follicular helper cells, Type 1 T helper, Type 2 T helper cells, T helper 17 cells) were obtained from DNA Data Bank of Japan (DDBJ) [274] under accession number JGAS000486, following the approval of our application to use the data. The R package edgeR (version 3.38.4) [275] was used to pre-process the data and conduct differential expression analysis between healthy controls and SLE patients for each T cell subtype. The change of expression of *IKZF2* in different cell types are compared.

6.2.3 Gene set enrichment analysis of target genes of Helios in DEGs between controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells

Gene set enrichment analysis (GSEA) of the predicted target genes of Helios to DEGs between healthy controls and SLE patients in CD4+ naïve T cells, Th1 cells, and Th17 cells were

conducted. Specifically, we assessed the differential expression of genes in each of these three cell types, applying a significance threshold of an FDR-adjusted p-value of 0.05. To perform GSEA, we ranked the differentially expressed genes (DEGs) based on their log-fold change (log2FC) and utilized the R packages clusterProfiler and enrichplot (version 1.16.2) for the analysis and visualization of results. Gene sets with a normalized enrichment score (NES) exceeding 1.0 and a p-value below 0.05 were considered as significantly enriched.

6.2.4 PPI network and functional annotation of core enrichment genes in Th1 cells

The core enrichment genes among Helios-predicted target genes associated with differentially expressed genes (DEGs) in Th1 cells when comparing healthy controls to individuals with SLE were extracted from GSEA result in Th1 cells. Metascape [224] was used to perform functional annotation to these core enrichment genes in Th1 cells. To explore protein-protein interactions, we employed the STRING 12.0 database and established a protein-protein interaction (PPI) network. We applied a medium confidence threshold of 0.400 and controlled the false discovery rate (FDR) at 0.05. Subsequently, we conducted module analysis using the MCODE [276] plugin within Cytoscape 3.7.2[277], with a cutoff MCODE score of >2. Additionally, we identified the top hub genes within each module using the CytoHubba [278]plugin in Cytoscape 3.7.2, employing the Maximum Margin Criterion (MMC) algorithm.

6.2.5 PPI network and functional annotation of core enrichment genes in Th17 cells

The core enrichment genes among Helios-predicted target genes associated with differentially expressed genes (DEGs) in Th17 cells when comparing healthy controls to individuals with SLE

were extracted from GSEA result in Th17 cells. Metascape [224] was used to perform functional annotation to these core enrichment genes in Th17 cells. To explore protein-protein interactions, we employed the STRING 12.0 database and established a protein-protein interaction (PPI) network. We applied a medium confidence threshold of 0.400 and controlled the false discovery rate (FDR) at 0.05. Subsequently, we conducted module analysis using the MCODE [276] plugin within Cytoscape 3.7.2 [277], with a cutoff MCODE score of >2. Additionally, we identified the top hub genes within each module using the CytoHubba [278] plugin in Cytoscape 3.7.2, employing the Maximum Margin Criterion (MMC) algorithm.

6.2.6 Correlation of expression of *IKZF2* with the target genes of Helios in Th1 cells cells

In this step, we aimed to elucidate the correlation between *IKZF2* and its predicted target genes in Th1 cells, particularly focusing on genes core-enriched in differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients. We extracted gene expression datasets for Th1 cells from SLE patients, which included read counts for 26,353 genes across 152 Th1 cell samples from SLE patients from the data I obtained in Chapter **6.2.2.** The raw read counts were normalized using edgeR package [262]. I specifically selected *IKZF2* and the predicted target genes of Helios in Th1 cells as the genes of interest. I calculated the Pearson correlation coefficients and p-values between the expression levels of *IKZF2* and each of these predicted target genes. To visualize these correlations, we produced a bar plot showing the top 20 genes based on their correlation with *IKZF2*, using R's ggplot2 [227] library. The correlations with the threshold of p value<0.05 are considered significant.

6.2.7 Correlation of expression of *IKZF2* with predicted target genes of Helios in Th17 cells cells

In this step, I aimed to elucidate the correlation between *IKZF2* and its predicted target genes in Th17 cells, particularly focusing on genes core-enriched in differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients. We extracted gene expression datasets for Th17 cells from SLE patients, which included read counts for 26,353 genes across 157 Th17 cell samples from SLE patients from the data I obtained in **Chapter 6.2.2.** The raw read counts were normalized using edgeR package [262]. I specifically selected *IKZF2* and the predicted target genes of Helios in Th17 cells as the genes of interest. I calculated the Pearson correlation coefficients and p-values between the expression levels of *IKZF2* and each of these predicted target genes. To visualize these correlations, I produced a bar plot showing the top 20 genes based on their correlation with *IKZF2*, using R's ggplot2 [227] library. The correlations with the threshold of p value<0.05 are considered significant.

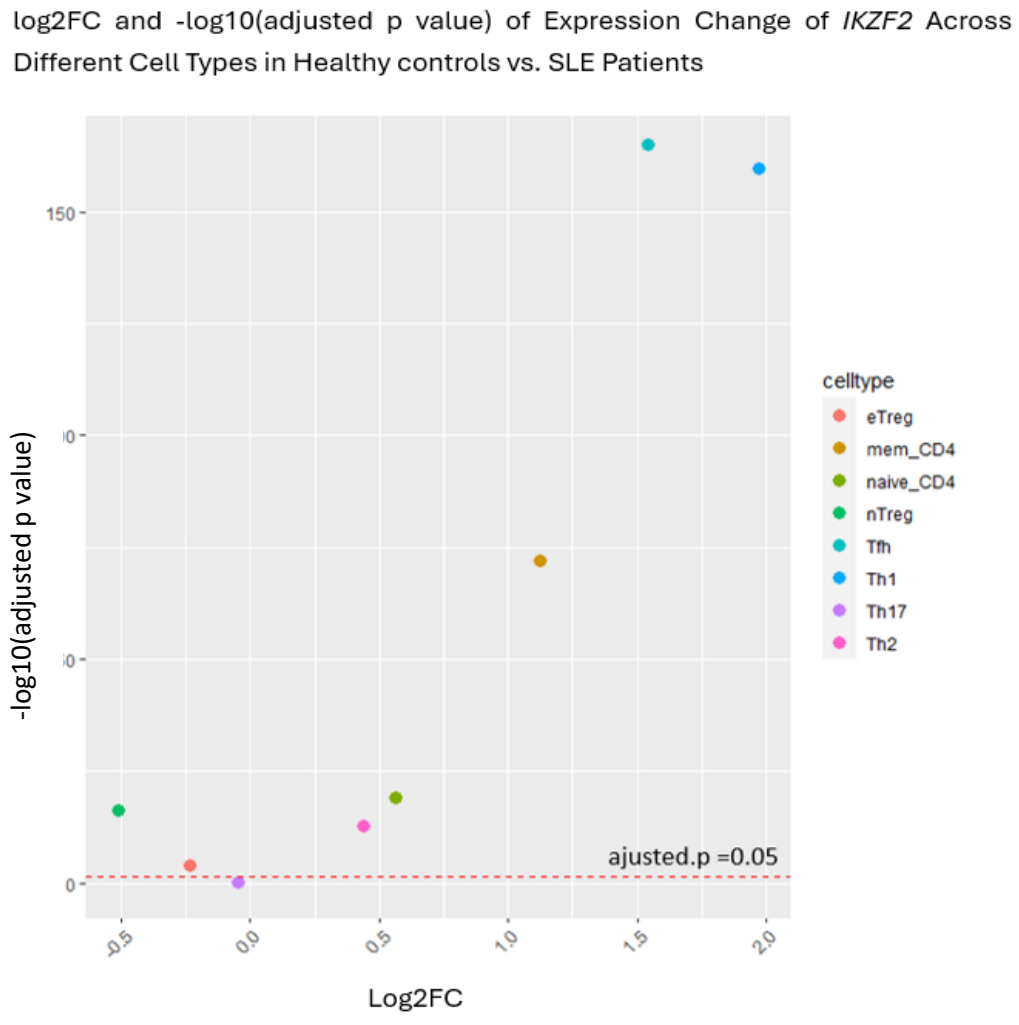
6.3 Results

6.3.1 Change of Expression of *IKZF2* in different T cell subtypes in SLE compared to Healthy controls

The changes of the expression of *IKZF2* in different T cell subtypes are shown in **Figure 6. 1.** The expression of *IKZF2* underwent significant changes in numerous T cell subtypes between controls and SLE patients. Specifically, there was a notable increase in *IKZF2* expression in SLE patients within CD4+ naïve T cells, CD4+ memory T cells, Th1 cells, Th2 cells, T follicular helper cells. Conversely, there was a significant decrease in *IKZF2* expression in SLE patients within

effector regulatory T cells and naïve regulatory T cells. However, the observed changes in *IKZF2* expression did not reach significance in Th17 cells. The expression changes in three specific cell types, namely Th1 cells, T follicular helper cells, and CD4+ memory T cells, all exhibited a |log2FC| > 1, indicating increased expression of at least 2-fold in SLE patients. Among these, the most substantial change, characterized by the highest |log2FC| value, was observed in Th1 cells.

Figure 6. 1 Differential Expression of *IKZF2* Across Various T Cell Subtypes between Healthy controls and SLE patients



The dotted red line represents the threshold for statistical significance (adjusted p-value = 0.05), with points above this line indicating significant changes in gene expression. The x-axis shows the log fold change (log2FC) in gene expression, while the y-axis represents the negative logarithm of the adjusted p-value (-log10(adjusted p-value)). Significant expression changes with a |log2FC| > 1 are most prominent in Th1 cells, Tfh cells, and CD4+ memory T cells, as indicated by their position on the plot. Th17 cells show changes in expression that do not cross the significance threshold. In this context, log2FC > 0 indicates an increase in gene expression in SLE, while log2FC < 0 indicates a decrease in gene expression in SLE. eTreg (Red): Effector Regulatory T Cells, mem_CD4 (Orange): CD4+ Memory T Cells, naïve _CD4 (Green): CD4+ Naïve T Cells, nTreg (Dark Green): Naïve Regulatory T Cells, Tfh (Cyan): T Follicular Helper Cells, Th1 (Blue): T Helper 1 Cells, Th17 (Purple): T Helper 17 Cells, Th2 (Pink): T Helper 2 Cells.

6.3.2 Change of Expression of Helios's target genes in different T cell subtypes in SLE compared to Healthy Controls

Among 875 predicted target genes of Helios in CD4+ naïve T cells, 273 overlap with differentially expressed genes (DEGs) (adjusted p-value <0.05) between healthy controls and SLE patients. Of these, 145 genes are upregulated (indicated as log2FC > 0), while 128 are downregulated (log2FC < 0) in SLE patients (see **Supplementary Table 5.1** https://www.dropbox.com/scl/fi/9iffjv5kg4f29kb4xvs6j/Supplementary-Table-5.1.cd4_predicted_targetgenes.xlsx?rlkey=keza81xiin19vcmcuwwdsr6i4&st=6606i5ri&dl=0). Among 1193 predicted target genes of Helios in Th1 cells, 518 overlap with DEGs between healthy controls and SLE patients. Specifically, 330 genes are upregulated (indicated as log2FC > 0), and 188 are downregulated (log2FC < 0) in SLE patients (see **Supplementary Table 5.2** https://www.dropbox.com/scl/fi/yvmpl7p5d5hlqvoxcqf59/Supplementary-Table-5.2.Th1cells_predicted_targetgenes.xlsx?rlkey=bjnwxv8zv0mupbb9kp9gwzqju&st=2bqpvczt&dl=0). Among 843 predicted target genes of Helios in Th17 cells, 496 overlap with DEGs between

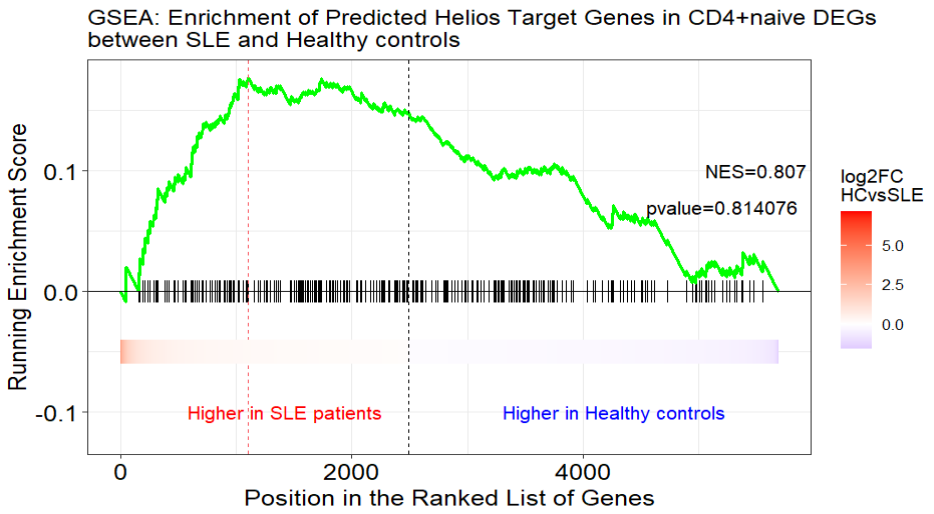
SLE patients and healthy controls. Here, 124 genes are upregulated (indicated as log2FC > 0), whereas 372 are downregulated (log2FC < 0) in SLE patients (see **Supplementary Table 5.3** https://www.dropbox.com/scl/fi/7ykb3nvcswwozh91fxgu0/Supplementary-Table-5.3.Th17cells_predicted_targetgenes.xlsx?rlkey=0qfawwuuiqrk7sj3avqlhllok&st=152ftaxr&dl=0).

6.3.3 Gene set enrichment analysis of predicted target genes of Helios in DEGs between Healthy controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells

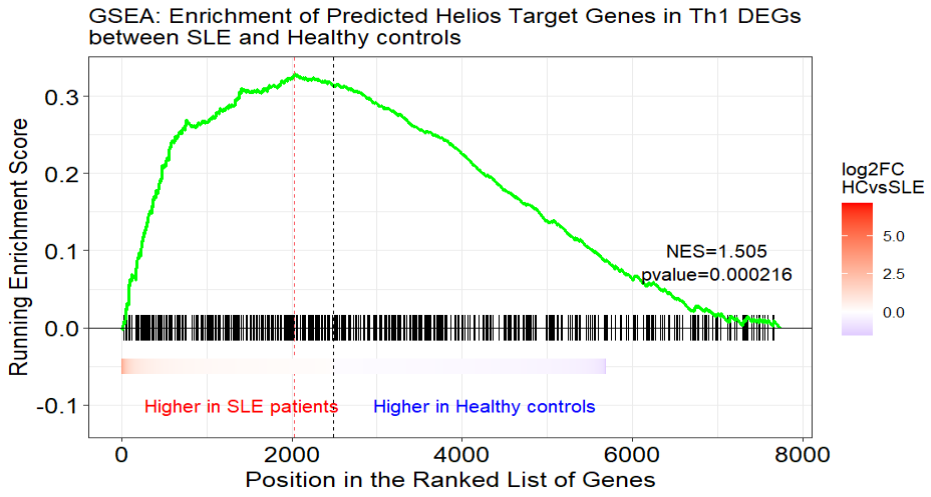
The predicted target genes in CD4+ naïve cells didn't show a significant enrichment for the DEGs between healthy controls versus SLE patients in CD4+ naïve cells (**Figure 6. 2 A**). The predicted target genes in Th1 cells exhibited a remarkable enrichment pattern at the top of the DEGs associated with healthy controls versus SLE patients in Th1 cells, resulting in a normalized enrichment score (NES) of 1.505 and a highly significant p-value of 0.000216 (**Figure 6. 2 B**). Similarly, the predicted target genes in Th17 cells demonstrated significant enrichment at the top of the DEGs between healthy controls and SLE patients in Th17 cells, yielding an NES of 1.41 with a p-value of 0.036 (**Figure 6. 2 C**).

Figure 6. 2 Results of gene set enrichment analysis of predicted target genes of Helios in DEGs between Healthy controls and SLE in CD4+naïve T cells, Th1 cells and Th17 cells

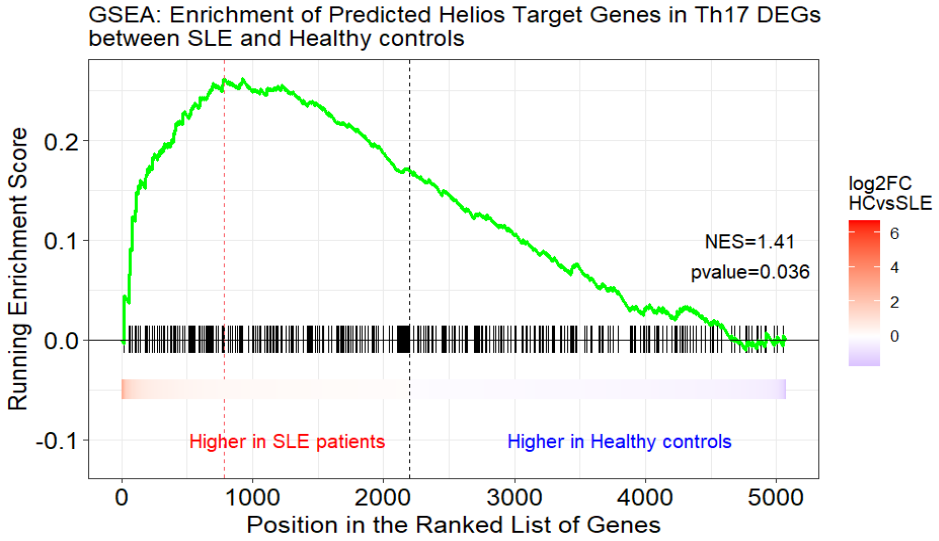
A


GSEA: Enrichment of Predicted Helios Target Genes in CD4+naive DEGs between SLE and Healthy controls
Running Enrichment Score
0.1
0.0
-0.1
NES=0.807
pvalue=0.814076
log2FC HCvsSLE
5.0
2.5
0.0
Higher in SLE patients
Higher in Healthy controls
0
2000
4000
Position in the Ranked List of Genes

B


GSEA: Enrichment of Predicted Helios Target Genes in Th1 DEGs between SLE and Healthy controls
Running Enrichment Score
0.3
0.2
0.1
0.0
-0.1
NES=1.505
pvalue=0.000216
log2FC HCvsSLE
5.0
2.5
0.0
Higher in SLE patients
Higher in Healthy controls
0
2000
4000
6000
8000
Position in the Ranked List of Genes

C


GSEA: Enrichment of Predicted Helios Target Genes in Th17 DEGs between SLE and Healthy controls
Running Enrichment Score
0.2
0.1
0.0
-0.1
NES=1.41
pvalue=0.036
log2FC HCvsSLE
6
4
2
0
Higher in SLE patients
Higher in Healthy controls
0
1000
2000
3000
4000
5000
Position in the Ranked List of Genes

A) Enrichment plot for CD4+ naïve T cells. It shows the enrichment score (ES) across the ranked list of differentially expressed genes (DEGs) between healthy controls and SLE patients in CD4+ naïve T cells, with a peak score indicating where the predicted Helios target genes are most concentrated. However, the enrichment is not significant (P>0.05). B) Enrichment plot for Th1 cells. It highlights the distribution of predicted Helios target genes within the ranked DEGs, with the peak representing significant enrichment among the top ranked DEGs between healthy controls and SLE patients in Th1 cells. C) Enrichment plot for Th17 cells. Like the other panels, this shows the enrichment of Helios target genes, with the peak indicating significant enrichment at the top of the DEGs list between healthy controls and SLE patients in Th17 cells. In each panel: The x-axis shows the position in the ranked list of DEGs, with genes higher in SLE patients on the left (indicated by the red arrow) and genes higher in healthy controls on the right (indicated by the blue arrow). The y-axis (left side) shows the running enrichment score for the set of Helios target genes. The vertical red dotted line indicates the point of maximum ES. The Vertical black dashed line indicates the position where log2FC is closest to zero. Black Bars below the plot show the position of Helios target genes in the ranked list of DEGs. Colour bar at the bottom of the plot shows the log2 fold change (log2FC) of the genes, with red indicating higher expression in SLE patients and blue indicating higher expression in healthy controls (HC). NES (Normalized Enrichment Score) and p-values are provided for each cell type, indicating the statistical significance of the enrichment.

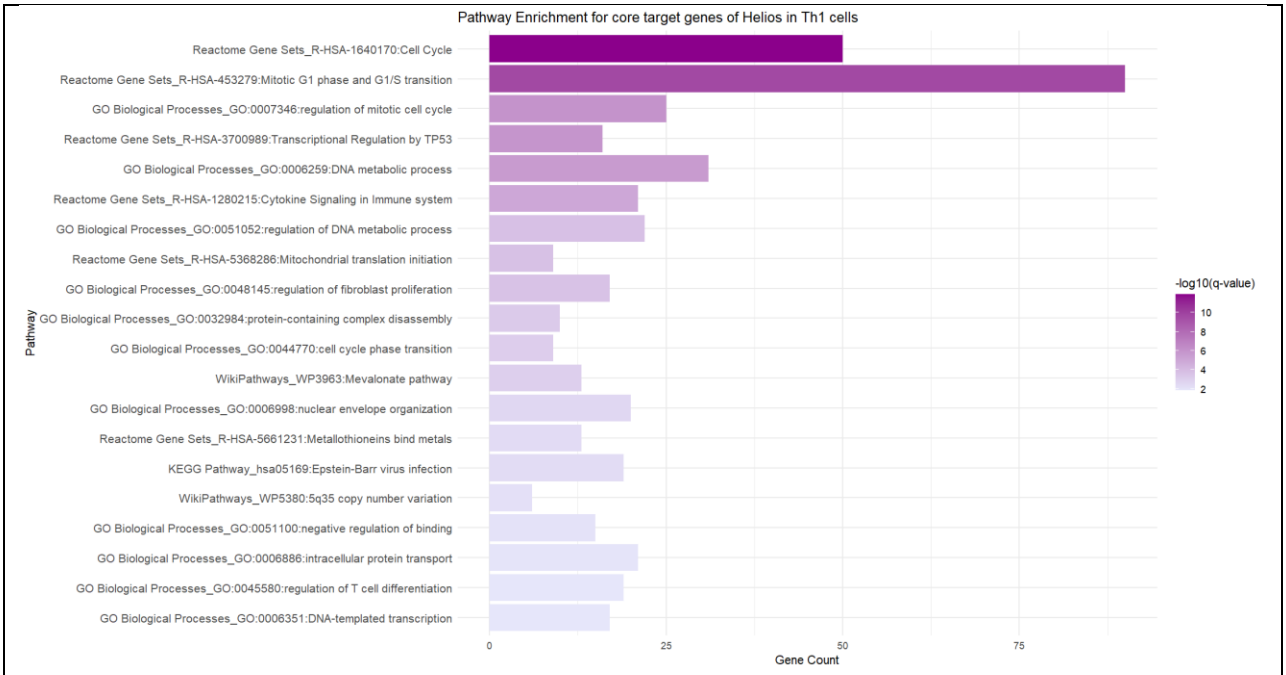
6.3.4 Functional annotation and Hub genes identification of core enrichment genes in Th1 cells

190 predicted target genes of Helios in Th1 cells were found to be core enriched to DEGs between healthy controls and SLE patients. This indicates that these genes are the primary contributors to the significant enrichment signal observed in the GSEA, suggesting they are the most important set of Helios target genes in Th1 cells that that significantly overlap with the differentially expressed genes between SLE and healthy controls. Further functional annotation

analysis indicates these genes are enriched for the pathways include cell cycle regulation, mitotic G1 phase and G1/S transition, regulation of the mitotic cell cycle, transcriptional regulation by TP53, DNA metabolic processes, cytokine signaling in the immune system, and regulation of DNA metabolic process. Other significant pathways identified are mitochondrial translation initiation, regulation of fibroblast proliferation, protein-containing complex disassembly, cell cycle phase transition, the mevalonate pathway, nuclear envelope organization, metallothioneins binding metals, Epstein-Barr virus infection, 5q35 copy number variation, negative regulation of binding, and intracellular protein transport. Additionally, pathways related to T cell differentiation and DNA-templated transcription were also enriched.

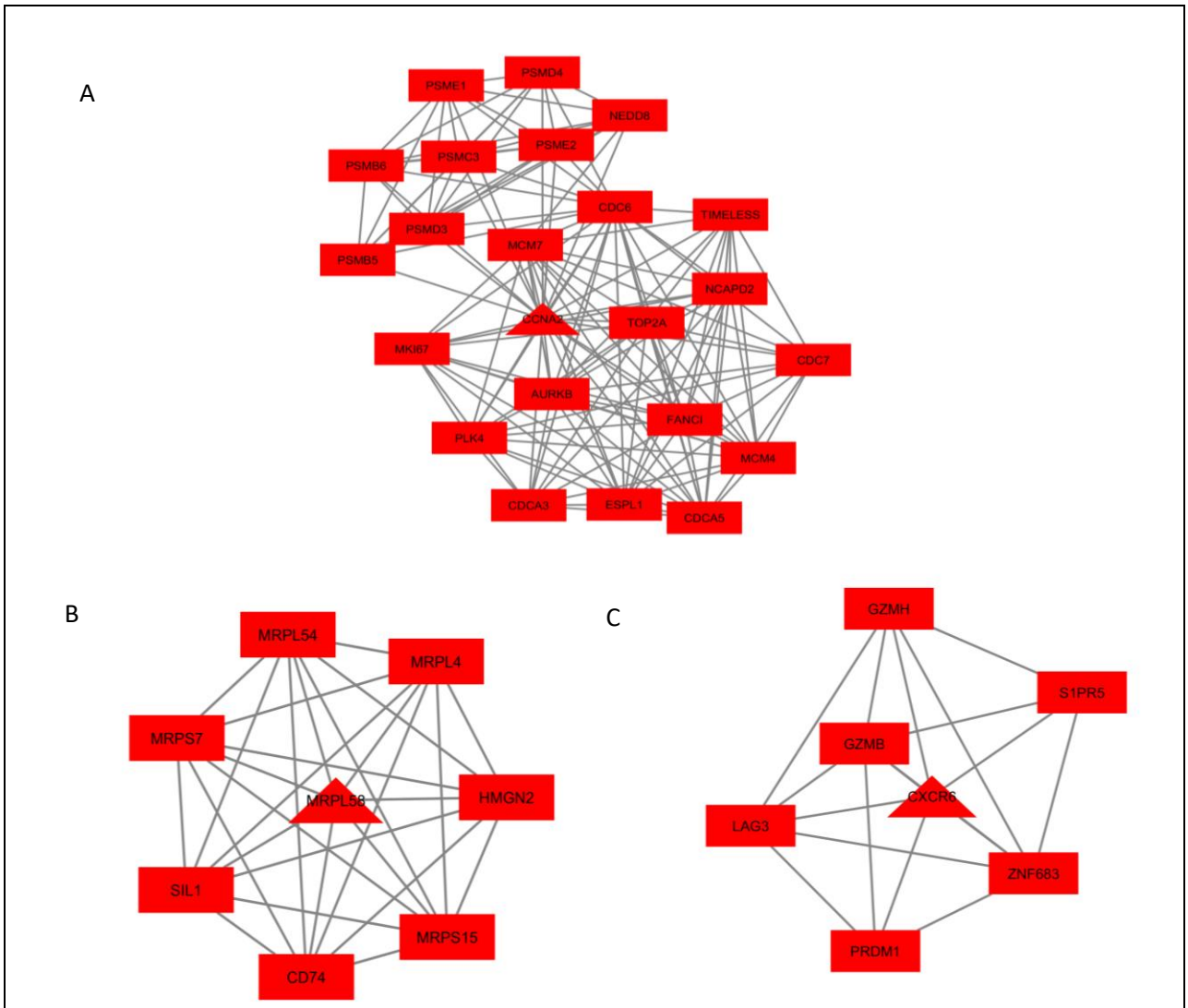
A network with 147 nodes and 583 edges. Three modules were found by MCODE plugin in Cytoscape with the following criteria: degree cutoff of 2, node score cutoff of 0.2, k-core of 2, maximum depth of 100, and nodes>5. The module 1 has 23 nodes and 144 edges, of which the hub gene is *CCNA2* (Cyclin A2) (**Figure 6. 4 A**). The module 2 contains 8 nodes and 28 edges, of which the hub gene is *MRPL58* (mitochondrial ribosomal protein L58) (**Figure 6. 4 B**). The module 3 contains 7 nodes and 18 edges, of which the hub gene is *CXCR6* (CXC motif chemokine receptor 6) (**Figure 6. 4 C**).

Figure 6. 3 Pathway enrichment of core enriched Helios target genes in DEGs between HC and SLE in Th1 cells



The horizontal bars represent the different pathways enriched. The length of each bar corresponds to the gene count associated with each pathway, reflecting the number of genes from the core target list that are involved in that specific pathway. The color intensity of the bars indicates the significance level of enrichment, with darker shades of purple corresponding to higher -log10(p-value), hence more significant enriched Pathways at the top, with the longest bars and darkest colors, such as Cell Cycle and Mitotic G1 phase and G1/S transition, are the most significantly enriched and contain the highest number of Helios target genes.The negative logarithm of the p-value (-log10(p-value)) is used to assess the statistical significance of the pathway enrichment, with values presented in the color scale on the right, where a higher value indicates a lower p-value and, therefore, a more significant enrichment.

Figure 6. 4 Modules identified from the core enriched genes of predicted Helios target genes to DEGs between healthy controls and SLE patients in Th1 cells



Three modules within the network of Helios target genes that are core enriched in DEGs between healthy controls and SLE patients within Th1 cells. Nodes represent genes, and edges illustrate the interactions or relationships between the genes. All these genes are found to be upregulated in SLE compared to healthy controls, as indicated by the red color of the nodes. Central triangle nodes denote the hub genes within each module. These are identified as *CCNA2* (Cyclin A2) in Module 1, *MRPL58* (Mitochondrial Ribosomal Protein L58) in Module 2, and *CXCR6* (CXC Motif Chemokine Receptor 6) in Module 3.

Module 1 (A): Comprises 23 nodes and 144 edges, highlighting a dense network of gene interactions with *CCNA2* as the hub gene.

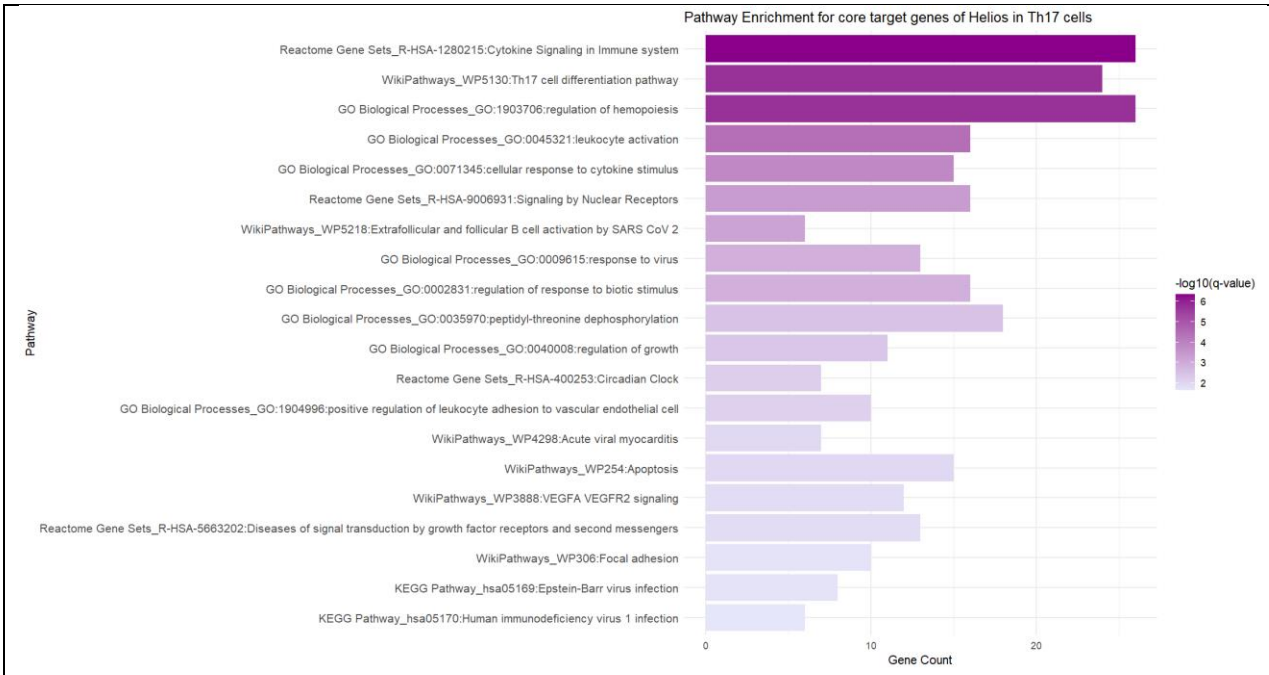
Module 2 (B): Consists of 8 nodes and 28 edges, with *MRPL58* as the central hub gene.

Module 3 (C): Contains 7 nodes and 18 edges, centered around *CXCR6* as the hub gene.

6.3.5 Hub genes identification of core enrichment genes in Th17 cells

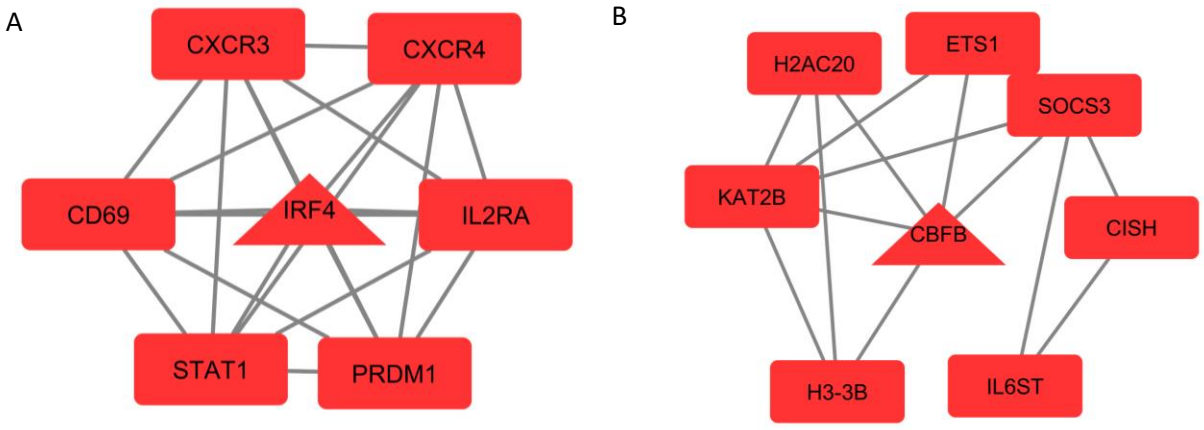
70 predicted target genes of Helios in Th17 cells were found to be core enriched to DEGs between healthy controls and SLE. A network with 52 nodes and 137 edges. Two modules were found by MCODE plugin in Cytoscape. The functional annotaion analysis reveals significant enrichment of these genes in pathways related to immune system regulation, particularly in cytokine signaling, hematopoiesis, Th17 cell differentiation, leukocyte activation, and response to cytokines. Key pathways also include signaling by nuclear receptors, response to viruses, regulation of growth, and apoptosis. The analysis identifies involvement in specific pathways like Epstein-Barr virus infection, HIV-1 infection, and VEGFA-VEGFR2 signaling. These findings highlight the integral role of these genes in various immune responses and regulatory mechanisms. Two modules were found by MCODE plugin in Cytoscape with the following criteria: degree cutoff of 2, node score cutoff of 0.2, k-core of 2, maximum depth of 100, and nodes >5. The module 1 has 7 nodes and 21 edges, of which the hub gene is *IRF4* (interferon regulatory factor 4) (**Figure 6. 4** A). The module 2 contains 8 nodes and 13 edges, of which the hub gene is *CBFB* (core-binding factor subunit beta) (**Figure 6. 4** B).

Figure 6. 5 Pathway enrichment of core enriched Helios target genes in DEGs between HC and SLE in Th1 cells



The horizontal bars represent the different pathways enriched. The length of each bar corresponds to the gene count associated with each pathway, reflecting the number of genes from the core target list that are involved in that specific pathway. The color intensity of the bars indicates the significance level of enrichment, with darker shades of purple corresponding to higher -log10(p-value), hence more significant enriched Pathways at the top, with the longest bars and darkest colors, such as Cytokine Signaling in Immune system and Th17 cell differentiation pathway, are the most significantly enriched and contain the highest number of Helios target genes.The negative logarithm of the p-value (-log10(p-value)) is used to assess the statistical significance of the pathway enrichment, with values presented in the color scale on the right, where a higher value indicates a lower p-value and, therefore, a more significant enrichment.

Figure 6. 6 Modules identified from the core enriched genes of predicted Helios target genes to DEGs between healthy controls and SLE patients in Th17 cells



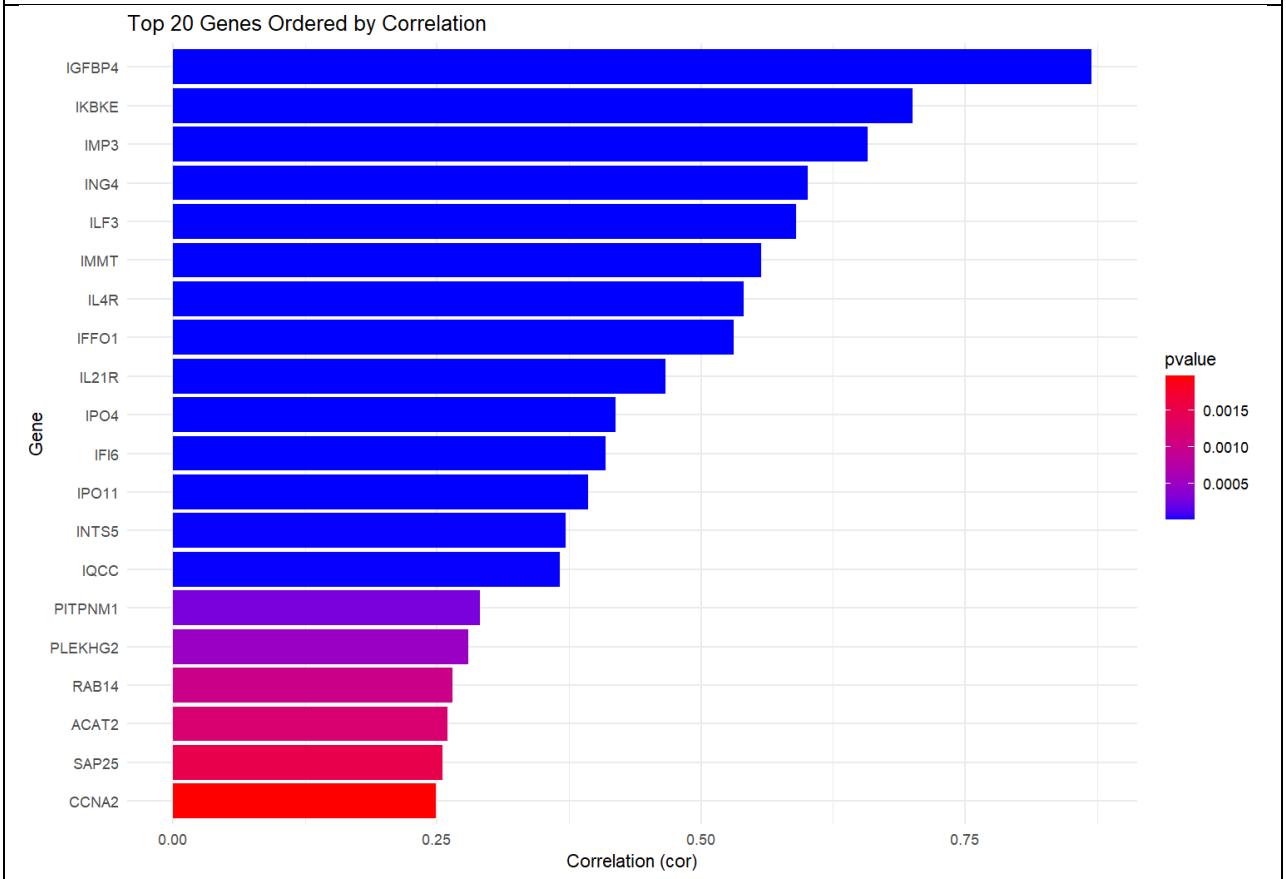
Three modules within the network of Helios target genes that are core enriched in DEGs between healthy controls and SLE patients within Th17 cells. Nodes represent genes, edges illustrate the interactions or relationships between the genes. All these genes are found to be upregulated in SLE compared to healthy controls, as indicated by the red color of the nodes. Central triangle nodes denote the hub genes within each module. These are identified as *IRF4* (Interferon Regulatory Factor 4) in Module 1, *CBFB* (Core-Binding Factor Subunit Beta) in Module 2.

6.3.6 Correlation between *IKZF2* and Helios' core predicted target genes related to SLE in Th1 and Th17 cells

Out of the 1193 predicted target genes of Helios in within Th1 cells, 103 demonstrate a significant correlation with *IKZF2*, with the criteria of a p-value < 0.05. The range of correlation coefficients for these genes spans from -0.17 to 0.87. The top 20 of these core target genes, based on their correlation strength, are displayed in **Figure 6.7.**

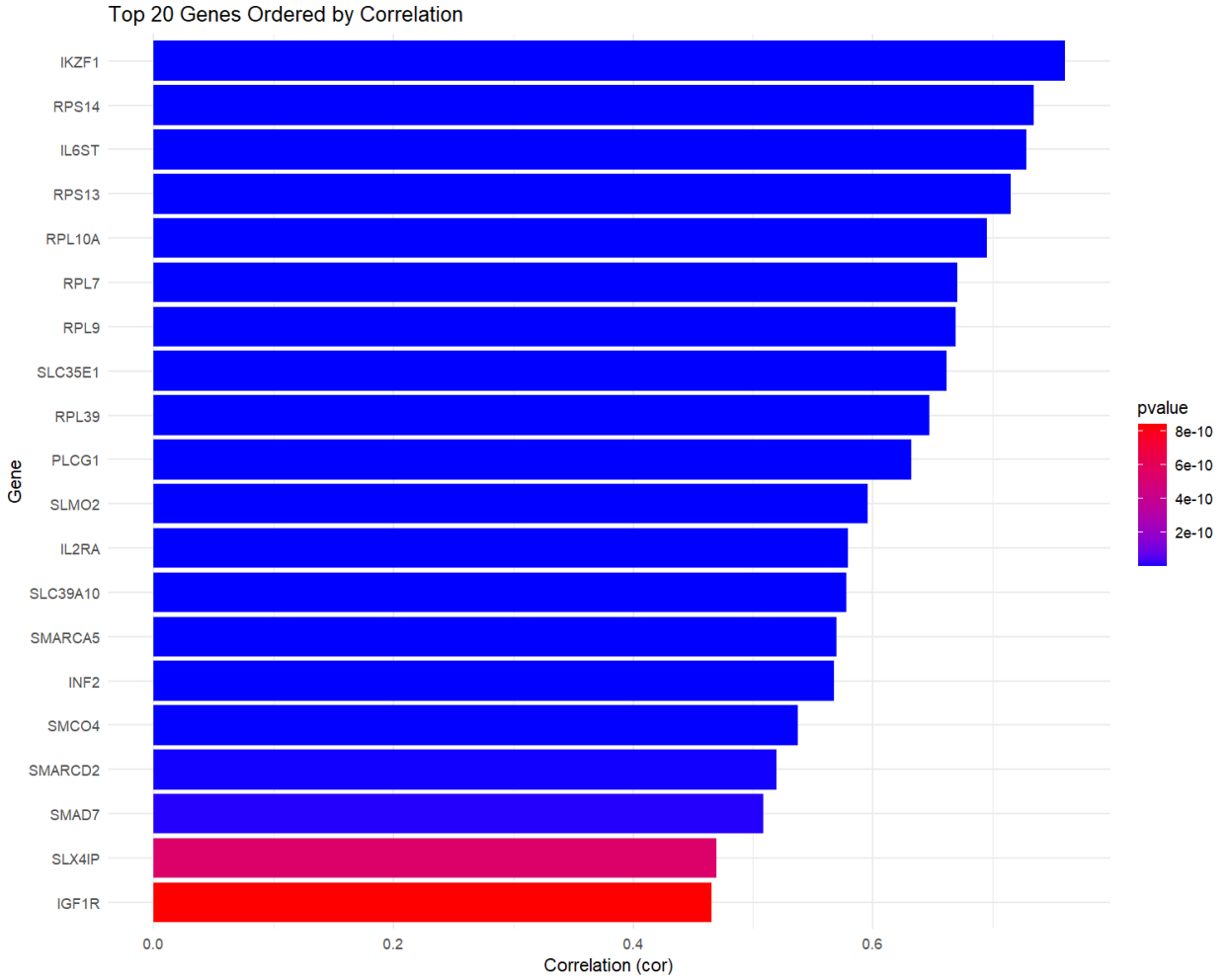
Out of the 843 predicted target genes of Helios in Th17 cells, 104 demonstrate a significant correlation with *IKZF2*, with the criteria of a p-value < 0.05 (**Figure 6. 8**). The range of correlation coefficients for these genes spans from -0.10 to 0.76.

Figure 6. 7 Top 20 predicted target genes of Helios in Th1 T cells based on the correlation with *IKZF2* in Th1 cells of SLE patients



Each bar indicates the correlation coefficient (cor) for a specific gene with *IKZF2*, with the length of the bar representing the strength of the correlation. The color of the bars corresponds to the p-value of the correlation, with blue shades indicating a higher level of significance and red shades indicating lower significance. The genes are ordered from top to bottom by decreasing correlation coefficient, with the highest correlation at the top. A p-value less than 0.05 is considered statistically significant.

Figure 6. 8 Top 20 predicted target genes of Helios in Th17 T cells based on the correlation with *IKZF2* in Th17 cells of SLE patients

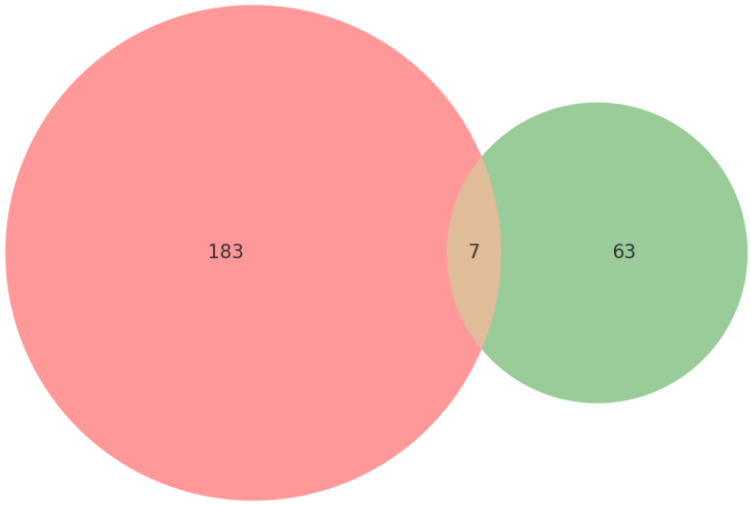


Each bar indicates the correlation coefficient (cor) for a specific gene with *IKZF2*, with the length of the bar representing the strength of the correlation. The color gradient from blue to red corresponds to the p-value associated with each correlation, with darker shades indicating lower p-values. Bars in dark blue signify correlations with p-values less than 0.01, and bars in red represent correlations with p-values approaching 0.02. All depicted correlations are statistically significant, with p-values less than 0.05.

6.3.7 Shared Th1 and Th17 cells core enriched target genes of Helios in DEGs between controls and SLE

There are 7 genes shared between two gene sets: Core Predicted Target Genes of Helios Enriched in Th1 Cells DEGs between controls vs. SLE and Core Predicted Target Genes of Helios Enriched in Th17 Cells DEGs between controls vs. SLE. The shared genes are: *IFI6* (interferon alpha inducible protein 6)*, FKBP5* (FKBP5 FKBP prolyl isomerase 5), *TRIB1* (tribbles pseudokinase 1), *PRDM1* (PR/SET Domain 1), *TRAT1* (T Cell Receptor Associated Transmembrane Adaptor 1)*, LINC00426, CCND3* (cyclin D3).

Figure 6. 9 Venn Diagram Illustrating Shared Target Genes of Helios in Th1 and Th17 Cells Related to SLE



Red Circle represents the core predicted target genes of Helios enriched in Th1 cell DEGs between healthy controls and SLE patients.Green Circle denotes the core predicted target genes of Helios enriched in Th17 cell DEGs between healthy controls and SLE patients. Overlap Area (in Tan) Indicates the seven genes shared between the two gene sets, which are *IFI6, FKBP5, TRIB1, PRDM1, TRAT1, LINC00426*, and *CCND3*. Numbers Show the count of unique and shared genes within each subset or between subsets, highlighting the intersection and differences in gene enrichment related to Helios's influence in these T cell types.

6.3.8 Expression Changes and Correlation with *IKZF2* of Recognized Key Predicted Target Genes of Helios related to SLE

The key predicted target genes of Helios related to SLE were identified by combining three groups: shared core enrichment predicted target genes associated with differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients, hub target genes of Helios specific to SLE-related targets in Th1 cells, and hub target genes of Helios in the context of SLE in Th17 cells. The log2 fold changes of these genes, along with their correlation with *IKZF2*, have been extracted and are presented in **Table 6.1**.

Table 6. 1 Expression Changes and Correlation with *IKZF2* of Recognized Key Target Genes of Helios in Th1 and Th17 Cells

Gene	Info	log2FC in Th1	log2FC in Th17	Correlation Coefficient with *IKZF2* in Th1	Correlation with *IKZF2* in Th17
IFI6	Shared core genes	**2.13**	**2.23**	**0.41**	**0.38**
FKBP5	Shared core genes	**1.42**	**1.00**	not sig	not sig
TRIB1	Shared core genes	**1.48**	**0.69**	not sig	not sig
PRDM1	Shared core genes	**0.41**	**0.69**	not sig	not sig
TRAT1	Shared core genes	**0.36**	**0.58**	not sig	not sig
LINC00426	Shared core genes	**0.46**	**0.33**	not sig	not sig

CCND3	Shared core genes	**0.42**	**0.32**	not sig	not sig
CCNA2	Hub gene in Th1cells	**1.90**	**0.25**	**0.25**	not sig
MRPL58	Hub gene in Th1cells	**0.42**	**0.12**	**0.19**	not sig
CXCR6	Hub gene in Th1cells	**0.93**	**-0.47**	not sig	not sig
IRF4	Hub gene in Th17 cells	**0.53**	**-0.15**	not sig	not sig
CBFB	Hub gene in Th17 cells	**0.30**	**0.28**	not sig	not sig

log2FC, log2 fold change (log2FC) of differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients. A positive log2FC indicates that the gene is upregulated in SLE patients, while a negative log2FC indicates downregulation. Log2FC values that reached significance (adjusted p value < 0.05 in DEG analysis) are shown in bold. Correlation Coefficient with *IKZF2*: Pearson Correlation Coefficient of the expression of the gene with the expression of *IKZF2*. A positive correlation coefficient means that as the expression of *IKZF2* increases, the expression of the gene also increases. A negative correlation coefficient means that as the expression of *IKZF2* increases, the expression of the other gene decreases. Values that reached significance (p value < 0.05 in Correlation test) are shown in bold. not sig, not significant (i.e p value >0.05) means the correlation is considered not statistically significant. Shared core genes: shared core enrichment predicted target genes associated with differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients. Hub gene in Th1cells: hub target genes of Helios specific to SLE-related targets in Th1 cells. Hub gene in Th1cells: hub target genes of Helios specific to SLE-related targets in Th17 cells

6.4 Discussion

The expression data of *IKZF2* across different T cell subtypes in SLE patients compared to controls indicates that *IKZF2* may play a significant role in SLE pathogenesis. The downregulation of *IKZF2* in regulatory T cells (Tregs), both natural (nTreg) and effector (eTreg),

with statistically significant false discovery rates (FDRs), suggests a possible disruption in immune regulation, a key factor in the development of SLE. The marked upregulation in Th1 and T follicular helper (Tfh) cells, cells known for their roles in promoting inflammation and B cell help respectively, points to a potential contribution of *IKZF2* in the inflammatory and autoantibody-producing processes of SLE. Additionally, the increased expression in naïve and memory CD4+ T cells could indicate a role in sustaining immune memory that may contribute to disease flares and progression, while the slight increase in Th2 cells could be involved in the antibody-mediated aspects of the disease. The minimal change observed in Th17 cells indicates that *IKZF2*'s role in these cells may not be central or might involve complex regulatory mechanisms not directly related to expression levels.

Gene set enrichment analysis (GSEA) of the predicted target genes of Helios to DEGs between healthy controls and SLE patients show the predicted target genes of Helios exhibited a remarkable enrichment pattern at the top of the DEGs between healthy controls versus SLE patients in Th1 cells and Th17 cells. The high NES and low p-value for DEGs in both cell types suggest that the dysregulation of Helios's target genes in Th1 cells and Th17 cells contribute to the pathogenesis of SLE. Enrichment at the top of the DEGs list suggests that Helios might play a role in activating or enhancing the expression of genes that contribute to the pathogenesis or progression of SLE in these two cell types. The analysis identified seven pivotal genes—*IFI6, FKBP5, TRIB1, PRDM1, TRAT1, LINC00426*, and *CCND3*—that are shared among the core predicted targets of Helios in Th1 and Th17 cell subsets. The presence of shared core genes between the predicted target genes of Helios in Th1 and Th17 cell DEGs in SLE patients compared to healthy controls suggests a fundamental role for Helios in influencing these genes across different T cell subsets that are critical in SLE pathogenesis. These genes play crucial roles in immune regulation and cell function. *IFI6,* recognized for its role in type I interferon signaling—a pathway crucial to immune responses—is implicated in the etiology and progression of SLE. Significant up-regulation of *IFI6* in various samples from SLE patients makes it as a potential biomarker for the disease [279, 280]. *FKBP5*, which regulates glu*coco*rticoid

receptor function and stress responses, could affect immune system reactivity in SLE. *FKBP5* gene polymorphisms has been found to be associated with depression and glucocorticoids efficacy of SLE patients[281]. *TRIB1* is implicated in signaling pathways governing cell survival and proliferation, influencing immune cell behaviour. *PRDM1 (BLIMP-1)* is a key regulator in the development of plasma cells from B cells and affects T cell regulation, pointing to a role in immune tolerance and autoantibody production. *TRAT1* contributes to T cell receptor signalling, essential for T cell activation and function. *LINC00426*, a long non-coding RNA, may influence gene expression and impact immune function. *CCND3* plays a critical role in cell cycle control, particularly in lymphocyte proliferation and differentiation. The commonality of these genes suggests a role for Helios in mediating inflammatory processes across T cell subtypes, potentially contributing to the autoimmune phenomena observed in SLE and providing a link between Helios function and disease progression. Detailed studies and experimental validation on how Helios regulate these target genes in healthy controls and SLE patients are needed.

I identified 190 Helios target genes are core enriched in differentially expressed genes (DEGs) between healthy controls and SLE patients, in Th1 cells, underscores Helios's significant role in SLE pathophysiology. The network, comprising 110 nodes and 321 edges, reveals a complex interaction pattern, suggesting a broad regulatory impact of Helios in Th1 cell function. This is further elucidated by six distinct modules identified using the MCODE plugin in Cytoscape, each with a specific hub gene such as *FANCI, MRPL4, PSMC3, ZNF683, EIF3I*, and *IFI6*. These modules, enriched in diverse pathways like cellular translation, RNA metabolism, stress response, mitochondrial translation, protein catabolism, T cell differentiation, cell killing, interferon signaling, and innate immune response, highlight the multifaceted nature of Helios's involvement. The enrichment in such varied pathways indicates Helios's influence in key aspects of Th1 cell function, from fundamental cellular processes to specific immune responses. Particularly notable is the role in regulating pathways critical for immune response, like T cell receptor signalling, cytokine interactions, and interferon pathways, pointing to Helios's central role in the autoimmune dynamics of SLE. This analysis suggests Helios's potential contribution

to both the development and exacerbation of autoimmune processes in SLE, highlighting its importance in Th1 cell dysregulation. The diverse range of pathways influenced by Helios and its target genes in Th1 cells not only provides deeper insight into the complex pathogenesis of SLE but also suggests potential therapeutic targets for modulating immune responses in the disease, offering new avenues for research and treatment strategies.

In Th17 cells, 70 of Helios' predicted target genes were found to be centrally contributed to differentially expressed genes (DEGs) between healthy controls and SLE patients. These genes form a complex network comprising 52 nodes and 137 edges, as revealed through network analysis. Functional annotation analysis of these core enriched genes underscored their significant enrichment in pathways pivotal to immune system regulation. Key among these were pathways involved in cytokine signaling, hematopoiesis, Th17 cell differentiation, leukocyte activation, and response to cytokines. Additionally, these genes were implicated in signaling by nuclear receptors, responses to viruses, regulation of growth, and apoptosis, highlighting their diverse roles in immune response and cellular regulation. Specific pathways like Epstein-Barr virus infection, HIV-1 infection, and VEGFA-VEGFR2 signaling were also identified, underlining the genes' involvement in varied biological processes and diseases. I delineated three distinct modules within the group of these genes and identified three hub genes within these modules, namely *IRF4, CBFB* and *CXCR6*. *MUM1 /IRF4* has been implicated in lupus pathogenesis, particularly through its involvement in dendritic cell dysfunction [240], and has been associated with specific gene expression signatures in SLE [241]. IRF4 deficiency has been observed to reduce lupus nephritis, albeit with increased systemic cytokine production, indicating its complex influence in SLE [242]. The role of IRF4 in SLE is further underscored by its association with differential microRNA expression patterns in CD4+ and CD19+ cells from asymptomatic SLE patients [243] and thalidomide has been found to exert anti-inflammatory effects in cutaneous lupus by inhibiting the IRF4/NF-κB and AMPK1/mTOR pathways [244]. Notable differential expression (log2FC = 0.67, adjusted.p <0.05) of *IRF4* following Helios knockdown in Jurkat T cells was observed in **Chapter 3 (Supplementary Table**

3.1). This significant expression, coupled with the identification of it as a hub gene in the Helios-regulated network specifically linked to differentially expressed genes in SLE compared to healthy controls, highlights the potential of Helios to influence SLE progression through its impact on *IRF4*. The CBFβ subunit, also known as CBFB, is a crucial component of the Runx1-Cbfβ transcription complex, which plays a significant role in the suppressive function of regulatory T cells and in maintaining Foxp3 expression. This complex is essential for the in vivo-suppressive function of FoxP3+ regulatory T cells, indicating its involvement in immune regulation [282]. The findings from the current study suggest it's involvement as a target gene of Helios in the differential expression of genes in SLE compared to healthy controls. Future studies should explore the regulatory relationship between Helios and CBFB in greater detail. This could involve studying how Helios influences *CBFB* expression and function especially in Treg cells and Th17 cells, and how this, in turn, affects the transcriptional landscape and immune cell behaviour in SLE.

Shared core enrichment predicted target genes associated with differentially expressed genes (DEGs) between healthy controls (HC) and SLE patients, hub target genes of Helios specific to SLE-related targets in Th1 cells, and hub target genes of Helios specific to SLE-related targets in Th1 cells made a list of key predicted target genes of Helios in these two cell types. This list was curated to prioritize genes based on their differential expression across these two cell types and their correlation with the transcription factor *IKZF2*. *IFI6*, a shared core gene, demonstrates heightened expression in SLE, as indicated by its log2FC values of 2.13 in Th1 and 2.23 in Th17 cells, coupled with a significant correlation with *IKZF2* (0.41 in Th1 and 0.38 in Th17 cells). Similarly, *CCNA2*, identified as a hub gene in Th1 cells, not only shows notable expression changes in SLE compared to healthy controls but also exhibits a modest yet significant correlation with *IKZF2* (correlation coefficient of 0.25). These findings underscore the importance of these genes in the context of SLE and their potential regulatory interaction with *IKZF2.*

7. Conclusion

The study provides an in-depth analysis of the transcription factor Helios, encoded by *IKZF2*, and its role in T cells, also its impact on SLE. Helios is a key player in SLE pathogenesis, predominantly expressed in T cells. The study used bioinformatics to examine Helios binding sites and target genes in T cells to understand its function in T cells and its possible role in SLE.

Key findings include the identification of 5,068 Helios binding sites in Jurkat T cells, linked to 1444 genes, highlighting Helios's extensive role in gene regulation. After Helios knockdown, changes in 1,072 genes were observed, affecting processes like cholesterol biosynthesis, apoptosis, and T cell receptor regulation. Shared pathways enriched for these two sets of genes emphasizes the role of Helios in diverse range of biological processes related to maintaining genomic integrity and cellular homeostasis, transcriptional regulation and immune system-related pathways.

Gene Set Enrichment Analysis was used to link the Helios with SLE by exploring the enrichment of predicted target genes of Helios in the differentially expression pattern of SLE. The study revealed the enrichment of Helios' predicted target genes at the top of the list of differentially expressed genes between healthy controls and SLE patients in Th1 and Th17 cells, indicating its potential role in the development of this autoimmune disease within these cell types. Notably, 190 and 70 Helios' predicted target genes were identified in Th1 and Th17 cells, respectively, as central contributors to these differentially expressed genes. Key hub genes associated with SLE, identified using tools like MCODE and Cytohubba in Cytoscape, included *CCNA2, MRPL58, CXCR6* in Th1 cells, and IRF4 and CBFB in Th17 cells. Moreover, seven genes – *IFI6, FKBP5, TRIB1, PRDM1, TRAT1, LINC00426*, and *CCND3* – were core predicted target genes of Helios in both

Th1 and Th17 cells, suggesting their fundamental role in SLE pathogenesis. Gene Set Enrichment Analysis revealed that the predicted target genes were enriched at the "top" of the list of differentially expressed genes when comparing healthy controls and SLE patients in both Th1 and Th17 cells (with a significance threshold set at p-value <0.05).

The study's findings underscore the significant role of Helios in regulating gene expression and immune function in T cells, particularly in the context of SLE. These insights pave the way for further research into the molecular mechanisms underlying SLE and highlight the potential of targeting Helios and its network of genes for therapeutic interventions in this complex autoimmune disease.

8. Reference

1.Wang L, Wang F-S, Gershwin ME**.** Human autoimmune diseases: a comprehensive update. Journal of Internal Medicine. 2015;278(4):369-95.

2.Barber MRW, Falasinnu T, Ramsey-Goldman R, et al. The global epidemiology of SLE: narrowing the knowledge gaps. Rheumatology. 2023;62(Supplement_1):i4-i9.

3.Lim SS, Helmick CG, Bao G, et al. Racial Disparities in Mortality Associated with Systemic Lupus Erythematosus - Fulton and DeKalb Counties, Georgia, 2002-2016. MMWR Morb Mortal Wkly Rep. 2019;68(18):419-22.

4.Tsokos GC**.** Systemic lupus erythematosus. N Engl J Med. 2011;365(22):2110-21.

5.Lahita RG**.** The role of sex hormones in systemic lupus erythematosus. Current Opinion in Rheumatology. 1999;11(5):352-6.

6.Guga S, Wang Y, Graham DC, et al. A review of genetic risk in systemic lupus erythematosus. Expert Rev Clin Immunol. 2023;19(10):1247-58.

7.Morris DL, Sheng Y, Zhang Y, et al. Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. Nat Genet. 2016;48(8):940-6.

8.López-Cortegano E, Caballero A**.** Inferring the Nature of Missing Heritability in Human Traits Using Data from the GWAS Catalog. Genetics. 2019;212(3):891-904.

9.MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. 2017;45(D1):D896-d901.

10.Almlöf JC, Nystedt S, Leonard D, et al. Whole-genome sequencing identifies complex contributions to genetic risk by variants in genes causing monogenic systemic lupus erythematosus. Hum Genet. 2019;138(2):141-50.

11.Chen L, Wang YF, Liu L, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet. 2020;29(10):1745-56.

12.Wang Y-F, Zhang Y, Lin Z, et al. Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. Nature Communications. 2021;12(1):772.

13.Knevel R, le Cessie S, Terao CC, et al. Using genetics to prioritize diagnoses for rheumatology outpatients with inflammatory arthritis. Sci Transl Med. 2020;12(545).

14.Barbhaiya M, Costenbader KH**.** Environmental exposures and the development of systemic lupus erythematosus. Curr Opin Rheumatol. 2016;28(5):497-505.

15.Fan Y-H, Leong P-Y, Chiou J-Y, et al. Association between endometriosis and risk of systemic lupus erythematosus. Scientific Reports. 2021;11(1):532.

16.Hsiao Y-P, Tsai J-D, Muo C-H, et al. Atopic Diseases and Systemic Lupus Erythematosus: An Epidemiological Study of the Risks and Correlations. International Journal of Environmental Research and Public Health. 2014;11(8):8112-22.

17.Wongtrakul W, Charoenngam N, Ponvilawan B, et al. Allergic rhinitis and risk of systemic lupus erythematosus: A systematic review and meta-analysis. Int J Rheum Dis. 2020;23(11):1460-7.

18.Gagliano Taliun SA, Evans DM**.** Ten simple rules for conducting a mendelian randomization study. PLoS Comput Biol. 2021;17(8):e1009238.

19.Bae SC, Lee YH**.** Causal association between periodontitis and risk of rheumatoid arthritis and systemic lupus erythematosus: a Mendelian randomization. Zeitschrift für Rheumatologie. 2020;79(9):929-36.

20.Inamo J. Association between celiac disease and systemic lupus erythematosus: a Mendelian randomization study. Rheumatology (Oxford). 2020;59(9):2642-4.
21.Xiang K, Wang P, Xu Z, et al. Causal Effects of Gut Microbiome on Systemic Lupus Erythematosus: A Two-Sample Mendelian Randomization Study. Frontiers in Immunology. 2021;12.
22.Bae SC, Lee YH. Coffee consumption and the risk of rheumatoid arthritis and systemic lupus erythematosus: a Mendelian randomization study. Clin Rheumatol. 2018;37(10):2875-9.
23.Wang P, Dan YL, Wu Q, et al. Non-causal effects of smoking and alcohol use on the risk of systemic lupus erythematosus. Autoimmun Rev. 2021;20(9):102890.
24.Jiang X, Zhu Z, Manouchehrinia A, et al. Alcohol Consumption and Risk of Common Autoimmune Inflammatory Diseases-Evidence From a Large-Scale Genetic Analysis Totaling 1 Million Individuals. Front Genet. 2021;12:687745.
25.Sang N, Gao RC, Zhang MY, et al. Causal Relationship Between Sleep Traits and Risk of Systemic Lupus Erythematosus: A Two-Sample Mendelian Randomization Study. Front Immunol. 2022;13:918749.
26.Mo X, Guo Y, Qian Q, et al. Mendelian randomization analysis revealed potential causal factors for systemic lupus erythematosus. Immunology. 2020;159(3):279-88.
27.Tas SW, Quartier P, Botto M, et al. Macrophages from patients with SLE and rheumatoid arthritis have defective adhesion in vitro, while only SLE macrophages have impaired uptake of apoptotic cells. Ann Rheum Dis. 2006;65(2):216-21.
28.Hejrati A, Rafiei A, Soltanshahi M, et al. Innate immune response in systemic autoimmune diseases: a potential target of therapy. Inflammopharmacology. 2020;28(6):1421-38.
29.Li F, Yang Y, Zhu X, et al. Macrophage Polarization Modulates Development of Systemic Lupus Erythematosus. Cell Physiol Biochem. 2015;37(4):1279-88.
30.Bruschi M, Bonanni A, Petretto A, et al. Neutrophil Extracellular Traps Profiles in Patients with Incident Systemic Lupus Erythematosus and Lupus Nephritis. J Rheumatol. 2020;47(3):377-86.
31.Villanueva E, Yalavarthi S, Berthier CC, et al. Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. J Immunol. 2011;187(1):538-52.
32.Sutanto H, Yuliasih Y. Disentangling the Pathogenesis of Systemic Lupus Erythematosus: Close Ties between Immunological, Genetic and Environmental Factors. Medicina (Kaunas). 2023;59(6).
33.Jeremic I, Djuric O, Nikolic M, et al. Neutrophil extracellular traps-associated markers are elevated in patients with systemic lupus erythematosus. Rheumatology International. 2019;39(11):1849-57.
34.Klarquist J, Zhou Z, Shen N, et al. Dendritic Cells in Systemic Lupus Erythematosus: From Pathogenic Players to Therapeutic Tools. Mediators Inflamm. 2016;2016:5045248.
35.Robak E, Smolewski P, Woźniacka A, et al. Clinical significance of circulating dendritic cells in patients with systemic lupus erythematosus. Mediators Inflamm. 2004;13(3):171-80.
36.Farkas L, Beiske K, Lund-Johansen F, et al. Plasmacytoid dendritic cells (natural interferon-alpha/beta-producing cells) accumulate in cutaneous lupus erythematosus lesions. Am J Pathol. 2001;159(1):237-43.
37.Fiore N, Castellano G, Blasi A, et al. Immature myeloid and plasmacytoid dendritic cells infiltrate renal tubulointerstitium in patients with lupus nephritis. Mol Immunol. 2008;45(1):259-65.
38.Tucci M, Quatraro C, Lombardi L, et al. Glomerular accumulation of plasmacytoid dendritic cells in active lupus nephritis: role of interleukin-18. Arthritis Rheum. 2008;58(1):251-62.
39.Mok MY. Tolerogenic dendritic cells: role and therapeutic implications in systemic lupus erythematosus. Int J Rheum Dis. 2015;18(2):250-9.

40.Postal M, Vivaldo JF, Fernandez-Ruiz R, et al. Type I interferon in the pathogenesis of systemic lupus erythematosus. Curr Opin Immunol. 2020;67:87-94.
41.Niewold TB, Hua J, Lehman TJ, et al. High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. Genes Immun. 2007;8(6):492-502.
42.Oke V, Gunnarsson I, Dorschner J, et al. High levels of circulating interferons type I, type II and type III associate with distinct clinical features of active systemic lupus erythematosus. Arthritis Res Ther. 2019;21(1):107.
43.Baechler EC, Batliwalla FM, Karypis G, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A. 2003;100(5):2610-5.
44.Bennett L, Palucka AK, Arce E, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J Exp Med. 2003;197(6):711-23.
45.Yao Y, Richman L, Higgs BW, et al. Neutralization of interferon-alpha/beta-inducible genes and downstream effect in a phase I trial of an anti-interferon-alpha monoclonal antibody in systemic lupus erythematosus. Arthritis Rheum. 2009;60(6):1785-96.
46.Obermoser G, Pascual V**.** The interferon-alpha signature of systemic lupus erythematosus. Lupus. 2010;19(9):1012-9.
47.Feng Y, Yang M, Wu H, et al. The pathological role of B cells in systemic lupus erythematosus: From basic research to clinical. Autoimmunity. 2020;53(2):56-64.
48.Iwata S, Tanaka Y**.** B-cell subsets, signaling and their roles in secretion of autoantibodies. Lupus. 2016;25(8):850-6.
49.Jacobi AM, Reiter K, Mackay M, et al. Activated memory B cell subsets correlate with disease activity in systemic lupus erythematosus: delineation by expression of CD27, IgD, and CD95. Arthritis Rheum. 2008;58(6):1762-73.
50.Kang N, Liu X, You X, et al. Aberrant B-Cell Activation in Systemic Lupus Erythematosus. Kidney Dis (Basel). 2022;8(6):437-45.
51.Yurasov S, Wardemann H, Hammersen J, et al. Defective B cell tolerance checkpoints in systemic lupus erythematosus. J Exp Med. 2005;201(5):703-11.
52.Tsokos GC, Lo MS, Costa Reis P, et al. New insights into the immunopathogenesis of systemic lupus erythematosus. Nat Rev Rheumatol. 2016;12(12):716-30.
53.Crotty S**.** T follicular helper cell differentiation, function, and roles in disease. Immunity. 2014;41(4):529-42.
54.Kim SJ, Lee K, Diamond B**.** Follicular Helper T Cells in Systemic Lupus Erythematosus. Front Immunol. 2018;9:1793.
55.Liarski VM, Kaverina N, Chang A, et al. Cell distance mapping identifies functional T follicular helper cells in inflamed human renal tissue. Sci Transl Med. 2014;6(230):230ra46.
56.Choi JY, Ho JH, Pasoto SG, et al. Circulating follicular helper-like T cells in systemic lupus erythematosus: association with disease activity. Arthritis Rheumatol. 2015;67(4):988-99.
57.Le Coz C, Joublin A, Pasquali JL, et al. Circulating TFH subset distribution is strongly affected in lupus patients with an active disease. PLoS One. 2013;8(9):e75319.
58.Ohl K, Tenbrock K**.** Regulatory T cells in systemic lupus erythematosus. Eur J Immunol. 2015;45(2):344-55.
59.Dolff S, Bijl M, Huitema MG, et al. Disturbed Th1, Th2, Th17 and T(reg) balance in patients with systemic lupus erythematosus. Clin Immunol. 2011;141(2):197-204.
60.Giang S, La Cava A**.** Regulatory T Cells in SLE: Biology and Use in Treatment. Curr Rheumatol Rep. 2016;18(11):67.

61.Lieberman LA, Tsokos GC**.** The IL-2 defect in systemic lupus erythematosus disease has an expansive effect on host immunity. J Biomed Biotechnol. 2010;2010:740619.
62.Chen PM, Tsokos GC**.** T Cell Abnormalities in the Pathogenesis of Systemic Lupus Erythematosus: an Update. Curr Rheumatol Rep. 2021;23(2):12.
63.Enghard P, Humrich JY, Rudolph B, et al. CXCR3+CD4+ T cells are enriched in inflamed kidneys and urine and provide a new biomarker for acute nephritis flares in systemic lupus erythematosus patients. Arthritis Rheum. 2009;60(1):199-206.
64.Shah D, Kiran R, Wanchu A, et al. Oxidative stress in systemic lupus erythematosus: relationship to Th1 cytokine and disease activity. Immunol Lett. 2010;129(1):7-12.
65.Torell F, Eketjäll S, Idborg H, et al. Cytokine Profiles in Autoantibody Defined Subgroups of Systemic Lupus Erythematosus. J Proteome Res. 2019;18(3):1208-17.
66.Boedigheimer MJ, Martin DA, Amoura Z, et al. Safety, pharmacokinetics and pharmacodynamics of AMG 811, an anti-interferon-γ monoclonal antibody, in SLE subjects without or with lupus nephritis. Lupus Sci Med. 2017;4(1):e000226.
67.Charles N, Hardwick D, Daugas E, et al. Basophils and the T helper 2 environment can promote the development of lupus nephritis. Nat Med. 2010;16(6):701-7.
68.Kim CJ, Lee CG, Jung JY, et al. The Transcription Factor Ets1 Suppresses T Follicular Helper Type 2 Cell Differentiation to Halt the Onset of Systemic Lupus Erythematosus. Immunity. 2018;49(6):1034-48.e8.
69.Xiang S, Zhang J, Zhang M, et al. Imbalance of helper T cell type 1, helper T cell type 2 and associated cytokines in patients with systemic lupus erythematosus: A meta-analysis. Front Pharmacol. 2022;13:988512.
70.Liu Z, Sun T, Song L, et al. Interleukin Polymorphisms and Predisposition of Systemic Lupus Erythematosus: A Meta-Analysis. Int Arch Allergy Immunol. 2021;182(10):1008-16.
71.Mohammadoo-Khorasani M, Salimi S, Tabatabai E, et al. Interleukin-1β (IL-1β) & IL-4 gene polymorphisms in patients with systemic lupus erythematosus (SLE) & their association with susceptibility to SLE. Indian J Med Res. 2016;143(5):591-6.
72.Elewa EA, Zakaria O, Mohamed EI, et al. The role of interleukins 4, 17 and interferon gamma as biomarkers in patients with Systemic Lupus Erythematosus and their correlation with disease activity. The Egyptian Rheumatologist. 2014;36(1):21-7.
73.Yu HH, Liu PH, Lin YC, et al. Interleukin 4 and STAT6 gene polymorphisms are associated with systemic lupus erythematosus in Chinese patients. Lupus. 2010;19(10):1219-28.
74.Nakajima A, Hirose S, Yagita H, et al. Roles of IL-4 and IL-12 in the development of lupus in NZB/W F1 mice. J Immunol. 1997;158(3):1466-72.
75.Carneiro JR, Fuzii HT, Kayser C, et al. IL-2, IL-5, TNF-α and IFN-γ mRNA expression in epidermal keratinocytes of systemic lupus erythematosus skin lesions. Clinics (Sao Paulo). 2011;66(1):77-82.
76.Mao Y-M, Zhao C-N, Leng J, et al. Interleukin-13: A promising therapeutic target for autoimmune disease. Cytokine & Growth Factor Reviews. 2019;45:9-23.
77.Wang R, Lu YL, Huang HT, et al. Association of interleukin 13 gene polymorphisms and plasma IL 13 level with risk of systemic lupus erythematosus. (1096-0023 (Electronic)).
78.Xu Z, Chen Y**.** Determination of serum interleukin-13 and nerve growth factor in patients with systemic lupus erythematosus and clinical significance. (1672-0733 (Print)).
79.Wong CK, Ho CY, Li EK, et al. Elevation of proinflammatory cytokine (IL-18, IL-17, IL-12) and Th2 cytokine (IL-4) concentrations in patients with systemic lupus erythematosus. Lupus. 2000;9(8):589-93.

80.Wong CK, Lit LC, Tam LS, et al. Hyperproduction of IL-23 and IL-17 in patients with systemic lupus erythematosus: implications for Th17-mediated inflammation in auto-immunity. Clin Immunol. 2008;127(3):385-93.
81.Henriques A, Inês L, Couto M, et al. Frequency and functional activity of Th17, Tc17 and other T-cell subsets in Systemic Lupus Erythematosus. Cell Immunol. 2010;264(1):97-103.
82.Yang J, Chu Y, Yang X, et al. Th17 and natural Treg cell population dynamics in systemic lupus erythematosus. Arthritis Rheum. 2009;60(5):1472-83.
83.Talaat RM, Mohamed SF, Bassyouni IH, et al. Th1/Th2/Th17/Treg cytokine imbalance in systemic lupus erythematosus (SLE) patients: Correlation with disease activity. Cytokine. 2015;72(2):146-53.
84.Wen Z, Xu L, Xu W, et al. Detection of dynamic frequencies of Th17 cells and their associations with clinical parameters in patients with systemic lupus erythematosus receiving standard therapy. Clin Rheumatol. 2014;33(10):1451-8.
85.Yin R, Xu R, Ding L, et al. Circulating IL-17 Level Is Positively Associated with Disease Activity in Patients with Systemic Lupus Erythematosus: A Systematic Review and Meta-Analysis. Biomed Res Int. 2021;2021:9952463.
86.Hsu HC, Yang P, Wang J, et al. Interleukin 17-producing T helper cells and interleukin 17 orchestrate autoreactive germinal center development in autoimmune BXD2 mice. Nat Immunol. 2008;9(2):166-75.
87.Kleczynska W, Jakiela B, Plutecka H, et al. Imbalance between Th17 and regulatory T-cells in systemic lupus erythematosus. Folia Histochem Cytobiol. 2011;49(4):646-53.
88.Xing Q, Wang B, Su H, et al. Elevated Th17 cells are accompanied by FoxP3+ Treg cells decrease in patients with lupus nephritis. Rheumatol Int. 2012;32(4):949-58.
89.Yuliasih Y, Rahmawati LD, Putri RM**.** Th17/Treg Ratio and Disease Activity in Systemic Lupus Erythematosus. Caspian J Intern Med. 2019;10(1):65-72.
90.Blanco P, Pitard V, Viallard JF, et al. Increase in activated CD8+ T lymphocytes expressing perforin and granzyme B correlates with disease activity in patients with systemic lupus erythematosus. Arthritis Rheum. 2005;52(1):201-11.
91.Couzi L, Merville P, Deminière C, et al. Predominance of CD8+ T lymphocytes among periglomerular infiltrating cells and link to the prognosis of class III and class IV lupus nephritis. Arthritis Rheum. 2007;56(7):2362-70.
92.Stohl W**.** Impaired polyclonal T cell cytolytic activity. A possible risk factor for systemic lupus erythematosus. Arthritis Rheum. 1995;38(4):506-16.
93.Comte D, Karampetsou MP, Yoshida N, et al. Signaling Lymphocytic Activation Molecule Family Member 7 Engagement Restores Defective Effector CD8+ T Cell Function in Systemic Lupus Erythematosus. Arthritis Rheumatol. 2017;69(5):1035-44.
94.Katsuyama E, Suarez-Fueyo A, Bradley SJ, et al. The CD38/NAD/SIRTUIN1/EZH2 Axis Mitigates Cytotoxic CD8 T Cell Function and Identifies Patients with SLE Prone to Infections. Cell Rep. 2020;30(1):112-23.e4.
95.Kang I, Quan T, Nolasco H, et al. Defective control of latent Epstein-Barr virus infection in systemic lupus erythematosus. J Immunol. 2004;172(2):1287-94.
96.Larsen M, Sauce D, Deback C, et al. Exhausted cytotoxic control of Epstein-Barr virus in human lupus. PLoS Pathog. 2011;7(10):e1002328.
97.Dean GS, Anand A, Blofeld A, et al. Characterization of CD3+ CD4- CD8- (double negative) T cells in patients with systemic lupus erythematosus: production of IL-4. Lupus. 2002;11(8):501-7.

98.Nose M, Nishihara M, Kamogawa J, et al. Genetic basis of autoimmune disease in MRL/lpr mice: dissection of the complex pathological manifestations and their susceptibility loci. Rev Immunogenet. 2000;2(1):154-64.
99.Crispín JC, Tsokos GC**.** Human TCR-alpha beta+ CD4- CD8- T cells can derive from CD8+ T cells and display an inflammatory effector phenotype. J Immunol. 2009;183(7):4675-81.
100.Rodríguez-Rodríguez N, Apostolidis SA, Penaloza-MacMaster P, et al. Programmed cell death 1 and Helios distinguish TCR-αβ+ double-negative (CD4-CD8-) T cells that derive from self-reactive CD8 T cells. J Immunol. 2015;194(9):4207-14.
101.Crispín JC, Oukka M, Bayliss G, et al. Expanded double negative T cells in patients with systemic lupus erythematosus produce IL-17 and infiltrate the kidneys. J Immunol. 2008;181(12):8761-6.
102.Crispín JC, Tsokos GC**.** Interleukin-17-producing T cells in lupus. Curr Opin Rheumatol. 2010;22(5):499-503.
103.Li H, Adamopoulos IE, Moulton VR, et al. Systemic lupus erythematosus favors the generation of IL-17 producing double negative T cells. Nature Communications. 2020;11(1):2859.
104.Adamichou C, Nikolopoulos D, Genitsaridi I, et al. In an early SLE cohort the ACR-1997, SLICC-2012 and EULAR/ACR-2019 criteria classify non-overlapping groups of patients: use of all three criteria ensures optimal capture for clinical studies while their modification earlier classification and treatment. Ann Rheum Dis. 2020;79(2):232-41.
105.Bombardier C, Gladman DD, Urowitz MB, et al. Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. Arthritis Rheum. 1992;35(6):630-40.
106.Connelly K, Morand EF**.** Systemic lupus erythematosus: a clinical update. Intern Med J. 2021;51(8):1219-28.
107.Mathias LM, Stohl W**.** Systemic lupus erythematosus (SLE): emerging therapeutic targets. Expert Opinion on Therapeutic Targets. 2020;24(12):1283-302.
108.Luscombe NM, Austin SE, Berman HM, et al. An overview of the structures of protein-DNA complexes. Genome Biol. 2000;1(1):Reviews001.
109.Panigrahi A, O'Malley BW**.** Mechanisms of enhancer action: the known and the unknown. Genome Biology. 2021;22(1):108.
110.Mattaini K**.** Introduction to Molecular and Cell Biology. Roger Williams University; 2020.
111.Henley MJ, Koehler AN**.** Advances in targeting 'undruggable' transcription factors with small molecules. Nature Reviews Drug Discovery. 2021;20(9):669-88.
112.Dowen JM, Fan ZP, Hnisz D, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell. 2014;159(2):374-87.
113.Ghavi-Helm Y, Jankowski A, Meiers S, et al. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. Nature Genetics. 2019;51(8):1272-82.
114.Lambert SA, Jolma A, Campitelli LF, et al. The Human Transcription Factors. Cell. 2018;172(4):650-65.
115.Köhler S, Doelken SC, Mungall CJ, et al. The Human Phenotype Ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014;42(Database issue):D966-74.
116.Singh A, Kumari R, Kirtonia A, et al. Chapter 1 - Current perspective of transcriptional regulators in human health and diseases. In: Garg M, Sethi G, Pandey AK, eds. Transcription and Translation in Health and Disease: Academic Press; 2023:3-16.
117.Odom DT, Zizlsperger N, Gordon DB, et al. Control of pancreas and liver gene expression by HNF transcription factors. Science. 2004;303(5662):1378-81.

118.Badis G, Berger MF, Philippakis AA, et al. Diversity and complexity in DNA recognition by transcription factors. Science. 2009;324(5935):1720-3.
119.Hellman LM, Fried MG. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc. 2007;2(8):1849-61.
120.Carey M. The enhanceosome and transcriptional synergy. Cell. 1998;92(1):5-8.
121.Galas DJ, Schmitz A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. 1978;5(9):3157-70.
122.Tullius TD, Dombroski BA, Churchill ME, et al. Hydroxyl radical footprinting: a high-resolution method for mapping protein-DNA contacts. Methods Enzymol. 1987;155:537-58.
123.Jolma A, Taipale J. Methods for analysis of transcription factor DNA-binding specificity in vitro. A Handbook of Transcription Factors. 2011:155-73.
124.Boyle AP, Davis S, Shulha HP, et al. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008;132(2):311-22.
125.Johnson DS, Mortazavi A, Myers RM, et al. Genome-wide mapping of in vivo protein-DNA interactions. Science. 2007;316(5830):1497-502.
126.Kuo MH, Allis CD. In vivo cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. Methods. 1999;19(3):425-33.
127.Park PJ. ChIP–seq: advantages and challenges of a maturing technology. Nature Reviews Genetics. 2009;10(10):669-80.
128.Furey TS. ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nature Reviews Genetics. 2012;13(12):840-52.
129.Bailey T, Krajewski P, Ladunga I, et al. Practical guidelines for the comprehensive analysis of ChIP-seq data. PLoS Comput Biol. 2013;9(11):e1003326.
130.Feng C, Song C, Liu Y, et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. Nucleic Acids Research. 2020;48(D1):D93-D100.
131.Fire A, Xu S, Montgomery MK, et al. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998;391(6669):806-11.
132.Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012;337(6096):816-21.
133.Rebollo A, Schmitt C. Ikaros, Aiolos and Helios: transcription regulators and lymphoid malignancies. Immunol Cell Biol. 2003;81(3):171-5.
134.John LB, Ward AC. The Ikaros gene family: transcriptional regulators of hematopoiesis and immunity. Mol Immunol. 2011;48(9-10):1272-8.
135.Fan Y, Lu D. The Ikaros family of zinc-finger proteins. Acta Pharm Sin B. 2016;6(6):513-21.
136.Powell MD, Read KA, Sreekumar BK, et al. Ikaros Zinc Finger Transcription Factors: Regulators of Cytokine Signaling Pathways and CD4(+) T Helper Cell Differentiation. Front Immunol. 2019;10:1299.
137.Joshi I, Yoshida T, Jena N, et al. Loss of Ikaros DNA-binding function confers integrin-dependent survival on pre-B cells and progression to acute lymphoblastic leukemia. Nat Immunol. 2014;15(3):294-304.
138.Medeiros BC. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med. 2009;360(17):1787; author reply -8.
139.Cunninghame Graham DS, Morris DL, Bhangale TR, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet. 2011;7(10):e1002341.

140.Franke A, McGovern DP, Barrett JC, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010;42(12):1118-25.
141.Ueta M, Sawai H, Sotozono C, et al. IKZF1, a new susceptibility gene for cold medicine-related Stevens-Johnson syndrome/toxic epidermal necrolysis with severe mucosal involvement. J Allergy Clin Immunol. 2015;135(6):1538-45.e17.
142.Swafford AD, Howson JM, Davison LJ, et al. An allele of IKZF1 (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes. Diabetes. 2011;60(3):1041-4.
143.Bentham J, Morris DL, Graham DSC, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet. 2015;47(12):1457-64.
144.Alexander T, Sattler A, Templin L, et al. Foxp3+ Helios+ regulatory T cells are expanded in active systemic lupus erythematosus. Ann Rheum Dis. 2013;72(9):1549-58.
145.Morgan B, Sun L, Avitahl N, et al. Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. Embo j. 1997;16(8):2004-13.
146.Romero F, Martínez AC, Camonis J, et al. Aiolos transcription factor controls cell death in T cells by regulating Bcl-2 expression and its cellular localization. Embo j. 1999;18(12):3419-30.
147.Sun J, Matthias G, Mihatsch MJ, et al. Lack of the transcriptional coactivator OBF-1 prevents the development of systemic lupus erythematosus-like phenotypes in Aiolos mutant mice. J Immunol. 2003;170(4):1699-706.
148.Cai X, Qiao Y, Diao C, et al. Association between polymorphisms of the IKZF3 gene and systemic lupus erythematosus in a Chinese Han population. PLoS One. 2014;9(10):e108661.
149.Marinho S, Custovic A, Marsden P, et al. 17q12-21 variants are associated with asthma and interact with active smoking in an adult population from the United Kingdom. Ann Allergy Asthma Immunol. 2012;108(6):402-11.e9.
150.Kurreeman FA, Stahl EA, Okada Y, et al. Use of a multiethnic approach to identify rheumatoid-arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet. 2012;90(3):524-32.
151.Molnár A, Georgopoulos K**.** The Ikaros gene encodes a family of functionally diverse zinc finger DNA-binding proteins. Mol Cell Biol. 1994;14(12):8292-303.
152.Font J, Cervera R, Ramos-Casals M, et al. Clusters of clinical and immunologic features in systemic lupus erythematosus: analysis of 600 patients from a single center. Semin Arthritis Rheum. 2004;33(4):217-30.
153.Tabayashi T, Ishimaru F, Takata M, et al. Characterization of the short isoform of Helios overexpressed in patients with T-cell malignancies. Cancer Sci. 2007;98(2):182-8.
154.Tonnelle C, Calmels B, Maroc C, et al. Ikaros gene expression and leukemia. Leuk Lymphoma. 2002;43(1):29-35.
155.Nakase K, Ishimaru F, Avitahl N, et al. Dominant negative isoform of the Ikaros gene in patients with adult B-cell acute lymphoblastic leukemia. Cancer Res. 2000;60(15):4062-5.
156.Tokunaga K, Yamaguchi S, Iwanaga E, et al. High frequency of IKZF1 genetic alterations in adult patients with B-cell acute lymphoblastic leukemia. Eur J Haematol. 2013;91(3):201-8.
157.Nakase K, Ishimaru F, Fujii K, et al. Overexpression of novel short isoforms of Helios in a patient with T-cell acute lymphoblastic leukemia. Exp Hematol. 2002;30(4):313-7.
158.Zhang Z, Swindle CS, Bates JT, et al. Expression of a non-DNA-binding isoform of Helios induces T-cell lymphoma in mice. Blood. 2007;109(5):2190-7.
159.Kelley CM, Ikeda T, Koipally J, et al. Helios, a novel dimerization partner of Ikaros expressed in the earliest hematopoietic progenitors. Current Biology. 1998;8(9):S1.

160.Hahm K, Cobb BS, McCarty AS, et al. Helios, a T cell-restricted Ikaros family member that quantitatively associates with Ikaros at centromeric heterochromatin. Genes Dev. 1998;12(6):782-96.
161.Dovat S, Montecino-Rodriguez E, Schuman V, et al. Transgenic expression of Helios in B lineage cells alters B cell properties and promotes lymphomagenesis. J Immunol. 2005;175(6):3508-15.
162.Thornton AM, Shevach EM**.** Helios: still behind the clouds. Immunology. 2019;158(3):161-70.
163.Fujii K, Ishimaru F, Nakase K, et al. Over-expression of short isoforms of Helios in patients with adult T-cell leukaemia/lymphoma. Br J Haematol. 2003;120(6):986-9.
164.Sun L, Kerawalla H, Wu X, et al. Expression of a unique helios isoform in human leukemia cells. Leuk Lymphoma. 2002;43(4):841-9.
165.Park SM, Cho H, Thornton AM, et al. IKZF2 Drives Leukemia Stem Cell Self-Renewal and Inhibits Myeloid Differentiation. Cell Stem Cell. 2019;24(1):153-65.e7.
166.Thul PJ, Lindskog C**.** The human protein atlas: A spatial map of the human proteome. Protein Sci. 2018;27(1):233-44.
167.Asanuma S, Yamagishi M, Kawanami K, et al. Adult T-cell leukemia cells are characterized by abnormalities of Helios expression that promote T cell growth. Cancer Sci. 2013;104(8):1097-106.
168.Cai Q, Dierich A, Oulad-Abdelghani M, et al. Helios deficiency has minimal impact on T cell development and function. J Immunol. 2009;183(4):2303-11.
169.Akimova T, Beier UH, Wang L, et al. Helios expression is a marker of T cell activation and proliferation. PLoS One. 2011;6(8):e24226.
170.Serre K, Bénézech C, Desanti G, et al. Helios is associated with CD4 T cells differentiating to T helper 2 and follicular helper T cells in vivo independently of Foxp3 expression. PLoS One. 2011;6(6):e20731.
171.Szurek E, Cebula A, Wojciech L, et al. Differences in Expression Level of Helios and Neuropilin-1 Do Not Distinguish Thymus-Derived from Extrathymically-Induced CD4+Foxp3+ Regulatory T Cells. PLoS One. 2015;10(10):e0141161.
172.Kim HJ, Barnitz RA, Kreslavsky T, et al. Stable inhibitory activity of regulatory T cells requires the transcription factor Helios. Science. 2015;350(6258):334-9.
173.Baine I, Basu S, Ames R, et al. Helios induces epigenetic silencing of IL2 gene expression in regulatory T cells. J Immunol. 2013;190(3):1008-16.
174.Naluyima P, Lal KG, Costanzo MC, et al. Terminal Effector CD8 T Cells Defined by an IKZF2(+)IL-7R(-) Transcriptional Signature Express FcγRIIIA, Expand in HIV Infection, and Mediate Potent HIV-Specific Antibody-Dependent Cellular Cytotoxicity. J Immunol. 2019;203(8):2210-21.
175.Ng MSF, Roth TL, Mendoza VF, et al. Helios enhances the preferential differentiation of human fetal CD4(+) naïve T cells into regulatory T cells. Sci Immunol. 2019;4(41).
176.Fontenot JD, Rasmussen JP, Williams LM, et al. Regulatory T cell lineage specification by the forkhead transcription factor foxp3. Immunity. 2005;22(3):329-41.
177.Gottschalk RA, Corse E, Allison JP**.** Expression of Helios in peripherally induced Foxp3+ regulatory T cells. J Immunol. 2012;188(3):976-80.
178.Thornton AM, Korty PE, Tran DQ, et al. Expression of Helios, an Ikaros transcription factor family member, differentiates thymic-derived from peripherally induced Foxp3+ T regulatory cells. J Immunol. 2010;184(7):3433-41.
179.Getnet D, Grosso JF, Goldberg MV, et al. A role for the transcription factor Helios in human CD4(+)CD25(+) regulatory T cells. Mol Immunol. 2010;47(7-8):1595-600.

180.Sugita K, Hanakawa S, Honda T, et al. Generation of Helios reporter mice and an evaluation of the suppressive capacity of Helios(+) regulatory T cells in vitro. Exp Dermatol. 2015;24(7):554-6.
181.Thornton AM, Lu J, Korty PE, et al. Helios(+) and Helios(-) Treg subpopulations are phenotypically and functionally distinct and express dissimilar TCR repertoires. Eur J Immunol. 2019;49(3):398-412.
182.Zabransky DJ, Nirschl CJ, Durham NM, et al. Phenotypic and functional properties of Helios+ regulatory T cells. PLoS One. 2012;7(3):e34547.
183.Fuhrman CA, Yeh WI, Seay HR, et al. Divergent Phenotypes of Human Regulatory T Cells Expressing the Receptors TIGIT and CD226. J Immunol. 2015;195(1):145-55.
184.Floess S, Freyer J, Siewert C, et al. Epigenetic control of the foxp3 locus in regulatory T cells. PLoS Biol. 2007;5(2):e38.
185.Kim YC, Bhairavabhotla R, Yoon J, et al. Oligodeoxynucleotides stabilize Helios-expressing Foxp3+ human T regulatory cells during in vitro expansion. Blood. 2012;119(12):2810-8.
186.Dawson NAJ, Rosado-Sánchez I, Novakovsky GE, et al. Functional effects of chimeric antigen receptor co-receptor signaling domains in human regulatory T cells. Science Translational Medicine. 2020;12(557):eaaz3866.
187.Yin X, Kim K, Suetsugu H, et al. Meta-analysis of 208370 East Asians identifies 113 susceptibility loci for systemic lupus erythematosus. Ann Rheum Dis. 2021;80(5):632-40.
188.París-Muñoz A, Aizpurua G, Barber DF**.** Helios Expression Is Downregulated on CD8(+) Treg in Two Mouse Models of Lupus During Disease Progression. Front Immunol. 2022;13:922958.
189.Golding A, Hasni S, Illei G, et al. The percentage of FoxP3+Helios+ Treg cells correlates positively with disease activity in systemic lupus erythematosus. Arthritis Rheum. 2013;65(11):2898-906.
190.Khunsriraksakul C, Li Q, Markus H, et al. Multi-ancestry and multi-trait genome-wide association meta-analyses inform clinical risk prediction for systemic lupus erythematosus. Nat Commun. 2023;14(1):668.
191.Zhou VW, Goren A, Bernstein BE**.** Charting histone modifications and the functional organization of mammalian genomes. Nature Reviews Genetics. 2011;12(1):7-18.
192.Calo E, Wysocka J**.** Modification of enhancer chromatin: what, how, and why? Mol Cell. 2013;49(5):825-37.
193.Ge X, Zhang H, Xie L, et al. EpiAlign: an alignment-based bioinformatic tool for comparing chromatin state sequences. Nucleic Acids Res. 2019;47(13):e77.
194.Mammana A, Chung H-R**.** Chromatin segmentation based on a probabilistic model for read counts explains a large portion of the epigenome. Genome Biology. 2015;16(1):151.
195.Ernst J, Kellis M**.** Chromatin-state discovery and genome annotation with ChromHMM. Nature Protocols. 2017;12(12):2478-92.
196.Pehrsson EC, Choudhary MNK, Sundaram V, et al. The epigenomic landscape of transposable elements across normal human development and anatomy. Nature Communications. 2019;10(1):5640.
197.Ward LD, Kellis M**.** HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Research. 2016;44(D1):D877-D81.
198.Zhu J, Adli M, Zou JY, et al. Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell. 2013;152(3):642-54.
199.Javierre BM, Burren OS, Wilder SP, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell. 2016;167(5):1369-84.e19.

200.Pérez RF, Tejedor JR, Bayón GF, et al. Distinct chromatin signatures of DNA hypomethylation in aging and cancer. Aging Cell. 2018;17(3):e12744.
201.Vu H, Ernst J**.** Universal annotation of the human genome through integration of over a thousand epigenomic datasets. Genome Biology. 2022;23(1):9.
202.Chronis C, Fiziev P, Papp B, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell. 2017;168(3):442-59.e20.
203.Boix CA, James BT, Park YP, et al. Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature. 2021;590(7845):300-7.
204.Sollis E, Mosaku A, Abid A, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Research. 2023;51(D1):D977-D85.
205.Patwardhan RP, Lee C, Litvin O, et al. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat Biotechnol. 2009;27(12):1173-5.
206.Lu X, Chen X, Forney C, et al. Global discovery of lupus genetic risk variant allelic enhancer activity. Nature Communications. 2021;12(1):1611.
207.Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.
208.Yu G, Wang LG, He QY**.** ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics. 2015;31(14):2382-3.
209.Lawrence M, Huber W Fau - Pagès H, Pagès H Fau - Aboyoun P, et al. Software for computing and annotating genomic ranges. (1553-7358 (Electronic)).
210.Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature. 2015;518(7539):317-30.
211.Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68-74.
212.Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4(1).
213.Yang J, Yang W, Hirankarn N, et al. ELF1 is associated with systemic lupus erythematosus in Asian populations. Human Molecular Genetics. 2011;20(3):601-7.
214.Kwong A, Boughton AP, Wang M, et al. FIVEx: an interactive eQTL browser across public datasets. Bioinformatics. 2022;38(2):559-61.
215.Schmiedel BJ, Singh D, Madrigal A, et al. Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell. 2018;175(6):1701-15.e16.
216.Akbarali Y, Oettgen P, Boltax J, et al. ELF-1 interacts with and transactivates the IgH enhancer pi site. J Biol Chem. 1996;271(42):26007-12.
217.Juang YT, Tenbrock K, Nambiar MP, et al. Defective production of functional 98-kDa form of Elf-1 is responsible for the decreased expression of TCR zeta-chain in patients with systemic lupus erythematosus. J Immunol. 2002;169(10):6048-55.
218.Nunes-Santos CJ, Kuehn HS, Rosenzweig SD**.** IKAROS Family Zinc Finger 1-Associated Diseases in Primary Immunodeficiency Patients. Immunol Allergy Clin North Am. 2020;40(3):461-70.
219.Zhao S, Fung-Leung WP, Bittner A, et al. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLoS One. 2014;9(1):e78644.
220.Aleksander SA, Balhoff J, Carbon S, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1).
221.Kanehisa M, Goto S**.** KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27-30.

222.Martens M, Ammar A, Riutta A, et al. WikiPathways: Connecting Communities. Nucleic Acids Research. 2020.
223.Croft D, O'Kelly G, Wu G, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research. 2011;39(suppl_1):D691-D7.
224.Zhou Y, Zhou B, Pache L, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
225.Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47-e.
226.Durinck S, Spellman PT, Birney E, et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols. 2009;4(8):1184-91.
227.Wickham H**.** Ggplot2; 2016.
228.Yu G**.** Enrichplot: Visualization of functional enrichment result. R package version 1.16.2. 2022.
229.Magasanik B**.** Global regulation of gene expression. Proceedings of the National Academy of Sciences. 2000;97(26):14044-5.
230.Ahmed M, Min DS, Kim DR**.** Integrating binding and expression data to predict transcription factors combined function. BMC Genomics. 2020;21(1):610.
231.Abugessaisa I, Noguchi S, Carninci P, et al. The FANTOM5 Computation Ecosystem: Genomic Information Hub for Promoters and Active Enhancers. Methods Mol Biol. 2017;1611:199-217.
232.Gao T, Qian J**.** EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Research. 2020;48(D1):D58-D64.
233.Vaughan HWaRFaLHaKMaD**.** dplyr: A Grammar of Data Manipulation. 2023.
234.Chen H**.** VennDiagram: Generate High-Resolution Venn and Euler Plots. 2022.
235.Russo G, Zegar C, Giordano A**.** Advantages and limitations of microarray technology in human cancer. Oncogene. 2003;22(42):6497-507.
236.King HC, Sinha AA**.** Gene Expression Profile Analysis by DNA MicroarraysPromise and Pitfalls. JAMA. 2001;286(18):2280-8.
237.Wang Z, Gerstein M, Snyder M**.** RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics. 2009;10(1):57-63.
238.Sheng Y-J, Gao J-P, Li J, et al. Follow-up study identifies two novel susceptibility loci PRKCB and 8p11.21 for systemic lupus erythematosus. Rheumatology. 2011;50(4):682-8.
239.Zhu Z, Yang L, Zhang Y, et al. Increased expression of PRKCB mRNA in peripheral blood mononuclear cells from patients with systemic lupus erythematosus. Ann Hum Genet. 2018;82(4):200-5.
240.Manni M, Gupta S, Nixon BG, et al. IRF4-Dependent and IRF4-Independent Pathways Contribute to DC Dysfunction in Lupus. PLoS One. 2015;10(11):e0141927.
241.Rodríguez-Carrio J, López P, Alperi-López M, et al. IRF4 and IRGs Delineate Clinically Relevant Gene Expression Signatures in Systemic Lupus Erythematosus and Rheumatoid Arthritis. Frontiers in Immunology. 2019;9.
242.Lech M, Weidenbusch M, Kulkarni OP, et al. IRF4 Deficiency Abrogates Lupus Nephritis Despite Enhancing Systemic Cytokine Production. Journal of the American Society of Nephrology. 2011;22(8).
243.Martínez-Ramos R, García-Lozano J-R, Lucena J-M, et al. Differential expression pattern of microRNAs in CD4+ and CD19+ cells from asymptomatic patients with systemic lupus erythematosus. Lupus. 2014;23(4):353-9.
244.Domingo S, Solé C, Moliné T, et al. Thalidomide Exerts Anti-Inflammatory Effects in Cutaneous Lupus by Inhibiting the IRF4/NF-κB and AMPK1/mTOR Pathways. Biomedicines. 2021;9(12).

245.Nijnik A, Ferry H, Lewis G, et al. Spontaneous B cell hyperactivity in autoimmune-prone MRL mice. International Immunology. 2006;18(7):1127-37.
246.Zeng Y, Peng X, Wang Y, et al. Therapeutic effect of modified zengye decoction on primary Sjogren's syndrome and its effect on plasma exosomal proteins. Front Pharmacol. 2022;13:930638.
247.Jayaram N, Usvyat D, R. Martin AC**.** Evaluating tools for transcription factor binding site prediction. BMC Bioinformatics. 2016;17(1):547.
248.Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89.
249.John S, Sabo PJ, Thurman RE, et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat Genet. 2011;43(3):264-8.
250.Zhang Q, Teng P, Wang S, et al. Computational prediction and characterization of cell-type-specific and shared binding sites. Bioinformatics. 2023;39(1):btac798.
251.Pique-Regi R, Degner JF, Pai AA, et al. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 2011;21(3):447-55.
252.Gusmao EG, Dieterich C, Zenke M, et al. Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. Bioinformatics. 2014;30(22):3143-51.
253.Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nature Genetics. 2016;48(10):1193-203.
254.Awdeh A, Turcotte M, Perkins TJ**.** Cell Type Specific DNA Signatures of Transcription Factor Binding. bioRxiv. 2022:2022.07.15.500259.
255.Alasoo K, Rodrigues J, Mukhopadhyay S, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. Nature Genetics. 2018;50(3):424-31.
256.Calderon D, Nguyen MLT, Mezger A, et al. Landscape of stimulation-responsive chromatin across diverse human immune cells. Nature Genetics. 2019;51(10):1494-505.
257.Kent WJ, Sugnet Cw Fau - Furey TS, Furey Ts Fau - Roskin KM, et al. The human genome browser at UCSC. (1088-9051 (Print)).
258.Lawrence M, Huber W, Pagès H, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol. 2013;9(8):e1003118.
259.Heinz S, Benner C Fau - Spann N, Spann N Fau - Bertolino E, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. (1097-4164 (Electronic)).
260.Luo Y, Hitz BC, Gabdank I, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. (1362-4962 (Electronic)).
261.Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47.
262.Chen Y, Lun AT, Smyth GK**.** From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res. 2016;5:1438.
263.Love MI, Huber W, Anders S**.** Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15(12):550.

264.Raplee ID, Evsikov AV, Marín de Evsikova C**.** Aligning the Aligners: Comparison of RNA Sequencing Data Alignment and Gene Expression Quantification Tools for Clinical Breast Cancer Research. Journal of Personalized Medicine. 2019;9(2):18.
265.Song Y, Liu C, Finegold SM**.** Real-Time PCR Quantitation of Clostridia in Feces of Autistic Children. Applied and Environmental Microbiology. 2004;70(11):6459-65.
266.Govindarajan R, Duraiyan J, Kaliyappan K, et al. Microarray and its applications. J Pharm Bioallied Sci. 2012;4(Suppl 2):S310-2.
267.Qian J, Fang Y, Yuan N, et al. Innate immune remodeling by short-term intensive fasting. Aging Cell. 2021;20(11):e13507.
268.Zhang J, Wang J, Wu Y, et al. Identification of SLED1 as a Potential Predictive Biomarker and Therapeutic Target of Post-Infarct Heart Failure by Bioinformatics Analyses. International Heart Journal. 2021;62(1):23-32.
269.Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2(3):100141.
270.Zeng H, Zhao S, Pang Z, et al. Identification of Key Genes and Pathways in the Hippocampus after Traumatic Brain Injury: Bioinformatics Analysis and Experimental Validation. JIN. 2023;22(2).
271.Merico D, Isserlin R, Stueker O, et al. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One. 2010;5(11):e13984.
272.van der Sande M, Frölich S, van Heeringen SJ**.** Computational approaches to understand transcription regulation in development. Biochemical Society Transactions. 2023;51(1):1-12.
273.Ota M, Nagafuchi Y, Hatano H, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell. 2021;184(11):3006-21.e17.
274.Fukuda A, Kodama Y, Mashima J, et al. DDBJ update: streamlining submission and access of human data. Nucleic Acids Res. 2021;49(D1):D71-d5.
275.Robinson MD, McCarthy DJ, Smyth GK**.** edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139-40.
276.Bader GD, Hogue CW**.** An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics. 2003;4:2.
277.Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.
278.Chin C-H, Chen S-H, Wu H-H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Systems Biology. 2014;8(4):S11.
279.Zhao X, Zhang L, Wang J, et al. Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. J Transl Med. 2021;19(1):35.
280.Bing PF, Xia W, Wang L, et al. Common Marker Genes Identified from Various Sample Types for Systemic Lupus Erythematosus. PLoS One. 2016;11(6):e0156234.
281.Lou QY, Li Z, Teng Y, et al. Associations of FKBP4 and FKBP5 gene polymorphisms with disease susceptibility, glucocorticoid efficacy, anxiety, depression, and health-related quality of life in systemic lupus erythematosus patients. Clin Rheumatol. 2021;40(1):167-79.
282.Kitoh A, Ono M, Naoe Y, et al. Indispensable role of the Runx1-Cbfbeta transcription complex for in vivo-suppressive function of FoxP3+ regulatory T cells. Immunity. 2009;31(4):609-20.