THE UNIVERSITY OF CHICAGO

MAPPING THE GENETIC DETERMINANTS OF IMMUNE DISEASE SUSCEPTIBILITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY

SILVIA NJERI KARIUKI

CHICAGO, ILLINOIS

JUNE 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTARY FIGURES

LIST OF SUPPLEMENTARY TABLES

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Anna Di Rienzo, for her enormous support and mentorship of my thesis work. Her devotion to my development as a graduate student in her lab, and as a scientist in general, is truly humbling, and I am grateful for her mentorship and encouragement to continue pursuing my scientific interests. I would also like to thank the members of my thesis committee, Carole Ober, Marcelo Nobrega and John Novembre, for their fruitful discussions and valuable guidance throughout my thesis project, as well as their constant encouragement throughout my years at grad school. I would also like to thank the members of the Di Rienzo lab for their valuable contributions to my projects on Vitamin D, and for their helpful discussions that enabled further understanding of the data. The friendships I have gained in this lab, as well as in the department of Human Genetics, have truly enriched my graduate school experience, and will always be treasured.

I would also like to acknowledge and thank Dr. Tim Niewold, to whom I am grateful for spiking my interest in the field of human genetics. My experiences in his lab were truly invaluable, and gave me the solid foundation for pursuing this PhD program in Human Genetics. I am also grateful to members of the Niewold Lab who helped me with data collection and data analysis on my thesis projects.

Finally, I would like to thank my family for their unrelenting support throughout my years as a graduate student. This dissertation is especially dedicated to them, for instilling in me the values of obtaining a good education, and for constantly encouraging my pursuit of a scientific career.

ABSTRACT

In this dissertation, I focus on mapping intermediate phenotypes, or endophenotypes, to characterize the molecular basis of inter-individual variation in immune disease susceptibility. In chapter 2, mapping serum interferon-alpha (IFN-α) activity, a stable heritable molecular sub-phenotype, enabled identification of novel loci associated with systemic lupus erythematosus (SLE), a chronic autoimmune disorder. These loci were replicated in an independent cohort of SLE cases, and represent novel loci underlying variation in SLE susceptibility, through dysregulation of the IFN-α pathway. Due to the genetic and phenotypic heterogeneity of SLE, which reduces the power of overall case-control studies, the endophenotype mapping approach was particularly useful for identification of novel disease-associated loci. In chapter 3, I shift focus to characterizing the transcriptional effects of vitamin D which plays an important immunomodulatory role, and is a modifiable environmental factor for autoimmune diseases. I examined the patterns of transcriptional response to the active, hormonal form of vitamin D, 1,25-dihydroxyvitamin D_3 (1,25D), in primary human monocytes, both in the presence and absence of bacterial lipopolysaccharide (LPS), a potent immune stimulant. A joint Bayesian analysis enabled clustering of genes into patterns of shared transcriptional response across treatments. The biological pathways enriched within these expression patterns highlighted the opposite effects of 1,25D and LPS on the transcriptome, and the potential molecular mechanisms through which 1,25D exerts its immunomodulatory role, such as through induction of genes in the mTOR signaling and EIF2 signaling pathways. Dysregulation of these pathways could contribute to the risk of the several immune-mediated diseases that

are linked to vitamin D deficiency, such as SLE. The processed gene expression values and differential expression analysis results from this chapter are included in the online **Supplementary File 3.1**. Finally, in chapter 4, I focus on mapping the molecular mechanisms underlying inter-individual variation in response to the immunomodulatory effects of 1,25D both at the cellular and at the transcriptional level. Two intergenic SNPs were associated, at genome-wide significant levels, with variation in percent inhibition of cell proliferation (I_{max}) induced by 1,25D treatment of peripheral blood mononuclear cells (PBMCs). I also identified several expression quantitative trait loci (eQTLs), which underlie variation in transcriptional response to 1,25D. Combining the information from the cellular and transcriptional endophenotypes in this study enabled identification of loci that putatively mediate the anti-proliferative activity of 1,25D in the immune system. Overall, the work described in this dissertation demonstrates that it is possible to detect the genetic determinants of intermediate endophenotypes, such as IFN-α activity in SLE, and cellular and transcriptional response to vitamin D, using relatively small sample sizes. These loci may not only underlie inter-individual variation in susceptibility to immune-mediated diseases, but they may also provide potential therapeutic targets for these diseases.

CHAPTER 1: INTRODUCTION

Inter-individual variation in susceptibility to immune disease can be attributed to several underlying factors, including both environmental and genetic factors. Historically, family-based studies and twin studies have been used to measure heritability in complex traits [1]. Linkage studies have been useful in identifying genes associated with 'Mendelian' types of diseases, attributable to single genes with large effect sizes. However, for variants with modest effect sizes, linkage studies have been very limited in their power and resolution [2, 3]. More recent studies of the genetic contribution to complex disease susceptibility have used genome-wide association studies (GWAS), which assay several hundreds of thousands to millions of single nucleotide polymorphisms (SNPs), in thousands of individuals, to deduce the genetic architecture of complex traits [1, 4, 5]. So far, hundreds of complex disease variants have been identified using GWAS, but these still do not fully explain the phenotypic variation that is attributable to genetic components in these complex diseases [1]. Several strategies have been postulated to increase the power to detect additional disease-associated variants in order to fully explain the heritability of these complex traits, including larger sample sizes for GWAS, since most of these disease-associated variants tend to have low effect sizes. However, with the expected modest effect sizes of variants associated with common disease susceptibility, very large sample sizes, reaching the tens of thousands of cases and controls, would be required to detect these variants with reasonable power [6, 7]. These sample sizes begin to reach the limits of feasibility, especially for complex autoimmune diseases, which affect a smaller proportion of the population compared to other common diseases, such as asthma or heart disease. It

is hence crucial to consider more practical approaches in designing GWAS for complex traits.

One such practical approach, which has been the focus of my thesis studies, is the evaluation of intermediate biological phenotypes, or 'endophenotypes', namely quantifiable heritable traits, such as gene expression or activity of specific proteins, that are related to the complex disease [8]. Since complex diseases represent the end points of several pathologic and physiological processes, mapping these intermediate endophenotypes presents a more powerful tool to detect additional variants that underlie disease susceptibility [9]. The advantage of studying endophenotypes is that they are less heterogeneous than the complex diseases, and the effect sizes of the variants associated with these endophenotypes are greater due to the simplicity of these endophenotypes. For these reasons, mapping disease-related quantitative traits is expected to increase the power to detect additional loci associated with complex diseases even within a smaller sample cohort.

Mapping serum interferon-alpha activity: an endophenotype in Systemic Lupus Erythematosus pathogenesis

Systemic lupus erythematosus (SLE) is a complex genetically and clinically heterogeneous trait that is influenced by a combination of genetic and environmental factors that lead to an irreversible break in immunologic self-tolerance [10]. It is characterized by multi-system involvement commonly affecting the skin, renal, musculoskeletal, and hematopoietic systems. Case-control genetic studies in SLE have been successful in identifying more than 30 loci linked to SLE susceptibility, with the HLA locus

providing the strongest evidence for association. The HLA locus contains greater than 100 genes [11, 12] that have functions in the immune system, including antigen presentation to adaptive immune cells, an event that is central to the manifestation of SLE and other autoimmune diseases. Several non-HLA loci have also been identified through GWAS, with many of these loci having roles in both the innate and adaptive immune system [13, 14]. Using statistical linear models to estimate the variance in liability explained by individual variants, several studies have indicated that these variants account for less than 20% of the heritability of SLE [13, 15-21]. This is similar to other complex diseases such as Crohn's disease, where more than 30 loci have been identified so far, yet these explain less than half of the heritability [20, 22].

Increasingly, mapping intermediate biological phenotypes, or endophenotypes, in GWAS studies has been shown to be more powerful in uncovering the underlying genetic and molecular mechanism of disease [8, 9]. This is especially important in SLE, which has a large amount of genetic and phenotypic heterogeneity, which greatly reduces the power of overall case-control genetic studies in SLE. Furthermore, previous work has shown that some of the established SLE-risk loci are characterized by strong effects on endophenotypes, providing more evidence for the increased power to detect pathogenic loci when endophenotype effects are incorporated into GWAS designs [23].

The molecular endophenotype that I focused on in my studies on SLE was serum interferon-alpha (IFN-α) activity, which is a stable heritable trait that is central to the pathogenesis of SLE. IFN-α is a type I interferon cytokine that plays an important role in viral defense. It activates dendritic cells and other antigen-presenting cells, and increases the expression of MHC class I and II molecules upon viral nucleic acid uptake. IFN-α has the

potential to break self-tolerance by potentially lowering the threshold for productive pro-inflammatory antigen presentation after uptake of nucleic acid material from the host, or 'self'-material [24, 25]. Serum IFN-α is elevated in many SLE patients, and elevations often correlate with disease activity [26, 27]. In addition, serum IFN-α is abnormally high in healthy first degree relatives of SLE patients as compared to healthy unrelated individuals, suggesting that high serum IFN-α is a heritable risk factor for SLE [28].

Mapping this important molecular endophenotype was therefore a useful tool to address the challenge posed by heterogeneity in SLE. Using a relatively small sample size compared to standard GWAS, we performed a GWAS of serum IFN-α activity using only SLE cases comparing SLE patients with high and low IFN-α activity. Using this study design increased our power to detect additional loci underlying SLE pathogenesis that have not been previously reported in case-control SLE studies. This study underscored the complex genetic architecture of SLE, and the importance of molecular sub-phenotyping in deciphering this complex architecture.

Vitamin D and Systemic Lupus Erythematosus risk

To further dissect the mechanisms underlying autoimmune disease susceptibility, I focused on studying vitamin D, a modifiable environmental factor in autoimmune disease with a well-known role as an immune system modulator [29-41]. The primary source of vitamin D is an inactive compound found in the skin, 7-dehydrocholesterol, which is converted to vitamin D precursors through exposure to ultraviolet (UV) light. As UV light exposure is central to the primary production of vitamin D, populations living at higher latitudes, where sun exposure is lower, have a higher prevalence of vitamin D deficiency.

Furthermore, individuals with darker skin pigmentation living in higher latitudes are especially prone to vitamin D deficiency, as is the case in the United States, where individuals of African-American ancestry have the highest prevalence of vitamin D deficiency [42-45]. Vitamin D deficiency is implicated as one of the environmental factors that contribute to SLE prevalence [46, 47], attributable to the role of vitamin D as a modulator of the immune system, where it attenuates the pro-inflammatory immune response. Inadequate vitamin D levels could hence contribute to an unchecked pro-inflammatory state that could lead to the pathogenesis of SLE. Indeed, numerous epidemiological studies have reported associations between deficiency in the circulating stable form of vitamin D, 25-hydroxyvitamin D_3 (25D), and risk of SLE [47-49]. It is also interesting to note that African-Americans, who have the highest prevalence of vitamin D deficiency, also have a higher prevalence and severity of SLE [10, 50-52].

Various studies have attempted to elucidate the mechanisms underlying the link between vitamin D deficiency and SLE risk. Studies on dendritic cells, which produce IFN-α after stimulation by nucleic acid-containing immune complexes, provide some clues on the mechanisms through which vitamin D could confer protection against SLE pathogenesis. These studies show that vitamin D suppresses dendritic cell differentiation and activity by inducing a tolerogenic phenotype [41], and it also suppresses the expression of IFN-α inducible genes, or the "interferon signature", in monocyte-derived dendritic cells from SLE patients [48, 49]. Given the central role of IFN-α in SLE pathogenesis, targeting of the IFN-α pathway provides a crucial link to the protective role of vitamin D.

Further knowledge on the manner in which vitamin D modulates the immune system is needed for potential use of vitamin D as a therapeutic agent for SLE, as well as

other immune-mediated diseases. The primary aim of my thesis studies on vitamin D was to examine the mechanisms through which vitamin D exerts its role in immune cells and the genetic architecture underlying inter-individual variation in the modulatory functions of vitamin D, in order to identify novel loci and pathways whose dysregulation could lead to pathogenesis of immune-mediated diseases like SLE.

Transcriptional effects of vitamin D in the immune system

The immunomodulatory role of vitamin D is mediated by its active hormonal substrate, 1,25-dihydroxyvitamin D_3 (1,25D), through a transcriptional mechanism [32, 37, 38]. Circulating inactive 25-hydroxyvitamin D_3 (25D) is converted to 1,25D by the enzyme CYP27B1, which is expressed in the kidney [53, 54]. The classic systemic role of 1,25D is to promote calcium homeostasis and bone health by enhancing absorption of calcium in the small intestine, and stimulating osteoclast differentiation and calcium reabsorption of the bone [38, 55]. Extra-renal tissues that express CYP27B1, such as cells of the immune system, are able to locally synthesize the active 1,25D intracellularly from the circulating 25D, in response to organismal demands such as infections [56, 57]. Immune cells also express the vitamin D receptor (VDR), which when bound by 1,25D forms a heterodimer with the retinoid-X-receptor (RXR) [30, 33, 37, 38, 58]. This heterodimer translocates into the nucleus and acts as a transcriptional regulator of vitamin D-responsive genes. Systemic and local intracellular 1,25D levels are regulated by CYP24A1, an enzyme that initiates the degradation of 1,25D into an inactive metabolite that is excreted [53, 54].

Vitamin D exerts its immunomodulatory role in innate immune cells by enhancing the antimicrobial response through induction of antimicrobial genes such as cathelicidin

antimicrobial peptide (*CAMP*), β-defensin 4A (*DEFB4A*) and autophagy related 5 (*ATG5*) [32, 39, 59]. The antimicrobial effects of vitamin D have been demonstrated through epidemiological studies that have linked low levels of the inactive 25D with increased susceptibility to tuberculosis (Tb), which is caused by the bacterium *Mycobacterium tuberculosis* [60, 61]. Vitamin D has historically been used as a treatment for bacterial infections in diseases like Tb through ingestion of cod liver oil, a rich source of vitamin D_2, [62], or through the use of UV light, the principal source of the cutaneous production of vitamin D_3, to treat lupus vulgaris, a cutaneous form of Tb, which earned Niels Friensen a Nobel Prize in Medicine in 1903 [56, 63]. More recent studies on the anti-microbial activity of vitamin D in monocytes indicate that activation of Toll-like receptors, which are primarily expressed on monocyte cell membranes, by bacterial stimuli like lipopolysaccharide (LPS), induces the expression of *CYP27B1* and *VDR*. Induction of CYP27B1 leads to the localized synthesis of 1,25D, which in turn up-regulates antimicrobial genes like *CAMP*, leading to the subsequent intracellular killing of bacteria [32, 59]. In addition to its important antimicrobial role, several studies have also shown that 1,25D attenuates the pro-inflammatory immune response through induction of a tolerogenic phenotype in monocyte-derived dendritic cells, which lose their capacity to induce autoreactive T cell proliferation [41, 64-66].

Previous studies have examined the transcriptional effects of 1,25D in innate immune cells to elucidate the molecular mechanisms underlying its immunomodulatory effects. Several of these studies have used a targeted gene approach to quantify the induction of antimicrobial gene production by 1,25D [32, 59], but such targeted gene approaches are limited in their capacity to fully delineate various pathways and molecular

mechanisms underlying the function of 1,25D in the immune system. Subsequent studies have profiled the transcriptome-wide effects of 1,25D in a human THP-1 macrophage cell line [67-69], which was originally obtained from an infant with acute monocytic leukemia [70]. A significant limitation to using this cell line for characterizing genome-wide transcriptional effects in monocytes, is the failure to capture the physiological context-specific effects of primary immune cells, such as inter-cellular cross-talk and antigen presentation [71]. Recognizing this limitation, more recent studies have shifted to using primary human immune cells to profile genome-wide transcriptional response to 1,25D. A study on primary monocyte-derived dendritic cells elucidated the mechanisms underlying the role of 1,25D in maintaining a tolerogenic phenotype in the dendritic cells, through transcriptional regulation of metabolic pathways [41]. However, this study did not examine the role of 1,25D in the context of an inflammatory challenge, which would give a better picture of how 1,25D modulates the immune response. Another study on peripheral blood mononuclear cells (PBMCs) measured the genome-wide transcriptional response to 1,25D in the presence of phytohemagglutinin (PHA), an immune cell stimulant [72], and observed a significant enrichment of immune-related processes such as interferon signaling among down-regulated genes, while metabolic processes were enriched among up-regulated genes. While the use of PHA to stimulate PBMCs was useful for highlighting the pathways modulated by 1,25D in stimulated immune cells, it has been shown that PHA stimulation might be more effective for T and B lymphocyte cells, and may not be as adequate a stimulant for innate immune cell subtypes such as monocytes.

My thesis work in chapter 3 therefore focused on characterizing the genome-wide transcriptional response patterns to 1,25D in monocytes, the primary innate immune cell

type, in the context of LPS, which activates a pro-inflammatory response by signaling through the TLR4 and CD14 receptors expressed on the membrane surfaces of monocytes. Stimulating the monocytes with LPS enabled examination of how an inflammatory stimulus modifies the transcriptional response to 1,25D. I analyzed the genome-wide transcriptional response to 1,25D and LPS using two statistical approaches: a linear mixed-effects analysis, which identified significantly differentially expressed genes in each treatment category, and a Bayesian approach which assigned all the differentially expressed genes into distinct differential expression patterns. This analysis highlighted several biological pathways that are modulated by 1,25D in the absence of LPS, such as oxidative phosphorylation and mitochondrial dysfunction, which were significantly enriched among up-regulated genes. The study also highlighted various immune response pathways such as pro-inflammatory cytokine signaling, which were strongly induced by LPS, and this response was then reversed by 1,25D, which down-regulated the genes in these immune response pathways. The strength of this study is that it highlights the context-specific transcriptional regulation of several functional pathways by 1,25D, both in the presence and absence of LPS stimulation, which might mediate its immunomodulatory effects in primary monocytes.

Mapping cellular and transcriptional response to vitamin D

The important immunomodulatory role of 1,25D also extends to inhibition of activation and proliferation of T and B lymphocytes, and attenuation of production of pro-inflammatory cytokines [29, 38, 73-77], leading to an overall tempering of an intense pro-inflammatory response, which, if left unchecked, can have toxic consequences such as sepsis and septic shock [78-80], or can lead to autoimmune disease pathogenesis. With

several epidemiological studies linking vitamin D deficiency with risk of autoimmune diseases [29, 47, 48, 81-84], there has been an increased interest in the use of vitamin D as a potential therapeutic in immune-mediated diseases, which has led to several randomized vitamin D supplementation trials [49, 84-89]. However, these trials have yielded mixed results. For example, a randomized trial testing the efficacy of vitamin D in attenuating the IFN-α-induced gene signature in SLE patients showed no significant differences between the vitamin D-recipients and the placebo-recipients [90]. Several other vitamin D supplementation trials in other immune-mediated diseases have also shown mixed results [85-89, 91]. The underwhelming success of vitamin D as a potential therapeutic agent could be due to several factors, such as small study population sizes, short duration of the trials, and insufficient dosage of 25D [92, 93]. Another important factor is the inter-individual differences in the response to vitamin D, irrespective of its concentration in circulation or within the cells at the level of the target organ. Little is known about the contribution of genetics to the inter-individual variation in response to vitamin D.

The main objective of chapter 4 was to map the genetic bases of inter-individual variation in the response to 1,25D. This study built upon a previous study that characterized the molecular basis for inter-individual variation in the response to glucocorticoids, which are steroid hormones that are widely used as therapeutic agents for a variety of diseases [94]. Since vitamin D is a fat-soluble steroidal hormone with anti-proliferative effects [37, 38, 73, 74], I was interested in similarly characterizing the molecular basis for inter-individual variation in 1,25D response, both at the cellular and transcriptional level, in peripheral blood mononuclear cells (PBMCs). Using PBMCs was

appropriate and practical for this study design as they are an abundant and easily accessible primary cell type.

To this end, I carried out a GWAS to map genetic variants underlying inter-individual variation in the percent inhibition of cell proliferation (I_{max}) by a single, high dose of 1,25D treatment of PBMCs obtained from 88 African-American healthy individuals. By measuring the proportion of African ancestry in this African-American cohort, I was able to directly test the relationship between African ancestry and response to 1,25D. While there were no significant associations between I_{max} and the proportion of African ancestry, there was a negative trend in the relationship between the proportion of African ancestry and serum 25D levels, which suggests a genetic contribution to the higher prevalence of vitamin D deficiency and insufficiency observed in African Americans [42]. Furthermore, majority of the GWAS of complex traits have been performed in individuals of European ancestry. While there are some shared variants underlying complex traits between populations, including the variants associated with variation in serum IFN-α activity described in chapter 2, many of the SNPs identified in European ancestry populations do not replicate in other non-European populations, resulting in an incomplete picture of the genetic architecture of complex traits. Individuals of recent African ancestry have the greatest genetic diversity and lower levels of linkage disequilibrium (LD) between alleles at different loci compared to non-African populations [4, 95]. Due to the shorter haplotype blocks in African ancestry populations, identifying disease-associated variants in these populations is likely to increase the resolution of putative-disease associated loci. Furthermore, vitamin D deficiency and insufficiency is most prevalent in individuals of African-American ancestry, and yet they are the most under-studied population in vitamin

D GWAS studies [42-45]. Understanding the genetic architecture of immune response to vitamin D, particularly in individuals of African-American ancestry, will be crucial in informing therapeutic supplementation interventions for immune-mediated diseases that have a higher prevalence in African-Americans, such as SLE.

In addition to mapping the immune cellular proliferation response to 1,25D, I also mapped genome-wide transcriptional response to 1,25D in the same individuals and cell culture system. Expression quantitative trait loci (eQTL) mapping is another powerful technique which utilizes variation in transcript abundance as an intermediate phenotype, or an endophenotype, to elucidate the genetic bases of complex traits [71]. Indeed, it has been shown that most of the single nucleotide polymorphisms (SNPs) identified in GWAS of a broad spectrum of complex traits are enriched for eQTLs [96]. Since most GWAS SNPs are non-coding and may affect gene regulation, incorporating eQTL mapping provides context to these non-coding SNPs, highlighting their transcriptional regulatory role on specific genes and pathways that underlie disease pathogenesis. For this study, I mapped *cis*-eQTLs within 100kb of the transcriptional start site of genes that were responsive to a single high dose of 1,25D treatment.

Intersecting the information from the eQTL mapping and GWAS of I_{max} analyses enabled identification not only of putative candidate genes that mediate the anti-proliferative properties of 1,25D in immune cells, but also enabled identification of variants that may influence inter-individual variation in response to 1,25D. Incorporating information on these genetic variants promises to be informative for future supplementation trials involving vitamin D, particularly in immune-mediated diseases like SLE.

CHAPTER 2: GENETIC ANALYSIS OF THE PATHOGENIC MOLECULAR SUB-PHENOTYPE INTERFERON-ALPHA IDENTIFIES MULTIPLE NOVEL LOCI INVOLVED IN SYSTEMIC LUPUS ERYTHEMATOSUS[1]

2.1: ABSTRACT

Systemic Lupus Erythematosus (SLE) is a chronic autoimmune disorder characterized by inflammation of multiple organ systems, loss of tolerance to self-antigens, and dysregulated interferon responses. SLE is both genetically and phenotypically heterogeneous, and we hypothesize that this greatly reduces the power of overall case-control studies in SLE. Increased circulating level of the cytokine interferon alpha (IFN-α) is a stable, heritable trait which has been implicated in SLE pathogenesis. To study genetic heterogeneity in SLE, we performed a case-case genome-wide association study comparing patients with high vs. low IFN-α in over 1550 SLE cases in both the discovery and replication cohorts. In the meta-analysis, the top associations in European ancestry subjects were rs7897633, an intronic SNP in protein kinase, cyclic GMP-dependent, type I (*PRKG1*) ($P_{Meta} = 2.75 \times 10^{-8}$), and rs1049564, a missense SNP in purine nucleoside phosphorylase (*PNP*) ($P_{Meta} = 1.24 \times 10^{-7}$). We also found evidence for cross-ancestral background associations in SNPs within the genes *ANKRD44* and *PLEKHF2*. These loci have not been previously identified in case-control SLE genetics studies. Bioinformatic analyses implicate these loci functionally in dendritic cells and natural killer cells, both of which are

[1] Citation for chapter: Kariuki SN, Ghodke-Puranik Y, et. al (2015). "Genetic analysis of the pathogenic molecular sub-phenotype interferon-alpha identifies multiple novel loci involved in systemic lupus erythematosus." Genes Immun 16(1): 15-23.

involved in IFN-α production in SLE. As case-control studies of heterogeneous diseases reach a limit of feasibility with respect to subject number and detectable effect size, the study of informative pathogenic sub-phenotypes becomes an attractive strategy for genetic discovery in complex human disease.

2.2: INTRODUCTION

Systemic lupus erythematosus (SLE) is a systemic autoimmune disorder characterized by involvement of multiple organ systems including skin, musculoskeletal, renal and hematologic systems. The pathogenesis of SLE is driven by a combination of both genetic and environmental risk factors, which lead to an irreversible break in immunologic self-tolerance [10]. SLE is four times more common in African-Americans compared with European-Americans [10], and both immunologic and genetic differences are appreciated between SLE patients from these ancestral backgrounds [50, 51, 97]. Familial aggregation and monozygotic twin studies strongly support the idea that SLE has a genetic component. There is a 50% concordance between identical twins, while first-degree relatives of SLE cases have a 20-fold higher risk of getting SLE [10, 28]. Genetic studies in SLE in various world populations have identified numerous susceptibility loci, however these account for far less than half of the heritability of SLE [15-19, 98], and most of the genes described have modest overall effect sizes (odds ratio (OR) ~1.5 to 1.2) [19, 98].

Further characterizing the heritability of SLE is challenging because of the large amount of genetic and phenotypic heterogeneity. Different genetic variations and molecular pathways may be of varying importance in different patients. Previous work

from our group has shown that some of the established SLE-risk loci are characterized by strong sub-phenotype effects, which are much greater than the overall case-control effect size [99]. This heterogeneity between patients greatly reduces the power of case-control studies in SLE, and is a potential explanation for much of the "missing heritability" in this disease. Designing genetic studies for SLE focusing on molecular endophenotypes should greatly increase our power to detect pathogenic loci.

Interferon alpha (IFN-α) is a molecular sub-phenotype which is central to the pathogenesis of SLE. IFN-α is a cytokine which works at the interface of the innate and adaptive immune systems, with the potential to break self-tolerance by activating antigen-presenting cells after the uptake of self-material [24]. Serum IFN-α is elevated in many SLE patients, and levels are stable over time [26, 100, 101]. Many lines of investigation support IFN-α as a primary causal factor in human SLE [102]. We have previously demonstrated familial aggregation of high IFN-α in SLE families [28], suggesting that high IFN-α is a heritable risk factor for SLE. Additionally, recombinant human IFN-α administered to humans as a therapy for chronic viral hepatitis and malignancy can induce de novo SLE in some cases. This IFN-α-induced SLE typically resolves after the IFN-α therapy is discontinued, which supports the idea that IFN-α is causal [103, 104]. Case-control genome-wide association studies (GWAS) in SLE have demonstrated remarkable over-representation of genes involved in type I interferon (IFN) signaling, production and response [98]. We have shown that many of these SLE-risk loci in the IFN-α pathway are associated with increased IFN-α pathway activity in SLE patients [105-108], supporting the idea that these loci are gain-of-function in humans. High circulating levels of IFN-α correspond to particular clinical manifestations [100], and thus activation of this pathway

contributes to both susceptibility and heterogeneity in SLE [109]. We suspect that heterogeneity in the molecular pathogenesis of SLE between patients is a major factor in the unexplained heritability of the disease to date. In this study, we directly address this heterogeneity by mapping the causal IFN-α molecular trait, which allowed for detection of novel genetic variations underlying SLE disease pathogenesis. In addition, over-activity of the IFN-α pathway has been implicated in other autoimmune diseases such as Sjogren's syndrome and inflammatory myositis [110, 111], and it is possible that these IFN-related loci underlie some of the genetic architecture of these conditions as well.

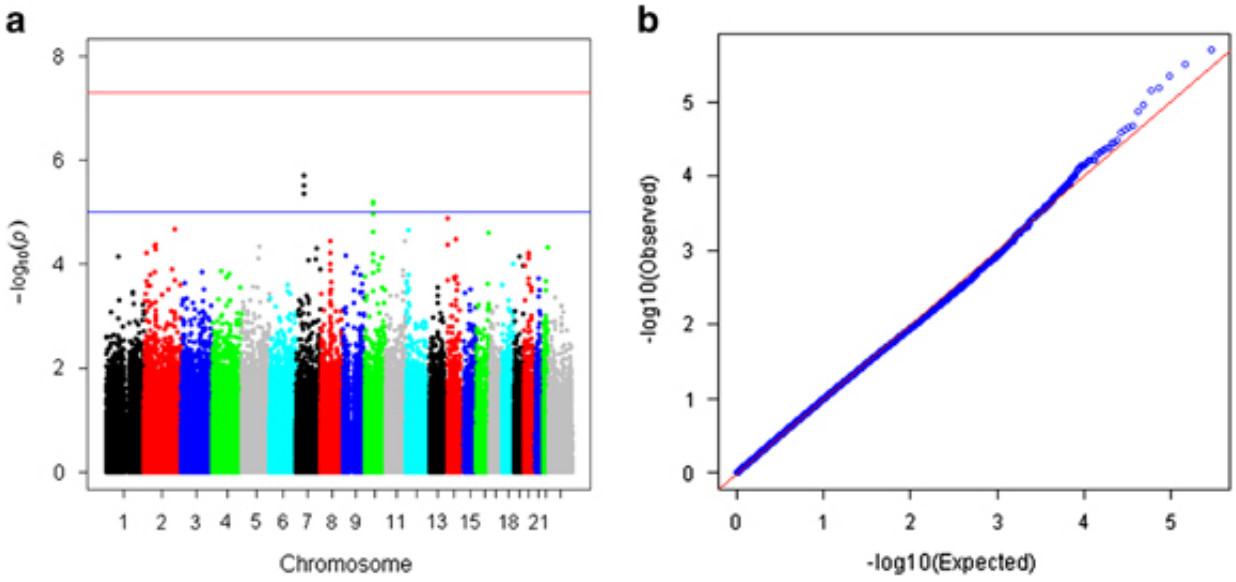
2.3: RESULTS

SNPs associated with IFN-α in the discovery cohort

We generated serum IFN-α activity data (using reporter cell assay described in Materials and Methods section to detect functional IFN-α activity) in the SLE cases who were genotyped in the SLE Genetics (SLEGEN) consortium genome-wide association study (GWAS) study for whom there was a serum sample available (n = 400) [16]. This group was used as our discovery cohort. Reanalyzing the GWAS data in a case − case analysis stratified by high vs low serum IFN-α, we found a number of strong associations (OR > 2.0) with serum IFN-α activity. These included single nucleotide polymorphisms (SNPs) in genes such as chromosome 7 open reading frame 57 (*C7orf57*), protein kinasecyclic GMPdependent type I (*PRKG1*), purine nucleoside phosphorylase (*PNP*), activating transcription factor 7 interacting protein (*ATF7IP*) and ankyrin repeat domain 44 (*ANKRD44*) (**Supplementary Table 2.1, Figure 2.1**). We conducted a pathway analysis to

identify canonical functional pathways that are enriched in the genes nearby these SNPs, and the results from this analysis are described later. The top SNPs identified in this analysis did not share any SNPs or loci in common with the known case-control SLE GWAS associations, supporting the ability of this approach to detect novel associations. Many of the underlying genetic variations with SLE could impact particular pathways or subsets of this heterogeneous disease, and these genetic variants can be missed by large case-control SLE GWAS in which all patients are grouped together. We then planned to replicate all SNPs identified in the discovery phase with $p < 10^{-4}$ (323 SNPs). In this replication list, there was one SNP which has been previously reported in a case-control SLE GWAS (rs1143678 in the gene *ITGAM*, p = 0.044) [16, 112], and there were two loci on the replication list which had previously been associated with serum IFN-α levels in SLE patients (*EFNA5* and *ZKSCAN1*/*LAMTOR4*, p = 0.036) [23, 113] (**Supplementary Table 2.2**).

Figure 2.1: Top signals of association with increased serum IFN-α activity in SLE cases in the discovery phase. A) Manhattan plot showing top GWAS association signals by chromosome. B) Q-Q Plot showing association of SLE GWAS SNPs with serum IFN-α. P-values that would be expected under the null hypothesis (no association between SNPs and serum IFN-α activity) are represented by the red line, and the observed P-values are represented by blue dots, one for each tested SNP- IFN-α activity association.



Validation of SNPs associated with serum IFN-α activity in an independent cohort

The 323 top SNPs which had a $p < 10^{-4}$ were genotyped in an independent cohort of 1165 SLE cases of European-American and African-American ancestry (see **Supplementary Table 2.3** for the characteristics of the replication cohort). We used logistic regression analysis to test SNPs from the discovery cohort for association with serum IFN-α, and European-American and African-American ancestral groups were analyzed separately. SNPs in the *PRKG1* (rs7897633, rs7906944) and *PNP* (rs1049564) loci showed strong evidence for association (**Table 2.1**) in the European-American patients. In meta-analysis, both *PRKG1* rs7897633 and *PNP* rs1049564 were associated with serum IFN-α in European ancestry with p-values that exceeded a conservative Bonferroni correction for multiple comparisons ($p < 1.71 \times 10^{-7}$, **Table 2.1**). Thus, the novel loci identified in the current study achieve genome-wide significance in the overall meta-analysis of discovery and replication sets. **Table 2.2** shows a list of the top SNP associations in African-Americans. No significant SNP-SNP interactions were detected. Haplotype analysis was performed when evidence for association was observed for two nearby SNPs, but none of the haplotype models were superior to the individual SNP models of association. For the SNPs which demonstrated evidence for association in both European-American and African-American ancestral backgrounds, those with homogenous effects by Breslow-Day testing were analyzed in meta-analysis assuming a fixed-effect model. The two SNPs included in this cross-ancestral background meta-analysis were rs4850410, an intronic SNP in *ANKRD44* (OR = 0.64; 95%CI (0.48 – 0.84); $P_{Meta} = 1.3 \times 10^{-6}$] and rs297573, which is downstream of the pleckstrin homology domain containing, family F member 2 gene (*PLEKHF2*) (OR=0.70; 95%CI (0.50 – 0.98); $P_{Meta} = 1.2 \times 10^{-4}$).

Table 2.1: List of top replicated SNPs associated with IFN-α in European-Americans

Chr*	Locus	SNP	SNP type	Assoc. allele*	Odds Ratio (95% CI)	P-discovery	P-replication	P_{Meta}
10	*PRKG1*	rs7897633	intron	C	0.59 (0.44 - 0.78)	1.07×10^{-5}	2.96×10^{-4}	2.75×10^{-8}
14	*PNP*	rs1049564	missense	T	2.08 (1.34 - 3.21)	1.32×10^{-5}	9.88×10^{-4}	1.24×10^{-7}
6	*DLL*	rs1028488	intergenic*	A	0.51 (0.38 - 0.70)	8.50×10^{-4}	3.12×10^{-5}	2.21×10^{-7}
7	*CALD1*	rs6467557	intron	T	1.50 (1.15 - 1.97)	5.00×10^{-5}	3.12×10^{-3}	1.40×10^{-6}
14	*PNP*	rs1713420	intron	C	1.82 (1.21 - 2.73)	4.25×10^{-5}	3.90×10^{-3}	1.58×10^{-6}
4	*GRXCR1*	rs6850606	intergenic*	A	0.64 (0.50 - 0.83)	4.75×10^{-4}	5.88×10^{-4}	1.81×10^{-6}
19	*ZNF536*	rs1549951	intergenic*	T	0.62 (0.45 - 0.85)	7.28×10^{-5}	3.10×10^{-3}	1.91×10^{-6}
10	*PRKG1*	rs7906944	intron	A	0.74 (0.57 - 0.93)	6.50×10^{-6}	1.83×10^{-2}	2.54×10^{-6}
1	*CHIA*	rs7411387	intron	C	1.61 (1.24 - 2.1)	1.23×10^{-3}	3.80×10^{-4}	3.07×10^{-6}
11	*TMPRSS5*	rs3934007	intergenic*	T	1.55 (1.19 - 2.00)	4.86×10^{-4}	9.98×10^{-4}	3.12×10^{-6}

Chr*: Chromosome
Assoc. allele*: Associated allele/minor allele
Intergenic*: The corresponding genes that are listed are those that are found nearest to the intergenic SNPs.

Table 2.2: List of top SNPs associated with serum IFN-α in African-Americans

Chr*	Locus	SNP	SNP type	Assoc. allele*	Odds Ratio (95% CI)	P-value
10	*NRG3*	rs1649949	intron	C	1.60 (1.20 - 2.15)	1.37×10^{-3}
2	*ANKRD44*	rs4850410	intron	T	0.64 (0.48 - 0.85)	1.69×10^{-3}
5	*LOC729506*	rs1666793	intron	C	1.5 (1.10 - 2.12)	1.10×10^{-2}
8	*ASPH*	rs7812327	intron	T	0.66 (0.48 - 0.93)	1.59×10^{-2}
20	*PLCB4*	rs2299676	intron	G	0.70 (0.50 - 0.95)	2.47×10^{-2}
5	*FGF18*	rs7711912	near 3’	A	1.45 (1.04 - 2.02)	2.90×10^{-2}
16	*RBFOX1*	rs4608354	intron	A	1.57 (1.03 - 2.40)	3.44×10^{-2}
8	*PLEKHF2*	rs297573	near 3'	C	0.70 (0.50 - 0.98)	3.83×10^{-2}
12	*KCNA5*	rs526654	near 3’	G	0.75 (0.57 - 1.00)	4.00×10^{-2}

Chr*: Chromosome
Assoc. allele*: Associated allele/minor allele

Association of GWAS candidates with autoantibody subsets in the replication cohort

Because the presence of particular autoantibodies has been strongly associated with high IFN-α in SLE [100], we also tested the SNPs which were replicated from the GWAS study for association with SLE autoantibodies. **Supplementary Tables 2.4 and 2.5** show the autoantibody associations observed in different ancestral groups in the replication cohort. These include the SNPs in *PRKG1* and *PLEKHF2* which were associated with IFN-α, as well as a SNP in a locus that we have previously found to be associated with autoantibodies in SLE (*EFNA5*) [23]. None of these serological associations withstood correction for multiple comparisons.

Canonical pathway analysis of GWAS candidate SNPs

A pathway analysis of the networks enriched among the top SNPs in the discovery cohort was generated through the use of IPA (Ingenuity Systems, www.ingenuity.com). All SNPs from the discovery cohort with $P < 10^{-4}$ were included. The top canonical pathways related to IFN-α-associated SNPs which pass a Benjamini-Hochberg false discovery rate of 0.05 are shown in **Table 2.3**. There was prominent representation of pathways associated with neural signaling and transmission, purine metabolism, and T cell signaling. Some of the key molecules defining these pathways were also some of the top validated serum IFN-α-associated loci in our replication cohort, such as *PNP* and *PRKG1*. Networks enriched in our study included those with various cellular functions such as cell morphology, cellular assembly and organization (*PRKG1*), cellular development and cell-mediated immune response (*PNP*) (**Supplementary Table 2.6**).

Table 2.3: Top 10 canonical pathways from IFN-α associated SNPs in initial discovery GWAS data

Canonical Pathways	Ratio	P-value
Axonal Guidance Signaling	0.03	4.04×10^{-4}
Synaptic Long Term Depression	0.04	4.27×10^{-3}
Dopamine-DARPP32 Feedback in cAMP Signaling	0.03	7.21×10^{-3}
Xanthine and Xanthosine Salvage	1.00	7.46×10^{-3}
Guanine and Guanosine Salvage I	0.50	1.49×10^{-2}
Adenine and Adenosine Salvage I	0.50	1.49×10^{-2}
Cellular Effects of Sildenafil (Viagra)	0.03	1.60×10^{-2}
Antiproliferative Role of TOB in T Cell Signaling	0.08	1.60×10^{-2}
Caveolar-mediated Endocytosis Signaling	0.04	1.67×10^{-2}
Cardiac Î²-adrenergic Signaling	0.03	1.77×10^{-2}

Ratio and P-value are calculated as described in the Methods section.

Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT)

Because the top loci identified in this study were not classical type I IFN pathway genes, we used the GIANT software to query potential relevance of the gene products encoded by these loci in various immune cell subsets. **Figure 2.2** shows the networks produced by the GIANT algorithm when the top hits from our study are used as the input data in the various immune cell subsets available for analysis. Networks with the highest density were observed in dendritic cells and natural killer (NK) cells, and low density networks were seen in T and B lymphocytes (**Table 2.4**). Similarly, the top associations with serum IFN-α generally demonstrated the greatest network strength in plasmacytoid dendritic cells and NK cells. These data support biological relevance of the transcripts in dendritic cells, which have been implicated as the major IFN-α producing cell type in SLE [114], and NK cells, which have been reported to play a critical cooperative role with dendritic cells in the production of IFN-α [115]. In addition, when examining the other molecules functionally implicated in these networks, a number of SLE-associated molecules are observed in the network diagrams, including IL12, TLR7, and the JAK/STAT pathways.

Figure 2.2: Tissue specific analysis of gene networks in different immune cells. Networks demonstrate relationships between *PNP*, *PRKG1*, *ANKRD44* and *PLEKHF2* to other molecules in immune cells. Edges with weight (relative confidence) greater than 0.4 are shown. Each network diagram represents a different immune cell type as follows: A: B lymphocyte, B: Dendritic cell, C: Monocyte, D: Neutrophil, E: NK cell, F: T lymphocyte.

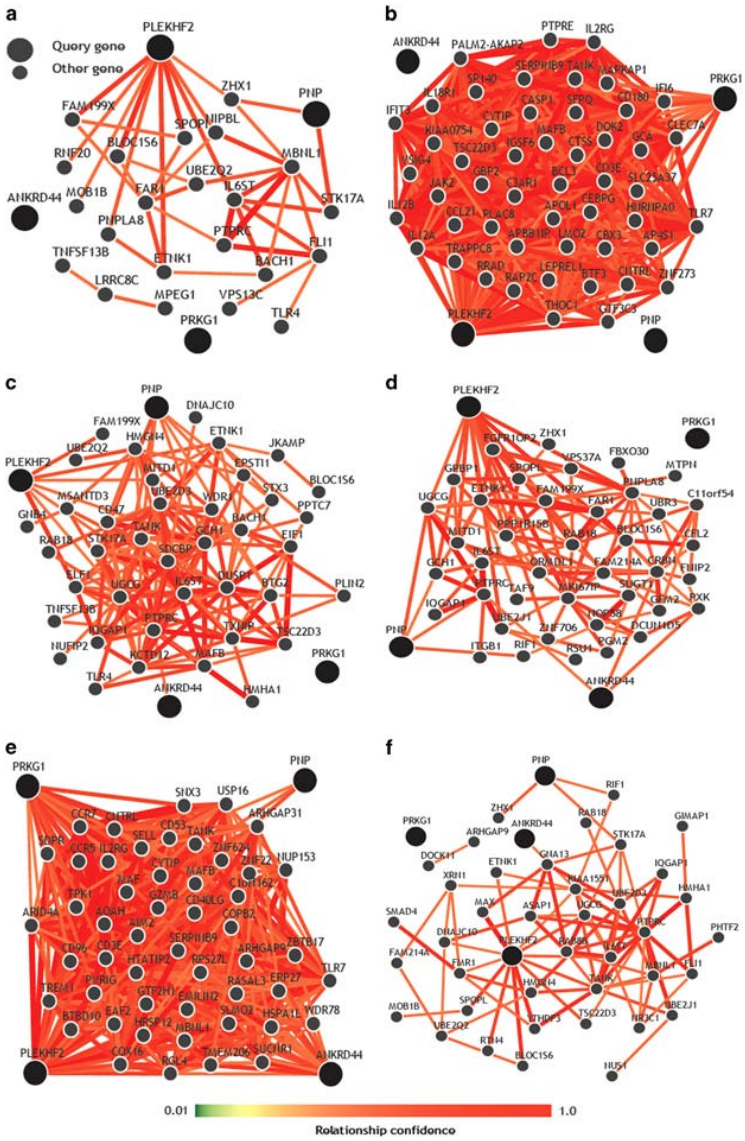


Table 2.4: Network density and network strength analysis for tissue specific gene networks in different immune cells

Cells	Network Density	Network Strength			
		ANKRD44	PNP	PRKG1	PLEKHF2
B lymphocyte	0.12		1.1		5.5
Dendritic cell	0.58			9.3	27.4
Monocyte	0.18	0.4	3.7		3.4
Neutrophil	0.13	1.4	1.9		8.7
T lymphocyte	0.09	0.4	1.4		7.3
NK cell	0.52	11.7	1.4	17.0	24.6

Networks generated by the GIANT software program for each immune cell type. Network density and strength calculated as described in the Methods. Density is calculated for the overall network in the cell, and strength is calculated for each of the loci entered in the analysis.

2.4: DISCUSSION

SLE is a highly heterogeneous disease, hence it is likely that certain genetic factors will be related to particular disease phenotypes and pathogenic pathways [10, 116, 117], and that genetic associations will not be shared between all SLE patients. We suspect this is a major factor in the unexplained heritability of the disease to date. Directly studying this heterogeneity by mapping a causal molecular trait greatly enhanced our power to detect novel genetic variations underlying SLE disease pathogenesis. The top loci in our study have not been previously reported in other case-control GWAS of SLE, and were not top loci in the initial case-control GWAS data set that we used in this study [16]. Thus, our alternative strategy was capable of finding genetic variants associated with disease that are not readily apparent in case-control designs, supporting a complexity in genetic architecture that will require molecular sub-phenotyping to fully delineate.

rs7897633, an intronic SNP in *PRKG1* ($p = 2.75 \times 10^{-8}$) was the strongest association observed in our study. This gene encodes the soluble isoforms of the cyclic GMP-dependent protein kinase (Iα and Iβ), which are important components of signal transduction processes in diverse cell types [118]. Canonical pathway analysis revealed this gene was associated with pathways such as synaptic long term depression, Dopamine-DARPP32 feedback in cyclic AMP signaling and netrin signaling; pathways in which modulation of cyclic GMP and cyclic AMP plays an important role in signaling and function. GIANT analysis supported biological function for PRKG1 in dendritic cells and NK cells, two cell types which cooperate to generate IFN-α in SLE [115]. PRKG1 function was not as strongly supported in other immune cells such as T and B lymphocytes, which are not

thought to be major IFN-α producing cells. It is not immediately clear how PRKG1 might impact type I IFN production or signaling, but mechanistic experiments directed at the IFN pathway in both dendritic and NK cells are likely to be informative.

The other strongly associated polymorphism in our study was a missense SNP in PNP (rs1049564, $P = 1.24 \times 10^{-7}$). PNP encodes the nucleoside phosphorylase enzyme, which is involved in purine metabolism. PNP together with adenosine deaminase (ADA), serve a key role in purine catabolism in the salvage pathway. Deficiency in this pathway leads to build up of elevated deoxy-GTP levels, resulting in T-cell toxicity and deficiency [119, 120]. Rare autosomal deficiency of the PNP gene results in a metabolic disorder characterized by defective T-cell and B-cell immunity as well as defective antibody responses [121, 122]. Interestingly, PNP-deficient patients have also been reported to develop autoimmune disorders, such as SLE, autoimmune hemolytic anemia, and idiopathic thrombocytopenic purpura [123]. The SNP identified in our study is a common coding-change variant which does not cause complete deficiency, and whether this variant results in some change in enzyme function is not currently known. In silico bioinformatic analysis using Polymorphism Phenotyping 2 (PolyPhen2) and Sorting Intolerant From Tolerant (SIFT) predicts this SNP as non-damaging, but an effect on enzyme activity would still be possible. There was strong representation of the purine metabolic pathway in our canonical pathway analysis, and *PNP* was the key molecule associated with this pathway. Some rare, highly penetrant variants in genes involved in nucleic acid metabolism have been associated with SLE, such as three prime repair exonuclease 1 (TREX1) and deoxyribonuclease I-like 3 (DNASE1L3) [124-126]. Given this precedent, PNP is a fascinating genetic association with SLE.

The non-European ancestral backgrounds studied were smaller, and did not allow for strong independent significance. Our discovery set was exclusively of European ancestry, and thus variants specific to other ancestral backgrounds could not have been discovered. Despite these limitations, we observed some interesting evidence for associations, which were of similar effect in European-American and African-American ancestral backgrounds. Intronic SNPs in *ANKRD44* and *PLEKHF2* were associated with IFN-α in African-American and European ancestral backgrounds (rs4850410, $P_{Meta} = 1.3 \times 10^{-6,}$ and rs297573, $P_{Meta} = 1.2 \times 10^{-4}$, respectively). *PLEKHF2* is an endosome-associated protein responsible for modulating the structure and function of endosomes, as well as the endocytotic process [127]. *PLEKHF2* can increase the activity of caspase 12, and a role in ER-related apoptotic pathway has been suggested [128]. ENCODE ChIP-seq data demonstrate that rs297573, the SNP downstream of *PLEKHF2* which was associated with IFN-α in our study, resides in the NFκB transcription factor binding site. *ANKRD44* has not been extensively studied, but it binds to the catalytic subunit of protein phosphatase 6 [129], which plays a role in cell cycle progression.

Our initial discovery cohort showed association of two SNPs in the *C7orf57* locus with serum IFN-α activity; however, this locus failed to replicate. One of the possible reasons for lack of replication could be that this locus was related to some unique feature of the discovery cohort that was not present in the replication cohort. One previous GWAS study of amyotrophic lateral sclerosis which used a very similar Illumina genotyping platform found evidence for association between these two SNPs and ALS which then failed to replicate in an additional independent replication cohort [130]. It is possible that some peculiarity of the earlier Illumina genotyping platform made it more likely for these SNPs

to be spuriously associated, although this locus was not associated in the original SLEGEN GWAS case-control study [16]. We used an entirely different rtPCR-based genotyping method for our replication cohort to eliminate potential platform-related biases. Another possibility is that this could have been a false-positive result due to statistical noise in our discovery cohort, which is why we could not replicate it in the independent replication cohort.

As referenced above, discovery methods followed by replication in non-European ancestral backgrounds would be an important next step to this work. It is likely that some polymorphisms will be ancestry specific, and will not be evident until a discovery strategy is used in that particular ancestral background. This would be especially important for African-Americans who have a higher incidence of SLE and more severe clinical manifestations [10]. African-American SLE cases also have higher levels of serum IFN-α activity [100], which could be one factor related to the increased incidence and severity of the disease. Our findings could have pharmacogenomic implications, as therapeutics targeting the IFN-α pathway are currently in development for SLE. Knowledge of the functional genetic factors underlying IFN-α dysregulation in a given patient could be useful in individualizing therapy with these agents.

2.5: METHODS

Samples and Genotyping

Discovery cohort:

Genome wide association study (GWAS) data from 755 SLE cases were obtained from multiple study centers as part of the international consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) [16]. The cohort studied by the SLEGEN consortium for GWAS in SLE consists of unrelated women of self-reported European ancestry and has been described in detail [16]. Out of 755 SLE cases, 400 cases had serum available for IFN-α analysis and were included in the discovery GWAS phase.

Samples were genotyped at 317,000 SNPs on the Illumina Infinium HumanHap300 genotyping Beadchip (Illumina Inc., San Diego, CA, USA). SNPs that failed the Hardy-Weinberg equilibrium test ($p<0.001$) were excluded, as were SNPs with a genotyping success rate < 95% or with a minor allele frequency < 0.05, resulting in 291,943 SNPs that were used in the analysis.

Replication cohort:

The independent multi-ethnic replication cohort of 1165 SLE patients was obtained from the Lupus Family Registry and Repository (LFRR) at the Oklahoma Medical Research Foundation and consisted of the following self-reported ancestral backgrounds: 715 European-Americans and 450 African- Americans. We incorporated 238 Hispanic/Native-American and 40 Asian-American SLE cases, in addition to the 1165 SLE cases, in the principal components analysis (PCA) to determine population stratification in the replication cohort. Clinical characteristics and demographic details for the patients in the replication cohort are summarized in **Supplementary Table 2.3**. Informed consent was obtained from all patients in both cohorts included in this study, and the study was approved by the institutional review boards at the respective institutions.

We followed up IFN-α associated SNPs, which had a $p < 1 \times 10^{-4}$ from the initial discovery GWAS analysis. SNPs that failed SNP assay design were excluded, resulting in 323 SNPs which were genotyped in the replication cohort. A separate panel of 334 ancestry-informative markers [131] was also genotyped in the replication cohort. SNPs were genotyped in genomic DNA using the Fluidigm Biomark microfluidic qPCR system (Fluidigm Corp, South San Francisco, CA, USA). All DNA samples were pre-amplified using the SNP-type primers from the genotyping assays, according to the manufacturer's protocol. PCR data were analyzed using the BioMark SNP Genotyping Analysis software version 3 (Fluidigm Corp) to obtain genotype calls. Scatter plots were all reviewed individually for quality, and SNPs that deviated significantly from the expected Hardy-Weinberg proportions ($P < 0.001$) or with < 95% genotyping success were excluded from the analysis.

Measurement of serum IFN-α activity

Enzyme-linked immunosorbent assay (ELISA) methods for the measurement of type I IFN in human sera have been complicated by low sensitivity and low specificity [132]. We used a well-documented sensitive and reproducible reporter cell assay to generate IFN-α activity data from patient sera [28, 133]. The reporter cells in this assay (WISH cells, ATCC #CCL-25, Manassas, VA, USA) measure the ability of patient sera to cause IFN-induced gene expression. These cells are an epithelial-derived cell line that is highly responsive to IFN-α. Cells are incubated with patient serum for 6 hours. Then real-time PCR is used to quantify three canonical IFN-α-induced transcripts in the WISH cell lysates (IFIT1, MX1 and PKR). Sera from healthy unrelated controls (n=200) were tested to establish a normal value for

the assay. Results from patient samples are expressed as the number of standard deviation (s.d.) above the mean of healthy unrelated control sera. The sum of the number of s.d. above healthy controls for the three transcripts is used as the quantitative output from the assay, representing a serum IFN-α activity score. This assay has been extremely informative in SLE and other autoimmune diseases [100, 111, 134].

Statistical Analysis

Control for Population Structure

To account for potential differences in admixture or population structure within self-reported ancestral backgrounds in the discovery and replication cohort, we performed a PCA using the GWAS SNPs and 334 independent ancestry-informative marker SNPs, respectively. PCA in the discovery cohort was carried out on all of the GWAS SNPs that passed quality control thresholds. This cohort is composed of SLE cases with self-reported European ancestry, and as shown in the principal components analysis plots (**Figure 2.3a**), there were no major population outliers. As expected, cases with varying proportions of Northern- and Southern-European ancestry were included in the study, and some cases cluster with the Ashkenazi Hap Map reference population, suggesting Jewish ancestry (**Figure 2.3b**).

PCA of the ancestry-informative markers genotyped in the replication cohort revealed that the PC1 obtained in this analysis provided a strong separation between subjects of self-reported African-American ancestry and the non-African ancestral backgrounds, while PCs 2 and 3 provided a separation between subjects of self-reported Asian-, Hispanic-, and European-American ancestry (**Figures 2.3c and d**). Self-reported

Hispanic-or Native-American ancestry subjects were largely overlapping in this analysis, and are considered together in these analyses. Association analyses were not performed in the Hispanic- or Native American (n=238) and Asian-American cases (n=40), due to the small number of subjects. These subjects were included in the principal component analysis of the AIMs to assist with the determination of population structure. Correction for population structure within the discovery and replication cohort was done using the first three PCs as covariates in the logistic regression association analyses. PCA analysis was performed using Cluster 3.0 software [135].

Association Analyses

Logistic regression analysis was used to detect associations between the SNPs and serum IFN-α in both stages of the study. IFN-α activity was studied as a categorical trait because the trait distribution is highly skewed, such that log transformation does not result in a normal distribution and the highly skewed data did not allow for linear modeling in a quantitative trait locus analysis. We used a binning strategy that has been highly informative in previous large scale studies and multivariate analyses of the serum IFN trait in SLE [28, 97, 100], in which subjects with a value > 2 s.d. above the mean of healthy controls are binned as high IFN-α, and the rest are binned as low IFN-α. Using this binning strategy prevents high outlying values from exerting an inordinate amount of influence in the model. In the discovery cohort, 88 were categorized as high IFN-α, and 322 were categorized as low IFN-α. Logistic regression analysis was carried out using PLINK v.1.07 software [136]. The first three PCs from the PCA of the GWAS SNPs were used as covariates in the logistic regression to control for population structure in the discovery cohort.

In the replication cohort, each self-reported ancestral background was analyzed separately, and the first three PCs were included as covariates to correct for population structure and admixture. Regression analysis was also performed to detect any potential associations between the presence of particular autoantibodies and SNPs in the replication cohort, because autoantibodies have been associated with high IFN-α in SLE patients [100]. In the replication cohort, we used the Benjamini–Hochberg procedure to control the false discovery rate at 0.05, and the SNPs which passed this threshold were considered for meta-analysis. The P-value threshold used for significance in the overall meta-analysis corrects for the number of SNPs which were analyzed for association in the initial GWAS discovery analysis, controlling the family-wise error rate at the 0.05 level.

For SNPs that demonstrated a homogenous effect across the discovery and replication sets by Breslow–Day testing, meta-analysis was performed using the weighted Z-score method [137] using R 2.11.1 statistical analysis software (www.r-project.org). For statistical correction of multiple comparisons, we applied a Bonferroni correction to the meta-analysis results using the number of SNPs that passed quality control in the discovery GWAS (n = 291 943), resulting in a threshold P-value for this study of $p < 1.71 \times 10^{-7}$. In the cross-ancestral background analysis, SNPs that demonstrated a homogenous effect across both ancestral backgrounds were meta-analyzed using the same weighted Z-score method [137], assuming a fixed-effect model. Enrichment P-values were calculated using a Fisher's exact test with the following parameters: for the SNP-wise calculation for the SLE-associated SNP, the number of possible confirmed SLE-risk SNPs in European ancestry was estimated at 40, and the number of SNPs that passed quality control in our GWAS screen was used as the denominator to establish the null proportion. The observed proportion

was one SNP out of the 323 SNPs in our replication list. For the locus-wise comparison for genes associated with circulating IFN in SLE, we estimated 18 loci, which have been previously associated with IFN in SLE, and 20 000 as the number of human gene loci to estimate the null proportion. The observed proportion was 2 loci out of the 277 loci represented by the 323 SNPs in the replication list.

Canonical pathway analysis

From the initial discovery GWAS data, IFN-α-associated SNPs (n = 323) with a $p < 1x10^{-4}$ were analyzed further using Ingenuity Pathway Analysis (IPA; Ingenuity Systems, www.ingenuity.com) to identify the top canonical pathways related to IFN-α-associated SNPs. SNPs were attributed to the nearby gene, and the genes were then compared with curated functional attribution lists organized by canonical pathway function. The magnitude of over-representation of a particular canonical pathway in the gene list from our study was calculated as the ratio of the number of molecules from our data set that map to the pathway divided by the total number of reference molecules in that pathway in the IPA database (a list of genes belonging to major canonical pathways is curated in IPA based on published literature). Statistical significance was determined using the Fisher's exact test, comparing the observed ratio of genes in a particular pathway to the null expectation (that the genes would assort proportionally across all IPA pathways), to estimate the probability that the observed over-representation of the particular pathway would arise by chance.

Genome-scale Integrated Analysis of gene Networks in Tissues (GIANT)

The top genes from the replication cohort were queried using the GIANT software program to determine likely functional relationships of these genes in various types of immune cells. GIANT is a public, web-based software program that uses tissue-specific gene expression databases to predict tissue-specific gene interactions (http://giant.princeton.edu/about/). About 145 tissues/cell types are available to be queried, including major immune cell subsets. The software generates functional networks based on the genes queried via the integration of thousands of publicly available gene expression datasets, sequence data, transcription factor binding sites and protein-protein interaction data to generate gene association matrix. Bayesian weights derived from the gold-standard tissue-specific datasets are then applied, and networks are generated for each tissue queried, which illustrate the most probable functional relationships between the queried genes and other molecules in that particular tissue. Network relationship confidence (edge weight) was set at a minimum of 0.4 for our analyses. After the networks were generated, we calculated overall network density and network strength of each of our study genes in each immune cell subset network. Network density (D) was defined as a ratio of the number of edges (E) to the number of possible edges, given by the binomial coefficient $\binom{N}{2}$, giving D=2E/N (N-1); where N=number of nodes. In these weighted networks, we calculated strength as the sum of a node's edge weights.

Prediction of the impact of coding-change SNPs

Prediction of consequences on protein structure and/or function of non-synonymous SNPs were evaluated using the prediction programs SIFT (http://sift.bii.a-star.edu.sg/) and PolyPhen (http://genetics.bwh.harvard.edu/pph/). These two programs

use algorithms to determine the likelihood that a particular coding-change polymorphism impacts protein-folding based upon local protein structure as well as the particular amino acid substitution.

2.6: Appendix: Supplementary Materials

Supplementary Table 2.1: List of top 10 SNPs associated with IFN-α in the GWAS discovery cohort

Chromosome	Locus	SNP	SNP type	Associated allele	Odds Ratio (95% CI)	P-value
7	*C7orf57*	rs2708912	missense	G	4.21 (2.33 – 7.62)	2.01×10^{-6}
7	*C7orf57*	rs2686792	intron	G	4.17 (2.29 – 7.59)	3.13×10^{-6}
7	*C7orf57*	rs2708890	missense	G	4.02 (2.22 – 7.29)	4.44×10^{-6}
10	*PRKG1*	rs7906944	intron	A	0.43 (0.30 – 0.62)	6.50×10^{-6}
10	*PRKG1*	rs7905063	intron	G	0.42 (0.29 – 0.62)	6.97×10^{-6}
10	*PRKG1*	rs7897633	intron	G	0.43 (0.29 – 0.62)	1.07×10^{-5}
14	*PNP*	rs1049564	missense	A	2.49 (1.65 – 3.75)	1.32×10^{-5}
2	*ANKRD44*	rs6730027	intron	T	2.57 (1.67 – 3.98)	2.11×10^{-5}
12	*ATF7IP*	rs10772783	intron	C	2.21 (1.53 – 3.18)	2.22×10^{-5}
10	*PRKG1*	rs9415777	intron	A	0.45 (0.31 – 0.65)	2.40×10^{-5}

Supplementary Table 2.2: List of top 323 SNPs from discovery GWAS genotyped in the replication cohort

Chr	SNP	P-value	Gene	Feature	Left gene	Right gene
1	rs7541937	0.000817	*DLGAP3*	Intron	*C1orf212*	*LOC653160*
1	rs3911861	0.008714	*C1orf164*	intron	*PRNPIP*	*TMEM53*
1	rs357210	0.009949	*NEGR1*	intron	*ZRANB2*	*LOC100132353*
1	rs11577464	7.08×10^{-5}	*NA*	NA	*LOC100133118*	*ST6GALNAC5*
1	rs17449554	0.000872	*NA*	NA	*LOC729779*	*ADH5P2*
1	rs1334336	0.000489	*NA*	NA	*LMO4*	*PKN2*
1	rs10489944	0.005257	*NA*	NA	*LMO4*	*PKN2*
1	rs7411387	0.001232	*CHIA*	intron	*RP11-165H20.1*	*C1orf88*
1	rs6537810	0.004346	*SYT6*	intron	*LOC100132906*	*MRP63P1*
1	rs946817	0.00037	*NA*	NA	*SEC16B*	*LOC100131700*
1	rs7544563	0.000578	*NA*	NA	*IVNS1ABP*	*HMCN1*
1	rs10926978	0.009515	*SDCCAG8*	intron	*CEP170*	*LOC729199*
1	rs11589847	0.000605	*PLD5*	intron	*LOC200149*	*LOC391183*
1	rs10924309	0.000799	*KIF26B*	intron	*LOC100128825*	*SMYD3*
2	rs9636493	0.000511	*NA*	NA	*MYT1L*	*LOC729897*
2	rs9287725	0.000371	*NA*	NA	*LOC645054*	*FLJ33534*
2	rs2380595	0.000727	*NA*	NA	*TRIB2*	*FAM84A*
2	rs4047462	6.17×10^{-5}	*NA*	NA	*LOC100128475*	*FAM49A*
2	rs875974	0.000792	*NA*	NA	*FLJ41481*	*OSR1*
2	rs13011502	0.000212	*OSR1*	near 3'	*FLJ41481*	*OSR1*
2	rs41462149	0.000255	*KLHL29*	intron	*FLJ14126*	*ATAD2B*
2	rs6547906	0.000504	*LOC165186*	intron	*WDR43*	*C2orf71*
2	rs2754530	0.004128	*SRD5A2*	intron	*XDH*	*AK2P2*
2	rs2123774	0.000658	*NA*	NA	*LTBP1*	*RASGRP3*
2	rs4670532	0.000885	*NA*	NA	*MRPL50P1*	*CRIM1*
2	rs7559001	0.000254	*EML4*	intron	*SGK493*	*COX7A2L*
2	rs2216784	0.002363	*NRXN1*	intron	*LOC130728*	*LOC730100*
2	rs1016387	0.000573	*NA*	NA	*NRXN1*	*LOC100128029*
2	rs746784	4.32×10^{-5}	*NA*	NA	*DNMT3AP1*	*LOC644838*
2	rs6546353	0.003299	*NA*	NA	*ETAA1*	*C1D*
2	rs7593084	0.000928	*EXOC6B*	intron	*LOC100128605*	*SPR*
2	rs4553845	0.000943	*NA*	NA	*C2orf3*	*LOC100129863*
2	rs885187	0.000762	*NA*	NA	*LOC647275*	*LRRTM4*
2	rs10519329	0.000766	*NA*	NA	*LOC647275*	*LRRTM4*
2	rs924901	4.25×10^{-5}	*NA*	NA	*LOC647275*	*LRRTM4*
2	rs1016347	5.20×10^{-5}	*NA*	NA	*LOC647275*	*LRRTM4*
2	rs7558427	0.009702	*ATOH8*	intron	*GNLY*	*ST3GAL5*
2	rs1192795	0.000754	*RNF149*	intron	*SNORD89*	*CREG2*
2	rs10188630	0.008033	*SH3RF3*	intron	*EDAR*	*LOC100132457*
2	rs1025736	0.000387	*NA*	NA	*EN1*	*MARCO*
2	rs1838999	0.002995	*NAP5*	intron	*LYPD1*	*LOC100130315*
2	rs1030599	0.000972	*NA*	NA	*LOC100128759*	*NR4A2*
2	rs956986	0.000302	*CCDC148*	intron	*UPP2*	*LOC100128061*
2	rs7574002	0.000811	*NA*	NA	*DLX2*	*ITGA6*
2	rs17777566	0.000993	*NA*	NA	*ZNF385B*	*KIAA1604*
2	rs1449264	0.000127	*ITGA4*	intron	*LOC100127923*	*CERKL*
2	rs4850410	0.006391	*ANKRD44*	intron	*PGAP1*	*LOC729342*

Supplementary Table 2.2 – continued.

Chr	SNP	P-value	Gene	Feature	Left gene	Right gene
2	rs1036542	0.000201	*ANKRD44*	near 3'	*PGAP1*	*ANKRD44*
2	rs6730027	2.11×10^{-5}	*ANKRD44*	intron	*PGAP1*	*LOC729342*
2	rs1429411	0.000956	*NA*	NA	*ANKRD44*	*LOC729342*
2	rs10168275	0.009742	*CAB39*	intron	*LOC645870*	*ITM2C*
2	rs1797399	0.000857	*NA*	NA	*NMUR1*	*LOC391490*
2	rs7577137	0.000252	*LOC339766*	intron	*UGT1A10*	*HJURP*
3	rs9854602	0.000602	*NA*	NA	*LOC402123*	*CNTN6*
3	rs6764561	0.000959	*GRM7*	intron	*MRPS36P1*	*LMCD1*
3	rs250403	0.000522	*RAD18*	near 3'	*OXTR*	*RAD18*
3	rs17009067	0.000232	*ZNF385D*	intron	*VENTXP7*	*LOC728516*
3	rs4245878	0.000448	*NA*	NA	*RPSAP11*	*LOC100129194*
3	rs9833530	0.000707	*NA*	NA	*RPSAP11*	*LOC100129194*
3	rs4642086	0.000553	*NA*	NA	*CCDC137P*	*RYBP*
3	rs6764864	0.000723	*NA*	NA	*LOC643766*	*HTR1F*
3	rs2399441	0.000879	*NA*	NA	*CD200R1L*	*CD200R1*
3	rs10511349	0.001417	*LSAMP*	intron	*GAP43*	*BZW1L1*
3	rs6766694	0.002214	*IQCB1*	intron	*GOLGB1*	*EAF2*
3	rs9813363	0.00014	*RAB6B*	intron	*SRPRB*	*C3orf36*
3	rs6439563	0.000805	*EPHB1*	intron	*LOC645218*	*PPP2R3A*
3	rs931726	0.000503	*EPHB1*	intron	*LOC645218*	*PPP2R3A*
3	rs9881418	0.003068	*SLC9A9*	intron	*LOC100128739*	*LOC257039*
3	rs9842818	0.000298	*NA*	NA	*PLSCR4*	*LOC440981*
3	rs2688692	0.000918	*PLSCR1*	intron	*PLSCR2*	*PLSCR5*
3	rs6777677	0.000887	*NA*	NA	*LOC646849*	*LOC344741*
3	rs4507220	0.00692	*NA*	NA	*LOC646849*	*LOC344741*
3	rs9846083	0.00089	*TNIK*	intron	*SLC2A2*	*PLD1*
3	rs1201292	0.000997	*TBL1XR1*	intron	*LOC730168*	*LOC339845*
3	rs13086642	0.000624	*NA*	NA	*BCL6*	*FLJ42393*
4	rs6446401	0.000821	*CRMP1*	intron	*EVC*	*LOC100128651*
4	rs4447863	0.009131	*SLC2A9*	intron	*LOC100131256*	*WDR1*
4	rs10517228	0.000923	*NA*	NA	*LOC645716*	*LOC642305*
4	rs6850606	0.000475	*NA*	NA	*ATP8A1*	*GRXCR1*
4	rs1531289	0.000838	*KDR*	intron	*LOC100132311*	*LOC100128865*
4	rs7654599	0.000138	*KDR*	intron	*LOC100132311*	*LOC100128865*
4	rs441785	0.000923	*NA*	NA	*LOC100131356*	*LOC644682*
4	rs524907	0.000183	*NA*	NA	*TIGD2*	*GPRIN3*
4	rs1377918	0.002944	*NA*	NA	*MGC48628*	*TMSL3*
4	rs1514733	0.000635	*MGC48628*	intron	*MMRN1*	*TMSL3*
4	rs10516939	0.000889	*GRID2*	intron	*LOC133083*	*ATOH1*
4	rs17625855	0.000516	*NA*	NA	*ZBED1P*	*LOC100133103*
4	rs1472076	0.000152	*NA*	NA	*LOC391686*	*LOC132719*
4	rs4438820	0.000488	*NA*	NA	*NT5C3P1*	*NDST3*
4	rs6535930	0.000887	*KIAA0922*	intron	*MND1*	*WDR45p*
4	rs6536595	0.000862	*FSTL5*	intron	*RAPGEF2*	*LOC729725*
4	rs2279932	0.000275	*AGA*	intron	*NEIL3*	*LOC285500*
4	rs2613024	0.000731	*NA*	NA	*LOC391719*	*hCG_2025798*
4	rs3796644	0.005047	*SORBS2*	intron	*PDLIM3*	*TLR3*
4	rs1879724	0.007361	*NA*	NA	*MGC39584*	*LOC728339*
4	rs13119686	0.00089	*NA*	NA	*MGC39584*	*LOC728339*
5	rs7702501	0.000411	*BRD9*	intron	*LOC100132536*	*TRIP13*

Supplementary Table 2.2 – continued.

5	rs30483	0.000529	NA	NA	*IRX1*	*LOC340094*
5	rs10512926	0.007331	*ADCY2*	intron	*LOC442132*	*C5orf49*
5	rs1666793	0.000722	NA	NA	*LOC729506*	*LOC100128382*
5	rs40687	0.000872	*SEMA5A*	intron	*LOC100128382*	*SNORD123*
5	rs1019810	0.000658	*FBXL7*	intron	*LOC391741*	*MARCH11*
5	rs6878131	0.009759	*FBXL7*	intron	*LOC391741*	*MARCH11*
5	rs2291114	0.001937	*PDZD2*	synonymous	*LOC100129608*	*GOLPH3*
5	rs2289876	0.000793	*UGT3A2*	intron	*UGT3A1*	*LMBRD2*
5	rs583595	0.000925	*UGT3A2*	intron	*UGT3A1*	*LMBRD2*
5	rs2287934	0.000596	*SKP2*	intron	*LMBRD2*	*C5orf33*
5	rs27130	0.000952	*SKP2*	intron	*LMBRD2*	*C5orf33*
5	rs1990977	0.000492	NA	NA	*LOC345645*	*LOC441073*
5	rs16893364	0.000543	NA	NA	*SDCCAG10*	*ADAMTS6*
5	rs2441109	0.00944	*MAST4*	intron	*LOC100129571*	*LOC100128443*
5	rs17732825	0.000271	*RGNEF*	intron	*UTP15*	*ENC1*
5	rs12514694	0.000909	*PDE8B*	intron	*ALDH7A1P1*	*WDR41*
5	rs10491245	0.000696	NA	NA	*FLJ41309*	*LOC100127911*
5	rs26521	0.000284	*HISPPD1*	intron	*GIN1*	*C5orf30*
5	rs250253	7.88×10^{-5}	*EFNA5*	intron	*LOC100129233*	*LOC345576*
5	rs26054	4.55×10^{-5}	NA	NA	*STARD4*	*C5orf13*
5	rs4957975	0.003668	NA	NA	*C5orf13*	*C5orf26*
5	rs1389849	0.007989	NA	NA	*C5orf13*	*C5orf26*
5	rs6897947	0.000867	*SEMA6A*	intron	*LOC100128691*	*LOC644146*
5	rs6887255	0.000544	NA	NA	*LOC100129374*	*LOC100130699*
5	rs884623	0.009212	*CYFIP2*	intron	*C5orf40*	*ICHTHYIN*
5	rs7711912	0.000501	NA	NA	*FGF18*	*C5orf50*
5	rs4074670	0.000498	NA	NA	*LOC100132848*	*LOC100129457*
5	rs3828686	0.000372	*GFPT2*	intron	*MAPK9*	*CNOT6*
6	rs17379732	0.000452	*NA*	NA	*F13A1*	*RP3-398D13.1*
6	rs9328444	0.009399	*BMP6*	intron	*RPL29P1*	*TXNDC5*
6	rs1753290	0.000646	*FGD2*	intron	*MTCH1*	*RP3-405J24.3*
6	rs3846755	0.009725	*CD109*	near 5'	*SLC17A5*	*CD109*
6	rs10457255	0.000331	NA	NA	*PA2G4P5*	*LOC643884*
6	rs2810169	0.000378	NA	NA	*LOC728590*	*LOC100132053*
6	rs6908717	0.000247	NA	NA	*LOC643954*	*hCG_1820801*
6	rs761840	0.000376	NA	NA	*TMEM200A*	*LOC285733*
6	rs9494022	0.000913	NA	NA	*FAM8A6P*	*LOC645175*
6	rs2846546	0.006682	*PARK2*	intron	*LOC100129958*	*PACRG*
6	rs9365514	0.000946	*PACRG*	intron	*PARK2*	*LOC729658*
6	rs9458956	0.000959	*LOC728275*	intron	*QKI*	*LOC728316*
6	rs1912668	0.007588	NA	NA	*LOC728275*	*LOC728316*
6	rs4709060	0.007301	NA	NA	*LOC728275*	*LOC728316*
6	rs7775504	0.008524	*WDR27*	intron	*LOC100130617*	*C6orf120*
6	rs1028488	0.00085	*DLL1*	near 3'	*LOC154449*	*DLL1*
7	rs1992025	0.000772	*TMEM195*	intron	*LOC100128217*	*MEOX2*
7	rs2191892	0.000494	NA	NA	*FERD3L*	*TWISTNB*
7	rs227951	0.000739	NA	NA	*LOC442517*	*CLK2P*
7	rs2107124	0.000505	NA	NA	*NPVF*	*LOC100131016*
7	rs2717907	0.000985	NA	NA	*NPVF*	*LOC100131016*
7	rs13224312	0.000587	*RALA*	intron	*LOC646999*	*LOC349114*
7	rs7780837	0.000881	*PKD1L1*	intron	*FLJ21075*	*HUS1*

Supplementary Table 2.2 – continued.

7	rs2708912	2.01 x 10^{-6}	*C7orf57*	missense	*SUNC1*	*UPP1*
7	rs2686792	3.13 x 10^{-6}	*C7orf57*	intron	*SUNC1*	*UPP1*
7	rs7794902	0.000402	NA	NA	*STAG3L4*	*AUTS2*
7	rs12698713	0.000965	NA	NA	*STAG3L4*	*AUTS2*
7	rs215276	0.000482	*SEMA3E*	intron	*PCLO*	*LOC100130572*
7	rs2371877	8.28 x 10^{-5}	NA	NA	*LOC100130572*	*SEMA3A*
7	rs6967487	0.009988	NA	NA	*LOC100128334*	*ZKSCAN1*
7	rs4727499	0.009263	NA	NA	*EMID2*	*MYLC2PL*
7	rs1017607	7.99 x 10^{-5}	NA	NA	*SND1*	*LOC100131212*
7	rs12706827	0.000969	NA	NA	*SND1*	*LOC100131212*
7	rs6467557	5.00 x 10^{-5}	*CALD1*	intron	*LOC100130187*	*AGBL3*
7	rs10250570	0.009738	*CNTNAP2*	intron	*LOC643308*	*tcag7.1231*
7	rs916514	0.000851	*DPP6*	intron	*DPP6*	*LOC100132707*
7	rs13221118	0.000744	*MNX1*	intron	*NOM1*	*LOC645249*
7	rs1049329	0.000128	*PTPRN2*	3' UTR	*tcag7.1023*	*LOC100127991*
8	rs11137053	0.000365	*LOC100132301*	intron	*MCPH1*	*AGPAT5*
8	rs9918794	0.000769	NA	NA	*CHMP7*	*R3HCC1*
8	rs6988827	0.000977	NA	NA	*NKX2-6*	*STC1*
8	rs13256023	0.000717	*CHD7*	intron	*RAB2A*	*LOC442389*
8	rs2279572	0.000372	*RLBP1L1*	intron	*NPM1P6*	*ASPH*
8	rs2350620	5.98 x 10^{-5}	*ASPH*	intron	*hCG_1988300*	*LOC645551*
8	rs2882460	3.60 x 10^{-5}	*ASPH*	intron	*hCG_1988300*	*LOC645551*
8	rs6549	0.000516	*ASPH*	3' UTR	*hCG_1988300*	*LOC645551*
8	rs7812327	0.000312	*ASPH*	intron	*hCG_1988300*	*LOC645551*
8	rs11990408	9.84 x 10^{-5}	*ASPH*	intron	*hCG_1988300*	*LOC645551*
8	rs11783343	0.000967	NA	NA	*ASPH*	*LOC645551*
8	rs1434937	0.000218	*C8orf34*	intron	*LOC728774*	*LOC100129096*
8	rs7016101	0.000626	*C8orf34*	intron	*LOC728774*	*LOC100129096*
8	rs1481278	0.000897	NA	NA	*LOC100129096*	*LOC100129809*
8	rs1866897	0.001001	*SULF1*	intron	*LOC100129809*	*SLCO5A1*
8	rs1440333	0.009587	*KCNB2*	intron	*LOC100129527*	*TERF1*
8	rs297573	0.000791	NA	NA	*PLEKHF2*	*C8orf37*
8	rs7839523	0.000792	*DDEF1*	intron	*FAM49B*	*DDEF1IT1*
8	rs2649127	0.00343	NA	NA	*KHDRBS3*	*LOC100129367*
9	rs2380941	0.000733	*GLIS3*	intron	*C9orf70*	*SLC1A1*
9	rs303723	6.73 x 10^{-5}	NA	NA	*NFIB*	*ZDHHC21*
9	rs4961497	0.000681	*BNC2*	intron	*LOC648570*	*CNTLN*
9	rs883966	0.009408	*PAX5*	intron	*MELK*	*LOC100128706*
9	rs2768659	0.007333	*GRHPR*	intron	*ZCCHC7*	*LOC100132896*
9	rs662975	0.000565	NA	NA	*TRPM3*	*TMEM2*
9	rs1329778	0.000146	NA	NA	*TRPM3*	*TMEM2*
9	rs2771090	0.000116	NA	NA	*OR7E116P*	*LOC340515*
9	rs7872276	0.000896	NA	NA	*OR13C4*	*OR13C3*
9	rs6477693	0.000358	*C9orf4*	intron	*C9orf5*	*EPB41L4B*
9	rs1887521	0.000159	*PALM2*	intron	*LOC402375*	*LOC100131672*
9	rs12555920	0.006403	*CTNNAL1*	intron	*C9orf6*	*C9orf5*
9	rs2767762	0.000235	NA	NA	*LHX2*	*NEK6*
9	rs4962060	0.000429	NA	NA	*NTNG2*	*SETX*
9	rs4363274	0.000895	NA	NA	*RXRA*	*COL5A1*
9	rs1891999	0.000978	NA	NA	*LOC401557*	*C9orf62*
10	rs3750685	0.000172	*ADARB2*	intron	*C10orf109*	*LOC100129465*

Supplementary Table 2.2 – continued.

10	rs11256581	0.000483	NA	NA	*TCEB1P3*	*LOC254312*
10	rs749232	0.000411	*FAM107B*	intron	*FRMD4A*	*ARMETL1*
10	rs1904694	6.40×10^{-5}	*PRKG1*	intron	*A1CF*	*CSTF2T*
10	rs1904683	8.86×10^{-5}	*PRKG1*	intron	*A1CF*	*CSTF2T*
10	rs7897633	1.07×10^{-5}	*PRKG1*	intron	*A1CF*	*CSTF2T*
10	rs7906944	6.50×10^{-6}	*PRKG1*	intron	*A1CF*	*CSTF2T*
10	rs7097412	0.009661	*KIAA1274*	intron	*NODAL*	*PRF1*
10	rs2031517	0.004761	*ZMIZ1*	intron	*LOC283050*	*PPIF*
10	rs1649949	0.00057	*NRG3*	intron	*LOC727960*	*LOC728027*
10	rs7069120	0.00013	*C10orf59*	intron	*LOC100128990*	*LIPJ*
10	rs6586129	0.000104	*C10orf59*	intron	*LOC100128990*	*LIPJ*
10	rs809812	0.00087	*FER1L3*	intron	*LOC643863*	*CEP55*
10	rs2094405	0.008948	*TCF7L2*	intron	*LOC143188*	*hCG_1776259*
10	rs1537685	7.28×10^{-5}	*ATRNL1*	intron	*TRUB1*	*GFRA1*
10	rs845079	0.000555	NA	NA	*LOC100131719*	*GPR26*
10	rs4363506	0.000913	NA	NA	*DOCK1*	*NPS*
10	rs7076452	0.000319	NA	NA	*DOCK1*	*NPS*
11	rs11043097	0.000939	NA	NA	*LOC729013*	*GALNTL4*
11	rs2938282	0.005736	*SOX6*	intron	*INSC*	*AKR1B1P3*
11	rs4944448	0.000972	NA	NA	*C11orf76*	*LOC100133306*
11	rs4923611	0.002851	*NELL1*	intron	*LOC100130160*	*ANO5*
11	rs1374616	0.000401	*MPPED2*	intron	*C11orf46*	*DCDC5*
11	rs3818229	0.00093	*TCP11L1*	intron	*DEPDC7*	*PIGCP1*
11	rs570098	0.000835	*MACROD1*	intron	*OTUB1*	*FLRT1*
11	rs542941	0.000399	NA	NA	*RBM7*	*REXO2*
11	rs7931871	0.008592	NA	NA	*ODZ4*	*LOC646112*
11	rs7949150	0.008506	NA	NA	*LOC100129203*	*FAM76B*
11	rs4922828	0.000163	*NELL1*	intron	*LOC100130160*	*ANO5*
11	rs2886189	0.000337	NA	NA	*DRD2*	*TMPRSS5*
11	rs4245155	0.000129	NA	NA	*DRD2*	*TMPRSS5*
11	rs11214985	0.00054	NA	NA	*RBM7*	*REXO2*
11	rs3934007	0.000486	NA	NA	*DRD2*	*TMPRSS5*
11	rs633745	0.000833	*PKNOX2*	intron	*LOC729492*	*FLJ30719*
11	rs3740898	0.000959	*PKNOX2*	intron	*LOC729492*	*FLJ30719*
11	rs2155314	3.58×10^{-5}	*KIRREL3*	intron	*ST3GAL4*	*PRR10*
11	rs1506876	0.002533	*OPCML*	intron	*LOC100128095*	*LOC646522*
12	rs10491958	0.00086	NA	NA	*ERC1*	*FBXL14*
12	rs4765914	0.009348	*CACNA1C*	intron	*DCP1B*	*LOC100129797*
12	rs887304	0.000235	*EFCAB4B*	3' UTR	*UNQ3104*	*PARP11*
12	rs720333	0.000766	NA	NA	*FGF23*	*FGF6*
12	rs526654	0.000212	NA	NA	*KCNA5*	*LOC387826*
12	rs1963810	0.00035	NA	NA	*KCNA5*	*LOC387826*
12	rs1047771	0.000737	*LEPREL2*	missense	*GPR162*	*GNB3*
12	rs7312042	0.000312	*ATF7IP*	intron	*LOC644693*	*FLJ22662*
12	rs10772783	2.22×10^{-5}	*ATF7IP*	intron	*LOC644693*	*FLJ22662*
12	rs2900333	0.000547	NA	NA	*ATF7IP*	*FLJ22662*
12	rs10841614	0.00063	NA	NA	*SLCO1C1*	*SLCO1B3*
12	rs163117	0.000834	NA	NA	*LHX5*	*LOC100129739*
13	rs9576827	0.000668	*LHFP*	intron	*TNAP*	*COG6*
13	rs599909	0.000401	NA	NA	*ATXN8OS*	*DACH1*
13	rs9573126	0.000555	NA	NA	*FABP5L1*	*LOC730242*

Supplementary Table 2.2 – continued.

13	rs7989815	0.000791	NA	NA	*RP11-114G1.1*	*LOC100129260*
13	rs9514046	0.000809	NA	NA	*C13orf39*	*FLJ40176*
14	rs1049564	1.32×10^{-5}	*NP*	missense	*TMEM55B*	*GAFA1*
14	rs1713420	4.25×10^{-5}	*NP*	intron	*TMEM55B*	*GAFA1*
14	rs10162514	0.00061	NA	NA	*RPS15AP3*	*RPL18P1*
14	rs17118957	0.00084	NA	NA	*RPS15AP3*	*RPL18P1*
14	rs1555233	0.000537	NA	NA	*RPS15AP3*	*RPL18P1*
14	rs8019172	0.001	NA	NA	*COX5AP2*	*PTGDR*
14	rs1951210	0.000182	*PPP2R5E*	intron	*GPHB5*	*LOC100129928*
14	rs10139749	0.000192	*PLEKHH1*	intron	*C14orf83*	*PIGH*
14	rs740505	3.30×10^{-5}	*COQ6*	intron	*FAM161B*	*ENTPD5*
14	rs4899503	0.000496	NA	NA	*LIN52*	*VSX2*
14	rs10484153	0.000318	NA	NA	*LOC730105*	*RNU3P3*
14	rs2110706	0.000239	*FOXN3*	intron	*CAP2P1*	*LOC400236*
14	rs3759722	0.007979	*CPSF2*	intron	*NDUFB1*	*SLC24A4*
14	rs10484068	0.000754	*BCL11B*	intron	*RPL3P4*	*SETD3*
15	rs1463408	0.000361	NA	NA	*TRPM1*	*LOC283710*
15	rs1157619	0.007113	*MEIS2*	intron	*LOC145845*	*LOC390576*
15	rs11071319	0.000598	*CGNL1*	intron	*LOC100128711*	*GCOM1*
15	rs875339	0.000756	*RORA*	intron	*CYCSP38*	*VPS13C*
15	rs290312	0.000302	NA	NA	*LOC100128015*	*MGC15885*
15	rs2053294	0.000773	*SMAD3*	intron	*LOC100131796*	*FLJ11506*
15	rs17526330	0.004241	*RGMA*	intron	*CHD2*	*LOC100124334*
15	rs4965671	0.000684	NA	NA	*LASS3*	*PRKXP1*
16	rs4608354	0.000641	*A2BP1*	intron	*LOC100131413*	*LOC100131080*
16	rs1573638	0.000596	NA	NA	*LOC729993*	*ERCC4*
16	rs1143678	0.007228	*ITGAM*	missense	*TRIM72*	*ITGAX*
16	rs7204044	0.007594	*SLC12A3*	intron	*NUP93*	*HERPUD1*
16	rs8044442	0.00076	*CFDP1*	intron	*BCAR1*	*TMEM170A*
16	rs11149991	0.000239	*ADAMTS18*	intron	*VN2R10P*	*NUDT7*
16	rs1155970	2.52×10^{-5}	*CDH13*	intron	*MPHOSPH6*	*HSBP1*
16	rs16959371	0.000691	*CDH13*	intron	*MPHOSPH6*	*HSBP1*
16	rs4782395	0.000775	*MVD*	intron	*CYBA*	*SNAI3*
16	rs11648894	0.000745	*RNF166*	intron	*SNAI3*	*C16orf84*
17	rs13422	0.000851	*PMP22*	3' UTR	*LOC441781*	*TEKT3*
17	rs9635758	0.00065	NA	NA	*LOC388401*	*CA10*
17	rs9904424	0.005124	*PSMD12*	intron	*HELZ*	*LOC729822*
17	rs7219896	0.006201	*KIAA1303*	intron	*LOC201259*	*LOC100128105*
18	rs11664521	0.000248	NA	NA	*SMCHD1*	*EMILIN2*
18	rs4602126	0.000707	*RNF165*	intron	*C18orf23*	*LOXHD1*
18	rs2046241	0.000489	*KIAA0427*	intron	*LOC100130666*	*SMAD7*
18	rs2045154	0.000803	*DCC*	intron	*LOC100132995*	*LOC100133176*
18	rs11151299	0.000805	NA	NA	*CDH19*	*DSEL*
18	rs7242877	0.000979	NA	NA	*MBP*	*GALR1*
18	rs2850855	9.92×10^{-5}	NA	NA	*MBP*	*GALR1*
19	rs1982074	0.000802	*KRI1*	missense	*ATG4D*	*CDKN2D*
19	rs1549951	7.28×10^{-5}	NA	NA	*ZNF536*	*TSHZ3*
19	rs757638	0.008229	*HIF3A*	intron	*IGFL1*	*PPP5C*
19	rs7259731	0.000109	NA	NA	*LILRA5*	*LILRA4*
19	rs2889010	0.000898	*C19orf18*	intron	*LOC646820*	*ZNF606*
20	rs6039134	0.008644	*ANGPT4*	intron	*FAM110A*	*RSPO4*

Supplementary Table 2.2 – continued.

20	rs2299676	0.003536	*PLCB4*	intron	*PLCB1*	*C20orf103*
20	rs6134059	0.000106	NA	NA	*JAG1*	*FAT1P1*
20	rs11907253	0.000185	*COX4I2*	missense	*ID1*	*BCL2L1*
20	rs6060627	0.001223	*BCL2L1*	intron	*COX4I2*	*TPX2*
20	rs6058381	6.05×10^{-5}	*BCL2L1*	intron	*COX4I2*	*TPX2*
20	rs6067709	0.000333	NA	NA	*RPSAP1*	*NFATC2*
20	rs856336	0.000975	NA	NA	*MRPS33P4*	*RPL36P1*
20	rs856327	0.000283	NA	NA	*MRPS33P4*	*RPL36P1*
21	rs723855	0.007739	NA	NA	*HSPA13*	*SAMSN1*
21	rs2284568	0.00019	*ITSN1*	intron	*CRYZL1*	*ATP5O*
21	rs2835561	0.000411	NA	NA	*DSCR6*	*PIGP*
21	rs6586230	0.000317	NA	NA	*C21orf129*	*RIPK4*
21	rs2839437	0.000796	NA	NA	*ZNF295*	*C21orf121*
21	rs4819077	0.000953	NA	NA	*C21orf93*	*COL18A1*
22	rs933241	0.000959	NA	NA	*CYTH4*	*ELFN2*
22	rs5750457	0.000314	NA	NA	*LGALS2*	*GGA1*
22	rs5757387	0.000214	NA	NA	*CBX6*	*APOBEC3A*
22	rs5768213	4.68×10^{-5}	NA	NA	*RP11-191L9.1*	*LOC388915*
23	rs6520279	0.00044	NA	NA	*NA*	*NA*
23	rs5924090	0.000653	NA	NA	*KLHL4*	*RPSAP15*
23	rs5924103	0.000621	NA	NA	*RPSAP15*	*MRPS22P1*
23	rs7062843	0.000842	NA	NA	NA	NA

Supplementary Table 2.3: Clinical and serologic characteristics of the replication cohort

	Clinical Feature	Cases			
		AA n=450	EA n=715	His n=238	Asian n=40
Demographic characteristics	Age, years*	43 (33-51)	46 (37-54)	40 (31-49)	30 (25-44)
	Female, no (%)	407 (90)	626 (88)	212 (89)	37 (92)
ACR Clinical Criteria for SLE	Malar rash, no (%)	182 (40)	408 (57)	124 (52)	25 (63)
	Discoid rash, no (%)	114 (25)	131 (18)	44 (19)	3 (8)
	Photosensitivity, no (%)	207 (46)	498 (70)	151 (63)	21 (53)
	Oral ulcers, no (%)	121 (27)	286 (40)	80 (34)	11 (27.5)
	Arthritis, no (%)	346 (77)	571 (80)	169 (71)	23 (58)
	Serositis, no (%)	190 (42)	281 (39)	83 (35)	14 (35)
	Renal disorder, no (%)	242 (54)	246 (34)	132 (56)	23 (58)
	Neurological disorder, no (%)	92 (20)	123 (17)	44 (19)	7 (18)
	Hematological disorder, no (%)	304 (68)	414 (58)	145 (61)	29 (73)
	Immunological disorder, no (%)	358 (80)	524 (73)	198 (83)	36 (90)
Prevalence of Specific Autoantibody Profiles	ANA, no (%)	434 (96)	653 (91)	221 (93)	39 (98)
	Ro, no (%)	110 (24)	139 (19)	51 (21)	11 (28)
	La, no (%)	24 (5)	44 (6)	19 (8)	2 (5)
	Sm, no (%)	56 (12)	12 (2)	15 (6)	2 (5)
	RNP, no (%)	176 (39)	57 (7)	41 (17)	11 (28)
	DNA, no (%)	143 (32)	173 (24)	54 (23)	17 (43)

* values are the median (interquartile range).
AA = African American, EA = European-American, His = Hispanic-American/Native American Association analyses were not performed in the Hispanic-American/Native American (n=238) and Asian-American cases (n=40), due to the small number of subjects, but these subjects were included in the ancestry analysis to help with the designation of genetic ancestry and admixture in our cohort.

Supplementary Table 2.4: Autoantibody associations observed in European SLE patients in the replication cohort

CHR	Locus	SNP	SNP type	OR (95% CI)	P value	Antibody
1	*LOC730102*	rs946817	Intron	1.76 (1.19-2.61)	4.68×10^{-3}	ANA
3	*SLC9A9*	rs9881418	Intron	0.53 (0.32-0.86)	1.09×10^{-2}	ANA
6	*LOC105378111*	rs4709060	Intergenic	0.79 (0.67-0.94)	7.63×10^{-3}	Anti_dsDNA
16	*CDH13*	rs16959371	Intron	1.39 (1.08-1.79)	9.78×10^{-3}	Anti_dsDNA
6	*LOC105378111*	rs1912668	Intergenic	0.8 (0.68-0.95)	1.01×10^{-2}	Anti_dsDNA
21	*RIPPLY3*	rs2835561	Intergenic	0.77 (0.63-0.95)	1.19×10^{-2}	Anti_dsDNA
1	*STARD4-AS1*	rs17449554	Intron	0.64 (0.45-0.91)	1.34×10^{-2}	Anti_dsDNA
5	*STARD4-AS1*	rs26054	Intron	0.65 (0.47-0.92)	1.49×10^{-2}	Anti_dsDNA
17	*RPTOR*	rs7219896	Intron	0.74 (0.58-0.95)	1.82×10^{-2}	Anti_dsDNA
12	*ATF7IP*	rs7312042	Intron	0.52 (0.36-0.74)	3.40×10^{-4}	Anti_La
12	*ATF7IP*	rs10772783	Intron	0.5 (0.34-0.73)	4.16×10^{-4}	Anti_La
4	*KDR*	rs7654599	Intron	0.61 (0.44-0.86)	4.11×10^{-3}	Anti_La
12	*LINC01234*	rs163117	Intergenic	0.52 (0.31-0.87)	1.32×10^{-2}	Anti_La
12	*ATF7IP*	rs2900333	3' UTR	2.41 (1.17-4.95)	1.71×10^{-2}	Anti_La
10	*FAM107B*	rs749232	Intron	0.45 (0.23-0.87)	1.77×10^{-2}	Anti_La
11	*PKNOX2*	rs3740898	Intron	0.65 (0.49-0.86)	2.25×10^{-3}	Anti_RNP
19	*LILRA4*	rs7259731	Near 3'	0.62 (0.45-0.85)	3.15×10^{-3}	Anti_RNP
7	*LAMTOR4*	rs7785392	Intron	0.75 (0.62-0.91)	4.41×10^{-3}	Anti_RNP
2	*ANKRD44*	rs6730027	Intron	0.73 (0.58-0.91)	5.70×10^{-3}	Anti_RNP
10	*NRG3*	rs1649949	Intron	1.36 (1.09-1.7)	7.53×10^{-3}	Anti_RNP
8	*STC1*	rs6988827	Intergenic	0.41 (0.21-0.81)	1.04×10^{-2}	Anti_RNP
10	*PRKG1*	rs1904683	Intron	1.32 (1.07-1.64)	1.13×10^{-2}	Anti_RNP
2	*CRIM1*	rs4670532	Intergenic	0.68 (0.5-0.93)	1.42×10^{-2}	Anti_RNP
10	*PRKG1*	rs7897633	Intron	1.44 (1.08-1.92)	1.44×10^{-2}	Anti_RNP

Supplementary Table 2.4 – continued.

CHR	*LOC102724145*	SNP	SNP type	OR (95% CI)	P value	Antibody
3	*PALD1*	rs17009067	Intron	0.77 (0.63-0.95)	1.65×10^{-2}	Anti_RNP
3	*ATF7IP*	rs6777677	Intergenic	0.75 (0.63-0.9)	1.63×10^{-3}	Anti_Ro
10	*EFNA5*	rs7097412	Intron	0.67 (0.5-0.89)	6.31×10^{-3}	Anti_Ro
12	*TENM4*	rs10772783	Intron	0.78 (0.65-0.94)	8.03×10^{-3}	Anti_Ro
5	*TLN2*	rs250253	Intron	1.32 (1.07-1.62)	8.37×10^{-3}	Anti_Ro
11	*ATF7IP*	rs7931871	Intron	0.77 (0.63-0.94)	8.62×10^{-3}	Anti_Ro
15	*LOC101927661*	rs290312	Intergenic	0.78 (0.64-0.94)	1.06×10^{-2}	Anti_Ro
12	*EPHB1*	rs7312042	Intron	0.79 (0.66-0.95)	1.10×10^{-2}	Anti_Ro
2	*FAM179A*	rs746784	Intron	0.76 (0.61-0.94)	1.35×10^{-2}	Anti_Ro
3	*PLEKHF2*	rs931726	Intron	0.72 (0.55-0.94)	1.41×10^{-2}	Anti_Ro
2	*ANKRD44*	rs6547906	Intron	0.8 (0.66-0.96)	1.67×10^{-2}	Anti_Ro
8	*ANKRD44*	rs297573	Near 3'	1.33 (1.05-1.68)	1.97×10^{-2}	Anti_Ro
2	*BCL11B*	rs1036542	Intron	0.54 (0.35-0.81)	3.26×10^{-3}	Anti_Sm
2	*LOC730102*	rs6730027	Intron	0.59 (0.41-0.86)	5.63×10^{-3}	Anti_Sm
14	*SLC9A9*	rs10484068	Intron	0.44 (0.24-0.8)	6.84×10^{-3}	Anti_Sm

Supplementary Table 2.5: Autoantibody associations observed in African American SLE patients in the replication cohort

CHR	Locus	SNP	SNP type	OR (95% CI)	P value	Antibody
15	*TRPM1*	rs1463408	near 5' UTR	3.08 (1.44-6.6)	3.74×10^{-3}	ANA
14	*MDGA2*	rs10162514	Intron	3.21 (1.32-7.79)	1.00×10^{-2}	ANA
8	*CLVS1*	rs2279572	Intron	2.91 (1.24-6.83)	1.42×10^{-2}	ANA
9	*GRHPR*	rs2768659	Intron	2.52 (1.34-4.74)	4.18×10^{-3}	Anti_dsDNA
8	*ASAP1*	rs7839523	Intron	0.69 (0.5-0.94)	1.90×10^{-2}	Anti_dsDNA
8	*TNFRSF10A*	rs9918794	Near 3'	0.4 (0.23-0.7)	1.36×10^{-3}	Anti_La
2	*CCDC85A*	rs1159916	Intron	0.4 (0.21-0.77)	5.90×10^{-3}	Anti_La
16	*ERCC4*	rs1573638	Intergenic	0.4 (0.21-0.78)	6.74×10^{-3}	Anti_La
4	*FSTL5*	rs6536595	Intron	0.38 (0.19-0.78)	8.29×10^{-3}	Anti_La
3	*TNIK*	rs9846083	Intron	0.34 (0.14-0.81)	1.49×10^{-2}	Anti_La
4	*SLC2A9*	rs4447863	Intron	0.48 (0.26-0.88)	1.88×10^{-2}	Anti_La
10	*NRG3*	rs1649949	Intron	0.61 (0.45-0.81)	6.29×10^{-4}	Anti_RNP
12	*EFCAB4B*	rs887304	3' UTR	0.5 (0.29-0.85)	1.12×10^{-2}	Anti_RNP
1	*ST6GALNAC5*	rs11577464	Intergenic	1.82 (1.15-2.9)	1.12×10^{-2}	Anti_RNP
2	*NRXN1*	rs2216784	Intron	0.43 (0.22-0.83)	1.24×10^{-2}	Anti_RNP
13	*LOC105370255*	rs599909	Intergenic	1.72 (1.12-2.65)	1.37×10^{-2}	Anti_RNP
10	*PALD1*	rs7097412	Intron	0.54 (0.37-0.78)	9.14×10^{-4}	Anti_Ro
2	*LOC101927661*	rs746784	Intron	0.58 (0.39-0.86)	7.32×10^{-3}	Anti_Ro
11	*TENM4*	rs7931871	Intron	1.52 (1.11-2.09)	9.52×10^{-3}	Anti_Ro
7	*RALA*	rs13224312	Intron	0.63 (0.44-0.91)	1.40×10^{-2}	Anti_Ro
17	*PMP22*	rs13422	3'UTR	0.61 (0.41-0.91)	1.52×10^{-2}	Anti_Ro
14	*BCL11B*	rs10484068	Intron	0.37 (0.16-0.84)	1.81×10^{-2}	Anti_Sm

Supplementary Table 2.6: Top 5 enriched networks by Ingenuity Pathway Analysis

ID	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
1	AKAP12, AKAP13, ALS2, DCC, DYNC1I2, ERK1/2, FAIM3, FSH, GFPT2, GLIS3, GTPase, IFT140, Insulin, Lh, MGAT3, MYO10, NRG3, NRP2, PDE1C, PDE8B, PRKAC, PRKG1, Proinsulin, PTPRN2, RGMA, RNF165, SEMA3E, SH3BP4, SLC25A12, Smad1/5/8, SORBS1, Sos, SYN1, SYN3, ZNF423	57	26	Cell Morphology, Cellular Compromise, Cellular Assembly and Organization
2	Actin, Akt, BCL11B, CALD1, Calmodulin, CDH13, CNTNAP2, EPB41L2, ERK, estrogen receptor, EXOC6B, F Actin, Focal adhesion kinase, GRIA2, Histone h3, Histone h4, IL33, Immunoglobulin, LPP, Mapk, OSBPL10, PDGF BB, PI3K (complex), Pka, PNP, RALA, SCD5, SLFN12L, SRC (family), Tgf beta, TRPM3, Vegf, YEATS2, ZFPM2, ZMIZ1	34	18	Cancer, Cell-mediated Immune Response, Cellular Development
3	APBB2, CALM1 (includes others), CBX1, CCDC168, COQ6, CSNK1A1, CSNK2A1, DOPEY2, E2F1, ERC1, FAM114A1, FAM83B, FIS1, HSP90AA1, HSP90AB1, KIAA0922, KIF20A, LCORL, LDLRAD4, MAGEA1, MCTP2, MID1, MIEF1, MYH9, PADI2, PAK3, PARK2, PRR14L, RAB6B, RABGAP1, REXO2, SLC39A14, THUMPD3, UBC, ZBTB16	28	15	Cancer, Endocrine System Disorders, Organismal Injury and Abnormalities
4	ADRB2, ALDH6A1, BNC2, CALD1, CAPZA2, CEP76, CHD7, CWC22, DCTD, EDIL3, FSCN1, KANK1, KCNK10, KHDRBS2, miR-124-3p (and other miRNAs w/seed AAGGCAC), MYH9, NDEL1, PAFAH1B1, PAFAH1B2, PAFAH1B3, PINX1, PLS3, PPP1CA, PPP1R18, RB1, RBFOX1, RBM24, SASS6, SLC7A14, TBC1D15, TMOD3, TPM2, UBC, UHMK1, ZCCHC24	23	13	Lipid Metabolism, Small Molecule Biochemistry, Cellular Development
5	APP, ASPH, BDNF, beta-estradiol, CAB39L, CAMK1, CaMKII, CDADC1, CDC37, Ck2, corticosterone, CWF19L2, EIF3M,F10, Actin, FOXN3, KIRREL3, LINGO2, LURAP1L, MARCKS, Metalloprotease, MPPED2, NAPB, NPHS1, NXPH1, PCP4L1, PPP6R2, RAB3A, RELN, Ryr, SHC1, SYN2, TJP1, TPM1, UBXN7	19	12	Behavior, Nervous System Development and Function, Cell-To-Cell Signaling and Interaction

CHAPTER 3: Patterns of transcriptional response to 1,25-dihydroxyvitamin D_3 and bacterial lipopolysaccharide in primary human monocytes[2]

3.1: Abstract

The active form of vitamin D, 1,25-dihydroxyvitamin D3 (1,25D), plays an important immunomodulatory role, regulating transcription of genes in the innate and adaptive immune system. The present study examines patterns of transcriptome-wide response to 1,25D and the bacterial lipopolysaccharide (LPS) in primary human monocytes, to elucidate pathways underlying the effects of 1,25D on the immune system. Monocytes obtained from healthy individuals of African-American and European-American ancestry were treated with 1,25D, LPS, or both simultaneously. The addition of 1,25D during stimulation with LPS induced significant up-regulation of genes in the antimicrobial and autophagy pathways and down-regulation of pro-inflammatory response genes compared to LPS treatment alone. A joint Bayesian analysis enabled clustering of genes into patterns of shared transcriptional response across treatments. The biological pathways enriched within these expression patterns highlighted several mechanisms through which 1,25D could exert its immunomodulatory role. Pathways such as mTOR signaling, EIF2 signaling, IL-8 signaling and Tec Kinase signaling were enriched among genes with opposite transcriptional responses to 1,25D and LPS, respectively, highlighting the important roles of these pathways in mediating the immunomodulatory activity of 1,25D. Furthermore, a

[2] Citation for chapter: Kariuki SN, Blischak JD, *et al.* (2016). "Patterns of transcriptional response to 1,25-dihydroxyvitamin D3 and bacterial lipopolysaccharide in primary human monocytes." G3 (Bethesda). Epub 2016/03/16. doi: 10.1534/g3.116.028712.

subset of genes with evidence of inter-ethnic differences in transcriptional response was also identified, suggesting that in addition to the well-established inter-ethnic variation in circulating levels of vitamin D, the intensity of transcriptional response to 1,25D and LPS also varies between ethnic groups. We propose that dysregulation of the pathways identified in this study could contribute to immune-mediated disease risk.

3.2: Introduction

Vitamin D plays an important immunomodulatory role through a transcriptional mechanism [32, 37-39]. In the immune system, the active form of vitamin D, 1,25-dihydroxyvitamin D_3 (1,25D), binds the vitamin D receptor (VDR), which translocates into the nucleus where it modulates the transcription of genes with immune function such as cathelicidin antimicrobial peptide (*CAMP*) , defensin genes such as β-defensin 4A (*DEFB4A*) , and autophagy genes such as autophagy related 5 (*ATG5*) [32, 37-39, 59, 138, 139]. In monocytes/macrophages, 1,25D can be produced intracellularly from the inactive form, 25-hydroxyvitamin D_3 (25D), which is found abundantly in circulation. The circulating levels of 25D vary greatly across individuals and ethnic groups [42, 43, 45]. Attesting to the important role of vitamin D in immune response, low levels of 25D have been linked to increased susceptibility to tuberculosis (Tb) [60, 61]. Moreover, 25D supplementation in individuals with hypovitaminosis D resulted in an enhanced antimicrobial response [32, 59, 91]. Although many studies have been conducted on the inter-individual and inter-ethnic variation in the circulating inactive 25D levels, with corresponding epidemiological

links to immune-related diseases [47, 48, 81, 83, 84], little is known about inter-individual and inter-ethnic variation in the transcriptional response to active 1,25D.

Previous studies of 1,25D activity in immune cells highlight its complex immunomodulatory role, regulating activities such as enhancement of the response to *Mycobacterium tuberculosis* (*M. tb*) in THP-1 macrophage cell lines [67], down-regulation of immune-related pathways such as interferon signaling in peripheral blood mononuclear cells (PBMCs) [72], and induction of a tolerogenic phenotype as well as an attenuation of the pro-inflammatory response in dendritic cells [41, 65, 66]. Though the immunoregulatory role of 1,25D in different innate immune cell types is complex, it generally results in the attenuation of an intense pro-inflammatory response, which can have toxic consequences such as sepsis and septic shock [78-80].

In this study, we focused on characterizing the transcriptional response to 1,25D in primary monocytes in the presence or absence of a pro-inflammatory stimulus, bacterial lipopolysaccharide (LPS). Stimulating monocytes with LPS enabled examination of how an inflammatory stimulus modifies the transcriptional response to 1,25D in monocytes. This analysis highlighted several biological pathways that are modulated by 1,25D in the absence of LPS (e.g. oxidative phosphorylation and mitochondrial dysfunction) as well as others that are modulated by LPS and reversed by 1,25D (e.g. pro-inflammatory cytokine signaling pathways). In addition, we identified inter-ethnic differential expression patterns, suggesting that the well-established inter-ethnic variations in the vitamin D pathway extend to the intensity of transcriptional response to LPS and 1,25D.

3.3: METHODS

Ethics Statement

All donors to Research Blood Components (http://researchbloodcomponents.com/) and Sanguine Biosciences (https://www.sanguinebio.com/) sign an IRB-approved consent form giving permission to collect blood, and use it for research purposes. This study did not require IRB review at the University of Chicago because blood samples were not shipped with individually identifiable information.

Subjects

All subjects were healthy donors collected by Research Blood Components and Sanguine Biosciences. Self-reported ethnicity, age, gender, date, and time of blood drawing were recorded for each donor. Buffy coats from 10 African-American (AA) and 10 European-American (EA) subjects were shipped within 24 hours of collection. We processed samples in multiple batches, balanced by ethnic group. Serum samples from the donors were sent to the Clinical Chemistry Laboratory of the University of Chicago to determine 25-hydroxyvitamin D_3 (25D) levels and parathyroid hormone (PTH) levels. Total serum 25D and PTH levels were determined using electrochemiluminescence detection assays (cat. no. 06506780160 and cat. no. 11972103160 respectively, Roche Diagnostics Corporation, Indianapolis, IN, USA).

Monocyte culture and treatment

We isolated peripheral blood mononuclear cells (PBMCs) from the buffy coats of the 20 subjects by density gradient centrifugation using Ficoll-Paque PLUS medium (GE Healthcare Life Sciences, Pittsburgh, PA). We isolated monocytes from the PBMCs by positive selection using magnetic CD14 MicroBeads according to the supplier's protocol (Miltenyi Biotec, San Diego, CA). We cultured isolated monocytes ($1x10^6$ cells/mL) in RPMI 1640 medium (Gibco, Life Technologies, Grand Island, NY), 25mg/mL Gentamicin (Gibco) and 10% charcoal-stripped fetal bovine serum (Gibco) in 24-well plates. Monocytes were cultured in three replicates for 24 hours for each of the following treatments: 1) Vehicle solution containing 1% Ethanol and 99% culture medium, as a negative control, 2) 100nM of 1,25D, 3) 10ng/mL of LPS in the vehicle solution, and 4) 100nM of 1,25D and 10ng/mL of LPS (experimental design summarized in **Supplementary Figure 3.1**). These four treatments are abbreviated E, V, L, and V+L, respectively.

Transcriptome analysis

We pooled the three replicates for each treatment and extracted total RNA from the pool using Qiagen RNeasy Plus mini kit (Valencia, CA). We extracted RNA from 80 samples consisting of the 20 subjects that each received 4 treatments, in 10 batches each balanced by ethnic group. RNA concentration and RNA integrity score (RIN) were recorded for each sample on the 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA) (average RNA concentration and RIN scores in each ethnic group summarized in **Supplementary Table 3.1**). Total RNA was reverse transcribed into cDNA, labeled, hybridized to Illumina (San Diego, CA, USA) Human HT-12 v3 Expression Beadchips and scanned at the University of Chicago Functional Genomics Core facility. The microarrays

were hybridized in three batches, and we recorded the array batch number for each sample to be used as a covariate in subsequent analyses.

We performed low-level microarray analyses using the Bioconductor software package lumi [140] in R, as previously described [141]. Briefly, we annotated probes by mapping their sequence to RefSeq (GRCh37) transcripts using BLAT. We discarded probes that mapped to multiple genes to avoid ambiguity in the source of a signal due to cross-hybridization of similar RNA molecules. We also discarded probes containing one or more HapMap SNPs to avoid spurious associations between expression measurements and ethnicity, due to allele frequency differences between ethnic groups. We applied variance stabilization to all arrays, discarded poor quality probes, and quantile normalized the arrays using the default method implemented in the lumiN function. After these filters, probes mapping to 10,958 genes were used in downstream analyses (data available in **Supplementary File 3.1**).

Differential expression analysis

We tested each gene for differential expression (DE) using a linear mixed-effects model with the R package, lme4 [142]. The model included fixed effects for ancestry and the three treatment conditions (V, L, V+L), as well as interaction effects between ancestry and the treatments. It also included a random effect to model the differences between the individuals. Lastly, the model included covariates for the technical factors with the strongest effects on the expression data ($p < 0.05$), as determined by their association with the principal components described below, including array batch, age, baseline 25D levels, baseline PTH levels, RNA concentration and RIN scores. P-values were obtained using the R

package, lmerTest, which provides a summary function with p-values added for the t-test based on the Satterthwaite approximation for denominator degrees of freedom [143]. To correct for multiple testing, we estimated the false discovery rate (FDR) using the "qvalue" function in R, based on the Storey method [144]. The FDR for DE was set at 1%. To identify genes that were DE between the two ancestries, we tested the significance of the fixed interaction effects between ancestry and the treatments. Here we used a more relaxed FDR threshold of 10% to determine significance, due to the smaller sample size in the inter-ethnic comparison (10 AA's and 10 EA's).

We also performed a joint Bayesian analysis using the R package Cormotif [145], which jointly models expression data across different experiments enabling classification of genes into patterns of shared and distinct differential expression. Genes are assigned to correlation motifs, which are the main patterns of differential expression obtained from the shared information across experiments, which in our study are treatments and ethnic groups. We regressed out the technical covariates described above from the expression data using the limma package removeBatchEffect [146], and used the residuals as input. We used a modified version of Cormotif as described in [147] where the original code was modified to return the cluster likelihood for each gene to enable downstream analyses. Also, since Cormotif is non-deterministic, we ran each test 100 times and kept the result with the largest maximum likelihood estimate.

Gene set enrichment analysis

We performed gene set enrichment analyses using the commercially available software Ingenuity Pathway Analysis (IPA). We compared DE genes with curated

functional attribution lists organized by canonical pathway function. The magnitude of over-representation of a particular canonical pathway in the gene list from our study was calculated as the ratio of the number of genes from our data set that map to the pathway divided by the total number of reference genes in that pathway in the IPA database. Statistical significance of the observed enrichment of a particular pathway was determined using Benjamini-Hochberg multiple testing corrected p-values provided by IPA [148].

Identifying vitamin D receptor binding sites near DE genes

We reanalyzed published data sets of VDR ChIP-seq, which used THP-1 monocytic cell lines treated with 1,25D and LPS or 1,25D alone [149], and FAIRE-seq, which used THP-1 cells treated with 1,25D [150]. First, we aligned sequence reads to the human reference (GRCh37) using BWA backtrack 0.7.5 [151]. Second, we kept only sequence reads with phred-scaled mapping quality ≥ 30 using samtools v1.1 [152]. Third, PCR duplicates were removed with Picard v 1.130 (http://broadinstitute.github.io/picard/). For the ChIP-seq data sets, we confirmed the quality of data sets by strand cross-correlation (SCC) analysis [153] implemented in the R script "run_spp_nodups.R" packaged in phantompeakqualtools (https://code.google.com/p/phantompeakqualtools/). Statistically significant peaks were identified using MACS version 2 [154] with the following essential command line arguments: macs2 callpeak --bw X -g hs --qvalue=0.05 -m 5 50, where X is a length of the bandwidth that was defined as a fragment length calculated by SCC for the ChIP-seq data or as 200 bp for the FAIRE-seq data reported in Seuter *et al*. (2013).

To identify VDR response elements, we considered peaks that overlapped completely or partially between the ChIP-seq data after 1,25D and LPS treatment and the

FAIRE-seq data. We then annotated them using HOMER [155] to find the closest gene to each peak and, among these genes, we selected those that were DE genes in response to 1,25D (V) and the combined 1,25D + LPS (V+L) treatments from the linear mixed-effects analysis. Enrichment of VDR response elements was determined using Fisher's exact test, comparing peaks in DE genes to those in non-DE genes.

We also examined the enrichment of VDR binding sites among genes clustered in each expression pattern from the joint Bayesian analysis using Cormotif. To inclusively identify VDR binding sites, we merged the ChIP-seq data from THP-1 cells treated with 1,25D and LPS treatment with data from THP-1 cells treated with 1,25D alone. We examined overlap between the genes that were closest to the peaks, and the genes in each expression pattern. Enrichment of VDR peaks was then determined using Fisher's exact test, comparing VDR peaks in each expression pattern with peaks in the "Non-DE" pattern.

Data availability

The raw microarray data files have been deposited in NCBI's Gene Expression Omnibus (GEO) [156] and are accessible through GEO Series accession number GSE78083 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE78083). The normalized expression values, the results from the linear mixed-effects model, and the results from Cormotif are provided in **Supplementary File 3.1.**

3.4: Results

Sources of transcriptome-wide variation

To evaluate the variation in transcriptional response to 1,25D in the innate immune system in the presence and absence of an inflammatory stimulus, we cultured primary monocytes obtained from 10 African-American and 10 European-American healthy donors in four different conditions in parallel: i. EtOH (i.e. the vehicle control, or E), ii. LPS (L), iii. 1,25 D (V), and iv. 1,25D plus LPS (V + L); the experimental design is illustrated in **Supplementary Figure 3.1**. Transcript levels were measured with gene expression arrays for each treatment condition and each individual, resulting in a total of 80 transcriptome data sets. Relevant covariates, including serum levels of 25D, were measured or recorded and used in downstream analyses (see Methods). Although there was significant inter-individual and inter-ethnic variation in serum 25D levels in our sample of donors (**Supplementary Table 3.1**), this variable was not correlated with the transcriptional response to LPS, 1,25D or their combination (**Supplementary Figure 3.5**). This suggests that our *in vitro* system is not affected by 25D levels *in vivo*.

To evaluate the sources of variation in the overall transcriptome data, we performed a principal components analysis (PCA) of the variance-stabilized $\log_2$-transformed expression data using the prcomp function in R. Principal component 1 (PC1) separates the samples by LPS treatment, accounting for 22% of the total variation in gene expression and reflecting the large effect of LPS on the transcriptome (**Supplementary Figure 3.2 (A) and (C), Supplementary Table 3.2**), while PC2 separates the samples by 1,25D treatment, and accounts for 8.6% of the total variation in gene expression (**Supplementary Figure 3.2 (A) and (D), Supplementary Table 3.2**). PC3 and PC4, which account for 6.7% and 5.8% of variation respectively, separate the samples by the three array processing batches (**Supplementary Figure 3.2 (B), (E) and (F), Supplementary Table 3.2**). We also tested

for associations between the PCs and the different covariates recorded for each sample (average sample covariates are compared between ancestries in **Supplementary Table 3.1**). PC1 was associated with RNA concentration ($p = 1.54 \times 10^{-5}$, $r^2 = 0.219$), PC2 was weakly associated with the RNA integrity number (RIN) scores ($p = 9.0 \times 10^{-3}$, $r^2 = 0.086$), while PC3 was associated with age ($p = 1.36 \times 10^{-6}$, $r^2 = 0.266$), baseline 25D levels ($p = 1.0 \times 10^{-3}$, $r^2 = 0.136$), baseline PTH levels ($p = 6.1 \times 10^{-6}$, $r^2 = 0.237$), and RIN score ($p = 5.0 \times 10^{-3}$, $r^2 = 0.099$) (**Supplementary Table 3.2**). The effects of array processing batch, RNA concentration, RIN scores, serum 25D and PTH levels were subsequently included as covariates in the linear mixed-effects model for differential expression.

After regressing out the covariates using the limma package removeBatchEffect, and performing PCA on the residuals of the covariates-corrected expression data, we observed that PC1 and PC2 separated the samples by treatment (**Supplementary Figure 3.3 (A), (C) and (D), Supplementary Table 3.3**), but PC1 was still associated with RNA concentration ($p = 3.56 \times 10^{-5}$, $r^2 = 0.203$) while PC2 was still associated with RIN score ($p = 3 \times 10^{-3}$, $r^2 = 0.113$) (**Supplementary Table 3.3**). PC3, which accounted for 5% of the total variation in gene expression, was associated with sample ($p = 4.5 \times 10^{-7}$, $r^2 = 0.286$), and ancestry ($p = 1.04 \times 10^{-5}$, $r^2 = 0.227$), highlighting the effect of inter-individual and inter-ethnic variation on gene expression (**Supplementary Figure 3.4, Supplementary Table 3.3**).

Opposite effects of 1,25D and LPS on the transcriptome

Using the main effects for each treatment from the linear mixed-effects model, we identified genes that were DE in response to the different treatment conditions at a FDR of 1%. 2,888 genes were DE in response to 1,25D alone relative to vehicle (V vs. E). Gene set

enrichment analysis identified metabolic processes, such as oxidative phosphorylation and the tricarboxylic acid (TCA) cycle, enriched among up-regulated genes (**Supplementary Table 3.4**). Pathways that play important roles in regulating translation processes, such as EIF2 signaling and mTOR signaling, were also significantly enriched among up-regulated genes, indicating an important role of 1,25D in regulating translation. Immune responses involving chemokine signaling, B and T cell signaling, as well as various pro-inflammatory signaling cascades such as Tec kinase signaling, Phospholipase C signaling, and Integrin signaling were enriched amongst the down-regulated genes, consistent with the immunomodulatory function of 1,25D.

There was a strong transcriptomic response to LPS treatment relative to vehicle (L vs. E), with 4,461 genes DE at a FDR of 1%. Pathways enriched among LPS responsive genes highlight the opposite direction of transcriptional response to 1,25D and LPS, where pro-inflammatory immune response pathways were enriched among up-regulated genes, while oxidative phosphorylation and translational control pathways were enriched among down-regulated genes (**Supplementary Table 3.4)**, indicating the importance of these pathways in the pro-inflammatory effects induced by LPS stimulation.

Effects of combined 1,25D + LPS treatment on the transcriptome

The combined treatment of 1,25D + LPS resulted in 4,720 genes significantly DE relative to vehicle (V + L vs. E). We also examined the transcriptional response of the combined 1,25D+LPS treatment relative to LPS (V + L vs. L) in an attempt to isolate the effect of 1,25D on the transcriptome in the presence of LPS, and identified 2,404 genes significantly DE in this treatment category.

The pattern of response to V + L vs. E followed a similar pattern to LPS treatment alone (L vs. E), with similar pathways enriched among genes in these treatment categories (**Supplementary Table 3.4**), probably because of the overwhelming transcriptional response to LPS. Genes significantly DE in response to V + L vs. L, which effectively subtracts the transcriptional effects of LPS, were similar to the genes significantly DE in response to 1,25D treatment alone (V vs. E), with similar pathways enriched.

We detected additional pathways enriched among genes significantly DE in response to the combined V + L treatments, both relative to vehicle and relative to LPS. These included adipogenesis and insulin receptor signaling pathways, both involved in lipid metabolic processes, which were enriched among up-regulated genes. IL-4 signaling, which is associated with allergy and asthma through development of T cell mediated immune responses [157, 158], was significantly enriched among down-regulated genes (**Supplementary Table 3.4**). Pathways enriched among genes responsive to the combined V + L treatment indicate a regulatory role of 1,25D in these pathways specifically in the context of LPS stimulation.

Bayesian analysis of shared transcriptional response across treatments and ethnic groups

To further dissect the effects of 1,25D and LPS on the transcriptome, we sought to identify the shared and distinct patterns of transcriptional response across treatments and across ethnic groups. A popular approach to this question is to investigate the overlap of DE genes between conditions at a given FDR threshold. However, this approach fails to account for incomplete power to detect DE genes, thus exaggerating the differences in the

transcriptional response between the conditions. In order to identify shared patterns of transcriptional response across treatments and ancestry while accounting for incomplete power, we implemented a joint Bayesian analysis with the R/Bioconductor package Cormotif [145]. Genes were classified into different response patterns, or correlation motifs, across treatments and ancestry (**Figure 3.1**). Since Cormotif does not distinguish the direction of effect across treatments, we used the results of the linear mixed-effects model in conjunction with the Cormotif approach to establish direction of response in the different response patterns (**Figure 3.2**).

A total of 5,737 genes were classified in the "No response" pattern, which includes genes whose expression levels were unchanged across all the treatments (**Figures 3.1 and 3.2A**). This is broadly consistent with the results of the linear mixed-effects model, with 80% of these genes being also classified as non-DE for any treatment in the linear mixed-effects model.

Genes that responded to all the treatments were classified in the "All" pattern and included 265 genes whose expression levels changed across all treatments and ancestries (**Figures 3.1 and 3.2B**). Genes classified in this Cormotif had response patterns to 1,25D and LPS that were both concordant (i.e. up- or down-regulated in both treatments) and discordant (i.e. up-regulated in one treatment and down-regulated in the other). Genes that were up-regulated in all treatments (top-right quadrant, **Figure 3.2B**) included *CD14*, which encodes a surface antigen expressed on monocytes that is involved in mediating response to bacterial LPS. Genes that were down-regulated in all treatments included chemokine signaling genes such as *CCL13* (bottom-left quadrant, **Figure 3.2B**). The discordant response patterns included genes that were up-regulated by 1,25D and down-

regulated by LPS (top-left quadrant, **Figure 3.2B**), with EIF2 signaling and mTOR signaling pathways significantly enriched among these genes (**Figure 3.3**). This is consistent with the opposite transcriptional effects of 1,25D and LPS on genes in these pathways that were highlighted in the linear mixed-effects analysis. Genes that were down-regulated by 1,25D and up-regulated by LPS (bottom-right quadrant, **Figure 3.2B**) included some cytokine receptor genes such as *IL7R* and *IL2RA* which are important components of the pro-inflammatory signaling cascade.

The "All except V+L" pattern included 1,364 genes whose expression levels changed in all treatments except the combined 1,25D+LPS relative to vehicle (V + L vs. E). All the genes in this Cormotif pattern were discordant in their response to 1,25D and LPS resulting in a neutral effect in the response to the combined V + L vs. E (**Figures 3.1 and 3.2C**). Genes that were responsive to the combined V + L vs. L followed a similar direction of response to the genes DE in response to the individual V vs. E treatment, suggesting that the response to 1,25D at these genes is not dramatically influenced by LPS. The genes that were up-regulated by 1,25D and down-regulated by LPS (top-left quadrant, **Figure 3.2C**) were enriched for EIF2 and mTOR signaling pathways, similar to the discordant genes in the "All" category (**Figure 3.3**). In addition, oxidative phosphorylation and mitochondrial dysfunction pathways were significantly enriched amongst these genes. On the other hand, genes that were down-regulated by 1,25D and up-regulated by LPS (bottom-right quadrant, **Figure 3.2C**) were enriched for various pro-inflammatory response pathways, including Granulocyte Adhesion and Diapedesis, IL-8 signaling, NF-kB signaling, TNFR2 signaling and Role of NFAT in regulation of the immune response (**Figure 3.3**).

Genes responsive to 1,25D were divided into three Cormotif patterns: “1,25D”, “1,25D+LPS”, and “1,25D all”. The “1,25D” pattern included 350 genes that were DE in response to V vs. E alone (**Figures 3.1 and 3.2D**). Oxidative phosphorylation and mitochondrial dysfunction pathways were significantly enriched among up-regulated genes (**Figure 3.3**). Interestingly, the oxidative phosphorylation pathway genes enriched in this Cormotif pattern responded similarly to the genes in the same pathway classified in the “All except V+L” Cormotif pattern, in that they are significantly induced by 1,25D. However, the oxidative phosphorylation pathway genes in the “1,25D” pattern respond exclusively to 1,25D, while those in the “All except V+L” Cormotif pattern are up-regulated by 1,25D and down-regulated by LPS (**Figure 3.3, Supplementary Figure 3.6**). This indicates a context-specific response profile among genes in the same pathway, where some genes in the oxidative phosphorylation pathway are uniquely regulated by 1,25D, whereas other genes in the same pathway are regulated by both 1,25D and LPS.

The “1,25D+LPS” pattern included 270 genes that responded to 1,25D only in the presence of LPS. This pattern captured genes that were DE in response to the combined V + L vs. E, and V + L vs. L (**Figures 3.1 and 3.2E**). Although there were no enriched pathways among genes in this Cormotif pattern at an FDR of 5%, some interesting pathways, such as eNOS signaling and cholesterol biosynthesis pathway, were represented among the down-regulated genes at a FDR of 27%, suggesting a role for 1,25D in modulating these pathways upon LPS stimulation.

The “1,25D all” pattern included 782 genes that were DE in response to 1,25D in the presence and in the absence of LPS (**Figures 3.1 and 3.2F**). Genes in the antimicrobial pathway were included in this category, such as the anti-bacterial peptide gene *CAMP*,

autophagy genes *ATG3*, *ATG5*, *ATG2A* and *ATG9A* and the intracellular pattern recognition receptor gene *NOD2*. These genes were significantly up-regulated in response to 1,25D alone or in combination with LPS. The Role of JAK family kinases in IL-6-type Cytokine Signaling was the most significantly enriched pathway amongst the up-regulated genes (**Figure 3.3**), and included genes such as *MAPK14*, *PTPN11* and *STAT5B*, all of which could be crucial for triggering antimicrobial responses in monocytes. Biological pathways enriched among down-regulated genes in this category included B cell receptor signaling, Tec kinase signaling and Leukocyte extravasation signaling (**Figure 3.3** and **Supplementary Figure 3.7**), highlighting the role of 1,25D in repressing pro-inflammatory response pathways. Interestingly, immunological and inflammatory diseases were among the most enriched disease categories from the IPA analysis among the down-regulated genes (**Supplementary Table 3.6** and **Supplementary Figure 3.7**), suggesting a protective role of 1,25D in immunological diseases. Overall, the "1,25D all" response pattern illustrates the important dual immunomodulatory role played by 1,25D in monocytes, where antimicrobial pathway genes are up-regulated, while pro-inflammatory pathway genes associated with immunological and inflammatory disease are down-regulated by 1,25D in the presence or absence of LPS stimulation.

The "LPS" pattern included 1,400 genes whose expression levels changed in response to L vs. E and the combined V + L vs. E (**Figures 3.1 and 3.2G)**. Consistent with the results from the linear mixed-effects model, pro-inflammatory pathways were significantly enriched among the up-regulated genes in this category, including IL-8 signaling, NF-kB signaling, IL-17 signaling, and TNFR2 signaling among others (**Figure 3.3**). Among the down-regulated genes, tRNA charging, mitochondrial dysfunction, the TCA

Cycle II, galactose metabolism pathway and folate transformation pathway were significantly enriched (**Figure 3.3** and **Supplementary Table 3.5)**, indicating that LPS modulates transcription of genes in these metabolic pathways.

Figure 3.1: **Transcriptional response patterns shared across the different treatments and ancestries identified by implementing a joint Bayesian analysis using Cormotif.** The shading of each box represents the posterior probability that a gene assigned to a given expression pattern (rows) is differentially expressed in individuals from a particular ancestry in response to each treatment (columns). **V**=Response to 1,25D, relative to vehicle; **L** = Response to LPS relative to vehicle; **V+L** = Response to 1,25D+LPS relative to vehicle; **V+L.vs.L** = Response to 1,25D+LPS relative to LPS; **EA** = European-American; **AA** = African-American.

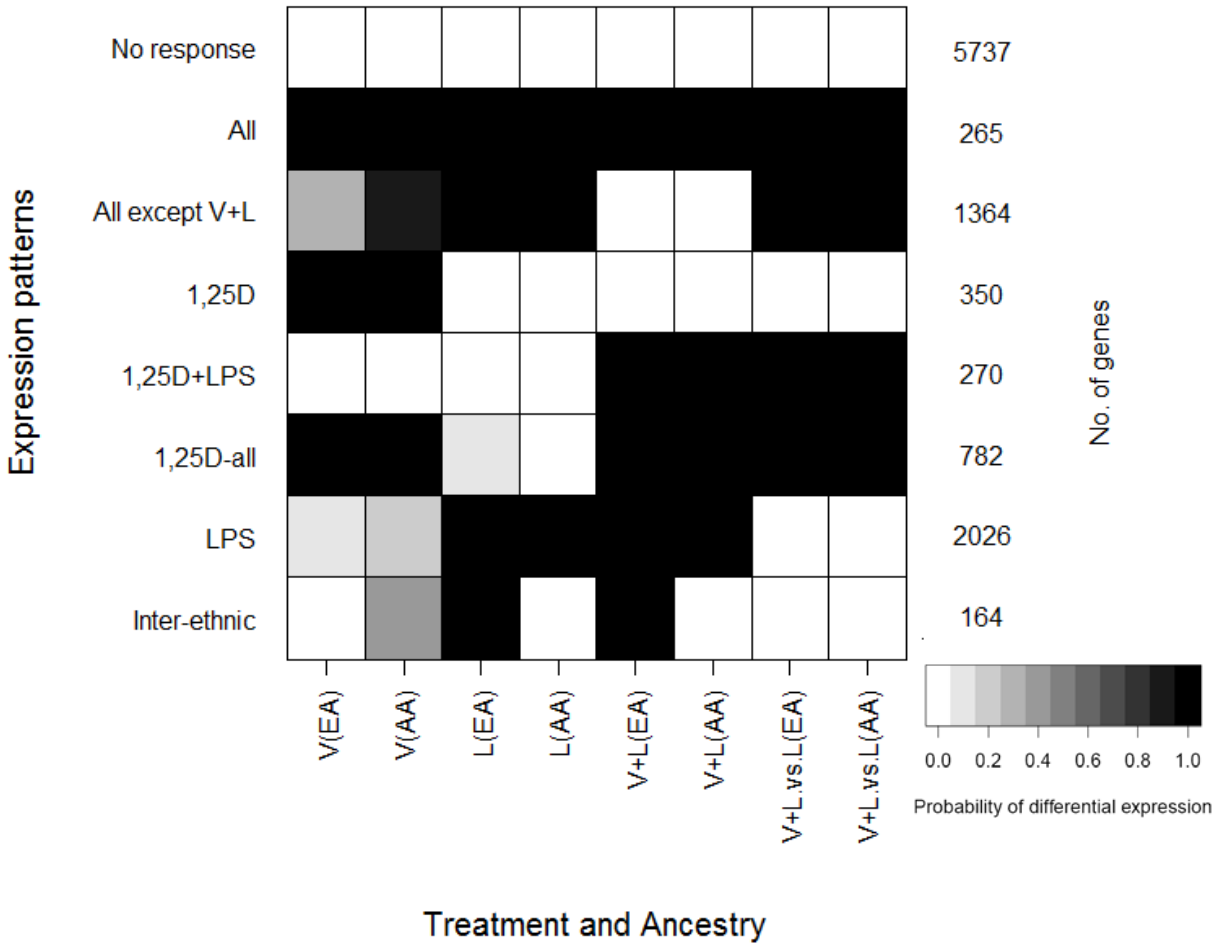


Figure 3.2: **Direction of response in the different correlation motifs.** Patterns of differential response to single treatment with 1,25D (vertical axis) or LPS (horizontal axis) for each correlation motif are shown in **A**-**D** and **F**-**H**. **E** shows patterns of differential response to the combined treatment with 1,25D and LPS relative to LPS (vertical axis) and 1,25D and LPS relative to vehicle (horizontal axis). Genes are color coded based on q-values < 0.01 from linear mixed-effects analysis as follows: **Red** = DE in response to both 1,25D and LPS; **Blue** = DE in response to 1,25D; **Green** = DE in response to LPS; **Grey** = not DE.

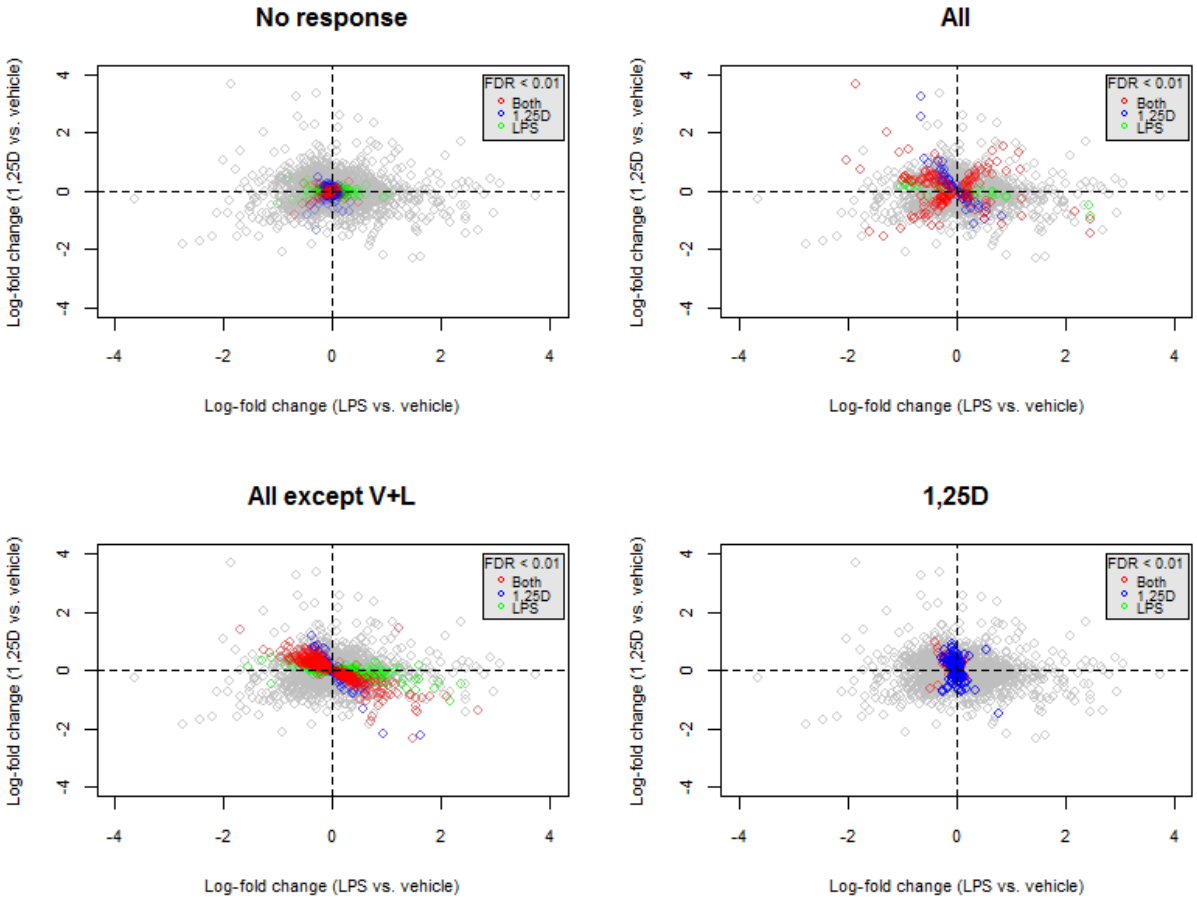


Figure 3.2 – continued.

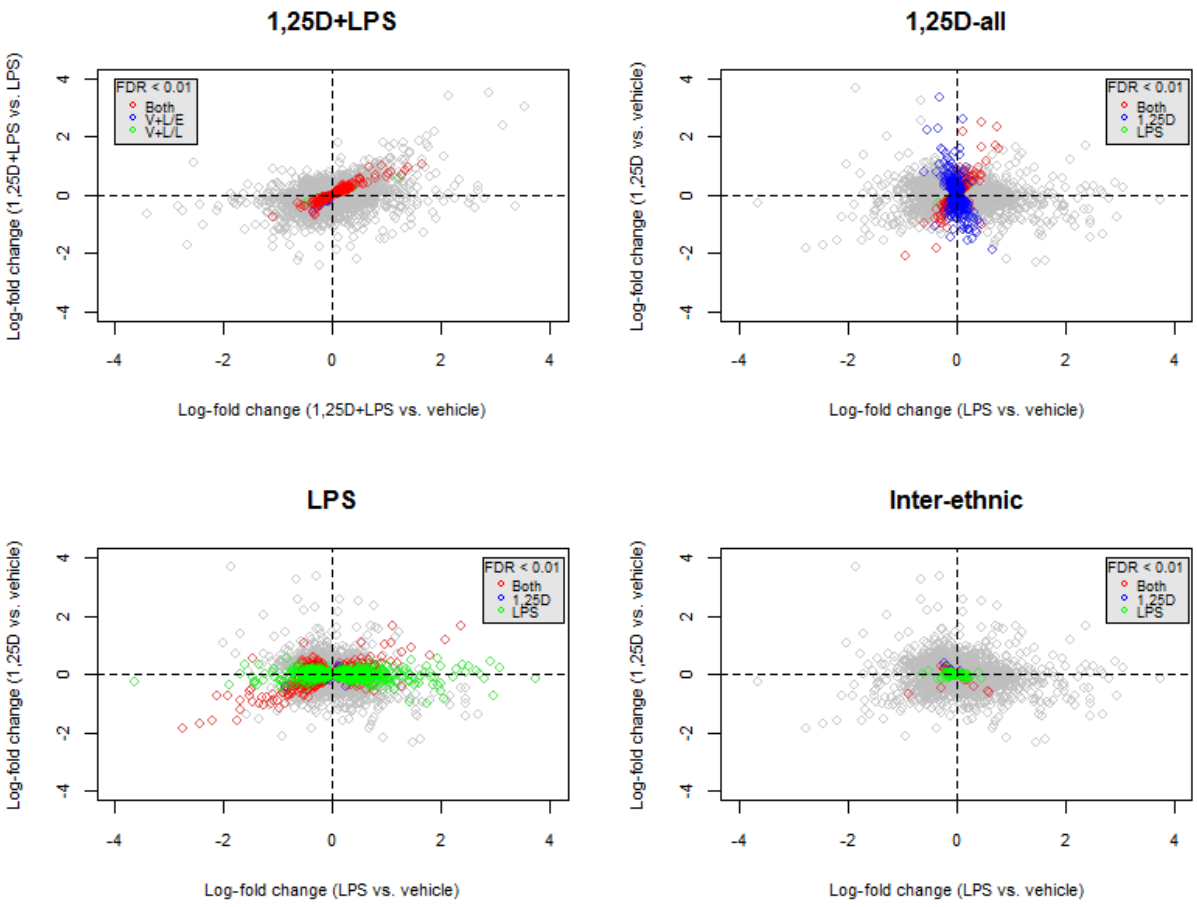
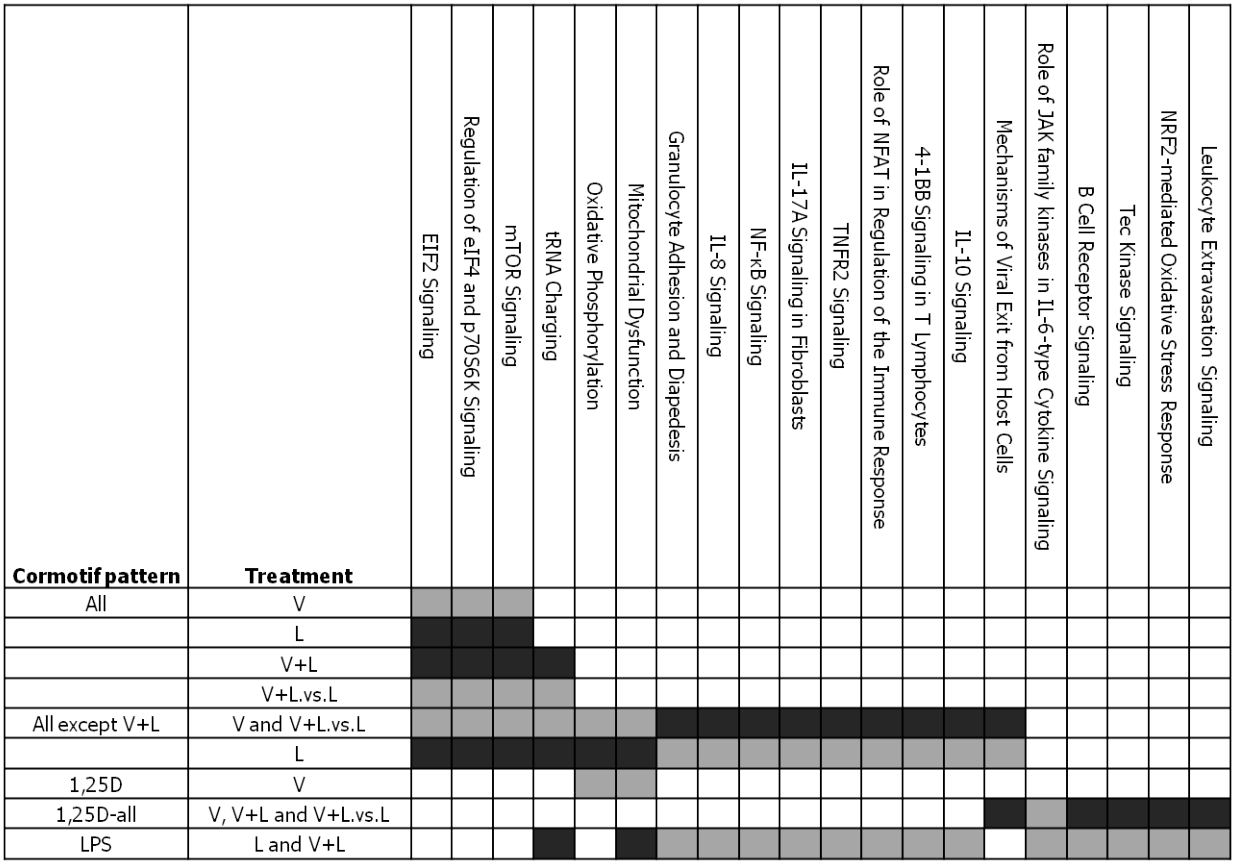


Figure 3.3: Sharing of enriched biological pathways across Cormotifs. The table shows the biological pathways that were enriched (FDR < 0.05) in more than one Cormotif subdivided based on the direction of transcriptional response (up-regulated genes in **light grey** and down-regulated genes in **dark grey**) and the treatment (**V**=Response to 1,25D, relative to vehicle; **L** = Response to LPS relative to vehicle; **V+L** = Response to 1,25D+LPS relative to vehicle; **V+L.vs.L** = Response to 1,25D+LPS relative to LPS).



Genes with inter-ethnic differential response

The "Inter-ethnic" pattern was of particular interest, as it identified 164 genes with evidence of differential responses to LPS treatments between AA's and EA's, with a stronger response in EA's compared to AA's (**Figures 3.1 and 3.2H**).

We also interrogated the degree of inter-ethnic differences in transcriptional response using the main interaction term for treatment and ancestry in the linear mixed-effects model. We identified 15 genes with strong inter-ethnic differences in response to V + L vs. E at a FDR < 10%. These genes include *PPAP2B* which encodes a member of the phosphatidic acid phosphatase (PAP) family and has been implicated in coronary artery disease risk [159, 160], *STEAP3* which encodes an endosomal ferrireductase required for efficient transferrin-dependent iron uptake, and *AKNA* which encodes a transcription factor that specifically activates the expression of the CD40 receptor and its ligand CD40L/CD154 on lymphocyte cell surfaces, which are critical for antigen-dependent-B-cell development (**Figure 3.4A**). Interestingly, 13 out of the 15 genes showed more significant differential responses in EA's (**Supplementary Figure 3.8**), similar to the pattern observed in the "Inter-ethnic" Cormotif pattern.

To account for the effect of LPS, and examine the extent to which the inter-ethnic differential response patterns were modulated by 1,25D, we examined inter-ethnic differential response to the combined V + L vs. L. *PPAP2B* and *KIAA1958* were the only statistically significant genes identified in this category ($p = 2.96 \times 10^{-6}$ and 9.62×10^{-6}, respectively), with both of these genes more significantly differentially expressed in EA's (**Figure 3.4B**).

Figure 3.4: **Inter-ethnic differential response patterns.** Genes with inter-ethnic differences in transcriptional responses to 1,25D+LPS relative to vehicle (**A**), and relative to LPS (**B**) were identified using the interaction term for treatment and ancestry in the linear mixed-effects model (FDR < 0.10). The boxplots show examples of these genes with different log-fold change in transcript levels between the two ethnic groups. **AA** = African-American; **EA** = European-American.

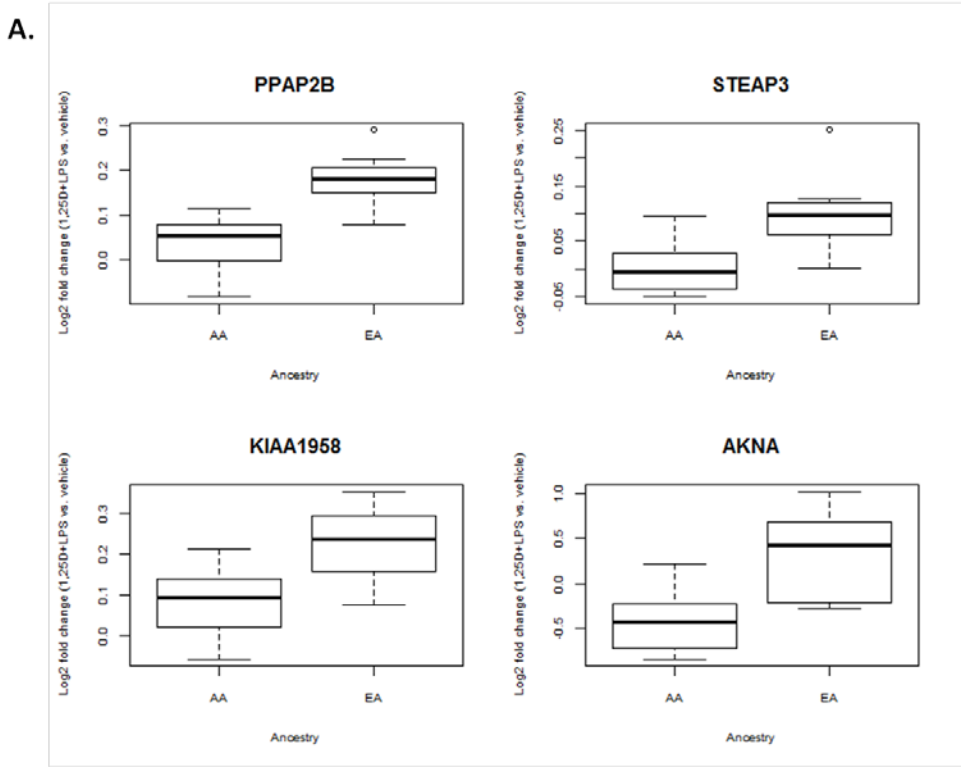
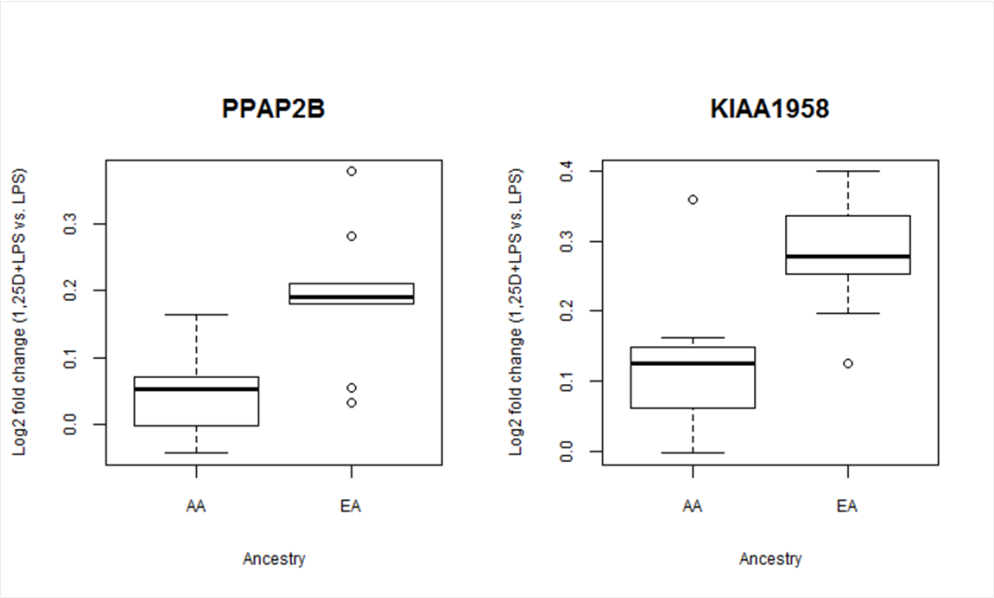


Figure 3.4 – continued.

B.



Regulatory elements near DE genes

We examined the overlap between genes DE in response to 1,25D and the combined 1,25D+LPS treatments in our primary monocytes, and published datasets for VDR ChIP-seq [149] and FAIRE-seq [150] performed in THP-1 monocytic cells, to examine whether there was enrichment of open chromatin regions and VDR binding sites near the transcription start sites of DE genes. We found a significant enrichment of VDR binding sites amongst genes DE in response to V vs. E ($p = 4.56 \times 10^{-11}$), V + L vs. E ($p = 3.97 \times 10^{-8}$) and V + L vs. L ($p = 1.54 \times 10^{-7}$) (**Supplementary Table 3.7**). There was an overlap of 201 genes between the DE genes, VDR ChIP-seq and FAIRE-seq datasets, highlighting genes such as *CAMP* and *CD14,* which contain open chromatin regions and VDR binding sites near the transcription start site; these 201 genes are potentially direct VDR targets.

In addition, we examined the enrichment of VDR binding sites across the different Cormotif patterns (**Table 3.1**). The genes in the "1,25D-all" and "All" Cormotif patterns had the highest enrichment of VDR binding sites ($p = 6.88 \times 10^{-13}$ and 2.57×10^{-8} respectively), indicating a higher proportion of potentially direct VDR targets represented in these Cormotif patterns. Genes in the "1,25D" Cormotif pattern were not significantly enriched for VDR binding sites, suggesting that the presence of LPS, in addition to 1,25D, is important to enable the 1,25D-VDR transcriptional activity in primary monocytes.

Table 3.1: Proportion of genes in each Cormotif pattern containing vitamin D receptor (VDR) binding sites. Enrichment of VDR peaks in each category was calculated using Fisher's exact test, comparing genes in each Cormotif pattern to those in the "Non-DE" Cormotif pattern.

Cormotif pattern	Total No. Genes	No. Genes with VDR binding site	Proportion of genes with VDR binding site	Enrichment p-value
Non-DE	5737	186	0.03	-
All	265	31	0.12	2.57×10^{-8}
All except V+L	1364	65	0.05	0.01
1,25D	350	17	0.05	0.13
1,25D+LPS	270	23	0.09	1.49×10^{-4}
1,25D-all	782	76	0.10	6.88×10^{-13}
LPS	2026	96	0.05	3.79×10^{-3}
Inter-ethnic	164	8	0.05	0.27

3.5: Discussion

We used a transcriptomic approach to characterize the immunomodulatory role of 1,25D in the presence of a pro-inflammatory stimulus to identify the mechanisms through which 1,25D exerts its immunomodulatory role. We analyzed differential expression patterns using both a linear mixed-effects analysis, which modeled individual treatment comparisons, and a Bayesian analysis using the Cormotif method, which jointly modeled differential expression across all treatments and ethnic groups, thereby accounting for incomplete power. A similar joint Bayesian framework has been successfully applied to expression quantitative trait loci (eQTL) mapping to distinguish between shared and context-specific eQTLs [161, 162]. Our joint Bayesian analysis enabled clustering of DE genes into distinct transcriptional response patterns, with pathways enriched within these transcriptional patterns highlighting mechanisms that mediate the immunomodulatory role of 1,25D.

Metabolic pathways involving oxidative phosphorylation were enriched among up-regulated genes in the "All except V+L" and "1,25D" Cormotif patterns (**Figure 3.3**). We highlight context-specific response pattern of genes within this pathway, where some genes were uniquely induced by 1,25D, while genes in other parts of the pathway were regulated by both 1,25D and LPS. The crucial role played by 1,25D in regulating oxidative phosphorylation was previously reported in PBMCs and dendritic cells [41, 66, 72], and this regulation of metabolic reprogramming by 1,25D is thought to be crucial for controlling function, growth, proliferation, and survival of various immune cell subsets [163, 164]. The fact that LPS down-regulated genes in the oxidative phosphorylation

pathway confirms previous reports indicating that LPS induces a metabolic shift away from oxidative phosphorylation to anaerobic glycolysis in macrophages and dendritic cells to enable ATP production [164]. This effect is similar to the Warbug effect in tumor cells whose high energy demand is met by switching the metabolic profile away from the tricarboxylic acid cycle and the oxidative phosphorylation pathway, towards glycolysis thereby enabling rapid ATP production [164, 165]. Previous work done in mouse macrophages and dendritic cells [166, 167] indicated that a metabolic shift towards glycolysis mediated the pro-inflammatory response, and this pro-inflammatory response could be attenuated by pharmacologic inhibition of glycolysis. From our study, this subset of oxidative phosphorylation pathway genes that were down-regulated by LPS, were then up-regulated by addition of 1,25D in combination with LPS (**Figure 3.3**). Therefore, oxidative phosphorylation could be one of the mechanisms through which 1,25D attenuates the pro-inflammatory response induced by LPS in monocytes, and the subset of genes in this pathway that we identified which were modulated by both LPS and 1,25D could be central to this mechanism.

The mTOR signaling pathway was consistently enriched among genes that were up-regulated by 1,25D and down-regulated by LPS in the “All” and “All except V+L” patterns. mTOR signaling was previously implicated in inhibition of pro-inflammatory response in LPS-stimulated monocytes/macrophages and dendritic cells, as well as in the maintenance of a tolerogenic phenotype in dendritic cells [41, 168-170]. Inhibition of mTOR resulted in increased pro-inflammatory cytokine production by LPS-stimulated monocytes/macrophages and dendritic cells [168, 170] and increased T cell proliferation [41, 170], implicating a role of mTOR in regulating the pro-inflammatory response. The

genes in this pathway were significantly down-regulated by LPS; however this direction of response was reversed by addition of 1,25D in combination with LPS (**Figure 3.3**), implying that 1,25D attenuates the pro-inflammatory response by up-regulating mTOR signaling. In addition, the genes in this pathway play important roles in regulating translation initiation, and include the ribosomal protein gene *RPS27*, and the eukaryotic translation initiation factor gene *EIF2A*, which encodes the eukaryotic initiation factor 2 (eIF-2α) that has been shown to be a downstream target of the vitamin D receptor [171]. Therefore, regulation of translation initiation through targeting the mTOR signaling pathway could be a novel mechanism for the attenuation of the pro-inflammatory response mediated by 1,25D in monocytes.

Furthermore, the EIF2 signaling pathway was also enriched among genes up-regulated by 1,25D and down-regulated by LPS in the "All" and "All except V+L" patterns, and this result is consistent with the individual treatment DE analysis using the linear mixed-effects model (**Table 3.1, Supplementary Tables 3.4 and 3.5**). EIF2 signaling plays an important role in regulating translation initiation in response to stress, and was implicated in regulating pro-inflammatory cytokine production and bacterial invasion in mouse embryonic fibroblast cells (MEFs) [172]. Shrestha *et al.* (2012) reported that the Yersinia-encoded virulence factor, YopJ, inhibited EIF2 signaling in MEFs. Similarly in our study, LPS consistently down-regulated genes in the EIF2 signaling pathway, in a mechanism that might be similar to that triggered by YopJ. In addition, Shrestha *et al.* (2012) observed that mutant MEFs with defective EIF2 signaling that were infected with different bacterial pathogens experienced enhanced cytotoxicity compared to wild type, due to increased bacterial invasion, indicating a direct role of EIF2 signaling in the

antimicrobial response. 1,25D could hence exert its antimicrobial role in monocytes by up-regulating genes in the EIF2 signaling pathway.

The dual immunomodulatory role of 1,25D was also highlighted by the genes clustered in the "1,25D-all" pattern. While 1,25D broadly down-regulated genes in the pro-inflammatory cytokine and signaling cascade pathways, it also played a crucial role in inducing important antimicrobial and autophagy genes in this Cormotif pattern. The most significantly enriched biological pathway among the up-regulated genes was the Role of JAK family kinases in IL-6-type cytokine signaling, which contained genes such as *STAT5B* which regulates signaling in diverse biological processes. Previous reports indicate that the TLR2/1-mediated induction of the vitamin D-dependent antimicrobial pathway requires IL-15 activity [173], which could be mediated via STAT5 activation which has been shown to be important for IL-15 signaling [174, 175]. 1,25D could hence regulate genes in this pathway to trigger antimicrobial responses in monocytes.

By profiling transcriptional response in monocytes from individuals of African-American and European-American ancestries, we identified some patterns of inter-ethnic variation in response to LPS, and the combined 1,25D+LPS treatment in both the linear mixed-effects analysis and the joint Bayesian analysis, while correcting for inter-individual variation in baseline levels of circulating 25D. This raises the intriguing possibility that inter-ethnic variation in the vitamin D pathway is not limited to the well-established differences in circulating levels of 25D [42, 43, 45], but it may extend to the intensity of the transcriptional response to LPS and 1,25D. Interestingly, most of the genes with inter-ethnic differential response showed more significant differential responses in EA's. The fact that most of the inter-ethnic transcriptional differences were detected in the response to

LPS or to the combined 1,25D+LPS, both relative to vehicle, suggests that these two ethnic groups differ in the pro-inflammatory transcriptional response. However, two genes had significant inter-ethnic differences in transcriptional response to the combined 1,25D+LPS relative to LPS (**Figure 3.4B**), suggesting that they differ more specifically in their response to vitamin D.

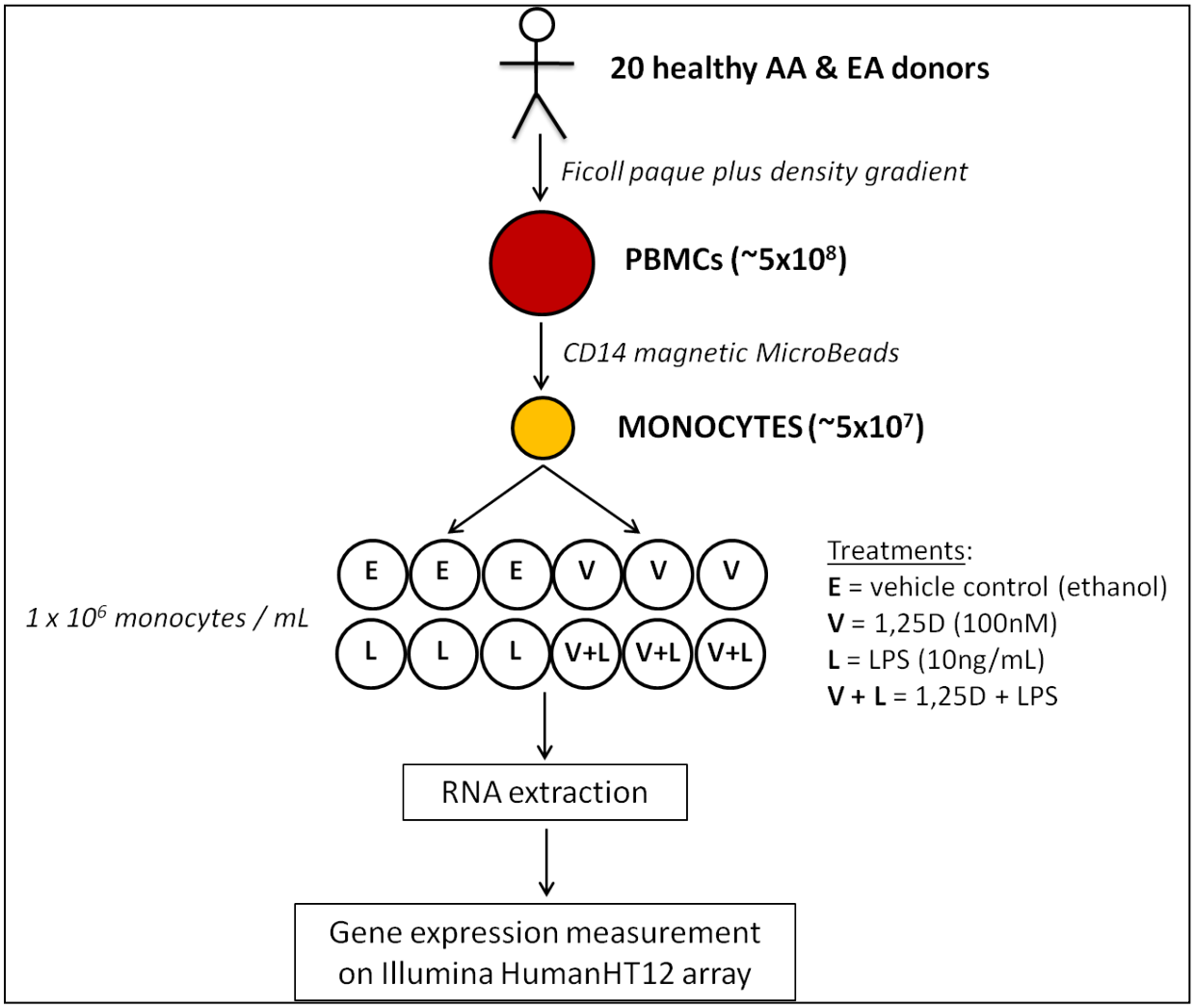
We further identified enrichment of VDR ChIP-seq and FAIRE-seq peaks among genes DE in response to the combined 1,25D+LPS treatments. This enrichment was particularly strong for genes in the "1,25D-all" and "All" Cormotif patterns, suggesting that a substantial proportion of these genes are under direct regulation of the 1,25D-VDR transcription factor complex. Intriguingly, we did not detect an enrichment of VDR binding sites near genes in the "1,25D" Cormotif. Different explanations could account for this observation. One is that the combination of both 1,25D and LPS is important for stimulating the transcriptional activity of the 1,25D-VDR transcriptional complex in human monocytes [32, 59, 149]. On the other hand, because the genes in the "1,25D" Cormotif are observed to respond only to one treatment condition, it is possible that they are enriched for false positives relative to genes in other Cormotifs that are found to respond to multiple treatment conditions. Another caveat to this analysis is that we examined the overlap of VDR ChIP-seq peaks from published data sets with experimental conditions that were different to ours. While we treated primary monocytes with 100nM 1,25D and 10 ng/mL LPS for 24 hours, the VDR ChIP-seq data was obtained from THP-1 monocytic cell lines cultured with 100 ng/mL LPS for 24 hours, and then treated with 10nM 1,25D for 80 minutes [149]. Future VDR ChIP-seq studies with uniform experimental conditions in

primary monocytes will enable better characterization of the regulatory architecture of 1,25D response genes.

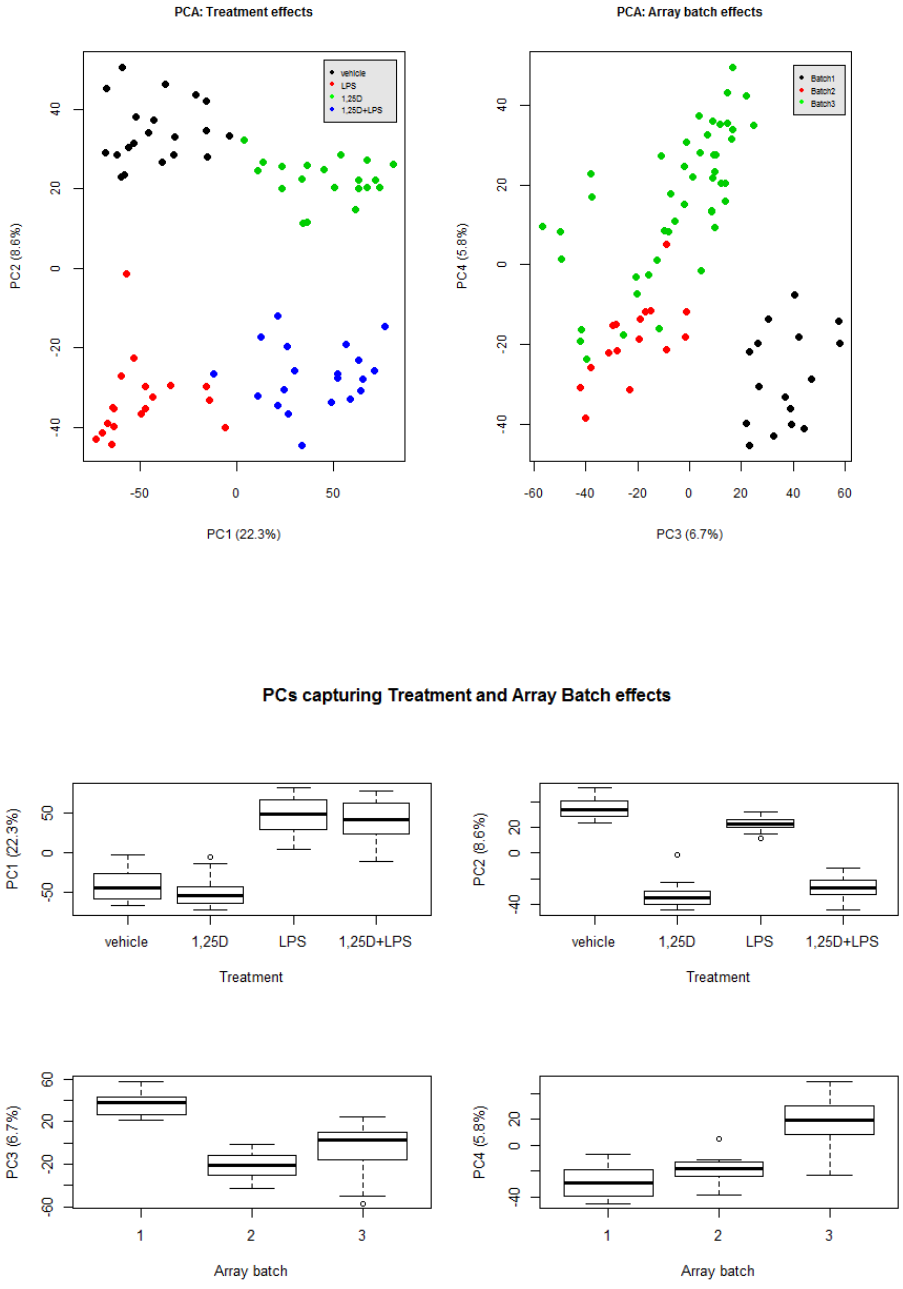
Overall, through transcriptomic profiling, our study characterizes the dual immunomodulatory role of 1,25D in primary human monocytes, highlighting the importance of biological pathways such as mTOR signaling and EIF2 signaling in mediating this immunomodulatory role. The pathways highlighted in this study may provide mechanistic clues for the observed associations between insufficient levels of circulating serum 25D and increased disease risk. The inter-individual and inter-ethnic variation in intracellular transcriptional response to 1,25D has not been previously characterized, and could serve as an additional contribution to disease risk.

3.6: Appendix: Supplementary Material

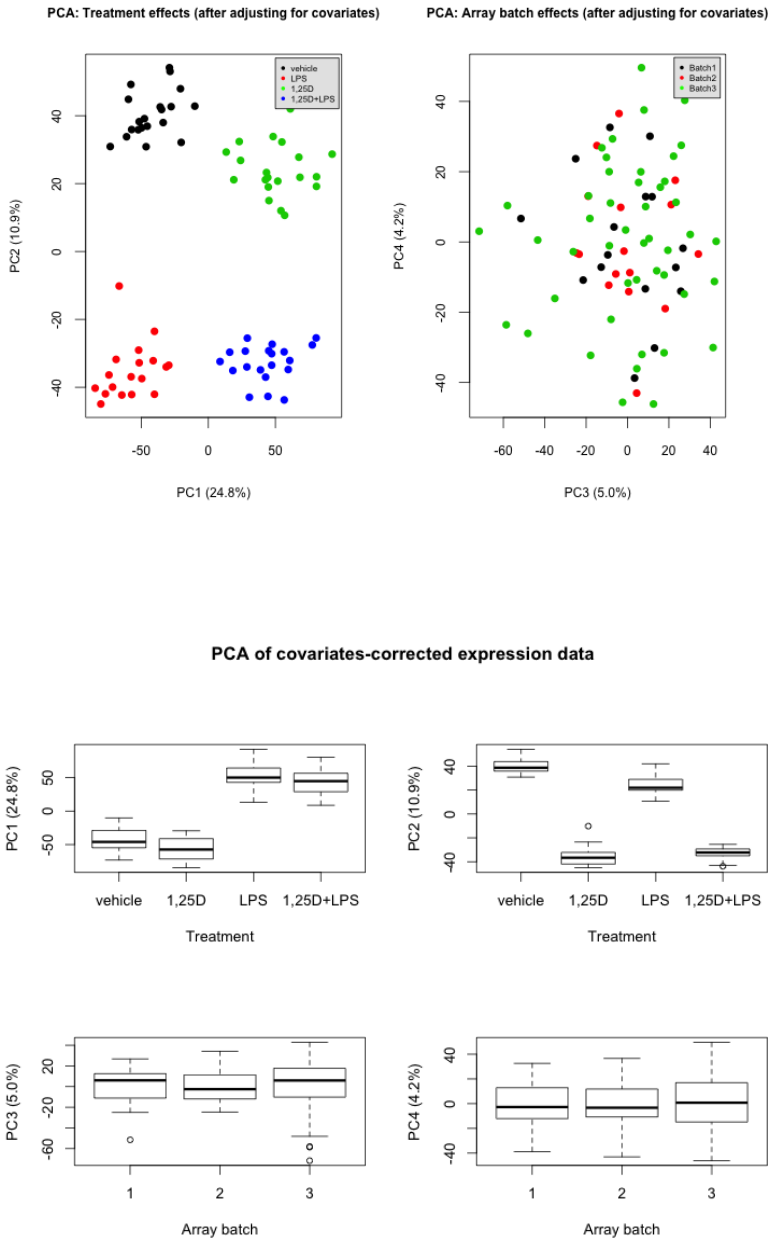
Supplementary Figure 3.1: Experimental Design. Primary monocytes were isolated from peripheral blood mononuclear cells (PBMCs) obtained from twenty healthy individuals of African-American (**AA**) and European-American (**EA**) ancestry. The monocytes were cultured in triplicate for 24 hours under four treatment conditions: i. the vehicle control (ethanol or **E**), ii. 1,25D (**V**), iii. LPS (**L**), and iv. 1,25D + LPS (**V + L**). The three replicates for each treatment were pooled for RNA extraction, and genome-wide gene expression was measured using Illumina microarrays.



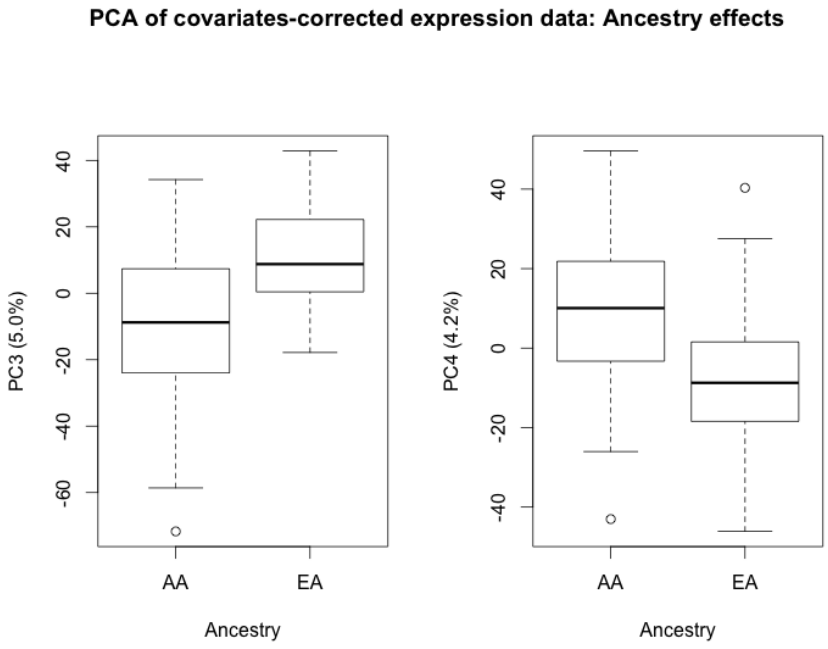
Supplementary Figure 3.2: Principal components analysis (PCA) of the expression data indicating the sources of transcriptome-wide variation. (**A**) Plot of PC1 vs. PC2, both of which captured the effects of treatment on the transcriptome. (**B**) Plot of PC3 vs. PC4, both of which captured array batch effects. (**C**) Boxplot showing effects of LPS treatment on expression data captured by PC1. (**D**) Boxplot showing effect of 1,25D treatment on expression data captured by PC2. (**E**) The array batch effects were captured by PC3 (**E**), and PC4 (**F**). The proportion of variation explained by the PCs is in parentheses.



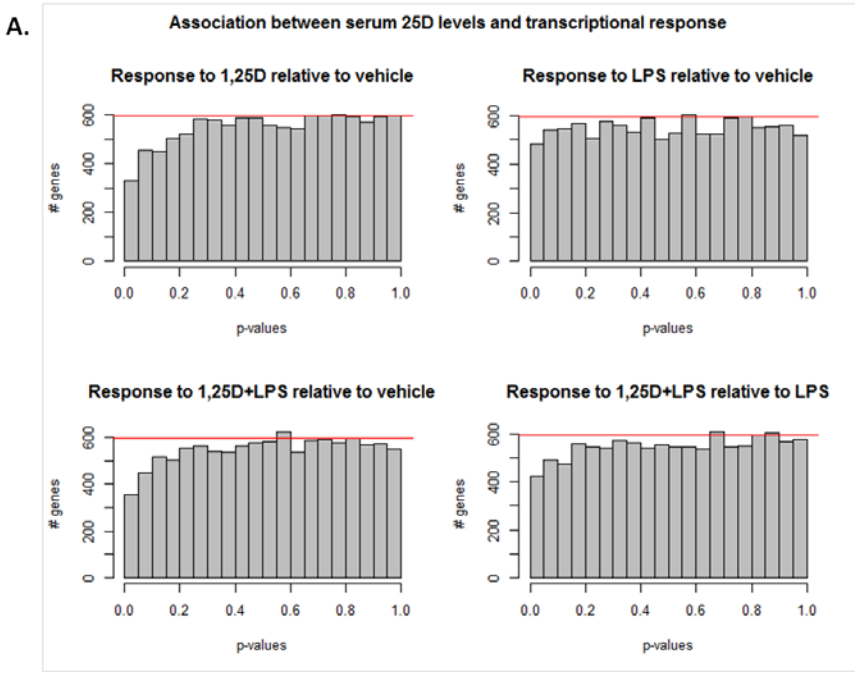
Supplementary Figure 3.3: **Principal components analysis (PCA) of covariates-corrected expression data indicating the sources of transcriptome-wide variation after correction for technical covariates.** (**A**) Plot of PC1 vs. PC2, both of which captured treatment effects. (**B**) Plot of PC3 vs. PC4 which shows that the array batch effects were corrected for. (**C**) Boxplot showing effect of LPS treatment on expression data captured by PC1. (**D**) Boxplot showing effect of 1,25D treatment on expression data captured by PC2. Array batch effects were no longer evident in the covariates-corrected expression data (**E**) and (**F**). The proportion of variation explained by the PCs is in parentheses.



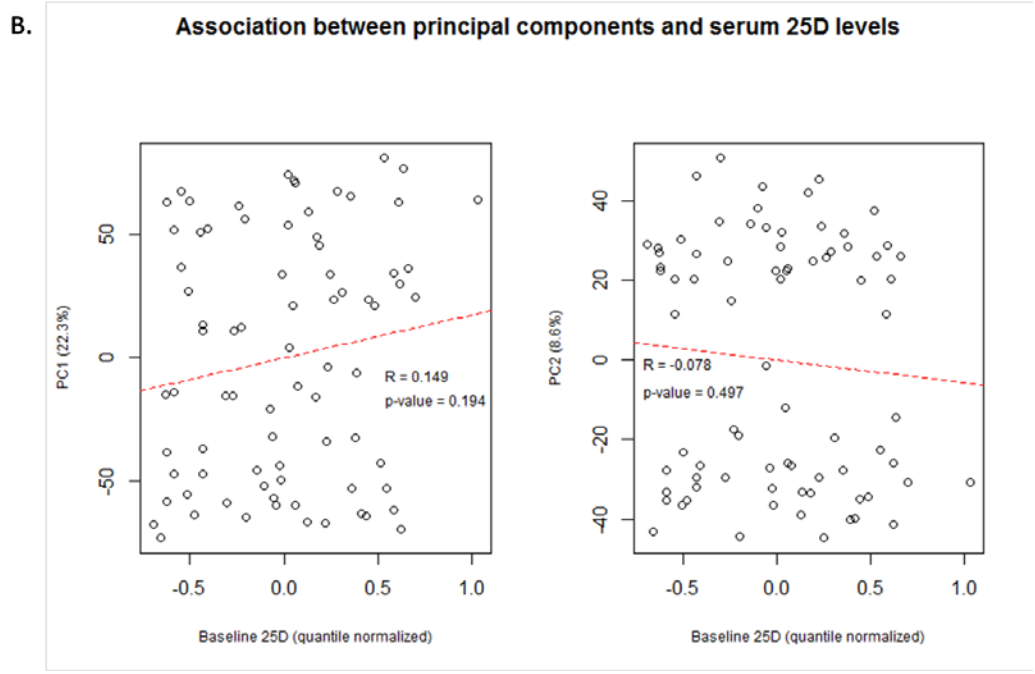
Supplementary Figure 3.4: Inter-ethnic variation. Boxplots showing inter-ethnic variation in covariates-corrected expression data captured by PC3 and PC4, with the proportion of variation explained in parenthesis.



Supplementary Figure 3.5: **Examining the effect of serum 25D levels on transcriptional response**. (**A**) Distribution of p-values from simple linear model measuring association between baseline 25D levels and log-fold change response to each of the four treatment conditions (1,25D, LPS, 1,25D+LPS relative to vehicle, and 1,25D+LPS relative to LPS). (**B**) Correlation between principal components 1 and 2 (PC1 and PC2), and baseline 25D levels. PC1 and PC2 captured the effect of LPS and 1,25D on the transcriptome.

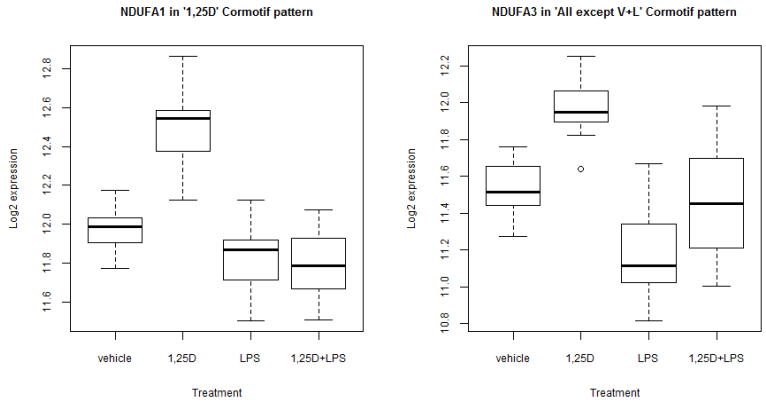


Supplementary Figure 3.5 - continued.

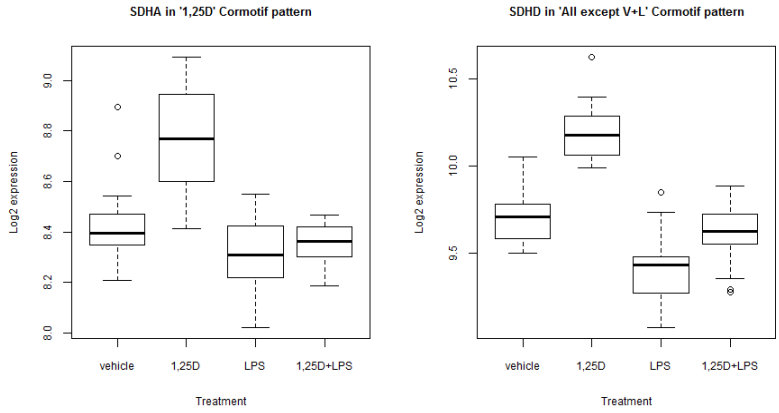


Supplementary Figure 3.6: **Treatment-specific response patterns in the oxidative phosphorylation pathway.** Boxplots of genes in the oxidative phosphorylation pathway clustered in the "1,25D" and "All except V+L" Cormotif patterns with treatment-specific response patterns. Genes representing the five respiratory complexes are shown in **A-E**. Oxidative phosphorylation pathway genes in the "1,25D" pattern were responsive only to 1,25D, while genes in the "All except V+L" pattern were responsive to both 1,25D and LPS in opposite directions.

A: Complex I

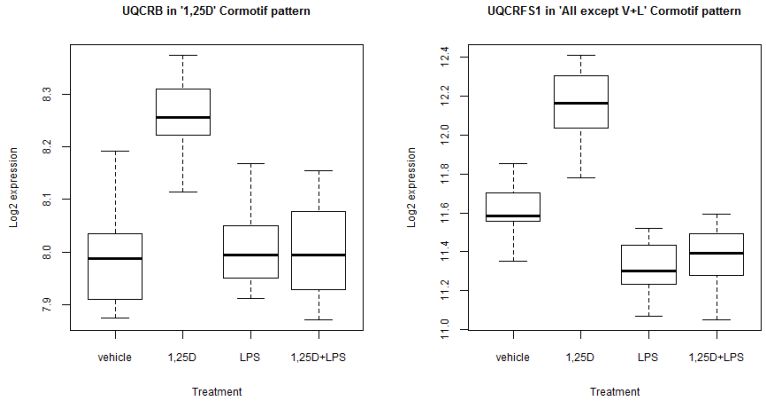


B: Complex II

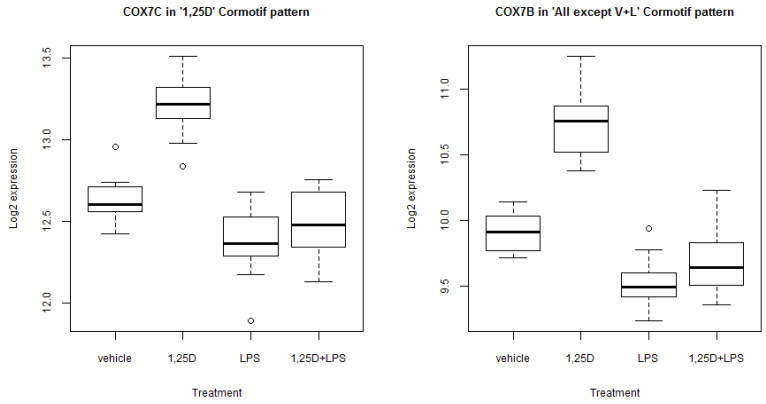


Supplementary Figure 3.6 - continued.

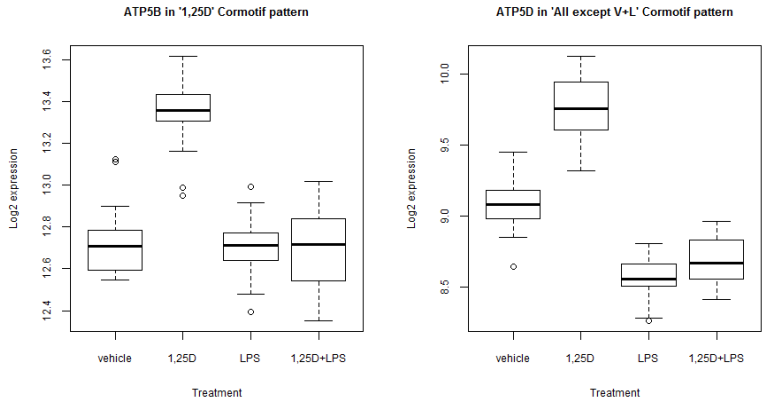
C: Complex III



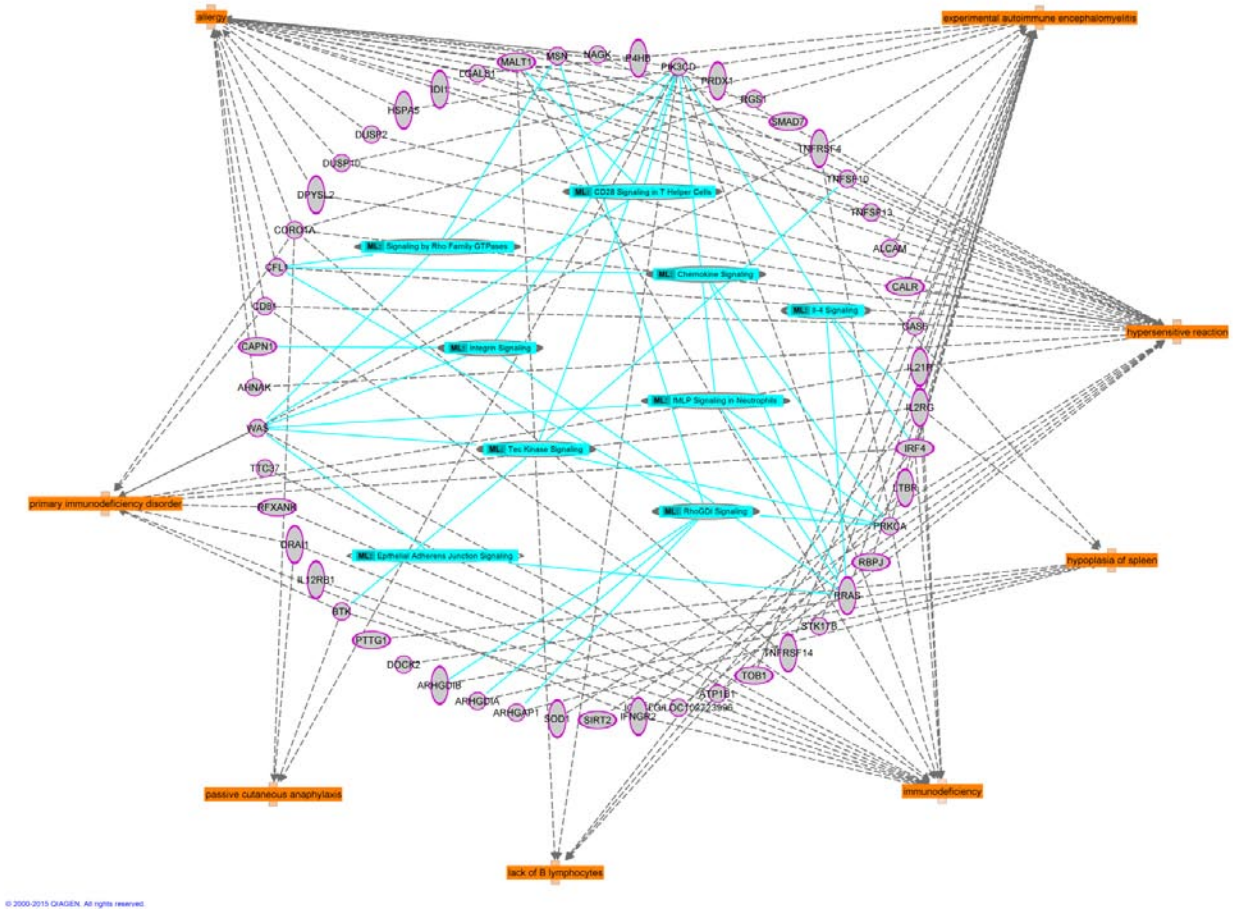
D: Complex IV



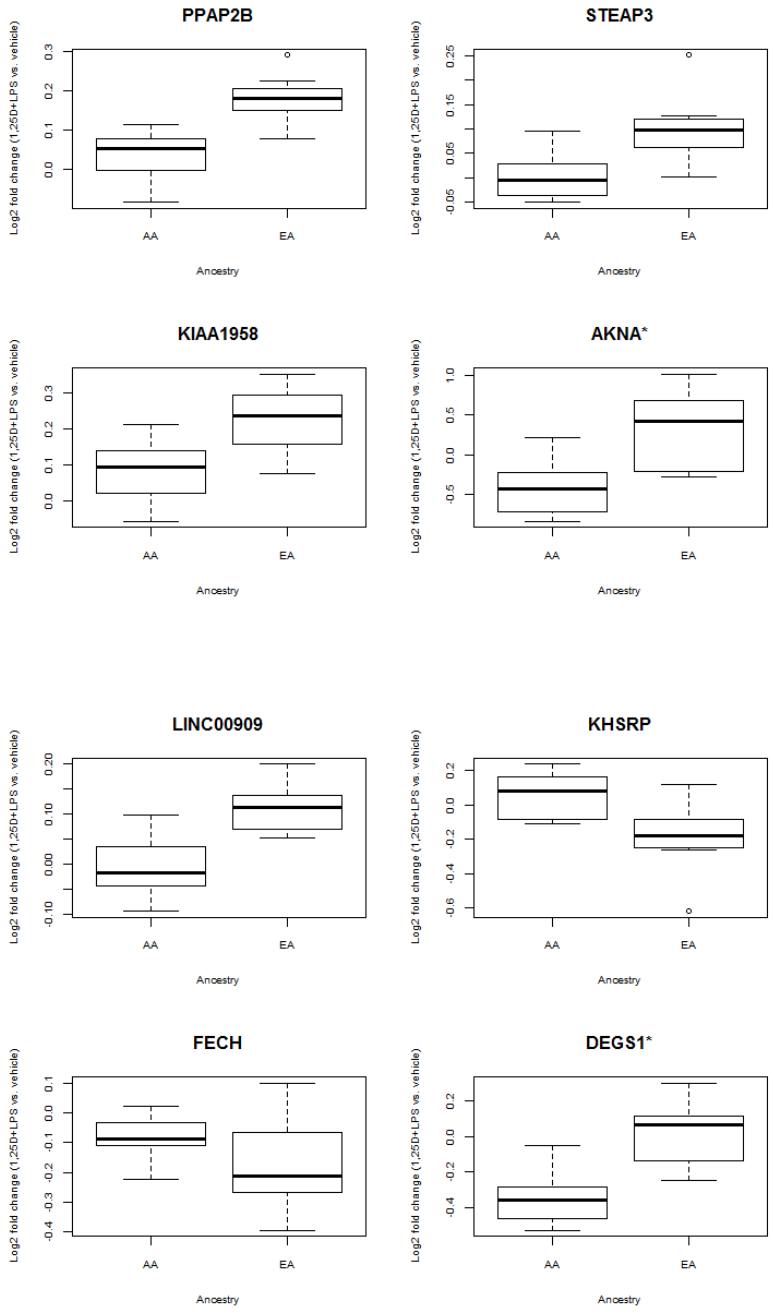
E: Complex V



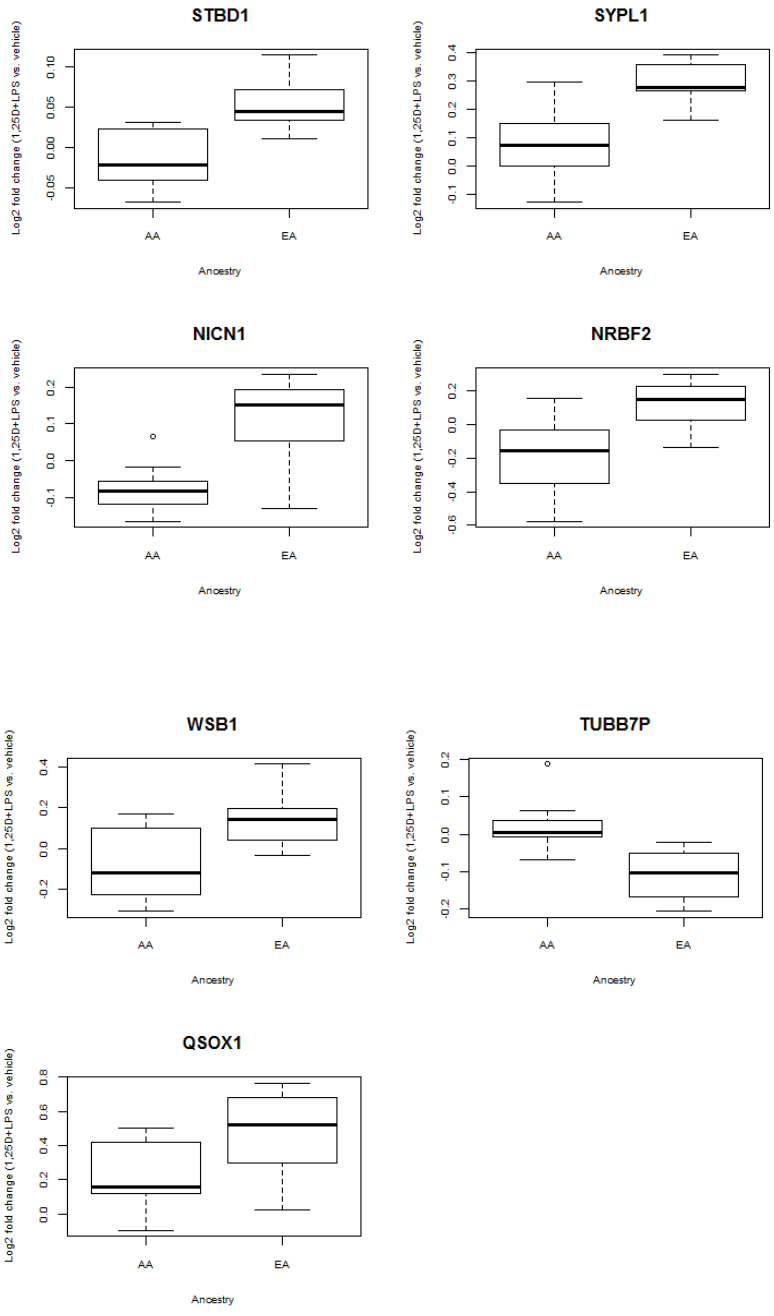
Supplementary Figure 3.7: **Network of down-regulated genes in the "1,25D-all" Cormotif pattern.** Biological pathways enriched among these genes are highlighted in cyan. Immunological and Inflammatory diseases enriched among these genes are highlighted in orange, while genes associated with diseases are circled in purple.



Supplementary Figure 3.8: **Genes with significant inter-ethnic differential expression.** Boxplots of genes with different log fold change in transcript levels between the two ethnic groups in response to 1,25D+LPS relative to vehicle (V + L vs. E) at a FDR < 0.10. 13 genes showed stronger response in EA's, while the 2 genes indicated with asterisks (*AKNA*, and *DEGS1*) showed stronger response in AA's. **AA** = African-American; **EA** = European-American.



Supplementary Figure 3.8 - continued.



Supplementary Table 3.1: **Sample characteristics**. Averaged sample covariates data are compared across ancestries, with p-values obtained from t-test.

Covariates	**EA**	**AA**	**p-value**
Age	29.7	30.1	0.90
Gender (F/M)	3/7	0/10	2.04×10^{-4}
Serum 25D concentration (nM)	52.8	35.3	6.92×10^{-6}
Serum PTH (pM)	27.6	32.7	0.26
RNA concentration (ng/mL)	53.6	44.0	0.16
RIN	8.2	8.6	0.90

EA = European-American, **AA** =African-American, **F** = Female, **M** = Male.

Supplementary Table 3.2: **Principal components analysis of variance-stabilized log2-transformed expression data.** (**A**) P-values for association between first six principal components (PCs) and sample covariates, with proportion of variance in expression data explained by each principal component at the bottom row. (**B**) R-squared values measuring effect sizes of associations between PCs and covariates.

A.

P-values	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Sample ID	0.966	0.993	0.205	0.094	5.01×10^{-5}	6.17×10^{-5}
Array Batch	0.164	0.265	4.95×10^{-6}	3.44×10^{-16}	0.392	0.757
Treatment	6.08×10^{-17}	4.68×10^{-6}	0.332	0.334	0.695	0.711
Month	0.303	0.801	0.834	0.001	0.043	1.0×10^{-3}
Age	0.853	0.946	1.36×10^{-6}	0.617	4.0×10^{-3}	0.097
Gender	0.314	0.547	0.272	0.863	0.124	0.739
Ancestry	0.711	0.911	0.399	0.403	0.047	0.004
Serum 25D	0.205	0.920	1.0×10^{-3}	0.147	0.743	0.929
Serum PTH	0.569	0.615	6.1×10^{-6}	0.055	0.758	0.336
RNA concentration	1.54×10^{-5}	0.986	0.029	0.158	0.935	0.061
RIN	0.160	9.0×10^{-3}	5.0×10^{-3}	0.277	0.524	0.973
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Proportion of Variance	0.223	0.086	0.067	0.058	0.039	0.034

B.

R-squared values	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Sample ID	2.47×10^{-5}	1.12×10^{-6}	0.021	0.037	0.196	0.191
Array Batch	0.025	0.016	0.241	0.586	0.010	1.0×10^{-3}
Treatment	0.604	0.242	0.012	0.012	2.0×10^{-3}	2.0×10^{-3}
Month	0.014	1×10^{-3}	1×10^{-3}	0.139	0.053	0.140
Age	4.56×10^{-3}	6.01×10^{-5}	0.266	3×10^{-3}	0.103	0.036
Gender	0.013	5.0×10^{-3}	0.016	3.95×10^{-4}	0.031	1×10^{-3}
Ancestry	0.002	1.65×10^{-4}	9×10^{-3}	9×10^{-3}	0.051	0.103
Serum 25D	0.021	1.35×10^{-4}	0.136	0.027	1.0×10^{-3}	1.05×10^{-4}
Serum PTH	0.004	3.0×10^{-3}	0.237	0.048	1.0×10^{-3}	0.012
RNA concentration	0.219	3.89×10^{-6}	0.061	0.026	8.80×10^{-5}	0.045
RIN	0.026	0.086	0.099	0.016	5×10^{-3}	1.0×10^{-4}

Supplementary Table 3.3: **Principal components analysis of variance-stabilized log2-transformed expression data adjusted for covariates.** (**A**) P-values for association between first six principal components (PCs) and sample covariates, with proportion of variance in expression data explained by each principal component at the bottom row. (**B**) R-squared values measuring effect sizes of associations between PCs and covariates.

A.

P-values	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Sample ID	0.937	0.716	4.5×10^{-7}	7.0×10^{-3}	0.041	3.50×10^{-4}
Array Batch	0.853	0.865	0.961	0.938	0.963	0.989
Treatment	1.62×10^{-17}	2.76×10^{-7}	0.715	0.402	0.865	0.996
Month	0.829	0.843	0.955	0.927	0.956	0.987
Age	0.634	0.714	0.810	0.886	0.882	0.799
Gender	0.969	0.940	0.081	0.151	0.299	0.056
Ancestry	0.866	0.689	1.04×10^{-5}	1.0×10^{-3}	0.017	1.66×10^{-6}
Serum 25D	0.518	0.888	0.037	0.129	0.253	0.026
Serum PTH	0.381	0.992	0.536	0.850	0.851	0.565
RNA concentration	3.56×10^{-5}	0.754	0.667	0.045	0.378	0.820
RIN score	0.345	3.0×10^{-3}	0.516	0.576	0.374	0.499
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Proportion of Variance	0.248	0.109	0.050	0.042	0.034	0.031

B.

R-squared values	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
Sample ID	8.34×10^{-5}	2.0×10^{-3}	**0.286**	0.090	0.054	0.156
Array Batch	4.52×10^{-4}	3.83×10^{-4}	3.16×10^{-5}	8.09×10^{-5}	2.92×10^{-5}	2.53×10^{-6}
Treatment	**0.617**	**0.295**	2.0×10^{-3}	9.0×10^{-3}	3.83×10^{-4}	3.50×10^{-7}
Month	0.001	1.0×10^{-3}	4.30×10^{-5}	1.10×10^{-4}	3.97×10^{-5}	3.44×10^{-6}
Age	0.003	2.0×10^{-3}	1.0×10^{-3}	2.72×10^{-4}	2.94×10^{-4}	1.0×10^{-3}
Gender	1.99×10^{-5}	7.58×10^{-5}	0.040	0.027	0.014	0.047
Ancestry	3.79×10^{-4}	2.0×10^{-3}	**0.227**	0.140	0.073	**0.262**
Serum 25D	6.0×10^{-3}	2.62×10^{-4}	0.056	0.030	0.017	0.063
Serum PTH	0.010	1.31×10^{-6}	5.0×10^{-3}	4.76×10^{-4}	4.64×10^{-4}	4.0×10^{-3}
RNA concentration	**0.203**	1.0×10^{-3}	2.0×10^{-3}	0.052	0.010	1.0×10^{-3}
RIN score	0.012	**0.113**	6.0×10^{-3}	4.0×10^{-3}	0.010	6.0×10^{-3}

Supplementary Table 3.4: **Biological pathways enriched at a FDR < 0.05 among genes significantly DE in response to single treatment with 1,25D or LPS, identified using linear mixed-effects model.** The biological pathways are stratified by direction of transcription response, where up-regulated genes are indicated in **green** while down-regulated genes are indicated in **red**. The biological pathways are also grouped according to the similar response patterns to 1,25D in the **V vs. E** (1,25D relative to vehicle) and **V + L vs. L** (1,25D+LPS relative to LPS) treatment categories, and the similar response patterns to LPS in the **L vs. E** (LPS relative to vehicle) and **V + L vs. E** (1,25D+LPS relative to vehicle) treatment categories.

Biological Pathway	**V vs. E**	**V + L vs. L**	**L vs. E**	**V + L vs. E**
EIF2 Signaling				
mTOR Signaling				
Oxidative Phosphorylation				
Mitochondrial Dysfunction				
Adipogenesis pathway				
Insulin Receptor Signaling				
fMLP Signaling in Neutrophils				
NRF2-mediated Oxidative Stress Response				
Signaling by Rho Family GTPases				
Role of NFAT in Regulation of the Immune Response				
Chemokine Signaling				
Remodeling of Epithelial Adherens Junctions				
Antigen Presentation Pathway				
Androgen Signaling				
Germ Cell-Sertoli Cell Junction Signaling				
Phagosome maturation				
Tec Kinase Signaling				
Phospholipase C Signaling				
Integrin Signaling				
Role of JAK1, JAK2 and TYK2 in Interferon Signaling				
T Helper Cell Differentiation				
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages				
OX40 Signaling Pathway				
PI3K Signaling in B Lymphocytes				
IL-8 Signaling				
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses				
Dendritic Cell Maturation				
CD28 Signaling in T Helper Cells				
B Cell Receptor Signaling				
Communication between Innate and Adaptive Immune Cells				
NF-kB Signaling				

Supplementary Table 3.4 – continued.

Mismatch Repair in Eukaryotes				
G-Protein Coupled Receptor Signaling				
Protein Ubiquitination Pathway				
IL-4 Signaling				
Leukocyte Extravasation Signaling				
IL-10 Signaling				
IL-6 Signaling				
Acute Phase Response Signaling				
TNFR2 Signaling				
TNFR1 Signaling				
Glucocorticoid Receptor Signaling				
CD40 Signaling				
IL-17 Signaling				
LPS-stimulated MAPK Signaling				
Toll-like Receptor Signaling				
JAK/Stat Signaling				
iNOS Signaling				
B Cell Activating Factor Signaling				
Apoptosis Signaling				
IL-15 Signaling				
CD27 Signaling in Lymphocytes				
IL-1 Signaling				
IL-2 Signaling				
Role of JAK family kinases in IL-6-type Cytokine Signaling				
T Cell Receptor Signaling				
IL-9 Signaling				
Oncostatin M Signaling				
IL-22 Signaling				
IL-12 Signaling and Production in Macrophages				
VDR/RXR Activation				
Differential Regulation of Cytokine Production in Macrophages and T Helper Cells by IL-17A and IL-17F				
tRNA Charging				
Fatty Acid beta-oxidation I				
Galactose Degradation I (Leloir Pathway)				

Supplementary Table 3.5: Biological pathways enriched at a FDR < 0.05 among genes in the different Cormotif patterns. The pathways are stratified by direction of transcription response in each Cormotif pattern, where up-regulated genes are indicated in **green** while down-regulated genes are indicated in **red**. For the pathways enriched among genes in the "All" and "All except V+L" patterns, **green** = genes up-regulated by 1,25D and down-regulated by LPS, **red** = genes down-regulated by 1,25D and up-regulated by LPS.

Biological Pathway	All	All except V+L	1,25D	1,25D-all	LPS
EIF2 Signaling					
Regulation of eIF4 and p70S6K Signaling					
mTOR Signaling					
Purine Nucleotides De Novo Biosynthesis II					
tRNA Charging					
Spermidine Biosynthesis I					
Adenine and Adenosine Salvage I					
Inosine-5'-phosphate Biosynthesis II					
Oxidative Phosphorylation					
Mitochondrial Dysfunction					
Granulocyte Adhesion and Diapedesis					
IL-8 Signaling					
NF-κB Signaling					
IL-17A Signaling in Fibroblasts					
PPAR Signaling					
TNFR2 Signaling					
Role of NFAT in Regulation of the Immune Response					
Unfolded protein response					
Mechanisms of Viral Exit from Host Cells					
4-1BB Signaling in T Lymphocytes					
phagosome maturation					
Induction of Apoptosis by HIV1					
STAT3 Pathway					
TWEAK Signaling					
UDP-N-acetyl-D-glucosamine Biosynthesis II					
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis					
CXCR4 Signaling					
IL-10 Signaling					
Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza					
Apoptosis Signaling					
Macropinocytosis Signaling					
Adipogenesis					
Role of JAK family kinases in IL-6-type Cytokine Signaling					
RhoGDI Signaling					
B Cell Receptor Signaling					

Supplementary Table 3.4 – continued.

Chemokine Signaling					
fMLP Signaling in Neutrophils					
Actin Nucleation by ARP-WASP Complex					
RhoA Signaling					
Integrin Signaling					
Tec Kinase Signaling					
Hereditary Breast Cancer Signaling					
Regulation of Actin-based Motility by Rho					
p70S6K Signaling					
Ephrin Receptor Signaling					
Role of BRCA1 in DNA Damage Response					
NRF2-mediated Oxidative Stress Response					
Signaling by Rho Family GTPases					
Actin Cytoskeleton Signaling					
CD28 Signaling in T Helper Cells					
Axonal Guidance Signaling					
FAK Signaling					
Non-Small Cell Lung Cancer Signaling					
N-acetylglucosamine Degradation II					
CCR3 Signaling in Eosinophils					
Remodeling of Epithelial Adherens Junctions					
Leukocyte Extravasation Signaling					
Fcγ Receptor-mediated Phagocytosis in Macrophages and Monocytes					
Epithelial Adherens Junction Signaling					
IL-4 Signaling					
Glioma Signaling					
IL-6 Signaling					
Role of IL-17A in Arthritis					
IL-17A Signaling in Airway Cells					
Role of IL-17F in Allergic Inflammatory Airway Diseases					
TREM1 Signaling					
CD40 Signaling					
Dendritic Cell Maturation					
IL-17 Signaling					
Acute Phase Response Signaling					
Type II Diabetes Mellitus Signaling					
Glucocorticoid Receptor Signaling					
Role of IL-17A in Psoriasis					
RANK Signaling in Osteoclasts					
Regulation of IL-2 Expression in Activated and Anergic T Lymphocytes					
Agranulocyte Adhesion and Diapedesis					
Type I Diabetes Mellitus Signaling					
Amyloid Processing					
HGF Signaling					
T Helper Cell Differentiation					
NF-κB Activation by Viruses					
Toll-like Receptor Signaling					
IL-22 Signaling					
Hypoxia Signaling in the Cardiovascular System					
Cholecystokinin/Gastrin-mediated Signaling					
EGF Signaling					

Supplementary Table 3.4 – continued.

Erythropoietin Signaling					
TNFR1 Signaling					
Gαq Signaling					
Role of Tissue Factor in Cancer					
Chondroitin and Dermatan Biosynthesis					
April Mediated Signaling					
PEDF Signaling					
PKCθ Signaling in T Lymphocytes					
Activation of IRF by Cytosolic Pattern Recognition Receptors					
Graft-versus-Host Disease Signaling					
JAK/Stat Signaling					
PI3K Signaling in B Lymphocytes					
LPS-stimulated MAPK Signaling					
B Cell Activating Factor Signaling					
GM-CSF Signaling					
Molecular Mechanisms of Cancer					
Pyridoxal 5'-phosphate Salvage Pathway					
G-Protein Coupled Receptor Signaling					
IL-2 Signaling					
MIF-mediated Glucocorticoid Regulation					
Gα12/13 Signaling					
Acute Myeloid Leukemia Signaling					
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis					
IL-15 Signaling					
Circadian Rhythm Signaling					
HMGB1 Signaling					
IL-1 Signaling					
Death Receptor Signaling					
RAR Activation					
IL-12 Signaling and Production in Macrophages					
Role of JAK1, JAK2 and TYK2 in Interferon Signaling					
IL-9 Signaling					
Antioxidant Action of Vitamin C					
Wnt/β-catenin Signaling					
Glutaryl-CoA Degradation					
Isoleucine Degradation I					
Tryptophan Degradation III (Eukaryotic)					
Fatty Acid β-oxidation I					
Galactose Degradation I (Leloir Pathway)					
Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate)					
Ketolysis					
Mevalonate Pathway I					
Heme Biosynthesis II					
Cell Cycle Control of Chromosomal Replication					
Folate Transformations I					
TCA Cycle II (Eukaryotic)					
Superpathway of Methionine Degradation					
Tetrapyrrole Biosynthesis II					
Folate Polyglutamylation					
Ketogenesis					
Dolichyl-diphosphooligosaccharide Biosynthesis					

Supplementary Table 3.6: **Diseases enriched among down-regulated genes in the "1,25D-all" Cormotif pattern at a FDR < 0.05.** B-H p-value* = Benjamini-Hochberg multiple testing corrected p-value.

Disease Category	B-H p-value*	Genes down-regulated in "1,25D-all" Cormotif pattern
Inflammatory Response	6.24×10^{-5}	*DPYSL2,GAS6,SGK1,DUSP3,TMSB10/TMSB4X,ICOSLG/LOC102723996,CFL1,RRAS,PLEC,ANXA2,RAP1A,ABCA7,OGG1,MTMR4,CAPN1,ALCAM,PIK3CD,ARHGAP1,ACTG1,CNN2,HAMP,RAB32,POU2F2,PRDX1,IL21R,HYOU1,TNFSF10,MMP25,ANPEP,DUSP2,SASH3,TNFSF12,DOCK2,ORAI1,RNASE2,C12orf4,IFNGR2,GNAQ,TOB1,CD58,NRROS,KIAA0226,GPR183,ABHD6,GNAI3,LY96,WAS,IMPDH1,TNFSF13,RBPJ,ARHGDIA,CD81,CH25H,RGS1,HYAL2,NAGK,TNFRSF4,PFN1,IDI1,IL12RB1,HSPA5,CD300A,RHOB,LTBR,SH3KBP1,STK17B,S100A10,NBEAL2,P4HB,IL2RG,PTPN6,BMP2K,MALT1,NFKBID,FZD5,MGAT2,ABCD1,SPRED1,CCL24,OTULIN,HSP90B1,CCDC88A,AHNAK,DUSP10,CD22,PRKCA,TPMT,CALR,NAIP,IRF4,PRMT2,AP3D1,SMAD7,SOD1,PLXND1,TNFRSF14,BTK,ZBTB46,CORO1A,TLR6,CTSC,DNM1L,LGALS1,MSN*
Immunological Disease	1.55×10^{-3}	*RFXANK,CD81,DPYSL2,RGS1,NAGK,TNFRSF4,IDI1,IL12RB1,GAS6,SGK1,PTTG1,DDB2,HSPA5,RB1,SEC24D,RHOB,OLIG2,XPO1,LTBR,ICOSLG/LOC102723996,STK17B,NBEAL2,P4HB,IL2RG,PTPN6,CFL1,RRAS,DLEU2,TSPAN33,MALT1,ARHGDIB,ABCA7,OGG1,DVL2,MTMR4,CAPN1,CECR6,BTG2,RASSF4,ALCAM,PIK3CD,ARHGAP1,ACTG1,POU2F2,PRDX1,IL21R,DAGLA,TNFSF10,DUSP2,TTC37,AHNAK,TNFSF12,DUSP10,DOCK2,EAF2,CD22,PALD1,BLM,ORAI1,PRKCA,CALR,ATP1B1,IRF4,TGFBI,SMAD7,IFNGR2,TOB1,CD58,SOD1,TNFRSF14,BTK,ZBTB46,CRELD2,SIRT2,MUM1,WAS,TNFSF13,TLR6,CORO1A,ARHGDIA,RBPJ,FEN1,DNM1L,SEPT6,MSN,LGALS1*
Inflammatory Disease	1.55×10^{-3}	*TNFRSF4,GAS6,IL21R,TNFSF10,HSPA5,HSP90B1,DUSP10,LTBR,STK17B,PRKCA,CALR,NBEAL2,IL2RG,PTPN6,IRF4,RRAS,SMAD7,TOB1,TSPAN33,SOD1,ABCA7,TNFRSF14,BTK,CORO1A,ALCAM,RBPJ,LGALS1*
Neurological Disease	1.55×10^{-3}	*SPRED1,PFN1,TNFRSF4,VPS35,GAS6,IL21R,CWF19L1,TNFSF10,SETX,HSPA5,HSP90B1,FANCD2,DUSP10,DHTKD1,FGD4,LTBR,RTN2,STK17B,PRKCA,CALR,IL2RG,IRF4,TGFBI,RRAS,GNAQ,TOB1,LRSAM1,LZTR1,SOD1,TNFRSF14,KIAA0226,SLC33A1,CORO1A,ALCAM,RBPJ,LGALS1*
Hematological Disease	4.78×10^{-3}	*HYAL2,TNFRSF4,SGK1,GAS6,PTTG1,DDB2,HSPA5,RB1,SEC24D,RHOB,OLIG2,XPO1,LTBR,STK17B,NBEAL2,IL2RG,CFL1,DLEU2,TSPAN33,ANXA2,MALT1,ABCA7,OGG1,DVL2,MTMR4,CECR6,CAPN1,BTG2,RASSF4,PIK3CD,ACTG1,ARHGAP1,POU2F2,PRDX1,IL21R,TNFSF10,DAGLA,MKL1,HSP90B1,DUSP10,EAF2,DOCK2,CD22,PALD1,BLM,TPMT,IRF4,TGFBI,IFNGR2,CD58,SOD1,TNFRSF14,BTK,CRELD2,SIRT2,WAS,MUM1,TLR6,FEN1,DNM1L,SEPT6,LGALS1*
Hereditary Disorder	2.32×10^{-2}	*RFXANK,PFN1,VPS35,PTTG1,CWF19L1,SETX,HSPA5,HSP90B1,AHNAK,TNFSF12,DOCK2,DHTKD1,FGD4,RTN2,ORAI1,CALR,IL2RG,PTPN6,TGFBI,PLEC,LRSAM1,SOD1,ARHGDIB,KIAA0226,BTK,SLC33A1,IMPDH1,PIK3CD,ARHGDIA,ARHGAP1*
Hypersensitivity Response	3.52×10^{-2}	*BTK,IL2RG,TNFRSF4,MTMR4,WAS,CAPN1,CORO1A,C12orf4,PIK3CD,DUSP2,CD300A,ORAI1*

Supplementary Table 3.6 – continued.

Cancer	3.52×10^{-2}	*CD81,TUBA1B,PHLDA1,ARHGAP26,TNFRSF4,GAS6,NDRG2,SGK1,PTTG1,KLF6,MCUR1,DDB2,POTEG,HSPA5,CACYBP,VASH1,RB1,SEC24D,RHOB,OLIG2,DHTKD1,XPO1,LTBR,PLCL1,TMSB10/TMSB4X,ICOSLG/LOC102723996,STK17B,S100A10,NBEAL2,PTPN6,IL2RG,CFL1,DLEU2,PLEC,TSPAN33,SIPA1L2,ANXA2,ARHGDIB,ABCA7,OGG1,DVL2,MTMR4,TES,CECR6,BTG2,RASSF4,KIDINS220,ALCAM,KLHL12,PIK3CD,ACTG1,EMILIN2,POU2F2,PRDX1,IL21R,HMGN1,HYOU1,DAGLA,TNFSF10,MKL1,ANPEP,USO1,CCDC88A,HSP90B1,ACTR3,FANCD2,TNFSF12,DOCK2,EAF2,CD22,PALD1,BLM,PRKCA,ORAI1,CALR,IRF4,MAP3K6,TGFBI,DROSHA,SMAD7,CD58,SOD1,TNFRSF14,BTK,LACC1,ZBTB46,CRELD2,SIRT2,WAS,MUM1,IMPDH1,TNFSF13,TLR6,FEN1,DNM1L,CTSC,SEPT6,LGALS1*
Organismal Injury and Abnormalities	3.52×10^{-2}	*TUBA1B,GAS6,SGK1,DDB2,CACYBP,RB1,SEC24D,OLIG2,TMSB10/TMSB4X,PLCL1,ICOSLG/LOC102723996,CFL1,PLEC,ANXA2,SIPA1L2,TSPAN33,ABCA7,OGG1,ARHGDIB,DVL2,MTMR4,BTG2,KIDINS220,ALCAM,PIK3CD,ARHGAP1,ACTG1,EMILIN2,PRDX1,POU2F2,IL21R,HMGN1,HYOU1,TNFSF10,DAGLA,ANPEP,SASH3,USO1,ACTR3,TNFSF12,DOCK2,PALD1,ORAI1,TGFBI,DROSHA,GNAQ,TOB1,CD58,ZG16B,MUM1,WAS,TNFSF13,IMPDH1,ARHGDIA,FEN1,RBPJ,CD81,PHLDA1,TNFRSF4,ARHGAP26,NDRG2,PTTG1,KLF6,MCUR1,POTEG (includes others),HSPA5,VASH1,RHOB,DHTKD1,XPO1,LTBR,STK17B,S100A10,NBEAL2,PTPN6,IL2RG,DLEU2,LZTR1,TES,CECR6,RASSF4,KLHL12,SPRED1,MKL1,CCDC88A,HSP90B1,FANCD2,EAF2,CD22,BLM,PRKCA,CALR,IRF4,MAP3K6,SMAD7,SOD1,TNFRSF14,BTK,LACC1,CRELD2,ZBTB46,SIRT2,TLR6,ARHGAP31,DNM1L,CTSC,SEPT6,LGALS1*
Skeletal and Muscular Disorders	4.43×10^{-2}	*CD81,CALR,RB1,HSP90B1,AHNAK,RRAS,PLEC,HSPA5,NDN*
Developmental Disorder	4.64×10^{-2}	*RFXANK,CALR,IL2RG,SPRED1,TGFBI,PTTG1,GNAQ,LZTR1,HSPA5,ARHGDIB,SASH3,BTK,HSP90B1,WAS,DOCK2,ARHGAP31,RBPJ,PIK3CD,ARHGDIA,ARHGAP1,ORAI1,PRKCA*

Supplementary Table 3.7: Enrichment of VDR ChIP-seq peaks among genes responsive to 1,25D treatment. VDR ChIP-seq peaks were obtained from published datasets, while 1,25D responsive genes were obtained among those that were significantly differentially expressed in response to different 1,25D treatment conditions, from the linear mixed-effects and Cormotif analyses. Enrichment of VDR peaks was calculated using Fisher's exact test, comparing DE genes to non-DE genes.

Treatment	DE genes	DE genes with VDR binding site	Proportion of DE genes with VDR binding site	Non-DE genes	Non-DE genes with VDR binding site	Proportion of Non-DE genes with VDR binding site	Enrichment p-values
Linear mixed-effects model							
V vs. E	2887	202	0.07	8071	300	0.04	4.56×10^{-11}
V + L vs. E	4720	335	0.07	6238	280	0.04	3.97×10^{-8}
V + L vs. L	2405	209	0.09	8554	466	0.05	1.54×10^{-7}
Cormotif Analysis							
1,25D response ("All", "All except V+L", "1,25D" and 1,25D-all" Cormotifs)	2761	189	0.07	5737	186	0.03	3.28×10^{-12}
1,25D response: ("1,25D" and 1,25D-all" Cormotifs)	1132	114	0.10	5737	186	0.03	3.33×10^{-18}

CHAPTER 4: MAPPING VARIATION IN CELLULAR AND TRANSCRIPTIONAL RESPONSE TO 1,25-DIHYDROXYVITAMIN D_3 IN PERIPHERAL BLOOD MONONUCLEAR CELLS

4.1: Abstract

The active hormonal form of vitamin D, 1,25-dihydroxyvitamin D (1,25D) is an important modulator of the immune system, inhibiting cellular proliferation and regulating transcription of immune response genes. In order to characterize the genetic basis of variation in the immunomodulatory effects of 1,25D, we mapped quantitative traits of 1,25D response at both the cellular and the transcriptional level. We carried out a genome-wide association scan of percent inhibition of cell proliferation (I_{max}) induced by 1,25D treatment of peripheral blood mononuclear cells from 88 healthy African-American individuals. Two genome-wide significant variants were identified: rs1893662 in a gene desert on chromosome 18 ($p=2.32 \times 10^{-8}$) and rs6451692 on chromosome 5 ($p=2.55 \times 10^{-8}$), which may influence the anti-proliferative activity of 1,25D by regulating the expression of nearby genes such as the chemokine gene, *CCL28*, and the translation initiation gene, *PAIP1*. We also identified 8 expression quantitative trait loci at a FDR<0.10 for transcriptional response to 1,25D treatment, which include the transcriptional regulator ets variant 3-like (*ETV3L*) and EH-domain containing 4 (*EHD4*). In addition, we identified response eQTLs in vitamin D receptor binding sites near genes differentially expressed in response to 1,25D, such as FERM Domain Containing 6 (*FRMD6*), which plays a critical role in regulating both cell proliferation and apoptosis. Combining information from the GWAS of I_{max} and the response eQTL mapping enabled identification of putative I_{max}-associated

candidate genes such as *PAIP1* and the transcriptional repressor gene *ZNF649*. Overall, the variants identified in this study are strong candidates for immune traits and diseases linked to vitamin D, such as multiple sclerosis.

4.2: Introduction

Epidemiological studies have linked variation in the circulating inactive form of vitamin D, 25-hydroxyvitamin D_3 (25D), to risk of autoimmune diseases such as multiple sclerosis, type 1 diabetes and systemic lupus erythematosus [29, 47, 48, 81-84], consistent with the known effects of vitamin D as an immune system modulator [30, 33, 37, 38, 77]. Furthermore, genetic variation in the vitamin D pathway is linked to autoimmune disease risk. For example, several studies have highlighted associations between variants in *CYP27B1*, which encodes the enzyme that activates 25D to 1,25-dihydroxyvitamin D_3 (1,25D), and risk for multiple sclerosis [176-178].

The fact that immune cells express CYP27B1 indicates that active vitamin D can be produced intra-cellularly in the immune system in response to organismal demands such as infections. Immune cells also express the vitamin D receptor (VDR), which when bound by the active 1,25D, forms a heterodimer with the retinoid X receptor (RXR) and translocates to the nucleus, resulting in transcriptional regulation of vitamin D-responsive genes [30, 33, 37, 38, 58]. The genes regulated by 1,25D are involved in various pathways including metabolic regulation, antimicrobial response and inflammatory cytokine response [29, 41, 49, 64, 73-76].

Extensive inter-individual and inter-ethnic variation in the circulating levels of 25D levels has been reported, with lower levels on average in African Americans compared to European Americans [42-45]. These differences are known to be influenced by various factors such as sun exposure, dietary intake, as well as genetic variations in critical genes in the vitamin D metabolic pathway [177, 179, 180]. Despite the strong epidemiological associations of 25D levels and disease risk, randomized clinical trials aimed at testing the efficacy of vitamin D supplementation as a therapeutic intervention [49, 84-88, 91] have yielded mixed results [92, 93]. In addition to environmental confounders, these results could be due to inter-individual differences in the response to vitamin D, irrespective of its concentration in circulation or within the cells at the level of the target organ. Indeed, at least one study identified a polymorphism in the *VDR* gene that influenced the response to vitamin D supplementation [89]. However, beyond the *VDR* gene, little – if anything – is known about the contribution of genetics to the inter-individual variation in response to vitamin D.

The aim of this study was to map the genetic bases of inter-individual variation in the transcriptional response to 1,25D and in the inhibition of cell proliferation induced by 1,25D in primary immune cells. To isolate the effects of genetic variation on the response to active vitamin D rather than on its concentration, we treated primary peripheral blood mononuclear cells cultured *in vitro* with a fixed amount of 1,25D and, in parallel, with a vehicle control. This allowed us to characterize the response to vitamin D both at the cellular and transcriptional level and to identify genetic variants associated with cellular and transcriptional response to 1,25D.

4.3: Methods

Samples

Peripheral blood was obtained from 88 African American (AA) donors collected by Research Blood Components (http://researchbloodcomponents.com/) as part of a larger study on transcriptional response [94]. All subjects were healthy donors and were not on any medication. All donors to Research Blood Components are required to sign an Institutional Review Board (IRB)-approved consent form giving permission to collect blood, and use it for research purposes. The IRB at the University of Chicago determined that this study is not human subject research because blood samples were not shipped with individually identifiable information. Self-reported ethnicity, age, gender, date, and time of blood drawing were recorded for each donor. Samples were processed in multiple successive batches. Batch number was recorded and used as a covariate.

Cell culture and treatment

The experimental design is illustrated in **Supplementary Figure 4.1**. We isolated peripheral blood mononuclear cells (PBMCs) from heparin-treated whole blood by density gradient centrifugation using Ficoll-Paque PLUS medium (GE Healthcare Life Sciences, Pittsburgh, PA). PBMCs were washed in PBS and transferred to RPMI supplemented with 10% charcoal-stripped fetal bovine serum. Each sample was then divided into one aliquot of 1.8×10^6 cells for measuring cell proliferation, and one aliquot of 9×10^6 cells for genome-wide transcriptional profiling. For the cell proliferation measurements, PBMCs were cultured at 2×10^5 cells per well in 10% charcoal-stripped media in 96-well plates.

Each donor was treated in triplicate with phytohemagglutinin (PHA) (2.5ug/ml) and either vehicle (EtOH) or 1,25-dihydroxyvitamin D3 (1,25D) (100nM) for 48 hours. For transcriptional profile measurements, PBMCs from each donor were cultured at 10^6 cells per well in 10% charcoal-stripped media in 24-well plates. As with the cellular proliferation measurements, each donor was treated in triplicate with PHA (2.5ug/ml) and either vehicle or 1,25D (100nM) for 6 hours. Cell type composition of the PBMCs was measured using flow cytometry as previously reported for these samples [94], where proportions of T cells including T helper cells ($CD4^+$) and cytotoxic T ($CD8^+$) cells, B cells, monocytes and neutrophils were measured using antibodies specifically targeting these cell types.

Cellular proliferation measurements

After 48 hours of treatment, cell proliferation was measured by H^3-thymidine incorporation using standard protocols as previously described [141]. The median value was taken from across the three replicates. Percent inhibition of proliferation by 1,25D (I_{max}) was calculated as 1 – [(proliferation in 1,25D+PHA)/(proliferation in EtOH+PHA)], and fit to a normal distribution. Associations between covariates and I_{max} were tested using a simple linear regression.

Transcriptional response profiling

After 6 hours of treatment with PHA and 1,25D or vehicle, the three replicates from each donor were pooled before RNA extraction. Total RNA was extracted from each pool with the RNeasy Plus Mini Kit (Qiagen 74134). Total RNA was reverse transcribed into

cDNA, labeled, hybridized to Illumina (San Diego, CA, USA) Human HT-12 v3 Expression Beadchips and scanned at the University of Chicago Functional Genomics Core facility. We performed low-level microarray analyses using the Bioconductor software package LUMI [140] in R, as previously described [141]. Briefly, we annotated probes by mapping their sequence to RefSeq (GRCh37) transcripts using BLAT. We discarded probes that mapped to multiple genes, or contained one or more HapMap SNPs. We applied variance stabilization transformation to all arrays, discarded poor quality probes, and quantile normalized the arrays using the default method implemented in the lumiN function. After these filters, probes mapping to 11,897 genes were used in downstream analyses. We used a paired t-test to identify genes that were differentially expressed between 1,25D- and vehicle-treated samples. False-discovery rates (FDR) were estimated using the q value function in R [144]. Gene set enrichment analysis was performed using the commercially available software Ingenuity Pathway Analysis (IPA).

Genome-wide association of inhibition of cellular proliferation by 1,25D (I_{max})

Samples were genotyped on two Illumina Omni BeadChip platforms, with a total of 884,015 SNPs across the genome genotyped for each donor, as previously described [94]. We then imputed genotypes at all SNPs identified in the 1000 Genomes Project [181] using IMPUTE2 [182], applying the output file flag option "-pgs_miss", which replaces the missing genotypes at typed SNPs with imputed genotypes. We filtered SNPs for minor allele frequency (>0.1), imputation quality (>0.9), and departure from Hardy Weinberg equilibrium (p> 0.001), resulting in a total of 4,047,158 SNPs available for all 88 samples.

We performed a genome-wide association scan (GWAS) of the cellular inhibition of proliferation by 1,25D (I_{max}) using a likelihood ratio test correcting for genome-wide proportions of African ancestry to control for spurious associations due to population structure. Genome-wide African ancestry proportions in each donor were estimated using STRUCTURE which uses multi-locus genotype data to investigate the genetic structure of populations [183]. Prior to the GWAS analysis, I_{max} was corrected for all covariates including age, gender, and cell type proportions.

Mapping variation in transcriptional response

We performed a genome-wide test for association between $\log_2$ fold change at every gene and SNPs within 100kb of the transcriptional start site of each gene. Transcriptional response profile data was not collected for 3 out of the 88 donors. For the 85 donors, the total number of genome-wide SNPs available for eQTL mapping that passed the filters described earlier was 4,100,242. eQTL mapping was performed using Matrix eQTL software, which performs a linear regression test for association between each SNP and each transcript, modeling the additive linear genotype effect on transcriptional response [184]. FDRs were calculated according to the Benjamini and Hochberg method [185]. We also corrected for genome-wide African ancestry proportions in this analysis.

As a complementary approach, we applied a Bayesian statistical framework that identifies different genotype-treatment interaction patterns, using the statistical software BRIdGE [161]. We mapped interaction eQTLs within 100kb of expressed genes, modeling four conditions through which SNPs could interact with transcriptional response phenotype under the two treatment conditions (1,25D and control): (i) Control-only model,

where genotype is associated with transcript levels in control-treated aliquots, but not in 1,25D-treated aliquots, (ii) 1,25D-only model, where genotype is associated with transcript levels in 1,25D-treated aliquots, but not in EtOH-treated aliquots, (iii) General interaction model, where genotype is associated with transcript levels in both conditions, but with different effects in each condition, and (iv) No interaction model, where genotype is associated with transcript levels in both conditions, with equal effect in each condition (baseline eQTLs). Using a hierarchical model, information across SNPs in each gene region and across genes was combined, and a posterior probability for each gene that it follows each of the models, and that it is affected by a SNP that follows that model, was calculated.

Identifying eQTLs within regulatory regions

We reanalyzed published data sets of VDR ChIP-seq obtained in THP-1 monocytic cell lines treated with 1,25D and LPS or 1,25D alone [149], and FAIRE-seq performed in THP-1 cells treated with 1,25D [150]. First, we aligned sequence reads to the human reference (GRCh37) using BWA backtrack 0.7.5. Second, we kept only sequence reads with phred-scaled mapping quality ≥ 30 using samtools v1.1 [152]. Third, PCR duplicate were removed with picard tool v 1.130 (http://broadinstitute.github.io/picard/). For the ChIP-seq data sets, we confirmed the quality of data sets by strand cross-correlation (SCC) analysis [153] implemented in the R script "run_spp_nodups.R" packaged in phantompeakqualtools (https://code.google.com/p/phantompeakqualtools/). Statistically significant peaks were identified using MACS version 2 [154] with the following essential command line arguments: macs2 callpeak --bw X -g hs --qvalue=0.05 -m 5 50, where X is a

length of the bandwidth that was defined as a fragment length calculated by SCC for the ChIP-seq data or as 200 bp for the FAIRE-seq data reported in Seuter *et al*. (2012).

Out of the 4,100,242 SNPs available for eQTL mapping, we identified subsets of these SNPs that were within ChIP-seq and FAIRE-seq peaks. We then used these subsets of SNPs to map response eQTLs using Matrix eQTL as described in the previous section.

Overlap between cellular and transcriptional response phenotypes

To identify genes whose transcriptional response to 1,25D may play a role in the inhibition of cell proliferation, we performed linear regression to test the association across individuals between the cellular response phenotype (I_{max}), and log-fold change response (1,25D-treated over vehicle-treated expression), and we estimated FDR using the q value function in R. We also applied a Bayesian method with the program Sherlock [186] to predict putative causal genes associated with I_{max}. This method predicts causal genes by identifying SNPs in these genes that are associated both with gene expression in *cis* and *trans*, and with the trait of interest, in our case, I_{max}. We used the results from the response *cis*-eQTL mapping and the GWAS of I_{max} to perform this analysis, setting the prior for association of each SNP with gene expression in *cis*, as well as association of each SNP with I_{max}, to 0.01. We chose this high prior due to the fact that we were examining transcriptional and cellular response phenotypes in primary cells obtained from the same individuals. The statistical significance of the Bayes factor for each gene was indicated by the corresponding p-values, which were calculated by permutation of the GWAS data, as detailed by He *et al.* (2013).

4.4: Results

Mapping variation in inhibition of cellular proliferation by 1,25D

To characterize inter-individual variation in cellular response to 1,25D, we measured cellular proliferation in PBMCs, which had been stimulated for 48 hours with PHA in the presence of either 1,25D or its vehicle (EtOH) as a control. I_{max} was calculated as the proportion of proliferation in 1,25D treated cells relative to proliferation in vehicle-treated cells. Using a simple linear regression, we measured the association between each donor's age, gender, time of collection, batch, serum 25D and cortisol levels, and found no significant correlations between these covariates and I_{max}. We also found no significant correlations between cell type proportions and I_{max}. However, to avoid any potential sources of confounding, we corrected I_{max} for all of these covariates before further downstream analyses.

To control for spurious associations potentially caused by population structure, we corrected for the proportion of genome-wide African ancestry in each donor, estimated using the program STRUCTURE. The median proportion of African ancestry in our donors was 81.4%, with an interquartile range of 14.7%. There were no significant correlations between I_{max}, or the other covariates, and proportion of African ancestry. However, there was a negative correlation between the genome-wide proportion of African ancestry and serum 25D levels ($p = 0.035$, $\beta = -0.034$) (**Supplementary Figure 4.2**), which suggests a genetic contribution to the higher prevalence of vitamin D insufficiency observed in African Americans [42]. The average serum 25D level in our African American donors was 20.81nM

with a standard deviation of 10.39nM, which is a level considered to be at risk for deficiency according to the Institute of Medicine definitions (less than 30nM) [187].

To investigate the genetic bases of variation in I_{max}, we carried out a genome-wide association scan for a total of 4,047,158 SNPs and identified genome-wide significant SNPs in chromosomes 5 and 18 (**Figures 4.1A and B**). The top signal of association was an intergenic SNP in chromosome 18 (rs1893662, $p = 2.32 \times 10^{-8}$) (**Figures 4.1A and C, Supplementary Table 4.1**). The A allele was associated with increased inhibition of proliferation (**Figure 4.1E**), and had a lower frequency in populations of African ancestry compared to European and Asian populations (allele frequency: 0.325, 0.811, and 0.648 respectively) (**Supplementary Figure 4.3A**). The next strongest signal of association was an intergenic SNP in chromosome 5 (rs6451692, $p = 2.55 \times 10^{-8}$) (**Figure 4.1A and D, Table S1**). The C allele was associated with increased inhibition of proliferation (**Figure 4.1F**), and had a higher frequency in populations of African ancestry compared to European and Asian populations (allele frequency: 0.839, 0.565, and 0.198 respectively) (**Supplementary Figure 4.3B**). The closest gene to this SNP is *CCL28*, which encodes a chemokine that recruits T cells, eosinophils, and B cells to mucosal sites; other genes within 100 kb of this SNP are two uncharacterized open reading frames (*C5orf28* and *C5orf35*) and *PAIP1*, which plays a role in stimulating translation initiation. Interestingly, we observed a marginal association between rs6451692 C allele and transcriptional response of *PAIP1* to 1,25D ($p = 0.02$, beta = -0.39) (**Supplementary Table 4.2**). In addition, this SNP lies less than 1 kb away from H3K4me1 enhancer-associated chromatin marks, DNase I hypersensitive sites and binding events for transcription factors such as TCF7L2, GATA3 and CEBPB in seven cell lines from the ENCODE project, including lymphoblastoid cell lines

[188] (**Supplementary Figure 4.4**). These chromatin marks highlight the potential regulatory activity of rs6451692 on transcriptional activity in immune cells.

To determine the proportion of variation of I_{max} explained by the top two SNPs in chromosomes 18 and 5, we examined the correlation coefficient from the linear model measuring the association between the top two associated SNPs and I_{max}. These two SNPs had a large effect on I_{max}, where rs1893662 explained 29.94% of the phenotypic variation in our samples, while rs6451692 explained 29.8% of the phenotypic variation in our samples. These top two SNPs explained ~45% of the variation in I_{max}.

Figure 4.1: **GWAS of inhibition of cellular proliferation by 1,25D (I_{max}). (A)** Manhattan plot of -$\log_{10}$ p-values of association of genome-wide variants with I_{max}. **(B)** Quantile-quantile (QQ) plot of distribution of observed -$\log_{10}$ p-values on the y-axis, versus the expected -$\log_{10}$ p-values on the x-axis. LocusZoom plots of the I_{max} GWAS associated regions in (**C**) chromosome 18 around rs1893662, and (**D**) chromosome 5 around rs6451692 (400kb windows, using 1000 genomes African populations as a reference). (**E**) Boxplots of I_{max} relative to genotypes of rs1893662 and rs6451692. I_{max} was corrected for age, gender, time of blood collection, batch, serum 25D levels, serum cortisol levels, and cell type proportions.

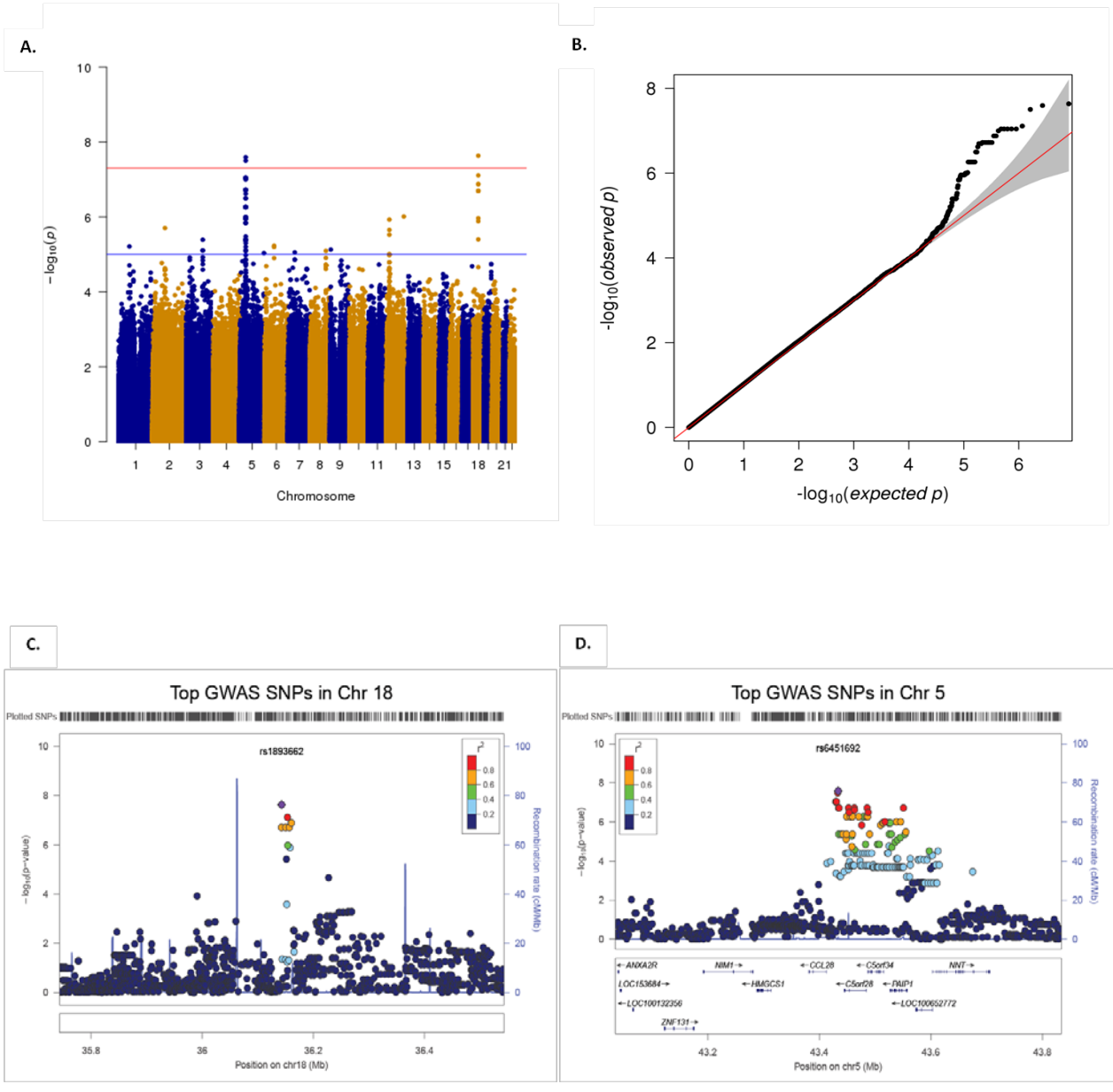
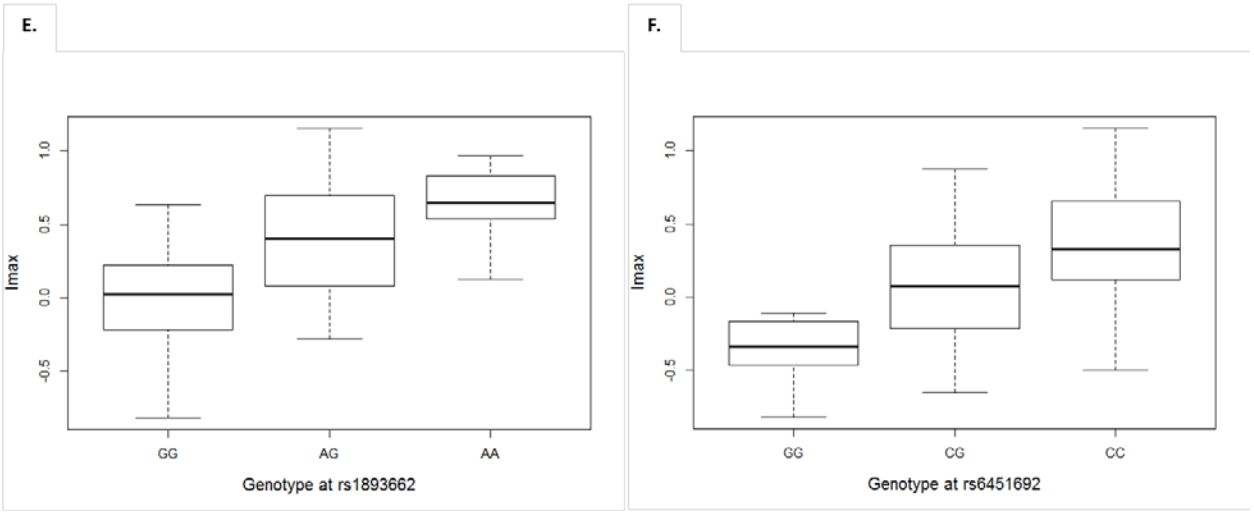


Figure 4.1 – continued.



Mapping variation in transcriptional response to 1,25D

We measured the expression of 11,897 genes in PBMCs from 85 donors treated with 100nM 1,25D and vehicle for 6 hours. We identified 720 genes differentially expressed (DE) in response to 1,25D at a FDR<0.01. Biological pathways significantly enriched among these genes included immune response pathways such as TREM1 signaling ($p = 4.0 \times 10^{-7}$, $FDR = 2 \times 10^{-4}$), Granulocyte differentiation and Diapedesis ($p = 2.0 \times 10^{-5}$ $FDR = 4 \times 10^{-3}$), and T Helper Cell Differentiation ($p = 6.0 \times 10^{-4}$, $FDR = 6.5 \times 10^{-2}$) (**Supplementary Table 4.3**), supporting the important role of 1,25D as an immunomodulator. In addition, there was an enrichment of the VDR/RXR activation pathway ($p = 7.0 \times 10^{-4}$, $FDR = 6.5 \times 10^{-2}$), including genes such as *CD14,* which encodes a monocyte surface antigen mediating innate immune response to bacterial lipopolysaccharide (LPS), *CAMP* which encodes an antimicrobial peptide, and *CYP24A1* which encodes the enzyme that initiates the degradation of 1,25D. A previous study characterizing patterns of transcriptional response to 1,25D and LPS in primary monocytes also found an overlapping list of immune response pathways identified in this study enriched among genes that were significantly down-regulated by 1,25D [189].

In order to identify polymorphisms that influence the transcriptional response to 1,25D, we tested the association between $\log_2$ fold change in transcript levels at each expressed gene and SNPs within 100kb of each gene using Matrix eQTL. Because DE genes tend to be those with consistent differences in transcript levels across all individuals, they may be biased against genes with common regulatory polymorphisms. For this reason, we did not limit our mapping analyses to the DE genes. We identified response *cis*-eQTLs for 8 genes at a FDR<0.10, with the most significant response eQTLs including the

transcriptional factor ets variant 3-line (*ETV3L*), and EH-domain containing 4 (*EHD4*), which plays a role in early endosomal transport (**Table 4.1A, Supplementary Figure 4.5**). Mapping $\log_2$ fold change does not distinguish among the types of genotype-by-treatment interactions that influence transcriptional response. To do that, we applied a Bayesian statistical framework using the BRIdGE software, which compares different interaction models to each other and to a null model of no genotypic effect in both treatment conditions. We identified 4 genes with high confidence interactions (posterior probability of interaction > 0.7) between 1,25D treatment and SNP genotype; all these interaction eQTLs followed a 1,25D-only model, namely genotype has an effect on transcript levels in the 1,25D-treated aliquot but not in the control-treated one (**Table 4.1B**). These interaction eQTLs included the top 2 most significant response eQTLs that had been identified by mapping $\log_2$ fold change: *ETV3L* and *EHD4*. In addition, we identified interaction eQTLs in leucine rich repeat containing 25 (*LRRC25*), which is involved in activation of various immune cell types, and the transcriptional regulator unkempt family zinc finger (*UNK*).

Table 4.1: *cis*-eQTLs for transcriptional response to 1,25D.

(**A**) *cis*-eQTL mapping of log-fold change expression using Matrix eQTL

SNP	Gene	T-Statistic	P-value	FDR	Beta
rs74116976	*ETV3L*	6.77	1.73×10^{-09}	1.17×10^{-4}	0.84
rs11070354	*EHD4*	6.11	3.16×10^{-8}	1.28×10^{-3}	0.86
rs7311057	*PARPBP*	5.73	1.56×10^{-7}	2.43×10^{-2}	0.79
rs59937851	*ZNHIT1*	5.41	5.98×10^{-7}	1.52×10^{-2}	1.02
rs7178702	*SPESP1*	5.12	1.97×10^{-6}	1.30×10^{-2}	0.69
rs10282056	*COBL*	-4.78	7.50×10^{-6}	3.10×10^{-2}	-0.92
rs62014366	*VWA9*	4.74	8.66×10^{-6}	4.67×10^{-2}	0.94
rs7779605	*CPED1*	4.31	4.38×10^{-5}	7.63×10^{-2}	0.75

(**B**) Interaction *cis*-eQTL mapping using BRIdGE

Gene	SNP	Posterior probability for each interaction model			
		Control-only	1,25D-only	General interaction	No interaction
EHD4	rs1648856	0	0.994	0	0.001
LRRC25	rs3848646	0	0.965	0	0.027
UNK	rs8081606	0	0.803	0	0.049
ETV3L	rs6689823	0	0.723	0	0.277

To evaluate additional response *cis*-eQTLs found in VDR response elements, we identified 988 SNPs within VDR ChIP-seq peaks from a dataset of published THP-1 monocytic cell lines treated with 1,25D [149], and mapped response eQTLs using this subset of SNPs. At a distance of 1Mb, we identified statistically significant response eQTLs (FDR < 0.10) in two genes: FERM Domain Containing 6 (*FRMD6*), a key activator of the Hippo kinase pathway with important roles in regulating cell proliferation and apoptosis [190], and the undefined *KIAA1211* (**Figure 4.2A and B**). In addition, we identified 17,417 SNPs within open chromatin regions, identified by FAIRE-seq from a published dataset of THP-1 monocytic cell lines treated with 1,25D [150]. Within this subset, we identified statistically significant response eQTLs (FDR < 0.10) in *ETV3L*, *EHD4* and *ZNHIT1* (**Supplementary Table 4.4**). These eQTLs were in strong linkage disequilibrium (LD) with the response eQTLs we had identified for the same genes (r^2 = 0.93, 0.69 and 0.95 for *ETV3L*, *EHD4* and *ZNHIT1*, respectively), raising the possibility that these response eQTLs are due to variants affecting open chromatin conformation.

Figure 4.2: **Associations between SNPs in vitamin D receptor (VDR) binding sites and transcriptional response.** (**A**) Boxplots showing the effect of genotype on $\log_2$ fold change of *FRMD6* and *KIAA1211* transcript levels, with genotypes of associated SNPs coded as the number of copies of the alternative allele. (**B**) Location of SNPs associated with transcription response of *FRMD6* (rs3783273, top panel) and *KIAA1211* (rs7698085, bottom panel) within VDR binding sites, indicated by the gray horizontal arrows. The SNP locations are indicated by the vertical orange arrows. VDR binding site information was obtained from a published ChIP-seq dataset from THP-1 monocytic cells treated with 1,25D and bacterial lipopolysaccharide (LPS).

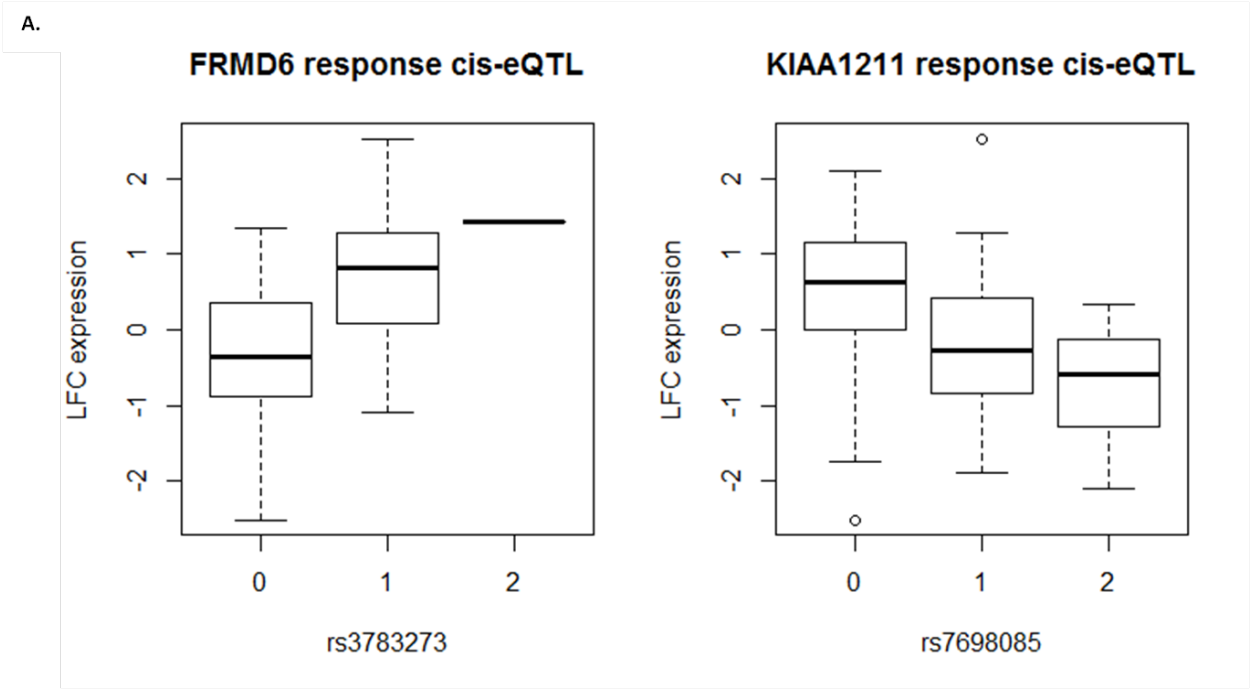
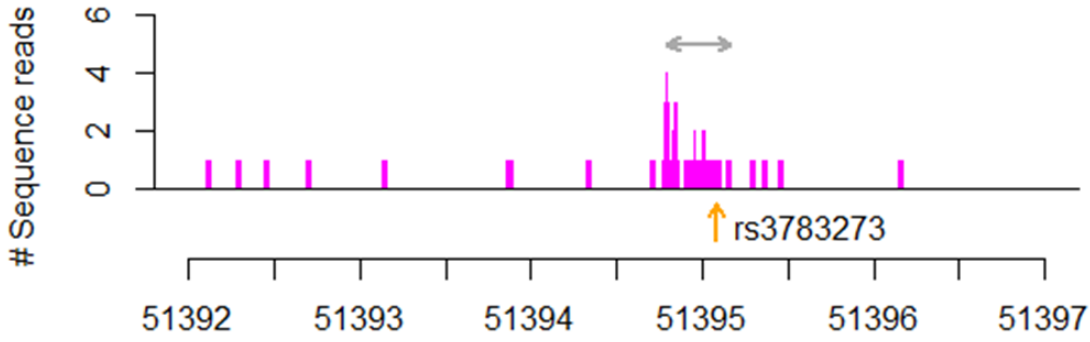
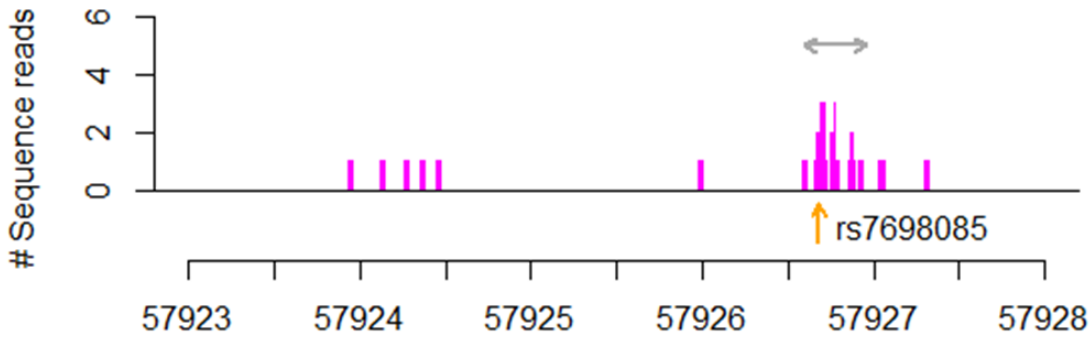


Figure 4.2 – continued.

B.



C.



Combined analysis of cellular and transcriptional response phenotypes

We examined the relationship between the two 1,25D response phenotypes: transcriptional response and the inhibition of cellular proliferation. To evaluate whether the SNPs associated with inhibition of cellular proliferation exerted their effects through regulation of transcriptional response, we first examined associations between the two most significant I_{max} GWAS SNPs and $\log_2$ fold change expression at all 11,897 genes expressed in the PBMCs. At a FDR < 0.10, we found no statistically significant associations. We then focused on the subset of genes where $\log_2$ fold change in expression was associated with I_{max}, reasoning that these genes are more likely to share genetic variation influencing both transcriptional response and inhibition of cell proliferation. Using a linear regression approach, we identified 16 associated genes at an FDR < 0.2 (**Supplementary Table 4.5**). When we considered only these genes, we found significant associations between two I_{max}-associated genes (*PCSK6* and *RASL11A*) and the top GWAS SNP in chromosome 18, rs1893662, and one I_{max}-associated gene (*KNCN*) with the second GWAS SNP in chromosome 5, rs6451692 (**Table 4.2**), at a Bonferroni-corrected $p < 3.125 \times 10^{-3}$. Both *PCSK6* and *KNCN* are involved in vesicular trafficking and secretory pathways, highlighting potential molecular mechanisms involved in inhibition of proliferation by vitamin D.

Table 4.2. Association between top I_{max} GWAS SNPs and transcriptional response

	rs1893662		rs6451692	
Gene Name	**Beta**	**P-value**	**Beta**	**P-value**
PCSK6	-2.09	**2.1×10^{-3}**	-0.78	0.26
SMARCD3	2.01	**2.1×10^{-3}**	0.93	0.16
RASL11A	2.08	**2.3×10^{-3}**	1.17	0.09
KNCN	-1.24	3.21×10^{-2}	-1.85	**9.95×10^{-4}**

We further predicted putative causal genes associated with I_{max} based on a Bayesian approach implemented in the program, Sherlock, using our response eQTL and GWAS of I_{max} data. At $p < 10^{-4}$ (FDR = 0.3), we identified three putative I_{max}-associated genes, including the translation initiation gene *PAIP1*, a transcriptional repressor gene *ZNF649*, and a golgin family gene *GORAB* (**Supplementary Table 4.6**). Interestingly, the top I_{max}-associated SNP in chromosome 5, rs6451692, was identified as being associated with transcriptional response of *PAIP1* using this method, which suggests that this SNP influences the inhibition of cell proliferation through a transcriptional mechanism in PBMCs.

4.5: Discussion

While the inter-individual variation in the circulating inactive form of vitamin D, 25D, has been well documented, little is known about the inter-individual variation in immune response to the active 1,25D. In this study, we identified several variants underlying variation in response to 1,25D both at the cellular and transcriptional level using primary peripheral blood mononuclear cells from a cohort of healthy individuals of African-American ancestry. These variants highlight genes with an important role in mediating the immunomodulatory effects of 1,25D, thereby providing a genetic basis for inter-individual variation in those aspects of the immune response influenced by vitamin D.

Intergenic SNPs in chromosome 5 that were significantly associated with inhibition of cellular proliferation by 1,25D are located close to several genes such as *CCL28*, which encodes a chemokine that recruits T cells, eosinophils, and B cells to mucosal sites [191-

193], and *PAIP1* which encodes a protein that interacts with poly(A)-binding protein and with the eIF4A cap-binding complex, stimulating translation initiation [194]. Interestingly, we found a marginal association between rs6451692 and down-regulation of *PAIP1*, raising the possibility that this polymorphism influences the inhibitory effects of 1,25D on immune cell proliferation by regulating the transcriptional response of a translation initiation gene.

We observed several regulatory marks near rs6451692 in seven cell lines from the ENCODE project, including an enrichment of H3K4me1 histone mark, which is associated with enhancers. There was also an abundance of transcription factor binding events in this region, where rs6451692 overlaps a TCF7L2 binding site. TCF7L2 is a member of the high mobility group DNA binding protein family of transcription factors which has been implicated in type 2 diabetes risk [195-197]. Other transcription factors with binding sites in the region include RXRA, which binds to the VDR, forming a heterodimer which then regulates transcription of vitamin D-responsive genes, GATA3 which has important roles in T cell development [198, 199], and CEBPB which plays an important role in regulating immune and inflammatory response genes [200-203]. The abundance of transcription factor binding events in this region suggests that the regulatory activity of rs6451692 on the surrounding genes could involve enhancer activity. Further functional validation assays specifically in PBMCs treated with vitamin D are needed to elucidate the regulatory mechanisms of this I_{max} GWAS interval.

In addition, from the Genotype-Tissue Expression (GTEx) project catalogue [204], we observed that rs6451692 is associated with variation in transcript levels of surrounding genes in multiple tissues. The C allele is associated with decreased expression of *CCL28* in the pancreas, decreased expression of *NNT* in skeletal muscle, and decreased

expression of the novel antisense long non-coding RNA *RP11-159F24.5* in multiple tissues such as subcutaneous adipose, tibial nerve, testis, thyroid and skin, suggesting that this variant influences the regulation of several genes in that genomic region. *RP11-159F24.5* was not covered by probes in our expression microarrays, therefore we cannot determine if rs6451692 has effects on the expression of this gene in PBMCs.

Enrichment of immune response pathways such as TREM1 signaling, which enhances innate immune responses to microbial infections and activates pro-inflammatory responses [205], and T helper cell differentiation among the genes that respond transcriptionally to 1,25D, underscores the important immunomodulatory role played by 1,25D [37, 38, 77]. This is consistent with the results of a previous study from our group investigating the transcriptional response to 1,25D and to bacterial lipopolysaccharide (LPS) in primary monocytes, where we also found an enrichment of immune response pathways, particularly among genes that were down-regulated by 1,25D and up-regulated by LPS [189]. This highlights the important immunomodulatory role played by 1,25D across cells in both the innate and adaptive immune system. In addition, among the genes that were up-regulated by 1,25D in monocytes, metabolic and translation initiation pathways were significantly enriched, consistent with previous reports in dendritic cells [41]. These pathways were not significantly enriched amongst the DE genes in PBMCs in this study (**Supplementary Table 4.3**), which could indicate that 1,25D regulates pathways involving metabolic reprogramming and translation particularly in innate immune response. It was however interesting to note the marginal association between one of the top I_{max} SNPs, rs6451692, and transcriptional response of *PAIP1*, a translation initiation gene.

Several studies have mapped genome-wide VDR binding sites in different immune cell lines [149, 206, 207]. Interestingly, one study examined VDR binding sites in primary $CD4^+$ T cells from nine individuals with varying 25D levels and reported a correlation between 25D levels and number of VDR binding sites [208], directly supporting the notion that vitamin D status affects the response to vitamin D. In addition, genome-wide maps of VDR binding sites allow identification of genetic variants within VDR binding sites that in turn may influence variation in the transcriptional response to vitamin D. Interestingly, one such study reported that many risk variants for autoimmune diseases detected in genome-wide association studies fall within VDR binding sites [70], suggesting that disease risk is influenced not only by inter-individual variation in 25D levels, but also by variation in the response to vitamin D. To build on these studies, we focused on mapping variants that regulate genome-wide transcriptional response to 1,25D in primary PBMCs. The *cis*-response eQTLs identified in this study highlighted several genes that could play an important role in mediating the effects of 1,25D in the immune response. Genes identified using both the linear regression and Bayesian eQTL mapping approaches included *ETV3L*, which is a transcriptional regulator that has been reported to play a role in inhibiting proliferation of neural progenitor cells [209], and *EHD4*, which plays a role in controlling early endosomal trafficking [210, 211]. Furthermore, we identified statistically significant response eQTLs in regions of open chromatin, marked by FAIRE-seq peaks, in *ETV3L*, *EHD4*, and *ZNHIT1* - a gene that is implicated in regulating the transcriptional activity of the orphan nuclear receptor Rev-erbbeta [212]. Interestingly, both *ETV3L* and *ZNHIT1* are transcriptional regulators, raising the possibility that these loci could play a role in modulating transcriptional response of other genes to 1,25D in immune cells.

We then identified variants within VDR binding sites that regulate transcriptional response possibly by altering the structure or accessibility of the VDR binding site. We did this by combining our *cis*-response eQTL data with a published VDR ChIP-seq dataset from a monocytic cell line [149]. We identified a response eQTL within a VDR binding site in *FRMD6*, which is part of the conserved Hippo pathway playing a critical role in controlling organ size by regulating both cell proliferation and apoptosis [213, 214]. *FRMD6* has been linked to various complex diseases such as asthma, Alzheimer's disease, and lung cancer [214-216], where it is thought to have tumor suppressor properties. The T allele of rs3783273, which is associated with increased *FRMD6* expression (**Figure 4.2**), could alter the binding properties of the VDR to its receptor elements in *FRMD6* and could affect the transcriptional response of this gene to 1,25D. Given its putative tumor suppressor properties, *FRMD6* may play a crucial role in mediating the role of 1,25D in inhibiting proliferation of immune cells.

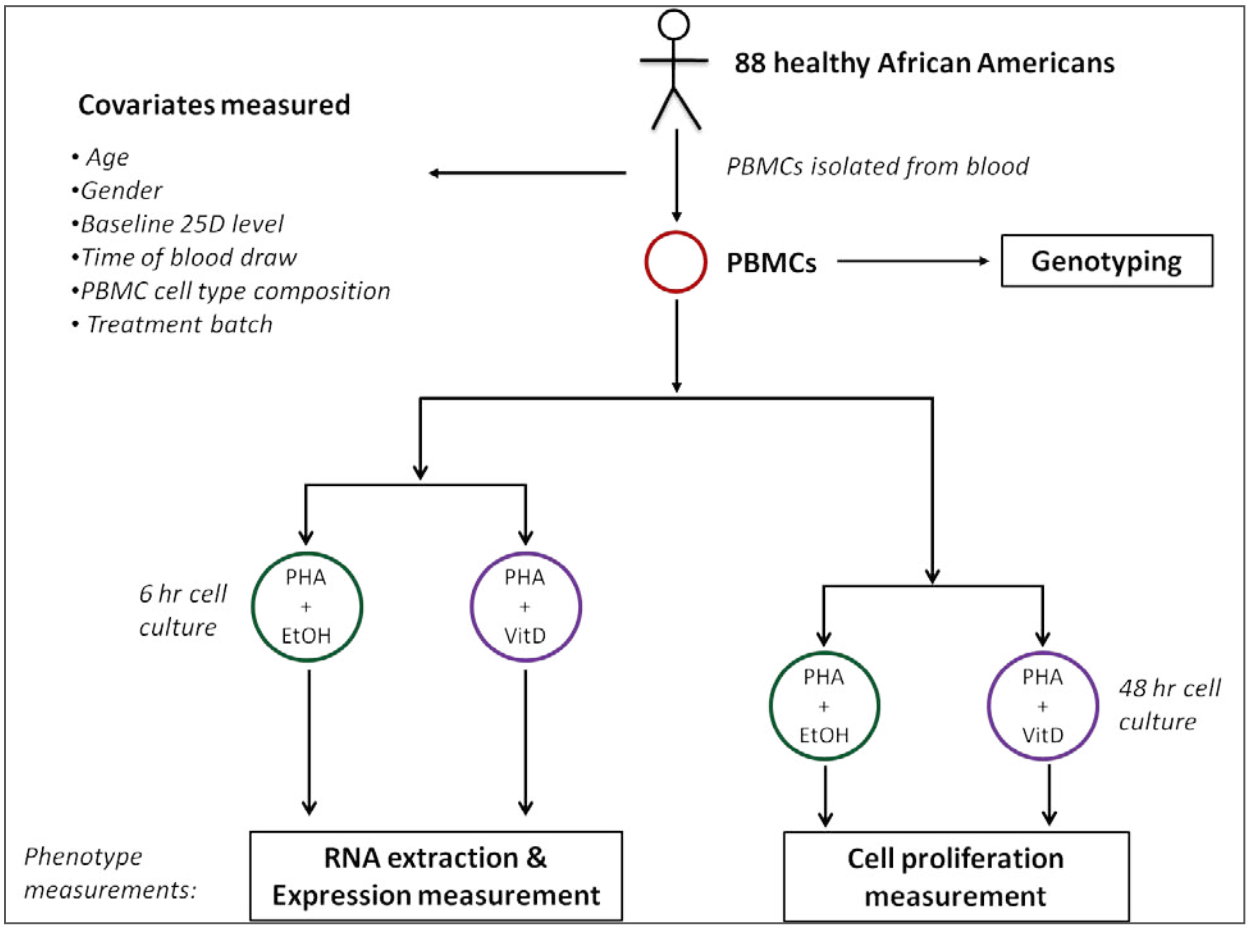
Using both simple linear regression analysis and a Bayesian approach, we combined the information from response *cis*-eQTL mapping and the GWAS of I_{max} to identify candidate genes mediating the inhibitory effects of cellular proliferation by 1,25D. Genes such as *PAIP1*, *ZNF649* and *GORAB* contained I_{max}-associated SNPs that also regulated transcriptional response of these genes in *cis*. While *PAIP1* encodes a protein that is involved in initiating translation, *ZNF649* encodes a transcriptional repressor that inhibits transcription factor complexes such as AP-1 which is involved in cellular proliferation and survival [217-219], and *GORAB* encodes a golgin family member with roles in the intracellular membrane trafficking and the secretory pathways of the Golgi apparatus [220, 221]. In addition, we identified *trans* effects of the top GWAS SNPs on transcriptional

response of genes such as *PCSK6* and *KNCN*, which both have roles in vesicular trafficking and secretory pathways, highlighting potential molecular mechanisms involved in the anti-proliferative activity of 1,25D. Increased *PCSK6* expression has been previously implicated in risk for rheumatoid arthritis [222]. Interestingly, knockdown of *PCSK6* by RNA interference significantly decreased proliferation, invasion, and migration of cultured rheumatoid arthritis synovial fibroblasts. It is plausible that the top I_{max}-associated SNP, rs1893662, regulates the anti-proliferative activity of 1,25D by regulating *PCSK6* transcription in immune cells. The potential mechanisms through which these putative I_{max}-associated candidate genes could mediate the inhibition of proliferation of immune cells by 1,25D should be further studied.

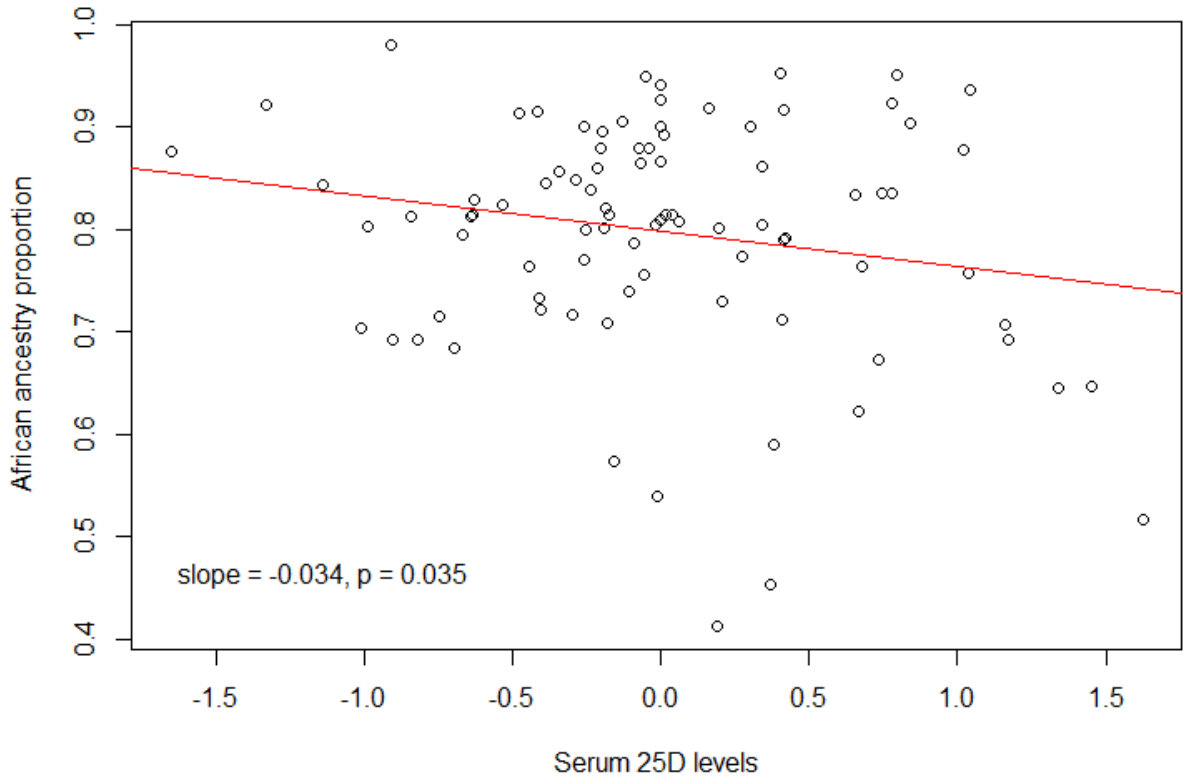
In summary, mapping response to 1,25D at both the cellular and transcriptional level in immune cells enabled identification of variants which may influence inter-individual variation in response to 1,25D, and identification of genes with potentially crucial roles in mediating the immunomodulatory role of 1,25D. Characterizing these genetic mediators of 1,25D activity in the immune system could inform additional therapeutic targets and markers for immune-related diseases in future randomized VD supplementation trials.

4.6: Appendix: Supplementary Material

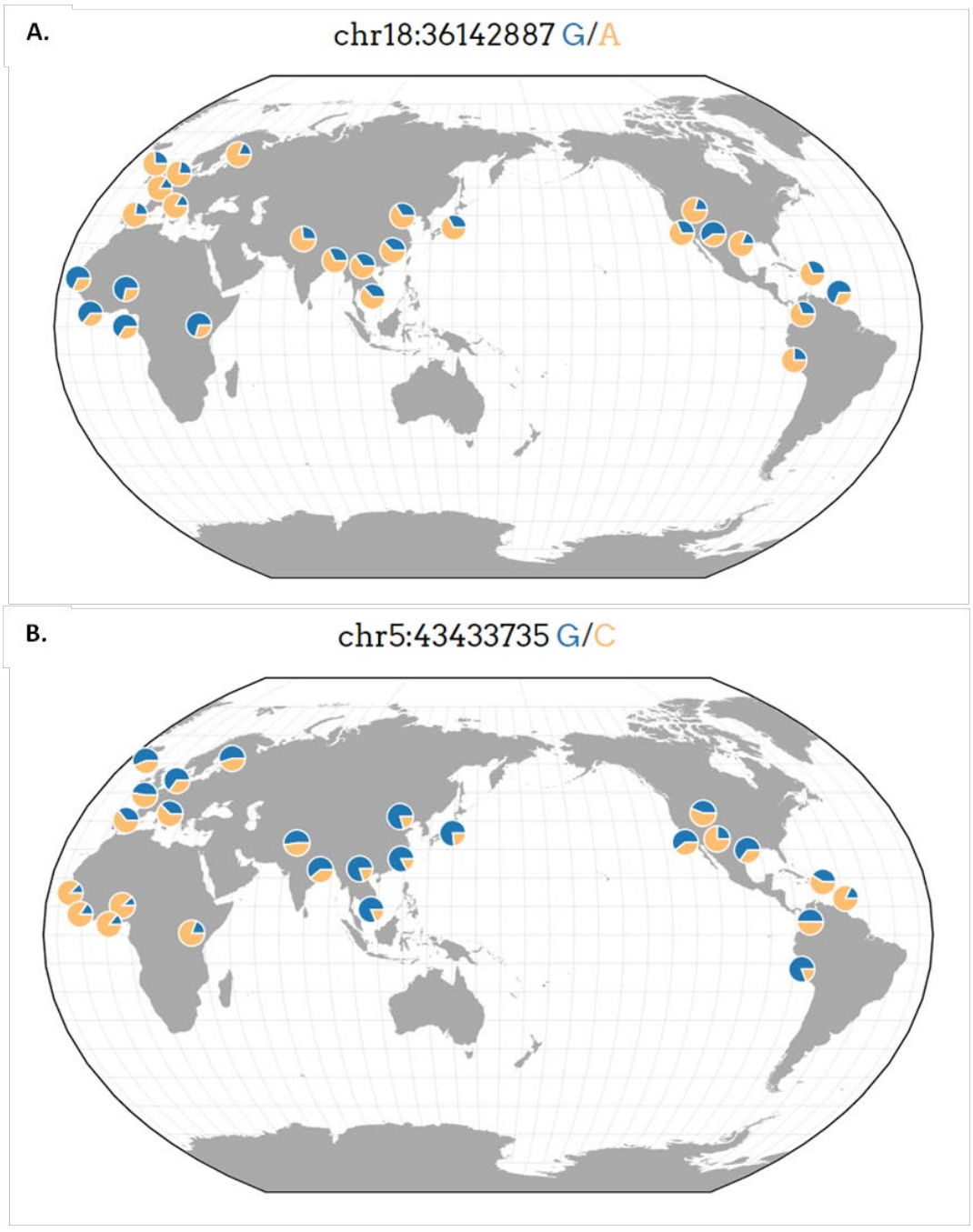
Supplementary Figure 4.1: Experimental Design. Peripheral blood mononuclear cells (PBMCs) were obtained from 88 healthy African American donors. PBMCs were cultured for 6 hours with phytohemagglutinin (PHA) and either vehicle (EtOH) or 1,25-dihydroxyvitamin D3 (1,25D), and RNA was extracted for gene expression measurements. PBMCs from the same samples were also cultured for 48 hours with PHA and either vehicle or 1,25D for cell proliferation measurements. DNA was also extracted from PBMCs for genotyping.



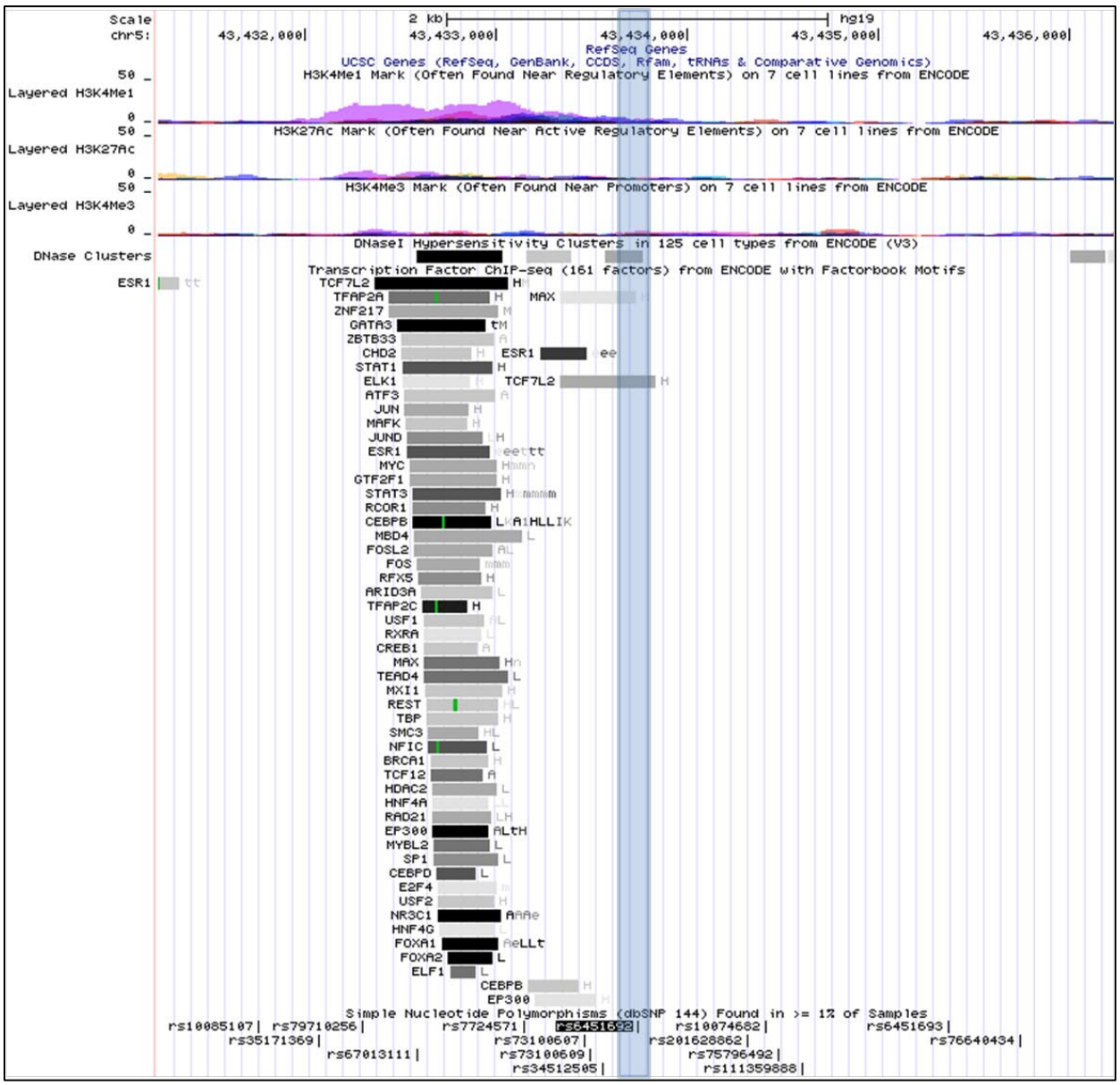
Supplementary Figure 4.2: Correlation between serum 25D levels and global ancestry. Serum levels of 25D are negatively correlated with global proportions of African ancestry. Serum 25D levels were corrected for age and batch effects.



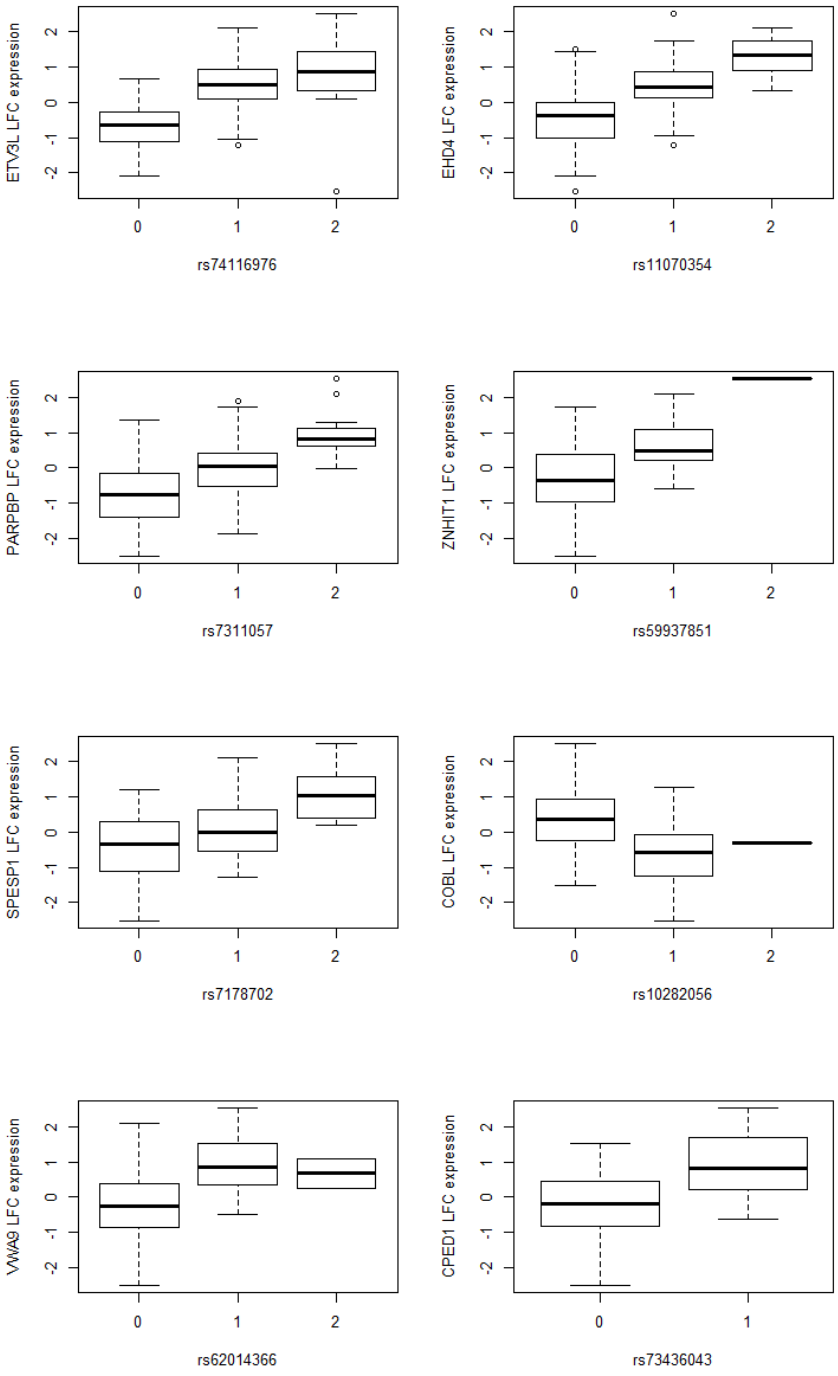
Supplementary Figure 4.3: Allele frequency distribution of top SNPs. Large inter-population allele frequency differentiation was observed in the top GWAS SNPs, rs1893662 (**A**) and rs6451692 (**B**). Image obtained from the Geography of Genetic Variants (GGV) browser [223].



Supplementary Figure 4.4: Regulatory marks near rs6451692. The image illustrates a magnified view of the GWAS interval in chromosome 5, with the location of rs6451692 highlighted by the blue rectangle, and nearby enhancer marks (H3K4me1), DNase I hypersensitive sites, and transcription factor binding sites obtained from seven cell lines from the ENCODE project [188].



Supplementary Figure 4.5: Mapping log-fold change response *cis*-eQTLs. Response eQTLs were identified using a linear model in Matrix eQTL [184], where the additive effect of genotype on $\log_2$ fold change in transcript levels in response to 1,25D was measured. The boxplots show the results of 8 response *cis*-eQTLs identified at an FDR < 10%, with genotype coded as the number of copies of the minor allele. All SNPs are within 100kb of the transcriptional start site (TSS) of their respective genes.



Supplementary Table 4.1: The top SNPs identified in the GWAS of I_{max}. The SNPs shown have p-values < 10^{-8}, which corresponds to a FDR of 0.036.

SNP	Chr	MAF	Nucleotide position	Beta	P
rs1893662	18	0.318	36142887	0.355	2.32×10^{-8}
rs6451692	5	0.778	43433735	0.361	2.55×10^{-8}
rs7724571	5	0.761	43433143	0.355	3.15×10^{-8}
rs4800030	18	0.318	36153493	0.352	7.78×10^{-8}
rs7707976	5	0.773	43429523	0.347	9.10×10^{-8}
rs7708443	5	0.773	43429761	0.347	9.10×10^{-8}
rs7708369	5	0.773	43429964	0.347	9.10×10^{-8}
rs750582	5	0.773	43430273	0.347	9.10×10^{-8}
rs750584	5	0.773	43430406	0.347	9.10×10^{-8}

Supplementary Table 4.2: Association between top I_{max}-associated SNPs in chromosome 5, and transcription response of nearby genes (within 100kb).

SNP	GeneName	Response eQTL p-value	Beta	I_{max} GWAS P-value	SNP-Gene distance (bp)
rs6451692	*PAIP1*	2.0×10^{-2}	-0.39	2.6×10^{-8}	92,634
rs10941640	*PAIP1*	3.5×10^{-2}	0.37	1.1×10^{-5}	0
rs6866325	*PAIP1*	1.2×10^{-2}	-0.39	2.0×10^{-5}	0
rs7708072	*C5orf34*	4.2×10^{-2}	0.37	3.1×10^{-5}	81,827
chr5:43598333:I	*C5orf34*	4.4×10^{-2}	0.37	6.5×10^{-5}	83,060

Supplementary Table 4.3: Gene set enrichment analysis of significantly differentially expressed (DE) genes at FDR < 0.01. Enrichment analyses were conducted using Ingenuity Pathway Analysis (IPA) software. The enriched pathways shown are at a p-value threshold of 0.05. The top 8 pathways were statistically significant at a FDR < 0.10. B-H p-value* = Benjamini-Hochberg multiple testing corrected p-value.

Ingenuity Canonical Pathways	p-value	B-H p-value	Genes
TREM1 Signaling	4×10^{-7}	2×10^{-4}	*TREM1,ICAM1,NLRP3,TLR8,CIITA,CCL3,TLR4,NOD2,PLCG2,TLR6,CASP1,CD86,IL1B,CCL7*
Granulocyte Adhesion and Diapedesis	2×10^{-5}	4×10^{-3}	*FPR3,ICAM1,C5AR1,FPR2,CCL22,CXCL5,MMP25,CCL3,CXCL6,FPR1,CLDN23,ITGAM,CCL8,CCL3L1,CCL3L3,IL1B,CXCL1,TNFRSF1B,CCL7*
T Helper Cell Differentiation	6×10^{-4}	6.5×10^{-2}	*ICOS,HLA-DMB,IL10RB,IL10RA,IFNGR2,CD86,IL2RA,TNFRSF1B,ICOSLG/LOC102723996*
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses	7×10^{-4}	6.5×10^{-2}	*PTX3,NLRP3,C5AR1,TLR8,TLR4,NOD2,PRKCD,PLCG2,TLR6,CASP1,OSM,IL1B,RIPK2*
VDR/RXR Activation	7×10^{-4}	6.5×10^{-2}	*SERPINB1,CAMP,CYP24A1,GADD45A,PRKCD,CD14,NCOR2,CEBPB,THBD,RXRA*
Role of IL-17A in Arthritis	1×10^{-3}	7.4×10^{-2}	*CXCL1,MAPK13,CXCL5,PTGS2,RPS6KA1,MAPKAPK2,CXCL6,CCL7*
Role of IL-17A in Psoriasis	1×10^{-3}	7.4×10^{-2}	*S100A9,CXCL1,CXCL5,CXCL6*
Sulfate Activation for Sulfonation	1×10^{-3}	8.1×10^{-2}	*PAPSS1,PAPSS2*
Primary Immunodeficiency Signaling	3×10^{-3}	0.13	*BTK,LCK,ICOS,CIITA,ADA,CD3D*
Role of IL-17F in Allergic Inflammatory Airway Diseases	5×10^{-3}	0.20	*IL1B,CXCL1,CXCL5,RPS6KA1,CXCL6,CCL7*
LPS/IL-1 Mediated Inhibition of RXR Function	7×10^{-3}	0.23	*CPT1A,CHST7,PAPSS2,CHST15,TLR4,LY96,CAT,IL1B,XPO1,CD14,PPARGC1B,HS3ST1,SLC27A3,RXRA,TNFRSF1B,ACSL1*
MIF-mediated Glucocorticoid Regulation	7×10^{-3}	0.23	*TLR4,PLA2G4A,LY96,CD14,PTGS2*
LXR/RXR Activation	7×10^{-3}	0.23	*TLR4,LYZ,LY96,CD36,CD14,IL1B,NCOR2,PTGS2,TNFRSF1B,RXRA,CCL7*
Agranulocyte Adhesion and Diapedesis	7×10^{-3}	0.23	*ICAM1,C5AR1,CCL22,CXCL5,MMP25,CCL3,CXCL6,CLDN23,CCL8,CCL3L1,CCL3L3,IL1B,CXCL1,CCL7*
Uracil Degradation II (Reductive)	9×10^{-3}	0.23	*DPYSL2,UPB1*
Thymine Degradation	9×10^{-3}	0.23	*DPYSL2,UPB1*
Mitochondrial Dysfunction	1.2×10^{-2}	0.30	*COX7B,ATP5G1,UCP2,CPT1A,CAT,COX5A,TRAK1,NDUFAF2,UQCRC1,CYB5R3,NDUFAB1,UQCRB*
CMP-N-acetylneuraminate Biosynthesis I (Eukaryotes)	1.4×10^{-2}	0.33	*NAGK,CMAS*
Macropinocytosis Signaling	1.6×10^{-2}	0.35	*PRKCD,PLCG2,HGF,USP6NL,CD14,ITGB8,PDGFB*
Type I Diabetes Mellitus Signaling	1.6×10^{-2}	0.35	*HLA-DMB,IFNGR2,CD86,BID,IL1B,SOCS2,MAPK13,TNFRSF1B,CD3D*
MIF Regulation of Innate Immunity	1.8×10^{-2}	0.38	*TLR4,PLA2G4A,LY96,CD14,PTGS2*

Supplementary Table 4.3 – continued.

Purine Ribonucleosides Degradation to Ribose-1-phosphate	2×10^{-2}	0.40	*ADA,PGM2*
Toll-like Receptor Signaling	2.2×10^{-2}	0.42	*TLR4,LY96,TLR6,TLR8,CD14,IL1B,MAPK13*
Communication between Innate and Adaptive Immune Cells	2.4×10^{-2}	0.42	*TLR4,CCL3L3,TLR6,TLR8,CD86,IL1B,CCL3*
iNOS Signaling	2.5×10^{-2}	0.42	*TLR4,LY96,IFNGR2,CD14,MAPK13*
γ-linolenate Biosynthesis II (Animals)	2.6×10^{-2}	0.42	*SLC27A3,CYB5R3,ACSL1*
Mitochondrial L-carnitine Shuttle Pathway	2.6×10^{-2}	0.42	*CPT1A,SLC27A3,ACSL1*
Differential Regulation of Cytokine Production in Macrophages and T Helper Cells by IL-17A and IL-17F	3×10^{-2}	0.47	*IL1B,CXCL1,CCL3*
IL-6 Signaling	3.6×10^{-2}	0.50	*TNFAIP6,CYP19A1,CD14,IL1B,MAPK13,CEBPB,TNFRSF1B,MAPKAPK2,MCL1*
p53 Signaling	3.6×10^{-2}	0.50	*RB1,GADD45B,GADD45A,THBS1,GNL3,HIF1A,DRAM1,PTEN*
iCOS-iCOSL Signaling in T Helper Cells	3.6×10^{-2}	0.50	*LCK,ICOS,HLA-DMB,IL2RA,VAV1,CD3D,ICOSLG/LOC102723996,PTEN*
Cardiolipin Biosynthesis II	3.9×10^{-2}	0.52	*PGS1*
Hepatic Cholestasis	4.4×10^{-2}	0.57	*TLR4,LY96,CYP27A1,PRKCD,ADCY3,CD14,OSM,IL1B,TNFRSF1B,RXRA,ADCY7*
IL-17A Signaling in Fibroblasts	4.5×10^{-2}	0.57	*MAPK13,CXCL5,CEBPB,CCL7*
Hepatic Fibrosis / Hepatic Stellate Cell Activation	4.9×10^{-2}	0.60	*TLR4,LY96,ICAM1,HGF,IL10RA,IFNGR2,CD14,IL1B,ECE1,TNFRSF1B,PDGFB,COL7A1*

Supplementary Table 4.4: Response *cis*-eQTLs found in open chromatin regions detected by FAIRE-seq. These response eQTLs are significant at a FDR < 0.10. The strength of the FAIRE-seq peaks is indicated by the Peak P-values and Peak FDR values.

SNP	Gene	T-Statistic	P-value	FDR	Beta	Peak P-value	Peak FDR
rs7520303	*ETV3L*	6.73	2.14×10^{-9}	6.13×10^{-5}	0.85	2.02×10^{-11}	1.51×10^{-19}
rs12913835	*EHD4*	5.21	1.36×10^{-6}	2×10^{-2}	0.72	2.22×10^{-5}	1.3×10^{-5}
rs6946706	*ZNHIT1*	4.94	4.07×10^{-6}	4×10^{-2}	0.87	1.76×10^{-6}	3.56×10^{-12}

Supplementary Table 4.5: Genes whose transcription responses are associated with inhibition of cellular proliferation by 1,25D at a FDR < 0.2.

Gene	Beta	P-value	FDR
ZNF571	-1.05	2.68×10^{-5}	0.17
FNTA	-1.16	4.26×10^{-5}	0.17
GALNT4	-0.82	9.74×10^{-5}	0.17
PYCRL	-1.48	1.04×10^{-4}	0.17
HARBI1	-1.62	1.16×10^{-4}	0.17
UQCRC2	0.96	1.25×10^{-4}	0.17
RASL11A	1.65	1.60×10^{-4}	0.17
NOTCH3	1.45	1.60×10^{-4}	0.17
ABCG1	2.01	1.65×10^{-4}	0.17
PPP2R1A	-1.11	1.90×10^{-4}	0.17
GEMIN7	-1.37	1.93×10^{-4}	0.17
SMARCD3	1.55	2.18×10^{-4}	0.17
SERPINA11	-2.50	2.20×10^{-4}	0.17
KNCN	-1.33	2.61×10^{-4}	0.19
PCSK6	-1.57	2.88×10^{-4}	0.19
PTGR2	1.20	2.97×10^{-4}	0.19

Supplementary Table 4.6: Genes associated with inhibition of cellular proliferation by 1,25D (I_{max}). These genes were detected using a Bayesian approach, implemented in the statistical program Sherlock [186]. The strength of the association between the genes and I_{max} is given by the overall $\log_{10}$ of Bayes factor (**LBF**).

Information on gene associated with I_{max}				**Information on SNP associated with transcriptional response of corresponding gene, and with I_{max}**				
Gene	**LBF**	**P-value**	**FDR**	**SNP**	**SNP Location**	**eQTL P-value**	**GWAS P-value**	**LBF of SNP**
PAIP1	2.51	5.59×10^{-6}	0.03	rs6451692	chr5: 43,433,735	2.01×10^{-2}	2.55×10^{-8}	2.51
ZNF649	2.32	5.59×10^{-6}	0.03	rs12459256	chr19: 52,323,088	3.84×10^{-4}	1.82×10^{-5}	2.32
GORAB	1.06	8.94×10^{-5}	0.27	rs6427252	chr1: 170,409,400	2.10×10^{-4}	8.42×10^{-4}	1.06
CAMK1G	0.85	1.90×10^{-4}	0.43	rs17014822	chr1: 209,756,470	4.44×10^{-4}	6.63×10^{-4}	0.85
RAD18	0.79	2.46×10^{-4}	0.44	rs73132887	chr3: 8,823,195	3.27×10^{-4}	3.66×10^{-4}	0.79
FGD2	0.69	3.80×10^{-4}	0.57	rs831504	chr6: 36,988,364	9.28×10^{-3}	1.04×10^{-4}	0.69
LIN7A	0.60	5.59×10^{-4}	0.61	rs1163656	chr12: 81,337,458	1.63×10^{-3}	1.32×10^{-4}	0.60
SMIM14	0.58	5.81×10^{-4}	0.61	rs11937734	chr4: 39,482,848	5.70×10^{-4}	2.41×10^{-3}	0.58
TMEM8C	0.56	6.37×10^{-4}	0.61	rs3793627	chr9: 136,407,659	3.14 x10-4	1.69×10^{-2}	0.56
B3GNT2	0.55	6.82×10^{-4}	0.61	rs2122382	chr2: 62,326,484	7.03×10^{-4}	1.18×10^{-3}	0.55
ZNF385D	0.47	8.94×10^{-4}	0.67	rs6774929	chr3: 21,480,184	1.11×10^{-3}	4.66×10^{-4}	0.47
ZFR	0.47	8.94×10^{-4}	0.67	rs11948227	chr5: 32,273,114	9.96×10^{-3}	1.70×10^{-4}	0.47

CHAPTER 5: CONCLUSIONS

Endophenotype mapping enables identification of novel loci associated with Systemic Lupus Erythematosus risk

The main goal of this dissertation was to investigate the feasibility of immune endophenotype mapping to characterize the molecular basis of inter-individual variation in immune disease susceptibility. I describe the utility of this approach in chapter 2, where I focused on mapping variation in serum IFN-α activity in individuals with SLE. Focusing on this endophenotype enabled a case-case study design, which directly addresses the heterogeneity in SLE, where cases on the tail end of the distribution of serum IFN-α were assessed. This greatly enhanced our power to detect variants with odds ratios > 2, supporting the idea that genetic variants associated with molecular sub-phenotypes have higher effect sizes, and can therefore be detected in smaller sample cohorts.

Importantly, this approach greatly enhanced our power to detect novel genetic variations that provided information on the underlying biological etiology of SLE pathogenesis. For example, one of the top signals identified was a missense SNP in purine nucleoside phosphorylase (*PNP*), which encodes an enzyme with important roles in purine metabolism. Deficiency in *PNP* is characterized by defective T-cell and B-cell immunity as well as defective antibody responses [123]. The amino acid change in the missense SNP identified in this study could hence have significant effects on the activity of this enzyme in immune cells.

The novel loci identified in this study provided information on molecular processes that could lead to dysregulation of the IFN-α pathway and subsequent pathogenesis of SLE. These loci have not been reported in previous case-control genetic association studies in SLE, supporting the notion that genetic mapping of pathogenic molecular sub-phenotypes can address some of the unexplained heritability in complex heterogeneous diseases.

Immunomodulatory transcriptional effects of vitamin D in the immune cells

The pathogenesis of SLE is driven by a combination of both genetic and environmental risk factors. My work in chapter 2 focused mainly on the genetic bases for variation in SLE susceptibility. In chapters 3 and 4, I shifted my focus to vitamin D, which is a modifiable environmental factor for immune-mediated diseases like SLE. In chapter 3, I characterized the genome-wide transcriptional effects of the active hormonal form of vitamin D, 1,25-dihydroxyvitamin D_3 (1,25D). This work was done in the primary innate immune cell type, monocytes, in the context of a pro-inflammatory stimulant, bacterial lipopolysaccharide (LPS), which enabled examination of how inflammation modifies transcriptional response to 1,25D.

The main strength of the study design in chapter 3 was the ability to discern how 1,25D reverses the transcriptional effects of LPS on specific pathways in monocytes. Genes in pro-inflammatory signaling pathways, such as IL-8, TNFR2 and NF-kB signaling, which were significantly up-regulated by LPS, were conversely down-regulated by 1,25D. This pattern was also illustrated among genes in the mTOR signaling and EIF2 signaling pathways, which were significantly down-regulated by LPS, while 1,25D significantly up-

regulated the same genes, reversing the transcriptional effects of LPS. The mTOR and EIF2 signaling pathways are particularly interesting as they have been previously implicated in regulating pro-inflammatory response in various cell types [41, 168-170, 172], and they thus highlight some mechanisms through which 1,25D attenuates the pro-inflammatory response to LPS.

This study characterized the transcriptional effects of 1,25D in the presence or absence of LPS, highlighting several biological pathways that mediate the immunomodulatory function of 1,25D. Dysregulation of these pathways could lead to immune disease pathogenesis.

Molecular basis for inter-individual variation in response to vitamin D

The main goal of my final thesis project described in chapter 4 was to understand the genetic basis for inter-individual variation in response to 1,25D. The increased interest in the potential use of 1,25D as a therapeutic agent in immune-mediated diseases has resulted in randomized supplementation trials of vitamin D. The results of these trials have so far been mixed, and it is important to consider inter-individual differences in the response to vitamin D, as well as the genetic basis of these differences, in the design of these studies. At least one study supports this notion, where a polymorphism in the vitamin D receptor gene (*VDR*) was identified as a modulator of response to vitamin D supplementation in Tuberculosis patients [89].

An additional strength of endophenotype mapping is the ability to incorporate multiple levels of phenotypic information to increase the power to detect novel loci in

relatively small sample sizes. Chapter 4 describes the joint mapping of cellular and transcriptional response phenotypes in peripheral blood mononuclear cells (PBMCs) from 88 individuals treated with 1,25D or vehicle control for both cellular and transcriptional response experiments, resulting in identification of novel loci underlying the immunomodulatory effects of 1,25D. The GWAS on the percent inhibition of PBMC proliferation (I_{max}) by 1,25D identified genome-wide significant SNPs in intergenic regions in two loci. Combining the information on cellular and transcriptional effects enabled identification of potential *trans*-effects of the two top GWAS SNPs, revealing a regulatory cascade where the top GWAS SNPs could influence cellular response to 1,25D through *trans*-regulation of genes that mediate the anti-proliferative activity of 1,25D.

Overall, this dissertation demonstrates that focusing on endophenotypes may provide useful insights into the genes underlying variation not only in disease susceptibility, but also variation in response to therapeutic agents for these diseases. Using this powerful endophenotype mapping approach to understand the etiology of disease pathogenesis and drug response may be informative not only for supplementation trials, but it may also provide additional potential therapeutic targets for these diseases.

Future Directions: Functional validation of novel loci underlying variation in immune disease susceptibility

Future work could involve incorporation of functional experiments, such as assays for transcription factor binding sites, enhancers, and other epigenetic markers, to the endophenotype mapping analyses in relevant primary immune cell subtypes, under the same experimental conditions. Ideally, carrying out this type of multi-level phenotypic analysis in a longitudinal study prior to, and after disease onset, will enable further understanding of the genetic architecture of the events that lead to final disease pathogenesis. Full knowledge of the molecular mechanisms underlying disease pathogenesis will be especially useful in predicting individuals who are more prone to complex immune-mediated diseases like SLE, and who are more likely to benefit from therapies specifically targeting the pathogenic endophenotypes that lead to disease susceptibility.

REFERENCES

1. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461(7265):747-53. Epub 2009/10/09. doi: 10.1038/nature08494. PubMed PMID: 19812666; PubMed Central PMCID: PMCPMC2831613.

2. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996;273(5281):1516-7. Epub 1996/09/13. PubMed PMID: 8801636.

3. Risch NJ. Searching for genetic determinants in the new millennium. Nature. 2000;405(6788):847-56. Epub 2000/06/24. doi: 10.1038/35015718. PubMed PMID: 10866211.

4. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449(7164):851-61. Epub 2007/10/19. doi: 10.1038/nature06258. PubMed PMID: 17943122; PubMed Central PMCID: PMCPMC2689609.

5. Hardy J, Singleton A. Genomewide association studies and human disease. N Engl J Med. 2009;360(17):1759-68. Epub 2009/04/17. doi: 10.1056/NEJMra0808700. PubMed PMID: 19369657; PubMed Central PMCID: PMCPMC3422859.

6. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet. 2009;5(5):e1000477. Epub 2009/06/06. doi: 10.1371/journal.pgen.1000477. PubMed PMID: 19492015; PubMed Central PMCID: PMCPMC2688469.

7. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007;447(7145):661-78. Epub 2007/06/08. doi: 10.1038/nature05911. PubMed PMID: 17554300; PubMed Central PMCID: PMCPMC2719288.

8. Weidinger S, Baurecht H, Naumann A, Novak N. Genome-wide association studies on IgE regulation: are genetics of IgE also genetics of atopic disease? Curr Opin Allergy Clin Immunol. 2010;10(5):408-17. Epub 2010/08/26. doi: 10.1097/ACI.0b013e32833d7d2d. PubMed PMID: 20736732.

9. Freimer N, Sabatti C. The human phenome project. Nat Genet. 2003;34(1):15-21. Epub 2003/05/02. doi: 10.1038/ng0503-15. PubMed PMID: 12721547.

10. Harley JB, Kelly JA, Kaufman KM. Unraveling the genetics of systemic lupus erythematosus. Springer Semin Immunopathol. 2006;28(2):119-30. Epub 2006/10/06. doi: 10.1007/s00281-006-0040-5. PubMed PMID: 17021721.

11. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5(12):889-99. Epub 2004/12/02. doi: 10.1038/nrg1489. PubMed PMID: 15573121.

12. Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. PLoS Genet. 2007;3(11):e192. Epub 2007/11/14. doi: 10.1371/journal.pgen.0030192. PubMed PMID: 17997607; PubMed Central PMCID: PMCPMC2065882.

13. Moser KL, Kelly JA, Lessard CJ, Harley JB. Recent insights into the genetic basis of systemic lupus erythematosus. Genes Immun. 2009;10(5):373-9. Epub 2009/05/15. doi: 10.1038/gene.2009.39. PubMed PMID: 19440199; PubMed Central PMCID: PMCPMC3144759.

14. Harley IT, Kaufman KM, Langefeld CD, Harley JB, Kelly JA. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. Nat Rev Genet. 2009;10(5):285-90. Epub 2009/04/02. doi: 10.1038/nrg2571. PubMed PMID: 19337289; PubMed Central PMCID: PMCPMC2737697.

15. Graham RR, Kozyrev SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW, et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet. 2006;38(5):550-5. Epub 2006/04/28. doi: 10.1038/ng1782. PubMed PMID: 16642019.

16. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet. 2008;40(2):204-10. Epub 2008/01/22. doi: 10.1038/ng.81. PubMed PMID: 18204446; PubMed Central PMCID: PMCPMC3712260.

17. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med. 2007;357(10):977-86. Epub 2007/09/07. doi: 10.1056/NEJMoa073003. PubMed PMID: 17804842; PubMed Central PMCID: PMCPMC2630215.

18. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009;41(11):1228-33. Epub 2009/10/20. doi: 10.1038/ng.468. PubMed PMID: 19838195; PubMed Central PMCID: PMCPMC2925843.

19. Rullo OJ, Tsao BP. Recent insights into the genetic basis of systemic lupus erythematosus. Ann Rheum Dis. 2013;72 Suppl 2:ii56-61. Epub 2012/12/21. doi:

10.1136/annrheumdis-2012-202351. PubMed PMID: 23253915; PubMed Central PMCID: PMCPMC3780983.

20. So HC, Gui AH, Cherny SS, Sham PC. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet Epidemiol. 2011;35(5):310-7. Epub 2011/03/05. doi: 10.1002/gepi.20579. PubMed PMID: 21374718.

21. Yang W, Lau YL. Solving the genetic puzzle of systemic lupus erythematosus. Pediatr Nephrol. 2015;30(10):1735-48. Epub 2014/09/23. doi: 10.1007/s00467-014-2947-8. PubMed PMID: 25239301.

22. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008;40(8):955-62. Epub 2008/07/01. doi: 10.1038/ng.175. PubMed PMID: 18587394; PubMed Central PMCID: PMCPMC2574810.

23. Kariuki SN, Franek BS, Kumar AA, Arrington J, Mikolaitis RA, Utset TO, et al. Trait-stratified genome-wide association study identifies novel and diverse genetic associations with serologic and cytokine phenotypes in systemic lupus erythematosus. Arthritis Res Ther. 2010;12(4):R151. Epub 2010/07/28. doi: 10.1186/ar3101. PubMed PMID: 20659327; PubMed Central PMCID: PMCPmc2945049.

24. Blanco P, Palucka AK, Gill M, Pascual V, Banchereau J. Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. Science. 2001;294(5546):1540-3. Epub 2001/11/17. doi: 10.1126/science.1064890. PubMed PMID: 11711679.

25. Takaoka A, Yanai H. Interferon signalling network in innate defence. Cell Microbiol. 2006;8(6):907-22. Epub 2006/05/10. doi: 10.1111/j.1462-5822.2006.00716.x. PubMed PMID: 16681834.

26. Hooks JJ, Moutsopoulos HM, Geis SA, Stahl NI, Decker JL, Notkins AL. Immune interferon in the circulation of patients with autoimmune disease. N Engl J Med. 1979;301(1):5-8. Epub 1979/07/05. doi: 10.1056/nejm197907053010102. PubMed PMID: 449915.

27. Kirou KA, Lee C, George S, Louca K, Peterson MG, Crow MK. Activation of the interferon-alpha pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. Arthritis Rheum. 2005;52(5):1491-503. Epub 2005/05/10. doi: 10.1002/art.21031. PubMed PMID: 15880830.

28. Niewold TB, Hua J, Lehman TJ, Harley JB, Crow MK. High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. Genes Immun.

2007;8(6):492-502. Epub 2007/06/22. doi: 10.1038/sj.gene.6364408. PubMed PMID: 17581626; PubMed Central PMCID: PMCPMC2702174.

29. Deluca HF, Cantorna MT. Vitamin D: its role and uses in immunology. Faseb j. 2001;15(14):2579-85. Epub 2001/12/01. doi: 10.1096/fj.01-0433rev. PubMed PMID: 11726533.

30. Lin R, White JH. The pleiotropic actions of vitamin D. Bioessays. 2004;26(1):21-8. Epub 2003/12/30. doi: 10.1002/bies.10368. PubMed PMID: 14696037.

31. Cantorna MT, Zhu Y, Froicu M, Wittke A. Vitamin D status, 1,25-dihydroxyvitamin D3, and the immune system. Am J Clin Nutr. 2004;80(6 Suppl):1717s-20s. Epub 2004/12/09. PubMed PMID: 15585793.

32. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, et al. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. Science. 2006;311(5768):1770-3. Epub 2006/02/25. doi: 10.1126/science.1123933. PubMed PMID: 16497887.

33. Bikle D. Nonclassic actions of vitamin D. J Clin Endocrinol Metab. 2009;94(1):26-34. Epub 2008/10/16. doi: 10.1210/jc.2008-1454. PubMed PMID: 18854395; PubMed Central PMCID: PMCPmc2630868.

34. Gombart AF. The vitamin D-antimicrobial peptide pathway and its role in protection against infection. Future Microbiol. 2009;4(9):1151-65. Epub 2009/11/10. doi: 10.2217/fmb.09.87. PubMed PMID: 19895218; PubMed Central PMCID: PMCPmc2821804.

35. Hewison M. Vitamin D and the intracrinology of innate immunity. Mol Cell Endocrinol. 2010;321(2):103-11. Epub 2010/02/17. doi: 10.1016/j.mce.2010.02.013. PubMed PMID: 20156523; PubMed Central PMCID: PMCPmc2854233.

36. Zittermann A, Gummert JF. Nonclassical vitamin D action. Nutrients. 2010;2(4):408-25. Epub 2010/04/01. doi: 10.3390/nu2040408. PubMed PMID: 22254030; PubMed Central PMCID: PMCPmc3257656.

37. Baeke F, Takiishi T, Korf H, Gysemans C, Mathieu C. Vitamin D: modulator of the immune system. Curr Opin Pharmacol. 2010;10(4):482-96. Epub 2010/04/30. doi: 10.1016/j.coph.2010.04.001. PubMed PMID: 20427238.

38. Aranow C. Vitamin D and the immune system. J Investig Med. 2011;59(6):881-6. Epub 2011/04/30. doi: 10.231/JIM.0b013e31821b8755. PubMed PMID: 21527855; PubMed Central PMCID: PMCPmc3166406.

39. Hewison M. Antibacterial effects of vitamin D. Nat Rev Endocrinol. 2011;7(6):337-45. Epub 2011/01/26. doi: 10.1038/nrendo.2010.226. PubMed PMID: 21263449.

40. Chun RF, Liu PT, Modlin RL, Adams JS, Hewison M. Impact of vitamin D on immune function: lessons learned from genome-wide analysis. Front Physiol. 2014;5:151. Epub 2014/05/06. doi: 10.3389/fphys.2014.00151. PubMed PMID: 24795646; PubMed Central PMCID: PMCPmc4000998.

41. Ferreira GB, Vanherwegen AS, Eelen G, Gutierrez AC, Van Lommel L, Marchal K, et al. Vitamin D3 Induces Tolerance in Human Dendritic Cells by Activation of Intracellular Metabolic Pathways. Cell Rep. 2015. Epub 2015/02/11. doi: 10.1016/j.celrep.2015.01.013. PubMed PMID: 25660022.

42. Looker AC, Johnson CL, Lacher DA, Pfeiffer CM, Schleicher RL, Sempos CT. Vitamin D status: United States, 2001-2006. NCHS Data Brief. 2011;(59):1-8. PubMed PMID: 21592422.

43. Rostand SG. Vitamin D, blood pressure, and African Americans: toward a unifying hypothesis. Clin J Am Soc Nephrol. 2010;5(9):1697-703. Epub 2010/07/24. doi: 10.2215/cjn.02960410. PubMed PMID: 20651156.

44. Freedman BI, Register TC. Effect of race and genetics on vitamin D metabolism, bone and vascular health. Nat Rev Nephrol. 2012;8(8):459-66. Epub 2012/06/13. doi: 10.1038/nrneph.2012.112. PubMed PMID: 22688752.

45. Murphy AB, Kelley B, Nyame YA, Martin IK, Smith DJ, Castaneda L, et al. Predictors of serum vitamin D levels in African American and European American men in Chicago. Am J Mens Health. 2012;6(5):420-6. Epub 2012/03/09. doi: 10.1177/1557988312437240. PubMed PMID: 22398989; PubMed Central PMCID: PMCPmc3678722.

46. Singh A, Kamen DL. Potential benefits of vitamin D for patients with systemic lupus erythematosus. Dermatoendocrinol. 2012;4(2):146-51. Epub 2012/08/29. doi: 10.4161/derm.20443. PubMed PMID: 22928070; PubMed Central PMCID: PMCPMC3427193.

47. Kamen DL, Cooper GS, Bouali H, Shaftman SR, Hollis BW, Gilkeson GS. Vitamin D deficiency in systemic lupus erythematosus. Autoimmun Rev. 2006;5(2):114-7. Epub 2006/01/25. doi: 10.1016/j.autrev.2005.05.009. PubMed PMID: 16431339.

48. Kamen D, Aranow C. Vitamin D in systemic lupus erythematosus. Curr Opin Rheumatol. 2008;20(5):532-7. Epub 2008/08/14. doi: 10.1097/BOR.0b013e32830a991b. PubMed PMID: 18698173.

49. Ben-Zvi I, Aranow C, Mackay M, Stanevsky A, Kamen DL, Marinescu LM, et al. The impact of vitamin D on dendritic cell function in patients with systemic lupus

erythematosus. PLoS One. 2010;5(2):e9193. Epub 2010/02/20. doi: 10.1371/journal.pone.0009193. PubMed PMID: 20169063; PubMed Central PMCID: PMCPmc2821911.

50. Sanchez E, Comeau ME, Freedman BI, Kelly JA, Kaufman KM, Langefeld CD, et al. Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. Arthritis Rheum. 2011;63(11):3493-501. Epub 2011/07/28. doi: 10.1002/art.30563. PubMed PMID: 21792837; PubMed Central PMCID: PMCPMC3205224.

51. Ko K, Koldobskaya Y, Rosenzweig E, Niewold TB. Activation of the Interferon Pathway is Dependent Upon Autoantibodies in African-American SLE Patients, but Not in European-American SLE Patients. Front Immunol. 2013;4:309. Epub 2013/10/09. doi: 10.3389/fimmu.2013.00309. PubMed PMID: 24101921; PubMed Central PMCID: PMCPMC3787392.

52. Gilkeson G, James J, Kamen D, Knackstedt T, Maggi D, Meyer A, et al. The United States to Africa lupus prevalence gradient revisited. Lupus. 2011;20(10):1095-103. Epub 2011/09/16. doi: 10.1177/0961203311404915. PubMed PMID: 21917913; PubMed Central PMCID: PMCPMC3535487.

53. Inouye K, Sakaki T. Enzymatic studies on the key enzymes of vitamin D metabolism; 1 alpha-hydroxylase (CYP27B1) and 24-hydroxylase (CYP24). Biotechnol Annu Rev. 2001;7:179-94. Epub 2001/11/01. PubMed PMID: 11686044.

54. Sakaki T, Kagawa N, Yamamoto K, Inouye K. Metabolism of vitamin D3 by cytochromes P450. Front Biosci. 2005;10:119-34. Epub 2004/12/03. PubMed PMID: 15574355.

55. Holick MF. Vitamin D deficiency. N Engl J Med. 2007;357(3):266-81. Epub 2007/07/20. doi: 10.1056/NEJMra070553. PubMed PMID: 17634462.

56. Kamen DL, Tangpricha V. Vitamin D and molecular actions on the immune system: modulation of innate and autoimmunity. J Mol Med (Berl). 2010;88(5):441-50. Epub 2010/02/02. doi: 10.1007/s00109-010-0590-9. PubMed PMID: 20119827; PubMed Central PMCID: PMCPMC2861286.

57. Cross HS. Extrarenal vitamin D hydroxylase expression and activity in normal and malignant cells: modification of expression by epigenetic mechanisms and dietary substances. Nutr Rev. 2007;65(8 Pt 2):S108-12. Epub 2007/09/18. PubMed PMID: 17867383.

58. Christakos S, Dhawan P, Liu Y, Peng X, Porta A. New insights into the mechanisms of vitamin D action. J Cell Biochem. 2003;88(4):695-705. Epub 2003/02/11. doi: 10.1002/jcb.10423. PubMed PMID: 12577303.

59. Adams JS, Ren S, Liu PT, Chun RF, Lagishetty V, Gombart AF, et al. Vitamin d-directed rheostatic regulation of monocyte antibacterial responses. J Immunol. 2009;182(7):4289-95. Epub 2009/03/21. doi: 10.4049/jimmunol.0803736. PubMed PMID: 19299728; PubMed Central PMCID: PMCPmc2683618.

60. Nnoaham KE, Clarke A. Low serum vitamin D levels and tuberculosis: a systematic review and meta-analysis. Int J Epidemiol. 2008;37(1):113-9. Epub 2008/02/05. doi: 10.1093/ije/dym247. PubMed PMID: 18245055.

61. White JH. Vitamin D signaling, infectious diseases, and regulation of innate immunity. Infect Immun. 2008;76(9):3837-43. Epub 2008/05/29. doi: 10.1128/iai.00353-08. PubMed PMID: 18505808; PubMed Central PMCID: PMCPmc2519414.

62. Williams CJB. Cod-liver Oil in Phthisis. BMJ. 1849;s2-1(1):1-18. doi: 10.1136/bmj.s2-1.1.1.

63. Martineau AR, Honecker FU, Wilkinson RJ, Griffiths CJ. Vitamin D in the treatment of pulmonary tuberculosis. J Steroid Biochem Mol Biol. 2007;103(3-5):793-8. Epub 2007/01/16. doi: 10.1016/j.jsbmb.2006.12.052. PubMed PMID: 17223549.

64. Rigby WF, Waugh MG. Decreased accessory cell function and costimulatory activity by 1,25-dihydroxyvitamin D3-treated monocytes. Arthritis Rheum. 1992;35(1):110-9. Epub 1992/01/01. PubMed PMID: 1370618.

65. van Halteren AG, Tysma OM, van Etten E, Mathieu C, Roep BO. 1alpha,25-dihydroxyvitamin D3 or analogue treated dendritic cells modulate human autoreactive T cells via the selective induction of apoptosis. J Autoimmun. 2004;23(3):233-9. Epub 2004/10/27. doi: 10.1016/j.jaut.2004.06.004. PubMed PMID: 15501394.

66. Ferreira GB, Kleijwegt FS, Waelkens E, Lage K, Nikolic T, Hansen DA, et al. Differential protein pathways in 1,25-dihydroxyvitamin d(3) and dexamethasone modulated tolerogenic human dendritic cells. J Proteome Res. 2012;11(2):941-71. Epub 2011/11/23. doi: 10.1021/pr200724e. PubMed PMID: 22103328.

67. Verway M, Bouttier M, Wang TT, Carrier M, Calderon M, An BS, et al. Vitamin D induces interleukin-1beta expression: paracrine macrophage epithelial signaling controls M. tuberculosis infection. PLoS Pathog. 2013;9(6):e1003407. Epub 2013/06/14. doi: 10.1371/journal.ppat.1003407. PubMed PMID: 23762029; PubMed Central PMCID: PMCPmc3675149.

68. Neme A, Nurminen V, Seuter S, Carlberg C. The vitamin D-dependent transcriptome of human monocytes. J Steroid Biochem Mol Biol. 2015. Epub 2015/11/03. doi: 10.1016/j.jsbmb.2015.10.018. PubMed PMID: 26523676.

69. Seuter S, Neme A, Carlberg C. Epigenome-wide effects of vitamin D and their impact on the transcriptome of human monocytes involve CTCF. Nucleic Acids Res. 2015. Epub 2015/12/31. doi: 10.1093/nar/gkv1519. PubMed PMID: 26715761.

70. Tsuchiya S, Kobayashi Y, Goto Y, Okumura H, Nakae S, Konno T, et al. Induction of maturation in cultured human monocytic leukemia cells by a phorbol diester. Cancer Res. 1982;42(4):1530-6. Epub 1982/04/01. PubMed PMID: 6949641.

71. Fairfax BP, Knight JC. Genetics of gene expression in immunity to infection. Curr Opin Immunol. 2014;30:63-71. Epub 2014/08/01. doi: 10.1016/j.coi.2014.07.001. PubMed PMID: 25078545; PubMed Central PMCID: PMCPMC4426291.

72. Kupfer SS, Maranville JC, Baxter SS, Huang Y, Di Rienzo A. Comparison of cellular and transcriptional responses to 1,25-dihydroxyvitamin d3 and glucocorticoids in peripheral blood mononuclear cells. PLoS One. 2013;8(10):e76643. Epub 2013/10/12. doi: 10.1371/journal.pone.0076643. PubMed PMID: 24116131; PubMed Central PMCID: PMCPmc3792986.

73. Lemire JM, Adams JS, Sakai R, Jordan SC. 1 alpha,25-dihydroxyvitamin D3 suppresses proliferation and immunoglobulin production by normal human peripheral blood mononuclear cells. J Clin Invest. 1984;74(2):657-61. Epub 1984/08/01. doi: 10.1172/jci111465. PubMed PMID: 6611355; PubMed Central PMCID: PMCPMC370520.

74. Lemire JM, Adams JS, Kermani-Arab V, Bakke AC, Sakai R, Jordan SC. 1,25-Dihydroxyvitamin D3 suppresses human T helper/inducer lymphocyte activity in vitro. J Immunol. 1985;134(5):3032-5. Epub 1985/05/01. PubMed PMID: 3156926.

75. Jeffery LE, Burke F, Mura M, Zheng Y, Qureshi OS, Hewison M, et al. 1,25-Dihydroxyvitamin D3 and IL-2 combine to inhibit T cell production of inflammatory cytokines and promote development of regulatory T cells expressing CTLA-4 and FoxP3. J Immunol. 2009;183(9):5458-67. Epub 2009/10/22. doi: 10.4049/jimmunol.0803217. PubMed PMID: 19843932; PubMed Central PMCID: PMCPMC2810518.

76. Chen S, Sims GP, Chen XX, Gu YY, Lipsky PE. Modulatory effects of 1,25-dihydroxyvitamin D3 on human B cell differentiation. J Immunol. 2007;179(3):1634-47. Epub 2007/07/21. PubMed PMID: 17641030.

77. Peelen E, Knippenberg S, Muris AH, Thewissen M, Smolders J, Tervaert JW, et al. Effects of vitamin D on the peripheral adaptive immune system: a review. Autoimmun Rev. 2011;10(12):733-43. Epub 2011/05/31. doi: 10.1016/j.autrev.2011.05.002. PubMed PMID: 21621002.

78. Opal SM. The host response to endotoxin, antilipopolysaccharide strategies, and the management of severe sepsis. Int J Med Microbiol. 2007;297(5):365-77. Epub 2007/04/25. doi: 10.1016/j.ijmm.2007.03.006. PubMed PMID: 17452016.

79. Zhang Y, Leung DY, Richers BN, Liu Y, Remigio LK, Riches DW, et al. Vitamin D inhibits monocyte/macrophage proinflammatory cytokine production by targeting MAPK phosphatase-1. J Immunol. 2012;188(5):2127-35. Epub 2012/02/04. doi: 10.4049/jimmunol.1102412. PubMed PMID: 22301548; PubMed Central PMCID: PMCPmc3368346.

80. Lehmann V, Freudenberg MA, Galanos C. Lethal toxicity of lipopolysaccharide and tumor necrosis factor in normal and D-galactosamine-treated mice. J Exp Med. 1987;165(3):657-63. Epub 1987/03/01. PubMed PMID: 3819645; PubMed Central PMCID: PMCPmc2188282.

81. Correale J, Ysrraelit MC, Gaitan MI. Immunomodulatory effects of Vitamin D in multiple sclerosis. Brain. 2009;132(Pt 5):1146-60. Epub 2009/03/27. doi: 10.1093/brain/awp033. PubMed PMID: 19321461.

82. Pakpoor J, Ramagopalan S. Evidence for an Association Between Vitamin D and Multiple Sclerosis. Curr Top Behav Neurosci. 2015;26:105-15. Epub 2014/12/17. doi: 10.1007/7854_2014_358. PubMed PMID: 25502544.

83. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. Lancet Neurol. 2010;9(6):599-612. Epub 2010/05/25. doi: 10.1016/s1474-4422(10)70086-7. PubMed PMID: 20494325.

84. Hypponen E, Laara E, Reunanen A, Jarvelin MR, Virtanen SM. Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study. Lancet. 2001;358(9292):1500-3. Epub 2001/11/14. doi: 10.1016/s0140-6736(01)06580-1. PubMed PMID: 11705562.

85. Smolders J, Peelen E, Thewissen M, Cohen Tervaert JW, Menheere P, Hupperts R, et al. Safety and T cell modulating effects of high dose vitamin D3 supplementation in multiple sclerosis. PLoS One. 2010;5(12):e15235. Epub 2010/12/24. doi: 10.1371/journal.pone.0015235. PubMed PMID: 21179201; PubMed Central PMCID: PMCPMC3001453.

86. Prietl B, Pilz S, Wolf M, Tomaschitz A, Obermayer-Pietsch B, Graninger W, et al. Vitamin D supplementation and regulatory T cells in apparently healthy subjects: vitamin D treatment for autoimmune diseases? Isr Med Assoc J. 2010;12(3):136-9. Epub 2010/08/06. PubMed PMID: 20684175.

87. Wejse C, Gomes VF, Rabna P, Gustafson P, Aaby P, Lisse IM, et al. Vitamin D as supplementary treatment for tuberculosis: a double-blind, randomized, placebo-controlled trial. Am J Respir Crit Care Med. 2009;179(9):843-50. Epub 2009/01/31. doi: 10.1164/rccm.200804-567OC. PubMed PMID: 19179490.

88. Li-Ng M, Aloia JF, Pollack S, Cunha BA, Mikhail M, Yeh J, et al. A randomized controlled trial of vitamin D3 supplementation for the prevention of symptomatic upper respiratory tract infections. Epidemiol Infect. 2009;137(10):1396-404. Epub 2009/03/20. doi: 10.1017/s0950268809002404. PubMed PMID: 19296870.

89. Martineau AR, Timms PM, Bothamley GH, Hanifa Y, Islam K, Claxton AP, et al. High-dose vitamin D(3) during intensive-phase antimicrobial treatment of pulmonary tuberculosis: a double-blind randomised controlled trial. Lancet. 2011;377(9761):242-50. Epub 2011/01/11. doi: 10.1016/s0140-6736(10)61889-2. PubMed PMID: 21215445; PubMed Central PMCID: PMCPMC4176755.

90. Aranow C, Kamen DL, Dall'Era M, Massarotti EM, Mackay MC, Koumpouras F, et al. Randomized, Double-Blind, Placebo-Controlled Trial of the Effect of Vitamin D3 on the Interferon Signature in Patients With Systemic Lupus Erythematosus. Arthritis Rheumatol. 2015;67(7):1848-57. Epub 2015/03/18. doi: 10.1002/art.39108. PubMed PMID: 25777546; PubMed Central PMCID: PMCPMC4732716.

91. Martineau AR, Wilkinson RJ, Wilkinson KA, Newton SM, Kampmann B, Hall BM, et al. A single dose of vitamin D enhances immunity to mycobacteria. Am J Respir Crit Care Med. 2007;176(2):208-13. Epub 2007/04/28. doi: 10.1164/rccm.200701-007OC. PubMed PMID: 17463418.

92. Zhao LJ, Zhou Y, Bu F, Travers-Gustafson D, Ye A, Xu X, et al. Factors predicting vitamin D response variation in non-Hispanic white postmenopausal women. J Clin Endocrinol Metab. 2012;97(8):2699-705. Epub 2012/05/16. doi: 10.1210/jc.2011-3401. PubMed PMID: 22585090; PubMed Central PMCID: PMCPMC3410260.

93. Aloia JF, Patel M, Dimaano R, Li-Ng M, Talwar SA, Mikhail M, et al. Vitamin D intake to attain a desired serum 25-hydroxyvitamin D concentration. Am J Clin Nutr. 2008;87(6):1952-8. Epub 2008/06/11. PubMed PMID: 18541590.

94. Maranville JC, Baxter SS, Witonsky DB, Chase MA, Di Rienzo A. Genetic mapping with multiple levels of phenotypic information reveals determinants of lymphocyte glucocorticoid sensitivity. Am J Hum Genet. 2013;93(4):735-43. Epub 2013/09/24. doi: 10.1016/j.ajhg.2013.08.005. PubMed PMID: 24055111; PubMed Central PMCID: PMCPMC3791266.

95. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9:403-33. Epub 2008/07/03. doi: 10.1146/annurev.genom.9.081307.164258. PubMed PMID: 18593304; PubMed Central PMCID: PMCPMC2953791.

96. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS

Genet. 2010;6(4):e1000888. Epub 2010/04/07. doi: 10.1371/journal.pgen.1000888. PubMed PMID: 20369019; PubMed Central PMCID: PMCPMC2848547.

97. Ko K, Franek BS, Marion M, Kaufman KM, Langefeld CD, Harley JB, et al. Genetic ancestry, serum interferon-alpha activity, and autoantibodies in systemic lupus erythematosus. J Rheumatol. 2012;39(6):1238-40. Epub 2012/04/17. doi: 10.3899/jrheum.111467. PubMed PMID: 22505704; PubMed Central PMCID: PMCPMC3381952.

98. Ghodke-Puranik Y, Niewold TB. Genetics of the type I interferon pathway in systemic lupus erythematosus. International Journal of Clinical Rheumatology. 2013;8(6):657-69.

99. Niewold TB, Kelly JA, Kariuki SN, Franek BS, Kumar AA, Kaufman KM, et al. IRF5 haplotypes demonstrate diverse serological associations which predict serum interferon alpha activity and explain the majority of the genetic association with systemic lupus erythematosus. Ann Rheum Dis. 2012;71(3):463-8. Epub 2011/11/18. doi: 10.1136/annrheumdis-2011-200463. PubMed PMID: 22088620; PubMed Central PMCID: PMCPmc3307526.

100. Weckerle CE, Franek BS, Kelly JA, Kumabe M, Mikolaitis RA, Green SL, et al. Network analysis of associations between serum interferon-alpha activity, autoantibodies, and clinical features in systemic lupus erythematosus. Arthritis Rheum. 2011;63(4):1044-53. Epub 2010/12/17. doi: 10.1002/art.30187. PubMed PMID: 21162028; PubMed Central PMCID: PMCPMC3068224.

101. Petri M, Singh S, Tesfasyone H, Dedrick R, Fry K, Lal P, et al. Longitudinal expression of type I interferon responsive genes in systemic lupus erythematosus. Lupus. 2009;18(11):980-9. Epub 2009/09/19. doi: 18/11/980 [pii]10.1177/0961203309105529. PubMed PMID: 19762399.

102. Niewold TB, Clark DN, Salloum R, Poole BD. Interferon alpha in systemic lupus erythematosus. J Biomed Biotechnol. 2010;2010:948364. Epub 2010/07/24. doi: 10.1155/2010/948364. PubMed PMID: 20652065; PubMed Central PMCID: PMCPMC2896914.

103. Niewold TB. Interferon alpha-induced lupus: proof of principle. J Clin Rheumatol. 2008;14(3):131-2. Epub 2008/06/06. doi: 10.1097/RHU.0b013e318177627d. PubMed PMID: 18525429; PubMed Central PMCID: PMC2743115.

104. Niewold TB, Swedler WI. Systemic lupus erythematosus arising during interferon-alpha therapy for cryoglobulinemic vasculitis associated with hepatitis C. Clin Rheumatol. 2005;24(2):178-81. Epub 2004/11/27. doi: 10.1007/s10067-004-1024-2. PubMed PMID: 15565395.

105. Kariuki SN, Kirou KA, MacDermott EJ, Barillas-Arias L, Crow MK, Niewold TB. Cutting edge: autoimmune disease risk variant of STAT4 confers increased sensitivity to IFN-alpha in lupus patients in vivo. J Immunol. 2009;182(1):34-8. Epub 2008/12/26. PubMed PMID: 19109131; PubMed Central PMCID: PMCPmc2716754.

106. Niewold TB, Kelly JA, Flesch MH, Espinoza LR, Harley JB, Crow MK. Association of the IRF5 risk haplotype with high serum interferon-alpha activity in systemic lupus erythematosus patients. Arthritis Rheum. 2008;58(8):2481-7. Epub 2008/08/01. doi: 10.1002/art.23613. PubMed PMID: 18668568; PubMed Central PMCID: PMC2621107.

107. Robinson T, Kariuki SN, Franek BS, Kumabe M, Kumar AA, Badaracco M, et al. Autoimmune disease risk variant of IFIH1 is associated with increased sensitivity to IFN-alpha and serologic autoimmunity in lupus patients. J Immunol. 2011;187(3):1298-303. Epub 2011/06/28. doi: 10.4049/jimmunol.1100857. PubMed PMID: 21705624; PubMed Central PMCID: PMCPmc3304466.

108. Agik S, Franek BS, Kumar AA, Kumabe M, Utset TO, Mikolaitis RA, et al. The autoimmune disease risk allele of UBE2L3 in African American patients with systemic lupus erythematosus: a recessive effect upon subphenotypes. J Rheumatol. 2012;39(1):73-8. Epub 2011/11/03. doi: 10.3899/jrheum.110590. PubMed PMID: 22045845; PubMed Central PMCID: PMCPMC3304461.

109. Eliopoulos E, Zervou MI, Andreou A, Dimopoulou K, Cosmidis N, Voloudakis G, et al. Association of the PTPN22 R620W polymorphism with increased risk for SLE in the genetically homogeneous population of Crete. Lupus. 2011;20(5):501-6. Epub 2011/05/06. doi: 10.1177/0961203310392423. PubMed PMID: 21543514; PubMed Central PMCID: PMCPMC3312778.

110. Mavragani CP, Niewold TB, Moutsopoulos NM, Pillemer SR, Wahl SM, Crow MK. Augmented interferon-alpha pathway activation in patients with Sjogren's syndrome treated with etanercept. Arthritis Rheum. 2007;56(12):3995-4004. PubMed PMID: 18050196.

111. Niewold TB, Wu SC, Smith M, Morgan GA, Pachman LM. Familial aggregation of autoimmune disease in juvenile dermatomyositis. Pediatrics. 2011;127(5):e1239-46. Epub 2011/04/20. doi: peds.2010-3022 [pii]10.1542/peds.2010-3022. PubMed PMID: 21502224; PubMed Central PMCID: PMC3081190.

112. Maiti AK, Kim-Howard X, Motghare P, Pradhan V, Chua KH, Sun C, et al. Combined protein- and nucleic acid-level effects of rs1143679 (R77H), a lupus-predisposing variant within ITGAM. Hum Mol Genet. 2014;23(15):4161-76. Epub 2014/03/13. doi: 10.1093/hmg/ddu106. PubMed PMID: 24608226; PubMed Central PMCID: PMCPMC4082363.

113. Koldobskaya Y, Ko K, Kumar AA, Agik S, Arrington J, Kariuki SN, et al. Gene-expression-guided selection of candidate loci and molecular phenotype analyses enhance genetic discovery in systemic lupus erythematosus. Clin Dev Immunol. 2012;2012:682018. Epub 2012/09/19. doi: 10.1155/2012/682018. PubMed PMID: 22988468; PubMed Central PMCID: PMCPmc3439981.

114. Farkas L, Beiske K, Lund-Johansen F, Brandtzaeg P, Jahnsen FL. Plasmacytoid dendritic cells (natural interferon- alpha/beta-producing cells) accumulate in cutaneous lupus erythematosus lesions. Am J Pathol. 2001;159(1):237-43. PubMed PMID: 11438470.

115. Hagberg N, Berggren O, Leonard D, Weber G, Bryceson YT, Alm GV, et al. IFN-alpha production by plasmacytoid dendritic cells stimulated with RNA-containing immune complexes is promoted by NK cells via MIP-1beta and LFA-1. Journal of immunology. 2011;186(9):5085-94. Epub 2011/03/25. doi: 10.4049/jimmunol.1003349. PubMed PMID: 21430220.

116. Kim K, Brown EE, Choi CB, Alarcon-Riquelme ME, Kelly JA, Glenn SB, et al. Variation in the ICAM1-ICAM4-ICAM5 locus is associated with systemic lupus erythematosus susceptibility in multiple ancestries. Ann Rheum Dis. 2012;71(11):1809-14. Epub 2012/04/24. doi: 10.1136/annrheumdis-2011-201110. PubMed PMID: 22523428; PubMed Central PMCID: PMCPMC3466387.

117. Rullo OJ, Woo JM, Wu H, Hoftman AD, Maranian P, Brahn BA, et al. Association of IRF5 polymorphisms with activation of the interferon alpha pathway. Ann Rheum Dis. 2010;69(3):611-7. Epub 2009/10/27. doi: 10.1136/ard.2009.118315. PubMed PMID: 19854706; PubMed Central PMCID: PMCPMC3135414.

118. Orstavik S, Natarajan V, Tasken K, Jahnsen T, Sandberg M. Characterization of the human gene encoding the type I alpha and type I beta cGMP-dependent protein kinase (PRKG1). Genomics. 1997;42(2):311-8. Epub 1997/06/01. doi: 10.1006/geno.1997.4743. PubMed PMID: 9192852.

119. Snyder FF, Jenuth JP, Mably ER, Mangat RK. Point mutations at the purine nucleoside phosphorylase locus impair thymocyte differentiation in the mouse. Proc Natl Acad Sci U S A. 1997;94(6):2522-7. Epub 1997/03/18. PubMed PMID: 9122228; PubMed Central PMCID: PMC20121.

120. Toro A, Grunebaum E. TAT-mediated intracellular delivery of purine nucleoside phosphorylase corrects its deficiency in mice. J Clin Invest. 2006;116(10):2717-26. Epub 2006/09/12. doi: 10.1172/JCI25052. PubMed PMID: 16964310; PubMed Central PMCID: PMC1560347.

121. Lipsky PE. Systemic lupus erythematosus: an autoimmune disease of B cell hyperactivity. Nat Immunol. 2. United States2001. p. 764-6.

122. Sasaki Y, Iseki M, Yamaguchi S, Kurosawa Y, Yamamoto T, Moriwaki Y, et al. Direct evidence of autosomal recessive inheritance of Arg24 to termination codon in purine nucleoside phosphorylase gene in a family with a severe combined immunodeficiency patient. Hum Genet. 1998;103(1):81-5. Epub 1998/09/16. PubMed PMID: 9737781.

123. Markert ML. Purine nucleoside phosphorylase deficiency. Immunodefic Rev. 1991;3(1):45-81. Epub 1991/01/01. PubMed PMID: 1931007.

124. Lee-Kirsch MA, Gong M, Chowdhury D, Senenko L, Engel K, Lee YA, et al. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 are associated with systemic lupus erythematosus. Nat Genet. 2007;39(9):1065-7. Epub 2007/07/31. doi: ng2091 [pii]10.1038/ng2091. PubMed PMID: 17660818.

125. Al-Mayouf SM, Sunker A, Abdwani R, Abrawi SA, Almurshedi F, Alhashmi N, et al. Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. Nature genetics. 2011;43(12):1186-8. Epub 2011/10/25. doi: 10.1038/ng.975. PubMed PMID: 22019780.

126. Namjou B, Kothari PH, Kelly JA, Glenn SB, Ojwang JO, Adler A, et al. Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort. Genes and immunity. 2011;12(4):270-9. Epub 2011/01/29. doi: 10.1038/gene.2010.73. PubMed PMID: 21270825; PubMed Central PMCID: PMC3107387.

127. Lin WJ, Yang CY, Lin YC, Tsai MC, Yang CW, Tung CY, et al. Phafin2 modulates the structure and function of endosomes by a Rab5-dependent mechanism. Biochem Biophys Res Commun. 2010;391(1):1043-8. Epub 2009/12/10. doi: 10.1016/j.bbrc.2009.12.016. PubMed PMID: 19995552.

128. Li C, Liu Q, Li N, Chen W, Wang L, Wang Y, et al. EAPF/Phafin-2, a novel endoplasmic reticulum-associated protein, facilitates TNF-alpha-triggered cellular apoptosis through endoplasmic reticulum-mitochondrial apoptotic pathway. J Mol Med (Berl). 2008;86(4):471-84. Epub 2008/02/22. doi: 10.1007/s00109-007-0298-7. PubMed PMID: 18288467.

129. Stefansson B, Ohama T, Daugherty AE, Brautigan DL. Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. Biochemistry. 2008;47(5):1442-51. Epub 2008/01/12. doi: 10.1021/bi7022877. PubMed PMID: 18186651.

130. Chio A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai SL, et al. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet. 2009;18(8):1524-32. Epub 2009/02/06. doi: 10.1093/hmg/ddp059. PubMed PMID: 19193627; PubMed Central PMCID: PMC2664150.

131. Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat. 2009;30(1):69-78. Epub 2008/08/08. doi: 10.1002/humu.20822. PubMed PMID: 18683858; PubMed Central PMCID: PMC3073397.

132. Jabs WJ, Hennig C, Zawatzky R, Kirchner H. Failure to detect antiviral activity in serum and plasma of healthy individuals displaying high activity in ELISA for IFN-alpha and IFN-beta. J Interferon Cytokine Res. 1999;19(5):463-9. Epub 1999/07/01. doi: 10.1089/107999099313901. PubMed PMID: 10386858.

133. Hua J, Kirou K, Lee C, Crow MK. Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies. Arthritis Rheum. 2006;54(6):1906-16. Epub 2006/06/01. doi: 10.1002/art.21890. PubMed PMID: 16736505.

134. Feng X, Han D, Kilaru BK, Franek BS, Niewold TB, Reder AT. Inhibition of interferon-beta responses in multiple sclerosis immune cells associated with high-dose statins. Arch Neurol. 2012;69(10):1303-9. Epub 2012/07/18. doi: 10.1001/archneurol.2012.465. PubMed PMID: 22801747.

135. Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics. 2001;17(9):763-74. Epub 2001/10/09. PubMed PMID: 11590094.

136. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559-75. Epub 2007/08/19. doi: 10.1086/519795. PubMed PMID: 17701901; PubMed Central PMCID: PMC1950838.

137. Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J Evol Biol. 2005;18(5):1368-73. Epub 2005/09/02. doi: 10.1111/j.1420-9101.2005.00917.x. PubMed PMID: 16135132.

138. Liu PT, Schenk M, Walker VP, Dempsey PW, Kanchanapoomi M, Wheelwright M, et al. Convergence of IL-1beta and VDR activation pathways in human TLR2/1-induced antimicrobial responses. PLoS One. 2009;4(6):e5810. Epub 2009/06/09. doi: 10.1371/journal.pone.0005810. PubMed PMID: 19503839; PubMed Central PMCID: PMCPMC2686169.

139. Yuk JM, Shin DM, Lee HM, Yang CS, Jin HS, Kim KK, et al. Vitamin D3 induces autophagy in human monocytes/macrophages via cathelicidin. Cell Host Microbe. 2009;6(3):231-43. Epub 2009/09/15. doi: 10.1016/j.chom.2009.08.004. PubMed PMID: 19748465.

140. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008;24(13):1547-8. Epub 2008/05/10. doi: 10.1093/bioinformatics/btn224. PubMed PMID: 18467348.

141. Maranville JC, Baxter SS, Torres JM, Di Rienzo A. Inter-ethnic differences in lymphocyte sensitivity to glucocorticoids reflect variation in transcriptional response. Pharmacogenomics J. 2013;13(2):121-9. Epub 2011/12/14. doi: 10.1038/tpj.2011.55. PubMed PMID: 22158329; PubMed Central PMCID: PMCPmc3774530.

142. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. 2015. 2015;67(1):48. Epub 2015-10-07. doi: 10.18637/jss.v067.i01.

143. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-25. http://CRAN.R-project.org/package=lmerTest 2015.

144. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A. 2003;100(16):9440-5. Epub 2003/07/29. doi: 10.1073/pnas.1530509100. PubMed PMID: 12883005; PubMed Central PMCID: PMCPmc170937.

145. Wei Y, Tenzen T, Ji H. Joint analysis of differential gene expression in multiple studies using correlation motifs. Biostatistics. 2015;16(1):31-46. Epub 2014/08/22. doi: 10.1093/biostatistics/kxu038. PubMed PMID: 25143368; PubMed Central PMCID: PMCPmc4263229.

146. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47. Epub 2015/01/22. doi: 10.1093/nar/gkv007. PubMed PMID: 25605792; PubMed Central PMCID: PMCPMC4402510.

147. Blischak JD, Tailleux L, Mitrano A, Barreiro LB, Gilad Y. Mycobacterial infection induces a specific human innate immune response. Sci Rep. 2015;5:16882. Epub 2015/11/21. doi: 10.1038/srep16882. PubMed PMID: 26586179; PubMed Central PMCID: PMCPMC4653619.

148. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Statist Soc B. 1995;57(1):289-300.

149. Tuoresmaki P, Vaisanen S, Neme A, Heikkinen S, Carlberg C. Patterns of genome-wide VDR locations. PLoS One. 2014;9(4):e96105. Epub 2014/05/03. doi: 10.1371/journal.pone.0096105. PubMed PMID: 24787735; PubMed Central PMCID: PMCPmc4005760.

150. Seuter S, Heikkinen S, Carlberg C. Chromatin acetylation at transcription start sites and vitamin D receptor binding regions relates to effects of 1alpha,25-dihydroxyvitamin D3 and histone deacetylase inhibitors on gene expression. Nucleic Acids Res. 2013;41(1):110-24. Epub 2012/10/25. doi: 10.1093/nar/gks959. PubMed PMID: 23093607; PubMed Central PMCID: PMCPmc3592476.

151. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMCPMC2705234.

152. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.

153. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res. 2012;22(9):1813-31. Epub 2012/09/08. doi: 10.1101/gr.136184.111. PubMed PMID: 22955991; PubMed Central PMCID: PMCPMC3431496.

154. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. Epub 2008/09/19. doi: 10.1186/gb-2008-9-9-r137. PubMed PMID: 18798982; PubMed Central PMCID: PMCPMC2592715.

155. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89. Epub 2010/06/02. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMCPMC2898526.

156. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207-10. Epub 2001/12/26. PubMed PMID: 11752295; PubMed Central PMCID: PMCPMC99122.

157. Chatila TA. Interleukin-4 receptor signaling pathways in asthma pathogenesis. Trends Mol Med. 2004;10(10):493-9. Epub 2004/10/07. doi: 10.1016/j.molmed.2004.08.004. PubMed PMID: 15464449.

158. LaPorte SL, Juo ZS, Vaclavikova J, Colf LA, Qi X, Heller NM, et al. Molecular and structural basis of cytokine receptor pleiotropy in the interleukin-4/13 system. Cell. 2008;132(2):259-72. Epub 2008/02/05. doi: 10.1016/j.cell.2007.12.030. PubMed PMID: 18243101; PubMed Central PMCID: PMCPMC2265076.

159. Schunkert H, Konig IR, Kathiresan S, Reilly MP, Assimes TL, Holm H, et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet. 2011;43(4):333-8. Epub 2011/03/08. doi: 10.1038/ng.784. PubMed PMID: 21378990; PubMed Central PMCID: PMCPmc3119261.

160. Dichgans M, Malik R, Konig IR, Rosand J, Clarke R, Gretarsdottir S, et al. Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. Stroke. 2014;45(1):24-36. Epub 2013/11/23. doi: 10.1161/strokeaha.113.002707. PubMed PMID: 24262325; PubMed Central PMCID: PMCPmc4112102.

161. Maranville JC, Luca F, Richards AL, Wen X, Witonsky DB, Baxter S, et al. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. PLoS Genet. 2011;7(7):e1002162. Epub 2011/07/14. doi: 10.1371/journal.pgen.1002162. PubMed PMID: 21750684; PubMed Central PMCID: PMCPmc3131293.

162. Flutre T, Wen X, Pritchard J, Stephens M. A statistical framework for joint eQTL analysis in multiple tissues. PLoS Genet. 2013;9(5):e1003486. Epub 2013/05/15. doi: 10.1371/journal.pgen.1003486. PubMed PMID: 23671422; PubMed Central PMCID: PMCPMC3649995.

163. Williams A, Henao-Mejia J, Harman CC, Flavell RA. miR-181 and metabolic regulation in the immune system. Cold Spring Harb Symp Quant Biol. 2013;78:223-30. Epub 2013/10/29. doi: 10.1101/sqb.2013.78.020024. PubMed PMID: 24163395.

164. Kelly B, O'Neill LA. Metabolic reprogramming in macrophages and dendritic cells in innate immunity. Cell Res. 2015;25(7):771-84. Epub 2015/06/06. doi: 10.1038/cr.2015.68. PubMed PMID: 26045163; PubMed Central PMCID: PMCPmc4493277.

165. Warburg O, Wind F, Negelein E. THE METABOLISM OF TUMORS IN THE BODY. J Gen Physiol. 1927;8(6):519-30. Epub 1927/03/07. PubMed PMID: 19872213; PubMed Central PMCID: PMCPmc2140820.

166. Freemerman AJ, Johnson AR, Sacks GN, Milner JJ, Kirk EL, Troester MA, et al. Metabolic reprogramming of macrophages: glucose transporter 1 (GLUT1)-mediated glucose metabolism drives a proinflammatory phenotype. J Biol Chem. 2014;289(11):7884-96. Epub 2014/02/05. doi: 10.1074/jbc.M113.522037. PubMed PMID: 24492615; PubMed Central PMCID: PMCPmc3953299.

167. Krawczyk CM, Holowka T, Sun J, Blagih J, Amiel E, DeBerardinis RJ, et al. Toll-like receptor-induced changes in glycolytic metabolism regulate dendritic cell activation. Blood. 2010;115(23):4742-9. Epub 2010/03/31. doi: 10.1182/blood-2009-10-249540. PubMed PMID: 20351312; PubMed Central PMCID: PMCPmc2890190.

168. Weichhart T, Costantino G, Poglitsch M, Rosner M, Zeyda M, Stuhlmeier KM, et al. The TSC-mTOR signaling pathway regulates the innate inflammatory response. Immunity. 2008;29(4):565-77. Epub 2008/10/14. doi: 10.1016/j.immuni.2008.08.012. PubMed PMID: 18848473.

169. Wang H, Brown J, Gu Z, Garcia CA, Liang R, Alard P, et al. Convergence of the mammalian target of rapamycin complex 1- and glycogen synthase kinase 3-beta-signaling pathways regulates the innate inflammatory response. J Immunol. 2011;186(9):5217-26. Epub 2011/03/23. doi: 10.4049/jimmunol.1002513. PubMed PMID: 21422248; PubMed Central PMCID: PMCPmc3137265.

170. Ohtani M, Nagai S, Kondo S, Mizuno S, Nakamura K, Tanabe M, et al. Mammalian target of rapamycin and glycogen synthase kinase 3 differentially regulate lipopolysaccharide-induced interleukin-12 production in dendritic cells. Blood. 2008;112(3):635-43. Epub 2008/05/22. doi: 10.1182/blood-2008-02-137430. PubMed PMID: 18492954; PubMed Central PMCID: PMCPMC2481549.

171. Wu S, Xia Y, Liu X, Sun J. Vitamin D receptor deletion leads to reduced level of IkappaBalpha protein through protein translation, protein-protein interaction, and post-translational modification. Int J Biochem Cell Biol. 2010;42(2):329-36. Epub 2009/11/26. doi: 10.1016/j.biocel.2009.11.012. PubMed PMID: 19931640; PubMed Central PMCID: PMCPMC2818560.

172. Shrestha N, Bahnan W, Wiley DJ, Barber G, Fields KA, Schesser K. Eukaryotic initiation factor 2 (eIF2) signaling regulates proinflammatory cytokine expression and bacterial invasion. J Biol Chem. 2012;287(34):28738-44. Epub 2012/07/05. doi: 10.1074/jbc.M112.375915. PubMed PMID: 22761422; PubMed Central PMCID: PMCPMC3436510.

173. Krutzik SR, Hewison M, Liu PT, Robles JA, Stenger S, Adams JS, et al. IL-15 links TLR2/1-induced macrophage differentiation to the vitamin D-dependent antimicrobial pathway. J Immunol. 2008;181(10):7115-20. Epub 2008/11/05. PubMed PMID: 18981132; PubMed Central PMCID: PMCPMC2678236.

174. Musikacharoen T, Matsuguchi T, Kikuchi T, Yoshikai Y. NF-kappa B and STAT5 play important roles in the regulation of mouse Toll-like receptor 2 gene expression. J Immunol. 2001;166(7):4516-24. Epub 2001/03/20. PubMed PMID: 11254708.

175. Pandiyan P, Yang XP, Saravanamuthu SS, Zheng L, Ishihara S, O'Shea JJ, et al. The role of IL-15 in activating STAT5 and fine-tuning IL-17A production in CD4 T lymphocytes. J Immunol. 2012;189(9):4237-46. Epub 2012/09/21. doi: 10.4049/jimmunol.1201476. PubMed PMID: 22993203; PubMed Central PMCID: PMCPMC3647038.

176. Ramagopalan SV, Dyment DA, Cader MZ, Morrison KM, Disanto G, Morahan JM, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Ann

Neurol. 2011;70(6):881-6. Epub 2011/12/23. doi: 10.1002/ana.22678. PubMed PMID: 22190362.

177. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet. 2009;41(7):824-8. Epub 2009/06/16. doi: 10.1038/ng.396. PubMed PMID: 19525955.

178. Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9. Epub 2011/08/13. doi: 10.1038/nature10251. PubMed PMID: 21833088; PubMed Central PMCID: PMCPMC3182531.

179. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012;491(7422):119-24. Epub 2012/11/07. doi: 10.1038/nature11582. PubMed PMID: 23128233; PubMed Central PMCID: PMCPMC3491803.

180. Wang TJ, Zhang F, Richards JB, Kestenbaum B, van Meurs JB, Berry D, et al. Common genetic determinants of vitamin D insufficiency: a genome-wide association study. Lancet. 2010;376(9736):180-8. Epub 2010/06/15. doi: 10.1016/s0140-6736(10)60588-0. PubMed PMID: 20541252; PubMed Central PMCID: PMCPMC3086761.

181. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65. Epub 2012/11/07. doi: 10.1038/nature11632. PubMed PMID: 23128226; PubMed Central PMCID: PMCPMC3498066.

182. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5(6):e1000529. Epub 2009/06/23. doi: 10.1371/journal.pgen.1000529. PubMed PMID: 19543373; PubMed Central PMCID: PMCPMC2689936.

183. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-59. Epub 2000/06/03. PubMed PMID: 10835412; PubMed Central PMCID: PMCPMC1461096.

184. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353-8. Epub 2012/04/12. doi: 10.1093/bioinformatics/bts163. PubMed PMID: 22492648; PubMed Central PMCID: PMCPMC3348564.

185. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.

186. He X, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. Am J Hum Genet. 2013;92(5):667-80. Epub 2013/05/07. doi: 10.1016/j.ajhg.2013.03.022. PubMed PMID: 23643380; PubMed Central PMCID: PMCPMC3644637.

187. Medicine Io. Dietary reference intakes for calcium and vitamin D. Washington (DC): National Academies Press; 2011.

188. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. Epub 2012/09/08. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMCPMC3439153.

189. Kariuki SN, Blischak JD, Nakagome S, Witonsky DB, Di Rienzo A. Patterns of transcriptional response to 1,25-dihydroxyvitamin D3 and bacterial lipopolysaccharide in primary human monocytes. bioRxiv. 2015. doi: 10.1101/030759.

190. Angus L, Moleirinho S, Herron L, Sinha A, Zhang X, Niestrata M, et al. Willin/FRMD6 expression activates the Hippo signaling pathway kinases in mammals and antagonizes oncogenic YAP. Oncogene. 2012;31(2):238-50. Epub 2011/06/15. doi: 10.1038/onc.2011.224. PubMed PMID: 21666719.

191. Wang W, Soto H, Oldham ER, Buchanan ME, Homey B, Catron D, et al. Identification of a novel chemokine (CCL28), which binds CCR10 (GPR2). J Biol Chem. 2000;275(29):22313-23. Epub 2000/04/27. doi: 10.1074/jbc.M001461200. PubMed PMID: 10781587.

192. Pan J, Kunkel EJ, Gosslar U, Lazarus N, Langdon P, Broadwell K, et al. A novel chemokine ligand for CCR10 and CCR3 expressed by epithelial cells in mucosal tissues. J Immunol. 2000;165(6):2943-9. Epub 2000/09/07. PubMed PMID: 10975800.

193. Hieshima K, Ohtani H, Shibano M, Izawa D, Nakayama T, Kawasaki Y, et al. CCL28 has dual roles in mucosal immunity as a chemokine with broad-spectrum antimicrobial activity. J Immunol. 2003;170(3):1452-61. Epub 2003/01/23. PubMed PMID: 12538707.

194. Craig AW, Haghighat A, Yu AT, Sonenberg N. Interaction of polyadenylate-binding protein with the eIF4G homologue PAIP enhances translation. Nature. 1998;392(6675):520-3. Epub 1998/04/21. doi: 10.1038/33198. PubMed PMID: 9548260.

195. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet. 2006;38(3):320-3. Epub 2006/01/18. doi: 10.1038/ng1732. PubMed PMID: 16415884.

196. Cauchi S, El Achhab Y, Choquet H, Dina C, Krempler F, Weitgasser R, et al. TCF7L2 is reproducibly associated with type 2 diabetes in various ethnic groups: a global meta-analysis. J Mol Med (Berl). 2007;85(7):777-82. Epub 2007/05/04. doi: 10.1007/s00109-007-0203-4. PubMed PMID: 17476472.

197. Savic D, Ye H, Aneas I, Park SY, Bell GI, Nobrega MA. Alterations in TCF7L2 expression define its role as a key regulator of glucose metabolism. Genome Res. 2011;21(9):1417-25. Epub 2011/06/16. doi: 10.1101/gr.123745.111. PubMed PMID: 21673050; PubMed Central PMCID: PMCPMC3166827.

198. Wei G, Abraham BJ, Yagi R, Jothi R, Cui K, Sharma S, et al. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity. 2011;35(2):299-311. Epub 2011/08/27. doi: 10.1016/j.immuni.2011.08.007. PubMed PMID: 21867929; PubMed Central PMCID: PMCPMC3169184.

199. Ho IC, Tai TS, Pai SY. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. Nat Rev Immunol. 2009;9(2):125-35. Epub 2009/01/20. doi: 10.1038/nri2476. PubMed PMID: 19151747; PubMed Central PMCID: PMCPMC2998182.

200. Akira S, Isshiki H, Sugita T, Tanabe O, Kinoshita S, Nishio Y, et al. A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. Embo j. 1990;9(6):1897-906. Epub 1990/06/01. PubMed PMID: 2112087; PubMed Central PMCID: PMCPMC551896.

201. Pope RM, Leutz A, Ness SA. C/EBP beta regulation of the tumor necrosis factor alpha gene. J Clin Invest. 1994;94(4):1449-55. Epub 1994/10/01. doi: 10.1172/jci117482. PubMed PMID: 7929820; PubMed Central PMCID: PMCPMC295278.

202. Wedel A, Sulski G, Ziegler-Heitbrock HW. CCAAT/enhancer binding protein is involved in the expression of the tumour necrosis factor gene in human monocytes. Cytokine. 1996;8(5):335-41. Epub 1996/05/01. PubMed PMID: 8726660.

203. Poli V. The role of C/EBP isoforms in the control of inflammatory and native immunity functions. J Biol Chem. 1998;273(45):29279-82. Epub 1998/10/29. PubMed PMID: 9792624.

204. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5. Epub 2013/05/30. doi: 10.1038/ng.2653. PubMed PMID: 23715323; PubMed Central PMCID: PMCPMC4010069.

205. Colonna M, Facchetti F. TREM-1 (triggering receptor expressed on myeloid cells): a new player in acute inflammatory responses. J Infect Dis. 2003;187 Suppl 2:S397-401. Epub 2003/06/07. doi: 10.1086/374754. PubMed PMID: 12792857.

206. Ramagopalan SV, Heger A, Berlanga AJ, Maugeri NJ, Lincoln MR, Burrell A, et al. A ChIP-seq defined genome-wide map of vitamin D receptor binding: associations with disease and evolution. Genome Res. 2010;20(10):1352-60. Epub 2010/08/26. doi: 10.1101/gr.107920.110. PubMed PMID: 20736230; PubMed Central PMCID: PMCPMC2945184.

207. Heikkinen S, Vaisanen S, Pehkonen P, Seuter S, Benes V, Carlberg C. Nuclear hormone 1alpha,25-dihydroxyvitamin D3 elicits a genome-wide shift in the locations of VDR chromatin occupancy. Nucleic Acids Res. 2011;39(21):9181-93. Epub 2011/08/19. doi: 10.1093/nar/gkr654. PubMed PMID: 21846776; PubMed Central PMCID: PMCPMC3241659.

208. Handel AE, Sandve GK, Disanto G, Berlanga-Taylor AJ, Gallone G, Hanwell H, et al. Vitamin D receptor ChIP-seq in primary CD4+ cells: relationship to serum 25-hydroxyvitamin D levels and autoimmune disease. BMC Med. 2013;11:163. Epub 2013/07/16. doi: 10.1186/1741-7015-11-163. PubMed PMID: 23849224; PubMed Central PMCID: PMCPMC3710212.

209. Janesick A, Abbey R, Chung C, Liu S, Taketani M, Blumberg B. ERF and ETV3L are retinoic acid-inducible repressors required for primary neurogenesis. Development. 2013;140(15):3095-106. Epub 2013/07/05. doi: 10.1242/dev.093716. PubMed PMID: 23824578.

210. George M, Ying G, Rainey MA, Solomon A, Parikh PT, Gao Q, et al. Shared as well as distinct roles of EHD proteins revealed by biochemical and functional comparisons in mammalian cells and C. elegans. BMC Cell Biol. 2007;8:3. Epub 2007/01/20. doi: 10.1186/1471-2121-8-3. PubMed PMID: 17233914; PubMed Central PMCID: PMCPMC1793994.

211. Sharma M, Naslavsky N, Caplan S. A role for EHD4 in the regulation of early endosomal transport. Traffic. 2008;9(6):995-1018. Epub 2008/03/12. doi: 10.1111/j.1600-0854.2008.00732.x. PubMed PMID: 18331452; PubMed Central PMCID: PMCPMC2795359.

212. Wang J, Li Y, Zhang M, Liu Z, Wu C, Yuan H, et al. A zinc finger HIT domain-containing protein, ZNHIT-1, interacts with orphan nuclear hormone receptor Rev-erbbeta and removes Rev-erbbeta-induced inhibition of apoCIII transcription. Febs j. 2007;274(20):5370-81. Epub 2007/09/26. doi: 10.1111/j.1742-4658.2007.06062.x. PubMed PMID: 17892483.

213. Zhao B, Li L, Lei Q, Guan KL. The Hippo-YAP pathway in organ size control and tumorigenesis: an updated version. Genes Dev. 2010;24(9):862-74. Epub 2010/05/05. doi: 10.1101/gad.1909210. PubMed PMID: 20439427; PubMed Central PMCID: PMCPMC2861185.

214. Ungvari I, Hullam G, Antal P, Kiszel PS, Gezsi A, Hadadi E, et al. Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. PLoS One. 2012;7(3):e33573. Epub 2012/03/21. doi: 10.1371/journal.pone.0033573. PubMed PMID: 22432035; PubMed Central PMCID: PMCPMC3303848.

215. Hong MG, Reynolds CA, Feldman AL, Kallin M, Lambert JC, Amouyel P, et al. Genome-wide and gene-based association implicates FRMD6 in Alzheimer disease. Hum Mutat. 2012;33(3):521-9. Epub 2011/12/23. doi: 10.1002/humu.22009. PubMed PMID: 22190428; PubMed Central PMCID: PMCPMC3326347.

216. Zhang R, Chu M, Zhao Y, Wu C, Guo H, Shi Y, et al. A genome-wide gene-environment interaction analysis for tobacco smoke and lung cancer susceptibility. Carcinogenesis. 2014;35(7):1528-35. Epub 2014/03/25. doi: 10.1093/carcin/bgu076. PubMed PMID: 24658283; PubMed Central PMCID: PMCPMC4076813.

217. Yang H, Yuan W, Wang Y, Zhu C, Liu B, Yang D, et al. ZNF649, a novel Kruppel type zinc-finger protein, functions as a transcriptional suppressor. Biochem Biophys Res Commun. 2005;333(1):206-15. Epub 2005/06/14. doi: 10.1016/j.bbrc.2005.05.101. PubMed PMID: 15950191.

218. Karin M, Liu Z, Zandi E. AP-1 function and regulation. Curr Opin Cell Biol. 1997;9(2):240-6. Epub 1997/04/01. PubMed PMID: 9069263.

219. Shaulian E, Karin M. AP-1 as a regulator of cell life and death. Nat Cell Biol. 2002;4(5):E131-6. Epub 2002/05/04. doi: 10.1038/ncb0502-e131. PubMed PMID: 11988758.

220. Egerer J, Emmerich D, Fischer-Zirnsak B, Chan WL, Meierhofer D, Tuysuz B, et al. GORAB Missense Mutations Disrupt RAB6 and ARF5 Binding and Golgi Targeting. J Invest Dermatol. 2015;135(10):2368-76. Epub 2015/05/23. doi: 10.1038/jid.2015.192. PubMed PMID: 26000619.

221. Yildirim Y, Tolun A, Tuysuz B. The phenotype caused by PYCR1 mutations corresponds to geroderma osteodysplasticum rather than autosomal recessive cutis laxa type 2. Am J Med Genet A. 2011;155a(1):134-40. Epub 2011/01/05. doi: 10.1002/ajmg.a.33747. PubMed PMID: 21204221.

222. Wang F, Wang L, Jiang H, Chang X, Pan J. Inhibition of PCSK6 may play a protective role in the development of rheumatoid arthritis. J Rheumatol. 2015;42(2):161-9. Epub 2014/12/01. doi: 10.3899/jrheum.140435. PubMed PMID: 25433529.

223. Marcus J, Novembre J. Geography of Genetic Variants (GGV) v0.1. http://www.popgen.uchicago/ggv/.