2019

Interrogation of epigenetic and genetic determinants of complex diseases

Megan Elizabeth Breitbach

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor, or, in his absence, by the Chair of the Department or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this dissertation.

(Student Signature)

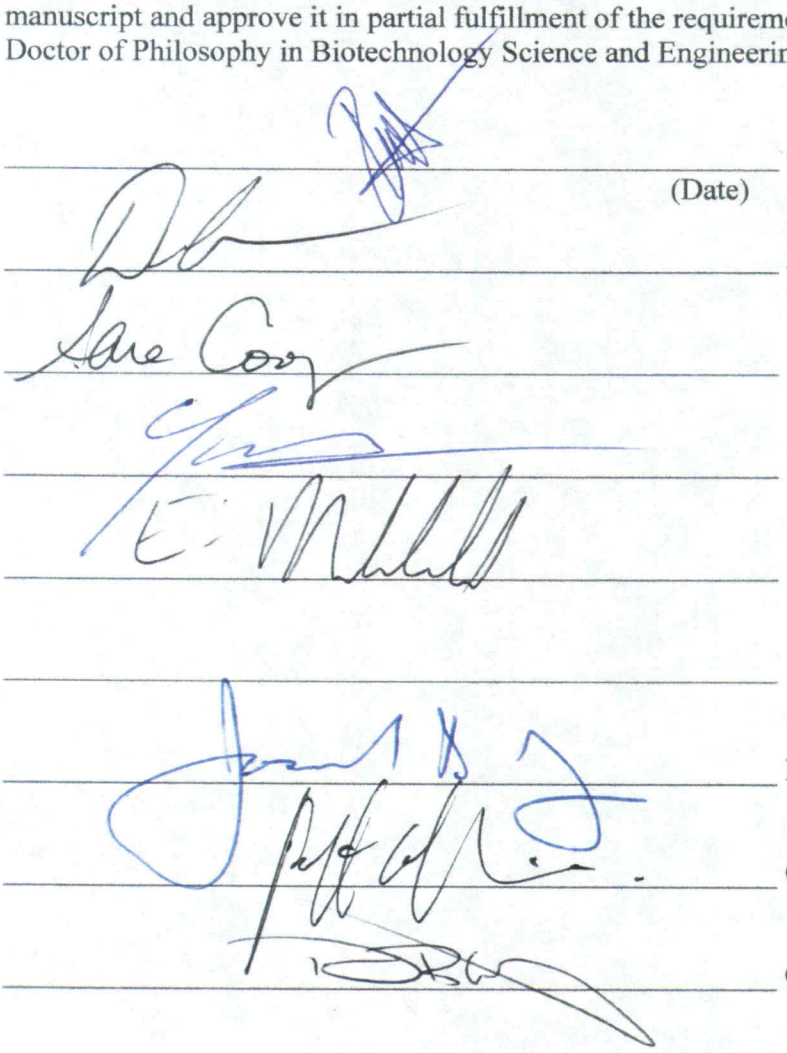
8-12-2019

(Date)

THESIS APPROVAL FORM

Submitted by Megan E. Breitbach in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biotechnology Science and Engineering and accepted on behalf of the Faculty of the School of Graduate Studies by the dissertation committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that I have advised and/or supervised the candidate on the work described in this dissertation. I further certify that I have reviewed the dissertation manuscript and approve it in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biotechnology Science and Engineering.



Committee Chair

(Date)

Advisor

Department Chair

College Dean

Graduate Dean

ABSTRACT
The School of Graduate Studies
The University of Alabama in Huntsville

Degree Doctor of Philosophy Program Biotechnology Science and Engineering

Name of Candidate Megan Elizabeth Breitbach

Title Interrogation of Epigenetic and Genetic Determinants of Complex Diseases

Systemic Lupus Erythematosus (SLE) is a complex autoimmune disease characterized by altered immune cells that induce systemic inflammation, organ damage and, in some cases, premature mortality. DNA methylation studies focused on SLE have revealed widespread hypomethylation of CpGs within interferon-related genes in SLE patients relative to controls. However, most studies to date have focused on European ethnicities leaving out the most at-risk population, African American females. African American females are affected two to three times more often than females of European ancestry. Furthermore, SLE tends to occur at an earlier age in African American females with increased severity compared to any other affected population. This work aimed to address the biological explanation for the racial disparity in SLE through interrogation of genome-wide DNA methylation data from five sorted B cell lineages in a cohort of both SLE patients and control females of both African and European ancestry. I aimed to identify at what stage in B cell development aberrant epigenetic patterns arise and to develop multivariate epigenetic signatures of SLE and its severity. I discovered that epigenetic defects in African American female SLE patients are present in immature B cells emerging from bone marrow (transitional B cells), while epigenetic defects appear

to develop later during B cell development in European American female SLE patients. The most associated African American-specific CpGs occur at interferon-regulating genes and are enriched for binding of transcription factors involved in immune regulation, such as *EBF1*. Lastly, these epigenetic changes proved to be highly predictive of disease status in all African American patient immune cell populations.

Abstract Approval: Committee Chair

Department Chair

Graduate Dean

For Jeffrey Breitbach

ACKNOWLEDGEMENTS

First and foremost, I'd like to acknowledge my family. To my sisters, thank you for being role models for me from the very start. Thank you for teaching me to be strong, fierce and steadfast. Without my parent's endless support and encouragement, I would not be who I am or where I am today. Thank you for teaching me to work hard, chase my dreams, and most importantly, find joy along the way.

To Jeffrey Breitbach, for whom I dedicate my thesis to, I owe many thanks. Many people would say "if only he had been born 30 years later" because of the groundbreaking and lifechanging genomic research that has led to tremendous advances in muscular dystrophy therapy. Yet, in my heart I know everything happens for a reason. Your disease brought my parents together and ultimately influenced me to pursue genomic research. I can only hope that you see the family that resulted from your amazing but short life and the work that you have inspired.

A special thank you to my mentor, Devin Absher. I cannot thank you enough for taking me in as your graduate student. I am so appreciative of the guidance and reassurance you offered me, and most importantly that your door was always open when I needed help.

I would also like to thank my fiancé, Ryne Ramaker. Your support and guidance throughout graduate school made an unsurmountable difference in my knowledge, as well as my confidence and perseverance. You are my true love and my best friend. I cannot wait to see the things we accomplish together.

Lastly, I would like to thank my friends and colleagues at HudsonAlpha. To Kevin Bowling, Andrew Hardigan and Matthew Neu in particularly, thank you for offering me camaraderie, kindness and laughter throughout my time in graduate school. I am honored to call you all my friends and am constantly astounded by the scientific achievements you all accomplish.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

LIST OF ABBREVIATIONS

AA: African American

ACR: American College of Rheumatology

ACR: American College of Rheumatology

ANA: Anti-Nuclear Antibody

APOE: Apolipoprotein E

APOE4: Apolipoprotein E Genotype ε4

AUC: Area Under the Curve

BCL2/6: B Cell CLL/Lymphoma 2/6

BCR: B Cell Receptor

BDD: Benign Breast Disease

BER: Base Excision Repair

BMI: Body Mass Index

bp: Base Pair

BRCA1/2: Breast Cancer Type 1/2 Susceptibility

CAD: Cardiovascular Disease

CADD: Combined Annotation Dependent Depletion

CATCH-Seq: Clone Adapted Template Capture Hybridization Sequencing

CD40L: Cluster of Differentiation 40 Ligand

CD80: Cluster of Differentiation 80 Antigen

CF: Cystic Fibrosis

CFTR: Cystic Fibrosis Transmembrane Conductance Regulator

CHAMP: Community Healthy Activities Model Program for Seniors

CHD1: Chromodomain Helicase DNA Binding Protein 1

ChIP-Seq: Chromatin Immunoprecipitation Sequencing

CLP: Common Lymphoid Progenitor

CMP: Common Myeloid Progenitor

CRISPR: Clustered Regularly Interspaced Palindromic Repeats

CS: Cockadyne Syndrome

DBD: Deoxyribose Nucleic Acid Binding Domain

DBP: Diastolic Blood Pressure

dCAS9: Deactivated CRISPR Associated Protein

DNA: Deoxyribonucleic acid

DNMT: DNA Methyltransferase

DSB: Double Strand Break

dsDNA: Double Stranded Deoxyribose Nucleic Acid

DSST: Digit Symbol Substitution Test

DXA: Dual Energy X-ray Absorptiometry

E2A: Transcription Factor E2-Alpha

EA: European American

EBF1: Early B Cell Factor 1

EBV: Epstein-Barr Virus

EHMT1: Euchromatic Histone Lysine Methyltransferase 1

EMS: Epigenetic Maintenance System

ENCODE: Encyclopedia of DNA Elements

eQTL: Expression Quantitative Trait Loci

ERCC1/4/5: Excision Cross Complementation Group 1/4/5

FANTOM5: Functional Annotation of the Mammalian Genome Phase 5

FDR: False Discovery Rate

FOXO3A: Forkhead Box O3A

GRASP: Genome-Wide Repository of Associations Between SNPs and Phenotypes

gRNA: Guide Ribonucleic Acid

GWA: Genome-wide Association

GWAS: Genome Wide Association Study

HCP: Histocompatibility Complex P5

HGPS: Hutchinson-Gilford Progeria Syndrome

HLA: Human Leukocyte Antigen

hQTL: Histone Quantitative Trait Loci

HR: Homologous Recombination

HSC: Hematopoietic Stem Cell

HTT: Huntingtin

IFI44L: Interferon Induced Protein 44 Like

IFIT1: Interferon Induced Protein with Tetratricopeptide Repeats 1

IFITM1: Interferon Induced Transmembrane Protein 1

IFN: Interferon

Ig: Immunoglobulin

IgH: Immunoglobulin Heavy Locus

IgL: Immunoglobulin Light Locus

IL-4: Interleukin 4

IRF4: Interferon Regulatory Factor 4

IRF8: Interferon Regulatory Factor 8

IZKF1: IKAROS Family Zinc Finger 1

kb: Kilobase

KLF4/9: Kruppel Like Factor 4/9

LASSO: L1 Penalized Log Partial Likelihood

LMNA: Lamin A

LMO2: LIM Domain Only 2

LMPP: Lymphoid-primed Multipotent Progenitor

LT HSC: Long Term Hematopoietic Stem Cell

MAF: Minor Allele Frequency

MBD: Methyl-CpG-Binding Domain

MC1R: Melanocortin 1 Receptor

MeDIP: Methylated DNA Immunoprecipitation

MEF2C: Myocyte Enhancer Factor 2C

methylQTL: Methylation Quantitative Trait Loci

MHC: Major Histocompatibility Complex

MLL: Lysine Methyltransferase 2A

MOCA: Montreal Cognitive Assessment

MSE: Mean Standard Error

MTA3: Metastasis Associated 1 Family Member 3

MX1: MX Dynamin Like GTPase

MYC: MYC Proto-Oncogene

NER: Nucleotide Excision Repair

NFKB1: Nuclear Factor Kappa B Subunit 1

NGS: Next Generation Sequencing

NLRC5: NLR Family CARD Domain Containing 5

OMIM: Online Mendelian Inheritance in Man

PARP1: Poly(ADP-Ribose) Polymerase 1

PAX5: B Cell-Specific Activator Protein

PCR: Polymerase Chain Reaction

PCR: Polymerase Chain Reaction

PFM: Position Frequency Matrix

PLZF: Promyelocytic Leukemia Zinc Finger

POLR2A: RNA Polymerase II Subunit A

POU2F2: POU Class 2 Homeobox 2

PU.1: Proto-Oncogene

qPCR: Quantitative Polymerase Chain Reaction

QQ: Quantile-Quantile

RELA: RELA Proto-Oncogene, NF-KB Subunit

RNA: Ribonucleic Acid

ROC: Receiver Operating Characteristic

RRBS: Reduce Representation Bisulfite Sequencing

RUNX1: Runt Related Transcription Factor 1

SCAN: SNP and CNV Annotation Database

SIFT: Sorting Intolerant From Tolerant

SLAM: Signaling Lymphocytic Activation Molecule

SLE: Systemic Lupus Erythematosus

SLEDAI: Systemic Lupus Erythematosus Disease Activity Index

SLICC: Systemic Lupus International Collaborating Clinics

SNP: Single Nucleotide Polymorphism

SNV: Single Nucleotide Variation

SPPB: Short Physical Performance Battery

ST HSC: Short Term Hematopoietic Stem Cell

STAT3: Signal Transducer and Activator of Transcription 3

SVM: Support Vector Machine

tAF: Total Allele Frequency

TET: Ten-Eleven Translocation

TFBS: Transcription Factor Binding Site

TFBS: Transcription Factor Sinding Site

TNFR: Tumor Necrosis Family Receptor B7

TSS: Transcription Start Site

UPMC: University of Pittsburgh Medical Center

WGBS: Whole-Genome Bisulfite Sequencing

WGS: Whole-genome Sequencing

WRN: Werner Syndrome ATP-dependent Helicase

WRNIP1: Werner Helicase Interacting Protein 1

XP: Xeroderma Pigmentosum

CHAPTER 1

GENETICS AND EPIGENETICS

1.1 An Introduction to Genetics

Deoxyribonucleic acid (DNA) is the hereditary material packaged within the cells of most all organisms, including humans ("Cells and DNA" 2018). DNA is made up of four chemical bases (adenine, thymine, guanine and cytosine) that together provide the information for the development, growth and reproduction of organisms. The chemical bases pair up in a double helical fashion that coils into larger structures referred to as chromosomes, for which humans have 23 pairs. Of the ~3 billion base pairs across all 46 chromosomes, about 99% is shared across the human race. The 0.1% of the human genome that exhibits variability across populations spans both regions that encode proteins (genes) and noncoding (intergenic) regulatory regions.

Specific sequences of DNA encode genes, which are basic functional units of heredity and can be thought of the instruction manuals for making ribonucleic acid (RNA) and proteins (National Institute of General Medical Sciences 2010; "Cells and DNA" 2018). The human genome contains between 20,000-25,000 genes ("Cells and DNA" 2018). Every human will receive two copies of each gene; one from each parent, which are referred to as alleles. Differences in alleles, in addition to de novo mutations that arise spontaneously, are what make organisms unique from one another. Variation at a nucleotide that occurs in only a minority of the population, or at a low allele frequency,

is referred to as a single nucleotide polymorphism (SNP). For example, SNPs in the Melanocortin 1 Receptor (MC1R) are responsible for variation in hair and skin color as well as freckling and mole count (Mengel-Jørgensen et al. 2006).

1.2 A Brief History of Genetic Discoveries

The advent of genomic analysis began in 1859 with the discovery of natural selection and the publication of Charles Darwin's classic, *On the Origin of Species by the means of Natural Selection, or the Preservation of Favored Races in the Struggle for Life*. Gregor Mendel provided further evidence for Darwin's theory of natural selection with his experimental work on peas in 1865. In subsequent years Frederick Miescher was the first to isolate DNA, and Walter Flemming's staining of chromatin allowed for the discovery of mitosis (1879). It was not until 1909 that the word "gene" was coined by Wilhelm Johannsen, who also is responsible for making the first distinction between one's hereditary disposition, or "genotype", and one's outward physical appearance, or "phenotype" (Johannsen 1911). The discovery that chromosomes contain genes was made by Thomas Hunt Morgan in 1911.

The 1950's saw a large jump in the understanding of DNA and genetics, beginning with Alfred Hershey and Martha Chase's experiments in virus and bacterium, proving that genes are made of DNA (1952). Francis H. Crick and James Watson performed the Nobel Prize winning work of determining the structure of DNA in 1953. Joe Hin Tjio defined the exact number of chromosomes in human cells in 1955, the same year that Arthur Kornberg and his team isolated DNA polymerase. The semiconservative nature of DNA replication was discovered in 1958 by Matthew Meselson and Franklin Stahl. In

1966 Marshall Nirenberg cracked the genetic code along with others by discovering that a 4-letter alphabet (A, T, C and G) encodes 20 difference proteins based on order alone.

The transition from discovery to manipulation and experimentation of genomics came in the 1960s. In 1968, the first restriction enzyme was discovered, paving the way for the development of genetic research tools. Recombinant DNA was produced for the first time in 1972, followed by the development of the Sanger method in 1975. The Sanger method allowed for DNA sequencing using specific dyes to identify each of the 4 nucleic acids in DNA. The development of polymerase chain reaction (PCR) method in 1983 allowed for more complex genotyping, including: tandem repeats, insertion and deletion polymorphisms, and single nucleotide polymorphisms (Hirschhorn et al. 2002). With PCR, genetic components could be associated with disease in a straightforward manner, paving the way for future research in genomic medicine. Using PCR, one could amplify regions of interest in those with disease and make comparisons with those that do not have the disease. By 2002, over 600 genomic associations with common complex diseases had been made using PCR technology (Hirschhorn et al. 2002).

Genetic sequencing only got bigger, better, and faster with the development of next generation sequencing (NGS) technologies. Briefly, NGS consists of randomly breaking DNA into fragments that are immobilized on a sequencing lane, allowing for thousands of sequencing reactions to occur simultaneously. Primers are then hybridized to the ends of each template for NGS. Dye-labeled nucleotides are incorporated during DNA synthesis reactions, which can then be aligned to a reference genome or assembled de novo.

NGS technology allowed for the completion of the monumental 13-year Human Genome Project in 2003 (Zaveri et al. 2001). Furthermore, rapid development in NGS technologies enabled the integration of basic translational and clinical research, paving the way for a more detailed genetic understanding of complex traits (Boyle, Li, and Pritchard 2017; Ziogas, Kyrochristos, and Roukos 2018). Companies like Illumina, Roche, and Thermo Fisher Scientific offered sequencing machines capable of whole exome, genome, transcriptome and targeted sequencing. With this, the cost of sequencing a genome decreased much faster than predicted (Wetterstrand KA 2016). The price of whole genome sequencing (WGS) began at approximately one million dollars. In 2009, Illumina offered WGS with 30-fold coverage for approximately $48,000. The price of NGS has rapidly decreased since then to approximately $1,000 in 2015.

1.3 DNA Methylation

In contrast to changes in DNA sequence, epigenetic modifications are dynamic, reversible changes that occur throughout development and affect gene expression (Yan et al. 2016). The word epigenetics means "in addition to changes in the genetic sequence" and encompasses methylation, acetylation, phosphorylation, ubiquitylation and sumoylation (Weinhold 2006). The most widely-studied epigenetic modification is methylation, which typically occurs at the 5' position on cytosine within CpG dinucleotides and is widely associated with altered gene expression (Guenette et al. 1992). DNA methylation influences various genetic processes, such as: chromatin structure modulation, transcriptional regulation, genomic stability, X chromosome

inactivation, and silencing of parasitic elements (Robertson 2002). Three DNA methyltransferases (DNMTs) exist that catalyze the incorporation of a methyl group from S-adenosyl-L-methionine to cytosine: DNMT1, DNMT3A and DNMT3B (B. Jin and Robertson 2013). DNMT1 is responsible for maintenance of DNA methylation, while DNMT3A and DNMT3B are responsible for de novo methylation. Additionally, hydroxymethylation can occur at the same 5' position of cytosine when a ten-eleven translocation (TET) protein catalyzes the oxidation of the methyl group (Yong, Hsu, and Chen 2016).

Approximately 80% of CpG sites in the genome are methylated, while the other 20% remain unmethylated and are typically located in CpG islands near the promoter region of genes (Yong, Hsu, and Chen 2016). CpG islands are stretches of approximately 500-1000 bps of DNA with a CG:GC ratio of greater than 0.6. Most CpG islands remain unmethylated so as to allow for the interaction of proteins with promoter regions. On the other hand, gene body methylation, typically occurring at repetitive sequence sites, has been shown to affect expression by altering intron-exon boundaries thus impacting splicing. Yet, the three dimensional nature of DNA allows for physical interactions between DNA regions far apart in linear genetic space further complicating canonical models of the effect of methylation on gene expression (Yan et al. 2016).

Altogether, DNA methylation is an intricate and dynamic process that must be tightly coordinated to ensure proper genomic stability as well as temporal and special gene expression throughout development. When these processes are altered, normal cell progression and development are impaired. For example, Breast Cancer Type 1 and 2 Susceptibility genes *(BRCA1/2*) are essential for homologous recombination- (HR)

mediated DNA repair that are often mutated in hereditary cases of breast and ovarian cancer (Powell and Kachnic 2003). Epigenetic hypermethylation of *BRCA1* and *BRCA2* is associated with tumorigenesis in both sporadic and hereditary cases of various cancer types (Anjum et al. 2014). Likewise, epigenetic silencing of Werner Syndrome ATP-Dependent Helicase (*WRN*) via promoter methylation occurs in several cancers and leads to loss of protein and enzyme activity causing chromosomal instability (B. Jin and Robertson 2013).

Several technologies exist for profiling DNA methylation. Restriction enzyme-based methods employ methylation-sensitive restriction enzymes (BstUI, HpII, NotI and SmaI) that cleave unmethylated target sequences leaving methylated DNA intact (Yong, Hsu, and Chen 2016). This method can be coupled with DNA sequencing for the identification of methylated DNA regions. Affinity enrichment methods utilize methyl-CpG-binding domain (MBD) proteins or antibodies specific for 5-methylcytosine to enrich for methylated DNA regions. Similarly, methylated DNA immunoprecipitation (MeDIP) uses an anti-methylcytosine antibody to immunoprecipitated DNA containing methylated CpG sites and can be coupled with an MeDIP-chip array or sequencing. Bisulfite conversion methods exploit the ability of sodium bisulfite to deaminate unmethylated cytosine to uracil, leaving methylated cytosine unaffected thus providing single base resolution, and can be coupled with NGS. Illumina has developed several the Infinium Beadchips, which have been the tool of choice for larger epigenome studies due to the high coverage and low cost of the arrays (Yan et al. 2016). The Infinium HumanMethylation450 Beadchip harnesses the interaction between sodium bisulfite and DNA to cover over 450,000 CpG sites through amplification of sodium bisulfite-

converted DNA and hybridization to arrays. This allows for coverage of most CpG islands (96%), CpG shores (92%) and CpG shelves (86%) across the genome. Illumina's newest array, the Infinium MethylationEPIC Beadchip expands upon the HumanMethylation450 Beadchip by covering approximately 90% of those CpG sites in addition to about 200,000 more CpG cites covering enhancer regions defined by the Encyclopedia of DNA Elements (ENCODE) and Function Annotation of the Mammalian Genome Phase 5 (FANTOM5) projects. Whole-genome bisulfite sequencing (WGBS) technologies, such as BS-seq and methyl-seq assess methylation at almost every CpG site including gene deserts or intronic regions at a much higher cost. Reduced-representation bisulfite sequencing (RRBS) uses Msp1 restriction enzyme digestion, bisulfite conversion and NGS for interrogation of the methylation pattern of specific DNA fragments. This method is more cost-effective than WGBS but lacks coverage at intergenic and distal regulatory elements.

1.4 Epigenetic and Genetic Mechanisms of Complex Diseases

The overarching goal of genomic medicine is to identify DNA sequence variation associated with specific phenotypes (Antonarakis 2001). The expanded ability to perform high-throughput genetic association studies of diseases brought about the realization that very few diseases can be fully explained by one or even a small number of genetic variants (Hirschhorn et al. 2002). We can now appreciate that ~80% of disease-causing genomic associations are rare and lie in non-coding regions making biological understanding of the these associations difficult to interpret (Manolio, Collins, Cox, Goldstein, Hindorr, et al. 2009). Therefore, much of the heritability remains

unexplained for most diseases despite a large number of identified variants from numerous genome association studies (Manolio, Collins, Cox, Goldstein, Hindorr, et al. 2009).

The most straightforward genetic diseases to interrogate are monogenic diseases, which are caused by an abnormal mutant allele in one gene can be identified simply by comparing disease DNA to controls (Antonarakis 2001). Most often, determination of monogenic disease genes is performed by first assaying candidate genes for which the phenotype is likely explained. For example, Cystic Fibrosis (CF) is a monogenic disorder, characterized by elevated sweat chloride concentrations, exocrine pancreatic insufficiency, progressive obstructive lung disease and male infertility (Gallati 2014). The causative CF gene was identified in 1989 as Cystic Fibrosis Transmembrane Conductance Regulator (*CFTR*) on chromosome 7. Since then, almost 2,000 variants have been reported in *CFTR* as causative of CF. Another monogenic disorder among the first to be genetically characterized is Huntington chorea, which results in motor disturbances, psychiatric symptoms, and cognitive decline. The disease is fully explained by a trinucleotide expansion (CAG) in the Huntingtin (*HTT*) gene, for which a repeat length over 40 confers a 100% risk of developing the disease (Brinkman et al. 1997). Although the genetic location of the disease-causing loci was determined in 1983, the function of *HTT* has yet to be determined further supporting the notion that investigation of genomic determinants of disease is a highly complex process (Gusella et al. 1983)

Unlike the few determined monogenic diseases, such as CF and Huntington Chorea, most diseases are polygenic in nature (Manolio, Collins, Cox, Goldstein, Hindorr, et al. 2009). As Figure 1.1 demonstrates, most diseases are caused by a large number of

variants exhibiting a small effect on disease. According to the Online Mendelian Inheritance in Man (OMIM), only ~1.4% of the total number of genes in the genome (20,000-25,000) are responsible for mendelian, monogenic disorders (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore n.d.). Variants that greatly increase disease risk and interfere with fitness are under greater selective pressure, meaning that variants which remain in the human population mostly consist of very rare variants with large effects on disease and common variants with very small effects on disease.

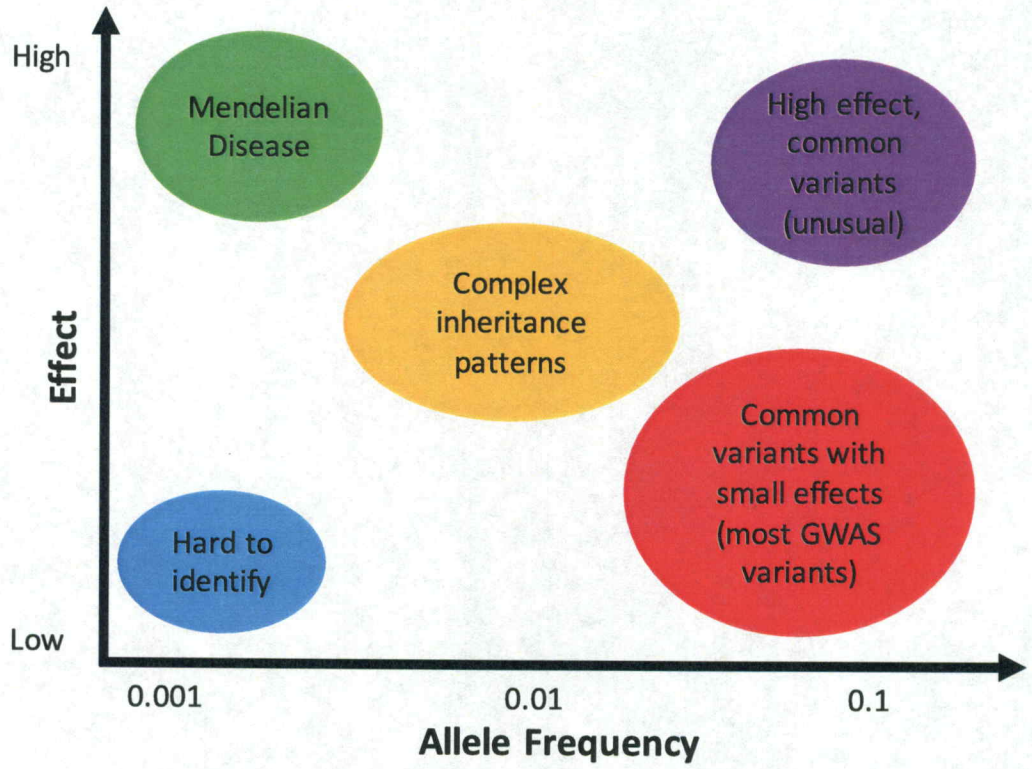


Figure 1.1 Genomic models of monogenic versus complex disorders.

Despite strong efforts in the field of genomic medicine, the determination of predisposing mutant alleles for common disorders has proven to be extremely difficult for multiple reasons, including: low allele frequency in the population, altered inheritance patters, uncertain phenotypes, variable age of onset, etc. (Antonarakis 2001). In order to detect rare variants with large effects on disease heritability, extremely large sample sizes are needed for adequate statistical power. Even with larger sample sizes, the ability to detect both rare variants, or those with a very small minor allele frequency (MAF < 0.05%), with large effects on disease heritability and common variants with very small effects on disease heritability pose the largest challenge to genomic analysis to date (Manolio, Collins, Cox, Goldstein, Hindorr, et al. 2009). Furthermore, most genome-wide association (GWA) studies thus far have focused on European populations so as to avoid false positives resulting from admixture and population structure, but in doing so have left out more genetically diverse populations, such as individuals of African ancestry. The failure to explore more genetically diverse populations further limits genetic association findings. Overall, these barriers make sample obtainment for GWA studies extremely difficult, meaning that unique and creative approaches to genetic analysis must be taken for successful genomic associations to be made in cohorts that are smaller than needed for appropriate genome-wide power.

Yet, researchers have proven that the determination of polygenic signatures, encompassing numerous sites across the genome, for complex diseases can be accomplished (Khera et al. 2018). For example, height is a highly heritable trait for which a study including over 250,000 individuals concluded that common genetic variants can explain ~60% of heritability (Wood et al. 2014). Researchers have since

expanded on that knowledge by building polygenic risk scores, capable of predicting one's height based purely on genetic variation (Krapohl et al. 2016). Additionally, polygenic risk scores have also been developed for cardiovascular disease, atrial fibrillation, type two diabetes, irritable bowel disorder and breast cancer among many other complex disease (Khera et al. 2018). In a cohort consisting of 184,305 European American individuals, Khera et al. was able to determine a polygenic risk score for cardiovascular disease (CAD) encompassing 6,630,150 variants which proved to be capable of successfully predicting CAD in a validation cohort with an Area Under the Curve (AUC) of 0.81 (Khera et al. 2018). The advantage of such polygenic scores is that they can be used to stratify individuals based on genetic variation as early as birth, which is most often much sooner than discriminative symptoms assessed clinically appear (Khera et al. 2018). Yet, the downside to undergoing polygenic risk score determination is that extremely large cohorts and subsequently large validation cohorts are required for proper analysis. Furthermore, most polygenic scores to date have been assessed and validated in cohorts of European ancestry, leaving out most genetically diverse individuals.

Despite large efforts in discovering the heritability of complex diseases, there still exists a large portion of "missing heritability" for most all complex diseases studied (Manolio, Collins, Cox, Goldstein, Hindorff, et al. 2009). Several researchers point to possible explanations for this missing heritability, which include the inability to account for all possible gene-gene and gene-environment interactions when conducting genomic studies (Trerotola et al. 2015). Epigenetic analysis of complex traits allows for the

interrogation of gene-environment interactions and explains much of the missing heritability for many complex diseases.

The first step in the initiation of gene expression involves alteration of chromatin structure, a process largely dependent on the local epigenetic landscape (Relton and Smith 2010). The epigenome is widely shaped by interactions with both genetics and the environment. These epigenetic responses are maintained and transmitted to next generations (Yan et al. 2016). Thus, epigenetic loci hold great potential for not only association studies but also for the determination of biomarkers and predictors of complex diseases.

A few drawbacks exist to such studies, with the first being the inability to determine the casual relationship between epigenetic changes and disease. To accomplish this, large cohorts with a longitudinal study design are needed. However, longitudinal studies require strong patient commitment that is often rare in the general population. Despite that drawback, many associations have been made between specific epigenetic changes and phenotypes such as: cancer, smoking, age, exercise and alcohol intake amongst many others (Relton and Smith 2010).

Another drawback to population-based epigenetic studies is the difficulty in determining the distinct epigenetic signature for individual cell types. Most samples assayed for methylation consist of multiple cell types, and so the epigenetic signature could be a result of differences in cell types rather than the phenotype of interest (Yan et al. 2016). Epigenetic studies performed on sorted cell types account for this, yet sorting cells sufficiently requires expensive equipment and can prove to be quite difficult. With continued research, technological advances and better patient recruitment strategies,

knowledge of the biological significance of epigenetic changes associated with disease will grow allowing for the exact relationship between epigenetic changes and disease to be determined and possibly lead to development of epigenetic targets and predictors for disease.

1.5 Aging as a Complex Disease

Aging is loosely defined as the time-dependent functional decline affecting all organisms, and the link between biological and phenotypic aging has gained much interest in the genomics field (Serrano et al. 2013). A common theme throughout all aging processes is the accumulation of genetic damage, which in theory should make studying the aging process relatively simple. However, the number of exogenous and endogenous factors challenging genetic integrity and stability throughout lifespan complicate aging research. While great strides have been made in the understanding of the genetic underpinnings of aging, much more work must be done in order to gain a complete understanding (Deelen et al. 2013). Traditional approaches to aging research have utilized model organisms, such as flies or mice, to identify hundreds of genes associated with aging (Costa-Reis and Sullivan 2013). This has led to a better understanding of pathways that are key to controlling homeostasis and ultimately longevity. Additional studies of model organisms have led to the formation the "hallmarks of aging" including: cellular senescence, mitochondrial dysfunction, deregulated nutrient sensing, loss of proteases, epigenetic alterations, stem cell exhaustion, altered intercellular communication, genomic instability, and telomere attrition (Serrano et al. 2013). One of the most researched pathways is insulin-like

signaling, as it has been established that longer-living organisms (i.e. humans) are more sensitive to insulin-harboring mutations causing lower expression levels of key proteins involved in the insulin-regulating pathway (Costa-Reis and Sullivan 2013).

With the advent of NGS, research has shifted from hypothesis-driven genetic analysis of aging in model organisms to GWA studies in humans consisting of large cohort sizes including those who have achieved a longer age. The increased life expectancy of family members of centenarians provides evidence for a strong genetic component of life expectancy, which could be captured in GWA studies (Santos-Lozano et al. 2016). However, the role that environmental factors plays is difficult to determine in itself, but also makes determination of genetic factors influencing aging throughout one's life more complex. Environmental factors known to influence aging include smoking, excessive alcohol intake, poor diet, toxic elements, and hard manual labor making it nearly impossible to control for these amongst many other factors that could confound GWA analysis of genetic factors associated with aging (Moskalev et al. 2014). Consequently, the only genetic factors with replicated evidence for genome-wide association with longevity are Forkhead Box O3 (*FOXO3A*) and Apolipoprotein E (*APOE*) (Erikson et al. 2016). *FOXO3A* is a transcription factor which plays an important role in cell cycle regulation (Eline Slagboom, van den Berg, and Deelen 2018). *APOE* is involved in lipid transfer throughout the bloodstream as well as injury repair in the brain (Shadyab and LaCroix 2015). Variation in *APOE* has also been widely associated with Alzheimer's disease and CAD, two aging-related diseases.

Interest has grown recently in the realm of "healthy aging" or "health span" which often has varied definitions but generally refers to the time span during which a person

lives without chronic illness and physical impairments (Martin, Bergman, and Barzilai 2007). Determining genetic associations with the healthy aging phenotype has proven to be extremely difficult. Most healthy aging and longevity-related studies have simply found "suggestive" genetic determinants (Erikson et al. 2016). For example, a recent study conducted on the Wellderly cohort of "well-aged individuals" failed to identify genetic associations with healthy aging but did identify associations with age-associated diseases, such as Alzheimer's disease and CAD. Together this has led to the hypothesis that healthy aging is more of a polygenic complex disease resulting from a genetically decreased risk of aging-related diseases (Erikson et al. 2016).

1.6 Summary

DNA provides the instructions for growth, development and reproduction of organisms. Approximately 1% of the genetic code differs across the human race. The advent of NGS technologies has allowed for high throughput interrogation of genetic variation across species. With this, we have obtained a deeper understanding of the genotype-phenotype relationships originally discovered by Gregor Mendel's experimental work in peas over 50 years ago. Yet, missing heritability for many phenotypes and in particular complex diseases still remains. Several lines of evidence supports the hypothesis that the missing heritability might be largely explained by epigenetic variation in individuals as a result of interaction between one's genetic makeup and environment. In this work I have expanded upon the analysis performed by others in pursuit of obtaining a better understanding the genetic and epigenetic

mechanisms of two complex diseases: accelerated aging and systemic lupus erythematosus (SLE).

CHAPTER 2

GENETIC VARIATION IN AGING-RELATED GENES AS A PREDICTOR OF AGING STATUS

2.1 Introduction

Exceptional longevity is influenced by a combination of environmental and genetic factors, and previous twin studies report that the heritability of biological aging is approximately 25% (Herskind et al. 1996). Familial studies have suggested that exceptional aging tends to run in families, yet the search for genetic determinants of longevity has produced inconsistent results (Sebastiani et al. 2012, 2013; Pilling et al. 2017). Several GWA studies have attempted to pinpoint genetic influences of healthy aging or longevity, yet only two loci, *APOE* and *FOXO3A*, have repeatedly reached genome-wide significance (Broer et al. 2015; Deelen et al. 2014). Thus, an alternative approach to understanding genetic factors underlying a complex phenotype like exceptional aging is warranted.

In my analysis, I utilized a comprehensive targeted sequencing approach designed to interrogate rare and common variants in both coding and non-coding regions within twenty key genes with strong evidence for involvement in aging related processes and have utilized traditional statistical and machine learning approaches to explore aging-related genetic variants. The twenty genes were chosen because they have previously been associated with various molecular functions involved in aging, such as DNA

damage response and repair, telomere maintenance, metabolism, and cellular stress resistance. There is ample evidence suggesting a causal role of DNA damage in aging and age-related diseases, for example most progeroid syndromes, including Werner Syndrome, Cockayne syndrome and Fanconi anemia, are characterized by accelerated aging possibly as a result of hypersensitivity to genotoxins predominantly due to problems with DNA repair and genome maintenance (Gensler and Bernstein 1981; Hoeijmakers 2009; Behrens et al. 2014; Sounni et al. 2011; Vermeij, Hoeijmakers, and Pathof 2016).

Several lines of evidence also suggest that levels of DNA damage increase with age, whereas DNA repair capacity in mammals reduces with age (Niedernhofer et al. 2018). Comparative studies in mammals further indicate that species longevity positively correlates with DNA repair efficiency (Hart and Setlow 1974). Long lived species such as the naked mole rat, *Heterocephalus glaber,* and bowhead whale, *Balaena mysticetus*, have a higher copy number of genes associated with DNA repair possibly allowing for decreased susceptibility to age-accumulated DNA damage (Macrae et al. 2015). Therefore, I hypothesized that variants associated with DNA repair, telomere maintenance and genomic stability could be predictive of biological age. To interrogate this hypothesis, 20 genes were picked for analysis based on 1) association with the aging phenotype, 2) association with aging-related phenotypes or 3) association based on animal studies. Table 2.1 provides each gene included in the study, along with biological function, biological association with aging or age-related pathology and literature references for study inclusion. In brief, *Alipoprotein E* (APOE) is a gene involved in cholesterol and lipid metabolism, which has been associated with both neurodegenerative disorders and human longevity (Soto et al. 2015; Broer et al.

2015; Pilling et al. 2017). Many of the genes chosen are associated with progeroid disorders, which are characterized by accelerated aging. *Fanconi Anemia Group A Protein* (FANCA) is a DNA repair protein, and mutations in this gene are responsible for a rare progeroid disorder, Fanconi Anemia (Soria-Valles and López-Otín 2016; Schumacher, Garinis, and Hoeijmakers 2008; Krishnan, Liu, and Zhou 2011). Likewise, *Werner Syndrome RecQ like Helicase* (WRN) is a gene involved in genomic maintenance and stability, and mutations in WRN lead to the progeroid disorder known as Werner syndrome (Multani and Chang 2007; Bendtsen, Juul, and Trusina 2012; Mohaghegh and Hickson 2002). Genes associated with telomere maintenance for which mutations are linked to telomeropathies were included in the panel of genes for this work, based on evidence supporting the necessity of telomere maintenance in aging (Aubert and Lansdorp 2008; Opresko and Shay 2017). *TERF1 Interacting Nuclear Factor (TINF2)* helps regulate telomere length and mutations lead to Revesz syndrome, which is characterized by symptoms characteristic of premature aging (Savage et al. 2008; Kim, Kaminker, and Campisi 1999; Rubelj and Vondraček 1999). Many genes involved in DNA repair were also included for analysis based on the previously mentioned evidence for the association of DNA repair and aging. For example, *Bloom Syndrome RecQ Like Helicase* (BLM) is involved in double-strand break repair and defects are associated with segmental aging of the immune system, an elevated risk for diabetes mellitus and cancer incidence (de Renty and Ellis 2017; Karow, Wu, and Hickson 2000; Coppede and Migliore 2012).

Table 2.1 Names, biological function and literature references for aging association of the 20 genes sequenced.

Gene	**Function**	**Biological association with aging/ age-related pathology**	**Literature reference for study inclusion**
Apolipoprotein E (APOE)	Combines with lipids to form lipoproteins which package cholesterol and other fats for transfer through the bloodstream.	Polymorphisms in APOE are associated with human longevity.	(Soto et al. 2015; Broer et al. 2015; Pilling et al. 2016)
Aprataxin (APTX)	Involved in DNA break repair and base excision repair.	Defects in Aprataxin cause the autosomal recessive neurodegenerative disorder Ataxia Oculomotor Apraxia 1 (AOA1).	(Krishnan, Liu, and Zhou 2011; Coppede and Migliore 2012; Katyal and McKinnon 2008)
Bloom Syndrome RecQ Like Helicase (BLM)	ATP-dependent DNA Helicase. Unwinds DNA in the 3'-5' direction. Involved in double-strand break repair.	Defects associated with segmental aging of immune system together with an elevated risk of otitis media and pneumonia, an elevated risk of diabetes mellitus, reduced fertility, and higher cancer incidence.	(de Renty and Ellis 2016; Karow, Wu, and Hickson 2000; Coppede and Migliore 2012)
Cyclin Dependent Kinase Inhibitor 2A (CDKN2A)	Induces cell cycle arrest and acts as a tumor suppressor.	Mutations near CDKN2A were particularly associated with disease of aging (e.g., cancer, atherosclerosis, type 2 diabetes, glaucoma). CDKN2A	(Baker, Jin, and Van Deursen 2008; Shiels 2010)

		expression increases with age. Removal of p16+ve cells in mouse models increases health span and lifespan.	
Sialic Acid Binding Ig-Like Lectin 3 (CD33)	Mediates cell-cell interactions and maintenance of immune cells in the resting state.	Mutations in CD33 are associated with AD risk.	(Griciuc et al. 2013; Estus et al. 2019)
Dyskerin Pseudouridine Synthase 1 (DKC1)	Stabilization and maintenance of telomerase.	Mutations in DKC1 causes premature aging, bone marrow failure and cancer.	(Gu et al. 2011; Blasco 2007)
Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 4 (ERCC4)	Catalytic component of a DNA repaid endonuclease responsible for 5' incision during DNA repair.	Loss of ERCC4 causes systemic accelerated aging (XFE) and neurodegeneration.	(Muñoz et al. 2005; Q. Yuan et al. 2014; Bogliolo et al. 2013)
Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 5 (ERCC5)	Endonuclease involved in single-strand DNA nucleotide excision repair at the 3' end.	Mutations in ERCC5 lead to Cockayne Syndrome (CS), which is characterized by premature aging.	(Coppede and Migliore 2012)
Excision Repair Cross-Complementing Rodent Repair Deficiency, Complementation Group 6 (ERCC6)	DNA-binding protein involved in transcription-coupled nucleotide excision repair.	Defects in ERCC6 cause CS and age-related macular degeneration.	(Tuo et al. 2006; Baas et al. 2010)
Fanconi Anemia Group A Protein (FANCA)	DNA repair protein involved in Interstrand Crosslink (ICL) repair.	Defects cause Fanconi Anemia, a progeroid syndrome with symptoms common in premature aging (sarcopenia,	(Soria-Valles and López-Otín 2016; Schumacher, Garinis, and Hoeijmakers 2008; Krishnan, Liu, and Zhou 2011)

		hypersensitivity to infectious agents, endocrine abnormalities, etc.).	
Lamin A/C (LMNA)	Component of the nuclear lamina.	LMNA mutations cause Hutchinson-Gilford syndrome (HGPS).	(Kawahara et al. 2011; Lopez-Mejia et al. 2011; Rodriguez et al. 2009)
Poly(ADP-Ribose) Polymerase 1 (PARP1)	Mediates poly-ADP-ribosylation of proteins and plays a role in DNA repair, chromatin remodeling, telomere maintenance and mediator of inflammation.	PARP1 activation increases with age in *C. elegans*. Increased activation has been associated with aging, neurodegeneration and metabolic abnormalities in humans.	(Coppede and Migliore 2012; Krishnan, Liu, and Zhou 2011; Maynard et al. 2015)
DNA Polymerase Beta (POLB)	DNA polymerase involved in base excision and repair.	*Polb*$^{+/-}$ mice have an increased age-related mortality rate and tumorigenesis.	(Cabelof et al. 2002; Strosznajder, Jęśko, and Strosznajder 2000)
DNA Polymerase Gamma (POLG)	Involved in mitochondrial DNA replication.	Increased mitochondrial mutation load in mice is associated with premature aging.	(Hiona and Leeuwenburgh 2008; Kujoth et al. 2005; Trifunovic et al. 2004)
Sirtuin 1 (SIRT1)	NAD-dependent protein deacetylase. Involved in cell cycle regulation, response to DNA damage, metabolism, apoptosis and autophagy.	SIRT1 overexpression extends lifespan in mice. Mutations are associated with age-related pathologies such as myocardial infarction (MI).	(Grabowska, Sikora, and Bielak-Zmijewska 2017; Y. Yuan et al. 2016; Satoh et al. 2013)
Sirtuin 6 (SIRT6)	NAD-dependent protein deacetylase. Deacetylase activity towards histone H3K9Ac and H3K56Ac. Required for genomic stability.	SIRT6 overexpression extends lifespan. Long-lived animals have highly efficient SIRT6 function.	(Moskalev et al. 2014; Schumacher, Garinis, and Hoeijmakers 2008; Berman et al. 2012; Serrano et al. 2013)

	Deacetylates telomeric DNA.		
Superoxide Dismutase 2 (SOD2)	Destroys superoxide anion radicals produced in cells.	SOD2 mutations are associated with heart disease and increased risk of malignancies.	(Fabrizio et al. 2004; Velarde et al. 2012; Patel 2002; Qiu et al. 2010)
Telomerase Reverse Transcriptase (TERT)	Ribonucleoprotein polymerase that maintains telomere ends by the addition of the telomere repeat TTAGGG.	Telomere attrition is highly associated with aging due to increased cellular senescence.	(Martinez and Blasco 2010; Blackburn, Epel, and Lin 2015; Aubert and Lansdorp 2008; Ghosh and Zhou 2014)
TERF1 Interacting Nuclear Factor 2 (TINF2)	Component of the telosome that is involved in telomere length regulation and protection.	Mutations in TINF2 are linked to Revesz syndrome, a telomeropathy with symptoms characteristic of accelerated aging.	(Savage et al. 2008; Kim, Kaminker, and Campisi 1999; Rubelj and Vondraček 1999)
Werner Syndrome RecQ like Helicase (WRN)	DNA helicase that is involved in maintenance of genomic stability, DNA repair, replication, transcription and telomere maintenance.	Mutations in WRN lead to Werner syndrome with systemic aging phenotypes	(Multani and Chang 2007; Ding et al. 2007; Bendtsen, Juul, and Trusina 2012; Mohaghegh and Hickson 2002)

Single variant association tests, such as linear regression, have been the statistical tools of choice for large GWA studies. In fact, most longevity-targeted GWA studies have taken this approach (Sebastiani et al. 2012; Broer et al. 2015; Pilling et al. 2017). However, such univariate models leave out epistatic effects that may be predictive of heterogeneous diseases, such as aging, meaning the actual number of genetic factors contributing to or predictive of polygenic diseases are often precluded (Stephan, Stegle, and Beyer 2015). More complex statistical approaches would account for genetic factors that alone have little

association but when considered in a multiplex manner hold great predictive power. Random forest and support vector machines (SVM) are just two of a multitude of ensemble learning methods capable of analyzing large data sets such as those obtained in GWA studies. Random forest couples bootstrap sampling and conditional inference trees for determining the importance of variables for classifying data (Lunetta et al. 2004). I sought to use random forest in my analysis of healthy aging as it is capable of handling sizable data sets, considers the interactions between variables, and provides importance measures for predictors. On the other hand, SVM is a type of supervised learning that not only supports high dimensional data but is robust against noise and sparsity in the data (Furey et al. 2000). SVM function by taking a set of input features or data and defining an optimal decision boundary or hyperplane that most accurately separates the input space based on assigned binary classifiers. Together these factors allow for better determination of genetic predictors in polygenic diseases that might be due to nonlinear interactions in both common and rare variants (Lunetta et al. 2004).

For this study, a panel of twenty aging-related genes was sequenced with a targeted sequencing method previously developed by the Devin Absher Lab at the HudsonAlpha Institute in a cohort of 200 individuals selected from the University of Pittsburgh Claude D. Pepper Older Americans Independence center (Day, Song, and Absher 2014). Half of the cohort was labeled as "early" agers, as determined by age (65-75 years old) and the inability to either walk up a flight of stairs or walk for 15 minutes without resting. The other half of the cohort is labeled as "late" agers due to age (>75 years old) and their ability to pass the walking tests performed on the "early" agers.

After applying univariate and multivariate analyses to the sequencing data, I show that a decision tree-based method, random forest, trained on genetic markers in the discovery cohort shows promise in predicting "phenotypic age". Despite the fact that a sample set of 200 is small for this genomic study, I show that this exploratory analysis to determine genetic predictors of aging provides a useful and novel mechanistic approach for investigating the association of polygenic risk variants with complex diseases. Further analyses with larger cohorts would find this approach valuable for determining a set of genetic variants which alone would not hold predictive value but in combination are highly predictive of accelerated aging. A predictive model of accelerated aging would not only give insight into key biological processes of this complex phenotype but could potentially be used in a clinical setting as a diagnostic tool to indicate patients that may be at risk for early onset of age-related diseases.

2.2 Methods

2.2.1 Discovery set University of Pittsburgh Medical Center (UPMC) Cohort participants

Participants were recruited from several sources with the help of Dr. Susan Greenspan and Dr. Neil M. Resnick with informed consent in accordance to the University of Pittsburgh Institutional Review Board IRB#: REN17120030 / PRO14010101. Funds were provided by the National Institutes of Health grant for the University of Pittsburgh Pepper Older Americans Independence Center (P30AG024827) and by discretionary monies from the Office of the Senior Vice Chancellor for the Health Sciences, University of Pittsburgh, a non- profit entity.

Most participants were obtained through the University of Pittsburgh Claude D. Pepper Older Americans Independence center, which maintains a registry of more than 2,500 older adults who live in the greater Pittsburgh area and are interested in participating in clinical research. Print and radio ads were also used. Respondents were screened with a standardized phone interview. Most respondents (~90%) were of self-reported Caucasian ethnic background (Figures A.1 and A.2). Assessments were performed under the guidance of Dr. Susan Greenspan, Dr. Neil M Resnick and Dr. Arthur S. Levin. Initial statistical analyses between groups were conducted with the help of Dr. Subashan Perera and Dr. Aditi U. Gurkar.

2.2.1.1 Assessments

- **Demographic information:** Age, gender, level of education, and smoking status.
- **Body composition:** Height, weight and dual x-ray absorptiometry (DXA) to measure total fat, lean body mass and calculate body mass index (BMI).
- **Cognitive function:** Montreal Cognitive Assessment (MOCA) and Digit Symbol Substitution Test (DSST). Higher scores indicate better cognitive function.
- **General health:** Comorbidities were assessed using a comorbidity index (Rigler et al. 2002) ; a higher score suggests a greater number of comorbidities and poorer health (Sangha et al. 2003). The SF-36 measured patients' self-reported health and wellness; higher scores indicate better health (J.E. Ware 1992). Finally, participants were characterized as frail, prefrail, or robust using

the five-item Fried Frailty Index; higher scores indicate frailty (Figure A.3) (Abellan Van Kan C, Geriatric, and Panel 2008).

- **Function and activity:** The Community Healthy Activities Model Program for Seniors (CHAMPS) Physical Activity Questionnaire was used to assess the frequency of activity and estimate calories per week involved in the activity (Stewart et al. 2001). Grip strength was assessed with a standard dynamometer. The short physical performance battery (SPPB) was used, which provides an integrated physical assessment based on several measures, including gait speed, chair stand, and balance; a higher score indicates better performance (Vasunilashorn et al. 2009).

2.2.2 Validation set (Wellderly cohort)

The Wellderly Cohort consists of individuals of at least 80 years of age with no chronic disease or need for chronic medications. Sample collection and processing for WGS as well as variant calling are previously described (Erikson et al. 2016). Individuals used in this study had an average age of 86 and consisted of less males (n = 195) than females (n = 316). Comparison of overlapping clinical features in the discovery and validation cohorts were assessed to ensure a similar population distribution (Figure A.4). Furthermore, the cohort contains no enrichment for longevity variants.

Table 2.2 Comparisons of variables between aging cohorts: mean ± standard *Computed using independent samples t-, Wilcoxon rank sum, or chi-square tests, as appropriate.

Demographics	Early Aged (n = 100)	Late Aged (n=100)	Early vs Late Aged p-value
Age	70.4±3.0	83.2±5.4	<0.0001
Sex (% female)	63 (63.0)	56 (56.0)	0.3133
Comorbidity Scale	4.4± 1.8	2.5±1.6	<0.0001
Gait Speed (m/s)	0.92 ± 0.24	1.08 ± 0.26	<0.0001
BMI	33.5± 8.3	27.2±4.6	<0.0001
Lean Body Mass (kg)	53.±11.8	47.4±9.9	0.0002
Total Mass (kg)	91.6±23.5	73.6±15.3	<0.0001
% Fat Body Mass	37.9± 8.6	32.2±7.8	<0.0001
MOCA	25.3±2.8	24.3±3.5	0.03
DSST Score	42.2 ± 9.5	39.7 ± 10.7	0.0808
Grip Strenth - dominant (kg)	26.7±10.8	26.7±10.6	0.9791
Chair Rise Time	14.7±13.8	12.4±11.8	0.0001
SPPB Total Score	9.1±2.5	10.2±1.8	0.0005
Balance Score	3.4±1.0	3.6±0.7	0.1873
Calories from all Activity Per Week	2320±2186	3585±3059	0.001
Calories from Moderate Activity Per Week	929±1495	2018±2322	0.0001
Freq. of all Activity Per Week	13.9±9.8	19.7±10.6	<0.0001
Freq. of Moderate Activity Per Week	4.3±5.0	7.2±6.4	0.0003
Frail Scale	2.6 ± 1.3	0.6 ± 0.9	<0.0001
Physical Function Index	37.3±19.1	77.2±17.2	<0.0001
General Health Perception	52.8±22.2	78.2±14.3	<0.0001
Bodily Pain	44.5±22.4	76.3±19.7	<0.0001
Social Function	69.0±24.5	92.1±15.7	<0.0001
Mental Health Index	65.2±14.3	75.8±9.4	<0.0001
Vitality	47.7±14.1	66.5±11.8	<0.0001

2.2.3 Participant group determination

The goal in participant group determination was to maximize the signal with respect to any genetic differences between the groups. Because there is no standard operational criterion for defining early and late agers, self-reported and performance-based measures of mobility we used (Abellan Van Kan C, Geriatric, and Panel 2008), as they are strongly associated with incident disability (Perera et al. 2014) and mortality (Perera et al. 2016) in the elderly. As such, "early aged" participants were defined as those 65-75 years of age who could not walk up a flight of stairs or walk for 15 minutes without resting; and "healthy aged" were defined as those age 75 years and older who could walk up a flight of stairs and walk for 15 minutes without resting. Participants with

a history of a major cancer were excluded. Table 2.2 depicts the differences in participant characteristics between groups.

2.2.4 Variant genotyping

Clone adapted template capture hybridization sequencing (CATCH-Seq) was used with the help of the Absher Lab as an alternative to other sequencing methods due to the low cost and high coverage ability of both coding and noncoding genomic regions (Day, Song, and Absher 2014). CATCH-Seq yield is comparable to whole genome sequencing (89% versus 98% at 100x) at a fraction of the cost. This allows for more samples to be included in the study when only a small set of genes are under investigation, as is the case in this study. CATCH-Seq probes were designed to capture ~150-200 kilobase (kb) regions around each of the 20 target genes (Table 2.2) (Day, Song, and Absher 2014). Standard Illumina sequencing libraries were hybridized to the CATCH-Seq probes and the target-enriched libraries were subjected to 2 x 100 base pair (bp) paired-end sequencing on HiSeq2500 sequencers. The resulting sequence data was aligned to the human reference genome (GRCh37) with BWA (H. Li and Durbin 2009) and variants were called using GATK v2 (McKenna et al. 2010) with exclusion filters for variants with low mapping quality (mapq<20) and low genotype quality (q<30).

2.2.4 Quality control

2.2.4.1 Variant inclusion criteria

Quality control analysis was performed with mentorship and guidance from Dr. Devin Absher. The initial datasets consisted of 25,273 variants in the discovery cohort

and 8,018 variants in the validation cohort (Table A.1). Variants with less than 8 alleles in the discovery cohort and over 10% missing data were excluded. Variants not covered in both the discovery and validation cohort were also excluded. Variants were then imputed across individual genes +/- 50kb using K-nearest neighbor imputation via the impute package in R (Hastie et al. 2001). A total of 5,896 variants was selected for further analysis.

2.2.5 Statistical Analysis

2.2.5.1 Total variance analysis

The sum of all variance between groups was analyzed using a Wilcoxon rank-sum test to determine whether "early" agers had more or less genetic variance in the target genes compared to "late" agers.

2.2.5.2 Single variant association

Logistic regression was utilized to assess the association of any single variant to the age group phenotype. A quantile-quantile (QQ) plot was used for evaluation of the distribution of p-values.

2.2.5.3 Gene association

Wilcoxon rank-sum tests were used to compare the distribution of CADD scores of non-reference alleles near target genes (+/- 50kb) between early and late agers. P-values were adjusted for multiple hypothesis via the Bonferroni method.

2.2.5.4 Predictive modeling

Four-fold cross-validation with four different seeds using a random forest regression model via the RandomForest package in R as well as SVM classification via the e1071 package in R were conducted for predictive modeling of the aging phenotype (Liaw and Wiener 2002; Dimitriadou et al. 2005). Default settings for number of trees grown (n = 500) and number of variables tried at each split (mtry = 6) were used for each random forest model. An SVM model was tuned using a range of costs (c = 0.1, 1.0, 10.0, 100.0) and gamma values (gamma = 0.5, 1, 2). Both random forest and SVM modeling were performed on 28 different stratifications of the data in addition to a control data set (Table A.2) resulting in 928 models in total. Most of the data subsets consisted of different groups of genomic spaces within the sequenced data as well as filters for frequency and deleteriousness. The first subsets of the data contained all sequence variants in addition to groups with different filters, including a subset of rare variants (tAF<0.1), very rare variants (tAF<0.01), mildly deleterious and highly deleterious variants as defined by the Combined Annotation-Dependent Depletion (CADD) score (CADD>10 and CADD>15 respectively). I then took subsets of only the variants within the start and end site of the target genes, and then applied the same filters as the first to analyze rare (tAF<0.1), very rare (tAF<0.01), mildly (CADD>10), and highly (CADD>15) deleterious variants. The next set of subsections contained target gene variants plus 50kb up- and downstream of the transcription start and end sites to capture regulatory genomic space within the analysis. Once again, the same cutoffs for allele frequency and CADD score were applied. The last genomic space stratification included variants within exons of the target gene isoforms, thus eliminating intronic

space from the models. Allele frequency and CADD score cutoffs further stratified the exonic variant subset. In addition to stratifications of the genomic space, publicly available databases such as the Genome-Wide Repository of Associations Between Phenotypes (GRASP), the SNP and copy number annotation (SCAN) database and software such as SIFT (Sorting Intolerant From Tolerant) were utilized for grouping the data based on variant effect (Lonsdale et al. 2013; Leslie, O'Donnell, and Johnson 2014; Gamazon et al. 2010; Ng and Henikoff 2003). For this I analyzed known versus unknown variant models, SIFT deleterious variants vs SIFT tolerated variants, variants effecting expression, and GWA variants. Lastly, I included a control set which was made by randomly shuffling all of the variants.

I assessed the performance of each model using receiver-operating characteristic (ROC). Additionally, I used Bayesian Classifier to determine the optimal cut-off between early and late agers in the random forest regression analysis. Top performing SVM and random forest models were tested on the validation (Wellderly) cohort of late agers, and the misclassification percentage, based the optimal cut-off, was used to rank each model rather than ROC-AUC since the cohort is made up of a single class (late agers) rather than the binary class available in the discovery cohort. Top classifiers in the best performing random forest model were determined by analyzing the Gini importance measures (Gini coefficient) for each split in the top models, which gives a measure of variable importance. In other words, the higher the Gini coefficient the better the classifier is at accurately splitting the data between two classes.

2.2.6 Enrichment Analysis

Enrichment for specific genomic domains and functions within the top variants was determined using a variety of tools. Enrichment of rare or severely deleterious variants was analyzed by assessing allele frequency and CADD scores of the top variants. I utilized the UCSC Genome Browser for determination of the specific location of each variant for analysis of intronic or exonic SNP enrichment (Kent et al. 1976). GRASP was used to discover whether the top classifying SNPs have been associated with specific phenotypes previously (Leslie, O'Donnell, and Johnson 2014). The Roadmap Epigenomics Project database was used to ascertain how many top variants were within regulatory regions via data from the HepG2 hepatocellular carcinoma cell line as well as GM12878 lymphoblastoid cells (Chadwick 2012). Lastly, enrichment for transcription factor binding sites within the top 50 variants was assessed using data from the ENCODE database (Encode Consortium, Carolina, and Hill 2013).

2.3 Results

2.3.1 Logistic regression

The data were first analyzed using logistic regression analyses to identify 1) single variants associated with the early or late aging groups, and 2) single genes carrying a combination of genetic variants associated with aging group. Neither the univariate, nor the gene-based multivariate analyses yielded statistically significant associations with aging group.

To identify high-impact aging-related variants, variants were tested for association with the aging group using logistic regression. Top variants were within

intronic and upstream regions of Lamin A (*LMNA)* (rs915180, p value = 0.0015) and *WRN* (rs6989940, p value = 0.0017), however none of the top hits reached significance beyond what would be expected by chance given the number of individual variant tests. A QQ plot of the logistic regression p-values indicated deflation as a result a lack of power owing to the small sample size in this study (Figure 2.1).

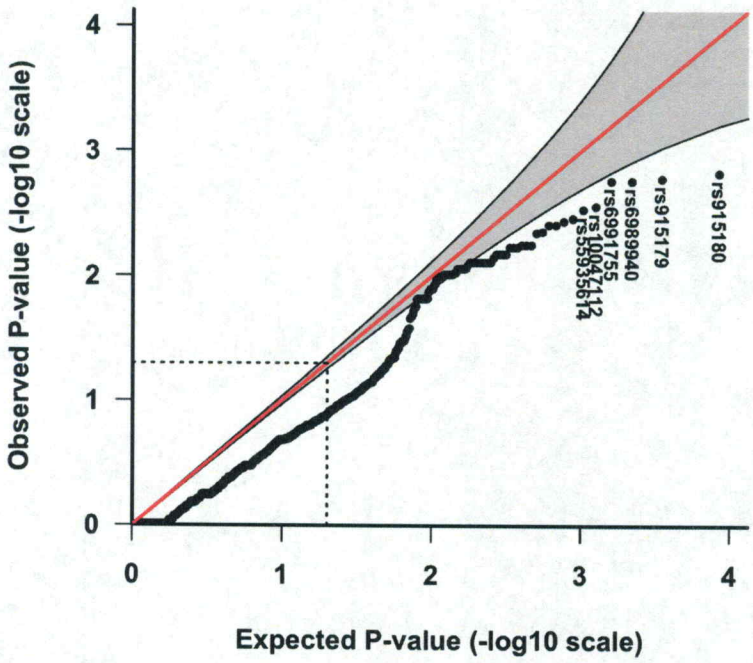


Figure 2.1 QQ plot of the single variant logistic regression p-values (-log10).

2.3.2 Variant burden

I combined the number of alternate alleles among all twenty genes in each subject following simple inclusion criteria of the variants for quality control to determine if early agers had a larger variant burden in aging-related genes compared to late agers and found

no significant difference (Wilcoxon p value = 0.75) (Figure 2.2). This method was then repeated for each individual gene, for which I compared the total amount of non-reference alleles in early agers compared to late agers in order to test whether the variant burden in that gene differed between groups. There was little difference in total non-reference allele count per target gene between early and late agers for most of the genes analyzed (Figure A.5). However, *LMNA* approached the Bonferroni corrected p-value of 0.003 according to a Wilcoxon rank-sum test (p-value = 0.006, FDR = 0.1).

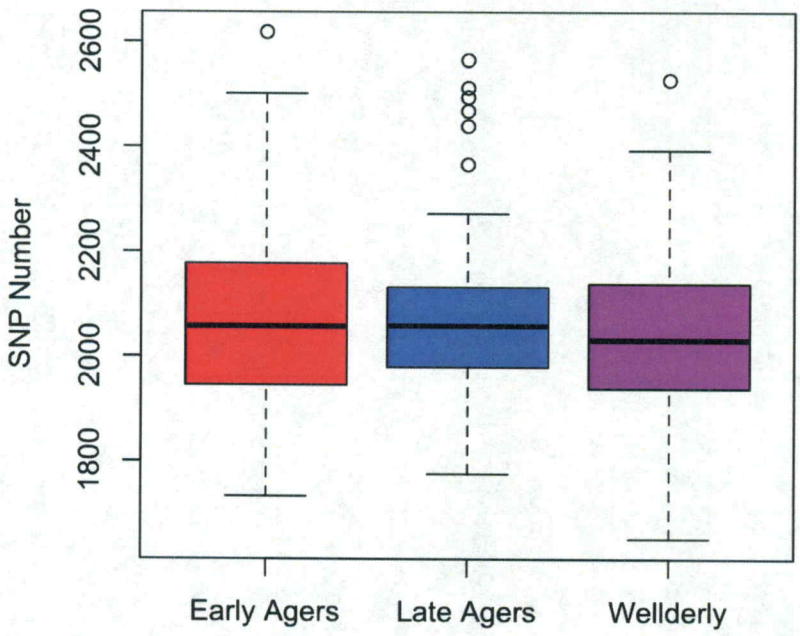


Figure 2.2 Preliminary Analysis Boxplots showing number of variants called per individual between groups. Wilcoxon rank-sum p-values: Late Agers vs Early Agers (p = 0.76), Late Agers vs Validation (p = 0.28), and Early Agers vs Validation (p = 0.14).

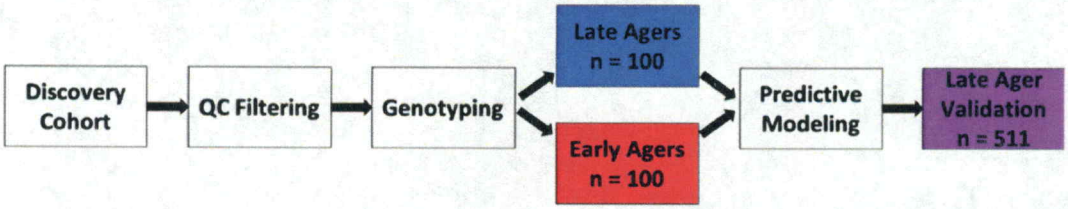


Figure 2.3 Flow chart of study design.

2.3.3 Machine learning

The next computational approach was geared at determining the predictive power of my sequencing data for aging status. Both random forest and SVM were applied to the variant data to determine the best genetic predictors of late aging. The ability of both the random forest algorithm and SVM to outperform other non-parametric classification methods led my use of these predictive modeling approaches in this study (Lunetta et al. 2004; Furey et al. 2000). As depicted in Figure 2.3, the training cohorts were divided into early and late agers for random forest model training, and top performing models according to the ROC-AUC were then tested for prediction of aging status in the validation cohort. Various stratifications of the data were fed into each algorithm to determine the best subset of predictors. These subsets included: variants of both low and high allele frequencies, variants that are known to effect expression (eQTL) defined by the SCAN database, variants previously associated with aging determined by the GRASP database, functional variants determined by ENCODE, variants with low and high levels of deleteriousness as defined by the CADD scores, as well as variants near or within the target genes. Four-fold cross-validated random forest at four different seeds was

performed on these various filters of the variant data as previously described resulting in a total of 16 models per filter, or 464 total models.

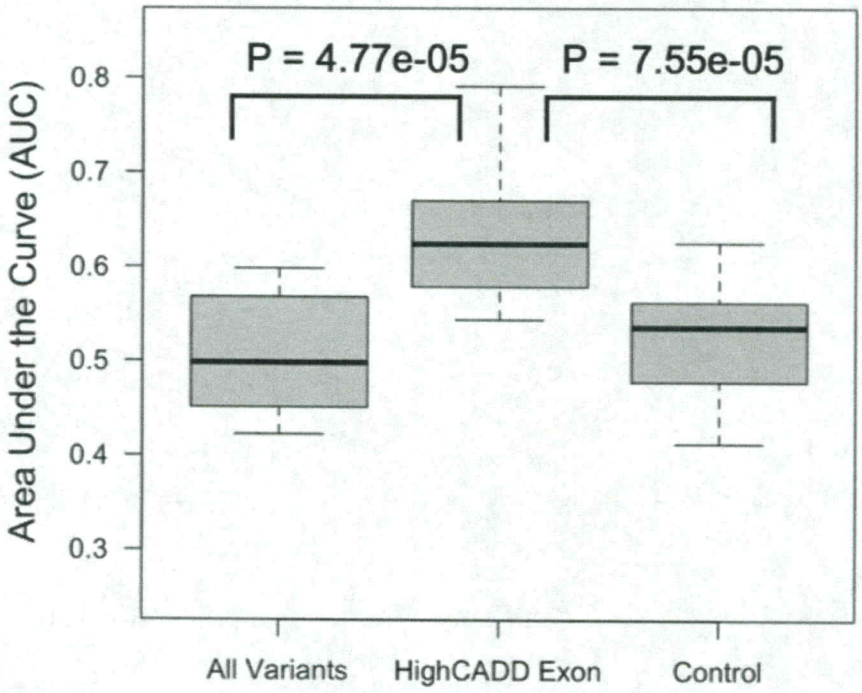


Figure 2.4 Boxplots of the area under the curve (AUC) for the random forest model performed using all variants (n = 5896), using alternate variants within exons that have a CADD score over 15 (n = 20), and the control dataset (shuffled all variants data frame, n =5896). Wilcoxon rank sum tests between all AUCs for each model were calculated showing a significant difference between both the all variant model (p = 4.77e-07) as well as the control model (P = 7.55e-05) and the model containing exonic variants with a CADD score over 15.

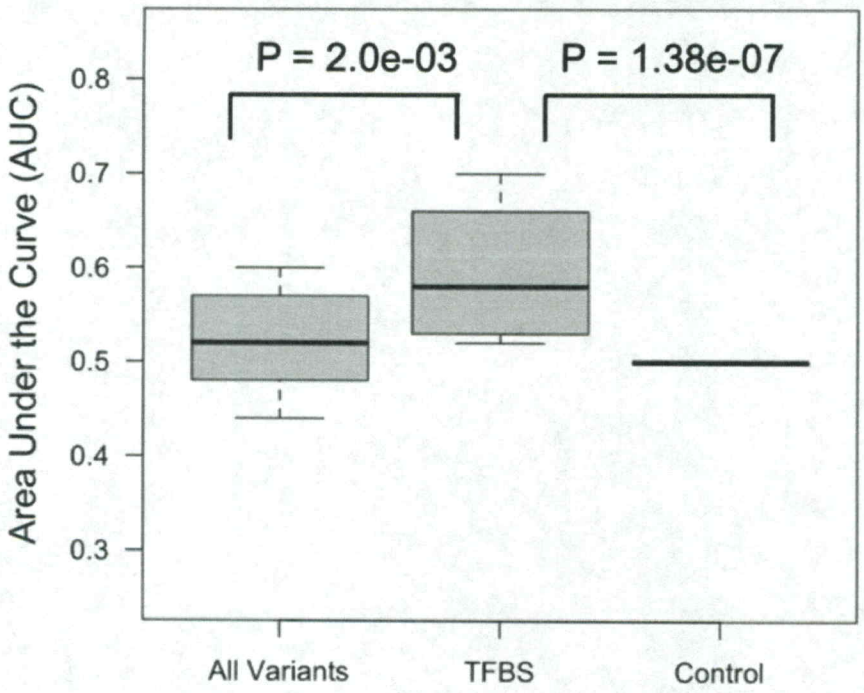


Figure 2.5 Boxplots of the AUC for the SVM model performed using all variants (n = 5896), using alternate variants within TFBSs (n = 1180), and the control dataset (shuffled all variants data frame, n =5896). Wilcoxon rank sum tests between all AUCs for each model were calculated showing a significant difference between both the all variant model (p = 2.0e-03) as well as the control model (P = 1.38e-07) and the model containing alternate variants within TFBSs.

The distribution of ROC-AUCs, a measure of model sensitivity and specificity, was compared to identify the top performing models (Table A.4). Random forest performed on the non-reference alleles within the exons of the twenty target genes having a CADD score greater than fifteen showed the greatest performance (mean ROC-AUC = 0.62) among random forest models, while the model trained on non-reference alleles within TFBSs proved to have the highest performance amongst all SVM models (Figures 2.4 and 2.5) but failed to outperform the top random forest model. This model proved to

outperform that of all sequenced variants (mean ROC-AUC = 0.51) (Figures 2.6 and 2.7). For analysis of model predictive power in an independent cohort, I tested the ability of the top random forest model to correctly identify the validation (Wellderly) cohort as late agers. As previously stated, because this cohort lacked any early agers I used percent misclassification rather than ROC-AUC to assess prediction accuracy as ROC-AUC assessment requires two groups. This analysis revealed that the top model performed well on the model validation (Wellderly) cohort (median misclassification = 0.02) (Figure 2.8). Additionally, smoking status, which is known to affect aging, was tested as a predictor of age group for comparison of genomic data to environment in predicting aging status, revealing that my model built on high CADD exon variants in aging-related genes performed comparably (Figure 2.9) (Valdes et al. 2005; Astuti et al. 2017; Csiszar 2009; Bosse et al. 1980). Lastly, because there is a significant difference in BMI between early and late agers ($p = 5.6 \times 10^{-8}$), I tested the correlation between the predictor value and BMI in the discovery cohort for the top performing model which revealed little correlation between age group prediction and BMI (Spearman Rho = 0.07) (Figure A.6). Furthermore, a scatterplot of the predicted age group from the best model (mean ROC-AUC = 0.62) versus BMI in both cohorts details a lack of trend between the two values further supporting that this is a model predictive of early versus late aging rather than BMI (Figure A.7).

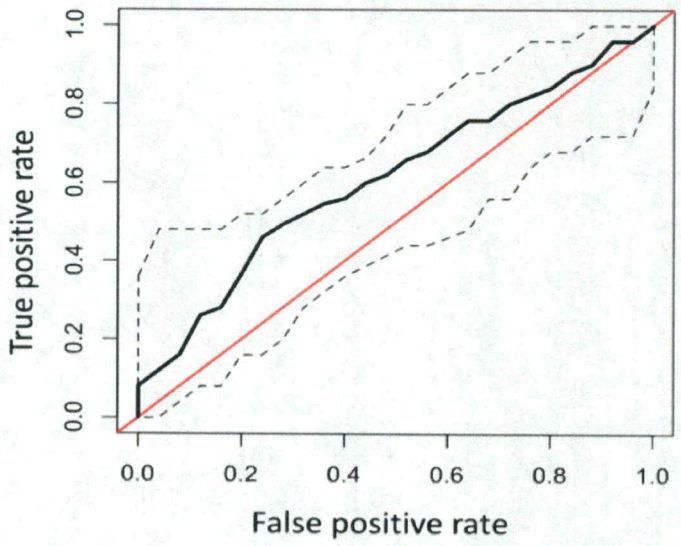


Figure 2.6 ROC curve for the model performed on the discovery cohort resulting in the median AUC (0.62) within the best performing data set (high CADD exon variants) with confidence intervals representing the best and worst AUC for the dataset.

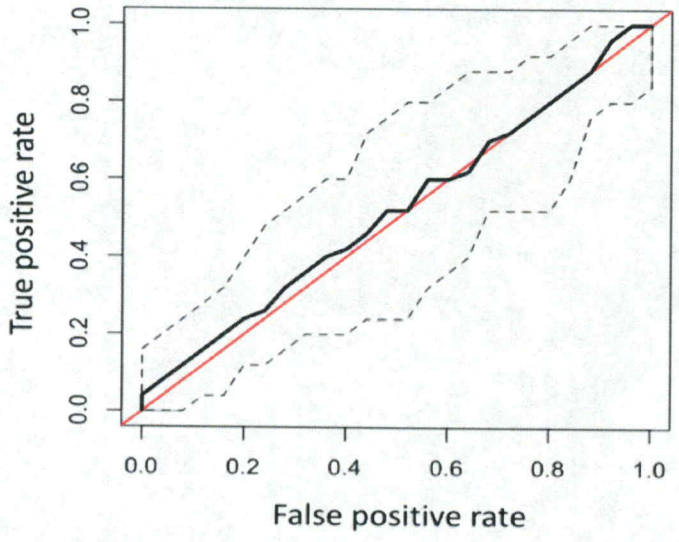


Figure 2.7 ROC curve for the all variant model resulting in the median AUC (0.51) with confidence intervals representing the best and worst AUC for the data set.

One of the most advantageous aspects of the random forest, especially when predicting phenotypes, is that it returns importance scores for each predictor in the model, allowing for the ranking of classifiers within the dataset and associations between predictors and phenotypes to be made. Classifiers in the top performing model were ordered by their Gini coefficient, a measure of how well the classifier contributed to accurately separating the classes. I found that most of the predictors within the top performing model (high CADD exon variants) were nonsynonymous mutations within genes that play a role in DNA repair and maintenance of genomic integrity.

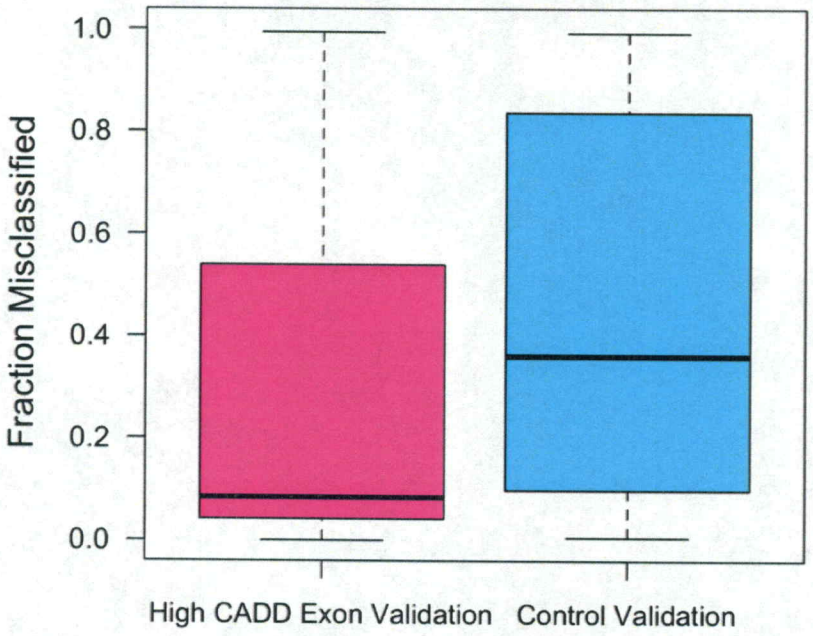


Figure 2.8 Boxplot of the misclassification error in the validation (Wellderly) cohort for the high CADD (>15) exon predictive model (median misclassification = 0.02) compared to that of the control, or randomly shuffled discovery data set, (median misclassification = 0.36).

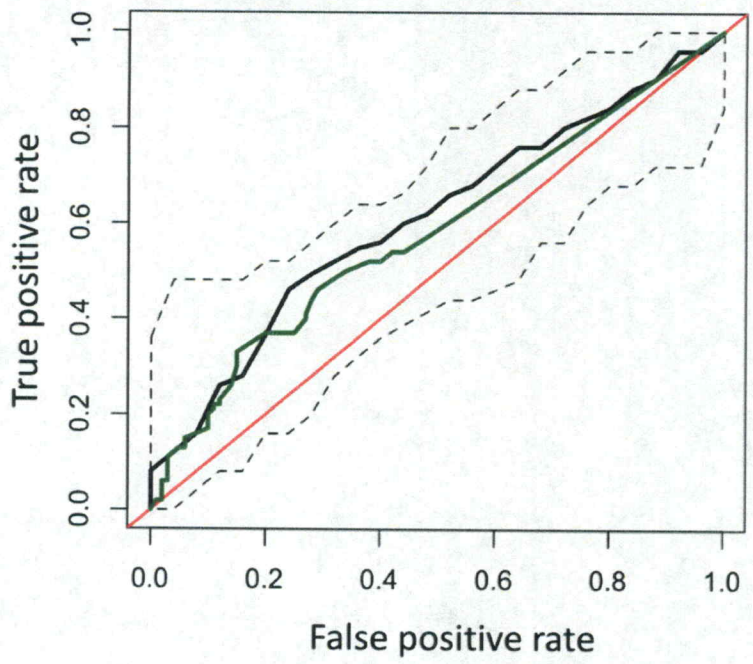


Figure 2.9 ROC curve for the model resulting in the median AUC (0.62, black line) within the best performing data set (high CADD exon variants) with confidence intervals representing the best (0.79) and worst (0.54) AUC (dash lines) for the data set compared to the ROC curve for the AUC (0.59) of smoking years as a predictor of aging status (green line).

2.3.4 Enrichment Analysis

Top variants were determined by averaging the Gini coefficients across the 16 models performed on the highly deleterious target gene exon data set. Enrichment analysis was then conducted on these variants in regard to gene and variant effect. I found that a majority of the top variants were located within and Excision Repair Cross Complementation Group 4 (*ERCC4),* Excision Repair Cross Complementation Group 5 *ERCC5, LMNA* and Pol(ADP-Ribose) Polymerase 1 *(PARP1)* (Figure 2.10 & Table A.5). Furthermore, six of the predictor's regions have previously been associated with over

fifteen different phenotypes in the GRASP database (Table A.6). Enrichment analysis of variant consequence effect revealed that predictors are enriched for those that cause a nonsynonymous change as well as a stop gain, or premature termination codon ($p<0.001$) and depleted for synonymous mutations (Figure 2.11).

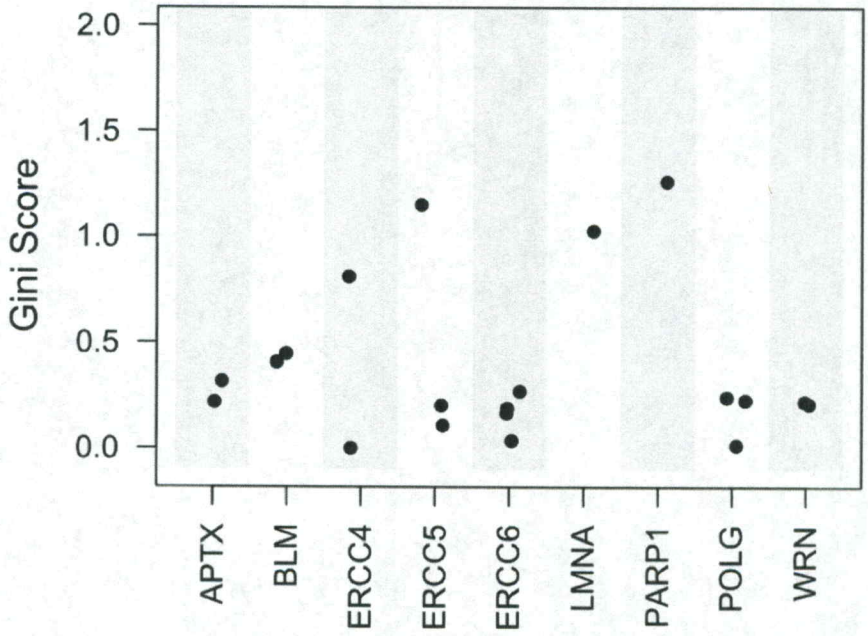


Figure 2.10 Scatter Plot of the mean Gini Scores for each variant by gene from the 16 replications of the best predictive model (High CADD Exon Variants).

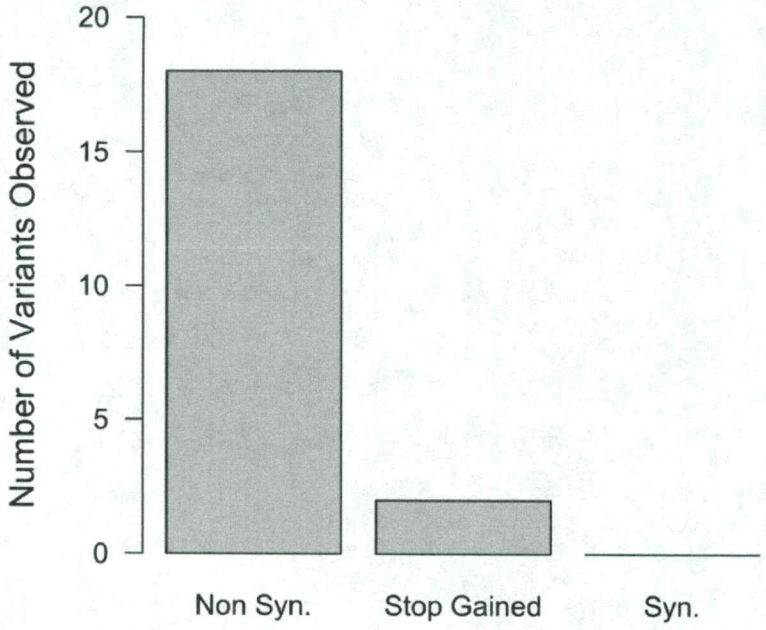


Figure 2.11 Bar plot of the variant consequences for the predictors used in the best predictive model. P-values represent the null expectation.

2.4 Discussion

Although aging is highly dependent on environmental, behavioral and social interactions, studies have shown that a quarter of the variance explaining aging is heritable (Herskind et al. 1996). Yet, only a handful of genetic determinants explaining a small portion of the heritability have been discovered thus far. This lack of discovery is due in part by the complexity of the disease but also because of the rarity of the longevity phenotype. Analysis of late aging rather than longevity allows for larger cohort sizes as late agers are more common in the general population than long-lived individuals (>100 years old), yet the lack of a clear definition for "healthy" or late aging makes genetic analysis and cross-study interpretation of this phenotype extremely difficult. Recently

Reed et al. defined "healthy" aging as living to the age of 70 in the absence of coronary surgery, heart attack, stroke, diabetes, or prostate cancer, and they found an approximate 50% heritability of the defined phenotype in a cohort of male twins (Reed et al. 2004). Several other late aging cohorts exist which are characterized by various definitions and have resulted in inconsistent heritability percentages and gene association results (Erikson et al. 2016; Brooks-Wilson 2013; Walter et al. 2011). Furthermore, large-scale aging GWA studies to date have failed to identify recurrent specific genomic regions that statistically associate with the longevity or late aging phenotypes. Albeit, combined analysis of SNPs have identified pathways and multi-allele signatures associated with aging phenotypes indicating that these studies should include polygenic or epistatic associations in addition to the more traditional analysis of single gene associations to more successfully discover genetic determinants of aging phenotypes (Brooks-Wilson 2013). This observation led us to design a unique approach for determining genetic predictors of late aging by conducting targeted sequencing of twenty previously determined aging-related genes in a cohort of "early" and "late" agers. This approach allowed for the identification of a set of genetic variants associated with various aspects of genomic integrity as possible predictors of late aging. While I point out that the size of the discovery cohort (n = 200) is not ideal for a genomic association study, this process of combining targeted sequencing and machine learning to identify a set of genetic factors that together act as predictive determinants for a complex disease will be useful in further genetic association studies of complex phenotypes for which individual variant association is insufficient.

My initial analysis of overall variant burden and individual variant association with early versus late aging failed to produce any variant with statistically significant association. While this is typical of GWA studies, especially those with either a complex phenotype or small sample sizes, single variant association does prove useful for prioritizing variants by p-value. In my analysis, two intronic variants within *LMNA* had the strongest association (rs915180 and rs915179) and were also the most predictive variants in the unfiltered data set random forest models (S8 Fig). Furthermore, these variants have also been previously associated with longevity (S3 Table). In fact, rs915179 is part of a haplotype within *LMNA* specifically associated with longevity (Conneely et al. 2012). Sebastiani et al. used rs915179 as part of a "genetic signature" of exceptional longevity and later found that this variant held up in a meta-analysis of longevity ($p = 0.0001$) (Conneely et al. 2012). *LMNA* encodes Lamin A and C, which are nuclear envelope proteins. These proteins are associated with Hutchinson-Gilford progeria syndrome (HGPS), an extremely rare disease causing premature aging leading to a life expectancy of about 13 years (Conneely et al. 2012). More interestingly, defective forms of *LMNA* is produced in small amounts within cells of healthy individuals, and there is evidence that this amount increases with age (Rodriguez et al. 2009). This variant was also one of the first to be associated with Alzheimer's disease in GWA studies indicating that it may play a pivotal role in cognitive function which is known to decline with increasing age. Lastly, rs915180 has been associated with suicide attempts in patients with mood disorders, cardiomyopathy, chronic kidney disease and birth weight in GWA studies (Perlis, Huang, and Purcell 2010; Köttgen et al. 2010; Horikoshi et al. 2013). Since this association failed to reach genome-wide significance, future

studies involving larger cohorts are needed further assess the association of rs915179 with late aging.

Because the individual variant association proved inadequate for determining variants within my data predictive of aging status I next focused my analysis on machine learning. Random forest and SVM were performed on various stratifications of the data, and assessment of the resulting ROC-AUC and misclassification percentages revealed that the random forest model built using variants with a CADD score over 15 (high CADD) proved to be the best performing predictor of aging status. As previously mentioned, one of the highlights of using random forest is that it ranks predictors based on how well they add to the purity of the model (Gini coefficient). The mean coefficient for each predictor in all trials of the high CADD exon variants was used as a metric with which to rank variants (Figure 2.10). The variant with the highest predictive power (rs1136410) in my top performing model of aging status is located in *PARP1* and causes an A>G alteration in the 17th exon (mean Gini = 1.26). *PARP1* is responsible for posttranslational modification of nuclear proteins in response to various types of DNA damage as well as oxidative stress (Muiras et al. 1998; Beneke and Bürkle 2007). With an essential role in base excision repair (BER) and double strand break (DSB) repair, *PARP1* has been known as the "sensor of nicks" within DNA (Czarny et al. 2017; Mao et al. 2011). Interestingly, comparative studies among 13 mammalian species found that the enzymatic activity of *PARP1* positively correlates with maximum lifespan in various mammals, including humans (Bükle, Grube, and Küpper 1992; Muiras et al. 1998; Piskunova et al. 2008; Noren Hooten et al. 2012). Additionally, this variant has previously been associated with survival in patients with early stage non-small-cell lung

cancer, depression, and baseline hippocampal volume loss in apolipoprotein E genotype ε4 (*APOE4)* patients (Nho et al. 2013).

The next strongest predictor in the top performing model is located within *ERCC5/XPG* (mean Gini = 1.15), located on chromosome 13q22-33 which causes a G>C (His1104Asp) change in the last (15th) exon of the gene (rs17655) (J. Zhao et al. 2018). *ERCC5* is an excision repair gene that is responsible for forming the 3' incision during Nucleotide Excision Repair (NER) and known to be extremely polymorphic (J. Zhao et al. 2018). The variant is located within the C-terminal of the gene and inhibits interactions of *ERCC5* with other DNA repair proteins (B. N. Xu et al. 2016). Damaging variants in this gene can lead to deficiencies in the NER pathway causing both xeroderma pigmentosum (XP) and Cockayne syndrome (CS), both of which result in symptoms shared with phenotypic aging (O'Donovan et al. 1994; Barnhoorn et al. 2014). Additionally, this specific variant, rs17655, is well-studied for its association with cancer risk, especially for gastric and colon cancer (J. Zhao et al. 2018). The well-established relationship between accelerated aging and deficient DNA damage repair (Gensler and Bernstein 1981) in addition to the high importance this variant has in my top performing model leads to the hypothesis that *ERCC5* is important for attenuating the aging process.

Next in importance within the predictors, is a variant within *LMNA* (rs513043) which causes a missense mutation (G>A) in the 2nd codon and has a CADD score of 18.44 indicating a high degree of deleteriousness (mean Gini = 1.03). *LMNA* encodes nuclear lamin proteins Lamin A and C for which mutations in this gene are associated with numerous diseases including cardiomyopathies, lipodystrophy, muscular dystrophies and progeroid (early aging) syndromes, such as HGPS. Again, the nuclear lamina has

been repeatedly linked to aging; in fact, Sebastiani et al. used numerous *LMNA* variants to build a "genetic signature" of longevity (Sebastiani et al. 2012).

Lastly, a variant in *ERCC4* (rs1800067) was also one of the top predictors in the best predictive model (mean Gini = 0.81). This variant causes a missense mutation (G>A) in the 8th exon, has a CADD score of 36 indicating a very high degree of deleteriousness within the gene, and has been associated with HDL cholesterol, and risk of glioma and lung cancer. *ERCC4* is an excision repair gene that forms a heterodimer with Excision Repair Cross Complementation Group 1 (*ERCC1*) for nucleotide excision repair (NER). Reduced expression of *ERCC4-ERCC1* leads to XPF-ERCC1 (XFE) progeria in humans that is characterized by systemic accelerated aging (Niedernhofer et al. 2006). Moreover, other studies examining genes under positive selection in the longest-lived mammalian species, the bowhead whale, identified *ERCC1* as a top hit, suggesting that this pathway may promote maintenance of health (Keane et al. 2015). Jorgensen et al. showed that this variant is significantly associated with benign breast disease (BDD), especially in patients with a family history of breast cancer (Jorgensen et al. 2009).

Like many genomic studies of longevity and late aging, several limitations of this study warrant comment (Martin, Bergman, and Barzilai 2007). First, in the absence of field-wide consensus regarding the definition of early versus late aging, I relied on physical function to differentiate the two groups. The parameters used to differentiate them- the ability to walk 15 minutes without stopping and to climb a flight of stairs- are well-validated (Abellan Van Kan C, Geriatric, and Panel 2008; Perera et al. 2014, 2016) and can be viewed as integrative, i.e., incorporating the impact of both physiological

decline and diseases. The advantage of using such standardized assessments of function is the ability to differentiate participants into non-overlapping groups. The disadvantage is that impaired function may reflect the effect of not only early aging but also comorbidity. However, because aging is characterized by both constriction of physiological reserve and the accumulation of diseases, it is difficult to disentangle the impact of early aging and disease. It is possible that subtle effects of genes or alleles on aging were masked by the impact of superimposed diseases but testing this hypothesis will require a study large enough to identify a sufficient number of participants who qualify as early agers in the absence of disease. It is also possible that conditions such as comorbidity, obesity, and frailty lie in the causal pathway from any genetic predispositions to functional outcomes. Therefore, efforts to control for them would attenuate any associations between genetics and the function-based group definition. Another limitation of this study, which is common amongst many genomic studies, is the cross-sectional design; future studies are needed to examine longitudinal trajectories. Furthermore, while the age cut-off of 75 years has been utilized for studies in older adults with fractures (Boonen et al. 2010; McClung et al. 2012; Boonen et al. 2006), an older age cut-off may alter the findings. Lastly, a validation cohort consisting of both early and late agers would improve my confidence in the constructiveness of this model for both early and late aging phenotypes.

2.5 Conclusion

Overall, this study found that more complex statistical analyses encompassing epistatic effects rather than traditional single gene association tests are useful for

interpretation of sparse or rare data. Random forest provided information complementary to more traditional statistical analyses, including the ability to correctly classify the validation cohort of "late" agers 90% of the time. The top predictors in the model were within genes involved in DNA repair and stability, reiterating previous accounts that the integrity of the genome is essential for "healthy" or late aging. I recognize that there are many genes and possibly intergenic regions of the genome engaged with genome stability and the biology of aging which were not included in this study, however the genes chosen for analysis here are those with which the authors have had the greatest familiarity and sequence knowledge. Additionally, I do point out that I did not account for admixture in my analysis however I believe this would not drastically alter my results as most of my discovery cohort and the entire validation cohort used were of self-reported EA descent. While I realize that the training set has a low number of patients to achieve statistical certainty, I propose that holistic analysis of rare variant data may have promise in a larger cohort. Thus, targeted sequencing of genes involved in aging in combination with machine learning should be considered as a method to determine predictors of complex phenotypes.

CHAPTER 3

EPIGENTICS REGULATION OF IMMUNE CELLS AND AUTOIMMUNITY

3.1 Epigenetic Mechanisms of Immune Cell Development

Hematopoietic cell development is a highly orchestrated process controlled by lineage-determining TFs and epigenetic changes that ultimately lead to the maturation of all cells (Oakes et al. 2016; Waddington 1956). Research involving immune, or lymphoid, cell development has focused on determining the cell of origin of various hematological diseases, such as myeloid malignancies, myeloproliferative neoplasms, leukemia and lymphoma. Lymphopoiesis, or the development of lymphoid cells, begins in the bone marrow where hierarchical epigenetic processes stimulate hematopoietic stems cell (HSC) differentiation into mature blood cells (Figure 3.1) (R. Li et al. 2018; Waddington 1956; Shapiro-Shelef and Calame 2005). External stimuli instigate alterations in the epigenetic landscape and activation of TFs, such as Soi-1 Proto-Oncogene (*PU.1),* Transcription Factor E2-Alpha *(E2A),* Early B Cell Factor 1 *(EBF1),* Interferon Regulatory Factor 4 (*IRF4),* Myocyte Enhancer Factor 2C *(MEF2C)* and B Cell-Specific Activator Protein *(PAX5),* that encourage HSCs to take specific cellular developmental pathways ultimately leading to maturation of B and T lymphocytes (Pancaldi et al. 2015).

In addition to transcriptional and epigenetic changes directing B and T cell development, lymphocyte-primed cells must also undergo multiple checkpoints to ensure adequate immune cell development. In healthy individuals, checkpoints throughout immune cell development control for the production of autoreactive immune cells. In patients with autoimmune diseases, immune cells escape these check points and go on to produce autoantibodies against self-tissue causing systemic inflammatory responses (Z. Jin et al. 2017; K. Chen, Liu, and Cao 2017; Q.-Z. Li et al. 2009; Baechler et al. 2003; Han et al. 2003; Garaud et al. 2011; Rhead et al. 2017; L. Wang et al. 2015; Absher et al. 2013; Toro-Domínguez, Carmona-Sáez, and Alarcón-Riquelme 2014). Determining the stage in cell differentiation, or cell of origin, in which aberrant epigenetic patterns as well as when autoreactivity first occurs would lead to a better understanding of the etiology of autoimmune diseases.

As depicted in Figure 3.1, differentiation from a long term- (LT) HSC to short term- (ST) HSCs involves the guidance of hematopoietic-specific TFs, such as Runt Related Transcription Factor 1 (*RUNX1),* Lysine Methyltransferase 2A *(MLL),* and LIM Domain Only 2 (*LMO2)* (Orkin and Zon 2008). The ST-HSC then decides between lymphoid-primed multipotent progenitor (LMPP) or a common myeloid progenitor (CMP) commitment, a decision highly guided by transcriptional and epigenetic mechanisms (Barneda-Zahonero et al. 2012; Pancaldi et al. 2015). Various lineage-determining TFs, such as *E2A, PU.1,* and *E47,* govern the decision between LMPP and CMP commitment (Laurenti and Göttgens 2018; Yang et al. 2008). *PU.1* is required for LMPP to CLP differentiation as it primes expression of lymphoid genes while suppressing neutrophil genes (Pang et al. 2018). *E2A*, a helix-loop-helix TF of the E-protein family, induces

PU.1 binding, which in turn alters chromatin accessibility to induce expression of lymphoid genes (Semerad et al. 2009; Heinz et al. 2010). *E47*, an isoform of *E2A*, is essential for lymphoid differentiation and V(D)J recombinase activity (Yang et al. 2008; Santos et al. 2011).

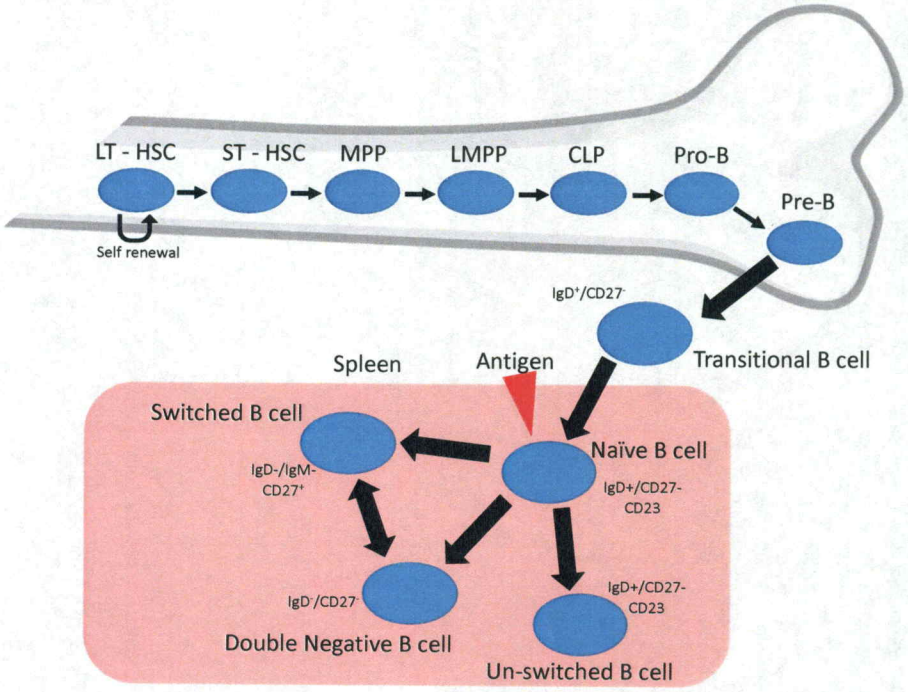


Figure 3.1 Development of mature B cell subsets from hematopoietic stem cells (HSCs).

Once the LMPP commitment is made, a cell can then differentiate into a common lymphoid progenitor (CLP), which is a precursor to B or T lymphocyte cell fate (Barneda-Zahonero et al. 2012). If the B cell fate is chosen, a CLP will first differentiate into a pro-B cell during which stage rearrangement of heavy chain (IgH) gene segments occurs (Jung et al. 2006). Both *EBF1* and *E2A* play a role in this stage of development. *E2A* regulates immunoglobulin (Ig) transcription and is essential for V(D)J

rearrangements within the pre-B cell receptor (BCR), which is required for the first B cell checkpoint. *E2A* and *EBF1* are epigenetic regulators that work in concert, regulating the expression of each other throughout B cell development (Hagman, Ramirez, and Lukin 2012). They also play an important role in pre-BCR rearrangements (Vilagos et al. 2012; Y. C. Lin et al. 2010). Pro-B cells that express an adequate pre-BCR, following V(D)J rearrangement become pre-B cells, while those that do not undergo pre-BCR editing or anergy (Melchers and Melchers 2015). Additionally, deletion of *EBF1* in the pro- to pre-B cell transition results in the transition to the T cell and depletion of B cells in mice (Nechanitzky et al. 2013). Only about 20% of pro-B cells pass this checkpoint in B cell maturation, indicating the stringency of immune development (Rajewsky 1996). Pre-B cells that do pass this checkpoint undergo clonal expansion and rearrangement of the light chain (IgL) gene segments prior to the next checkpoint. At this stage, Pre-B cells that do not present IgM on the cell surface undergo BCR editing or anergy while those that do differentiate into transitional B cells and exit the bone marrow (Noviski et al. 2018; Shlomchik 2008).

Transitional B cells develop into naïve B cells capable of recognizing antigen within the spleen. At this point in differentiation, B cells are morphologically similar to T cells (Alberts et al. 2002). It is only after antigen stimulation that epigenetic and genetic changes occur so that B and T cells are distinguishable. TFs involved in this stage of development include: *PAX5,* B Cell CLL/Lymphoma 6 *(BCL6),* Nuclear Factor Kappa B Subunit 1 *(NFKB1),* MYC Proto-Oncogene *(MYC), IRF4* and Interferon Regulatory Factor 8 (*IRF8)* (Recaldin and Fear 2016). Mouse studies show that *PAX5*, is necessary

for mature B cell development, as deletion decreases expression of mature B cell genes and mature B cell antigens (Horcher, Souabni, and Busslinger 2001).

Naïve B cells go on to form memory B cells with varied immune responses controlled by interaction with cytokines, T cells, and antigens. During this transition from naïve to memory B cells, transcriptional and epigenetic mechanisms lead to higher proliferative rates, increased Ig secretion, and enhanced survival (Seifert and Küppers 2016; Good, Avery, and Tangye 2014). Naïve cells may proliferate and undergo hypermutation and/or class switching, which results in BCR constant region switches from IgM/IgD to IgG, IgA, or IgE (Stavnezer and Schrader 2014). Repression of genes involved in quiescence, such as Kruppel Like Factor 4 (*KLF4*), Kruppel Like Factor 9 (*KLF9*) and Promyelocytic Leukemia Zinc Finger (*PLZF*), led to increased proliferation rates of memory B cells (Good and Tangye 2007). *BCL6* upregulation in memory B cells influences T cell interaction and consequently antibody response (Kitano et al. 2011). Lastly, expression of members of the Tumor Necrosis Family Receptor B7 (TNFR), Cluster of Differentiation 80 Antigen (*CD80*), Signaling Lymphocytic Activation Molecule (*SLAM*), and B Cell CLL/Lymphoma 2 (*BCL2*) family members influence memory B cell maintenance through various mechanisms (Good, Avery, and Tangye 2014).

Altogether, these various transcriptional changes influence the memory B cell pathway (switched, un switched and double-negative) that naïve B cells can take (Alberts et al. 2002; Brezinschek et al. 2012). Switched B cells are formed following BCR switching, and undergo germinal center dependent hypermutation (Shapiro-Shelef and Calame 2005). Non-switched memory B cells do not have switched BCR regions, and undergo GC independent hypermutation. Double-negative ($IgD^{neg}CD27^{neg}$) B cells are a

unique memory B cell in that double-negative B cells can differentiate into switched memory B cells, and switched memory B cells can differentiate into double-negative B cells (Y. C. B. Wu, Kipling, and Dunn-Walters 2011).

3.2 Systemic Lupus Erythematosus: An Immune Cell Disease

SLE is a historically complex autoimmune disease which lacks a clear definition despite being first defined in 916 AD (Rekvig 2018). Multiple lines of evidence support the hypothesis that dysregulated B cells are a major contributor to SLE (Zhang et al. 2001; Chan et al. 2013; Tipton et al. 2018). A more specific and commonly cited hypothesis regarding the origin of SLE is that an environmental trigger, such as an infection or drug exposure, elicits T cells to recognize self-antigens and induce B cells to produce autoantibodies (H. Wu et al. 2017). B cells play a pivotal role in adaptive immunity through their ability to produce and present antigens, proinflammatory cytokines, and costimulatory factors to T cells. Dysregulation of B cell function, signaling, or development can lead to excessive autoantibody production and B cell hyperactivity (De and Barnes 2014). Multiple lines of evidence have demonstrated the significant role that the epigenome plays in the etiology of several autoimmune diseases (Absher et al. 2013; Coit et al. 2013; Jeffries et al. 2011). Association studies between DNA methylation and SLE reveal widespread hypomethylation of CpGs within interferon (IFN)-related genes in SLE patients relative to controls in numerous populations, but the exact nature of the relationship has not been established (S. Chen et al. 2019; Absher et al. 2013; Coit, Yalavarthi, et al. 2015; Mok et al. 2016; Yeung et al. 2017).

3.2.1 Epidemiology of SLE

SLE is known to affect non-European ethnic populations more often, and at a higher severity than European ancestral populations (Danchenko, Satia, and Anthony 2006). In addition, females are affected 9 times more often than males (Yen et al. 2017; Jarukitsopa et al. 2015). Gender, in combination with ethnicity, further increases risk for the disease. Non-European females have a 3-fold higher prevalence for SLE than European females (Danchenko, Satia, and Anthony 2006; Menard et al. 2016). African American (AA) females have the most severe SLE-associated symptoms, often including severe organ damage (Mohan and Putterman 2015a). The average age of onset is 36 years, and the average duration is about 17 years (Leuchten et al. 2018).

The annual incidence of SLE ranges from 1-10 per 100,000 people, while the prevalence of this disease is estimated to range between 5.8-130 per 100,000 people indicating the difficulty in diagnosing patients (Jarukitsopa et al. 2015). Because symptoms affect each patient differently based on their genetic background and environment, diagnosing patients is often a difficult and prolonged process. A 46-year longitudinal study focused on SLE mortality revealed 50,249 SLE-caused deaths between 1968 and 2013 and that the reduction in SLE mortality was less than that of non-SLE mortality, indicating that treatment for the disease is lagging behind modern medicine. (Yen et al. 2017).

3.2.2 Clinical manifestations of SLE

SLE affects multiple organs and systems, and the defining symptoms of SLE have changed over time. Originally, SLE was only characterized by cutaneous symptoms, but as more and more cases arose additional symptoms have been added to the diagnosing criteria. Currently 11 criteria are used for clinical diagnosis of SLE, as defined by The American College of Rheumatology (ACR) and The Systemic Lupus Collaborating Clinics (SLICC), which together are termed the SLE Disease Activity Index (SLEDAI). Patients need only present 4 of the 11 symptoms during observations by a physician at any given time, which allows for hundreds of unique SLE phenotypes amongst patients (Rekvig 2018). Therefore, clinical presentations vary drastically amongst patients. Even in single patients, symptoms can be unpredictable and vary over time from mild to severe (Mohan and Putterman 2015a). This has led to high rates of misdiagnosis in SLE patients who consequently go years before being correctly diagnosed and treated.

The most prevalent early symptoms of SLE (fatigue and joint pain) are common symptoms shared by a multitude of common diseases making the early detection and diagnosis of the disease extremely difficult (Leuchten et al. 2018). Yet, an earlier diagnosis is key in SLE, as altered innate and adaptive immune responses can have detrimental effects to organs causing severe and irreversible damage. SLE patients fluctuate between symptomatic (flare) periods and quiescent periods. During quiescent periods patients are generally asymptomatic and have lower SLEDAI scores (1-3). Flares occur when the immune system attacks multiple tissues, such as the skin, kidneys, heart and lungs (Coit et al. 2016). The most common symptoms present in patients during a flare include: skin bleeding, severe inflammation, Raynaud's syndrome, photosensitivity and leg edema (Leuchten et al. 2018). Nephritis is the most prevalent severe

manifestation of SLE, affecting anywhere between 40-70% of SLE patients and leading to mortality in over 50% of those afflicted (H. Wu et al. 2017; Mohan and Putterman 2015a; Menard et al. 2016)

One of the most common clinic assessment for diagnosing SLE is elevated serum antibody levels, including antinuclear antibodies (ANA) or anti-double stranded DNA antibodies (anti-dsDNA) (M. Zhao et al. 2016b). Yet, current clinical diagnostic tests for these autoantibodies lack specificity (ANA) or sensitivity (anti-dsDNA) leading to insufficient clinical utility (H. Wu et al. 2017). Serum complement, and C-reactive protein levels are additional biomarkers commonly used by physicians for indication disease status in patients, however they too yield low clinical utility (M. Zhao et al. 2016a). Unfortunately, even if these tests result in positive outcomes for SLE, current treatment (corticosteroids and hydroxychloroquine) do not significantly improve patients' outcomes (H. Wu et al. 2017; Merrill et al. 2010).

3.2.3 Etiology of SLE

It is widely believed that SLE is caused by a combination of genetic predisposing factors and environmental triggers that result in dysregulation of both innate and adaptive immune responses (Figure 3.2) (H. Wu et al. 2017; Absher et al. 2013). As previously mentioned, a common hypothesis regarding the disease origin is that an environmental trigger causes T cells to recognize self-antigens resulting in the production of auto antibodies by B cells. Yet, a full understanding of the pathogenesis of SLE has yet to be determined.

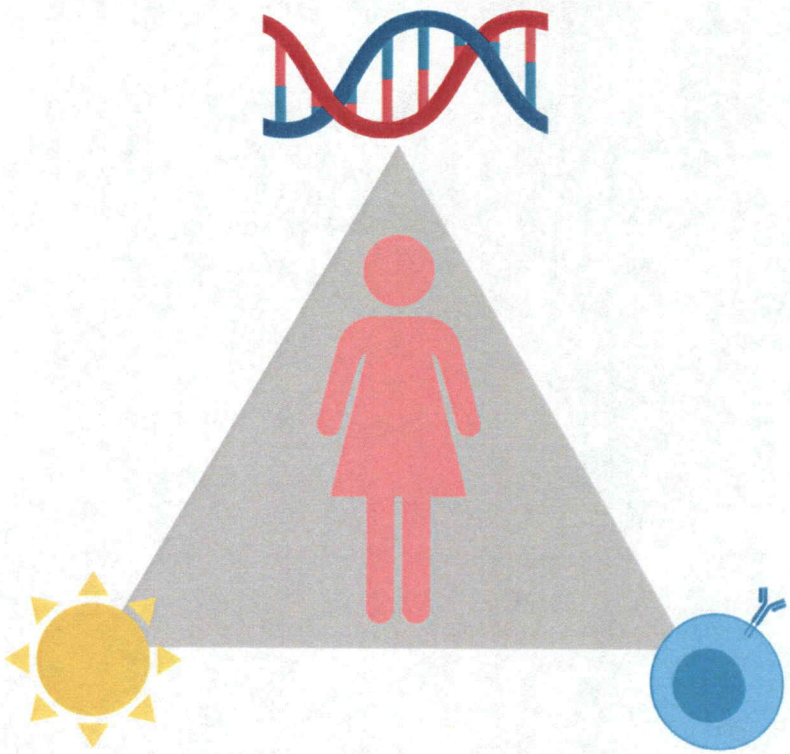


Figure 3.2. SLE is believed to be caused by genetic risk factors and environmental triggers leading to immune cell hyperactivity.

Several environmental triggers have been cited over the years for association with SLE risk, including excessive ultraviolet light exposure, silica dust, and smoking, yet the mechanisms and timeline for development of disease in response to such stimuli still remains unknown (Tsokos et al. 2016; Kamen 2014). Viral and microorganism infections, such as the Epstein-Barr virus (EBV), cytomegalovirus and herpes simplex virus type 2, have also been associated with SLE risk. The microbiome has recently gained interest as a risk factor for SLE, as it is thought that the collection of fungi, bacteria and viruses that coexist in the body may modulate the immune response (Tsokos et al. 2016). Vitamin D insufficiency is another environmental risk factor with compelling evidence as being associated with SLE, however this factor may be

confounded by patient intolerance to ultraviolet light, a main source of vitamin D (Kamen 2014). Other environmental influences on SLE with less compelling evidence include: metals, such as mercury, pesticides asbestos, industrial chemicals and solvents, air pollution, smoking and certain cosmetic products (Kamen 2014).

The identification of SLE-associated specific polymorphisms in genes involved in reactive oxidative species production leads to another hypothesis in regards to SLE etiology, which is that SLE is caused by genetic variation in genes encoding metabolic enzymes for pathways involved in SLE pathophysiology in conjunction with environmental triggers stressing those pathways (Kamen 2014). Understanding both the environmental and genetic risks, as well as the combined risks will allow for the determination of SLE etiology and pathophysiology and preventative measures that can be undertaken in populations with increased susceptibility to the disease such as AA females during childbearing years (Jarukitsopa et al. 2015).

3.2.4 Genetics of SLE

A strong genetic component for SLE is well known, but poorly understood owing to the multifactorial nature of the disease and it's disparate manifestations (Deng and Tsao 2017). Over 100 SLE risk loci encompassing over 40 genes have been identified thus far (Mohan and Putterman 2015a). Yet, the contribution of these loci to the heritability of SLE is undetermined leading to the hypothesis that epigenetic variations may explain the missing heritability of SLE. To date, the genomic region having the strongest association with SLE risk is the human leukocyte antigen (HLA) region (Costa-Reis and Sullivan 2013). However, the HLA region is the most gene-dense region of the genome,

consisting of over 100 genes, further complicating the genetics of SLE predisposition. Other susceptibility loci associated with SLE risk include genes involved in various pathways, such as Toll-like receptor and α-IFN signaling (*ACP5, ETS1, IRF5, STAT4*, etc.), NFκB signaling (*IRAK1, TNFAIP3, TNIP1*, etc.), apoptosis, and clearance of cellular debris (*TREX1, DNASE1,* etc.) (Costa-Reis and Sullivan 2013). Overall, most of the genetic variants identified to infer risk to SLE are common variants with small effect size (Costa-Reis and Sullivan 2013). Incomplete concordance in identical twins and sporadic, rather than familial, cases strengthen the hypothesis that other factors play a large part in disease risk (H. Wu et al. 2017).

In addition to genetic variation, vast epigenetic alterations in SLE patients compared to healthy individuals exist increasing evidence that epigenetic variation might explain the missing heritability of SLE (Absher et al. 2013; Jeffries et al. 2011; Coit, Yalavarthi, et al. 2015; S. Chen et al. 2019). Many factors influence methylation of DNA, including nutrition and diet, stress, age, and life experiences, allowing methylation-based studies to identify combined genetic and environmental risks for disease. Despite repeated observations that epigenetic variation plays a significant role in the etiology of several autoimmune diseases, the exact relationship has not yet been determined between the aberrant IFN epigenetic signature characteristic of SLE patients and disease (Absher et al. 2013).

Expression of IFN-associated genes is strongly correlated with SLE disease activity, as measured by SLEDAI score. Yet, epigenetic modifications remain stagnate between flare and quiescent periods, raising the question of exactly how epigenetic modifications affect expression of SLE-associated genes (Landolt-Marticorena et al. 2009; Feng et al.

2006). This is potentially explained by the nature of the methylation studies, which have generally been performed using whole blood or mixed immune cell populations. As previously mentioned, methylation analysis of mixed cell populations is well known to have confounding results as the epigenetic signature unique to individual cell types varies tremendously (Jaffe and Irizarry 2014). Furthermore, recent epigenetic research has shown that variation in epigenetic regulation is highly correlated with ethnicity (Wiley et al. 2013). Langefeld et al. performed one of the largest ethnicity-specific SLE genomic studies thus far revealing ethnicity-specific risk loci, concluding that ethnicities must be studied independently to fully understand the genetic mechanisms of SLE (Langefeld et al. 2017). A unique AA immune signature was recently established in CD4+ T cells compared to European American (EA) CD4+ T cells within healthy subjects (Coit, Ognenovski, et al. 2015). Cis-acting genetic variants were found to be correlated with this epigenetic variation within naïve T-cells in healthy AA patients. Therefore, exploring both genetic and epigenetic associations with disease in sorted cell types from single ethnic populations would allow for more precise interrogation of disease, especially in the case of SLE which predominately affects a specific ethnic population.

3.3 Summary

Development of lymphocytes from hematopoietic stem cells is tightly regulated by both genetic and epigenetic processes. Multiple checkpoints exist throughout this process to prevent the development of immune cells that produce autoantigens against self-tissues, or autoreactive immune cells. Patients with SLE are burdened by the production of autoreactive immune cells resulting in systemic inflammation which can

lead to organ failure. Several studies have identified epigenetic defects in SLE patients, and particularly in the immune cell lineage. These observations have led to the generalization that SLE is an immune cell-driven disease.

CHAPTER 4

EPIGENETIC DEFECTS IN THE B CELL LINEAGE OF SYSTEMIC LUPUS ERYTHEMATOSUS PATIENTS DISPLAY POPULATION-SPECIFIC PATTERNS

4. 1 Introduction

SLE is a complex autoimmune disease characterized as a dysregulated immune system causing chronic inflammation for which underlying mechanisms of pathogenesis and effective treatments remain elusive (Rekvig 2018; Morawski and Bolland 2017). AA females are the most at-risk population for developing SLE, and they also have the most severe SLE-associated symptoms often including acute and irreversible organ damage (Yen et al. 2017; Bentham et al. 2015; CDC 2018; Guillermo J. Pons-Estel, MD*, Graciela S. Alarcón, MD, MPH*, Lacie Scofield, MSPH†,‡ and Reinlib, PhD‡, and Glinda S. Cooper 2010; Somers, EC Marder, W Cagnoli, P Lewis and Deguire, P Gordon 2014; Lewis and Jawad 2017). Most SLE genomic studies to date have focused on European and Asian ethnicities leaving out the most at-risk population, thus I focus on AA females in this study (Iwamoto et al. 2018).

This study was designed to analyze whole genome DNA methylation data from B cell subsets in a cohort of both SLE and control females to identify where in B cell development aberrant epigenetic patterns arise and to identify multivariate epigenetic signatures of SLE and its severity. To accomplish these tasks, B cell subsets were sorted

from AA females with SLE (n=24) and without SLE (n=25) as well as EA females with SLE (n = 18) and without SLE (n = 13). B cell subsets analyzed include transitional, naive, un-switched, switched and IgD-/CD27- double negative B cells. I used regression analysis to test single CpG associations within each B cell type with SLE in addition to ethnicity-specific SLE-associated CpGs across B cell types. The strongest associated single CpGs were used in machine-learning approaches to build a multivariate model of SLE. The models were tested across ethnicity groups in an independent cohort consisting of both CD19+ pan-B cell and CD4+ pan-T cell samples to test whether this signature could predict SLE status in broader cell populations.

My results demonstrate that epigenetic defects in female AA SLE patients are already present in immature B cells emerging from bone marrow (transitional B cells), while epigenetic defects appear to develop later during B cell development in EA female SLE patients. Furthermore, I observed that AA-specific CpG sites associated with SLE are enriched for IFN-regulated genes and near EBF1 regulatory sites, and that AA-specific SLE CpGs are predictive of SLE status in mixed immune cell populations from AA females.

4.2 Methods

4.2.1 Discovery Set Participants

The discovery cohort was obtained through recruitment to the Rheumatology outpatient clinic at the University of Alabama at Birmingham. Patient samples were collected with consent and in compliance with the University of Alabama at Birmingham Institutional Review Board. A total of 80 patient samples were used in which 49 were of

self-reported AA ancestry and 31 were of self-reported EA ancestry (Table B.1). Age, gender, self-reported ethnicity, and smoking status were recorded for each patient regardless of SLE status. Additional clinical data was also obtained from SLE patients, which included: Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) score, as determined using the ACR guidelines, flare status, BMI, nephritis presence and stage, creatinine level, proteinuria, glucose level, and prednisone usage and dose (Table B.2).

4.2.2 Validation Set Participants

The validation cohort was obtained from the outpatient Rheumatology clinic at the University of Alabama in Birmingham and detailed in Absher et al. 2013 (Absher et al. 2013). Overlapping subjects from the discovery cohort were removed leaving 43 CD19+ pan-B cell samples and 45 CD4+ pan-T cell samples from AA female patients and controls, as well as 24 CD19+ pan-B cell and 23 CD4+ pan-T cell samples from female EA patients and controls (Table B.3).

4.2.3 Cell and DNA Isolation

Individual B cell subsets from the discovery cohort were sorted using flow cytometry. Pure subsets of 5 B cell lineages were separated using antibodies particular to lineage-specific cell surface receptors. Overall, the minimum number of samples extracted for a specific cell type was 42 (IgD-/CD27- double negative) and the maximum was 72 (Naïve) for all discovery samples (Figure B.1). Cells were then lysed and DNA was extracted using Qiagen DNAeasy kits according to the manufacturers protocol.

4.2.4 Methylation450 Assays, Quality Control and Batch Normalization

The Infinium HumanMethylation450 array (Illumina, San Diego, USA) was used for measuring DNA methylation levels. 500 ng of each sample was bisulfite converted (Zymo EZ DNA), amplified, hybridized and imaged. Raw data intensity files were processed using GenomeStudio. CpG probes with detection P-values over 0.01 and those affected by common variants were filtered out. ComBat was then used to correct for batch effects in which a single array (12 samples) was used to describe a batch. Infinium chemistry corrections were then made to correct for differences between probe types. Genome-wide data methylation analysis was then performed using the R statistical suite (version 3.5.1).

4.2.5 Regression Analysis

Multivariate linear regression was performed to identify CpG methylation changes associated with SLE status both independent and dependent of self-reported ethnicity. In detail, a simpler model was performed first for the interrogation of single CpGs associated with SLE status incorporating age and smoking status as covariates (Equation 1). A second, more complex model included age, gender, self-reported ethnicity and smoking status as covariates for the discovery of CpGs associated with the interaction of SLE status and ethnicity (Equation 2). Both models were performed on AA and EA female samples across all 5 B cell lineages resulting in a total of ten regression models in this stage of analysis.

Equation 1:

$$\beta \sim Age + Smoking + Ethnicity + SLE$$

Equation 2:

$$\beta \sim Age + Smoking + Ethnicity + SLE + Ethnicity * SLE$$

4.2.6 Predictive modeling

Random forest regression was performed on the AA female transitional B cell samples (n = 38), since that was the first B cell stage for which aberrant methylation associated with SLE status was detected. I tested various p-value cut-offs ($5x10^{-2}$-$5x10^{-6}$) resulting from multivariate linear regression as well as various variance cut-offs (0.02-0.1) with m-tries ranging from 1 to 20 and 1000 trees using the RandomForest package in R (Breiman 2001). For each iteration, a random sample of 25 was used for the training set, and the remaining samples were used as the test set. Using the same approach, multivariate linear regression using L1 penalized log partial likelihood (LASSO) was performed via the glmnet package in R (Friedman, Hastie, and Tibshirani 2010). Three-fold cross validation using an alpha parameter of 1 was performed to find the optimal lambda value. The lambda plus one standard deviation, and an alpha of 1 was then used to build the model. Ridge regression using AA female samples was performed similarly to LASSO, with the only difference being that the alpha parameter of 0 instead of 1 was used.

An optimal prediction cut-off for separation of SLE and control was determined for the model based on the training and used to calculate the misclassification percentage in the test set. Model prediction accuracy was then assessed in individual B cell lineages in EA and AA patient samples as well as CD19+ pan-B and CD4+ pan-T independent sample sets from EA and AA patients. ROC-AUC was used to assess the model's

sensitivity and specificity in each of the test sets described in addition to misclassification percentage as defined by the training set optimal cut-off. Top performing classifiers from each random forest model were determined based on their Gini coefficient, a function unique to the random forest algorithm which assigns the coefficient based on ability to accurately separate cases from controls during the model training phase so that the higher the coefficient, the better the CpG is at separating SLE cases from controls in the model.

Random forest, LASSO and Ridge regression were also performed on all sorted EA B cell samples combined, using 1/3 as the training set (n= 24) and the rest as a test set using the same p-value cut offs and m-tries tested in the AA modeling. Modeling was performed using the switched B cell stage since B cell stage for which substantial differences in methylation associated with SLE status was detected in EA samples. The top performing model was then determined using the same approach as was used for AA analysis.

4.2.7 Transcription Factor Enrichment

The GM12878 chromatin immunoprecipitation sequencing (ChIP-seq) from the ENCODE data portal was used for genome-wide transcription factor binding site (TFBS) interrogation (Davis et al. 2018). Using BEDtools, the intersection between the Methyl450 annotation file and the ChIP-seq bed file was performed (Quinlan and Hall 2010). Permutation testing then allowed for the determination of whether specific TFs binding within 500 bps of the top CpGs were enriched compared to all sites assayed. Orthogonal analysis was conducted on top TFBSs enriched using ChiP-seq data. Position frequency matrices (PFMs) for the most enriched TFs were obtained from JASPAR, and

the FIMO package within the MEME suite was used for analysis of motifs within 500 bps of CpGs of interest (Bailey et al. 2009; Grant, Bailey, and Noble 2011; Khan et al. 2018). Permutation testing allowed us to determine whether top CpGs were enriched for a specific binding motif compared to the other sites assayed.

4. 3 Results

4.3.1 CpG methylation is altered in immature B cells from SLE patients

To identify loci where DNA methylation levels were associated with SLE, I performed epigenome-wide regression analysis in sorted B cells for which I found statistically significant methylation differences in SLE patients compared to disease-free controls (Figures B.2.1-B.2.5 and Table B.4). Across all 5 B cell subsets, 60 CpGs reached genome-wide significance ($p \leq 1.07 \times 10^{-7}$) in which 13 occurred in transitional B cells, the earliest B cell stage assayed, indicating that epigenetic defects in SLE patients are already present as immature B cells emerge from bone marrow. The most significant CpGs were found in multiple genes, particularly near IFN-regulated genes (Figure 4.1). Top SLE-associated CpGs from transitional B cell regression displayed a distinct hypomethylation pattern within AA SLE patients compared to controls that is less apparent in EA patients and controls (Figure 4.2).

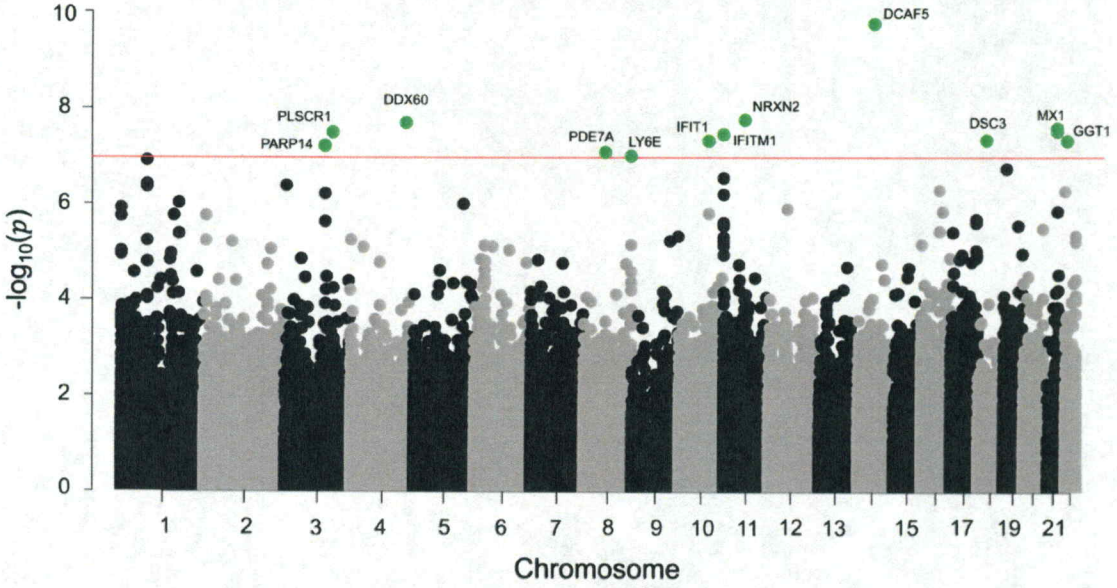


Figure 4.1 Multiple CpGs across the genome are associated with SLE in transitional B cells. Manhattan plot of transitional B cell regression SLE p-values ($-\log_{10}$).

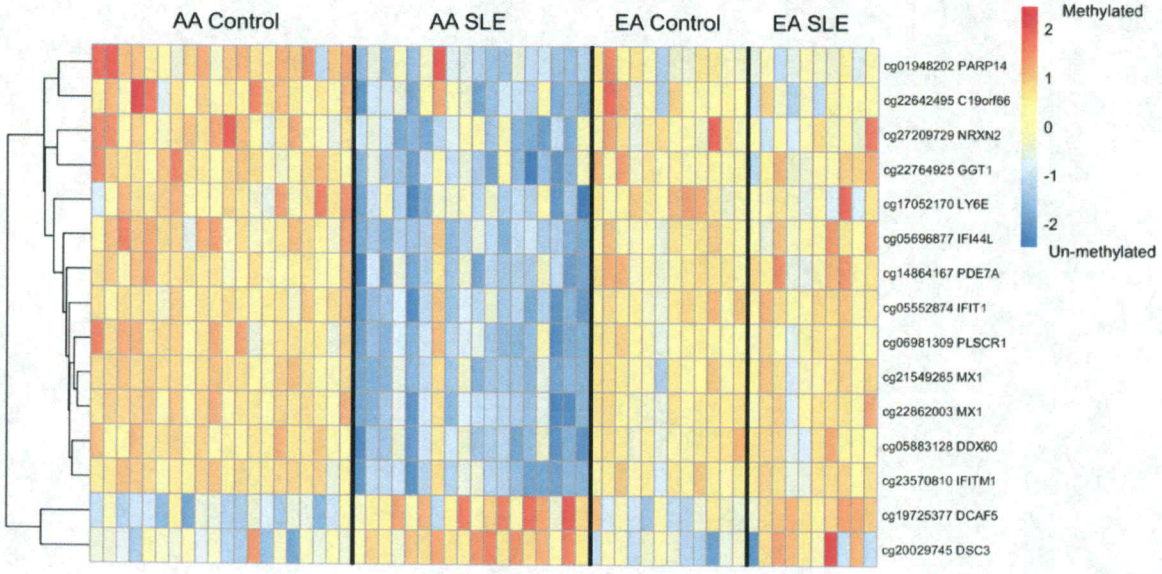


Figure 4.2 Hypomethylation of IFN-regulated genes is more severe in AA SLE patients compared to healthy AA patients. Heatmap of the methylation status across the top 15 CpGs associated with SLE status (based on p-value) from transitional B cell regression. Individual patients are represented by columns and grouped by ethnicity and disease

status. All subsequent rows represent top CpGs and are colored based on methylation. Hierarchical clustering of CpGs resulted in grouping representative of the methylation change in SLE patients versus controls.

The epigenetic defects observed indicated that those individuals carrying severe defects tended to carry severe defects across all of the significant CpGs, and that those with mild effects tended to carry mild effects across all loci suggesting that the epigenetic pattern behaves as a correlated module, rather than a set of independent, heterogeneous effects at each CpG. Further interrogation revealed that methylation specifically within the top 15 CpGs (based on transitional B cell regression) was highly correlated in AA SLE patients and in controls, but there was a strong anti-correlation between both groups indicating strong differences in the methylation at these sites correlated with disease (Figure B.3), a trend which was not seen in EA patients (Figure B.4). To further examine the population-specific nature of these epigenetic defects, I ran additional regression models using an interaction term between SLE and ethnicity revealing numerous sites across all 5 B cell subsets that approached and/or reached genome-wide significance (Figures B.5.1-B.5.5). Most of those top CpGs resulted from early B cell (transitional and naïve) analyses and were near IFN-regulated genes. The highest AA-dependent SLE association (cg17980508, $p = 1.12x10^{-9}$) resulted from transitional B cell regression and is located near Interferon Induced Protein 44 Like (*IFI44L*) (Table B.5). The highest EA-dependent SLE-association (cg13710613, $p = 8.13x10^{-8}$) resulted from naïve B cell regression and is located near Euchromatic Histone Lysine Methyltransferase 1 (*EHMT1*) (Table B.6).

Analysis of methylation at top individual AA- and EA-specific sites (cg17980508, cg21549285, and cg13710613) across B cell development in both AA and EA patients

and controls revealed that methylation strongly distinguishes AA but not EA disease patients from controls (Supplementary Figures 6-8). The difference in percent methylation within AA patients and controls is broader earlier during B cell development and decreases in later B cell stages further suggesting that aberrant hypomethylation occurs at the earliest stages of B cell development in SLE patients of AA ancestry.

4.3.2 DNA methylation patterns in transitional B cells can predict SLE status in AA females

The modular nature of the epigenetic defects I observed in immature B cells from SLE patients suggested that a multivariate predictive model could be used to distinguish patients from controls, and that a machine learning optimization approach would identify the strongest, and potentially most informative genes that carry these aberrant epigenetic states. Multiple machine learning approaches were tested using methylation status in transitional B cells to predict SLE in AA patient samples. Average ROC-AUCs across the discovery and validation cohorts for AA females was highest for the top performing ridge regression model (p-value cutoff = $1x10^{-4}$, average ROC-AUC = 0.97), yet the mean standard error (MSE) was lowest in the random forest regression model (0.02) (Table B.7). When considering performance across individual B cell subsets, the random forest regression model performed well in both early and later AA B cell stages (misclassification errors of naïve = 2%, un-switched = 3%, switched = 3%, and double negative = 6%). Applying the same models to EA B cell stages produced poor predictive success, indicating that the AA SLE signature is distinct from that in EA (Figure 4.3 A). For validation, each model's ability to distinguish SLE patients from controls in data

from broader cell populations using previously published data from CD4+ pan-T cells and CD19+ pan-B cells were tested. The top AA random forest regression model accurately predicted SLE status with ROC-AUCs of 0.967 and 0.945 respectively in AA patients and controls (Figure 4.3 B).

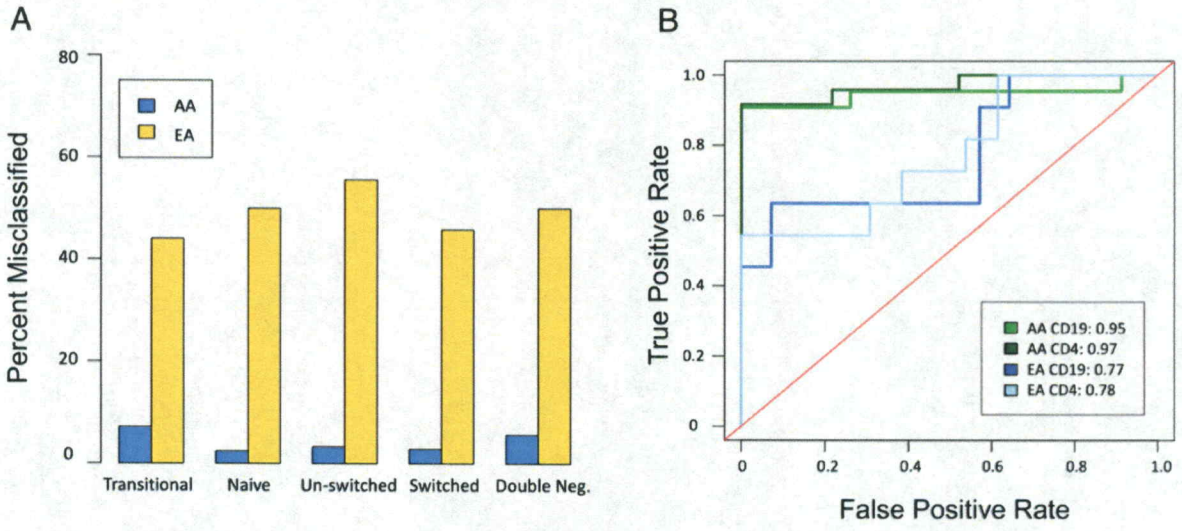


Figure 4.3 Machine learning identifies an epigenetic signature in transitional B cells that can predict SLE in AA patients. A) Model performance based on misclassification percentage across B cell stages in AA females (blue) and EA females (yellow). B) Validation ROC curves for AA CD19+ pan-B cells (green) and CD4+ pan-T cells (dark green) as well as EA CD19+ pan-B cells (blue) and CD4+ pan-T cells (light blue).

Unlike pure regression analysis, for which top results are often within correlated regions of the genome, random forest results present unique and highly predictive sites across the genome. Most of the top predictors (based on Gini importance) were associated with various genes scattered throughout the genome (Table B.8). The top predictor, cg07839457, is near the transcription start site (TSS) of NLR Family CARD Domain Containing 5 (*NLRC5*). Methylation analysis at cg07839457 across B cell development revealed perpetuation of the hypomethylated state in AA disease patients

compared to controls but not EA patients and controls (Figure B.9). Further interrogation of the methylation status at other highly predictive sites in transitional B cells revealed consistent hypomethylation in SLE cases for almost every CpG regardless of SLEDAI index.

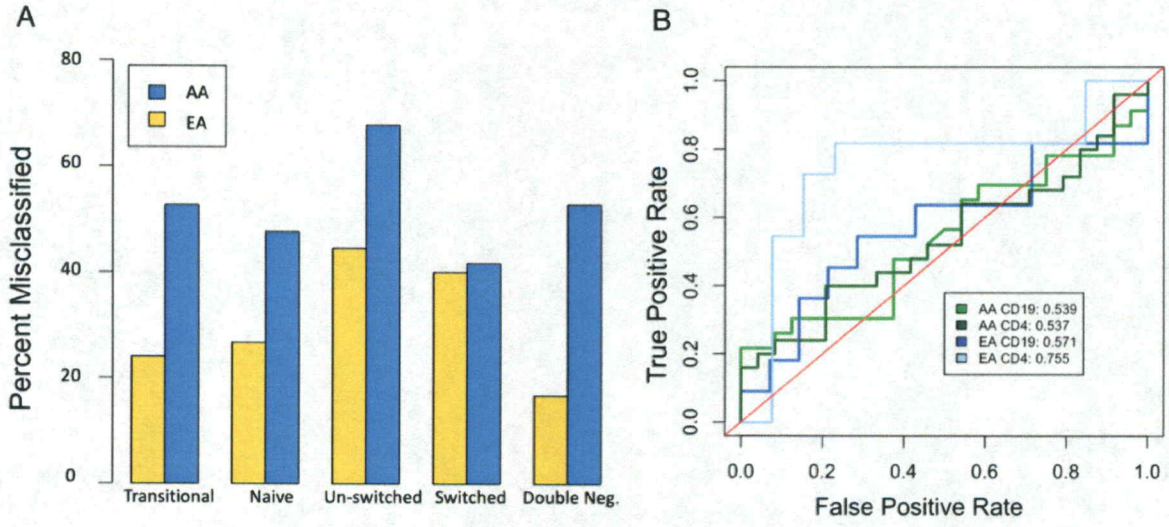


Figure 4.4 Machine learning in switched B cells modestly predicts SLE in EA patients. A) Model performance based on misclassification percentage across B cell stages in AA females (blue) and EA females (yellow). B) Validation ROC curves for AA CD19+ pan-B cells (green) and CD4+ pan-T cells (dark green) as well as EA CD19+ pan-B cells (blue) and CD4+ pan-T cells (light blue).

The optimal data reduction method for predictive modeling in EA patients utilized a p-value cut-off of $1x10^{-4}$ for CpGs associated with SLE status and dependent on EA ethnicity in the switched B cell subset. This model modestly distinguished EA SLE patients from controls, with <45% misclassification error across all B cell subsets and proved to have poor performance in AA B cells, resulting in misclassification percentages ranging from 42-68% across the 5 B cell stages assayed (Figure 4.4). As for

validation performance, the top EA model had a ROC-AUC of 0.571 and 0.755 for EA CD19+ pan-B cell and CD4 cells respectively (Figure 4.5). Performance was much lower for AA validation mixed immune cells, having a ROC-AUC of 0.539 and 0.537 for CD19+ pan-B cell and CD4 cells respectively. Predictors in the top EA random forest model had Gini coefficients ranging from 0.299-0.009 (Table B.9). There was a lack of methylation difference between EA SLE patient samples and controls (Figure B.10), in contrast to AA predictors for which severe differences in methylation exist.

4.3.3 SLE-associated CpGs near IFN-regulated genes are enriched for EBF1 binding

To determine whether CpGs significantly associated with SLE based on methylation status reside in regulatory regions of the genome, I performed an enrichment test of TFBS status near the top 100 CpGs (based on p-value) of the SLE-association analysis revealing enrichment for RNA Polymerase II Subunit A (*POLR2A*), *EBF1*, Chromodomain Helicase DNA Binding Protein 1 (*CHD1*), Werner Helicase Interacting Protein 1 (*WRNIP1*), Metastasis Associated 1 Family Member 3 (*MTA3*), Signal Transducer and Activator of Transcription 3 (*STAT3*) and IKAROS Family Zinc Finger 1 (*IZKF1*). A similar analysis of the ethnicity-specific SLE-associated sites identified enrichment of CpGs in binding sites for *POLR2A* ($p < 5x10^{-4}$), *EBF1* ($p < 8x10^{-4}$), POU Class 2 Homeobox 2 (*POU2F2*) ($p < 3.27x10^{-1}$), and RELA Proto-Oncogene, NF-KB Subunit (*RELA*) ($p < 4.46x10^{-1}$) (Figure 4.5). Amongst the most enriched TFs, *EBF1* is of particular interest, as it is an essential pioneer TF for B cell development (R. Li et al. 2018). Further interrogation of *EBF1* binding using a bioinformatics-based approach to identify *EBF1* motifs (5′-TCCCNNGGGA-3′) within the DNA sequence surrounding

those top SLE-associated CpGs strengthened this hypothesis, as hypermethylated sites significantly associated with SLE in transitional B cells were enriched for *EBF1* motifs. Amongst the top 100 SLE-associated population-specific CpGs based on p-value, 36 were near (+/- 250 bps) an *EBF1* binding motif, which was much greater than expected based on analysis of *EBF1* motifs near CpGs on a random subset of 100,000 CpGs covered by the Methyl450 array ($p < 0.015$). Genes associated with CpGs both near an *EBF1* ChiP-seq peak and an EBF1 motif included IFN-regulated genes such as, NLRC5, Interferon Induced Transmembrane Protein 1 (*IFITM1*), and the HLA class I Histocompatibility Complex P5 (*HCP5*) among others, further strengthening the hypothesis that EBF1 regulation is actively involved in SLE disease biology.

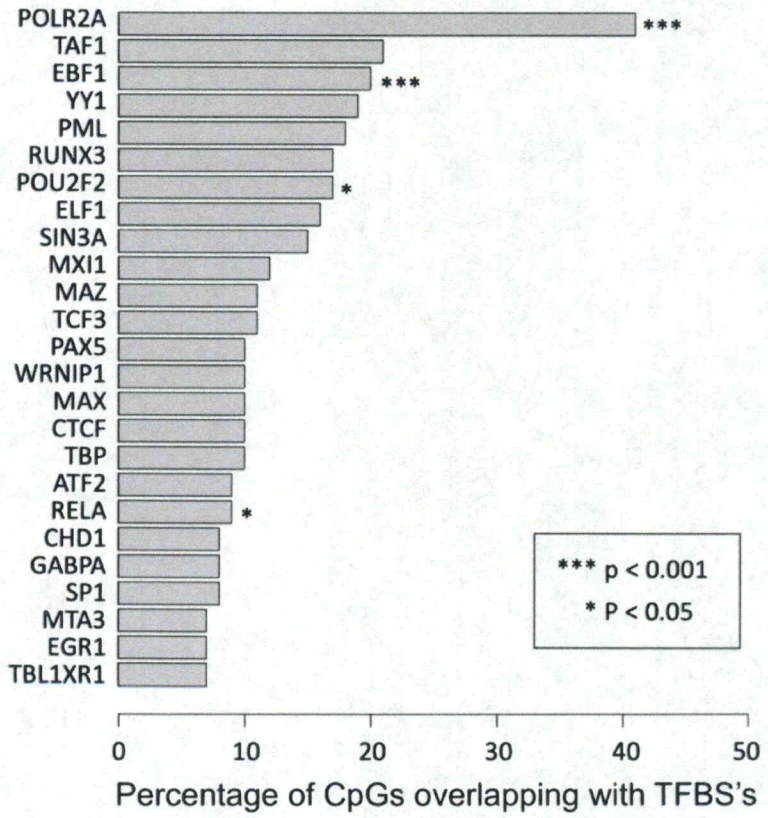


Figure 4.5 SLE-associated CpGs are enriched for EBF1 binding. TF enrichment within the top 100 ethnicity-dependent SLE associated CpGs using ENCODE ChIP-seq data. TFs that are statistically (based off of an empirical p-value) more enriched compared to the rest of the CpGs assayed are indicated by an asterisk (POLR2A, EBF1, POU2F2 and RELA).

4.4 Discussion

In this work, I used various statistical and machine learning methods to better understand the biological changes that occur throughout B cell development in SLE patients and the differences in these effects between AA and EA patients. My initial analysis was focused on following up previous studies that performed linear regression analysis in mixed cell populations. To discover CpG methylation associated with SLE status, I first implemented a simple regression model for SLE status within the discovery cohort across all 5 purified B cell subsets regardless of ethnicity. Several CpGs reached genome-wide significance ($1.07x10^{-7}$) across B cell subsets, including many that have been previously associated with the disease (Absher et al. 2013; Coit, Yalavarthi, et al. 2015; Jeffries et al. 2011; Coit et al. 2013; Park et al. 2017; Lugar et al. 2012; Wahadat et al. 2018). I was able to show that CpGs near IFN-regulated genes are hypomethylated in AA SLE patients from the earliest circulating B cell stage indicating that B cells might be epigenetically "primed" for an aberrant immune response in AA SLE patients prior to maturation. When analyzing EA SLE patients, I observed methylation changes in mature B cell stages, indicating potential differences in etiology between ethnicities. These epigenetic differences may explain the differences in disease presentation and severity between the ethnic groups as well. Epidemiology studies reveal that SLE risk and burden is significantly higher in AA females compared to other affected populations, which

could potentially be explained by the more severe epigenetic signature observed in AA females in this work. To further interrogate the ethnicity-specific methylation changes, we also performed a more complex regression model to identify ethnicity-specific SLE-associated CpGs. We note that many genes associated with our top AA-specific CpGs (based on p-value) are IFN-regulated and/or have previously been associated with SLE in other cohorts of various ancestral background (H. Wu et al. 2017; Bing et al. 2016; Mohan and Putterman 2015b). However, the same observation was not made for EA-specific CpGs.

The observed high level of within-patient correlation across genes suggests a modular epigenetic defect in SLE patients, prompting us to build multivariate predictive models of disease using machine learning optimization. The AA-specific predictor was extremely effective at distinguishing SLE patients from controls, with high sensitivity and specificity (ROC-AUC = 0.94). Of the 34 CpGs used in the top performing AA SLE prediction model, 7 had a statistically significant SLE association p-value ($p<1.07x10^{-7}$) and were near 6 different IFN-related genes, providing ancillary evidence that IFN plays a major role in SLE biology. A CpG at the TSS of *NLRC5*, and near a TFBS for *EBF1* among multiple other TFs, had the highest predictive power for SLE status (Gini = 0.461). *NLRC5* is one of 22 NLR family proteins, is highly expressed in the cytoplasm and nucleus of lymphocytes, and is a known regulator of major histocompatibility (MHC) class molecules (Meissner et al. 2010; Kobayashi and van den Elsen 2012). In Coit et al.'s analysis of genome-wide methylation in naïve CD8+ T-cells, hypomethylation at this same site near NLRC5 was observed (Coit et al. 2013). I hypothesize that hypomethylation in SLE patients at this site might alter regulation of

MHC class 1 molecules in response to upregulated IFN expression, which is characteristic of SLE. The next highest predictor (Gini coefficient = 0.453) is near Interferon Induced Protein With Tetratricopeptide Repeats 1 (*IFIT1*), an IFN stimulated gene that activates the innate immune system in response to inflammatory stimulation (McDermott et al. 2012). IFIT1 not only contains known risk alleles for SLE, but has also been repeatedly noted for having both decreased methylation and increased expression in SLE patients compared to control (Coit, Yalavarthi, et al. 2015; McDermott et al. 2012; Ye et al. 2003). A CpG near the 5' UTR of MX Dynamin Like GTPase 1 (*MX1*) (cg21549285) was also within the top predictors. This site is near a TFBS for *RUNX3*, MYC Associated Factor X (*MAX*) and *STAT3*, and is characterized by hypomethylation in AA cases compared to controls across B cell development. MX1 is activated by IFIT1, another top gene from my analysis, indicating that the model successfully found biologically relevant relationships between CpGs to build a superior predictive model of disease. MX1 is downstream in the type 1 IFN pathway and is an important component of the early innate immune system as it plays a role in the IFN-induces antiviral response against various viruses, but it's role in response to IFN stimulation in the absence of virus has yet to be determined (Melén et al. 1994; Nakayama et al. 1991; Haller and Kochs 1996). MX1 hypomethylation, increased RNA expression and higher protein concentrations are consistently observed in SLE patients (Shimizu et al. 2017; Coit, Ognenovski, et al. 2015; Watanabe et al. 2013). Multiple CpGs associated with MX1 were observed to be highly associated with SLE status in this analysis, and methylation at two MX1 CpGs reveals almost perfect separation of AA SLE

patients from controls leading us to hypothesize that a targeted assay for MX1 methylation would be fruitful in diagnosing SLE.

To assess whether SLE-associated CpG sites were in regulatory regions of the genome, the TFBSs near those regions using publicly available ChIP-seq data were interrogated. An increased number of binding sites for *POL2RA, EBF1, POU2F2* and *RELA* in the top SLE-associated CpGs compared to all other CpG sites assayed was observed. Of particular interest is the enrichment of *EBF1* interaction at highly associated IFN sites, as this TF is known as a pioneer TF especially in the B cell lineage (R. Li et al. 2018). Through a distinct DNA binding domain (DBD) at the N-terminus, *EBF1* binds DNA as a homodimer at a palindromic recognition site consisting of 18 base pairs (5′-TCCCNNGGGA-3′) stabilized by an α-α motif, referred to as a zinc knuckle (Treiber et al. 2010). *EBF1* operates as an epigenetic regulator that induces demethylation, nucleosome remodeling and active chromatin modifications to target genes (Vilagos et al. 2012). Binding sites for *EBF1* that are lowly methylated are enriched for enhancer sites that undergo TF-mediated changes in methylation (R. Li et al. 2018). A limitation to ChIP-sequencing is the inability to distinguish between direct and indirect protein-DNA interactions. To supplement the *EBF1* enrichment observed in ChIP-seq data, I performed a motif enrichment analysis in the top 100 SLE-associated CpGs, for which I observed an increased occurrence of EBF1 motifs ($p< 0.015$) within 500 bps the top 100 CpGs associated with SLE status in comparison to the rest of the CpG sites assayed. From these observations, I hypothesize that *EBF1* plays a role in the immune response through regulation of IFN-regulated gene expression. A possible mechanism in which this might occur is that hypomethylation at or near the *EBF1* motif

of IFN-regulated genes, as frequently observed in this data, increases chromatin accessibility at those regions allowing for higher affinity of *EBF1* binding. This would allow for increased expression of downstream genes involved in the IFN pathway and dysregulation of the immune response in AA SLE patients. ChIP-PCR experimentation in conjunction with methylation analysis in patient-derived B cells for which the methylation levels at *EBF1* binding regions should be performed to further investigate this hypothesis.

4.5 Conclusion

In conclusion, this work has allowed for the identification of an aberrant epigenetic signature that developed early in B cell development in AA patients. This AA-specific SLE signature is enriched for *EBF1* interaction and serves as an effective multivariate predictor of disease. A limitation to this study is the lack of genetic variant analysis. The presence of an SLE-related epigenetic signature, with a population-specific severity, could indicate either a shared set of methylation quantitative trait loci (methylQTLs) in each of the genes that are only carried by AA SLE patients, or alternatively, a defective signaling pathway that leads to correlated epigenetic effects at multiple genes. Given that the effects are seen at many unlinked genes and the unlikelihood of shared genetic variants at each of these unlinked genes among AA SLE patients, the former hypothesis is more likely. To add confirmation to this hypothesis, I analyzed CpG p-values within genes with known AA-specific SLE genetic associations from previous studies and found a lack of association with methylation and SLE status for those sites (Ghodke-puranik et al. 2019; Silvia N. Kariuki et al. 2010). If known SLE-related genetic risk factors were drivers of epigenetic effects, I would have expected

an overlap between those loci. Additionally, we point out that another limitation to this study was the number of EA patients included. The vast difference in methylation (effect size) observed in AA patients compared to controls allowed for the determination of several CpGs highly associated with SLE status, yet the observed effect size was much lower in EA patients compared to controls. While our EA sample size should have been sufficient to detect the effects seen in the AA patients, if they exist, power calculations indicate that we were well powered to detect an effect size of 0.09 (i.e. 9% change in methylation). Despite the limitations addressed, this study was the first to interrogate genome-wide methylation across individual B cell developmental stages in both AA and EA SLE patients. Overall, the epigenetic patterns appeared to be different in AA and EA patients, suggesting some mechanistic differences in etiology that may be related to the different clinical paths that these patient populations experience.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Summary of previous findings

SLE is a complex autoimmune disease characterized by altered immune cells which form immune complexes that induce systemic inflammation, organ damage and in some cases premature mortality (Papp et al. 2012). Many hypotheses regarding the initiating factors leading to SLE exist. A commonly cited hypothesis is that an environmental trigger, such as an infection or drug exposure, induces apoptosis causing T cells to recognize self-antigens and, in the presence of increased IFN, B cells to produce antibodies to self-antigens (autoantibodies) (H. Wu et al. 2017). This cycle of self-antigen recognition and increased IFN levels can result in dysregulation of B cell function, signaling, or development ultimately resulting in excessive autoantibody production and B cell hyperactivity (De and Barnes 2014). Recent work has found evidence that epigenetic dysregulation of B cell differentiation is an important mechanism underlying SLE pathogenesis (Absher et al. 2013; Coit et al. 2013; Jeffries et al. 2011). Yet, the exact nature of the relationship between epigenetic dysregulation and B cell autoreactivity has not been established.

DNA methylation studies focused on SLE have revealed widespread hypomethylation of CpGs within IFN-related genes in SLE patients relative to controls in

European and Asian populations (Absher et al. 2013). Population-specific risk loci have recently been identified, concluding that different ethnic populations must be studied independently to fully understand the genetic mechanisms of SLE (Langefeld et al. 2017). Through interrogation of genome-wide methylation across individual B cell developmental stages in both AA and EA SLE patients, I have identified an altered epigenetic signature enriched for IFN-regulated genes that is present in the earliest stage of B cell development and is specific to AA female patients. Specifically, *MX1, IFIT1, IFITM1,* and *IFI44L* are of interest as they reached genome-wide significance and, when incorporated into a multivariate predictive model of SLE, proved to be effective in discriminating AA SLE patients from controls regardless of which immune cell type was used for validation. Further interrogation of top associated IFN CpGs of AA patients revealed an enrichment for *EBF1* binding, indicating a role for *EBF1* in IFN signaling. *EBF1* is of specific interest because of its ability to demethylate promoter regions of target genes as well as its essential role in maintenance of B cell development (Vilagos et al. 2012).

The conclusions that: 1) AA and EA patients have unique epigenetic changes specific to disease, 2) epigenetic dysregulation exists in the first developmental stage (transitional) of B cells in AA patients, 3) the most associated AA-specific CpGs occur at IFN-regulating genes and are enriched for binding of TFs involved in immune regulation, such as *EBF1*, and 4) these epigenetic changes are highly predictive of disease in all AA patient immune cell populations instigate further hypotheses about the pathophysiology of SLE. Particularly with respect to AA patients, a population for which SLE research lags behind despite being the most at risk and severely affected demographic.

Specifically, these conclusions have generated the following hypotheses: 1) that hypomethylation of IFN-regulating genes primes B cells for an increased response to IFN, 2) that *EBF1* occupancy is required for an IFN response at important IFN-regulated genes, and 3) that the IFN epigenetic signature across mature immune cells in AA females is heritable. In the following sections I will provide more detailed evidence to support these hypotheses and propose experiments that could be performed to test these hypotheses.

5.2 Epigenetic Mechanisms of IFN Signaling in SLE

IFN's role in increasing antigen presentation, activating dendritic cells, and increasing MHC expression along with initial observations that treatments directly affecting the IFN pathway induced SLE led researchers to examine the role IFN perturbation plays in SLE pathophysiology (Ioannou and Isenberg 2000; Iwamoto et al. 2018; Blanco et al. 2001; Timothy B. Niewold 2008). High serum IFN is now a known heritable risk factor for SLE, and patients with high serum IFN tend to have increased downstream, IFN-induced gene expression (T. B. Niewold et al. 2007; Timothy B. Niewold and Swedler 2005; Mathian et al. 2015; Ivashkiv and Donlin 2015; Petri et al. 2009; Wahadat et al. 2018; Feng et al. 2006; Hoffman et al. 2017; Bennett et al. 2003; Baechler et al. 2003). Several GWA studies have identified multiple genetic variants within IFN genes across the genome unique to AA patients, such as *APOL1, MHC, HLA, ITGAM, BANK1, IRF8, IRF5, MW1* and *IFIH1*, that confer increased risk for SLE, solidifying the importance of this pathway in SLE etiology (Freedman et al. 2014; Matzaraki et al. 2017; Sánchez et al. 2011; Lodolce et al. 2010; J. Wu et al. 2014; Lessard

et al. 2012; Robinson et al. 2011; Kelly et al. 2008; Harley and Alarcón-Riquelme 2013; AlFadhli et al. 2016). A positive correlation between autoantibody production, serum IFN levels, and SLE activity has been shown, and the correlation is strongest in AA SLE patients compared to other ethnicities (Hamilton et al. 2018). Specifically, IFN-regulated gene expression correlates with SLE symptoms in AA patients more strongly than any other ethnicity (Iwamoto et al. 2018; Ko et al. 2013). Ko et al. showed that increased IFN-regulated gene expression patterns are directly related to the presence of autoantibodies in AA SLE patients while the same is not true for EA SLE patients (Ko et al. 2013). Furthermore, higher serum IFN levels in AA SLE patients compared to EA patients coincide with higher autoantibody levels (Weckerle et al. 2011).

The work I performed interrogating the methylation profile of sorted B cells revealed significant hypomethylation of IFN-regulated genes associated with SLE status in AA SLE patients. Together with previous observations, this leads to the conclusion that SLE patients have increased IFN levels and altered IFN-related gene expression. However, the causal relationship between the hypomethylated state of IFN-regulated genes and IFN levels in SLE patients has yet to be determined. The hypothesis that IFN-regulating genes are already hypomethylated allowing for more rigorous B cell activation in response to IFN can be tested by direct manipulation of DNA methylation in important regulatory regions of IFN-regulated genes, treatment with IFN, and exposure of B cells to activating stimuli, as depicted in figure 5.1. To manipulate the DNA methylation at specific sites in the genome of B cells, the clustered regularly interspersed short palindromic repeats (CRISPR) deactivated CRISPER associated protein (dCas9) system can be coupled with either the TET enzyme or a methyltransferase (Dnmt3a). Targeted

demethylation can be performed though the use of the Tet enzyme, which demethylates cytosines in the genome through dioxygenase-catalyzed 5-methylcytosine oxidation (Ito et al. 2011). Xu et al. recently demonstrated the catalytic domain of the Tet enzyme can be combined with the bacterial CRISPR system for efficient and accurate targeted demethylation of CpG sites throughout the genome (X. Xu et al. 2016). By tethering the dCas9, specific guide RNAs (gRNAs) to the target region of the genome, and the catalytic domain of the TET enzyme (Tet1-CD) specific demethylation of the CpG sites found to be hypomethylated near IFN-regulating genes in AA SLE patients could be performed. The same approach can be taken for targeted methylation, except instead of using the Tet enzyme a DNA methyltransferase responsible for de novo methylation (Dnmt3a) would be used (Liu et al. 2016).

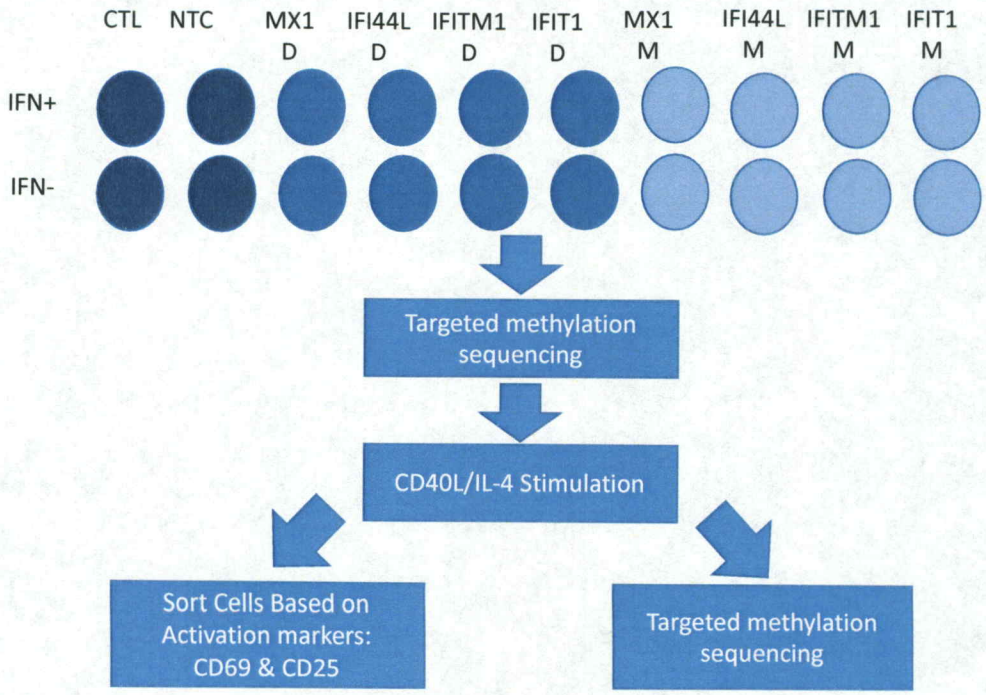


Figure 5.1 Experimental design for the interrogation of the epigenetic mechanisms of the IFN response.

After altering the methylation status of IFN-regulating sites, half of the cells would then be treated with IFN. Treatment with or without IFN in conjunction with Cluster of Differentiation 40 Ligand/Interleukin 4 (CD40L/IL-4) stimulation would be performed on CRISPR treated B cells in addition to untreated and non-targeting CRISPR control treated B cells (Donahue and Fruman 2014). To assess differences in activation between treatment groups, B cells can be sorted using flow cytometry based on activation markers (CD69 and CD25). Furthermore, targeted methylation sequencing prior to and following stimulation with CD40L and IL-4 can reveal changes in nearby DNA methylation of IFN-regulated genes following treatment. The outcome affirming the hypothesis that hypomethylated IFN-regulated genes increase B cell activation in response to IFN, would be that IFN treatment of B cells with the CRISPR demethylated target sites leads to increased activation (as measured by proportion of cells expressing CD25/69) relative to control B cells and stable DNA methylation post-IFN treatment compared to pre-IFN treatment. If IFN treatment in the control and CRISPR methylated and/pre demethylated B cells leads to increased activation relative to the B cells with no IFN treatment and target sites become demethylated following IFN treatment, then the outcome would support the alternative hypothesis, that increased IFN leads to hypomethylation of IFN-regulating genes in B cells. Multiple possible outcomes could occur to support the null hypothesis, that there is no relationship between IFN levels, hypomethylation of IFN-regulating genes in B cells, and B cells activation. These include the observations that 1) there are no activation differences across treatment and

control B cells, 2) B cells with methylated target sites have more activity than B cells with demethylation target sites, and 3) B cells without IFN treatment have more activity than those with IFN treatment.

5.3 EBF1 as a Regulator of IFN Response

The observation of enrichment for *EBF1* binding sites near significantly hypomethylated CpGs in my work begs the question of whether hypomethylation of IFN-regulated promoter regions influences *EBF1* binding, or *EBF1* binding demethylates IFN-regulated promoter regions. The EBF family of TFs consists of 4 members which are highly conserved evolutionarily, yet structurally unique compared to other TF families (H. Lin and Grosschedl 1995). Work over the past 20 years has led to a better understanding of not only the *EBF1* structure but also its function as a cell-type specific TF (Siponen et al. 2010; Nechanitzky et al. 2013; Barneda-Zahonero et al. 2012). Originally identified in B cells and olfactory neurons, *EBF1* is now appreciated for its involvement in various specialized cells, such as adipocytes and osteoclasts. Extensive work has shown the essential role that *EBF1* plays in promoting B cell maturation from the earliest progenitor stage (R. Li et al. 2018; Nechanitzky et al. 2013; Siponen et al. 2010). *EBF1* operates as an epigenetic regulator that induces demethylation, nucleosome remodeling and active chromatin modifications to target genes (Vilagos et al. 2012). Binding sites for *EBF1* that are poorly methylated are enriched for enhancer sites that undergo TF-mediated changes in methylation (R. Li et al. 2018). Time resolved analysis of *EBF1* activity shows that *EBF1* occupancy coincides with expression and precedes chromatin accessibility (R. Li et al. 2018).

Together, the observation that significantly demethylated IFN CpGs in AA SLE patients lie within *EBF1* binding sites, and the knowledge that EBF1 acts as a pioneer TF throughout B cell development and maturation, supports the hypothesis that: hypomethylation of IFN-regulated genes and *EBF1* binding leads to dysregulation of IFN signaling in B cells of AA patients with SLE. In order to test this hypothesis, a similar approach to the previous question could be taken utilizing the CRISPR system in sorted B cells, as depicted in figure 5.2. Use of the CRISPR system with the activated Cas9 endonuclease allows for the mutation of sites of the genome when tethered to a gRNA specific to the region of interest (Mali et al. 2013). In this case, the *EBF1* motif (5'-ATTCCCNNGGGAATT-3') within *MX1* regulatory region would be targeted as *MX1* methylation status proved to be highly indicative of SLE status in AA females in my work, and it is a known GTPase in response to IFN stimulation (J. Wang et al. 2012; Haller and Kochs 1996). For this experiment, the binding motif for *EBF1* would be mutated to inhibit binding of *EBF1* within the promoter region of *MX1*. In addition to a *MX1* mutated *EBF1* binding group, a control group and a nontargeting CRISPR control group would also be tested in the presence and absence IFN treatment. Targeted detection of *EBF1* binding can be performed using ChIP quantitative PCR (qPCR). ChIP qPCR allows for the interrogation of protein-DNA interactions of regions of interest. Furthermore, qPCR and targeted methylation sequencing of the *MX1* promoter region would give insight into *EBF1* motif ablation on the surrounding epigenetic state and expression level of *MX1*.

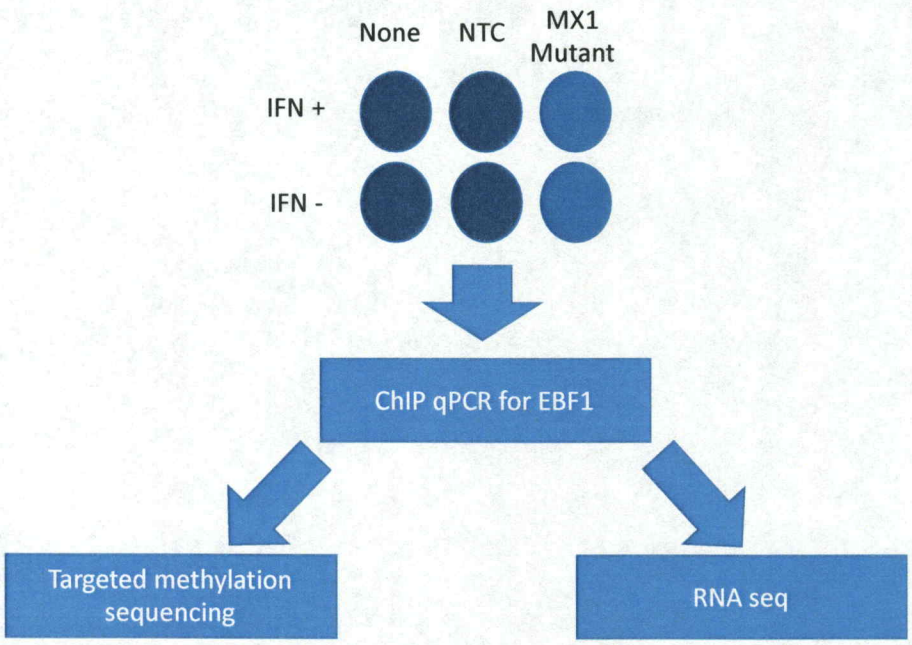


Figure 5.2 Experimental design for the interrogation of the role of EBF1 in the IFN response.

Evidence to support the hypothesis that *EBF1* is required for an IFN response at important IFN-regulating sites include the observations of 1) hypomethylation of *MX1* in control compared to *MX1* mutant B cells and 2) increased expression of *MX1* in control B cells compared to *MX1* mutant B cells. However, the alternative hypothesis, that *EBF1* is not required for an IFN response at important IFN-regulating sites, would be supported if the *MX1* mutation does not alter methylation or expression compared to control B cells.

5.4 Genetic influence of the Epigenetic IFN Signature

In my work, I observed that the epigenetic IFN signature seen in transitional B cells is shared amongst other mature B cell subsets as well as pan-T cells and monocytes of AA female SLE patients. As previously mentioned, GWA studies have identified multiple genetic variants within IFN genes across the genome unique to AA patients

(Freedman et al. 2014; Matzaraki et al. 2017; Sánchez et al. 2011; Lodolce et al. 2010; J. Wu et al. 2014; Lessard et al. 2012; Robinson et al. 2011; Kelly et al. 2008; Harley and Alarcón-Riquelme 2013; AlFadhli et al. 2016). Yet, these findings do not offer a complete understanding of the heritability of SLE, a disease for which multiple sub-phenotypes exist. Several lines of evidence suggest that identification of IFN-associated loci rather than SLE-associated risk loci may lead to a better understanding of SLE susceptibility since increased IFN is a widely shared trait among all SLE sub-phenotypes (Robinson et al. 2011; S N Kariuki et al. 2015; Rullo et al. 2011; Timothy B. Niewold et al. 2008). Recent work has identified genetic variants that increase SLE susceptibility through effect of the epigenome, known as methylation quantitative trait loci (methylQTL) and histone quantitative trait loci (hQTL) (Demirci et al. 2016; Pelikan et al. 2018; Imgenberg-Kreuz et al. 2016). These recent observations compliment the conclusion made from my work, that the IFN epigenetic signature is shared amongst diverse immune cells in AA SLE patients. Ultimately, these findings support the hypothesis that the IFN epigenetic signature in AA females with SLE is determined by an inherited genetic background or haplotype.

An ideal experiment to address this question would be a twin study conducted on discordant AA female twins, as the present work showed a more severe difference in methylation in AA females compared to EA females. SLE familial studies to date have misrepresented populations affected by SLE (Deafen et al. 1992; Alarcón-Segovia et al. 2005). One of the largest SLE familial studies thus far took a multiethnic approach by incorporating American populations of African and Caribbean descent as well as Europeans in the determination of familial aggregation of SLE amongst other

autoimmune diseases (Alarcón-Segovia et al. 2005). By doing so, Alarcón-Segovia et al. identified that familial aggregation of SLE is much higher in the AA population (0.375) compared to the European (0.010-0.081) and African Caribbean (0.110.25) populations, suggesting that inheritance of SLE should be studied for individual populations rather than multiethnic populations as a whole to better understand the hereditary nature of complex diseases such as SLE. Furthermore, twin studies offer an improved model for understanding of the heritability of complex diseases compared to more general familial approaches. Through interrogation of disease outcome in monozygotic twins, the degree of genetic versus environmental components of complex diseases can be identified. Data from the National Twin Registry demonstrates that the concordance rate for SLE in monozygotic twins is no more than 24% while that of dizygotic twins is approximately 2%, suggesting a strong environmental component for SLE susceptibility.

To answer the question, of whether the IFN epigenetic signature in AA females with SLE is determined by genetics, AA female twins with and without SLE would have both methylation and RNA sequencing performed on sorted B cells, as depicted in Figure 5.3. I hypothesize that the IFN epigenetic signature would be present only in twins affected by SLE. The observation of an IFN hypomethylated signature in affected twins only would support this hypothesis, and would indicate that the genetic background alone is not responsible for SLE risk but rather a nongenetic event, such as an environmental trigger, leads to the epigenetic IFN signature observed in SLE patients. If the IFN epigenetic signature is observed both in SLE twins and healthy twins, the alternative hypothesis would be supported indicating that the epigenetic defects observed in AA female SLE patients are the result of a shared genetic background.

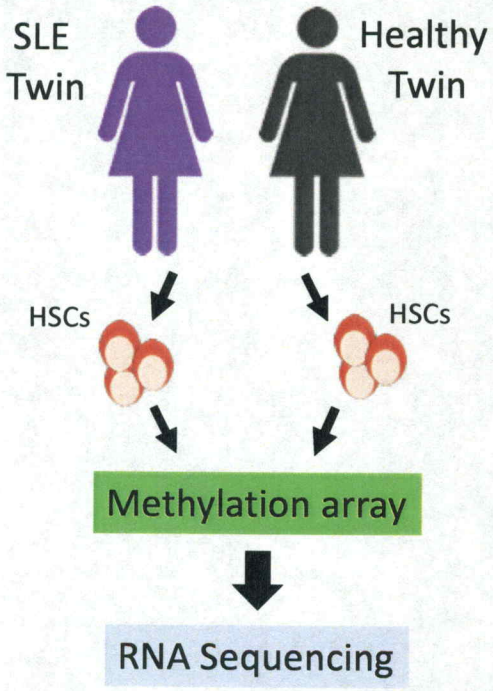


Figure 5.3 SLE twin study design.

Obtaining both gene expression and methylation data from twins allows for the ability to determine whether these hypomethylated IFN CpGs correspond with increased IFN gene expression. Based on the canonical model of methylation, CpG methylation near gene promoter regions and near TSSs are negatively correlated with gene expression (Coit et al. 2013; Coit, Yalavarthi, et al. 2015; Busslinger, Hurst, and Flavell 1983; McGhee and Ginder 1979). Therefore, I would expect to see higher expression of IFN-regulating genes in SLE patients who bear hypomethylation at the previously identified SLE-associated CpGs, especially since most are within the promoter regions and TSS of their corresponding genes. Alternatively, recent work has highlighted differences from the canonical model of methylation and expression (Olsson et al. 2014; Wagner et al.

2014; Hu et al. 2016). Lack of direct relationship between methylation at these sites and gene expression would indicate more complicated epigenetic regulation of gene expression at these IFN-regulating sites.

APPENDIX A

Table A.1 Number of variants before and after QC in both cohorts.

Data set	**Number of SNPs**
Initial Wellderly	**8018**
Initial UPM	**25273**
Final (both)	**5896**

Table A.2 Number of variants in each data set for Random Forest predictive modeling.

Subset	**Number of Variants**
Variants	5896
Rare Variants	2773
Very Rare Variants	998
Medium CADD Score Variants	530
High CADD Score Variants	140
Target Variants	1962

Rare Target Variants	963
Very Rare Target Variants	331
Medium CADD Score Target Variants	203
High CADD Score Target Variants	60
Variants +/- 50kb of Target Genes	4522
Rare Variants +/- 50kb of Target Genes	2102
Very Rare Variants +/- 50kb of Target Genes	751
Medium CADD Score Variants +/- 50kb of Target Genes	410
High CADD Score Variants +/- 50kb of Target Genes	113
Exon Variants	228
Rare Exon Variants	99
Very Rare Exon Variants	45
Medium CADD Score Exon Variants	51
High CADD Score Exon Variants	20
CodingAnnoType_variants	162
Known Variants	5067
Unknown Variants	829
TFBS Variants	1180
SIFT Deleterious Variants	28
SIFT Tolerant Variants	56
eQTL Variants	30

GWAS Variants	1540
Control	5896

Table A.3 Single Variant Association of the aging phenotype.

Chr	Pos	rsID	Gene	P-Value	Adj. P-Value	tAF
1	156079083	rs915180	LMNA	0.0015	1	0.56
1	156078249	rs915179	LMNA	0.0017	1	0.56
8	30910690	rs6989940	WRN	0.0017	1	0.07
8	30911082	rs6991755	WRN	0.0017	1	0.07
1	156045662	rs10047112	MEX3A	0.0028	1	0.59
1	155993678	rs55935614	SSR2	0.003	1	0.05
1	156074845	rs6661281	LMNA	0.0035	1	0.6
8	30909416	rs55932348	WRN	0.0037	1	0.04
8	30897476	rs55895301	WRN	0.0039	1	0.05
8	30907657	rs56111434	WRN	0.0039	1	0.06
8	30926637	rs11574211	WRN	0.0044	1	0.06
8	31026051	rs2553257	WRN	0.0045	1	0.9
9	21999800	rs3218007	RP11-149I24	0.0057	1	0.16
9	22000247	rs3218005	RP11-149I24	0.0057	1	0.16

9	22000841	rs3218002	RP11-149I24	0.0057	1	0.16
9	22053895	rs17756311	CDKN2B-AS1	0.0059	1	0.07
9	22054164	rs74655961	CDKN2B-AS1	0.0059	1	0.07
9	22054356	rs17694572	CDKN2B-AS1	0.0059	1	0.07
8	30993804	rs56359757	WRN	0.0067	1	0.04
8	31018962	rs67722242	WRN	0.0067	1	0.04

Table A.4 Labels for Random Forest Models based on data filter.

Label	**Subset**	**Number of Variants**
1	Variants	5896
2	Rare Variants	2773
3	Very Rare Variants	998
4	Medium CADD Score Variants	530
5	High CADD Score Variants	140
6	Target Variants	1962
7	Rare Target Variants	963
8	Very Rare Target Variants	331

9	Medium CADD Score Target Variants	203
10	High CADD Score Target Variants	60
11	Variants +/- 50kb of Target Genes	4522
12	Rare Variants +/- 50kb of Target Genes	2102
13	Very Rare Variants +/- 50kb of Target Genes	751
14	Medium CADD Score Variants +/- 50kb of Target Genes	410
15	High CADD Score Variants +/- 50kb of Target Genes	113
16	Exon Variants	228
17	Rare Exon Variants	99
18	Very Rare Exon Variants	45
19	Medium CADD Score Exon Variants	51
20	High CADD Score Exon Variants	20
21	CodingAnnoType_variants	162
22	Known Variants	5067
23	Unknown Variants	829
24	TFBS Variants	1180
25	SIFT Deleterious Variants	28
26	SIFT Tolerant Variants	56
27	eQTL Variants	30
28	GWAS Variants	1540

29	Control	5896

Table A.5 Genomic information of top predictors from the best performing model.

Chr	Pos	Ref. Allele	Alt. Allele	Gene	rsID	tAF	Exon	SIFT Category	CADD Score	GINI Score
1	226555302	A	G	PARP1	rs1136410	0.24	17/23	tolerated	20.9	1.262
13	103528002	G	C	ERCC5	rs17655	0.38	15/15	deleterious	18.44	1.150
1	156099669	T	G	LMNA	rs513043	0.17	2/13	deleterious	18.33	1.027
16	14029033	G	A	ERCC4	rs1800067	0.03	8/11	deleterious	36	0.810
15	91354521	G	A	BLM	rs7167216	0.07	19/20	deleterious	15.2	0.446
15	91326099	C	T	BLM	rs11852361	0.05	13/20	deleterious	19.45	0.406
9	32974493	C	T	APTX	rs104894103	NA	7/8	NA	35	0.316
10	50690821	G	A	ERCC6	rs114852424	NA	4/15	deleterious	35	0.270
15	89873364	C	G	POLG	rs61752784	0.0037	3/23	deleterious	32	0.243
15	89866691	C	G	POLG	rs121918054	5.00E-04	1/4	deleterious	25.5	0.229
8	31012237	C	G	WRN	rs78488552	9.00E-04	32/35	deleterious	19.22	0.222
9	32984657	A	T	APTX	rs141195622	0.0014	6/8	tolerated	18.05	0.218
8	31030535	C	T	WRN	rs11574410	0.0018	35/35	NA	40	0.211
13	103518693	G	A	ERCC5	rs142438319	NA	10/15	deleterious	22.5	0.203
10	50680422	C	T	ERCC6	rs145720191	0.0023	10/15	deleterious	32	0.190
10	50708599	C	T	ERCC6	rs41549213	NA	7/21	tolerated	21.9	0.167
13	103527849	G	C	ERCC5	rs9514066	1	15/15	deleterious	17.11	0.108
10	50732202	T	G	RP11-123B3.6	rs4253046	NA	5/6	deleterious	17.18	0.036
15	89873337	T	A	POLG	rs138929605	NA	3/23	tolerated	15.41	0.015

Table A.6. Information for GWAS variants in the top performing model.

snp ID	GWAS P-value	PMID	GWAS Phenotype
rs7167216	5.10E-05	23555315	Breast cancer
rs1136410	4.20E-04	20686565	LDL cholesterol
rs1136410	9.50E-04	20686565	Total cholesterol
rs1136410	1.00E-03	17554300	Hypertension, combined control dataset, gender differentiated
rs1136410	1.90E-03	20935629	Waist hip ratio
rs1136410	2.60E-03	22504419	Infant head circumference
rs17655	1.40E-02	21909115	Diastolic blood pressure (DBP)
rs7167216	1.70E-02	23722424	College completion
rs11852361	2.10E-02	23722424	College completion
rs1136410	2.10E-02	23202125	Coronary artery disease (CAD)
rs1800067	2.30E-02	20686565	HDL cholesterol
rs1136410	2.30E-02	22210626	Age at death with kuru exposure
rs1136410	2.60E-02	23474815	Refractive error

rs1136410	3.20E-02	19060906	LDL cholesterol
rs1136410	3.40E-02	21124317	Neuroblastoma (brain cancer)
rs11852361	3.50E-02	20339536	Triglycerides change with statins
rs7167216	3.80E-02	20339536	Triglycerides change with statins
rs7167216	3.90E-02	20339536	Total cholesterol change with statins
rs1136410	4.00E-02	22837397	Salmonella-induced pyroptosis
rs1136410	4.10E-02	18483556	Eye color
rs1136410	4.20E-02	20383146	Chronic kidney disease

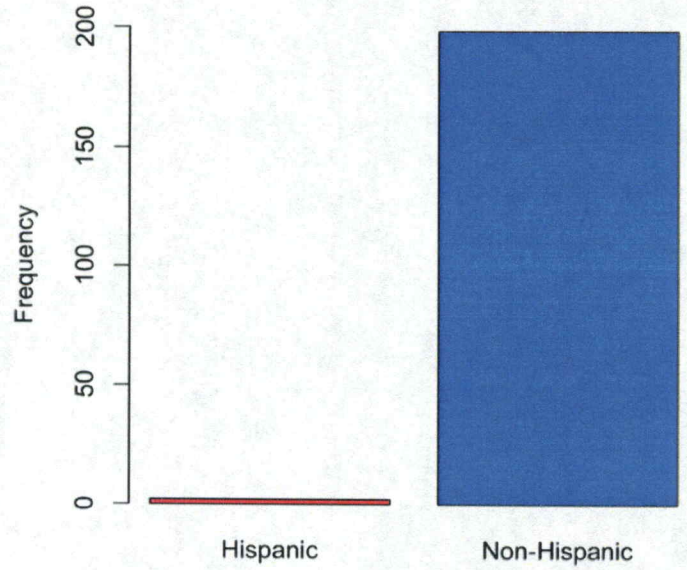


Figure A.1 Bar plot of the discovery cohort Hispanic background distribution.

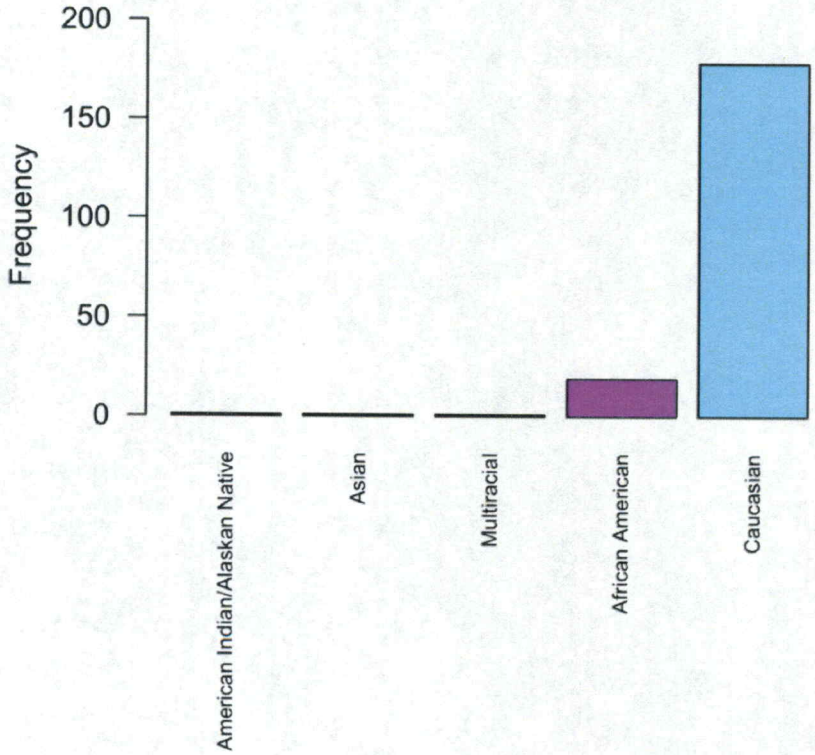


Figure A.2 Bar plots of the non-Hispanic discovery cohort ethnic background.

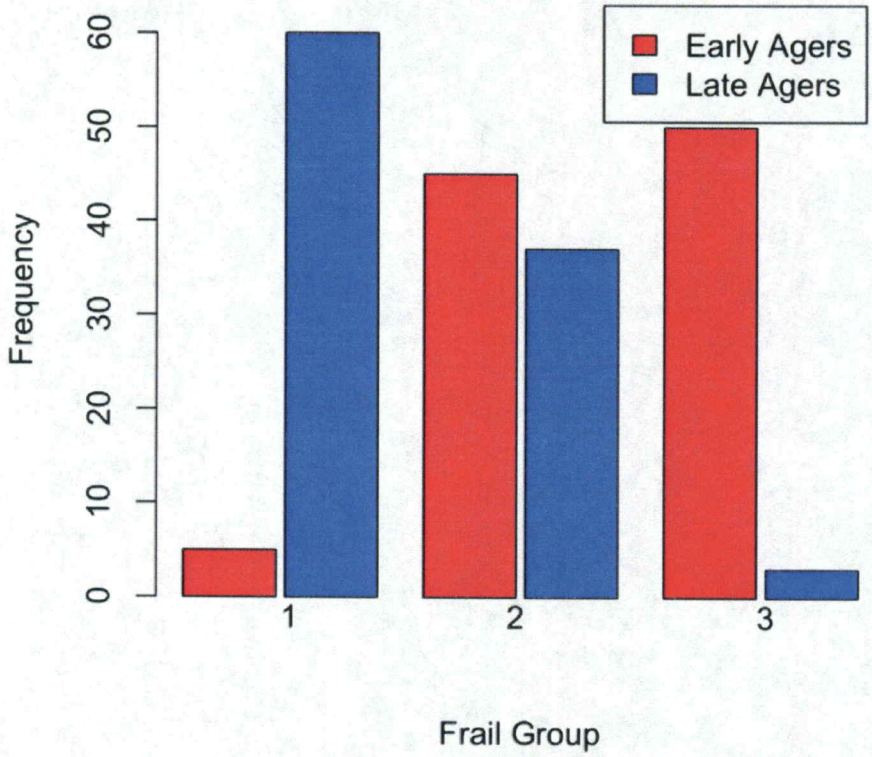


Figure A.3 Bar plot of discovery cohort frailty group (1-3) by age group (Early and Late).

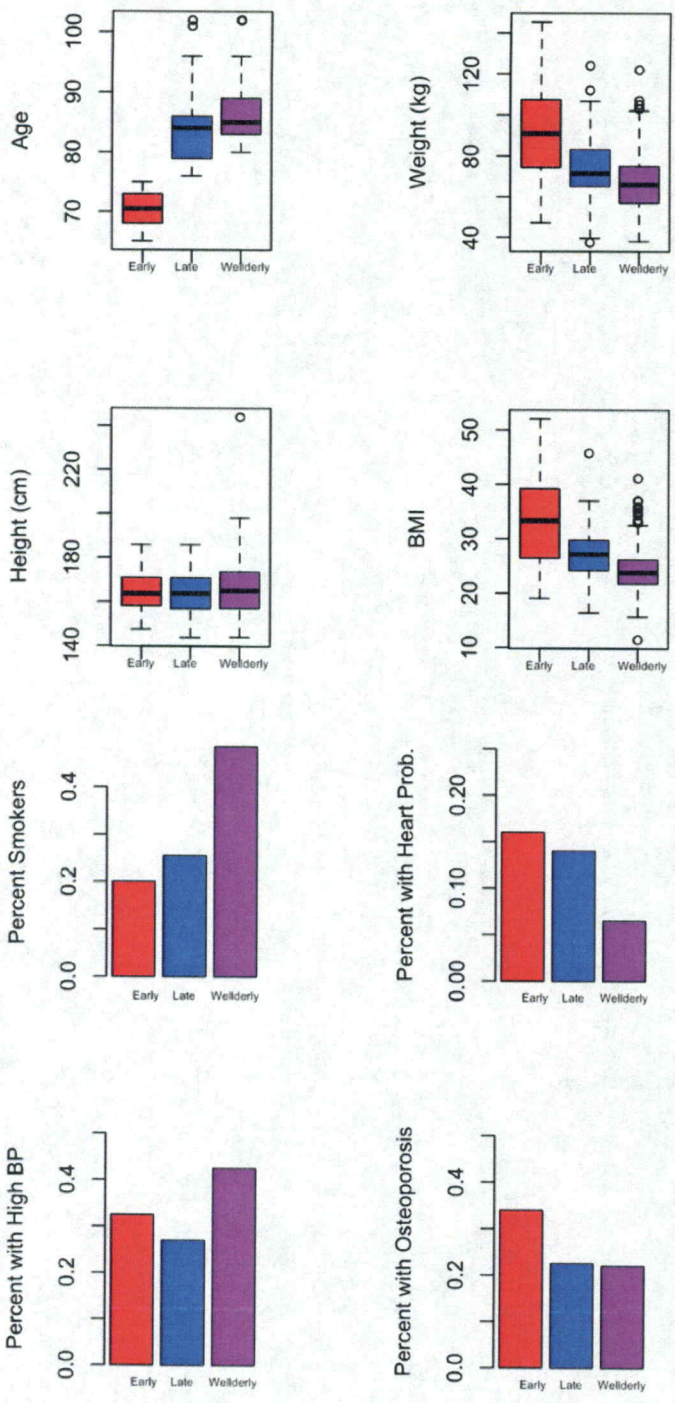


Figure A.4 Box and bar plots of overlapping phenotype data in the training and discovery cohorts.

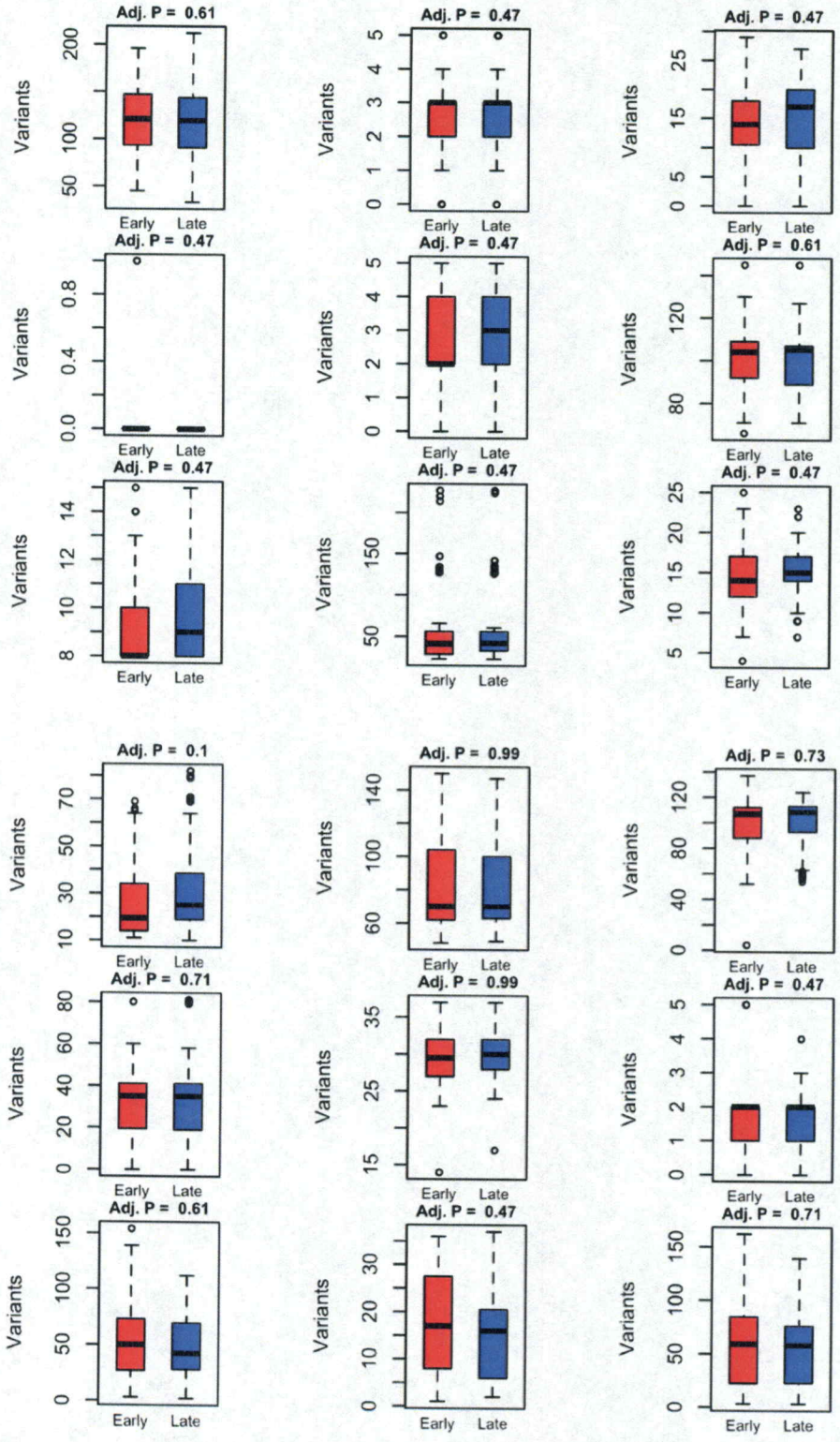


Figure A.5 Box plots on non-reference allele counts in each target gene by age group.

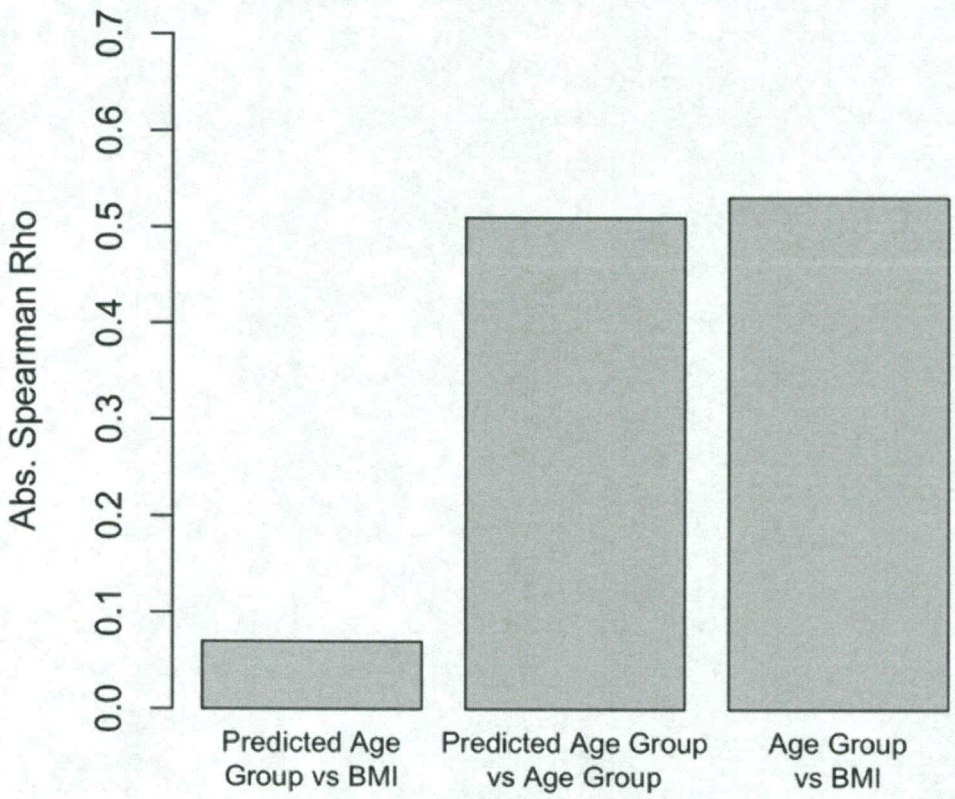


Figure A.6 Bar plot of the Spearman Rho correlation coefficient between the predicted age group and BMI, the predicted age group and the actual age group, and the actual age group and BMI.

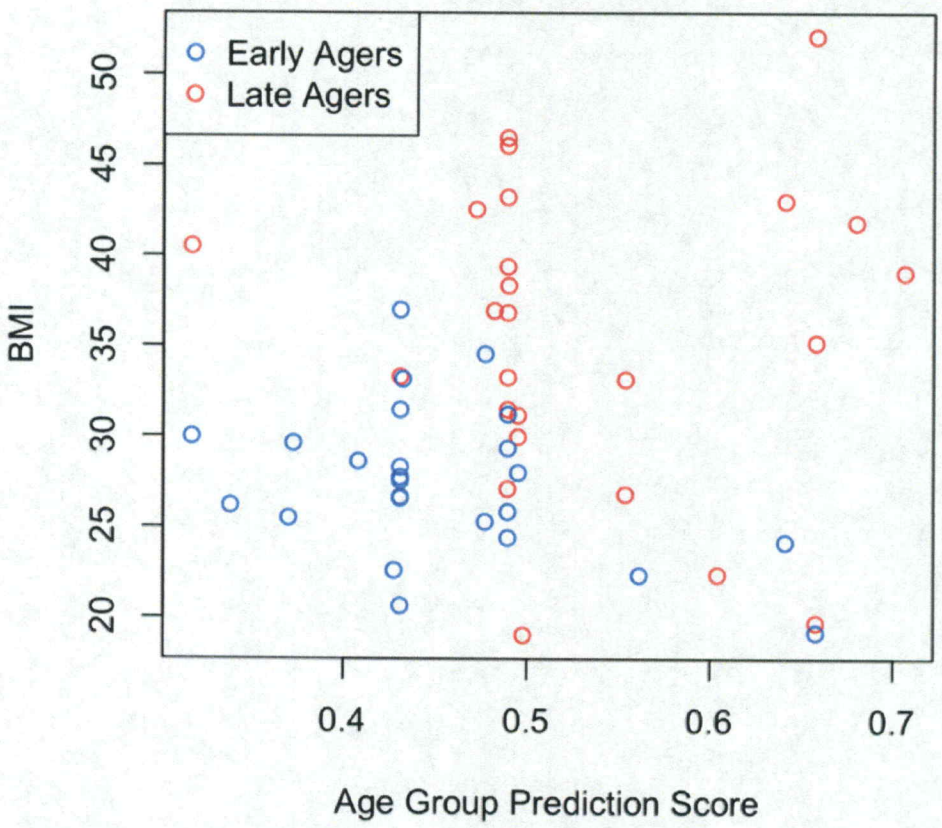


Figure A.7 Scatter Plot of the age group prediction versus BMI for the discovery and validation cohorts in the top model.

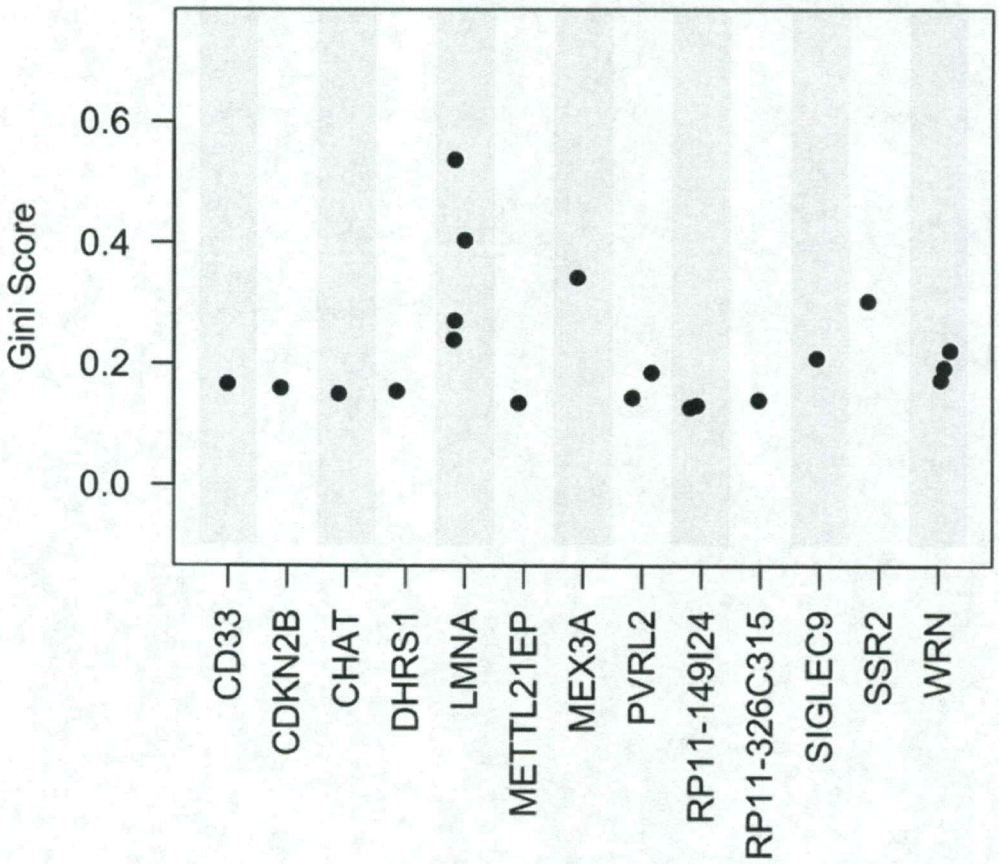


Figure A.8 Scatter Plot of the mean Gini Scores for each variant by gene from the non-filtered model.

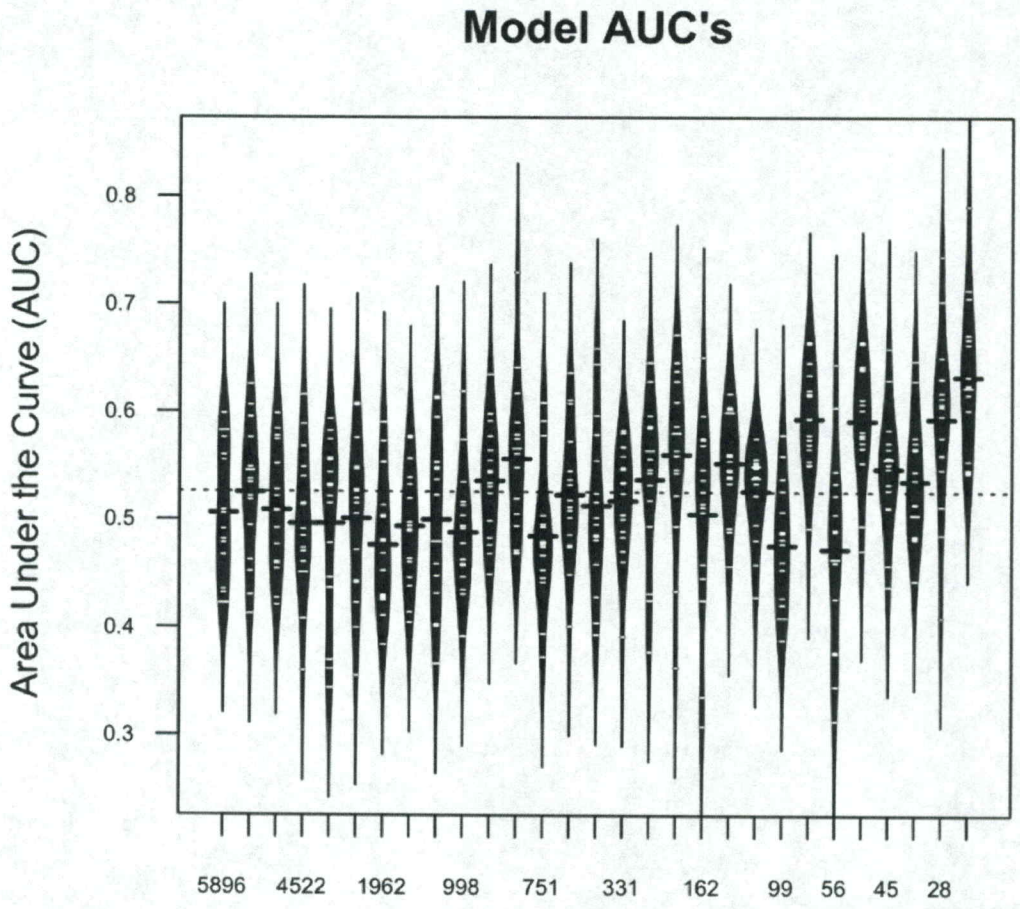


Figure A.9 Bean plot of random forest model AUCs

APPENDIX B

Table B.1 SLE discovery and validation demographic data.

Cohort	Discovery SLE (AA)	Discovery Control (AA)	Validation SLE (AA)	Validation Control (AA)	Discovery SLE (EA)	Discovery Control (EA)	Validation SLE (EA)	Validation Control (EA)
Number of Patients	24	25	23	22	18	13	11	13
Median Age (years)	37	37.37	28	29.5	38.145	37.13	37	25
Age Range (years)	22-49	23.96-47.13	19-49	22-48	22-54.47	24.13-48.13	25-47	21-65
Number of Smoked	1	0	N/A	N/A	1	3	N/A	N/A
Number of Patients in a Flare	N/A	N/A	7	0	N/A	N/A	4	0
Median BMI	28.45	N/A	N/A	N/A	30.11	20.23-44.27	N/A	N/A
BMI Range	17.97-52.18	N/A	N/A	N/A	20.23-44.27	N/A	N/A	N/A
Median SLEDAI	2	N/A	N/A	N/A	0	N/A	N/A	N/A
SLEDAI Range	0-12	N/A	N/A	N/A	0-4	N/A	N/A	N/A
Median Nephritis Stage	3	N/A	N/A	N/A	0	N/A	N/A	N/A
Nephritis Range	0-5	N/A	N/A	N/A	0	N/A	N/A	N/A
Number of Patients with Nephritis	14	N/A	N/A	N/A	0	N/A	N/A	N/A
Median Creatinine Level	0.85	N/A	N/A	N/A	0.75	N/A	N/A	N/A
Creatinine Level Range	0.6-2.1	N/A	N/A	N/A	0.6-1.0	N/A	N/A	N/A
Median Proteinuria	0	N/A	N/A	N/A	0	N/A	N/A	N/A
Proteinuria Range	0-2	N/A	N/A	N/A	0	N/A	N/A	N/A
Median Glucose	87	N/A	N/A	N/A	89	N/A	N/A	N/A
Glucose Range	50-312	N/A	N/A	N/A	81-103	N/A	N/A	N/A
Median Prednisone Dose	5	N/A	N/A	N/A	0	N/A	N/A	N/A
Prednisone Dose Range	0-40	N/A	N/A	N/A	0-10	N/A	N/A	N/A
Number of Patients on Prednisone	13	N/A	N/A	N/A	5	N/A	N/A	N/A

Table B.2 Individual discovery sample demographic and clinical information

Age	Gender	Ethnicity	SLE	Smoking	SLEDAI	Flare	Neph. Stage	Nephritis	Creatine	Proteinuria	BMI	Glucose	Pred. Dose	Prednisone	
SLE10643	24	F	EA	S	N	4	NA	0	N	0.7	0	36.22	103	0	0
SLE10797	24	F	AA	S	Y	12	NA	4	Y	1	2	19.89	84	0	0
SLE11227	35	F	AA	S	N	1	NA	0	N	0.7	0	31.27	82	0	0
SLE11439	24	F	AA	S	N	10	NA	4	Y	2.1	2	26.6	80	40	1
SLE11787	49	F	AA	S	N	3	NA	0	N	0.9	0	24.04	89	15	1
SLE11844	33	F	AA	S	N	4	NA	0	N	0.7	1	28.9	95	10	1
SLE1189	47	F	AA	S	N	0	NA	0	N	1.4	0	27.4	69	0	0
SLE11894	33	F	AA	S	N	0	NA	5	Y	0.8	0	29.87	68	0	0
SLE11897	22	F	EA	S	N	4	NA	0	N	0.7	0	25.4	81	10	1
SLE11906	22	F	EA	S	N	0	NA	0	N	0.7	0	34.77	90	0	0
SLE11913	45	F	EA	S	N	0	NA	0	N	0.6	0	25.9	89	0	0
SLE11929	39	F	AA	S	N	4	NA	0	N	0.8	0	34.33	144	4	1
SLE11933	31	F	AA	S	N	5	NA	3.5	Y	0.6	1	23.04	106	5	1
SLE11940	22	F	AA	S	N	3	NA	4	Y	0.7	0	22.41	81	20	1
SLE11946	41	F	AA	S	N	0	NA	5	Y	0.9	1	28.45	77	0	0
SLE1296	44	F	AA	S	N	0	NA	4	Y	1.8	2	42.93	96	5	1
SLE1796	28	F	AA	S	N	2	NA	0	N	0.9	0	30.35	100	0	0
SLE11947	22	F	AA	S	N	0	NA	4	Y	0.9	2	27.64	65	10	1
SLE4091	43	F	AA	S	N	0	NA	0	N	0.7	0	17.97	84	5	1
SLE4207	28	F	EA	S	N	0	NA	0	N	NA	NA	44.27	86	0	0
SLE5276	23	F	AA	S	N	7	NA	3.5	Y	0.7	2	28.7	50	10	1
SLE3759	43	F	AA	S	N	0	NA	3	Y	NA	NA	28.62	94	7.5	1
SLE10357	38	F	EA	S	N	0	N	0	N	0.8	0	43.91	87	5	1
SLE1846	48	F	AA	S	N	0	N	5	Y	1.4	0	52.18	94	0	0
SLE11928	48	F	EA	S	N	0	NA	0	N	0.7	0	24.6	93	0	0
SLE11955	44	F	EA	S	N	4	Y	0	N	0.7	0	23.79	98	0	0
SLE11956	43	F	EA	S	N	0	NA	0	N	0.9	0	20.23	89	0	0
SLE11957	41	F	AA	S	N	3	NA	3.5	Y	0.9	1	27.8	87	0	0
SLE10558	44	F	AA	S	N	0	NA	2	Y	1	0	28.15	89	5	1
SLE88	49	F	EA	S	N	NA	NA	0	N	0.9	0	31.41	86	5	1
SLE1182	48	F	AA	S	N	2	N	0	N	0.8	0	32.21	100	0	0
SLE11971	28	F	AA	S	N	2	NA	2	Y	0.8	NA	23	84	20	1
SLE11335	38	F	EA	S	N	NA	NA	0	N	0.8	0	26.26	98	0	0
SLE11453	37	F	EA	S	N	4	NA	0	N	0.9	0	30.11	87	10	1
SLE11972	48	F	EA	S	N	2	Y	0	N	0.6	0	34.56	103	0	0
SLE11977	36	F	EA	S	N	0	N	0	N	1	0	30.11	89	0	0
SLE11978	42	F	AA	S	N	0	Y	0	N	0.7	0	33.5	312	0	0
SLE11979	37	F	EA	S	N	4	NA	0	N	0.9	0	30.6	82	10	1
CNTL119	47.79	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL1241	47.71	F	EA	C	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL1246	47.13	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL4010	45.46	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL5170	37.37	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11369	32.21	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11538	40.13	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11542	36.47	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11560	23.96	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11568	24.96	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11717	48.13	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11875	36.29	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11891	39.13	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11896	39.71	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11925	24.13	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11927	24.46	F	EA	C	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11943	32.21	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11944	36.47	F	EA	C	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11945	26.13	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11948	28.71	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11949	41.96	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11950	26.55	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11951	45.88	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11952	29.38	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11961	30.05	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11962	26.96	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11963	44.96	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11966	41.79	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11970	44.3	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11973	33.38	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11975	39.05	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11976	37.38	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL120	44.3	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL139	34.8	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL4230	38.63	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL4791	38.38	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11571	37.13	F	EA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CNTL11954	32.38	F	AA	C	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SLE11539	54.47	F	EA	S	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SLE11888	40.21	F	EA	S	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SLE11862	38.29	F	EA	S	N	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
SLE11585	29.46	F	AA	S	Y	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table B.3 Individual validation sample demographic and clinical data.

SUBID	AGE	GENDER	ETH	SLE	FLARE	CD4	CD19
SLE11414	29	F	AA	S	Post	1	1
SLE5274	24	F	AA	S	Post	1	1
SLE10495	26	F	AA	S	Post	1	1
SLE10024	30	F	EA	S	Post	1	1
SLE1190	45	F	AA	S	Post	1	1
CNTL11423	28	F	EA	C	Control	1	1
CNTL10414	26	F	AA	C	Control	1	1
CNTL10412	33	F	AA	C	Control	1	1
CNTL3968	32	F	AA	C	Control	1	1
CNTL4122	21	F	EA	C	Control	1	1
SLE1167	28	F	AA	S	Flare	1	1
SLE11440	45	F	EA	S	Post	1	1
SLE4090	25	F	AA	S	Post	1	1
SLE11445	19	F	AA	S	Post	1	1
CNTL10897	27	F	AA	C	Control	1	1
CNTL10450	28	F	AA	C	Control	1	1
CNTL10576	33	F	AA	C	Control	1	1
SLE10494	31	F	EA	S	Post	1	1
SLE92	33	F	AA	S	Post	1	1
SLE4046	49	F	AA	S	Flare	1	1
SLE11476	46	F	AA	S	Post	1	1
CNTL4665	26	F	AA	C	Control	1	1
CNTL10413	25	F	AA	C	Control	1	1
SLE11473	41	F	AA	S	Post	1	1
SLE4048	48	F	AA	S	Post	1	1
SLE10374	20	F	AA	S	Flare	1	1
SLE1627	41	F	AA	S	Flare	1	1
CNTL10943	48	F	AA	C	Control	1	1
CNTL11394	21	F	EA	C	Control	1	1
CNTL4184	34	F	AA	C	Control	1	1
SLE10347	23	F	AA	S	Post	1	1
SLE11548	44	F	EA	S	Flare	1	1
SLE10593	37	F	EA	S	Flare	1	1
CNTL11432	25	F	AA	C	Control	1	1
CNTL11533	28	F	EA	C	Control	1	1
CNTL11577	25	F	AA	C	Control	1	1
CNTL11573	22	F	EA	C	Control	1	1
SLE10011	22	F	AA	S	Post	1	1
SLE5424	26	F	AA	S	Post	1	1
SLE1121	32	F	AA	S	Flare	1	0
SLE11586	25	F	EA	S	Flare	1	1
SLE3667	30	F	AA	S	Post	1	1
CNTL11558	32	F	AA	C	Control	1	1
CNTL11559	23	F	EA	C	Control	1	1
CNTL11578	34	F	AA	C	Control	1	1
CNTL11451	24	F	EA	C	Control	1	1
CNTL11575	22	F	AA	C	Control	1	1
SLE5422	22	F	AA	S	Flare	1	0
SLE1072	47	F	EA	S	Post	1	1
SLE11530	37	F	EA	S	Flare	1	1
CNTL11550	25	F	EA	C	Control	1	1
CNTL11463	38	F	EA	C	Control	1	1
CNTL11556	44	F	EA	C	Control	1	1
CNTL4136	32	F	AA	C	Control	1	1
SLE3686	25	F	EA	S	Post	1	1
SLE2534	29	F	AA	S	Post	1	1
SLE11352	32	F	EA	S	Post	1	1
SLE11555	26	F	AA	S	Post	1	1
CNTL11570	25	F	AA	C	Control	1	1
CNTL11561	31	F	AA	C	Control	1	1
CNTL11471	21	F	EA	C	Control	1	1
CNTL11080	27	F	AA	C	Control	1	1
SLE11546	39	F	EA	S	Post	1	1
CNTL11301	22	F	AA	C	Control	1	1
CNTL11406	29	F	EA	C	Control	1	1
CNTL1501	41	F	AA	C	Control	1	1
CNTL4492	34	F	AA	C	Control	1	1
SLE4440	22	F	AA	S	Flare	1	1
CNTL1690	65	F	EA	C	Control	0	1

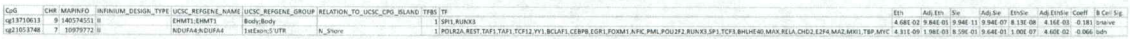
Table B.4 SLE Regression significant CpGs.

CpG	CHR	MAPINFO	INFINIUM_DESIGN_TYPE	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	TFBS	TF	Sle	Adj.Sle	Coeff	B Cell Sig. In
cg19725377	14	69535979	II	DCAF5	Body		0		0 1.76E-10	8.11E-05	0.078	btrans
cg27209729	11	64428925	II	NRXN2	Body;Body	S_Shore	0		0 1.75E-08	2.23E-03	-0.102	btrans
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	2.02E-08	2.23E-03	-0.211	btrans
cg21549285	21	42799141	II	MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	2.65E-08	2.23E-03	-0.365	btrans
cg22862003	21	42797588	II	MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	3.05E-08	2.23E-03	-0.219	btrans
cg06981309	3	146260954	II	PLSCR1	5'UTR	N_Shore	0		0 3.17E-08	2.23E-03	-0.269	btrans
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	3.52E-08	2.23E-03	-0.243	btrans
cg20029745	18	28621985	II	OSC3	Body;Body	Island	1	EZH2	4.66E-08	2.23E-03	0.041	btrans
cg05552874	10	91153143	II	IFIT1	Body		1	RUNX3,MAX,STAT3	4.78E-08	2.23E-03	-0.217	btrans
cg22764925	22	24979964	II	GGT1	5'UTR		0		0 4.85E-08	2.23E-03	-0.098	btrans
cg01948202	3	122400474	I	PARP14	Body	S_Shore	1	POLR2A,TAF1,YY1,STAT1	6.04E-08	2.53E-03	-0.101	btrans
cg14864167	8	66751182	II	PDE7A	Body	N_Shelf	0		0 8.34E-08	3.20E-03	-0.197	btrans
cg17052170	8	144099482	I	LY6L	TSS1500;TSS1500;Body	Island	1	POLR2A	9.97E-08	3.53E-03	-0.128	btrans
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	2.31E-10	6.45E-05	-0.253	bnaive
cg03038262	11	315262	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	2.80E-10	6.45E-05	-0.281	bnaive
cg22930808	3	122281881	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	N_Shore	0		0 5.30E-10	7.01E-05	-0.287	bnaive
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	6.72E-10	7.01E-05	-0.224	bnaive
cg01948202	3	122400474	I	PARP14	Body	S_Shore	1	POLR2A,TAF1,YY1,STAT1	7.62E-10	7.01E-05	-0.102	bnaive
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	1.15E-09	8.84E-05	-0.262	bnaive
cg21549285	21	42799141	II	MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	1.40E-09	9.18E-05	-0.353	bnaive
cg22862003	21	42797588	II	MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	1.94E-09	1.08E-04	-0.221	bnaive
cg01028142	2	7004578	II	CMPK2	Body	N_Shore	1	EBF1	2.11E-09	1.08E-04	-0.18	bnaive
cg01971407	11	313624	II	IFITM1	TSS1500	N_Shelf	1	CHD1,POLR2A,ELF1,SIX5,EBF1,EBF1,PML,MAX,RELA,CHD2,MAZ,MXI1,NFYB,TBP,WRNIP1	2.51E-09	1.12E-04	-0.127	bnaive
cg21686213	11	315118	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	2.77E-09	1.12E-04	-0.189	bnaive
cg10549986	2	7018153	I	RSAD2	1stExon		0		0 2.92E-09	1.12E-04	-0.193	bnaive
cg14864167	8	66751182	II	PDE7A	Body	N_Shelf	0		0 3.26E-09	1.15E-04	-0.23	bnaive
cg18686270	3	146258875	II	PLSCR1	5'UTR	N_Shelf	0		0 4.43E-09	1.45E-04	-0.166	bnaive
cg08122652	3	122281939	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	N_Shore	0		0 4.97E-09	1.52E-04	-0.253	bnaive
cg11842367	1	78703580	II	MIGC27382	Body		0		0 5.70E-09	1.64E-04	0.129	bnaive
cg09410045	3	150997688	II	MED12L	Body;TSS1500		0		0 6.13E-09	1.66E-04	0.114	bnaive
cg13304609	1	79085162	II	IFI44L	TSS1500		1	TAF1	6.92E-09	1.77E-04	-0.143	bnaive
cg01190666	20	62204908	II	PRIC285	5'UTR	N_Shore	1	CTCF	1.31E-08	3.18E-04	-0.088	bnaive
cg16626047	1	949893	II	ISG15	3'UTR	S_Shore	1	POLR2A,TAF1,YY1	1.95E-08	4.42E-04	-0.133	bnaive
cg05552874	10	91153143	II	IFIT1	Body		1	RUNX3,MAX,STAT3	2.02E-08	4.42E-04	-0.212	bnaive
cg07285983	1	174844490	II	RABGAP1L	Body;TSS200		1	EBF1,RUNX3,ZEB1	2.49E-08	5.21E-04	-0.186	bnaive
cg18749404	17	76975944	II	LGALS3BP	1stExon;5'UTR		1	ELF1,EBF1	2.80E-08	5.37E-04	-0.062	bnaive
cg04788999	1	949850	II	ISG15	Body	Island	1	POLR2A,TAF1,YY1	2.80E-08	5.37E-04	-0.124	bnaive
cg17990365	11	319718	II	IFITM3	3'UTR	S_Shore	0		0 3.38E-08	6.22E-04	-0.14	bnaive
cg20062691	1	949392	II	ISG15	Body	Island	1	POLR2A	3.65E-08	6.33E-04	-0.114	bnaive
cg20045320	11	319555	I			S_Shore	0		0 3.72E-08	6.33E-04	-0.139	bnaive
cg09026253	11	313267	II	IFITM1	TSS1500	S_Shore	1	CHD1,POLR2A,ELF1,EBF1,MAX,CHD2,MAZ,MXI1,WRNIP1	5.28E-08	8.67E-04	-0.167	bnaive
cg14870271	17	76976010	II	LGALS3BP	1stExon;5'UTR		1	ELF1,EBF1	5.56E-08	8.68E-04	-0.078	bnaive
cg14392283	8	144103587	I	LY6L	3'UTR;3'UTR	N_Shelf	1	POLR2A,LLF1,USF1,YY1,EBF1,LGR1,BHLHE40	5.66E-08	8.68E-04	-0.121	bnaive
cg01079652	1	791181951	II	IFI44	Body		0		0 6.17E-08	9.16E-04	-0.158	bnaive
cg03714676	8	48174649	II	KIAA0146	Body	S_Shore	0		0 6.46E-08	9.29E-04	-0.084	bnaive
cg26312951	21	42797847	II	MX1	TSS200;5'UTR	N_Shore	1	CHD1,POLR2A,ELF1,RLS1,TAF1,TCF12,ZBTB33,FOXM1,PAX5,PML,POU2F2,TCF3,RELA,CHD2,MAZ,MXI1,SIN3A,TBL1XR1,TBP	8.33E-08	1.15E-03	-0.137	bnaive
cg04582010	11	313120	II	IFITM1	TSS1500	S_Shore	1	CHD1,POLR2A,ELF1,MAX,CHD2,MAZ,MXI1	8.48E-08	1.15E-03	-0.15	bnaive
cg01028142	2	7004578	II	CMPK2	Body	N_Shore	1	EBF1	1.32E-10	4.64E-05	-0.112	bswitch
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	2.02E-10	4.64E-05	-0.189	bswitch
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	5.52E-10	8.47E-05	-0.203	bswitch
cg09843907	13	43563347	II	EPSTI1	Body;Body	N_Shelf	0		0 5.53E-09	5.86E-04	-0.134	bswitch
cg03038262	11	315262	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	6.37E-09	5.86E-04	-0.144	bswitch
cg01274660	7	100465625	II	TRIP6	Body	S_Shore	1	POLR2A,ZEB1	1.09E-08	8.35E-04	-0.077	bswitch
cg05696877	1	79088769	II	IFI44L	5'UTR		1	POLR2A,TBP,WRNIP1	1.60E-08	1.05E-03	-0.163	bswitch
cg21686213	11	315118	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	2.26E-08	1.30E-03	-0.19	bswitch
cg18580296	7	15726411	II	MEOX2	TSS200		0		0 2.61E-08	1.33E-03	-0.058	bswitch
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	3.69E-08	1.55E-03	-0.088	bswitch
cg12906975	8	144105259	II			N_Shore	1	POLR2A,EBF1,RELA	3.71E-08	1.55E-03	0.064	bswitch
cg06981309	3	146260954	II	PLSCR1	5'UTR	N_Shore	0		0 4.82E-08	1.72E-03	-0.127	bswitch
cg00676801	2	191876673	II	STAT1	5'UTR;5'UTR	N_Shore	1	POLR2A	4.85E-08	1.72E-03	-0.096	bswitch
cg12243034	17	40701872	II			S_Shore	0		0 6.74E-08	2.20E-03	-0.068	bswitch
cg18991321	7	100144475	II	AGFG2	Body		1	POLR2A,TAF1,BATF,EGR1,NFATC1,NFIC,SPI1,RUNX3,STAT5A,RELA,KLF3,MAZ,SMC3,STAT1,TBP,ZNF143	7.17E-08	2.20E-03	-0.093	bswitch
cg12047941	11	315908	I			Island	1	POLR2A	9.63E-08	2.53E-03	-0.07	bswitch
cg08090640	17	41159289	II	IFI35	Body		1	POLR2A,PML,RUNX3,RELA	1.05E-07	2.53E-03	-0.08	bswitch
cg08122652	3	122281939	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	N_Shore	0		0 1.06E-07	2.53E-03	-0.091	bswitch
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	1.09E-12	4.19E-07	-0.336	bdn
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	1.82E-12	4.19E-07	-0.343	bdn
cg03038262	11	315262	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	7.73E-12	9.98E-07	-0.294	bdn
cg21549285	21	42799141	II	MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	8.67E-12	9.98E-07	-0.353	bdn
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	1.54E-11	1.29E-06	-0.243	bdn
cg22862003	21	42797588	II	MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	1.69E-11	1.29E-06	-0.267	bdn
cg00474696	2	224821113	II	MRPL44	TSS1500	N_Shore	1	BATF,BCL11A,IRF4,RUNX3	2.40E-11	1.58E-06	-0.216	bdn
cg21686213	11	315118	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	1.02E-10	5.89E-06	-0.337	bdn
cg08122652	3	122281939	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	N_Shore	0		0 1.74E-10	8.91E-06	-0.242	bdn
cg06872964	1	79085250	II	IFI44L	TSS1500		1	TAF1	2.08E-10	9.56E-06	-0.25	bdn
cg01028142	2	7004578	II	CMPK2	Body	N_Shore	1	EBF1	3.22E-10	1.35E-05	-0.221	bdn
cg09026253	11	313267	II	IFITM1	TSS1500	S_Shore	1	CHD1,POLR2A,ELF1,EBF1,MAX,CHD2,MAZ,MXI1,WRNIP1	6.69E-10	2.32E-05	-0.232	bdn
cg07839457	16	57023022	II	NLRC5	TSS1500	N_Shore	1	POLR2A,LLF1,TAF1,USF1,YY1,ATF2,EBF1,FOXM1,TCF3,BHLHE40,MAX,MAZ,MXI1,SIN3A,SMC3,USF2,MYC	6.98E-10	2.32E-05	-0.247	bdn
cg05552874	10	91153143	II	IFIT1	Body		1	RUNX3,MAX,STAT3	7.05E-10	2.32E-05	-0.288	bdn
cg06981309	3	146260954	II	PLSCR1	5'UTR	N_Shore	0		0 1.18E-09	3.63E-05	-0.235	bdn
cg07285983	1	174844490	II	RABGAP1L	Body;TSS200		1	EBF1,RUNX3,ZEB1	1.47E-09	4.22E-05	-0.255	bdn
cg10549986	2	7018153	I	RSAD2	1stExon		0		0 1.80E-09	4.87E-05	-0.239	bdn
cg01971407	11	313624	II	IFITM1	TSS1500	N_Shelf	1	CHD1,POLR2A,ELF1,SIX5,EBF1,EBF1,PML,MAX,RELA,CHD2,MAZ,MXI1,NFYB,TBP,WRNIP1	3.13E-09	8.00E-05	-0.148	bdn
cg03607951	1	79085586	II	IFI44L	TSS1500		1	POLR2A,TAF1,ATF2,MTA3,NFIC,PML,RELA,STAT3	4.33E-09	1.05E-04	-0.275	bdn
cg14392283	8	144103587	I	LY6L	3'UTR;3'UTR	N_Shelf	1	POLR2A,ELF1,USF1,YY1,EBF1,LGR1,BHLHE40	5.78E-09	1.33E-04	-0.299	bdn
cg13130398	1	174844397	II	RABGAP1L	TSS1500;Body		1	EBF1,RUNX3,ZEB1,IKZF1	8.96E-09	1.96E-04	-0.202	bdn
cg01886988	11	312560	II	IFITM1	TSS1500	S_Shore	1	SPI1	9.83E-09	2.06E-04	-0.164	bdn
cg22930808	3	122281881	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	N_Shore	0		0 2.78E-08	5.56E-04	-0.281	bdn
cg25050332	11	316088	I			Island	1	POLR2A	3.24E-08	6.22E-04	-0.17	bdn
cg13304609	1	79085162	II	IFI44L	TSS1500		1	TAF1	3.86E-08	6.74E-04	-0.187	bdn
cg18686270	3	146258875	II	PLSCR1	5'UTR	N_Shelf	0		0 5.20E-08	9.08E-04	0.16	bdn
cg01190666	20	62204908	II	PRIC285	5'UTR	N_Shore	1	CTCF	5.33E-08	9.08E-04	-0.109	bdn
cg04696877	1	79088769	II	IFI44L	5'UTR		1	POLR2A,TBP,WRNIP1	6.88E-08	1.09E-03	-0.236	bdn
cg16945410	11	313354	II	IFITM1	TSS1500	S_Shore	1	CHD1,POLR2A,ELF1,EBF1,MAX,CHD2,MAZ,MXI1,WRNIP1	6.89E-08	1.09E-03	-0.147	bdn
cg17980508	1	79085713	II	IFI44L	TSS1500		1	POLR2A,TAF1,ATF2,MTA3,NFIC,PML,RUNX3,RELA,STAT1,STAT3	9.01E-08	1.38E-03	-0.2	bdn
cg02215171	4	89379156	II	HERC5	Body	S_Shore	1	POLR2A	9.58E-08	1.42E-03	-0.133	bdn

Table B.5 Significant AA-specific SLE-associated CpGs.

CpG	CHR	MAPINFO	INFINIUM_DESIGN_TYPE	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	TFBS	TF	Eth	Adj.Eth	Sle	Adj.Sle	EthSle	Adj.EthSle	Coeff	B Cell Sig.
cg17980508	1	79085713	II	IFI44L	TSS1500		1	POLR2A,TAF1,ATF2,MTA3,NFIC,PML,RUNX3,RELA,STAT1,STAT3	1.93E-01	1.00E+00	1.66E-13	9.56E-09	1.12E-09	5.14E-04	0.245	btrans
cg22862003	21	42797588	II	MX1;MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	2.14E-01	1.00E+00	2.99E-15	6.95E-10	1.44E-08	2.23E-03	0.36	btrans
cg21549285	21	42799141	II	MX1;MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	1.75E-01	1.00E+00	3.02E-15	6.95E-10	1.67E-08	2.23E-03	0.595	btrans
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	1.46E-01	1.00E+00	8.17E-14	6.26E-09	2.34E-08	2.23E-03	0.394	btrans
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	3.19E-01	1.00E+00	5.96E-15	9.14E-10	2.51E-08	2.23E-03	0.396	btrans
cg05552874	10	91153143	II	IFIT1	Body		1	RUNX3,MAX,STAT3	7.34E-01	1.00E+00	9.26E-15	9.71E-10	2.91E-08	2.23E-03	0.357	btrans
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	3.16E-01	1.00E+00	1.06E-14	9.71E-10	7.60E-08	4.99E-03	0.328	btrans
cg21549285	21	42799141	II	MX1;MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	8.96E-01	9.97E-01	1.17E-17	5.40E-12	1.11E-09	5.10E-04	0.569	bnaive
cg22862003	21	42797588	II	MX1;MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	4.61E-01	9.85E-01	4.87E-17	6.33E-12	3.36E-09	7.74E-04	0.352	bnaive
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	9.56E-01	9.99E-01	4.74E-17	6.33E-12	5.33E-09	8.18E-04	0.406	bnaive
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	4.31E-01	9.84E-01	5.50E-17	6.33E-12	2.66E-08	1.61E-03	0.358	bnaive
cg10549986	2	7018153	I	RSAD2	1stExon		0		0 2.80E-01	9.84E-01	3.42E-16	2.62E-11	1.60E-08	1.50E-03	0.302	bnaive
cg07285983	1	174844490	II	RABGAP1L;RABGAP1L	Body;TSS200		1	EBF1,RUNX3,ZEB1	3.65E-01	9.84E-01	3.79E-15	1.75E-10	2.31E-08	1.61E-03	0.314	bnaive
cg06981309	3	146260954	II	PLSCR1	5'UTR	N_Shore	0		0 4.42E-01	9.84E-01	2.10E-14	8.04E-10	2.80E-08	1.61E-03	0.436	bnaive
cg23213327	2	7016509	II	RSAD2	TSS1500		0		0 2.38E-01	9.84E-01	2.22E-13	4.63E-09	1.64E-08	1.50E-03	0.223	bnaive
cg22862003	21	42797588	II	MX1;MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	1.60E-01	1.00E+00	1.53E-12	7.04E-07	1.32E-08	6.06E-03	0.221	bunswitch
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	4.30E-01	1.00E+00	4.81E-12	1.11E-06	6.54E-08	1.50E-02	0.321	bunswitch

Table B.6 Significant EA-specific SLE-associated CpGs.



CpG	CHR	MAPINFO	INFINIUM_DESIGN_TYPE	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	TFBS	TF	Eth	Adj.Eth	Six	Adj.Six	EthSix	Adj.EthSix	Coeff	B Cell Sig
cg13710613	9	140574551	II	EHMT1;EHMT1	Body;Body		1	SP1,RUNX3	4.60E-02	9.84E-01	9.94E-11	9.94E-07	8.13E-08	4.16E-03	-0.181	Inactive
cg21053748	7	10979772	II	NDUFA4;NDUFA4	1stExon;5'UTR	N_Shore	1	POLR2A,REST,TAF1,TAF1,TCF12,YY1,BCLAF1,CEBPB,EGR1,FOXM1,AFIC,PML,POU2F2,RUNX3,SP1,TCF3,BHLHE40,MAX,RELA,CHD2,E2F4,MAZ,MXI1,TBP,MYC	4.31E-09	1.98E-03	8.59E-01	9.64E-01	1.00E-07	4.60E-02	-0.066	Indn

Table B.7 Model validation comparison results.

	RF Regression	RF Variance	LASSO Regression	LASSO Variance	Ridge Regression	Ridge Variance
Cut-off	P ² 1E-4	Variance ³ 4E-2	P ² 5E-2	Variance ³ 6E-2	P ² 1E-3	Variance ³ 7E-2
M Try	5	15	NA	NA	NA	NA
MSE	0.02	0.18	0.1	0.13	0.1	0.2
Lambda	NA	NA	0.09	0.23	7.85	5.79
Early Misclassification	0.03	0.06	0.03	0.03	0.04	0.05
Late Misclassification	0.07	0.19	0.05	0.07	0.05	0.46
Transitional Misclassification	0.07	0.07	0.77	0.07	0	0
Na•ve Misclassification	0.02	0.1	0.02	2	0.05	0.07
Un-switched Misclassification	0.03	0.26	0.03	0.07	0.03	0.48
Switched Misclassification	0.03	0.22	0.08	0.14	0.06	0.53
Double Neg. Misclassification	0.06	0.14	0.03	0	0.06	0.36
Discovery AUC	0.92	0.96	0.88	0.88	0.96	0.92
CD4 AUC	0.97	0.93	0.98	0.98	0.98	0.7
CD19 AUC	0.95	0.93	0.97	0.97	0.97	0.85

Table B.8 Top AA SLE Random Forest predictors (based on Gini Importance).

CpG	CHR	MAPINFO	INFINIUM_DESIGN_TYPE	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	TFBS	TF	Importance
cg07839457	16	57023022	II	NLRC5	TSS1500	N_Shore	1	POLR2A,ELF1,TAF1,USF1,YY1,ATF2,EBF1,FOXM1,TCF3,BHLHE40,MAX,MAZ,MXI1,SIN3A,SMC3,USF2,MYC	0.461
cg05552874	10	91153143	II	IFIT1	Body		1	RUNX3,MAX,STAT3	0.453
cg21549285	21	42799141	II	MX1	5'UTR;5'UTR	S_Shore	1	POLR2A,TAF1,YY1,PAX5,POU2F2	0.374
cg22862003	21	42797588	II	MX1	TSS1500;5'UTR	N_Shore	1	POLR2A,CHD2	0.363
cg05696877	1	79088769	II	IFI44L	5'UTR		1	POLR2A,TBP,WRNIP1	0.356
cg22764925	22	24979964	II	OGT1	5'UTR		0	0	0.345
cg22930808	3	122281881	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	Shore	0	0	0.320
cg05883128	4	169239131	II	DDX60	5'UTR	N_Shore	1	CTCF,POLR2A,YY1,WRNIP1	0.320
cg23570810	11	315102	II	IFITM1	Body	N_Shore	1	POLR2A,POU2F2,WRNIP1	0.313
cg03038262	11	315262	II	IFITM1	3'UTR	N_Shore	1	POLR2A,POU2F2,WRNIP1	0.302
cg07285983	1	174844490	II	RABGAP1L	Body;TSS200		1	EBF1,RUNX3,ZEB1	0.299
cg14864167	8	66751182	II	PDE7A	Body	N_Shelf	0	0	0.247
cg06872964	1	79085250	II	IFI44L	TSS1500		1	TAF1	0.198
cg14293575	22	18635460	II	USP18	5'UTR	S_Shelf	1	RUNX3,MAZ	0.183
cg06981309	3	146260954	II	PLSCR1	5'UTR	N_Shore	0	0	0.171
cg01971407	11	313624	II	IFITM1	TSS1500	N_Shelf	1	CHD1,POLR2A,ELF1,SIX5,EBF1,EBF1,PML,MAX,RELA,CHD2,MAZ,MXI1,NFYB,TBP,WRNIP1	0.167
cg01028142	2	7004578	II	CMPK2	Body	N_Shore	1	EBF1	0.151
cg10959651	2	7018020	II	RSAD2	1stExon		0	0	0.132
cg10549986	2	7018153	I	RSAD2	1stExon		0	0	0.105
cg17990365	11	319718	II	IFITM3	3'UTR	S_Shore	0	0	0.100
cg09251764	17	6659070	II	XAF1	TSS200;TSS200		1	POLR2A,TAF1,YY1,ATF2,MTA3,RELA,TBP	0.087
cg16411857	16	57023191	II	NLRC5	TSS1500	Island	1	POLR2A,ELF1,TAF1,TCF12,USF1,YY1,ATF2,EBF1,FOXM1,PML,POU2F2,TCF3,BHLHE40,MAX,MAZ,MXI1,SIN3A,SMC3,TBL1XR1,USF2,M	0.084
cg08122652	3	122281939	II	PARP9	5'UTR;5'UTR;TSS1500;5'UTR;5'UTR;5'UTR;5'UTR	Shore	0	0	0.075
cg26312951	21	42797847	II	MX1	TSS200;5'UTR	N_Shore	1	CHD1,POLR2A,ELF1,REST,TAF1,TCF12,ZBTB33,FOXM1,PAX5,PML,POU2F2,TCF3,RELA,CHD2,MAZ,MXI1,SIN3A,TBL1XR1,TBP	0.057
cg03607951	1	79085586	II	IFI44L	TSS1500		1	POLR2A,TAF1,ATF2,MTA3,NFIC,PML,RELA,STAT3	0.028
cg00855901	1	79085765	II	IFI44L	TSS1500		1	POLR2A,TAF1,TAF1,ATF2,MTA3,NFIC,PML,RUNX3,RELA,STAT1,STAT3,TBP	0.027
cg09122035	11	319667	II			S_Shore	0	0	0.025
cg01190666	20	62204908	II	PRJC285	5'UTR	N_Shore	1	CTCF	0.017
cg06085204	17	6659164	II	XAF1	5'UTR;5'UTR;1stExon;1stExon		1	POLR2A,TAF1,YY1,ATF2,MTA3,RELA,TBP	0.006
cg11971423	17	46653711	II	HOXB4	3'UTR	N_Shore	1	POLR2A	0.003
cg10616795	5	76464212	I				0	0	0.001

Table B.9 Top EA SLE Random Forest predictors (based on Gini Importance).

CpG	CHR	MAPINFO	INFINIUM_DESIGN_TYPE	UCSC_REFGENE_NAME	UCSC_REFGENE_GROUP	RELATION_TO_UCSC_CPG_ISLAND	Importance
cg25334892	8	15668801	II				0.299
cg26794885	8	12908302	II				0.283
cg09128529	2	227586520	II				0.265
cg01791778	2	583901	II			S_Shore	0.23
cg26948823	16	1255116	II	CACNA1H;CACNA1H	Body;Body	S_Shore	0.178
cg09563102	2	47889	II			S_Shore	0.171
cg10504436	22	24180492	II	DERL3;DERL3;DERL3	Body;Body;Body	Island	0.17
cg18163909	1	1897959	I	KIAA1751	Body	S_Shore	0.161
cg07262519	2	162657172	II	SLC4A10	Body		0.16
cg02070740	1	146763914	II	CHD1L	Body		0.157
cg12029281	6	35461818	II	TEAD3	5'UTR	N_Shelf	0.139
cg20177522	20	62410437	II	ZBTB46	Body	S_Shelf	0.138
cg20595846	14	76597468	II			N_Shore	0.136
cg03841832	4	9981202	II	SLC2A9;SLC2A9	Body;Body		0.131
cg17251609	3	65489574	II	MAGI1;MAGI1;MAGI1	Body;Body;Body		0.103
cg00283887	7	157935557	II	PTPRN2;PTPRN2;PTPRN2	Body;Body;Body	N_Shelf	0.101
cg21930668	1	17592170	II	PADI3	Body		0.097
cg20388256	9	34588432	II	CNTFR;CNTFR	5'UTR;5'UTR	N_Shore	0.096
cg00102726	21	46897181	II	COL18A1;COL18A1;COL18A1	Body;Body;Body	N_Shore	0.094
cg04830191	1	212868520	II	BATF3	Body	N_Shelf	0.088
cg26108416	1	49226683	II	BEND5;AGBL4	Body;Body		0.087
cg12807588	1	53688861	II			S_Shelf	0.086
cg17880816	19	3789435	II	MATK	5'UTR	S_Shelf	0.084
cg16640358	2	239892057	II			S_Shelf	0.084
cg07591515	1	51810132	II	TTC39A	Body	N_Shore	0.083

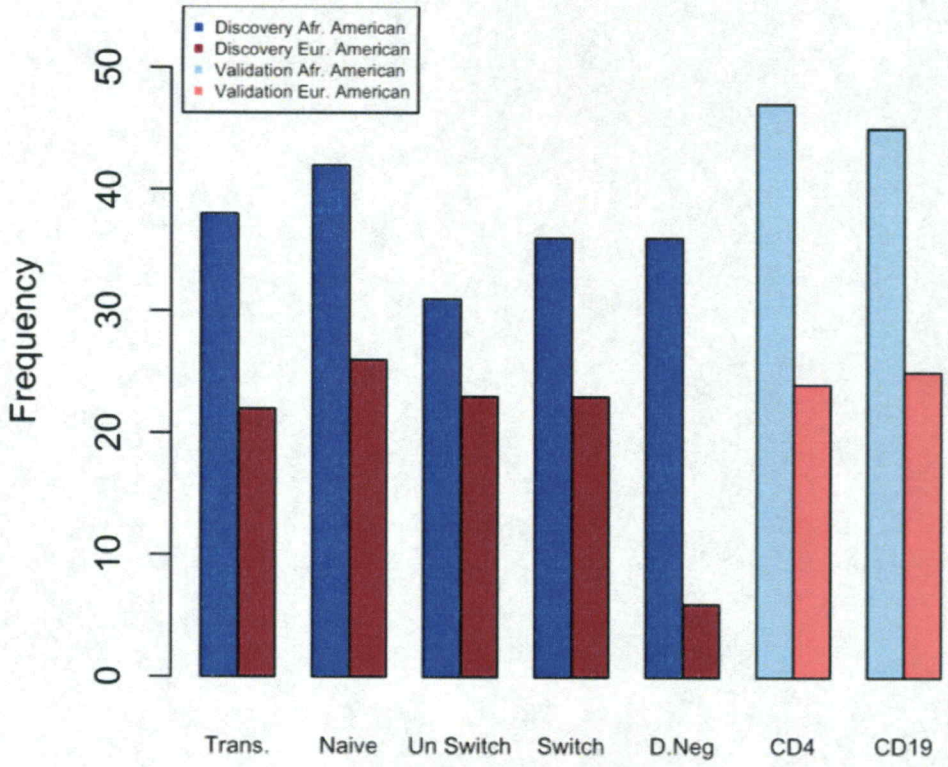


Figure B.1 Bar plot of cell coverage across discovery and validation cohorts.

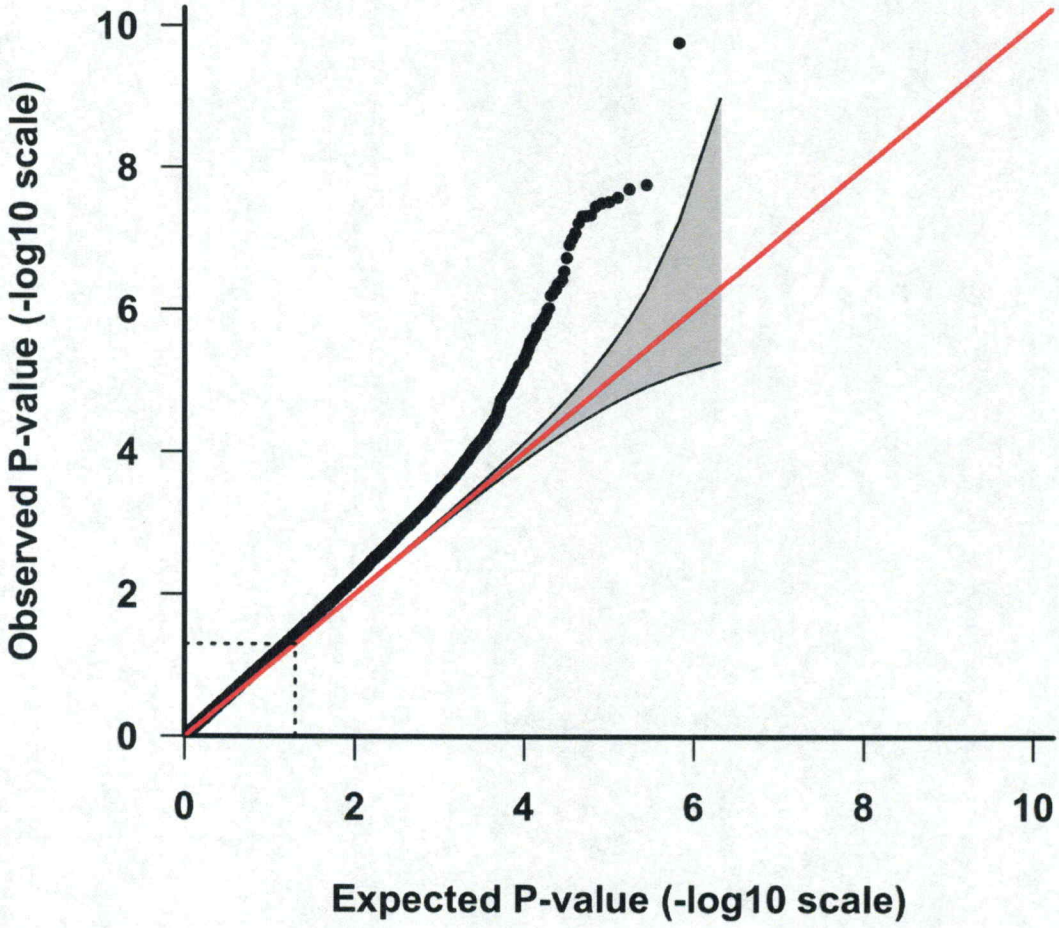


Figure B.2.1 Q-Q Plot of transitional B cell SLE regression p-values.

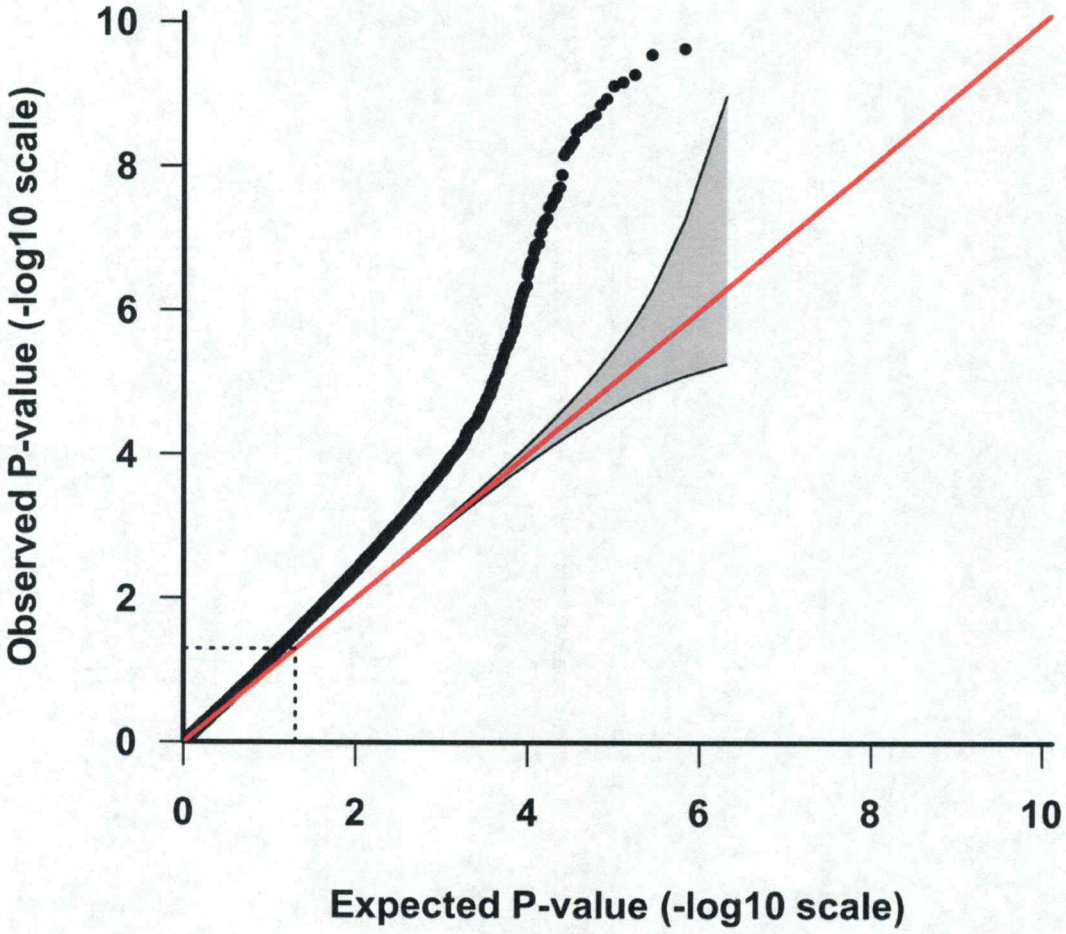


Figure B.2.2 Q-Q Plot of naive B cell SLE regression p-values.

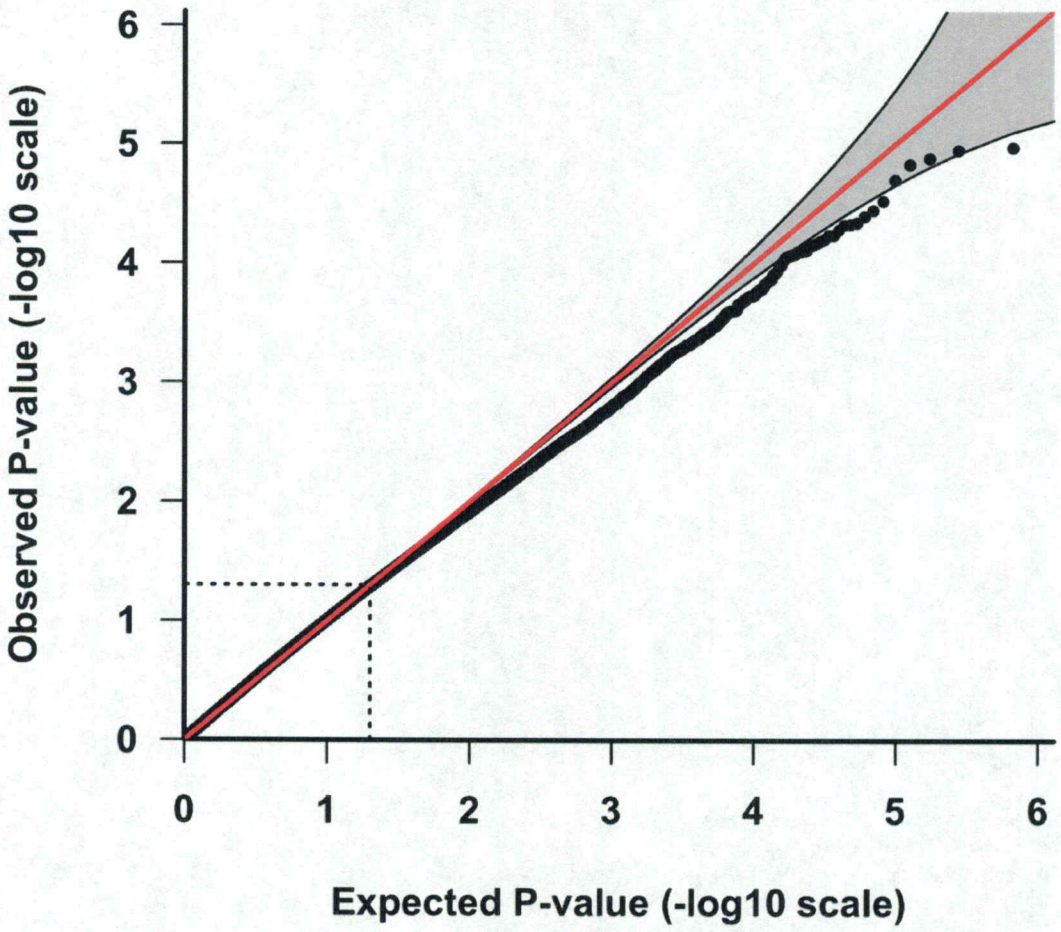


Figure B.2.3 Q-Q Plot of un-switched B cell SLE regression p-values.

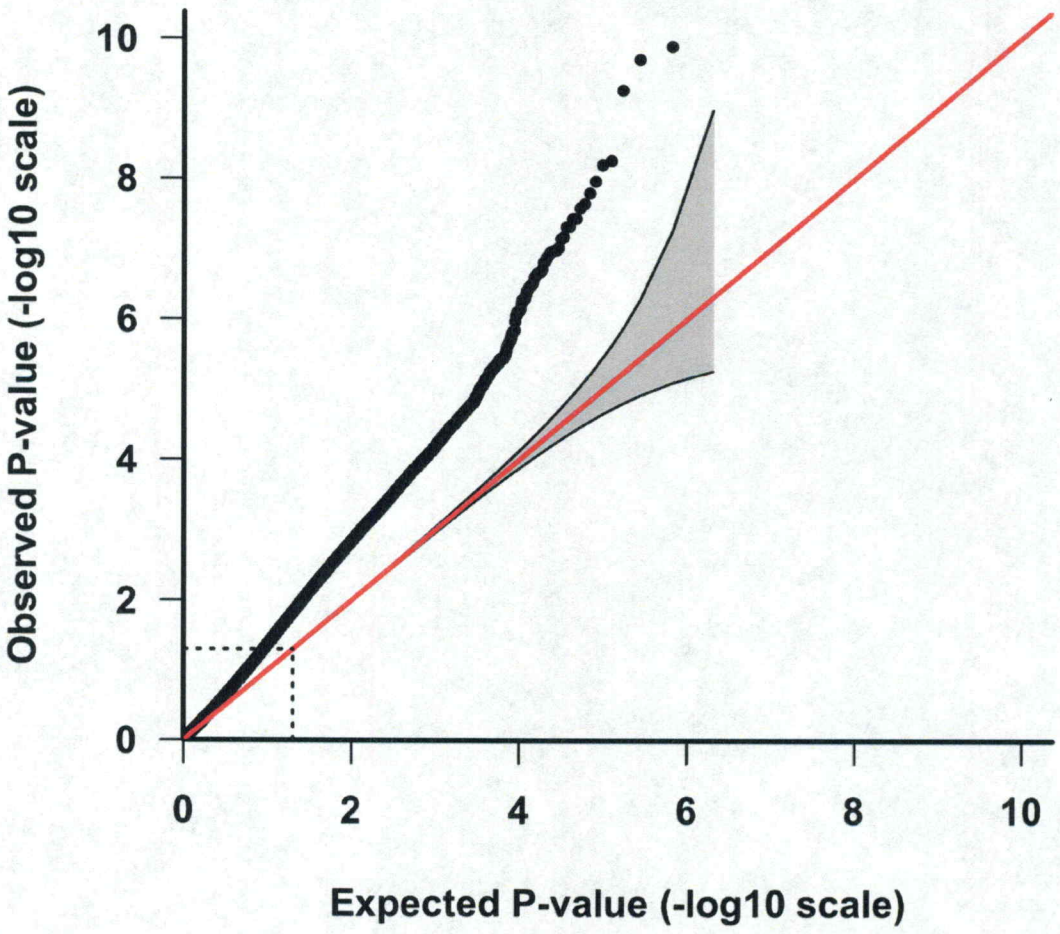


Figure B.2.4 Q-Q Plot of switched B cell SLE regression p-values.

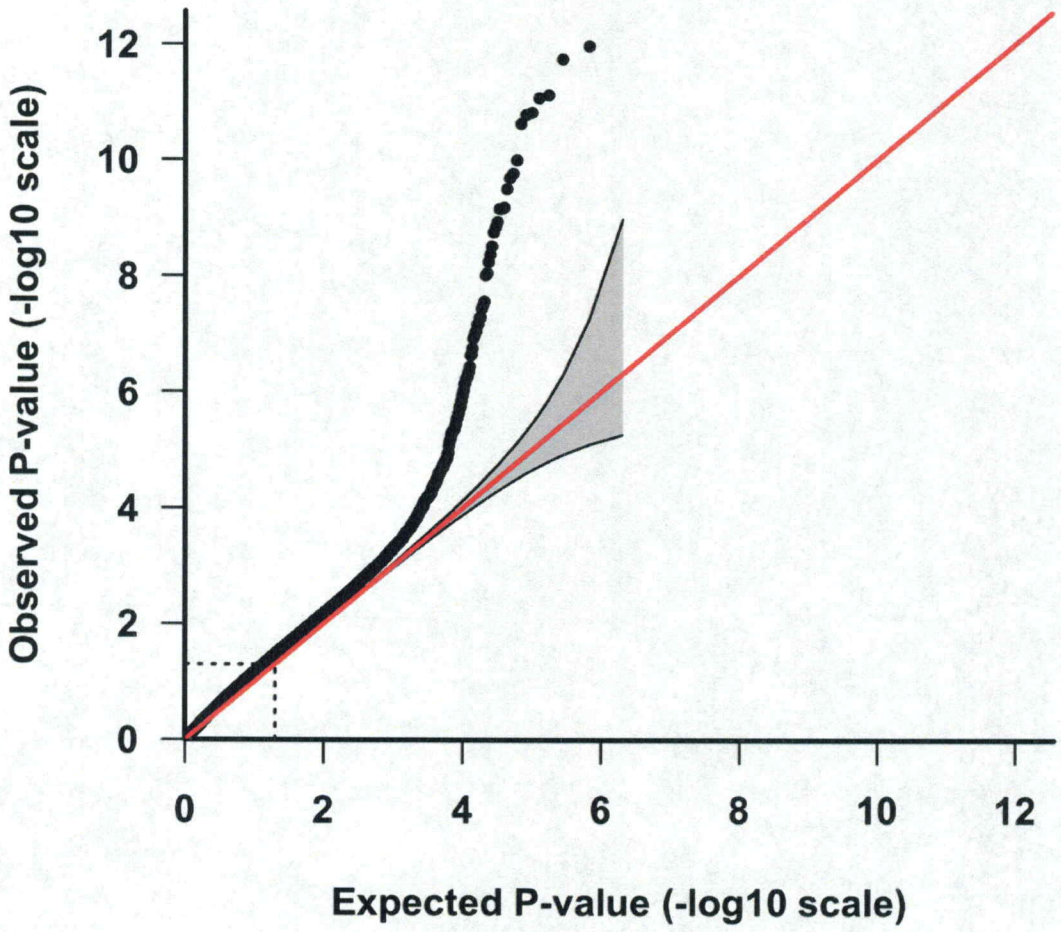

Observed P-value (-log10 scale)
Expected P-value (-log10 scale)
0
2
4
6
8
10
12

Figure B.2.5 Q-Q Plot of double negative B cell SLE regression p-values.

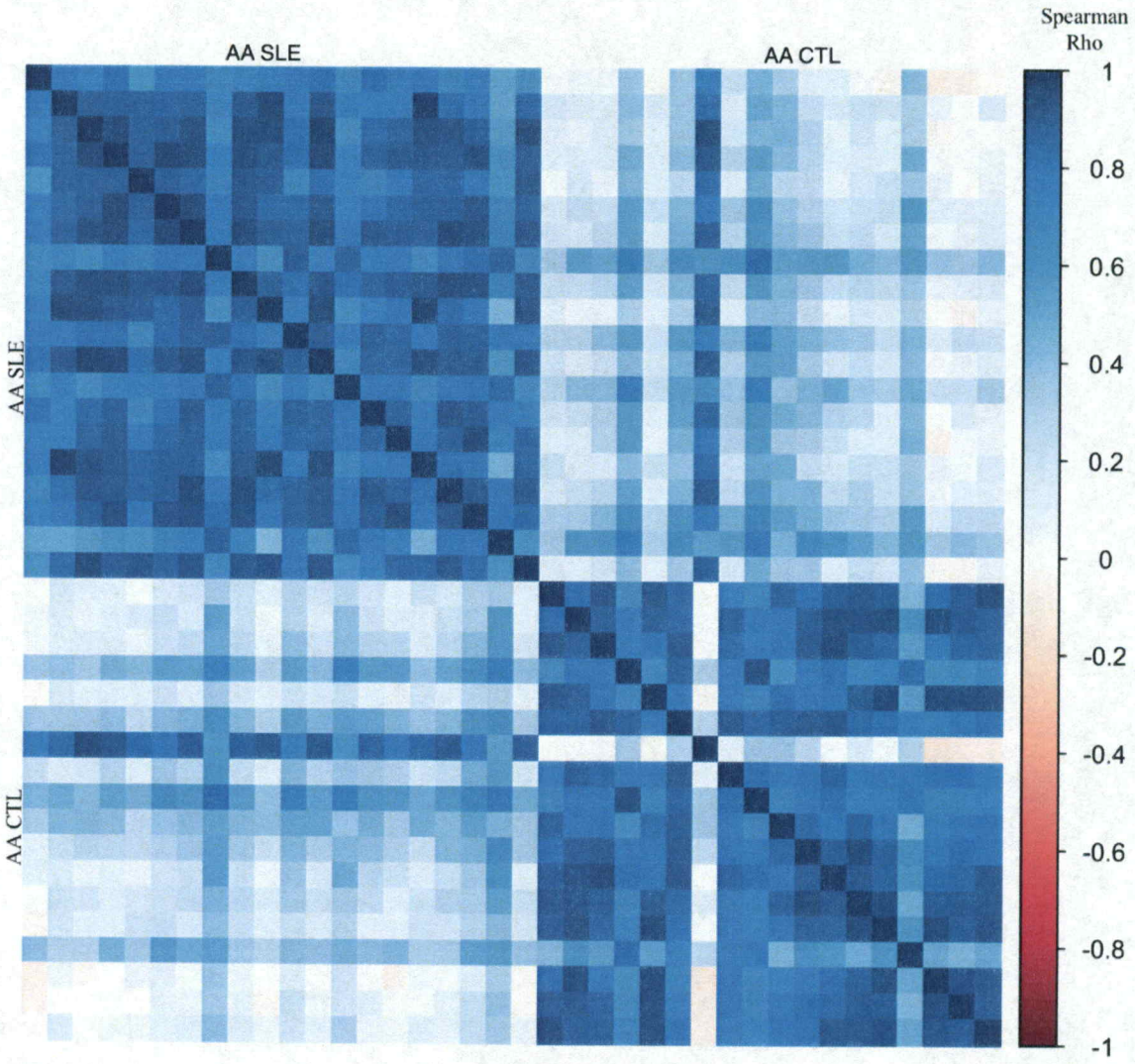


Figure B.3 AA transitional B cell top SLE-associated CpG methylation correlation.

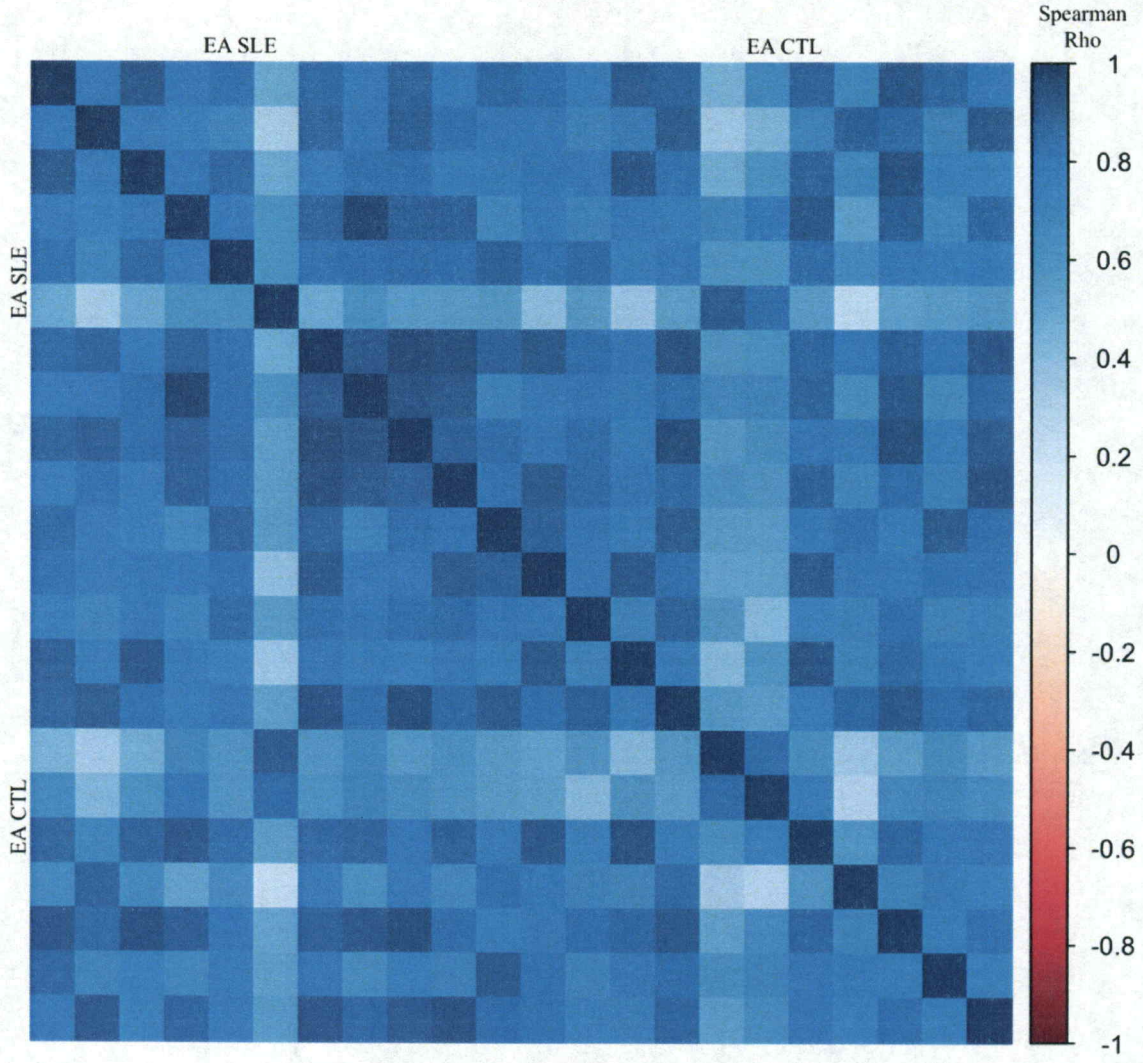


Figure B.4 EA transitional B cell top SLE-associated CpG methylation correlation.

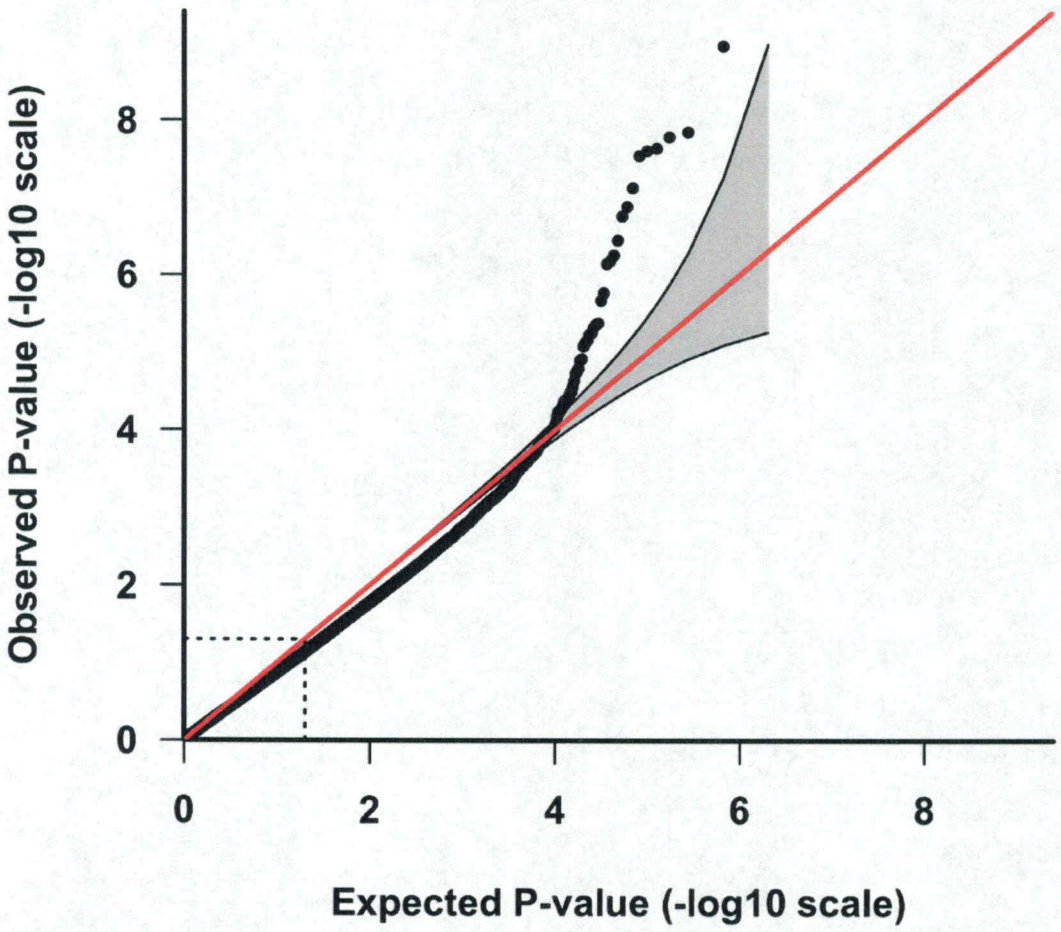


Figure B.5.1 Q-Q Plot of transitional B cell ethnicity-specific SLE regression p-values.

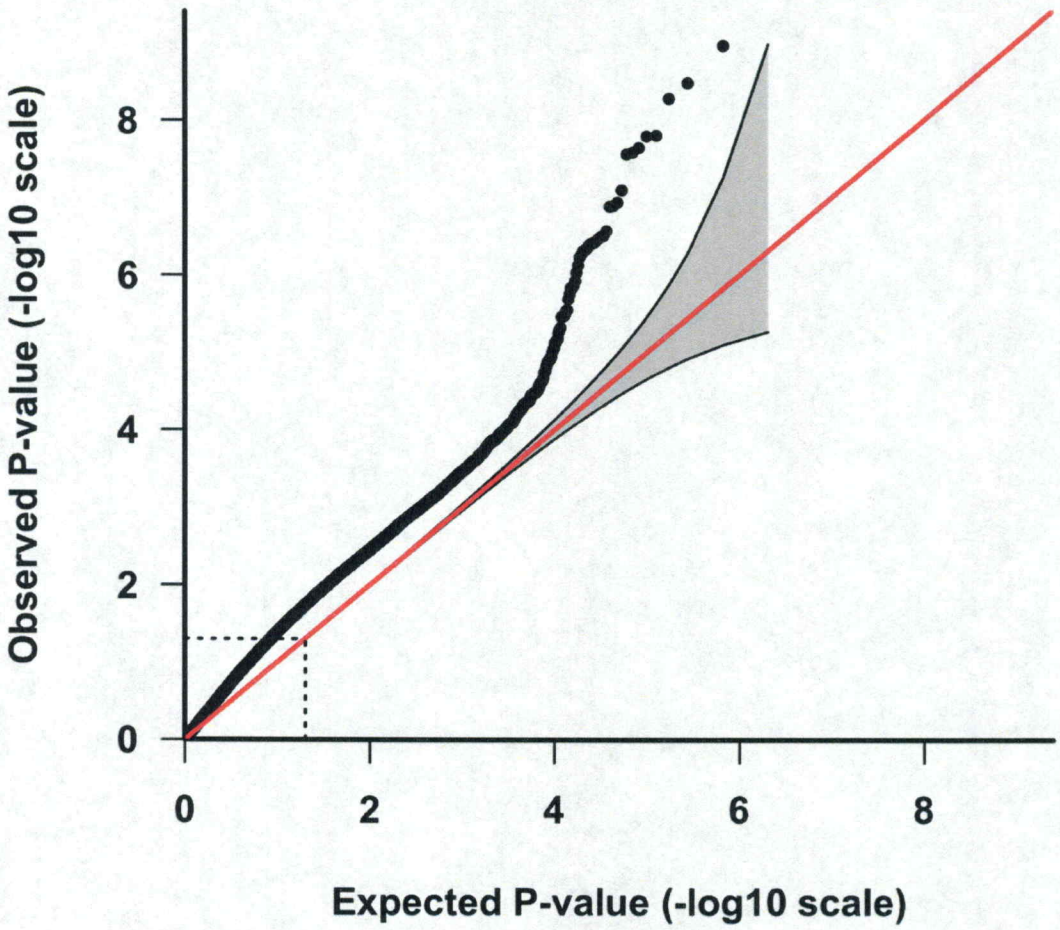


Figure B.5.2 Q-Q Plot of naive B cell ethnicity-specific SLE regression p-values.

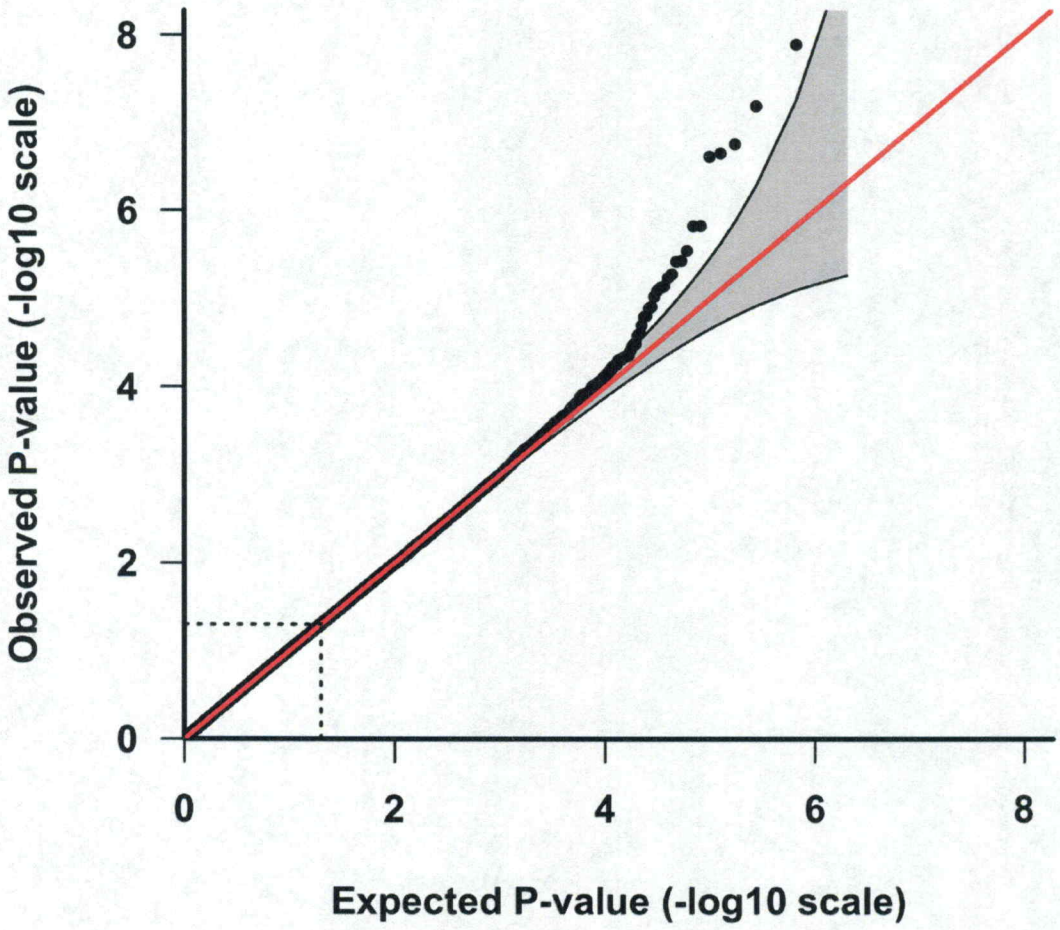


Figure B.5.3 Q-Q Plot of un-switched B cell ethnicity-specific SLE regression p-values.

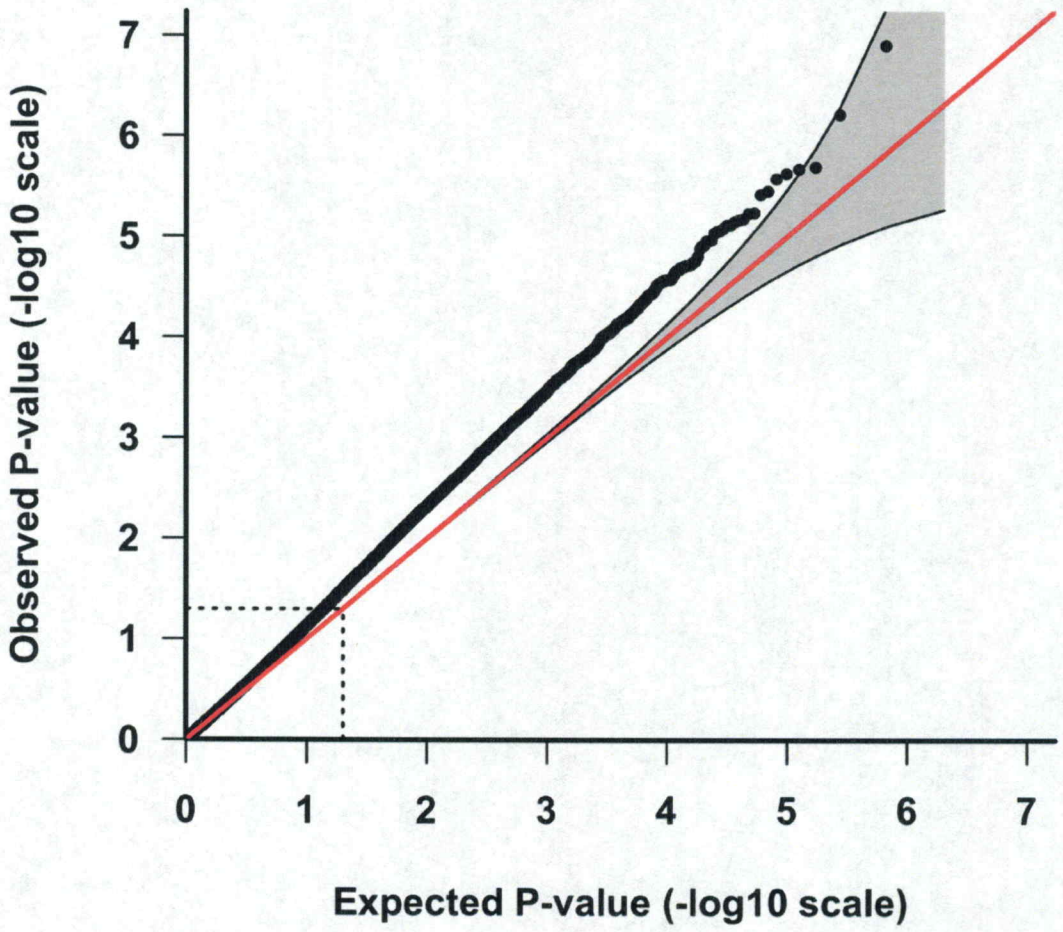


Figure B.5.4 Q-Q Plot of switched B cell ethnicity-specific SLE regression p-values.

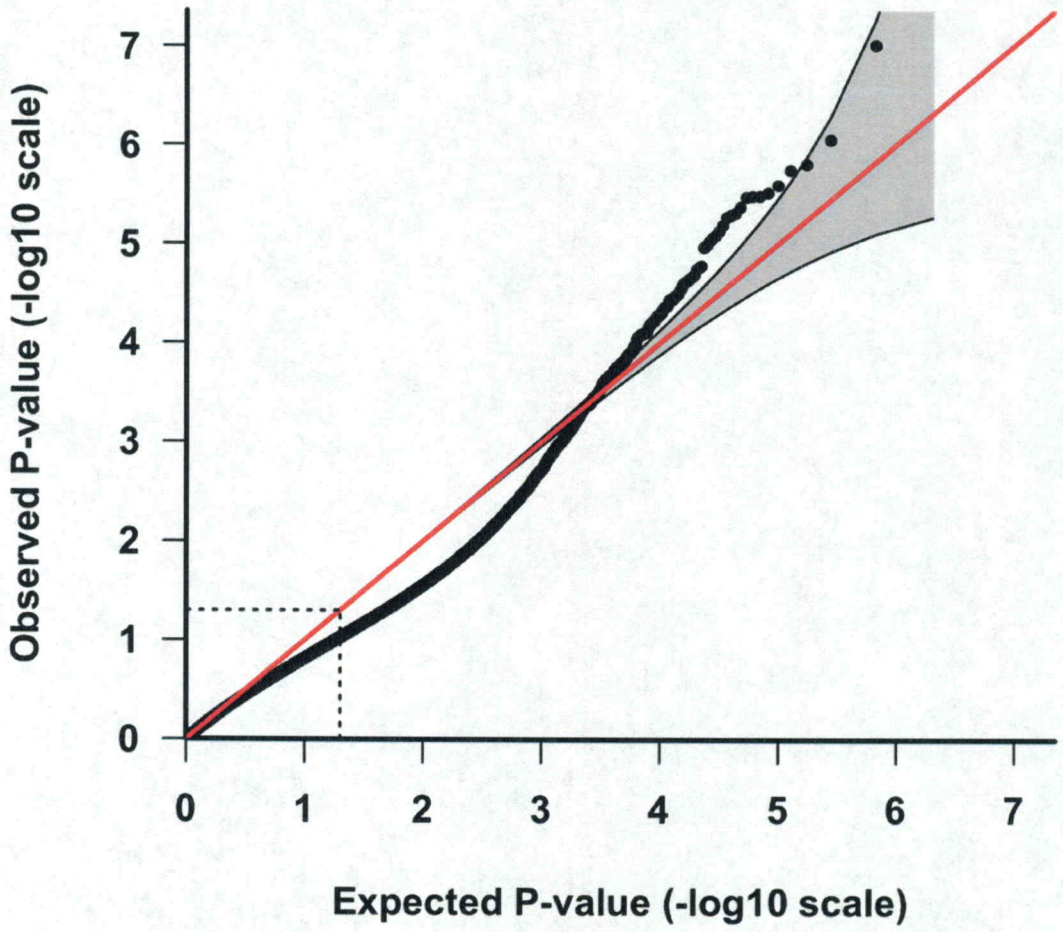


Figure B.5.5 Q-Q Plot of double negative B cell ethnicity-specific SLE regression p-values.

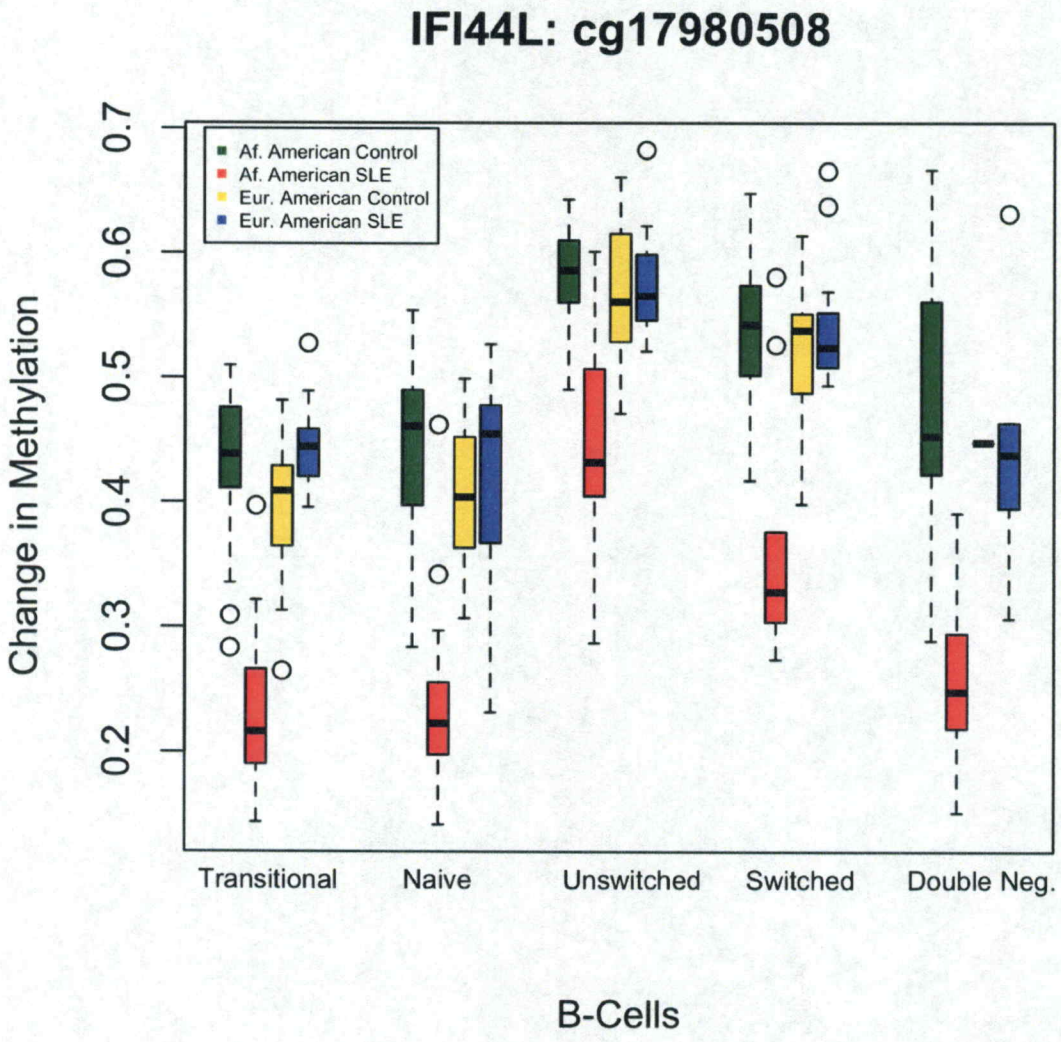


Figure B.6 Methylation levels across B cell development for the top AA-specific CpG in AA and EA control and SLE samples.

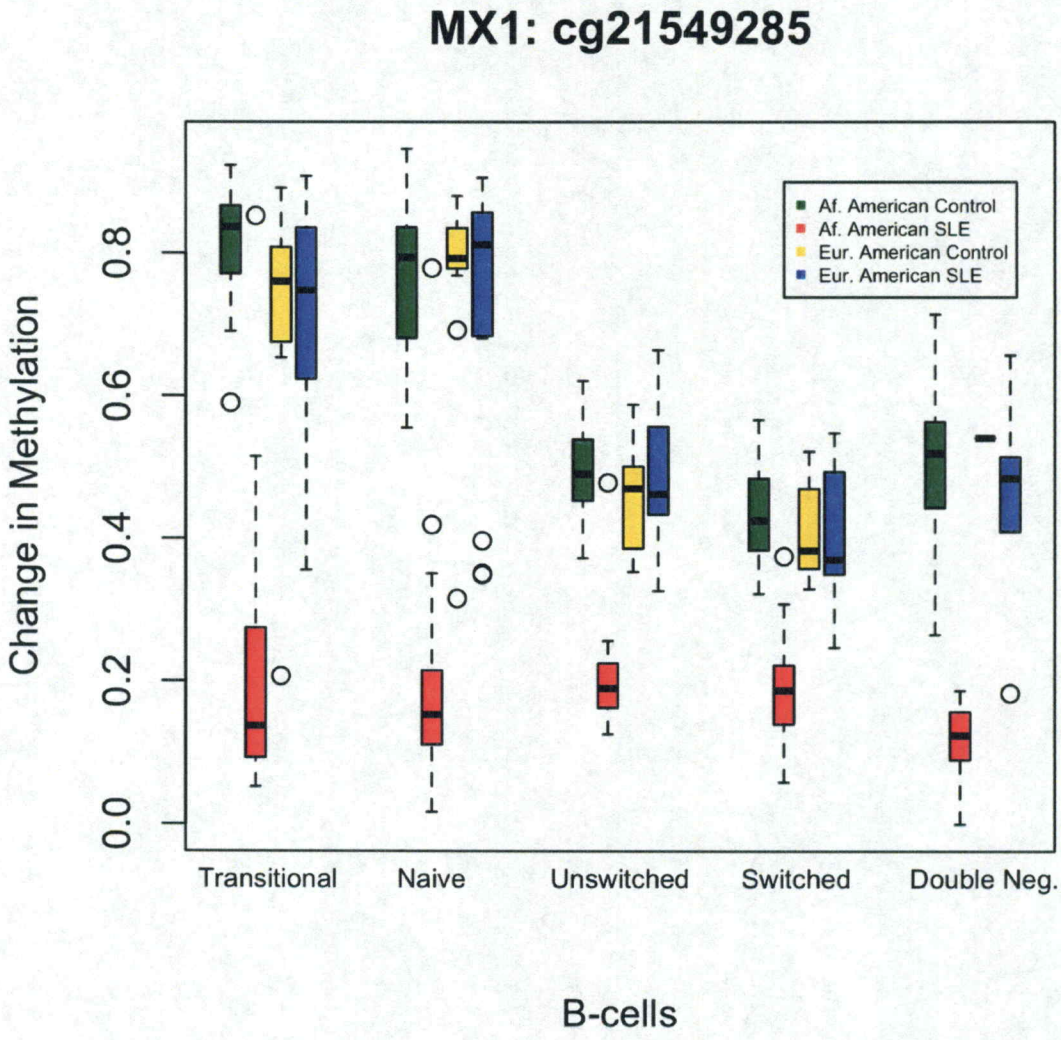


Figure B.7 Methylation levels across B cell development for the top MX1 CpG in AA and EA control and SLE samples.

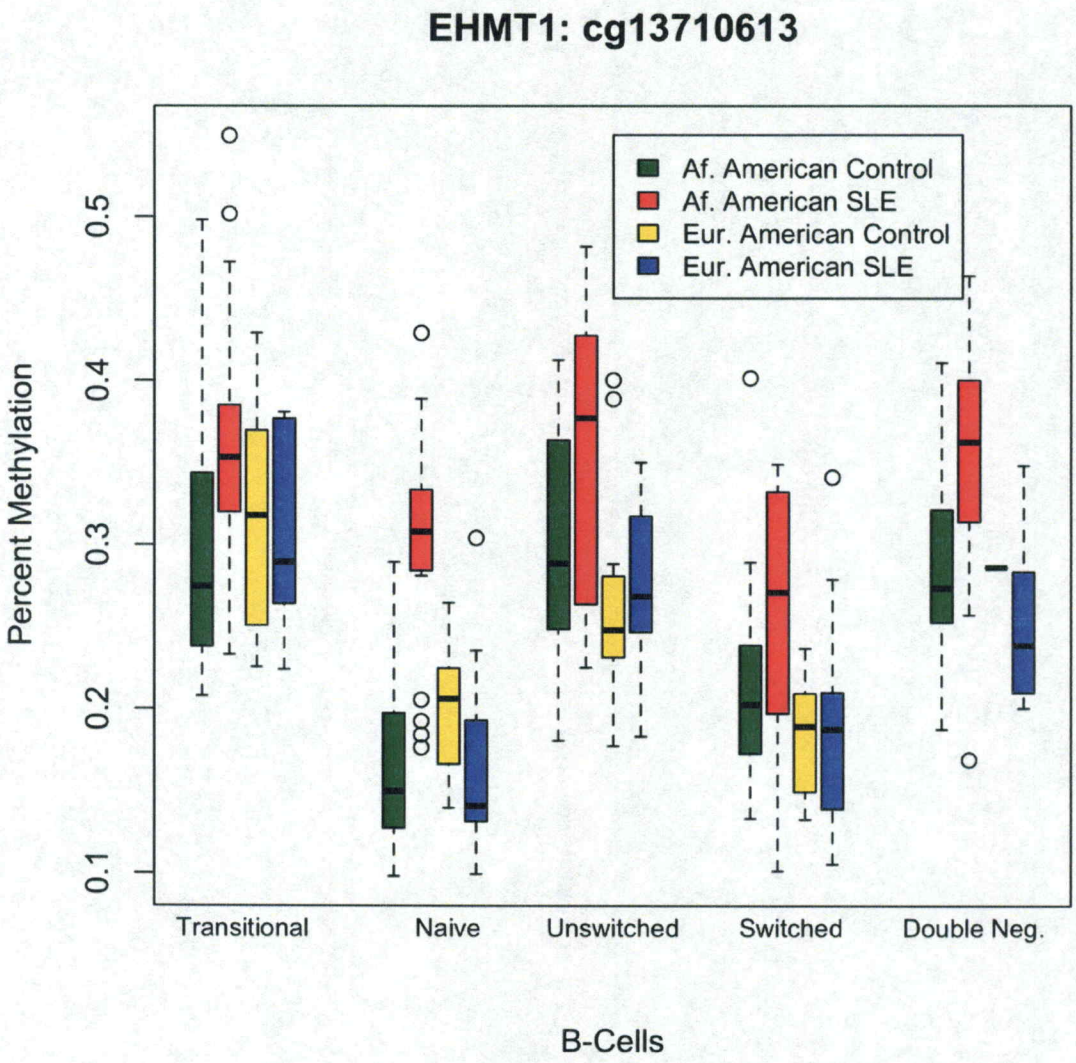


Figure B.8 Methylation levels across B cell development for the top EA-specific CpG in AA and EA control and SLE samples.

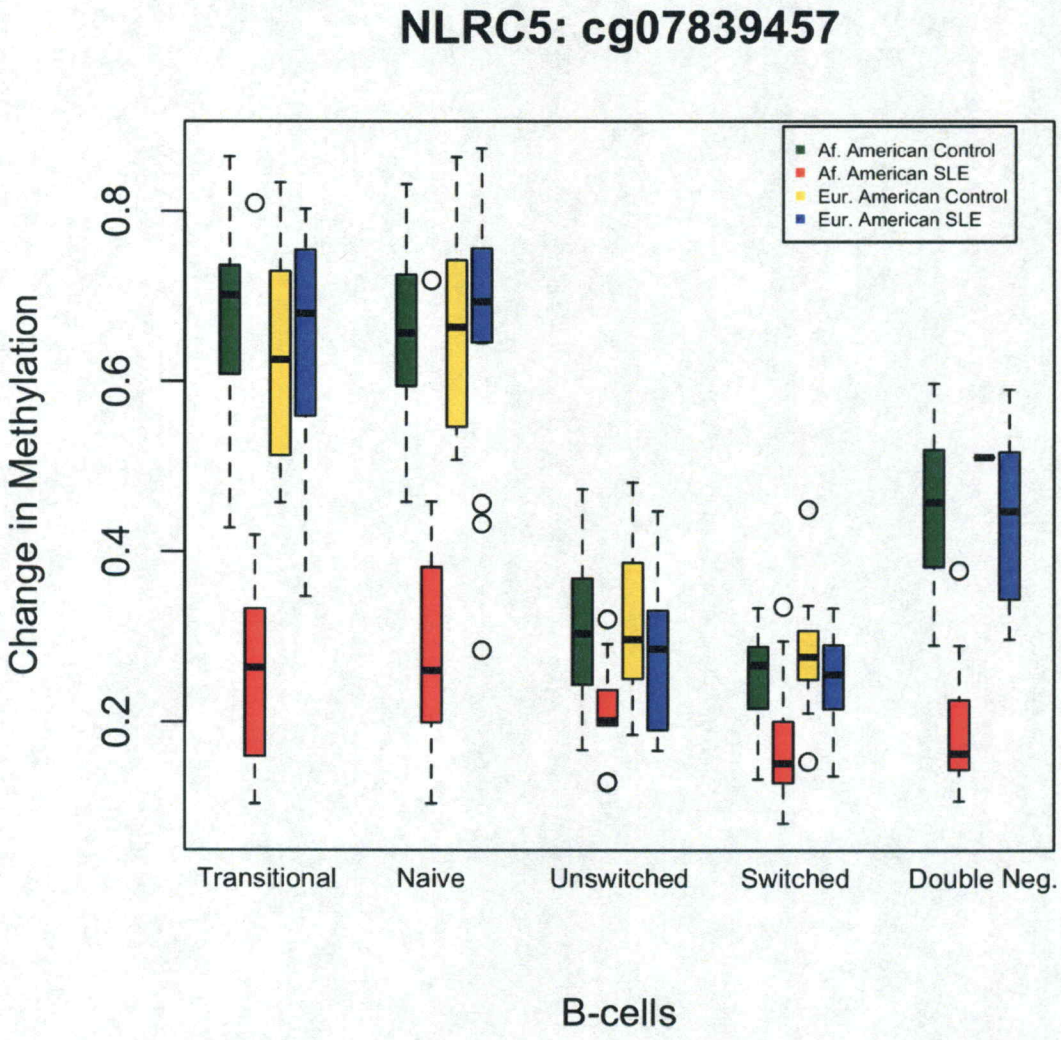


Figure B.9 Methylation levels across B cell development for the top AA SLE Random Forest predictor (CpG) in AA and EA control and SLE samples.

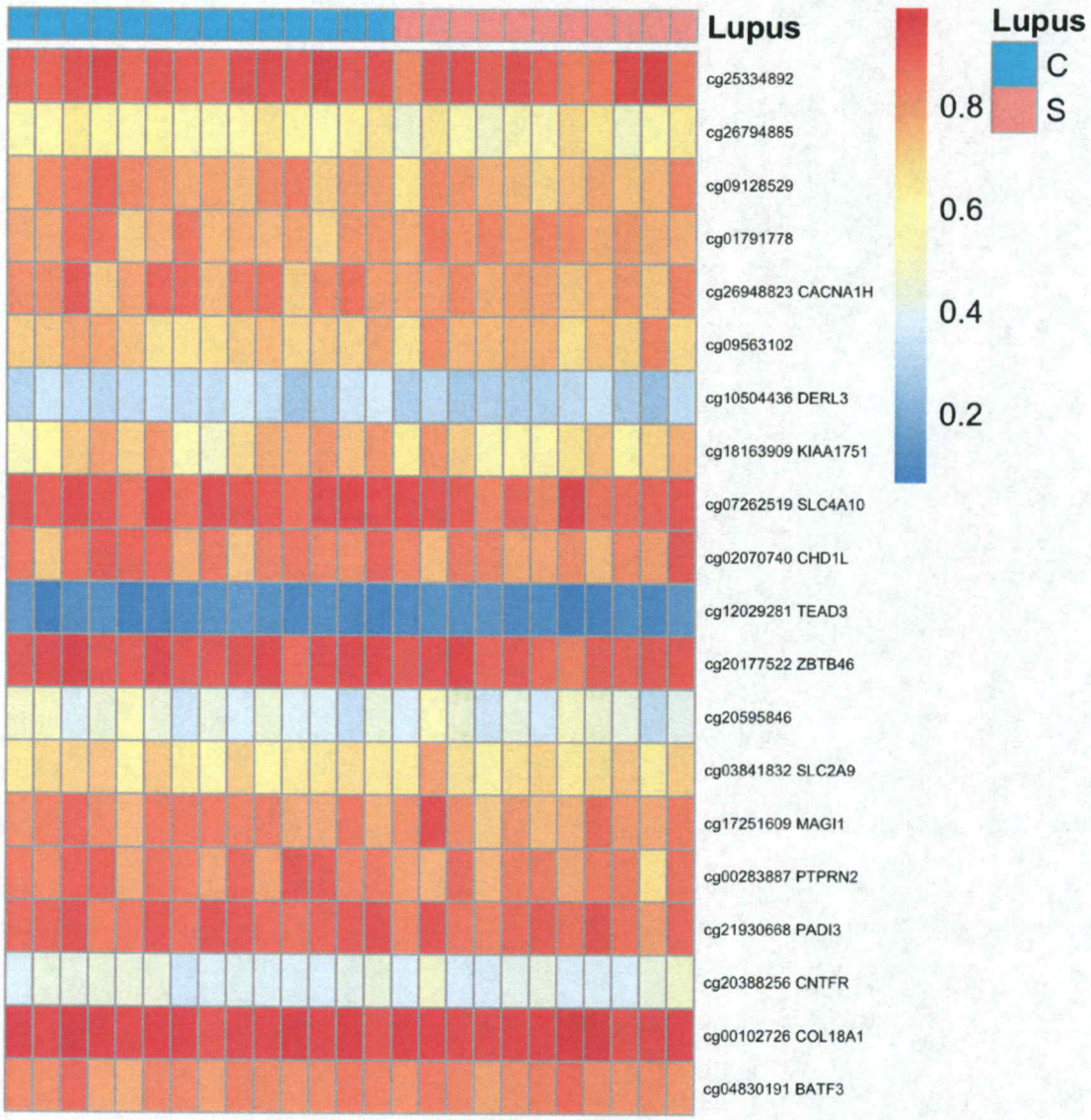


Figure B.10 Heatmap of methylation at top EA predictor CpGs in EA SLE patients and controls.

REFERENCES

Abellan Van Kan C, Behalf, T H E Geriatric, and Advisory Panel. 2008. "The I . a . N . a . Task Force on Frailty Assessment of Older People in Clinical Practice." *The Journal of Nutrition, Health & Aging*© 12 (1): 29–37. https://doi.org/10.1007/BF02982161.

Absher, Devin M., Xinrui Li, Lindsay L. Waite, Andrew Gibson, Kevin Roberts, Jeffrey Edberg, W. Winn Chatham, and Robert P. Kimberly. 2013. "Genome-Wide DNA Methylation Analysis of Systemic Lupus Erythematosus Reveals Persistent Hypomethylation of Interferon Genes and Compositional Changes to CD4+ T-Cell Populations." *PLoS Genetics* 9 (8). https://doi.org/10.1371/journal.pgen.1003678.

Alarcón-Segovia, Donato, Marta E. Alarcón-Riquelme, Mario H. Cardiel, Francisco Caeiro, Loreto Massardo, Antonio R. Villa, and Bernardo A. Pons-Estel. 2005. "Familial Aggregation of Systemic Lupus Erythematosus, Rheumatoid Arthritis, and Other Autoimmune Diseases in 1,177 Lupus Patients from the GLADEL Cohort." *Arthritis and Rheumatism* 52 (4): 1138–47. https://doi.org/10.1002/art.20999.

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. 2002. *Molecular Biology of the Cell*. 4th editio. New York: Garland Science.

AlFadhli, Suad, Mashael Al-Mutairi, Bader Al Tameemi, and Rasheeba Nizam. 2016. "Influence of MX1 Promoter Rs2071430 G/T Polymorphism on Susceptibility to Systemic Lupus Erythematosus." *Clinical Rheumatology* 35 (3): 623–29.

https://doi.org/10.1007/s10067-016-3179-z.

Anjum, Shahzia, Evangelia-Ourania Fourkala, Michal Zikan, Andrew Wong, Aleksandra Gentry-Maharaj, Allison Jones, Rebecca Hardy, et al. 2014. "A BRCA1-Mutation Associated DNA Methylation Signature in Blood Cells Predicts Sporadic Breast Cancer Incidence and Survival." *Genome Medicine* 6 (6): 47. https://doi.org/10.1186/gm567.

Antonarakis, S E. 2001. "The Search for Allelic Variants That Cause Monogenic Disorders or Predispose to Common, Complex Polygenic Phenotypes." *Dialogues in Clinical Neuroscience* 3 (1): 7–15. http://www.ncbi.nlm.nih.gov/pubmed/22034389%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3181641.

Astuti, Yuliana, Ardyan Wardhana, Johnathan Watkins, and Wahyu Wulaningsih. 2017. "Cigarette Smoking and Telomere Length: A Systematic Review of 84 Studies and Meta-Analysis." *Environmental Research* 158 (July): 480–89. https://doi.org/10.1016/j.envres.2017.06.038.

Aubert, Geraldine, and Peter M Lansdorp. 2008. "Telomeres and Aging." *Physiological Reviews* 88 (2): 557–79. https://doi.org/10.1152/physrev.00026.2007.

Baas, Dominique C., Dominiek D. Despriet, Theo G.M.F. Gorgels, Julie Bergeron-Sawitzke, André G. Uitterlinden, Albert Hofman, Cornelia M. van Duijn, et al. 2010. "The ERCC6 Gene and Age-Related Macular Degeneration." *PLoS ONE* 5 (11). https://doi.org/10.1371/journal.pone.0013786.

Baechler, Emily C, Franak M Batliwalla, George Karypis, Patrick M Gaffney, Ward A Ortmann, Karl J Espe, Katherine B Shark, et al. 2003. "Interferon-Inducible Gene

Expression Signature in Peripheral Blood Cells of Patients with Severe Lupus." *Proceedings of the National Academy of Sciences* 100 (5): 2610–15. https://doi.org/10.1073/pnas.0337679100.

Bailey, Timothy L., Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. 2009. "MEME Suite: Tools for Motif Discovery and Searching." *Nucleic Acids Research* 37 (SUPPL. 2): 202–8. https://doi.org/10.1093/nar/gkp335.

Baker, Darren J., Fang Jin, and Jan M. Van Deursen. 2008. "The Yin and Yang of the Cdkn2a Locus in Senescence and Aging." *Cell Cycle* 7 (18): 2795–2802. https://doi.org/10.4161/cc.7.18.6687.

Barneda-Zahonero, Bruna, Lidia Roman-Gonzalez, Olga Collazo, Tokameh Mahmoudi, and Maribel Parra. 2012. "Epigenetic Regulation of B Lymphocyte Differentiation, Transdifferentiation, and Reprogramming." *Comparative and Functional Genomics* 2012 (Figure 1). https://doi.org/10.1155/2012/564381.

Barnhoorn, Sander, Lieneke M. Uittenboogaard, Dick Jaarsma, Wilbert P. Vermeij, Maria Tresini, Michael Weymaere, Hervé Menoni, et al. 2014. "Cell-Autonomous Progeroid Changes in Conditional Mouse Models for Repair Endonuclease XPG Deficiency." *PLoS Genetics* 10 (10): 7–9. https://doi.org/10.1371/journal.pgen.1004686.

Behrens, Axel, Jan M van Deursen, K Lenhard Rudolph, and Björn Schumacher. 2014. "Impact of Genomic Damage and Ageing on Stem Cell Function." *Nature Cell Biology* 16 (3): 201–7. https://doi.org/10.1038/ncb2928.

Bendtsen, Kristian Moss, Jeppe Juul, and Ala Trusina. 2012. "Fragile DNA Repair

Mechanism Reduces Ageing in Multicellular Model." *PLoS ONE* 7 (5). https://doi.org/10.1371/journal.pone.0036018.

Beneke, Sascha, and Alexander Bürkle. 2007. "Survey and Summary: Poly(ADP-Ribosyl)Ation in Mammalian Ageing." *Nucleic Acids Research* 35 (22): 7456–65. https://doi.org/10.1093/nar/gkm735.

Bennett, Lynda, A. Karolina Palucka, Edsel Arce, Victoria Cantrell, Josef Borvak, Jacques Banchereau, and Virginia Pascual. 2003. "Interferon and Granulopoiesis Signatures in Systemic Lupus Erythematosus Blood." *The Journal of Experimental Medicine* 197 (6): 711–23. https://doi.org/10.1084/jem.20021553.

Bentham, James, David L Morris, Deborah S Cunninghame Graham, Christopher L Pinder, Philip Tombleson, Timothy W Behrens, Javier Martín, et al. 2015. "Genetic Association Analyses Implicate Aberrant Regulation of Innate and Adaptive Immunity Genes in the Pathogenesis of Systemic Lupus Erythematosus." *Nature Genetics 2015 47:12* 47 (12): 1457. https://doi.org/10.1038/ng.3434.

Berman, Albert E, Olga V Leontieva, Venkatesh Natarajan, James A McCubrey, Zoya N Demidenko, and Mikhail A Nikiforov. 2012. "Recent Progress in Genetics of Aging, Senescence and Longevity: Focusing on Cancer-Related Genes." *Oncotarget* 3 (12): 1522–32. https://doi.org/10.18632/oncotarget.889.

Bing, Peng Fei, Wei Xia, Lan Wang, Yong Hong Zhang, Shu Feng Lei, and Fei Yan Deng. 2016. "Common Marker Genes Identified from Various Sample Types for Systemic Lupus Erythematosus." *PLoS ONE* 11 (6): 1–12. https://doi.org/10.1371/journal.pone.0156234.

Blackburn, E. H., E. S. Epel, and J. Lin. 2015. "Human Telomere Biology: A

Contributory and Interactive Factor in Aging, Disease Risks, and Protection." *Science* 350 (6265): 1193–98. https://doi.org/10.1126/science.aab3389.

Blanco, Patrick, A Karolina Palucka, Michelle Gill, Virginia Pascual, and Jacques Banchereau. 2001. "Induction of Dendritic Cell Differentiation by IFN-α in Systemic Lupus Erythematosus." *Science* 294 (5546): 1540–43.

Blasco, Maria A. 2007. "Telomere Length, Stem Cells and Aging." *Nature Chemical Biology* 3 (10): 640–49. https://doi.org/10.1038/nchembio.2007.38.

Bogliolo, Massimo, Beatrice Schuster, Chantal Stoepker, Burak Derkunt, Yan Su, Anja Raams, Juan P. Trujillo, et al. 2013. "Mutations in ERCC4, Encoding the DNA-Repair Endonuclease XPF, Cause Fanconi Anemia." *American Journal of Human Genetics* 92 (5): 800–806. https://doi.org/10.1016/j.ajhg.2013.04.002.

Boonen, Steven, Dennis M. Black, Cathleen Colon-Emeric, Richard Eastell, Jay S. Magaziner, Erik Fink Eriksen, Peter Mesenbrink, Patrick Haentjens, and Kenneth W. Lyles. 2010. "Efficacy and Safety of a Once-Yearly Intravenous Zoledronic Acid 5 Mg for Fracture Prevention in Elderly Postmenopausal Women with Osteoporosis Aged 75 and Older Steven." *Journal of The American Geriatrics Society* 58 (2): 292–99. https://doi.org/10.1111/j.1532-5415.2009.02673.x.

Boonen, Steven, Fernando Marin, Dan Mellstrom, Li Xie, Durisala Desaiah, John H. Krege, and Clifford J. Rosen. 2006. "Safety and Efficacy of Teriparatide in Elderly Women with Established Osteoporosis: Bone Anabolic Therapy from a Geriatric Perspective." *Journal of the American Geriatrics Society* 54 (5): 782–89. https://doi.org/10.1111/j.1532-5415.2006.00695.x.

Bosse, Raymond, David Sparrow, Arthur J. Garvey, Paul T. Costa, Scott T. Weiss, and

John W. Rowe. 1980. "Cigarette Smoking Aging and Decline in Pulmonary Function: A Longitudinal Study." *Archives of Environmental Health.* https://doi.org/10.1080/00039896.1980.10667500.

Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard. 2017. "An Expanded View of Complex Traits: From Polygenic to Omnigenic." *Cell* 169 (7): 1177–86. https://doi.org/10.1016/j.cell.2017.05.038.

Breiman, Leo. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

Brezinschek, Hans Peter, Franz Rainer, Kerstin Brickmann, and Winfried B. Graninger. 2012. "B Lymphocyte-Typing for Prediction of Clinical Response to Rituximab." *Arthritis Research and Therapy* 14 (4): R161. https://doi.org/10.1186/ar3901.

Brinkman, R R, M M Mezei, J Theilmann, E Almqvist, and M R Hayden. 1997. "The Likelihood of Being Affected with Huntington Disease by a Particular Age, for a Specific CAG Size." *American Journal of Human Genetics* 60 (5): 1202–10. http://www.ncbi.nlm.nih.gov/pubmed/9150168%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1712445.

Broer, Linda, Aron S. Buchman, Joris Deelen, Daniel S. Evans, Jessica D. Faul, Kathryn L. Lunetta, Paola Sebastiani, et al. 2015. "GWAS of Longevity in CHARGE Consortium Confirms APOE and FOXO3 Candidacy." *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 70 (1): 110–18. https://doi.org/10.1093/gerona/glu166.

Brooks-Wilson, Angela R. 2013. "Genetics of Healthy Aging and Longevity." *Human Genetics* 132 (12): 1323–38. https://doi.org/10.1007/s00439-013-1342-z.

Bükle, A, K Grube, and JH Küpper. 1992. "No TitlePoly(ADP-Ribosyl)Ation: Its Role in Inducible DNA Amplification, and Its Correlation with the Longevity of Mammalian Species." *Experimental and Clinical Immunogenetics* 9 (4): 230–40.

Busslinger, M., J. Hurst, and R.A. Flavell. 1983. "DNA Methylation and the Regulation of Globin Gene Expression." *Cell* 34 (1): 197–206. https://doi.org/10.1016/0092-8674(83)90150-2.

Cabelof, Diane C, Julian J Raffoul, Sunitha Yanamadala, Cirlette Ganir, ZhongMao Guo, and Ahmad R Heydari. 2002. "Attenuation of DNA Polymerase Beta-Dependent Base Excision Repair and Increased DMS-Induced Mutagenicity in Aged Mice." *Mutation Research* 500 (1–2): 135–45. https://doi.org/10.1038/mp.2011.182.

CDC. 2018. "Lupus." Centers for Disease Control and Prevention. 2018. www.cdc.gov/lupus/facts/detailed.html.

"Cells and DNA." 2018. *National Institutes of Health.*

Chadwick, Lisa Helbling. 2012. "The NIH Roadmap Epigenomics Program Data Resource Lisa." *Epigenomics* 4 (3): 317–24. https://doi.org/10.2217/epi.12.18.The.

Chan, Vera Sau Fong, Helen Hoi Lun Tsang, Rachel Chun Yee Tam, Liwei Lu, and Chak Sing Lau. 2013. "B-Cell-Targeted Therapies in Systemic Lupus Erythematosus." *Cellular and Molecular Immunology* 10 (2): 133–42. https://doi.org/10.1038/cmi.2012.64.

Chen, Kun, Juan Liu, and Xuetao Cao. 2017. "Regulation of Type I Interferon Signaling in Immunity and Inflammation: A Comprehensive Review." *Journal of Autoimmunity* 83 (2017): 1–11. https://doi.org/10.1016/j.jaut.2017.03.008.

Chen, Sidi, Weilin Pu, Shicheng Guo, Li Jin, Dongyi He, and Jiucun Wang. 2019.

"Genome-Wide DNA Methylation Profiles Reveal Common Epigenetic Patterns of Interferon-Related Genes in Multiple Autoimmune Diseases." *Frontiers in Genetics* 10 (April). https://doi.org/10.3389/fgene.2019.00223.

Coit, Patrick, Mikhail G. Dozmorov, Joan T. Merrill, W. Joseph McCune, Kathleen Maksimowicz-McKinnon, Jonathan D. Wren, and Amr H. Sawalha. 2016. "Epigenetic Reprogramming in Naive CD4+ T Cells Favoring T Cell Activation and Non-Th1 Effector T Cell Immune Response as an Early Event in Lupus Flares." *Arthritis and Rheumatology* 68 (9): 2200–2209. https://doi.org/10.1002/art.39720.

Coit, Patrick, Matlock Jeffries, Nezam Altorok, Mikhail G. Dozmorov, Kristi A. Koelsch, Jonathan D. Wren, Joan T. Merrill, W. Joseph McCune, and Amr H. Sawalha. 2013. "Genome-Wide DNA Methylation Study Suggests Epigenetic Accessibility Andtranscriptional Poising of Interferon-Regulated Genes in Naïve CD4+ T Cellsfrom Lupus Patients." *Journal of Autoimmunity* 43 (2013): 78–84. https://doi.org/10.1016/j.jaut.2013.04.003.

Coit, Patrick, Mikhail Ognenovski, Elizabeth Gensterblum, Kathleen Maksimowicz-Mckinnon, Jonathan D. Wren, and Amr H. Sawalha. 2015. "Ethnicity-Specific Epigenetic Variation in Naive CD4+ T Cells and the Susceptibility to Autoimmunity." *Epigenetics and Chromatin* 8 (1): 1–13. https://doi.org/10.1186/s13072-015-0037-1.

Coit, Patrick, Srilakshmi Yalavarthi, Mikhail Ognenovski, Wenpu Zhao, Sarfaraz Hasni, Jonathan D. Wren, Mariana J. Kaplan, and Amr H. Sawalha. 2015. "Epigenome Profiling Reveals Significant DNA Demethylation of Interferon Signature Genes in Lupus Neutrophils." *Journal of Autoimmunity* 58: 59–66.

https://doi.org/10.1016/j.jaut.2015.01.004.

Conneely, Karen N., Brian C. Capell, Michael R. Erdos, Paola Sebastiani, Nadia Solovieff, Amy J. Swift, Clinton T. Baldwin, et al. 2012. "Human Longevity and Common Variations in the LMNA Gene: A Meta-Analysis." *Aging Cell* 11 (3): 475–81. https://doi.org/10.1111/j.1474-9726.2012.00808.x.

Coppede, Fabio, and Lucia Migliore. 2012. "DNA Repair in Premature Aging Disorders and Neurodegeneration." *Current Aging Science* 3 (1): 3–19. https://doi.org/10.2174/1874612811003010003.

Costa-Reis, Patrícia, and Kathleen E. Sullivan. 2013. "Genetics and Epigenetics of Systemic Lupus Erythematosus." *Current Rheumatology Reports* 15 (9). https://doi.org/10.1007/s11926-013-0369-4.

Csiszar, Anna. 2009. "Oxidative Stress and Accelerated Vascular Aging: Implications for Cigarette Smoking." *Frontiers in Bioscience* Volume (14): 3128. https://doi.org/10.2741/3440.

Czarny, Piotr, Dominik Kwiatkowski, Monika Toma, Joanna Kubiak, Agnieszka Sliwinska, Monika Talarowska, Janusz Szemraj, Michael Maes, Piotr Galecki, and Tomasz Sliwinski. 2017. "Impact of Single Nucleotide Polymorphisms of Base Excision Repair Genes on DNA Damage and Efficiency of DNA Repair in Recurrent Depression Disorder." *Molecular Neurobiology* 54 (6): 4150–59. https://doi.org/10.1007/s12035-016-9971-6.

Danchenko, N, JA Satia, and MS Anthony. 2006. "Epidemiology of Systematic Lupus Erythematosus: A Comparison of Worldwide Disease Burden." *Lupus* 12 (3): 308–18.

Davis, Carrie A., Benjamin C. Hitz, Cricket A. Sloan, Esther T. Chan, Jean M. Davidson, Idan Gabdank, Jason A. Hilton, et al. 2018. "The Encyclopedia of DNA Elements (ENCODE): Data Portal Update." *Nucleic Acids Research* 46 (D1): D794–801. https://doi.org/10.1093/nar/gkx1081.

Day, Kenneth, Jun Song, and Devin Absher. 2014. "Targeted Sequencing of Large Genomic Regions with CATCH-Seq." *PLoS ONE* 9 (10). https://doi.org/10.1371/journal.pone.0111756.

De, Saurav, and Betsy J. Barnes. 2014. "B Cell Transcription Factors: Potential New Therapeutic Targets for SLE." *Clinical Immunology* 152 (1–2): 140–51. https://doi.org/10.1016/j.clim.2014.03.009.

Deafen, Dennis, Agustin Escalante, Lisa Weinrib, David Horwitz, Barbara Bachman, Pradip Roy-Burman, Ann Walker, and Thomas M. Mack. 1992. "A Revised Estimate of Twin Concordance in Systemic Lupus Erythematosus." *Arthritis & Rheumatism* 35 (3): 311–18. https://doi.org/10.1002/art.1780350310.

Deelen, Joris, Marian Beekman, Miriam Capri, Claudio Franceschi, and P. Eline Slagboom. 2013. "Identifying the Genomic Determinants of Aging and Longevity in Human Population Studies: Progress and Challenges." *BioEssays* 35 (4): 386–96. https://doi.org/10.1002/bies.201200148.

Deelen, Joris, Marian Beekman, Hae Won Uh, Linda Broer, Kristin L. Ayers, Qihua Tan, Yoichiro Kamatani, et al. 2014. "Genome-Wide Association Meta-Analysis of Human Longevity Identifies a Novel Locus Conferring Survival beyond 90 Years of Age." *Human Molecular Genetics* 23 (16): 4420–32. https://doi.org/10.1093/hmg/ddu139.

Demirci, F. Yesim, Xingbin Wang, Jennifer A. Kelly, David L. Morris, M. Michael Barmada, Eleanor Feingold, Amy H. Kao, et al. 2016. "Identification of a New Susceptibility Locus for Systemic Lupus Erythematosus on Chromosome 12 in Individuals of European Ancestry." *Arthritis & Rheumatology* 68 (1): 174–83. https://doi.org/10.1002/art.39403.

Deng, Yun, and Betty P. Tsao. 2017. "Updates in Lupus Genetics." *Current Rheumatology Reports* 19 (11). https://doi.org/10.1007/s11926-017-0695-z.

Dimitriadou, Author Evgenia, Kurt Hornik, Friedrich Leisch, David Meyer, Andreas Weingessel, and Maintainer Friedrich Leisch Friedrichleischcituwienacat. 2005. "The E1071 Package." https://cran.r-project.org/web/packages/e1071/index.html.

Ding, S L, J C Yu, S T Chen, G C Hsu, and C Y Shen. 2007. "Genetic Variation in the Premature Aging Gene WRN: A Case-Control Study on Breast Cancer Susceptibility." *Cancer Epidemiol Biomarkers Prev* 16 (2): 263–69. https://doi.org/16/2/263 [pii]\r10.1158/1055-9965.EPI-06-0678.

Donahue, A. C., and D. A. Fruman. 2014. "Proliferation and Survival of Activated B Cells Requires Sustained Antigen Receptor Engagement and Phosphoinositide 3-Kinase Activation." *The Journal of Immunology* 170 (12): 5851–60. https://doi.org/10.4049/jimmunol.170.12.5851.

Eline Slagboom, P., Niels van den Berg, and Joris Deelen. 2018. "Phenome and Genome Based Studies into Human Ageing and Longevity: An Overview." *Biochimica et Biophysica Acta - Molecular Basis of Disease* 1864 (9): 2742–51. https://doi.org/10.1016/j.bbadis.2017.09.017.

Encode Consortium, North Carolina, and Chapel Hill. 2013. "An Integrated Encyclopedia

of DNA Elements in the Human Genome." *Nature* 489 (7414): 57–74. https://doi.org/10.1038/nature11247.An.

Erikson, Galina A, Dale L Bodian, Manuel Rueda, John E Niederhuber, Eric J Topol, Ali Torkamani, Galina A Erikson, et al. 2016. "Whole-Genome Sequencing of a Healthy Aging Resource Whole-Genome Sequencing of a Healthy Aging Cohort." *Cell* 165 (4): 1002–11. https://doi.org/10.1016/j.cell.2016.03.022.

Estus, Steven, Benjamin C. Shaw, Nicholas Devanney, Yuriko Katsumata, Eileen E. Press, and David W. Fardo. 2019. "Evaluation of CD33 as a Genetic Risk Factor for Alzheimer's Disease." *Acta Neuropathologica*, no. 0123456789. https://doi.org/10.1007/s00401-019-02000-4.

Fabrizio, P., S. D. Pletcher, N. Minois, J. W. Vaupel, and V. D. Longo. 2004. "Chronological Aging-Independent Replicative Life Span Regulation by Msn2/Msn4 and Sod2 in Saccharomyces Cerevisiae." *FEBS Letters* 557 (1–3): 136–42. https://doi.org/10.1016/S0014-5793(03)01462-5.

Feng, Xuebing, Hui Wu, Jennifer M. Grossman, Punchong Hanvivadhanakul, John D. FitzGerald, Grace S. Park, Xin Dong, et al. 2006. "Association of Increased Interferon-Inducible Gene Expression with Disease Activity and Lupus Nephritis in Patients with Systemic Lupus Erythematosus." *Arthritis and Rheumatism* 54 (9): 2951–62. https://doi.org/10.1002/art.22044.

Freedman, Barry I., Carl D. Langefeld, Kelly K. Andringa, Jennifer A. Croker, Adrienne H. Williams, Neva E. Garner, Daniel J. Birmingham, et al. 2014. "End-Stage Renal Disease in African Americans with Lupus Nephritis Is Associated with APOL1." *Arthritis and Rheumatology* 66 (2): 390–96. https://doi.org/10.1002/art.38220.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1). https://doi.org/10.18637/jss.v068.b04.

Furey, Terrence S, Nello Cristianini, Nigel Duffy, and W David. 2000. "Support Vector Machine Classification and Validation of Cancer Tissue Samples Using Microarray Expression Data." *Bioinformatics* 16 (10): 906–14.

Gallati, Sabina. 2014. "Disease-Modifying Genes and Monogenic Disorders: Experience in Cystic Fibrosis." *Application of Clinical Genetics* 7: 133–46. https://doi.org/10.2147/TACG.S18675.

Gamazon, Eric R, Wei Zhang, Anuar Konkashbaev, Shiwei Duan, Emily O Kistner, Dan L Nicolae, M Eileen Dolan, and Nancy J Cox. 2010. "SCAN : SNP and Copy Number Annotation" 26 (2): 259–62. https://doi.org/10.1093/bioinformatics/btp644.

Garaud, Jean Claude, Jean Nicolas Schickel, Gilles Blaison, Anne Marie Knapp, Doulaye Dembele, Julie Ruer-Laventie, Anne Sophie Korganow, Thierry Martin, Pauline Soulas-Sprauel, and Jean Louis Pasquali. 2011. "B Cell Signature during Inactive Systemic Lupus Is Heterogeneous: Toward a Biological Dissection of Lupus." *PLoS ONE* 6 (8): 4–12. https://doi.org/10.1371/journal.pone.0023900.

Gensler, Helen L, and Harris Bernstein. 1981. "DNA Damage as the Primary Cause of Aging." *The Quarterly Review of Biology* 56 (3): 279–303.

Ghodke-puranik, Yogita, Molly Imgruet, Jessica M Dorschner, Prakriti Shrestha, Kaci Mccoy, Jennifer A Kelly, Miranda Marion, et al. 2019. "Novel Genetic Associations with Interferon in Systemic Lupus Erythematosus Identi Fi Ed by Replication and Fi Ne-Mapping of Trait-Strati Fi Ed Genome-Wide Screen." *Cytokine*, no. October

2018: 0–1. https://doi.org/10.1016/j.cyto.2018.12.014.

Ghosh, Shrestha, and Zhongjun Zhou. 2014. "Genetics of Aging, Progeria and Lamin Disorders." *Current Opinion in Genetics and Development* 26: 41–46. https://doi.org/10.1016/j.gde.2014.05.003.

Good, K. L., D. T. Avery, and S. G. Tangye. 2014. "Resting Human Memory B Cells Are Intrinsically Programmed for Enhanced Survival and Responsiveness to Diverse Stimuli Compared to Naive B Cells." *The Journal of Immunology* 182 (2): 890–901. https://doi.org/10.4049/jimmunol.182.2.890.

Good, K. L., and S. G. Tangye. 2007. "Decreased Expression of Kruppel-like Factors in Memory B Cells Induces the Rapid Response Typical of Secondary Antibody Responses." *Proceedings of the National Academy of Sciences* 104 (33): 13420–25. https://doi.org/10.1073/pnas.0703872104.

Grabowska, Wioleta, Ewa Sikora, and Anna Bielak-Zmijewska. 2017. "Sirtuins, a Promising Target in Slowing down the Ageing Process." *Biogerontology* 18 (4): 447–76. https://doi.org/10.1007/s10522-017-9685-9.

Grant, Charles E., Timothy L. Bailey, and William Stafford Noble. 2011. "FIMO: Scanning for Occurrences of a given Motif." *Bioinformatics* 27 (7): 1017–18. https://doi.org/10.1093/bioinformatics/btr064.

Griciuc, Ana, Alberto Serrano-Pozo, Antonio R. Parrado, Andrea N. Lesinski, Caroline N. Asselin, Kristina Mullin, Basavaraj Hooli, Se Hoon Choi, Bradley T. Hyman, and Rudolph E. Tanzi. 2013. "Alzheimer's Disease Risk Gene Cd33 Inhibits Microglial Uptake of Amyloid Beta." *Neuron* 78 (4): 631–43. https://doi.org/10.1016/j.neuron.2013.04.014.

Gu, Bai Wei, Jian Meng Fan, Monica Bessler, and Philip J. Mason. 2011. "Accelerated Hematopoietic Stem Cell Aging in a Mouse Model of Dyskeratosis Congenita Responds to Antioxidant Treatment." *Aging Cell* 10 (2): 338–48. https://doi.org/10.1111/j.1474-9726.2011.00674.x.

Guenette, D K, J D Ritzenthaler, J Foley, J D Jackson, and B D Smith. 1992. "DNA Methylation Inhibits Transcription of Procollagen α 2(I) Promoters." *Biochemical Journal* 283 (3): 699–703. https://doi.org/10.1042/bj2830699.

Guillermo J. Pons-Estel, MD*, Graciela S. Alarcón, MD, MPH*, Lacie Scofield, MSPH†,‡, Leslie, and PhD‡ Reinlib, PhD‡, and Glinda S. Cooper. 2010. "Understanding the Epidemiology and Progression of Systemic Lupus Erythematosus" 39 (4): 1–23. https://doi.org/10.1016/j.semarthrit.2008.10.007.Understanding.

Gusella, James F., Nancy S. Wexler, P. Michael Conneally, Susan L. Naylor, Mary Anne Anderson, Rudolph E. Tanzi, Paul C. Watkins, et al. 1983. "A Polymorphic DNA Marker Genetically Linked to Huntington's Disease." *Nature* 306 (5940): 234–38. https://doi.org/10.1038/306234a0.

Hagman, James, Julita Ramirez, and Kara Lukin. 2012. "B Lymphocyte Lineage Specification, Commitment and Epigenetic Control of Transcription by Early B Cell Factor 1 James." *World Electric Vehicle Journal* Current To (356): 17–38. https://doi.org/10.1007/82.

Haller, Otto, and Georg Kochs. 1996. "Human MxA Protein: An Interferon-Induced Dynamin-Like GTPase with Broad Antiviral Activity." *Journal of Interferon and Cytokine Research* 50 (3): 248–49. https://doi.org/10.1089/jir.2010.0076.

Hamilton, Jennie A., Qi Wu, PingAr Yang, Bao Luo, Shanrun Liu, Jun Li, Alexa L. Mattheyses, et al. 2018. "Cutting Edge: Intracellular IFN-β and Distinct Type I IFN Expression Patterns in Circulating Systemic Lupus Erythematosus B Cells." *The Journal of Immunology* 201 (8): 2203–8. https://doi.org/10.4049/jimmunol.1800791.

Han, G-M, S-L Chen, N Shen, S Ye, C-D Bao, and Y-Y Gu. 2003. "Analysis of Gene Expression Profiles in Human Systemic Lupus Erythematosus Using Oligonucleotide Microarray." *Genes and Immunity* 4: 177–86. https://doi.org/10.1038/sj.gene.6363966.

Harley, JB, and ME Alarcón-Riquelme. 2013. "Genome-Wide Association Scan in Women with Systemic Lupus Erythematosus Identifies Susceptibility Variants in ITGAM, PXK, KIAA1542 and Other Loci." *Nature* ... 40 (2): 204–10. https://doi.org/10.1038/ng.81.Genome-wide.

Hart, R W, and R B Setlow. 1974. "Correlation between Deoxyribonucleic Acid Excision-Repair and Life-Span in a Number of Mammalian Species." *Proceedings of the National Academy of Sciences USA* 71 (6): 2169–73. https://doi.org/10.1073/pnas.71.6.2169.

Hastie, Trevor, Robert Tibshirani, Balasubramanian Narasimhan, and C Gilbert. 2001. "Impute: Imputation for Microarray Data." *Bioinformatics*. https://doi.org/10.18129/B9.bioc.impute.

Heinz, Sven, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. 2010. "Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities."

Molecular Cell 38 (4): 576–89. https://doi.org/10.1016/j.molcel.2010.05.004.

Herskind, Anne Maria, Matthew McGue, Niels V. Holm, Thorkild I.A. Sorensen, Bent Harvald, and James W. Vaupel. 1996. "The Heritability of Human Longevity : A Population-Based Study of 2872 Danish Twin Pairs Born 1870-1900." *Human Genetics* 97: 319–23.

Hiona, Asimina, and Christiaan Leeuwenburgh. 2008. "The Role of Mitochondrial DNA Mutations in Aging and Sarcopenia." *Experimental Gerontology* 43 (1): 24–33.

Hirschhorn, Joel N., Kirk Lohmueller, Edward Byrne, and Kurt Hirschhorn. 2002. "A Comprehensive Review of Genetic Association Studies." *Genetics in Medicine* 4 (2): 45–61. https://doi.org/10.1097/00125817-200203000-00002.

Hoeijmakers, Jan H.J. 2009. "DNA Damage, Aging, and Cancer." *New England Journal of Medicine* 361 (15): 1475–85. https://doi.org/10.1056/NEJMra0804615.

Hoffman, Robert W., Joan T. Merrill, Marta M.E. Alarcón-Riquelme, Michelle Petri, Ernst R. Dow, Eric Nantz, Laura K. Nisenbaum, et al. 2017. "Gene Expression and Pharmacodynamic Changes in 1,760 Systemic Lupus Erythematosus Patients From Two Phase III Trials of BAFF Blockade With Tabalumab." *Arthritis and Rheumatology* 69 (3): 643–54. https://doi.org/10.1002/art.39950.

Horcher, Markus, Abdallah Souabni, and Meinrad Busslinger. 2001. "Pax5/BSAP Maintains the Identity of B Cells in Late B Lymphopoiesis." *Immunity* 14 (6): 779–90. https://doi.org/10.1016/S1074-7613(01)00153-4.

Horikoshi, M, H Yaghootkar, DO Mook-Kanamori, U Sovio, HR Taal, BJ Hennig, JP Bradfield, et al. 2013. "New Loci Associated with Birth Weight Identify Genetic Links between Intrauterine Growth and Adult Height and Metabolism." *Nature*

Genetics 45 (1). https://doi.org/10.2337/db13-dd08.

Hu, Youjin, Kevin Huang, Qin An, Guizhen Du, Ganlu Hu, Jinfeng Xue, Xianmin Zhu, Cun-Yu Wang, Zhigang Xue, and Guoping Fan. 2016. "Simultaneous Profiling of Transcriptome and DNA Methylome from a Single Cell." *Genome Biology* 17 (1): 88. https://doi.org/10.1186/s13059-016-0950-z.

Imgenberg-Kreuz, Juliana, Johanna K. Sandling, Jonas Carlsson Almlöf, Jessica Nordlund, Linnea Signér, Katrine Braekke Norheim, Roald Omdal, et al. 2016. "Genome-Wide DNA Methylation Analysis in Multiple Tissues in Primary Sjögren's Syndrome Reveals Regulatory Effects at Interferon-Induced Genes." *Annals of the Rheumatic Diseases* 75 (11): 2029–36. https://doi.org/10.1136/annrheumdis-2015-208659.

Ioannou, Yiannis, and David A. Isenberg. 2000. "Current Evidence for the Induction of Autoimmune Rheumatic Manifestations by Cytokine Therapy." *Arthritis and Rheumatism* 43 (7): 1431–42. https://doi.org/10.1002/1529-0131(200007)43:7<1431::AID-ANR3>3.0.CO;2-E.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science* 333 (6047): 1300–1303. https://doi.org/10.1126/science.1210597.

Ivashkiv, Lionel B., and Laura T. Donlin. 2015. "Regulation of Type I Interferon Responses." *Nature Reviews Immunology* 14 (1): 36–49. https://doi.org/10.1038/nri3581.Regulation.

Iwamoto, Taro, Jessica Dorschner, Meenakshi Jolly, Xiangyang Huang, and Timothy B

Niewold. 2018. "Associations between Type I Interferon and Antiphospholipid Antibody Status Differ between Ancestral Backgrounds." *Lupus Science & Medicine* 5 (1): e000246. https://doi.org/10.1136/lupus-2017-000246.

J.E. Ware, C. D. Sherbourn. 1992. "The MOS 36-Item Short-Form Health Survey (SF-36): I . Conceptual Framework and Item Selection Author (s): John E . Ware , Jr . and Cathy Donald Sherbourne Published by : Lippincott Williams & Wilkins Stable URL : Http://Www.Jstor.Org/Stable/3765916 Ac." *Medical Care* 30 (6): 473–83.

Jaffe, Andrew E., and Rafael A. Irizarry. 2014. "Accounting for Cellular Heterogeneity Is Critical in Epigenome-Wide Association Studies." *Genome Biology* 15 (2): 1–9. https://doi.org/10.1186/gb-2014-15-2-r31.

Jarukitsopa, S, D D Hoganson, C S Crowson, O Sokumbi, M D Davis, C J Michet Jr., E L Matteson, H Maradit Kremers, and V R Chowdhary. 2015. "Epidemiology of Systemic Lupus Erythematosus and Cutaneous Lupus Erythematosus in a Predominantly White Population in the United States." *Arthritis Care Res (Hoboken)* 67 (6): 817–28. https://doi.org/10.1002/acr.22502.

Jeffries, Matlock, Mikhail Dozmorov, Yuhong Tang, Joan T. Merrill, Jonathan D. Wren, and Amr H. Sawalha. 2011. "Genome-Wide DNA Methylation Patterns in CD4+ T Cells from Patients with Systemic Lupus Erythematosus." *Epigenetics* 6 (5): 593–601. https://doi.org/10.4161/epi.6.5.15374.

Jin, Bilian, and Keith D. Robertson. 2013. "DNA Methyltransferases (DNMTs), DNA Damage Repair, and Cancer Bilian." *Advances in Experimental Medical Biology* 754 (754): 3–29. https://doi.org/10.1108/AMHID-01-2015-0001.

Jin, Zhongbo, Wei Fan, Mark A Jensen, Jessica M Dorschner, George F Bonadurer,

Danielle M Vsetecka, Shreyasee Amin, et al. 2017. "Single-Cell Gene Expression Patterns in Lupus Monocytes Independently Indicate Disease Activity, Interferon and Therapy." *Lupus Science & Medicine* 4 (1): e000202. https://doi.org/10.1136/lupus-2016-000202.

Johannsen, W. 1911. "The Genotype Conception of Heredity." *The American Naturalist* 45 (531): 129–59.

Jorgensen, Timothy J., Kathy J. Helzlsouer, Sandra C. Clipp, Judy Hoffman Bolton, Rosa M. Crum, and Kala Visvanathan. 2009. "NIH Public Access." *Cancer Epidemiol Biomarkers Prev* 18 (1): 346–50. https://doi.org/10.1097/MCA.0000000000000178.Endothelial.

Jung, David, Cosmas Giallourakis, Raul Mostoslavsky, and Frederick W. Alt. 2006. "Mechanism and Control of V(D)J Recombination At the Immunoglobulin Heavy Chain Locus." *Annual Review of Immunology* 24 (1): 541–70. https://doi.org/10.1146/annurev.immunol.23.021704.115830.

Kamen, Diane L. 2014. "Environmental Influences on Systemic Lupus Erythematosus Expression." *Rheumatic Disease Clinics of North America*. https://doi.org/10.1016/j.rdc.2014.05.003.

Kariuki, S N, J M Dorschner, B S Chrabot, J A Kelly, B P Tsao, R P Kimberly, and M E Alarcón-riquelme. 2015. "Genetic Analysis of the Pathogenic Molecular Sub-Phenotype Interferon-Alpha Identi Fi Es Multiple Novel Loci Involved in Systemic Lupus Erythematosus." *Genes and Immunity* 16 (September 2014): 15–23. https://doi.org/10.1038/gene.2014.57.

Kariuki, Silvia N., Beverly S. Franek, Akaash A. Kumar, Jasmine Arrington, Rachel A.

Mikolaitis, Tammy O. Utset, Meenakshi Jolly, Mary K. Crow, Andrew D. Skol, and Timothy B. Niewold. 2010. "Trait-Stratified Genome-Wide Association Study Identifies Novel and Diverse Genetic Associations with Serologic and Cytokine Phenotypes in Systemic Lupus Erythematosus." *Arthritis Research and Therapy* 12 (4). https://doi.org/10.1186/ar3101.

Karow, Julia K., Leonard Wu, and Ian D. Hickson. 2000. "RecQ Family Helicases: Roles in Cancer and Aging." *Current Opinion in Genetics and Development* 10 (1): 32–38. https://doi.org/10.1016/S0959-437X(99)00039-8.

Katyal, Sachin, and Peter J. McKinnon. 2008. "DNA Strand Breaks, Neurodegeneration and Aging in the Brain." *Mechanisms of Ageing and Development* 129 (7–8): 483–91. https://doi.org/10.1016/j.mad.2008.03.008.

Kawahara, Tiara L.A., Nicole A. Rapicavoli, Angela R. Wu, Kun Qu, Stephen R. Quake, and Howard Y. Chang. 2011. "Dynamic Chromatin Localization of Sirt6 Shapes Stress- and Aging-Related Transcriptional Networks." *PLoS Genetics* 7 (6): 1–12. https://doi.org/10.1371/journal.pgen.1002153.

Keane, Michael, Jeremy Semeiks, Andrew E. Webb, Yang I. Li, Víctor Quesada, Thomas Craig, Lone Bruhn Madsen, et al. 2015. "Insights into the Evolution of Longevity from the Bowhead Whale Genome." *Cell Reports* 10 (1): 112–22. https://doi.org/10.1016/j.celrep.2014.12.008.

Kelly, J. A., J. M. Kelley, K. M. Kaufman, J. Kilpatrick, G. R. Bruner, J. T. Merrill, J. A. James, et al. 2008. "Interferon Regulatory Factor-5 Is Genetically Associated with Systemic Lupus Erythematosus in African Americans." *Genes and Immunity* 9 (3): 187–94. https://doi.org/10.1038/gene.2008.4.

Kent, W., C. Sugnet, T. Furey, and K. Roskin. 1976. "The Human Genome Browser at UCSC W. [Available from: Http://Genome.Ucsc.Edu/]." *Journal of Medicinal Chemistry* 19 (10): 1228–31. https://doi.org/10.1101/gr.229102.

Khan, Aziz, Oriol Fornes, Arnaud Stigliani, Marius Gheorghe, Jaime A. Castro-Mondragon, Robin Van Der Lee, Adrien Bessy, et al. 2018. "JASPAR 2018: Update of the Open-Access Database of Transcription Factor Binding Profiles and Its Web Framework." *Nucleic Acids Research* 46 (D1): D260–66. https://doi.org/10.1093/nar/gkx1126.

Khera, Amit V., Mark Chaffin, Krishna G. Aragam, Mary E. Haas, Carolina Roselli, Seung Hoan Choi, Pradeep Natarajan, et al. 2018. "Genome-Wide Polygenic Scores for Common Diseases Identify Individuals with Risk Equivalent to Monogenic Mutations." *Nature Genetics* 50 (9): 1219–24. https://doi.org/10.1038/s41588-018-0183-z.

Kim, Sahn-ho, Patrick Kaminker, and Judith Campisi. 1999. "TIN2, a New Regulator of Telomere Length in Human Cells." *Nature Genetics* 23 (4): 405–12. https://doi.org/10.1038/70508.

Kitano, Masahiro, Saya Moriyama, Yoshikazu Ando, Masaki Hikida, Yasuo Mori, Tomohiro Kurosaki, and Takaharu Okada. 2011. "Bcl6 Protein Expression Shapes Pre-Germinal Center B Cell Dynamics and Follicular Helper T Cell Heterogeneity." *Immunity* 34 (6): 961–72. https://doi.org/10.1016/j.immuni.2011.03.025.

Ko, Kichul, Yelena Koldobskaya, Elizabeth Rosenzweig, and Timothy B Niewold. 2013. "Activation of the Interferon Pathway Is Dependent Upon Autoantibodies in African-American SLE Patients, but Not in European-American SLE Patients."

Frontiers in Immunology 4 (October): 1–9. https://doi.org/10.3389/fimmu.2013.00309.

Kobayashi, K. S., and P. J. van den Elsen. 2012. "NLRC5: A Key Regulator of MHC Class I-Dependent Immune Responses." *Nature Reviews Immunology* 12: 813–20.

Köttgen, Anna, Cristian Pattaro, Carsten A Böger, Christian Fuchsberger, Matthias Olden, Nicole L Glazer, Afshin Parsa, et al. 2010. "Multiple Loci Associated with Indices of Renal Function and Chronic Kidney Disease." *Nature Genetics* 42 (5): 376–84. https://doi.org/10.1038/ng.568.Multiple.

Krapohl, E., J. Euesden, D. Zabaneh, J. B. Pingault, K. Rimfeld, S. Von Stumm, P. S. Dale, G. Breen, P. F. O'Reilly, and R. Plomin. 2016. "Phenome-Wide Analysis of Genome-Wide Polygenic Scores." *Molecular Psychiatry* 21 (9): 1188–93. https://doi.org/10.1038/mp.2015.126.

Krishnan, Vaidehi, Baohua Liu, and Zhongjun Zhou. 2011. "DNA Repair, Human Diseases and Aging." *Intech*, 13. https://doi.org/http://dx.doi.org/10.5772/57353.

Kujoth, G. C., A Hiona, T D Pugh, S Someya, K Panzer, S E Wohlgemuth, T Hofer, et al. 2005. "Mitochondrial DNA Mutations, Oxidative Stress, and Apoptosis in Mammalian Aging." *Science (New York, N.Y.)* 309 (5733): 481–84. https://doi.org/10.1126/science.1112125.

Landolt-Marticorena, C, G Bonventi, a Lubovich, C Ferguson, T Unnithan, J Su, D D Gladman, M Urowitz, P R Fortin, and J Wither. 2009. "Lack of Association between the Interferon-α Signature and Longitudinal Changes in Disease Activity in Systemic Lupus Erythematosus." *Annals of the Rheumatic Diseases* 68 (9): 1440–46. https://doi.org/10.1136/ard.2008.093146.

Langefeld, Carl D., Hannah C. Ainsworth, Deborah S. Cunninghame Graham, Jennifer A. Kelly, Mary E. Comeau, Miranda C. Marion, Timothy D. Howard, et al. 2017. "Transancestral Mapping and Genetic Load in Systemic Lupus Erythematosus." *Nature Communications* 8 (May): 16021. https://doi.org/10.1038/ncomms16021.

Laurenti, Elisa, and Berthold Göttgens. 2018. "From Haematopoietic Stem Cells to Complex Differentiation Landscapes." *Nature*. https://doi.org/10.1038/nature25022.

Leslie, Richard, Christopher J. O'Donnell, and Andrew D. Johnson. 2014. "GRASP: Analysis of Genotype-Phenotype Results from 1390 Genome-Wide Association Studies and Corresponding Open Access Database." *Bioinformatics* 30 (12): 185–94. https://doi.org/10.1093/bioinformatics/btu273.

Lessard, Christopher J., Indra Adrianto, John A. Ice, Graham B. Wiley, Jennifer A. Kelly, Stuart B. Glenn, Adam J. Adler, et al. 2012. "Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as Susceptibility Loci for Systemic Lupus Erythematosus in a Large-Scale Multiracial Replication Study." *American Journal of Human Genetics* 90 (4): 648–60. https://doi.org/10.1016/j.ajhg.2012.02.023.

Leuchten, N., B. Milke, B. Winkler-Rohlfing, D. Daikh, T. Dörner, S. R. Johnson, and M. Aringer. 2018. "Early Symptoms of Systemic Lupus Erythematosus (SLE) Recalled by 339 SLE Patients." *Lupus*, 1431–36. https://doi.org/10.1177/0961203318776093.

Lewis, Myles J., and Ali S. Jawad. 2017. "The Effect of Ethnicity and Genetic Ancestry on the Epidemiology, Clinical Features and Outcome of Systemic Lupus Erythematosus." *Rheumatology (Oxford, England)* 56 (1): i67–77. https://doi.org/10.1093/rheumatology/kew399.

Li, Heng, and Richard Durbin. 2009. "Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform." *Bioinformatics* 25 (14): 1754–60. https://doi.org/10.1093/bioinformatics/btp324.

Li, Q.-Z., J Zhou, Y Lian, B Zhang, V K Branch, D R Karp, C Mohan, E K Wakeland, and N J Olsen. 2009. "Interferon Signature Gene Expression Is Correlated with Autoantibody Profiles in Patients with Incomplete Lupus Syndromes," 281–91. https://doi.org/10.1111/j.1365-2249.2009.04057.x.

Li, Rui, Pierre Cauchy, Senthilkumar Ramamoorthy, Sören Boller, Lukas Chavez, and Rudolf Grosschedl. 2018. "Dynamic EBF1 Occupancy Directs Sequential Epigenetic and Transcriptional Events in B-Cell Programming." *Genes and Development* 32 (2): 96–111. https://doi.org/10.1101/gad.309583.117.

Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest." *The Newsletter of the R Project*, 2002. https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf.

Lin, Haishan, and Rudolf Grosschedl. 1995. "Failure of B–Cell Differentiation in Mice Lacking the Transcription Factor EBF." *Nature* 376 (6537): 263–67. https://doi.org/10.1038/376263a0.

Lin, Yin C., Suchit Jhunjhunwala, Christopher Benner, Sven Heinz, Eva Welinder, Robert Mansson, Mikael Sigvardsson, et al. 2010. "A Global Network of Transcription Factors, Involving E2A, EBF1 and Foxo1, That Orchestrates B Cell Fate." *Nature Immunology* 11 (7): 635–43. https://doi.org/10.1038/ni.1891.

Liu, X. Shawn, Hao Wu, Xiong Ji, Yonatan Stelzer, Xuebing Wu, Szymon Czauderna, Jian Shu, Daniel Dadon, Richard A. Young, and Rudolf Jaenisch. 2016. "Editing

DNA Methylation in the Mammalian Genome." *Cell* 167 (1): 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056.

Lodolce, James P., Lauren E. Kolodziej, Lesley Rhee, Silvia N. Kariuki, Beverly S. Franek, Nancy M. McGreal, Mark F. Logsdon, et al. 2010. "African-Derived Genetic Polymorphisms in TNFAIP3 Mediate Risk for Autoimmunity James." *J Immunol.* 184 (12): 997–1003. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted.

Lonsdale, John, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, et al. 2013. "The Genotype-Tissue Expression (GTEx) Project." *Nature Genetics* 45 (6): 580–85. https://doi.org/10.1038/ng.2653.

Lopez-Mejia, Isabel C., Valentin Vautrot, Marion De toledo, Isabelle Behm-ansmant, Cyril F. Bourgeois, Claire L. Navarro, Fernando G. Osorio, et al. 2011. "A Conserved Splicing Mechanism of the LMNA Gene Controls Premature Aging." *Human Molecular Genetics* 20 (23): 4540–55. https://doi.org/10.1093/hmg/ddr385.

Lugar, Patricia L., Cassandra Love, Amrie C. Grammer, Sandeep S. Dave, and Peter E. Lipsky. 2012. "Molecular Characterization of Circulating Plasma Cells in Patients with Active Systemic Lupus Erythematosus." *PLoS ONE* 7 (9): 1–15. https://doi.org/10.1371/journal.pone.0044362.

Lunetta, Kathryn L, L Brooke Hayward, Jonathan Segal, and Paul Van Eerdewegh. 2004. "Screening Large-Scale Association Study Data: Exploiting Interactions Using Random Forests." *BMC Genetics* 5 (1): 32. https://doi.org/10.1186/1471-2156-5-32.

Macrae, Sheila L., Quanwei Zhang, Christophe Lemetre, Inge Seim, Robert B. Calder, Jan Hoeijmakers, Yousin Suh, et al. 2015. "Comparative Analysis of Genome

Maintenance Genes in Naked Mole Rat, Mouse, and Human." *Aging Cell* 14 (2): 288–91. https://doi.org/10.1111/acel.12314.

Mali, P., Luhan Yang, Kevin M Esvelt, John Aach, Marc Guell, J. E. DiCarlo, Julie E Norville, and George M Church. 2013. "RNA-Guided Human Genome Engineering via Cas9." *Science* 339 (6121): 823–26. https://doi.org/10.1126/science.1232033.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, et al. 2009. "Finding the Missing Heritability of Complex Diseases." *Nature* 461 (7265): 747–53. https://doi.org/10.1038/nature08494.

Manolio, Teri A., Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorr, David J. Hunter, Mark I. McCarthy, et al. 2009. "Finding the Missing Heritability of Complex Diseases Teri." *Nature Reviews* 461 (8): 747–53. https://doi.org/10.1038/nature08494.

Mao, Zhiyong, Chris Hine, Xiao TIan, Micheal Van Meter, Matthew Au, Amita Vaidya, Anrdrei Seluanov, and Vera Gorbunova. 2011. "SIRT6 Promotes DNA Repair under Stress by Activating PARP1 Zhiyong." *Science* 332: 1443–46. https://doi.org/10.1126/science.1202723.

Martin, George M., Aviv Bergman, and Nir Barzilai. 2007. "Genetic Determinants of Human Health Span and Life Span: Progress and New Opportunities." *PLoS Genetics* 3 (7): 1121–30. https://doi.org/10.1371/journal.pgen.0030125.

Martinez, Paula, and Maria A. Blasco. 2010. "Role of Shelterin in Cancer and Aging." *Aging Cell* 9 (5): 653–66. https://doi.org/10.1111/j.1474-9726.2010.00596.x.

Mathian, Alexis, Miguel Hie, Fleur Cohen-Aubart, and Zahir Amoura. 2015. "Targeting

Interferons in Systemic Lupus Erythematosus: Current and Future Prospects." *Drugs* 75 (8): 835–46. https://doi.org/10.1007/s40265-015-0394-x.

Matzaraki, Vasiliki, Vinod Kumar, Cisca Wijmenga, and Alexandra Zhernakova. 2017. "The MHC Locus and Genetic Susceptibility to Autoimmune and Infectious Diseases." *Genome Biology* 18 (1). https://doi.org/10.1186/s13059-017-1207-1.

Maynard, Scott, Evandro Fei Fang, Morten Scheibye-Knudsen, Deborah L. Croteau, and Vilhelm A Bohr. 2015. "DNA Damage, DNA Repair, Aging, and Neurodegeneration." *Cold Spring Harbor Perspectives in Medicine* 5 (10): a025130. https://doi.org/10.1101/cshperspect.a025130.

McClung, Michael R., Steven Boonen, Ove Törring, Christian Roux, René Rizzoli, Henry G. Bone, Claude Laurent Benhamou, et al. 2012. "Effect of Denosumab Treatment on the Risk of Fractures in Subgroups of Women with Postmenopausal Osteoporosis." *Journal of Bone and Mineral Research* 27 (1): 211–18. https://doi.org/10.1002/jbmr.536.

McDermott, Jason E., Keri B. Vartanian, Hugh Mitchell, Susan L. Stevens, Antonio Sanfilippo, and Mary P. Stenzel-Poore. 2012. "Identification and Validation of Ifit1 as an Important Innate Immune Bottleneck." *PLoS ONE* 7 (6): 1–8. https://doi.org/10.1371/journal.pone.0036465.

McGhee, James D., and Gordon D. Ginder. 1979. "Specific DNA Methylation Sites in the Vicinity of the Chicken Beta-Globin Genes" 280: 419–20.

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulksis, Andrew Kernytsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data."

Genome Research 20: 1297–1303. https://doi.org/10.1101/gr.107524.110.20.

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). n.d. "Online Mendelian Inheritance in Man, OMIM®." Accessed May 6, 2019. https://omim.org/.

Meissner, T. B., A. Li, A. Biswas, K.-H. Lee, Y.-J. Liu, E. Bayir, D. Iliopoulos, P. J. van den Elsen, and K. S. Kobayashi. 2010. "NLR Family Member NLRC5 Is a Transcriptional Regulator of MHC Class I Genes." *Proceedings of the National Academy of Sciences* 107 (31): 13794–99. https://doi.org/10.1073/pnas.1008684107.

Melchers, Fritz, and Fritz Melchers. 2015. "Checkpoints That Control B Cell Development Find the Latest Version : Checkpoints That Control B Cell Development" 125 (6): 2203–10. https://doi.org/10.1172/JCI78083.In.

Melén, Krister, Tapani Ronni, Timo Lotta, and Ilkka Julkunen. 1994. "Enzymatic Characterization of Interferon-Induced Antiviral GTPases Murine Mx1 and Human MxA Proteins." *Journal of Biological Chemistry* 269 (3): 2009–15.

Menard, Laurence C., Sium Habte, Waldemar Gonsiorek, Deborah Lee, Dana Banas, Deborah A. Holloway, Nataly Manjarrez-Orduno, et al. 2016. "B Cells from African American Lupus Patients Exhibit an Activated Phenotype." *JCI Insight* 1 (9): 1–14. https://doi.org/10.1172/jci.insight.87310.

Mengel-Jørgensen, J., H. Eiberg, C. Børsting, and N. Morling. 2006. "Genetic Screening of 15 SNPs in the MC1R Gene in Relation to Hair Colour in Danes." *International Congress Series* 1288: 55–57. https://doi.org/10.1016/j.ics.2005.11.037.

Merrill, Joan T., C. Michael Neuwelt, Daniel J. Wallace, Joseph C. Shanahan, Kevin M. Latinis, James C. Oates, Tammy O. Utset, et al. 2010. "Efficacy and Safety of

Rituximab in Moderately-to-Severely Active Systemic Lupus Erythematosus: The Randomized, Double-Blind, Phase II/III Systemic Lupus Erythematosus Evaluation of Rituximab Trial." *Arthritis and Rheumatism* 62 (1): 222–33. https://doi.org/10.1002/art.27233.

Mohaghegh, Payam, and Ian D. Hickson. 2002. "Premature Aging in RecQ Helicase-Deficient Human Syndromes." *International Journal of Biochemistry and Cell Biology* 34 (11): 1496–1501. https://doi.org/10.1016/S1357-2725(02)00039-0.

Mohan, Chandra, and Chaim Putterman. 2015a. "Genetics and Pathogenesis of Systemic Lupus Erythematosus and Lupus Nephritis." Nature Publishing Group.

———. 2015b. "Genetics and Pathogenesis of Systemic Lupus Erythematosus and Lupus Nephritis." *Nature Reviews Nephrology* 11 (6): 329–41. https://doi.org/10.1038/nrneph.2015.33.

Mok, Amanda, Olivia Solomon, Renuka R Nayak, Patrick Coit, Hong L Quach, Joanne Nititham, Amr H Sawalha, Lisa F Barcellos, Lindsey A Criswell, and Sharon A Chung. 2016. "Genome-Wide Profiling Identifies Associations between Lupus Nephritis and Differential Methylation of Genes Regulating Tissue Hypoxia and Type 1 Interferon Responses." *Lupus Science & Medicine* 3 (1): e000183. https://doi.org/10.1136/lupus-2016-000183.

Morawski, Peter A., and Silvia Bolland. 2017. "Expanding the B Cell-Centric View of Systemic Lupus Erythematosus." *Trends in Immunology* 38 (5): 373–82. https://doi.org/10.1016/j.it.2017.02.001.

Moskalev, Alexey A., Alexander M. Aliper, Zeljka Smit-McBride, Anton Buzdin, and Alex Zhavoronkov. 2014. "Genetics and Epigenetics of Aging and Longevity." *Cell*

Cycle 13 (7): 1063–77. https://doi.org/10.4161/cc.28433.

Muiras, M L, M Müller, F Schächter, and A Bürkle. 1998. "Increased Poly(ADP-Ribose) Polymerase Activity in Lymphoblastoid Cell Lines from Centenarians." *Journal of Molecular Medicine (Berlin, Germany)* 76 (5): 346–54. https://doi.org/10.1007/s001090050226.

Multani, Asha S, and Sandy Chang. 2007. "WRN at Telomeres: Implications for Aging and Cancer." *Journal of Cell Science* 120 (Pt 5): 713–21. https://doi.org/10.1242/jcs.03397.

Muñoz, Purificación, Raquel Blanco, Juana M. Flores, and María A. Blasco. 2005. "XPF Nuclease-Dependent Telomere Loss and Increased DNA Damage in Mice Overexpressing TRF2 Result in Premature Aging and Cancer." *Nature Genetics* 37 (10): 1063–71. https://doi.org/10.1038/ng1633.

Nakayama, M., K. Nagata, A. Kato, and A. Ishihama. 1991. "Interferon-Inducible Mouse Mx1 Protein That Confers Resistance to Influenza Virus Is GTPase." *Journal of Biological Chemistry* 266 (32): 21404–8. https://doi.org/10.1089/brain.2011.0064.

National Institute of General Medical Sciences. 2010. "National Institute of General Medical Sciences Publicatoins: The New Genetics," no. 10–662: 98. http://www.nigms.nih.gov.

Nechanitzky, Robert, Duygu Akbas, Stefanie Scherer, Ildiko Györy, Thomas Hoyler, Senthilkumar Ramamoorthy, Andreas Diefenbach, and Rudolf Grosschedl. 2013. "Transcription Factor EBF1 Is Essential for the Maintenance of B Cell Identity and Prevention of Alternative Fates in Committed Cells." *Nature Immunology* 14 (8): 867–75. https://doi.org/10.1038/ni.2641.

Ng, Pauline C., and Steven Henikoff. 2003. "SIFT: Predicting Amino Acid Changes That Affect Protein Function." *Nucleic Acids Research* 31 (13): 3812–14. https://doi.org/10.1093/nar/gkg509.

Nho, K., J. J. Corneveaux, S. Kim, H. Lin, S. L. Risacher, L. Shen, S. Swaminathan, et al. 2013. "Whole-Exome Sequencing and Imaging Genetics Identify Functional Variants for Rate of Change in Hippocampal Volume in Mild Cognitive Impairment." *Molecular Psychiatry* 18 (7): 781–87. https://doi.org/10.1038/mp.2013.24.

Niedernhofer, Laura J., George A. Garinis, Anja Raams, Astrid S. Lalai, Andria Rasile Robinson, Esther Appeldoorn, Hanny Odijk, et al. 2006. "A New Progeroid Syndrome Reveals That Genotoxic Stress Suppresses the Somatotroph Axis." *Nature* 444 (7122): 1038–43. https://doi.org/10.1038/nature05456.

Niedernhofer, Laura J, Aditi U. Gurkar, Yinsheng Wang, Jan Vijg, Jan H J Hoeijmakers, and Paul D. Robbins. 2018. "Nuclear Genomic Instability and Aging." *Annual Review of Biochemistry* 87: 295–322. https://doi.org/10.1146/annurev-biochem-062917-012239.

Niewold, T. B., J. Hua, T. J A Lehman, J. B. Harley, and M. K. Crow. 2007. "High Serum IFN-α Activity Is a Heritable Risk Factor for Systemic Lupus Erythematosus." *Genes and Immunity* 8 (6): 492–502. https://doi.org/10.1038/sj.gene.6364408.

Niewold, Timothy B. 2008. "Interferon Alpha-Induced Lupus." *JCR: Journal of Clinical Rheumatology* 14 (3): 131–32. https://doi.org/10.1097/RHU.0b013e318177627d.

Niewold, Timothy B., Jennifer A. Kelly, Marie H. Flesch, Luis R. Espinoza, John B.

Harley, and Mary K. Crow. 2008. "Association of the IRF5 Risk Haplotype with High Serum Interferon-α Activity in Systemic Lupus Erythematosus Patients." *Arthritis and Rheumatism* 58 (8): 2481–87. https://doi.org/10.1002/art.23613.

Niewold, Timothy B., and William I. Swedler. 2005. "Systemic Lupus Erythematosus Arising during Interferon-Alpha Therapy for Cryoglobulinemic Vasculitis Associated with Hepatitis C." *Clinical Rheumatology* 24 (2): 178–81. https://doi.org/10.1007/s10067-004-1024-2.

Noren Hooten, Nicole, Megan Fitzpatrick, Kari Kompaniez, Kimberly D. Jacob, Brittany R. Moore, Julia Nagle, Janice Barnes, Althaf Lohani, and Michele K. Evans. 2012. "Coordination of DNA Repair by NEIL1 and PARP-1: A Possible Link to Aging." *Aging* 4 (10): 674–85. https://doi.org/10.18632/aging.100492.

Noviski, Mark, James L Mueller, Anne Satterthwaite, Lee Ann Garrett-Sinha, Frank Brombacher, and Julie Zikherman. 2018. "IgM and IgD B Cell Receptors Differentially Respond to Endogenous Antigens and Control B Cell Fate." *ELife* 7: 1–29. https://doi.org/10.7554/elife.35074.

O'Donovan, Anne, Adelina A. Davies, Jonathan G. Moggs, Stephen C. West, and Richard D. Wood. 1994. "XPG Endonuclease Makes the 3′ Incision in Human DNA Nucleotide Excision Repair." *Nature*. https://doi.org/10.1038/371432a0.

Oakes, Christopher C., Marc Seifert, Yassen Assenov, Lei Gu, Martina Przekopowitz, Amy S. Ruppert, Qi Wang, et al. 2016. "DNA Methylation Dynamics during B Cell Maturation Underlie a Continuum of Disease Phenotypes in Chronic Lymphocytic Leukemia." *Nature Genetics* 48 (3): 253–64. https://doi.org/10.1038/ng.3488.

Olsson, Anders H., Petr Volkov, Karl Bacos, Tasnim Dayeh, Elin Hall, Emma A.

Nilsson, Claes Ladenvall, Tina Rönn, and Charlotte Ling. 2014. "Genome-Wide Associations between Genetic and Epigenetic Variation Influence MRNA Expression and Insulin Secretion in Human Pancreatic Islets." Edited by John M. Greally. *PLoS Genetics* 10 (11): e1004735. https://doi.org/10.1371/journal.pgen.1004735.

Opresko, Patricia L., and Jerry W. Shay. 2017. "Telomere-Associated Aging Disorders." *Ageing Research Reviews* 33: 52–66. https://doi.org/10.1016/j.arr.2016.05.009.

Orkin, Stuart H, and Leonard I Zon. 2008. "Hematopoiesis: An Evolving Paradigm for Stem Cell Biology." *Cell* 132 (4): 631–44. https://doi.org/10.1016/j.cell.2008.01.025.

Pancaldi, Vera, Guillem Clot, Reiner Siebert, Nuria Russiñol, Martí Duran-Ferrer, Néria Verdaguer-Dot, Angelika Merkel, et al. 2015. "Whole-Genome Fingerprint of the DNA Methylome during Human B Cell Differentiation." *Nature Genetics* 47 (7): 746–56. https://doi.org/10.1038/ng.3291.

Pang, Swee Heng Milon, Carolyn A. de Graaf, Douglas J. Hilton, Nicholas D. Huntington, Sebastian Carotta, Li Wu, and Stephen L. Nutt. 2018. "PU.1 Is Required for the Developmental Progression of Multipotent Progenitors to Common Lymphoid Progenitors." *Frontiers in Immunology* 9 (JUN). https://doi.org/10.3389/fimmu.2018.01264.

Papp, Krisztián, Péter Végh, Renáta Hóbor, Zoltán Szittner, Zoltán Vokó, János Podani, László Czirják, and József Prechl. 2012. "Immune Complex Signatures of Patients with Active and Inactive SLE Revealed by Multiplex Protein Binding Analysis on Antigen Microarrays." *PLoS ONE* 7 (9): 1–10.

https://doi.org/10.1371/journal.pone.0044824.

Park, Sung Ho, Kyuho Kang, Eugenia Giannopoulou, Yu Qiao, Keunsoo Kang, Geonho Kim, Kyung Hyun Park-Min, and Lionel B. Ivashkiv. 2017. "Type i Interferons and the Cytokine TNF Cooperatively Reprogram the Macrophage Epigenome to Promote Inflammatory Activation." *Nature Immunology* 18 (10): 1104–16. https://doi.org/10.1038/ni.3818.

Patel, Manisha N. 2002. "Oxidative Stress, Mitochondrial Dysfunction, and Epilepsy." *Free Radical Research* 36 (11): 1139–46. https://doi.org/10.1080/1071576021000016391.

Pelikan, Richard C., Jennifer A. Kelly, Yao Fu, Caleb A. Lareau, Kandice L. Tessneer, Graham B. Wiley, Mandi M. Wiley, et al. 2018. "Enhancer Histone-QTLs Are Enriched on Autoimmune Risk Haplotypes and Influence Gene Expression within Chromatin Networks." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-05328-9.

Perera, Subashan, Kushang V. Patel, Caterina Rosano, Susan M. Rubin, Suzanne Satterfield, Tamara Harris, Kristine Ensrud, et al. 2016. "Gait Speed Predicts Incident Disability: A Pooled Analysis." *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 71 (1): 63–71. https://doi.org/10.1093/gerona/glv126.

Perera, Subashan, Stephanie Studenski, Anne Newman, Eleanor Simonsick, Tamara Harris, Ann Schwartz, and Marjolein Visser. 2014. "Are Estimates of Meaningful Decline in Mobility Performance Consistent among Clinically Important Subgroups? (Health ABC Study)." *Journals of Gerontology - Series A Biological*

Sciences and Medical Sciences 69 (10): 1260–68. https://doi.org/10.1093/gerona/glu033.

Perlis, Rh, J Huang, and S Purcell. 2010. "Genome-Wide Association Study of Suicide Attempts in Mood Disorder Patients." *American Journal of ...* 167 (12): 1499–1507. https://doi.org/10.1176/appi.ajp.2010.10040541.Genome-Wide.

Petri, M, S Singh, H Tesfasyone, R Dedrick, K Fry, PG Lal, G Williams, et al. 2009. "Longitudinal Expression of Type I Interferon Responsive Genes in Systemic Lupus Erythematosus." *Lupus* 18 (11): 980–89. https://doi.org/10.1177/0961203309105529.

Pilling, Luke C., Chia Ling Kuo, Kamil Sicinski, Jone Tamosauskaite, George A. Kuchel, Lorna W. Harries, Pamela Herd, Robert Wallace, Luigi Ferrucci, and David Melzer. 2017. "Human Longevity: 25 Genetic Loci Associated in 389,166 UK Biobank Participants." *Aging* 9 (12): 2504–20. https://doi.org/10.18632/aging.101334.

Pilling, Luke C, Janice L Atkins, Kirsty Bowman, Samuel E Jones, Jessica Tyrrell, Robin N Beaumont, Katherine S Ruth, et al. 2016. "Human Longevity Is Influenced by Many Genetic Variants: Evidence from 75,000 UK Biobank Participants." *Aging* 8 (3): 1–24. https://doi.org/10.1101/038430.

Piskunova, Tatiana S., Maria N. Yurova, Anton I. Ovsyannikov, Anna V. Semenchenko, Mark A. Zabezhinski, Irina G. Popovich, Zhao-Qi Wang, and Vladimir N. Anisimov. 2008. "Deficiency in Poly(ADP-Ribose) Polymerase-1 (PARP-1) Accelerates Aging and Spontaneous Carcinogenesis in Mice." *Current Gerontology and Geriatrics Research* 2008: 1–11. https://doi.org/10.1155/2008/754190.

Powell, Simon N., and Lisa A. Kachnic. 2003. "Roles of BRCA1 and BRCA2 in

Homologous Recombination, DNA Replication Fidelity and the Cellular Response to Ionizing Radiation." *Oncogene* 22 (37 REV. ISS. 3): 5784–91. https://doi.org/10.1038/sj.onc.1206678.

Qiu, Xiaolei, Katharine Brown, Matthew D. Hirschey, Eric Verdin, and Danica Chen. 2010. "Calorie Restriction Reduces Oxidative Stress by SIRT3-Mediated SOD2 Activation." *Cell Metabolism* 12 (6): 662–67. https://doi.org/10.1016/j.cmet.2010.11.015.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.

Rajewsky, Klaus. 1996. "Clonal Selection and Learning in the Antibody System." *Nature* 381 (6585): 751–58. https://doi.org/10.1038/381751a0.

Recaldin, T., and D. J. Fear. 2016. "Transcription Factors Regulating B Cell Fate in the Germinal Centre." *Clinical and Experimental Immunology* 183 (1): 65–75. https://doi.org/10.1111/cei.12702.

Reed, Terry, Danielle M Dick, Sean K Uniacke, Tatiana Foroud, and William C Nichols. 2004. "Genome-Wide Scan for a Healthy Aging Phenotype Provides Support for a Locus near D4S1564 Promoting Healthy Aging." *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences* 59 (3): 227–32.

Rekvig, Ole Petter. 2018. "Systemic Lupus Erythematosus: Definitions, Contexts, Conflicts, Enigmas." *Frontiers in Immunology* 9 (MAR): 1–16. https://doi.org/10.3389/fimmu.2018.00387.

Relton, Caroline L., and George Davey Smith. 2010. "Epigenetic Epidemiology of

Common Complex Disease: Prospects for Prediction, Prevention, and Treatment." *PLoS Medicine* 7 (10). https://doi.org/10.1371/journal.pmed.1000356.

Renty, Christelle de, and Nathan A. Ellis. 2017. "Bloom's Syndrome: Why Not Premature Aging?" *Ageing Research Reviews* 33 (24): 36–51. https://doi.org/10.1016/j.arr.2016.05.010.

Renty, Christelle de, and Nathan A Ellis. 2016. "Bloom's Syndrome: Why Not Premature Aging?: A Comparison of the BLM and WRN Helicases." *Ageing Research Reviews* 33: 36–51. https://doi.org/10.1016/j.arr.2016.05.010.

Rhead, Brooke, Calliope Holingue, Michael Cole, Xiaorong Shao, Hong L. Quach, Diana Quach, Khooshbu Shah, et al. 2017. "Rheumatoid Arthritis Naive T Cells Share Hypermethylation Sites With Synoviocytes." *Arthritis and Rheumatology* 69 (3): 550–59. https://doi.org/10.1002/art.39952.

Rigler, Sally K., Stephanie Studenski, Dennis Wallace, Dean M. Reker, and Pamela W. Duncan. 2002. "Co-Morbidity Adjustment for Functional Outcomes in Community-Dwelling Older Adults." *Clinical Rehabilitation* 16 (4): 420–28. https://doi.org/10.1191/0269215502cr515oa.

Robertson, Keith D. 2002. "DNA Methylation and Chromatin - Unraveling the Tangled Web." *Oncogene* 21 (35 REV. ISS. 3): 5361–79. https://doi.org/10.1038/sj.onc.1205609.

Robinson, T., S. N. Kariuki, B. S. Franek, M. Kumabe, A. A. Kumar, M. Badaracco, R. A. Mikolaitis, et al. 2011. "Autoimmune Disease Risk Variant of IFIH1 Is Associated with Increased Sensitivity to IFN- and Serologic Autoimmunity in Lupus Patients." *The Journal of Immunology* 187 (3): 1298–1303.

https://doi.org/10.4049/jimmunol.1100857.

Rodriguez, Sofia, Fabio Coppedè, Hanna Sagelius, and Maria Eriksson. 2009. "Increased Expression of the Hutchinson-Gilford Progeria Syndrome Truncated Lamin A Transcript during Cell Aging." *European Journal of Human Genetics : EJHG* 17 (7): 928–37. https://doi.org/10.1038/ejhg.2008.270.

Rubelj, Ivica, and Zoran Vondraček. 1999. "Stochastic Mechanism of Cellular Aging - Abrupt Telomere Shortening as a Model for Stochastic Nature of Cellular Aging." *Journal of Theoretical Biology* 197 (4): 425–38. https://doi.org/10.1006/jtbi.1998.0886.

Rullo, Ornella J, Jennifer M P Woo, Hui Wu, Alice D C Hoftman, Paul Maranian, A Brahn, Deborah Mccurdy, Rita M Cantor, and Betty P Tsao. 2011. "Association of IRF5 Polymorphisms with Activation of the Interferon α Pathway" 69 (3): 611–17. https://doi.org/10.1136/ard.2009.118315.Association.

Sánchez, Elena, Mary E. Comeau, Barry I. Freedman, Jennifer A. Kelly, Kenneth M. Kaufman, Carl D. Langefeld, Elizabeth E. Brown, et al. 2011. "Identification of Novel Genetic Susceptibility Loci in African American Lupus Patients in a Candidate Gene Association Study." *Arthritis and Rheumatism* 63 (11): 3493–3501. https://doi.org/10.1002/art.30563.

Sangha, Oliver, Gerold Stucki, Matthew H. Liang, Anne H. Fossel, and Jeffrey N. Katz. 2003. "The Self-Administered Comorbidity Questionnaire: A New Method to Assess Comorbidity for Clinical and Health Services Research." *Arthritis & Rheumatism* 49 (2): 156–63. https://doi.org/10.1002/art.10993.

Santos-Lozano, Alejandro, Ana Santamarina, Helios Pareja-Galeano, Fabian Sanchis-

Gomar, Carmen Fiuza-Luces, Carlos Cristi-Montero, Aranzazu Bernal-Pino, Alejandro Lucia, and Nuria Garatachea. 2016. "The Genetics of Exceptional Longevity: Insights from Centenarians." *Maturitas* 90: 49–57. https://doi.org/10.1016/j.maturitas.2016.05.006.

Santos, Patricia, Fortuna Arumemi, Kyung Soo Park, Lisa Borghesi, and Christine Milcarek. 2011. "Transcriptional and Epigenetic Regulation of B Cell Development." *Immunologic Research* 50 (2–3): 105–12. https://doi.org/10.1007/s12026-011-8225-y.

Satoh, Akiko, Cynthia S. Brace, Nick Rensing, Paul Cliften, David F. Wozniak, Erik D. Herzog, Kelvin A. Yamada, and Shin-ichiro Imai. 2013. "Sirt1 Extends Life Span and Delays Aging in Mice through the Regulation of Nk2 Homeobox 1 in the DMH and LH." *Cell Metabolism* 18 (3): 416–30. https://doi.org/10.1016/j.cmet.2013.07.013.

Savage, Sharon A., Neelam Giri, Gabriela M. Baerlocher, Nick Orr, Peter M. Lansdorp, and Blanche P. Alter. 2008. "TINF2, a Component of the Shelterin Telomere Protection Complex, Is Mutated in Dyskeratosis Congenita." *American Journal of Human Genetics* 82 (2): 501–9. https://doi.org/10.1016/j.ajhg.2007.10.004.

Schumacher, Björn, George A. Garinis, and Jan H J Hoeijmakers. 2008. "Age to Survive: DNA Damage and Aging." *Trends in Genetics* 24 (2): 77–85. https://doi.org/10.1016/j.tig.2007.11.004.

Sebastiani, Paola, Harold Bae, Fangui X. Sun, Stacy L. Andersen, E. Warwick Daw, Alberto Malovini, Toshio Kojima, et al. 2013. "Meta-Analysis of Genetic Variants Associated with Human Exceptional Longevity." *Aging* 5 (9): 653–61.

https://doi.org/10.18632/aging.100594.

Sebastiani, Paola, Nadia Solovieff, Andrew T. DeWan, Kyle M. Walsh, Annibale Puca, Stephen W. Hartley, Efthymia Melista, et al. 2012. "Genetic Signatures of Exceptional Longevity in Humans." *PLoS ONE* 7 (1). https://doi.org/10.1371/journal.pone.0029848.

Seifert, M., and R. Küppers. 2016. "Human Memory B Cells." *Leukemia* 30 (12): 2283–92. https://doi.org/10.1038/leu.2016.226.

Semerad, Craig L., Elinore M. Mercer, Matthew A. Inlay, Irving L. Weissman, and Cornelis Murre. 2009. "E2A Proteins Maintain the Hematopoietic Stem Cell Pool and Promote the Maturation of Myelolymphoid and Myeloerythroid Progenitors." *Proceedings of the National Academy of Sciences* 106 (6): 1930–35. https://doi.org/10.1073/pnas.0808866106.

Serrano, Manuel, Carlos López-Otín, Maria A. Blasco, Guido Kroemer, and Linda Partridge. 2013. "The Hallmarks of Aging." *Cell* 153 (6): 1194–1217. https://doi.org/10.1016/j.cell.2013.05.039.

Shadyab, Aladdin H., and Andrea Z. LaCroix. 2015. "Genetic Factors Associated with Longevity: A Review of Recent Findings." *Ageing Research Reviews* 19: 1–7. https://doi.org/10.1016/j.arr.2014.10.005.

Shapiro-Shelef, Miriam, and Kathryn Calame Calame. 2005. "Regulation of Plasma-Cell Development." *Nature Reviews Immunology*. https://doi.org/10.1038/nri1572.

Shiels, Paul G. 2010. "Improving Precision in Investigating Aging: Why Telomeres Can Cause Problems." *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 65 A (8): 789–91. https://doi.org/10.1093/gerona/glq095.

Shimizu, Yuka, Shinsuke Yasuda, Taichi Kimura, Saori Nishio, Michihiro Kono, Kazumasa Ohmura, Sanae Shimamura, et al. 2017. "Interferon-Inducible Mx1 Protein Is Highly Expressed in Renal Tissues from Treatment-Naïve Lupus Nephritis, but Not in Those under Immunosuppressive Treatment." *Modern Rheumatology* 0 (0): 1–9. https://doi.org/10.1080/14397595.2017.1404711.

Shlomchik, Mark J. 2008. "Sites and Stages of Autoreactive B Cell Activation and Regulation." *Immunity* 28 (1): 18–28. https://doi.org/10.1016/j.immuni.2007.12.004.

Siponen, Marina I, Magdalena Wisniewska, Lari Lehtiö, Ida Johansson, Linda Svensson, Grzegorz Raszewski, Lennart Nilsson, Mikael Sigvardsson, and Helena Berglund. 2010. "Structural Determination of Functional Domains in Early B-Cell Factor (EBF) Family of Transcription Factors Reveals Similarities to Rel DNA-Binding Proteins and a Novel Dimerization Motif." *Journal of Biological Chemistry* 285 (34): 25875–79. https://doi.org/10.1074/jbc.C110.150482.

Somers, EC Marder, W Cagnoli, P Lewis, EE, and C Deguire, P Gordon. 2014. "Population-Based Incidence and Prevalence of Systemic Lupus Erythematosus." *Arthritis Rheumatol.* 66 (2): 369–78. https://doi.org/10.1002/art.38238.Population-Based.

Soria-Valles, Clara, and Carlos López-Otín. 2016. "IPSCs: On the Road to Reprogramming Aging." *Trends in Molecular Medicine* 22 (8): 713–24. https://doi.org/10.1016/j.molmed.2016.05.010.

Soto, Ileana, Leah C. Graham, Hannah J. Richter, Stephen N. Simeone, Jake E. Radell, Weronika Grabowska, W. Keith Funkhouser, Megan C. Howell, and Gareth R. Howell. 2015. "APOE Stabilization by Exercise Prevents Aging Neurovascular

Dysfunction and Complement Induction." *PLoS Biology* 13 (10): 1–33. https://doi.org/10.1371/journal.pbio.1002279.

Sounni, Nor Eddine, Alexandra Paye, Lorin Host, and Agnès Noël. 2011. "MT-MMPS as Regulators of Vessel Stability Associated with Angiogenesis." *Frontiers in Pharmacology* MAY (May): 1–11. https://doi.org/10.3389/fphar.2011.00111.

Stavnezer, Janet, and Carol E Schrader. 2014. "IgH Chain Class Switch Recombination: Mechanism and Regulation." *The Journal of Immunology* 193 (11): 5370–78. https://doi.org/10.4049/jimmunol.1401849.

Stephan, Johannes, Oliver Stegle, and Andreas Beyer. 2015. "A Random Forest Approach to Capture Genetic Effects in the Presence of Population Structure." *Nature Communications* 6 (May): 1–10. https://doi.org/10.1038/ncomms8432.

Stewart, Anita L., Kristin M. Mills, Abby C. King, William L. Haskell, Dawn Gillis, and Phillip L. Ritter. 2001. "CHAMPS Physical Activity Questionnaire for Older Adults: Outcomes for Interventions." *Medicine and Science in Sports and Exercise*, 1126–41. https://doi.org/10.1097/00005768-200107000-00010.

Strosznajder, Joanna B., Henryk Jęśko, and Robert P. Strosznajder. 2000. "Effect of Amyloid Beta Peptide on Poly(ADP-Ribose) Polymerase Activity in Adult and Aged Rat Hippocampus." *Acta Biochimica Polonica* 47 (3): 847–54.

Tipton, Christopher M., Jennifer R. Hom, Christopher F. Fucile, Alexander F. Rosenberg, and Inaki Sanz. 2018. "Understanding B-Cell Activation and Autoantibody Repertoire Selection in Systemic Lupus Erythematosus: A B-Cell Immunomics Approach." *Immunological Reviews* 284 (1): 120–31. https://doi.org/10.1111/imr.12660.

Toro-Domínguez, Daniel, Pedro Carmona-Sáez, and Marta E. Alarcón-Riquelme. 2014. "Shared Signatures between Rheumatoid Arthritis, Systemic Lupus Erythematosus and Sjögren's Syndrome Uncovered through Gene Expression Meta-Analysis." *Arthritis Research and Therapy* 16 (1): 1–8. https://doi.org/10.1186/s13075-014-0489-x.

Treiber, N., Thomas Treiber, Georg Zocher, and Rudolf Grosschedl. 2010. "Structure of an Ebf1:DNA Complex Reveals Unusual DNA Recognition and Structural Homology with Rel Proteins." *Genes & Development* 24 (20): 2270–75. https://doi.org/10.1101/gad.1976610.

Trerotola, Marco, Valeria Relli, Pasquale Simeone, and Saverio Alberti. 2015. "Epigenetic Inheritance and the Missing Heritability." *Human Genomics* 9: 17. https://doi.org/10.1186/s40246-015-0041-3.

Trifunovic, Aleksandra, Anna Wredenberg, Maria Falkenberg, Johannes N. Spelbrink, Anja T. Rovio, Carl E. Bruder, Mohammad Bohlooly-Y, et al. 2004. "Premature Ageing in Mice Expressing Defective Mitochondrial DNA Polymerase." *Nature* 429 (6990): 417–23. https://doi.org/10.1038/nature02517.

Tsokos, George C., Mindy S. Lo, Patricia Costa Reis, and Kathleen E. Sullivan. 2016. "New Insights into the Immunopathogenesis of Systemic Lupus Erythematosus." *Nature Reviews Rheumatology* 12 (12): 716–30. https://doi.org/10.1038/nrrheum.2016.186.

Tuo, Jingsheng, Baitang Ning, Christine M Bojanowski, Zhong-Ning Lin, Robert J Ross, George F Reed, Defen Shen, et al. 2006. "Synergic Effect of Polymorphisms in ERCC6 5' Flanking Region and Complement Factor H on Age-Related Macular

Degeneration Predisposition." *Proceedings of the National Academy of Sciences of the United States of America* 103 (24): 9256–61. https://doi.org/10.1073/pnas.0603485103.

Valdes, A. M., T. Andrew, J. P. Gardner, M. Kimura, E. Oelsner, L. F. Cherkas, A. Aviv, and T. D. Spector. 2005. "Obesity, Cigarette Smoking, and Telomere Length in Women." *Lancet* 366 (9486): 662–64. https://doi.org/10.1016/S0140-6736(05)66630-5.

Vasunilashorn, Sarinnapha, Antonia K. Coppin, Kushang V. Patel, Fulvio Lauretani, Luigi Ferrucci, Stefania Bandinelli, and Jack M. Guralnik. 2009. "Use of the Short Physical Performance Battery Score to Predict Loss of Ability to Walk 400 Meters: Analysis from the InCHIANTI Study." *Journals of Gerontology - Series A Biological Sciences and Medical Sciences* 64 (2): 223–29. https://doi.org/10.1093/gerona/gln022.

Velarde, Michael C., James M. Flynn, Nicholas U. Day, Simon Melov, and Judith Campisi. 2012. "Mitochondrial Oxidative Stress Caused by Sod2 Deficiency Promotes Cellular Senescence and Aging Phenotypes in the Skin." *Aging* 4 (1): 3–12. https://doi.org/10.18632/aging.100423.

Vermeij, WP, JH Hoeijmakers, and J Pathof. 2016. "Genome Integrity in Aging: Human Syndromes, Mouse Models, and Therapeutic Options." *Annual Review of Pharmacology and Toxicology* 56: 427–45.

Vilagos, Bojan, Mareike Hoffmann, Abdallah Souabni, Qiong Sun, Barbara Werner, Jasna Medvedovic, Ivan Bilic, et al. 2012. "Essential Role of EBF1 in the Generation and Function of Distinct Mature B Cell Types." *The Journal of*

Experimental Medicine 209 (4): 775–92. https://doi.org/10.1084/jem.20112422.

Waddington, Hal Conrad. 1956. "Genetic Assimilation of the Bithorax Phenotype." *Evolution* 10 (1): 1–13.

Wagner, James R, Stephan Busche, Bing Ge, Tony Kwan, Tomi Pastinen, and Mathieu Blanchette. 2014. "The Relationship between DNA Methylation, Genetic and Expression Inter-Individual Variation in Untransformed Human Fibroblasts." *Genome Biology* 15 (2): R37. https://doi.org/10.1186/gb-2014-15-2-r37.

Wahadat, M. Javad, Iris L.A. Bodewes, Naomi I. Maria, Cornelia G. van Helden-Meeuwsen, Annette van Dijk-Hummelman, Eline C. Steenwijk, Sylvia Kamphuis, and Marjan A. Versnel. 2018. "Type I IFN Signature in Childhood-Onset Systemic Lupus Erythematosus: A Conspiracy of DNA- and RNA-Sensing Receptors?" *Arthritis Research and Therapy* 20 (1): 1–8. https://doi.org/10.1186/s13075-017-1501-z.

Walter, Stefan, Gil Atzmon, Ellen W. Demerath, and Nora Franceschini. 2011. "A Genome-Wide Association Study of Aging." *Neurobiol Aging* 32 (11): 1–21. https://doi.org/10.1016/j.neurobiolaging.2011.05.026.A.

Wang, Jie, Jiali Zhuang, Sowmya Iyer, Xin Ying Lin, Troy W. Whitfield, Melissa C. Greven, Brian G. Pierce, et al. 2012. "Sequence Features and Chromatin Structure around the Genomic Regions Bound by 119 Human Transcription Factors." *Genome Research* 22 (9): 1798–1812. https://doi.org/10.1101/gr.139105.112.

Wang, Lan, Long Fei Wu, Xin Lu, Xing Bo Mo, Zai Xiang Tang, Shu Feng Lei, and Fei Yan Deng. 2015. "Integrated Analyses of Gene Expression Profiles Digs out Common Markers for Rheumatic Diseases." *PLoS ONE* 10 (9): 1–11.

https://doi.org/10.1371/journal.pone.0137522.

Watanabe, Shojiro, Tadaatsu Imaizumi, Kazushi Tsuruga, Tomomi Aizawa, Tatsuya Ito, Tomoh Matsumiya, Hidemi Yoshida, Kensuke Joh, Etsuro Ito, and Hiroshi Tanaka. 2013. "Glomerular Expression of Myxovirus Resistance Protein 1 in Human Mesangial Cells: Possible Activation of Innate Immunity in the Pathogenesis of Lupus Nephritis." *Nephrology* 18 (12): 833–37. https://doi.org/10.1111/nep.12155.

Weckerle, Corinna E., Beverly S. Franek, Jennifer A. Kelly, Marissa Kumabe, Rachel A. Mikolaitis, Stephanie L. Green, Tammy O. Utset, et al. 2011. "Network Analysis of Associations between Serum Interferon-α Activity, Autoantibodies, and Clinical Features in Systemic Lupus Erythematosus." *Arthritis & Rheumatism* 63 (4): 1044–53. https://doi.org/10.1002/art.30187.

Weinhold, Bob. 2006. "Epigenetics: The Science of Change." *Environmental Health Perspectives.* 114 (3): 160–67. https://doi.org/10.1289/ehp.114-a160.

Wetterstrand KA. 2016. "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Available at: Www.Genome.Gov/Sequencingcostsdata. Accessed 2016-09-05." 2016. www.genome.gov/sequencingcostsdata/.

Wiley, Kenneth L., Edward Treadwell, Kayihura Manigaba, Beverly Word, and Beverly D. Lyn-Cook. 2013. "Ethnic Differences in DNA Methyltransferases Expression in Patients with Systemic Lupus Erythematosus." *Journal of Clinical Immunology* 33 (2): 342–48. https://doi.org/10.1007/s10875-012-9803-z.

Wood, Andrew R, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, et al. 2014. "Defining the Role of Common Variation in the Genomic and Biological Architecture of Adult Human Height." *Nature Genetics*

46 (11): 1173–86. https://doi.org/10.1038/ng.3097.

Wu, Haijing, Jinrong Zeng, Jinghua Yin, Qiao Peng, Ming Zhao, and Qianjin Lu. 2017. "Organ-Specific Biomarkers in Lupus." *Autoimmunity Reviews* 16 (4): 391–97. https://doi.org/10.1016/j.autrev.2017.02.011.

Wu, J., C. Metz, X. Xu, R. Abe, A. W. Gibson, J. C. Edberg, J. Cooke, F. Xie, G. S. Cooper, and R. P. Kimberly. 2014. "A Novel Polymorphic CAAT/Enhancer-Binding Protein Element in the FasL Gene Promoter Alters Fas Ligand Expression: A Candidate Background Gene in African American Systemic Lupus Erythematosus Patients." *The Journal of Immunology* 170 (1): 132–38. https://doi.org/10.4049/jimmunol.170.1.132.

Wu, Yu Chang Bryan, David Kipling, and Deborah K. Dunn-Walters. 2011. "The Relationship between CD27 Negative and Positive B Cell Populations in Human Peripheral Blood." *Frontiers in Immunology* 2 (DEC): 1–12. https://doi.org/10.3389/fimmu.2011.00081.

Xu, Bao Ning, Ye Tian, Si Yang Liu, Xi Tong Zhao, Qing Wang, Da Li Tian, and Wen Jun Jiang. 2016. "The Relationship between XPG Rs17655 Polymorphism and the Risk of Lung Cancer." *International Journal of Clinical and Experimental Medicine* 9 (2): 4620–24.

Xu, Xingxing, Yonghui Tao, Xiaobo Gao, Lei Zhang, Xufang Li, Weiguo Zou, Kangcheng Ruan, Feng Wang, Guo-liang Xu, and Ronggui Hu. 2016. "A CRISPR-Based Approach for Targeted DNA Demethylation." *Cell Discovery* 2: 16009. https://doi.org/10.1038/celldisc.2016.9.

Yan, Huihuang, Shulan Tian, Susan L. Slager, Zhifu Sun, and Tamas Ordog. 2016.

"Genome-Wide Epigenetic Studies in Human Disease: A Primer on -Omic Technologies." *American Journal of Epidemiology* 183 (2): 96–109. https://doi.org/10.1093/aje/kwv187.

Yang, Qi, Lela Kardava, Anthony St Leger, Kathleen Martincic, Barbara Varnum-Finney, Irwin D Bernstein, Christine Milcarek, and Lisa Borghesi. 2008. "E47 Controls the Developmental Integrity and Cell Cycle Quiescence of Multipotential Hematopoietic Progenitors." *Journal of Immunology (Baltimore, Md. : 1950)* 181 (9): 5885–94. https://doi.org/10.1016/j.freeradbiomed.2008.10.025.

Ye, S., H. Pang, Y. Y. Gu, J. Hua, X. G. Chen, C. D. Bao, Y. Wang, et al. 2003. "Protein Interaction for an Interferon-Inducible Systemic Lupus Associated Gene, IFIT1." *Rheumatology* 42 (10): 1155–63. https://doi.org/10.1093/rheumatology/keg315.

Yen, Eric Y., Magda Shaheen, Jennifer M.P. Woo, Neil Mercer, Ning Li, Deborah K. McCurdy, Arun Karlamangla, and Ram R. Singh. 2017. "46-Year Trends in Systemic Lupus Erythematosus Mortality in the United States, 1968 to 2013: A Nationwide Population-Based Study." *Annals of Internal Medicine* 167 (11): 777–85. https://doi.org/10.7326/M17-0102.

Yeung, Kit San, Brian Hon-Yin Chung, Sanaa Choufani, Mo Yin Mok, Wai Lap Wong, Christopher Chun Yu Mak, Wanling Yang, et al. 2017. "Genome-Wide DNA Methylation Analysis of Chinese Patients with Systemic Lupus Erythematosus Identified Hypomethylation in Genes Related to the Type I Interferon Pathway." *Plos One* 12 (1): e0169553. https://doi.org/10.1371/journal.pone.0169553.

Yong, Wai Shin, Fei Man Hsu, and Pao Yang Chen. 2016. "Profiling Genome-Wide DNA Methylation." *Epigenetics and Chromatin* 9 (1): 1–16.

https://doi.org/10.1186/s13072-016-0075-3.

Yuan, Quan, Jing-Wei Liu, Cheng-Zhong Xing, and Yuan Yuan. 2014. "Associations of ERCC4 Rs1800067 Polymorphism with Cancer Risk: An Updated Meta-Analysis." *Asian Pacific Journal of Cancer Prevention : APJCP* 15 (18): 7639–44. https://doi.org/10.7314/APJCP.2014.15.18.7639.

Yuan, Yujia, Vinicius Fernandes Cruzat, Philip Newshome, Jingqiu Cheng, Younan Chen, and Yanrong Lu. 2016. "Regulation of SIRT1 in Aging: Roles in Mitochondrial Function and Biogenesis." *Mechanisms of Ageing and Development* 155 (1): 10–21. https://doi.org/10.1016/j.mad.2016.02.003.

Zaveri, K, J F Abril, R Guigo, M J Campbell, K V Sjolander, B Karlak, A Kejariwal, et al. 2001. "The Sequence of the Human Genome." *Science* 291 (5507): 1304–51.

Zhang, J, V Roschke, K P Baker, Z Wang, G S Alarcón, B J Fessler, H Bastian, R P Kimberly, and T Zhou. 2001. "Cutting Edge: A Role for B Lymphocyte Stimulator in Systemic Lupus Erythematosus." *Journal of Immunology (Baltimore, Md. : 1950)* 166 (1): 6–10. https://doi.org/10.4049/jimmunol.166.1.6.

Zhao, J., S. Chen, H. Zhou, T. Zhang, Y. Liu, J. He, J. Zhu, and J. Ruan. 2018. "XPG Rs17655 G > C Polymorphism Associated with Cancer Risk: Evidence from 60 Studies." *Aging* 10 (5). https://doi.org/10.18632/aging.101448.

Zhao, Ming, Yin Zhou, Bochen Zhu, Mengjie Wan, Tingting Jiang, Qiqun Tan, Yan Liu, et al. 2016a. "IFI44L Promoter Methylation as a Blood Biomarker for Systemic Lupus Erythematosus." *Annals of the Rheumatic Diseases* 75 (11): 1998–2006. https://doi.org/10.1136/annrheumdis-2015-208410.

———. 2016b. "IFI44L Promoter Methylation as a Blood Biomarker for Systemic Lupus

Erythematosus." *Annals of the Rheumatic Diseases* 75 (11): 1998–2006. https://doi.org/10.1136/annrheumdis-2015-208410.

Ziogas, Demosthenes E., Ioannis D. Kyrochristos, and Dimitrios H. Roukos. 2018. "Next-Generation Sequencing: From Conventional Applications to Breakthrough Genomic Analyses and Precision Oncology." *Expert Review of Medical Devices* 15 (1): 1–3. https://doi.org/10.1080/17434440.2017.1419059.