

Western Kentucky University
TopSCHOLAR®

Honors College Capstone Experience/Thesis
Projects

Honors College at WKU

Fall 12-14-2013

Two Genetic Effects at the IRF5/ TNPO3 Locus are Independently Associated with the Development of Specific Lupus Symptoms

Samantha Hawtrey

Western Kentucky University, samantha.hawtrey943@topper.wku.edu

Follow this and additional works at: http://digitalcommons.wku.edu/stu_hon_theses

 Part of the [Biology Commons](#)

Recommended Citation

Hawtrey, Samantha, "Two Genetic Effects at the IRF5/ TNPO3 Locus are Independently Associated with the Development of Specific Lupus Symptoms" (2013). *Honors College Capstone Experience/Thesis Projects*. Paper 429.
http://digitalcommons.wku.edu/stu_hon_theses/429

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Honors College Capstone Experience/Thesis Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.

TWO GENETIC EFFECTS AT THE *IRF5/TNPO3* LOCUS ARE INDEPENDENTLY
ASSOCIATED WITH THE DEVELOPMENT OF SPECIFIC LUPUS SYMPTOMS

A Capstone Experience/Thesis Project

Presented in Partial Fulfillment of the Requirements for

the Degree Bachelor of Science with

Honors College Graduate Distinction at Western Kentucky University

By

Samantha F. Hawtrey

Western Kentucky University
2014

CE/T Committee:

Professor Rodney King, Advisor

Professor Leah C Kottyan

Mister Derick Strode

Approved by:

Advisor
Department of Biology

Copyright by
Samantha F Hawtrey
2014

ABSTRACT

Systemic Lupus Erythematosus (SLE) is an autoimmune disorder that can affect every tissue in the body. Single nucleotide polymorphisms (SNPs) in the *IRF5* and *TNPO3* genes are statistically associated with the development of SLE. My research identified correlations between *IRF5/TNPO3* SNPs and specific lupus symptoms. Logistic regression analyses were conducted using 101 genetic variants in the *IRF5/TNPO3* region that were genotyped in over 6,000 lupus patients of different ethnicities, with admixture covariates applied. Three clinical phenotypes displayed significant correlation ($p < 1.6 \times 10^{-5}$) in subjects of European ancestry. For each of these phenotypes, a step-wise conditional analysis was conducted using two lupus associated single nucleotide polymorphisms (SNPs) at this genetic loci. In Europeans, lupus disease onset ($p\text{-value}_{\text{EU}} = 2.44 \times 10^{-16}$, $\text{OR} = 0.67^*$) and the presence of anti-Ro ($p\text{-value}_{\text{EU}} = 2.09 \times 10^{-7}$, $\text{OR} = 0.67$) and anti-dsDNA ($p\text{-value}_{\text{EU}} = 4.15 \times 10^{-7}$, $\text{OR} = 0.75$) antibodies were associated with SNPs in the *IRF5/TNPO3* genes. SNPs in the *IRF5* promoter and those spanning *IRF5* and *TNPO3* were both associated with disease onset. The presence of anti-Ro and anti-dsDNA antibodies is only associated with SNPs in the *IRF5* promoter. Genetic variants at the *IRF5/TNPO3* locus are associated with lupus disease onset and production of anti-dsDNA and anti-Ro antibodies in lupus patients. SNPs in the promoter region of

IRF5 (associated with rs4728142) and SNPs spanning the *IRF5* and *TNPO3* genes (associated with rs12534421) contribute independently to these symptoms.

Keywords: Lupus, *IRF5*, serology, sub-phenotype, SNP

Dedicated to Derick Strobe, who inspired me to try research in the first place.

ACKNOWLEDGEMENTS

I would like to thank Dr. Leah Kottyan and all the wonderful members of CAGE, who have played an integral part in making this research possible. Their insightful contributions have had no small impact on the outcome of this paper.

I am grateful to Derick Strode, who pushed me to get involved with research from the very start of my time at WKU. He never failed to set the standards high, but always equipped me to rise to the challenge. Thank you for seeing potential when I couldn't.

I would also like Dr. Rodney King, Dr. Claire Rinehart, and Dr. Ajay Srivastava, who have invested countless hours in developing and encouraging my research. They taught me how to think critically and inquisitively, along with imparting a wealth of knowledge and laboratory skills that have proven to be invaluable!

And of course, many thanks goes to my family, who not only loved me when I was overwhelmed with the research, but who also went the extra mile by striving to read and even understand my work! Their support has been invaluable.

VITA

Gatton Academy of Mathematics and Science, Bowling Green, Kentucky.....2010-2012

- ❖ One of only sixty recipients of a full scholarship to attend this residential, exclusive dual-enrollment academy for Kentucky's top math and science students to complete high school requirements while being an integrated freshman and sophomore at Western Kentucky University
- ❖ Siemens Competition Semifinalist: one of 300 entrants (out of 1500) to achieve this status in the most prestigious high school research competition in the nation
- ❖ National Merit Semifinalist
- ❖ Global Relief and Sustainable Societies Club president

Cincinnati Children's Hospital.....2012

- ❖ Conducted sub-phenotypic genetic variant research of lupus patients in a paid, 10 week long summer research fellowship in the Department of Rheumatology.

Western Kentucky University.....2010-2014

- ❖ President's Scholar
- ❖ GPA: 3.93
- ❖ Study abroad: Costa Rican ecology and diversity, January 2011. Partners in Caring: Medicine in Kenya, January 2014.
- ❖ Howard Hughes Medical Institute Genome Discovery and Exploration Program (2010-2011): isolated and characterized a novel bacteriophage, and annotated the genome of another one. This research was presented at the US Naval Academy (June 2011), WKU student research conference (April 2011), and at the HHMI's SEA symposium in Virginia (June 2011). Faculty mentors: Dr. Rodney King, Dr. Claire Rinehart
- ❖ *Drosophila Melanogaster* scalloped gene research (2011-2012): studied mutations in the sd58d (scalloped) gene in *Drosophila* (fruit flies). This research was presented at the WKU student research conference. Faculty mentor: Dr. Ajay Srivastava

PUBLICATIONS

Hawtrey S, Lovell L, King R. 2011. Isolation, Characterization, and Annotation: the Search for Novel Bacteriophage Genomes. In: DeVries SH, Golowich D, Thurston EPS, Haderlie S, Tenenbaum H, editors. Journal of Experimental Secondary Science. Available from: www.jes2s.com

FIELDS OF STUDY

Major Field 1: Biology

Major Field 2: Chemistry

TABLE OF CONTENTS

	<u>Page</u>
Abstract.....	ii
Dedication.....	iv
Acknowledgements.....	v
Vita.....	vi
List of Figures.....	viii
Chapters:	
1. Preface.....	1
2. Introduction.....	9
3. Materials and Methods.....	11
4. Results.....	14
5. Discussion.....	19
References.....	21

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. SLE criteria.....	2
2. SLE case ethnicities.....	3
3. SNPs associated with SLE in the <i>IRF5/TNPO3</i> genes.....	5
4. Age of onset conditional analysis.....	16
5. Anti-Ro conditional analysis.....	17
6. Anti-dsDNA conditional analysis.....	18

CHAPTER 1

PREFACE

During the summer of 2012, I was selected to be one of 120 students involved in a Summer Undergraduate Research Fellowship at Cincinnati Children's Hospital. It was a ten week, paid internship where students conducted research in the extensive research facilities of the third best children's hospital in the nation. Students in this program were placed in laboratories addressing almost every health issue available, from cancer to development to x-ray imaging and cardiac diseases. Because of my prior experience in bioinformatics analysis, I was chosen to analyze the genetic data of lupus patients in the rheumatology department.

Systemic Lupus Erythematosus (SLE or lupus) is a systemic autoimmune disorder that can affect every tissue in the human body. The body's immune system is designed to differentiate between self (cells of the body) and non-self (cells from the outside world, which could possibly be harmful). When functioning correctly, immune cells attack non-self cells while ignoring self cells. However, in an autoimmune disease like SLE, the immune system begins attacking the cells of the body.

Lupus is a heterogenous disease that can look very different in different people. There is no laboratory test for lupus, instead, there are eleven criteria that lead to a clinical diagnosis of lupus. Any four of the eleven must be manifested for a lupus diagnosis according to the American College of Rheumatology. They are: hematologic

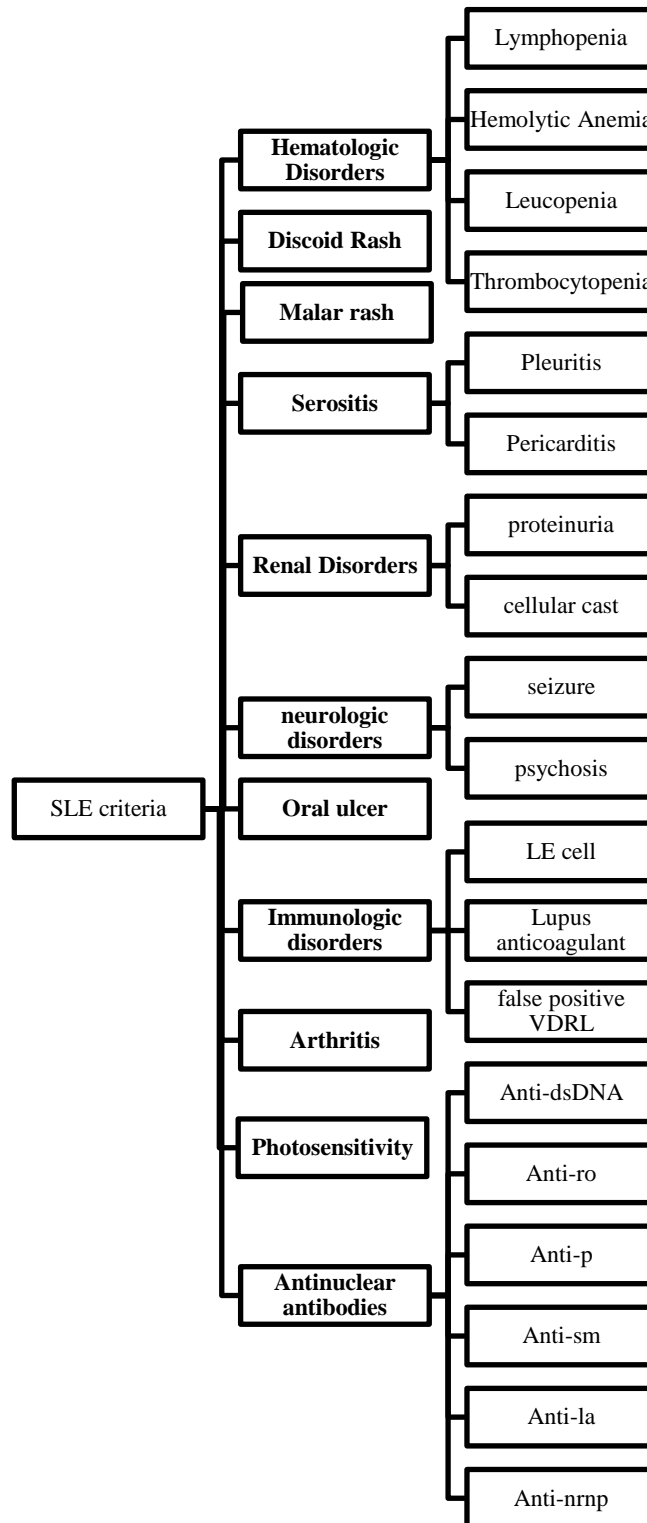


Figure 1. A chart of the thirty phenotypic characteristics of SLE. The bold items are the eleven main criteria defined by the American Rheumatology. In addition to these thirty characteristics, the age of onset of lupus symptoms was also studied.

(blood) disorders (anemia, leukopenia, lymphopenia, or thrombocytopenia), discoid rash, malar rash, serositis (inflammation of the lining around the lungs or heart), photosensitivity, renal (kidney) disorders, arthritis, neurologic disorders, oral ulcers, immunologic disorders (presence of certain autoantibodies, and abnormal antinuclear antibodies). These 11 symptoms have been broken into 30 subgroups, for which clinical data can be collected in lupus cases and controls (Figure 1). To study the genetic determinates of these sub-phenotypes, we used data from the Large Lupus Association Study 2 (LLAS2). Twenty-four different investigators contributed samples of DNA in addition to clinical information. Altogether, the LLAS2 dataset contained genetic and phenotypic data of over 6,000 people with lupus collected by researchers around the globe. Data from subjects of four different ancestral origins was studied: European American, Hispanic, African American, and Asian American (Figure 2). For each ancestry, covariates were calculated to account for admixture. Admixture is the concept

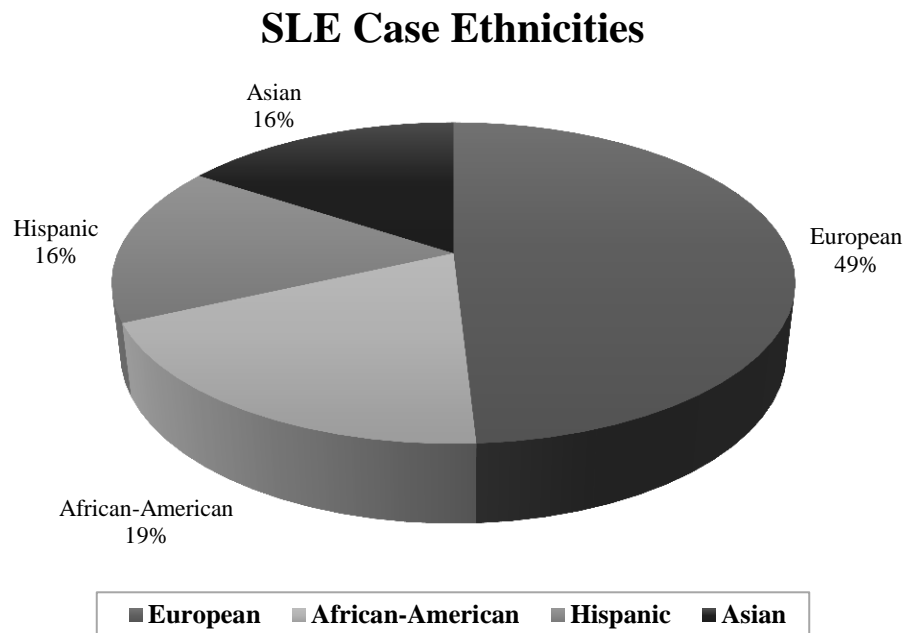


Figure 2. Percentage of cases of different ethnicities analyzed from the LLAS2 data set.

that individuals are rarely ever purely one race, but instead have a genetic history from multiple ancestries. The covariates determine and account for mixed genetics in each individual in the study.

Single Nucleotide Polymorphisms (SNPs) are variations in the genome that occur at a single base pair. These genetic variants are manifested in at least 5% of the population. The variations can be silent, causing no change in the gene's protein structure, or they can change the promoter region, causing the gene to be over or under expressed, or the amino acid sequence, impacting the entire function of the resulting protein.

There is no clearly defined single cause of lupus onset, rather, it is a result of both environmental and genetic causes. SNPs in over 45 genes have been repeatedly associated with lupus development, or etiology. Two of the most commonly associated loci includes the Interferon Regulatory Factor 5 (*IRF5*) gene and the nearby Transportin 3 (*TNPO3*) gene. The *IRF5* gene is especially noteworthy because it regulates chemicals called interferons. Interferons act as signals in the body that activate the immune response to viral infections. In lupus, interferons have been demonstrated to drive the autoimmune response and the pathology of SLE. In the LLAS2 study subjects, 101 genetic variants were genotyped in the *IRF5/TNPO3* gene region. Case-control analysis of these SNPs showed significant association with lupus development. In fact, the group at the Cincinnati Children's Hospital established that there are SNPs that are highly associated with lupus development in the promoter of *IRF5* and in a region spanning the *IRF5* and *TNPO3* genes (p-value of less than 10^{-8}).

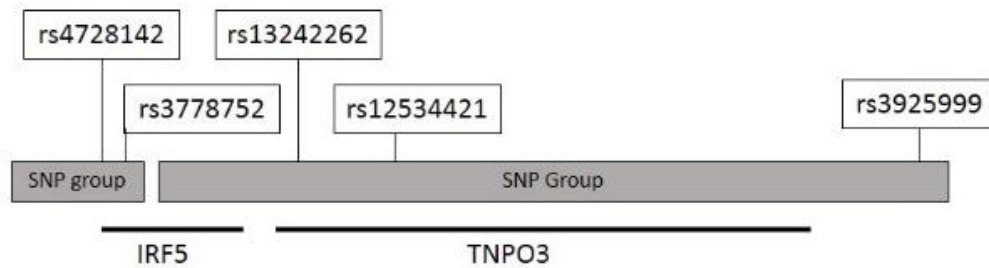


Figure 3. SNP locations in reference to the *IRF5* and *TNPO3* genes. The black bars show the gene locations, and the two gray boxes are two groups of SNPs that are in high linkage within the gray region, but display little association with the other SNP group. The relative locations of SNPs connected to SLE symptoms are shown.

SNPs in a certain area tend to be found together, an occurrence known as linkage disequilibrium (LD). In the case of lupus development, the most associated SNPs are located in the promoter region of *IRF5*, and were closely linked with each other (Figure 3). When the association of the SNP with the lowest p-value (rs4728142, located in the promoter of *IRF5*) was accounted for in a conditional analysis, the only remaining genetic association was found in a group of SNPs that span the *IRF5* and *TNPO3* genes. Upon further conditional analysis, removing the association of the next-most associated SNP (rs12534421) caused all association to be eliminated. This suggested that all significant association with lupus development was linked with two main SNPs in the *IRF5/TNPO3* area.

Although studies have shown the association of SNPs in the *IRF5* locus with lupus development, there has been little analysis of the possible association with specific symptoms in lupus. My research sought to identify correlations between SNPs and the development of lupus sub-phenotypes. I used the statistical software package PLINK version 1.07 to analyze the LLAS2 data. PLINK is designed to perform in-depth analyses of SNP associations, and it enabled me to conduct a sub-phenotypic logistic regression analysis (see Chapter 3) of SNPs in the *IRF5/TNPO3* region, identifying connections

between displayed lupus symptoms and the SNPs present in this gene area. After formatting the data appropriately and including covariates to account for differences in genetic admixture, we determined the p-value for each SNP's association with every clinical phenotype. In a logistic regression, the p-value gives the probability that the frequency of the minor allele of a SNP is the same in the people with a sub-phenotype, such as the people with anti-dsDNA, and the people without the sub-phenotype. The logistic regression analysis was first done on subjects of European ancestry, then on Asians, African Americans, and Hispanics. The accepted alpha error rate is 0.05, but because we performed many independent tests, a multiple testing correction had to be used. This correction takes into account the number of SNPs compared, as well as the number of clinical phenotypes analyzed. Using this correction, a significant p-value was determined to be 1.6×10^{-5} .

In European Americans, the logistic analysis identified three sub-phenotypes that were significantly associated with SNPs in and around the *IRF5/TNPO3* genes (p-value less than 1.6×10^{-5}). These were the age of onset of lupus (how old a person is when they develop lupus), the presence of anti-Ro autoantibodies, and the presence of anti-dsDNA autoantibodies. Anti-Ro autoantibodies target the Ro ribonucleoprotein, which is normally found in the nucleolus of the cell, while anti-dsDNA autoantibodies target double stranded DNA (dsDNA). These autoantibodies are not normally formed in absence of an autoimmune response. In African Americans and Asian Americans, there was a statistically significant association to age of onset, but not to the other symptoms.

A conditional analysis (see Chapter 3) was conducted for each of the three statistically associated symptoms, similar to the conditional analyses that demonstrated

the two independent SNP groups that were associated with SLE development. This analysis revealed an association with specific SNPs in the *IRF5/TNPO3* area. For age of onset of lupus, there was a statistically significant association with the SNP rs12534421. Nearly all phenotype association was removed when the impact of this SNP and all associated variants were removed in the analysis. When the symptom association with the next significant SNP, rs4728142, was removed, no SNP association with the age of onset remained. This demonstrated that all associated SNPs were linked to SNPs rs4728142 and rs12534421.

For the presence of anti-Ro and anti-dsDNA autoantibodies, the conditional analysis showed association to rs4728142, the primary variant associated with SNPs in the promoter region of *IRF5*. After conditioning on (removing the impact of all SNPs associated with) rs4728142, there was still association with a few SNPs spanning *IRF5* and *TNPO3* genes (those variants in linkage disequilibrium with rs12534421), but most of the association was with genetic variants in the promoter.

The fact that different groups of SNPs in the *IRF5/TNPO3* region were statistically associated with the development of different lupus symptoms provides support to the theory that there are two independent groups of variants in this region, and that they play separate roles in the development of SLE. This will allow future studies to specifically focus on the *IRF5/TNPO3* region. My results also support previous studies of these two SNP groups and could refine the study of the biological processes that are impacted by these variants by directing the focus towards each of the SNP groups separately.

Additionally, our work to identify variants associated with the development of specific symptoms in people with lupus has immediate therapeutic implications. Lupus is a relatively rare disease, and doctors are unlikely to prophylactically treat healthy people who are genetically predisposed to lupus. Once people have lupus, however, there are different therapeutic options that vary in their aggressiveness. This study may also contribute to the type of personalized medicine that is the future of health care.

At the end of the summer, I was able to present my research at the University of Cincinnati's Undergraduate Research Symposium, where I earned an honorable mention. In January 2013, I returned to the Rheumatology Department to complete additional data analyses, instruct a new graduate student in programming with PLINK, and assemble cohesive, straightforward instructions for future sub-phenotypic, conditional analysis research in this field.

CHAPTER 2

INTRODUCTION

Systemic Lupus Erythematosus (SLE or lupus) is an autoimmune disorder affecting every organ in the body [1]. Patients must display four of eleven clinical findings for official diagnosis according to the American College of Rheumatology [5, 6]. One in every 5,000 people will be diagnosed with lupus, with 90% of them being female (Figure 3). African Americans and Asians are three to four times more likely to develop lupus than persons of European descent, and African American, Asian, and Hispanic ancestries have a greater mortality rate from this disease [2-4]. Without sophisticated medical care, SLE has a 50% mortality rate [7].

Genetic factors play a significant role in the etiology of lupus [8]. Previous studies have demonstrated the effect of genetic variants on the development of specific symptoms in lupus patients [9-15]. The Interferon Regulatory Factor 5 (*IRF5*) functions within the interferon inflammatory pathway and is critical for lupus pathogenesis [15-19]. Single nucleotide polymorphisms (SNPs) within the *IRF5* and the nearby Transportin 3 (*TNPO3*) genes are highly associated with the etiology of lupus, but genetic variant associations with specific clinical phenotypes of SLE have not yet been identified [20-36]. Prior studies indicate that genetic association occurs through two independent groups of linked SNPs at the *IRF5/TNPO3* locus [37]. Variants in the promoter region are present across ancestries and are accounted for by the rs4728142 SNP, while variants

spanning the *IRF5/TNPO3* region are present only in populations with European admixture and are strongly linked to the SNP rs12534421.

The purpose of this study was to determine if either of these two genetic effects at the *IRF5/TNPO3* locus are statistically associated with specific clinical phenotypes in patients with lupus. The study also investigated how sub-phenotypic associations in *IRF5/TNPO3* vary across different ancestries. Our data indicate that the two effects are independently associated with three different lupus phenotypes in European Americans, but only one effect is evident in African Americans, Asians, and Hispanics.

CHAPTER 3

MATERIALS AND METHODS

Subjects

The study included 3,926 European, 1,257 Asian, 1,287 Hispanic, and 1,524 African-American SLE patients, all of whom met the American College of Rheumatology (ACR) SLE classification criteria. Phenotypic data and patient DNA were collected with informed consent and IRB approval by several institutions around the globe. Thirty-one clinical phenotypes were analyzed, each of which fall into one of the following categories: age of onset of lupus, presence of certain autoantibodies, false positive syphilis test, discoid rash, malar rash, renal disorders, neurologic disorders, oral ulcers, photosensitivity, and hematologic disorders.

Genotyping

Patient DNA was genotyped on an iSCAN array genotyping instrument using the Illumina custom bead system. One hundred and one lupus susceptibility loci in the *IRF5/TNPO3* region were identified for each patient. SNPs were located in a 437,000 base pair region, ranging from 167 kbp upstream of *IRF5* to 61 kbp downstream of *TNPO3*. Admixture markers, which indicate ancestral origins, were genotyped and analyzed for each sample to determine admixture from five ancestries: European American, African American, Asian and Asian American, and Hispanic American [38-41].

Statistical analysis

Two main types of analyses were used in this study. The first was a logistic regression analysis of all SLE symptoms. This analysis compared the presence of different risk alleles with the presence of each symptom, identifying any association. It then determined a p-value, which was the probability that the observed association could randomly occur. The logistic regression served to identify lupus symptoms for which SNPs in the *IRF5/TNPO3* region were significantly associated.

A sub-phenotypic logistic regression analysis was conducted using PLINK version 1.07, a program designed to analyze SNPs and their associations with phenotypic data. Admixture covariates were included in the program to account for variances in ancestry and each ancestry was individually assessed. Using the most stringent Bonferroni multiple testing correction ($p = 0.05/(\text{number of variants} \times \text{number of phenotypes}) = 0.05/(101 \times 31)$), a significant p-value was determined to be 1.6×10^{-5} . Additional label-swapping iterative permutation testing showed that empirically significant p-values for each SNP are actually closer to 10^{-4} . This test swaps the symptom affectedness status (from unaffected to affected, or vice versa) for a random selection of patients, and checks to see if SNP association still exists. It repeats this process several thousand times, and uses the results to calculate the probability of SNP association randomly occurring. The sub phenotypic logistic regression analysis of lupus patients identified three symptoms with significant p-values. The additional permutation testing confirmed that no other symptoms were significantly associated.

Aside from the logistic analysis, the second main test used in this study is a conditional analysis. This test determines how many SNPs are actually associated with a

specific symptom. For a single symptom, this analysis removes the most associated SNP, and all of the SNPs that are linked with it. It is then possible to see if any SNPs not connected with the first SNP are associated with the symptom. For a step-wise conditional analysis, the symptom is “conditioned” (the association is removed) for the most associated SNP, followed by the next remaining SNP, then both SNPs simultaneously, in a series of steps that result in the removal of all significant SNP association. The conditional analysis shows the top two most associated SNPs that are not linked to each other. It could be used to identify more associations, but after the association of the top two are removed, very little SNP association remains.

A conditional analysis was conducted in PLINK v 1.07 to determine which SNPs were statistically associated with these three phenotypes [42]. Phenotypic associations were conditioned step-wise using the SNPs that account for the two linked SNP groups in the *IRF5/TNPO3* area: the SNP rs4728142, which is associated with genetic variants in the promoter region of *IRF5*, and the SNP rs12534421, which is associated with variants spanning *IRF5* and *TNPO3*. In cases where the SNPs most associated with lupus development differed from those most associated with the specific symptoms, the SNPs most associated with the symptoms were used for the conditional analysis. This was the case for anti-Ro and anti-dsDNA. In these cases, conditional analyses were conducted with the two main SNPs for lupus development in addition to the two most associated symptomatic SNPs.

CHAPTER 4

RESULTS

The three clinical phenotypes that demonstrate statistically significant association with variants in *IRF5* and *TNPO3* are: age of onset of lupus (p value_{EU} = 2.44x10⁻¹⁶, Odds Ratio = 0.67*), the presence of anti-Ro antibodies (p value_{EU} = 2.09x10⁻⁷, Odds Ratio = 0.67), and the presence of anti-dsDNA antibodies (p value_{EU} = 4.15x10⁻⁷, Odds Ratio = 0.75). Of the three phenotypes, age of onset of lupus was the only one that was significantly associated across all three ethnicities (European American, African American, and Asian). Testing revealed no genetic association with any symptom in Hispanics. The presence of anti-Ro antibodies was associated with variants in Europeans only, and the presence of anti-dsDNA antibodies displayed SNP association in Europeans and African Americans. In Europeans, the SNP rs3778752 had the lowest p value for the presence of both anti-Ro and anti-dsDNA antibodies. For age of onset of lupus, the SNP rs4728142 was most significantly associated in both African Americans and Asians, but not in Europeans.

For all three phenotypes, genetic variant association was no longer significant when the phenotypic association resulting from the two most associated SNPs was removed. With age of onset, the group of SNPs spanning *IRF5/TNPO3* (closely linked to the SNP rs12534421) and the group of genetic variants in the promoter region of *IRF5* (closely linked with the SNP rs4728142) have significant association in Europeans, with

variants linked to rs12534421 having the lowest p values (Figure 4). This is shown by the sharp decline in statistical significance (-log p value) when each SNP is conditioned separately. For the presence of anti-Ro and anti-dsDNA antibodies, the promoter group (rs4728142 and associated SNPs) had the greatest effect (Figures 5-6).

Although rs4728142 and rs12534421 are most closely associated with the development of lupus and the two genetic effects in the *IRF5/TNPO3* region, rs3778752 was most closely linked to the presence of anti-Ro and anti-dsDNA antibodies. The rs3778752 risk allele is a part of the SNP group in the promoter region of IRF5. In association studies, rs3778752 had lower p-values than rs4728142 in both anti-dsDNA ($p_{rs3778752} = 4.154 \times 10^{-7}$, $p_{rs4728142} = 5.85 \times 10^{-6}$) and anti-Ro ($p_{rs3778752} = 2.09 \times 10^{-7}$, $p_{rs4728142} = 9.37 \times 10^{-7}$). We identified chromosomes in European Americans in which the two SNPs segregate (n = 1992) and the subjects have either anti-dsDNA or anti-Ro antibodies (n = 624). Our results show that subjects with anti-Ro or anti-dsDNA antibodies are 133% more likely to have the rs3778752 risk allele than the rs4728142 risk allele.

Age of Onset

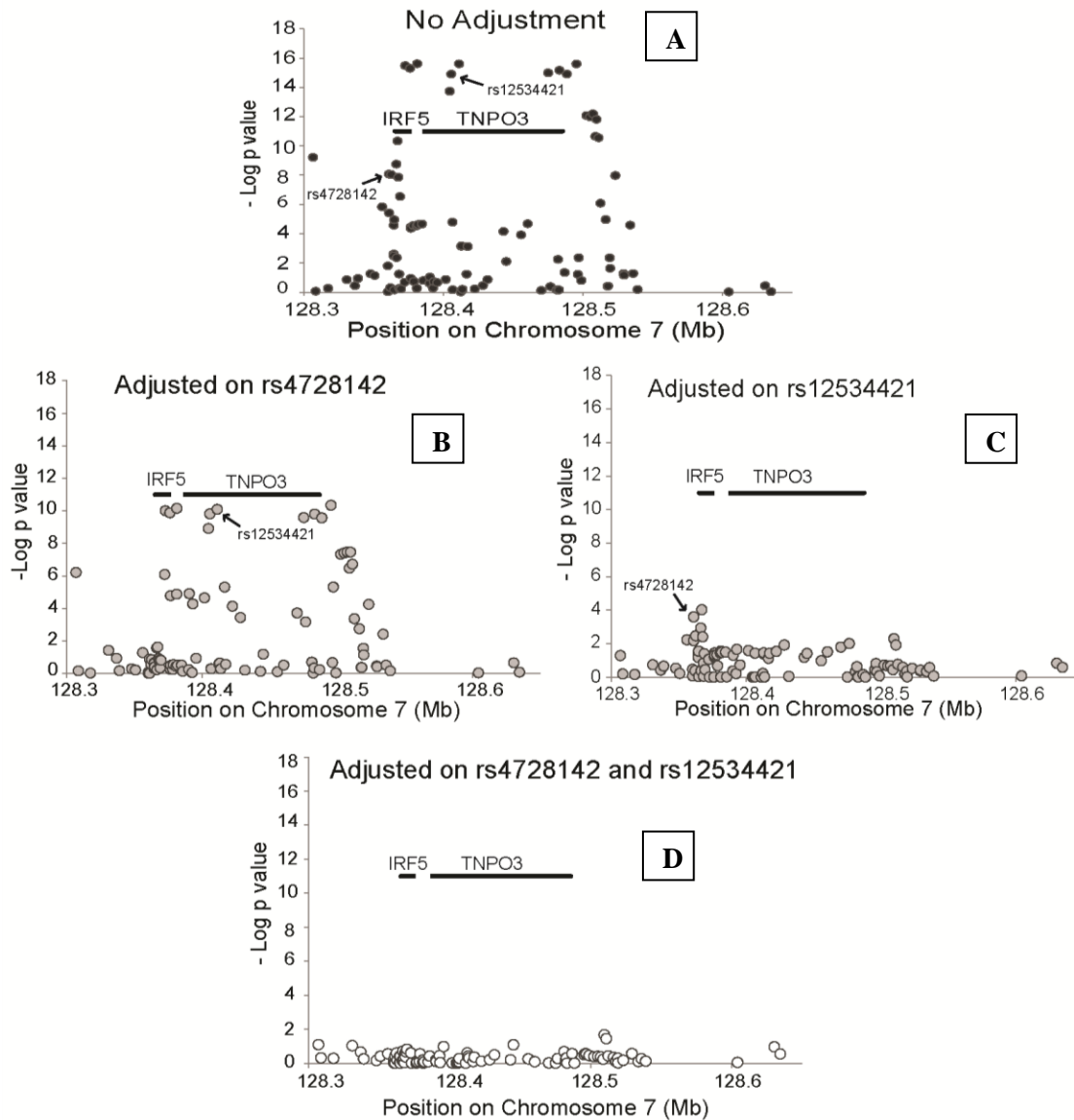


Figure 4. Conditional analysis of age-of-onset phenotype. The horizontal bars show the position on chromosome 7 of *IRF5* and *TNPO3*. The most associated SNPs are identified in A, B, and C. Graph A shows the statistical significance ($-\log_{10}[p\text{-value}]$) of each SNP in the *IRF5/TNPO3* region. For graphs B and C, we removed each of the top two most associated SNPs. Graph D removes the association of both SNPs. From this, we conclude that rs12534421 (the second group of linked SNPs) is responsible for most of the statistical association between genetic variants and the age of onset of lupus. When the contribution of both SNPs is removed, no association remains $p < 0.01$ (D).

Anti-Ro

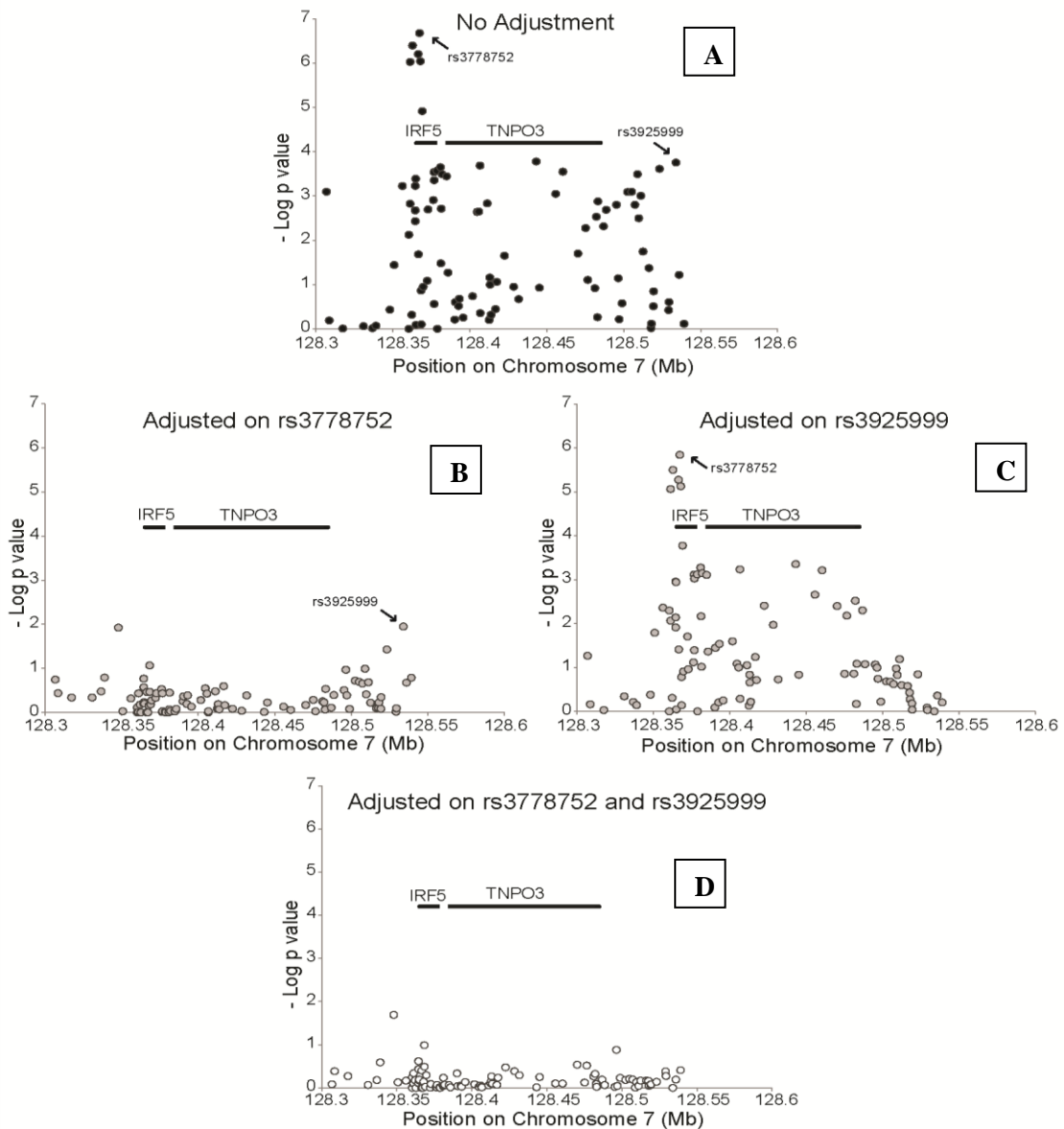


Figure 5. Conditional analysis of the presence of anti-Ro antibodies. Graph A shows the statistical significance ($-\log_{10}[\text{p-value}]$) of the association of each SNP in the *IRF5/TNPO3* region with the presence of anti-Ro antibodies. For graphs B and C, we removed the association of the two most significant SNPs (which did not include rs4728142 or rs12534421) separately. The bottom graph removes the association of both SNPs. From this, we can conclude that rs3778752 (the first SNP group) is responsible for most of the statistical association of genetic variants with the presence of anti-Ro antibodies.

Anti-dsDNA

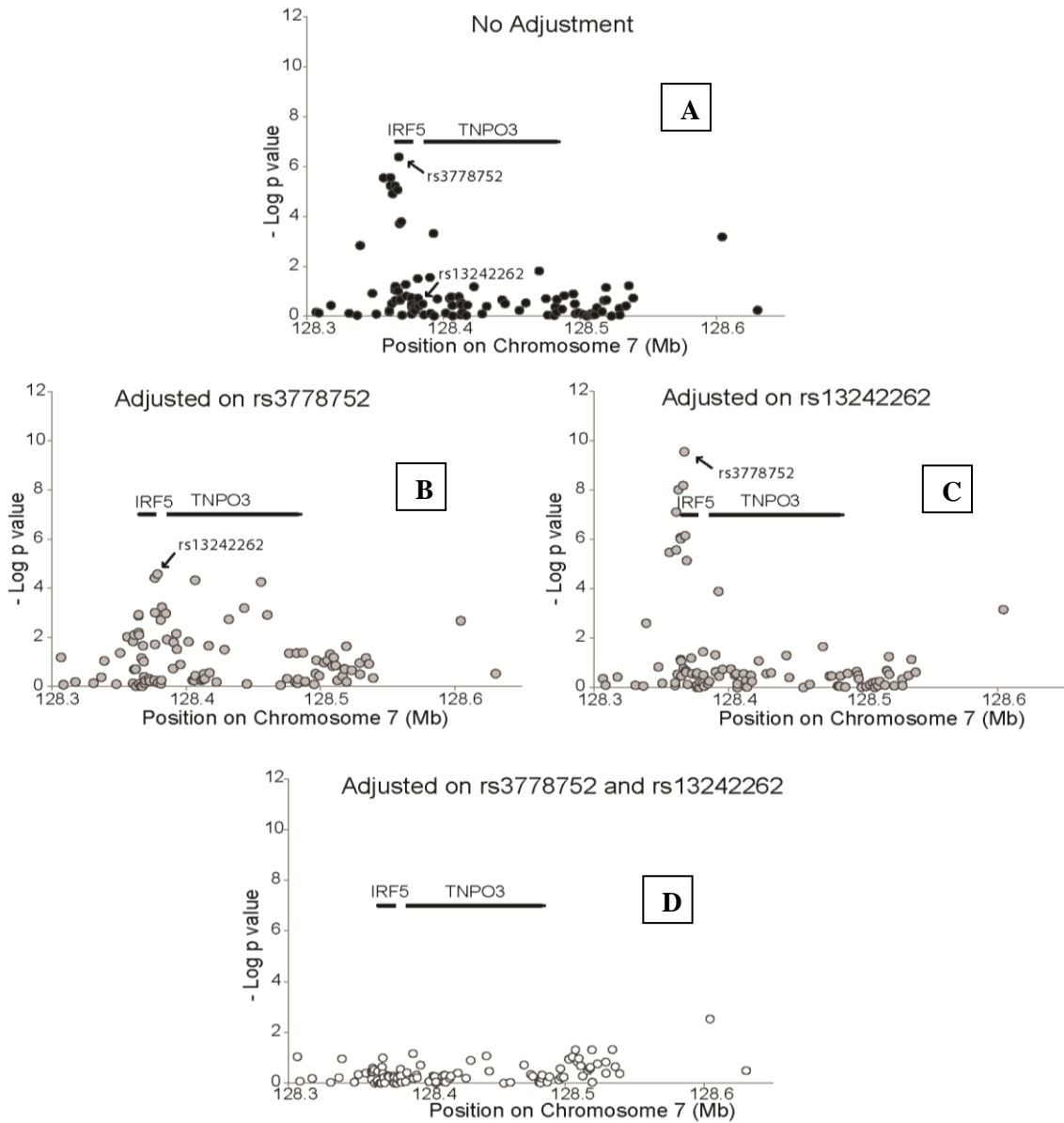


Figure 6. Conditional analysis of the presence of anti-dsDNA antibodies. Graph A shows the significance ($-\log_{10}[p\text{-value}]$) of the association of each SNP in the *IRF5/TNPO3* region with the presence of anti-dsDNA antibodies. For graphs B and C, we conditioned separately on the two most closely associated SNPs. Graph D adjusts for the association of both SNPs. From this, we can conclude that rs3778752 (the first SNP group) is responsible for most of the statistical association of genetic variants with the presence of anti-dsDNA antibodies.

CHAPTER 5

DISCUSSION

Genetic Variant Association

Defined connections between genetic variants and lupus sub-phenotypes may enable increased understanding of the etiology of lupus. These connections may also allow clinical practitioners to predict a patient's risk of developing specific symptoms. In this study, variants in *IRF5* and *TNPO3* that are already known to be statistically associated with the development of lupus were shown to be connected with specific sub phenotypes including age of onset of lupus and the production of anti-Ro or anti-dsDNA antibodies. Association was found largely in European American subjects but was also present in African Americans and Asians to a lesser degree.

Two Independent Genetic Effects

Although the two groups of linked SNPs in *IRF5* and *TNPO3* are highly associated with the development of lupus, genetic variants in the promoter of *IRF5* (linked to the SNP rs4728142) and genetic variants spanning *IRF5* and *TNPO3* (linked to the SNP rs12534421) are independently associated with specific symptoms of lupus. This supports the hypothesis that these are indeed two separate SNP groups which play different roles in SLE. The group spanning *IRF5/TNPO3* only displays significance in

European Americans, while the promoter group plays a role in all ethnicities. This could signify that SNPs in the promoter of *IRF5* are part of an ancient set of variants, while the group of associated SNPs spanning *IRF5* and *TNPO3* are the result of more recent variation.

SNP Association Differences

The rs3778752 risk allele shows 133% more association with the presence of anti-Ro and anti-dsDNA antibodies than the rs4728142 risk allele. Our results suggest that although rs4728142 is one of the most significant indicators of lupus risk, rs3778752 seems to be far more effective in predicting the production of anti-Ro and anti-dsDNA in the body.

Conclusion

In conclusion, our study shows that genetic variants in the *IRF5/TNPO3* genes are associated with specific clinical phenotypes of lupus. The two independent SNP groups in this region function separately in the different symptoms as well as in the different ethnicities. Our data support prior findings that the two etiological effects are associated with different ancestries, and indicate that genetic variant groups in *IRF5/TNPO3* function differently in the development of certain lupus phenotypes.

REFERENCES

- 1 Tsokos GC: Systemic lupus erythematosus. *The New England journal of medicine* 2011;22:2110-2121.
- 2 Lau CS, Yin G, Mok MY: Ethnic and geographical differences in systemic lupus erythematosus: an overview. *Lupus* 2006;11:715-719.
- 3 McCarty DJ, Manzi S, Medsger TA, Jr., Ramsey-Goldman R, LaPorte RE, Kwok CK: Incidence of systemic lupus erythematosus. Race and gender differences. *Arthritis Rheum* 1995;9:1260-1270.
- 4 Seligman VA, Lum RF, Olson JL, Li H, Criswell LA: Demographic differences in the development of lupus nephritis: a retrospective analysis. *The American journal of medicine* 2002;9:726-729.
- 5 Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ: The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982;11:1271-1277.
- 6 Guidelines for referral and management of systemic lupus erythematosus in adults. American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Guidelines. *Arthritis Rheum* 1999;9:1785-1796.
- 7 Merrell M, Shulman LE: Determination of prognosis in chronic disease, illustrated by systemic lupus erythematosus. *Journal of chronic diseases* 1955;1:12-32.
- 8 Vaughn SE, Kottyan LC, Munroe ME, Harley JB: Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. *Journal of leukocyte biology* 2012;3:577-91.
- 9 Sanchez E, Nadig A, Richardson BC, Freedman BI, Kaufman KM, Kelly JA, Niewold TB, Kamen DL, Gilkeson GS, Ziegler JT *et al*: Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus. *Annals of the rheumatic diseases* 2011;10:1752-1757.
- 10 Morris DL, Vyse TJ: Analysis of systemic lupus erythematosus sub-phenotype data for genetic association. *Current opinion in rheumatology* 2012;5:482-488.
- 11 Salloum R, Franek BS, Kariuki SN, Rhee L, Mikolaitis RA, Jolly M, Utset TO, Niewold TB: Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients. *Arthritis Rheum* 2010;2:553-561.
- 12 Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, Jacob CO, Kamboh MI, Alarcon-Riquelme ME, Tsao BP *et al*: Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. *PLoS genetics* 2011;2:e1001311.
- 13 Webb R, Kelly JA, Somers EC, Hughes T, Kaufman KM, Sanchez E, Nath SK, Bruner G, Alarcon-Riquelme ME, Gilkeson GS *et al*: Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. *Annals of the rheumatic diseases* 2011;1:151-156.
- 14 Niewold TB, Kelly JA, Kariuki SN, Franek BS, Kumar AA, Kaufman KM, Thomas K, Walker D, Kamp S, Frost JM *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. *Annals of the rheumatic diseases* 2012;3:463-468.
- 15 Cham CM, Ko K, Niewold TB: Interferon regulatory factor 5 in the pathogenesis of systemic lupus erythematosus. *Clinical and Developmental Immunology* Published Online First: 12 September 2012. doi:10.1155/2012/780436

- 16 Guthridge JM, Clark DN, Templeton A, Dominguez N, Lu R, Vidal GS, Kelly JA, Kauffman KM, Harley JB, Gaffney PM *et al*: Effects of *IRF5* lupus risk haplotype on pathways predicted to influence B cell functions. *Journal of biomedicine & biotechnology* 2012, 2012:594056.
- 17 Beal E: *IRF5* and B cells in LUPUS development. *Nature reviews Rheumatology* 2010;**8**:437.
- 18 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;**8**:2481-2487.
- 19 Kozyrev SV, Alarcon-Riquelme ME: The genetics and biology of *IRF5*-mediated signaling in lupus. *Autoimmunity* 2007, 40(8):591-601.
- 20 Cunninghame Graham DS, Manku H, Wagner S, Reid J, Timms K, Gutin A, Lanchbury JS, Vyse TJ: Association of *IRF5* in UK SLE families identifies a variant involved in polyadenylation. *Hum Mol Genet* 2007;**6**:579-591.
- 21 Demirci FY, Manzi S, Ramsey-Goldman R, Minster RL, Kenney M, Shaw PS, Dunlop-Thomas CM, Kao AH, Rhew E, Bontempo F *et al*: Association of a common interferon regulatory factor 5 (*IRF5*) variant with increased risk of systemic lupus erythematosus (SLE). *Annals of human genetics* 2007;**3**:308-311.
- 22 Ferreira-Neira I, Calaza M, Alonso-Perez E, Marchini M, Scorza R, Sebastiani GD, Blanco FJ, Rego I, Pullmann R, Jr., Pullmann R *et al*: Opposed independent effects and epistasis in the complex association of *IRF5* to SLE. *Genes Immun* 2007;**5**:429-438.
- 23 Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, Plenge RM, Koeuth T, Ortmann WA, Hom G *et al*: Three functional variants of IFN regulatory factor 5 (*IRF5*) define risk and protective haplotypes for human lupus. *Proc Natl Acad Sci U S A* 2007;**16**:6758-6763.
- 24 Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, Tsao BP, Vyse TJ, Langefeld CD: Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nature Genetics* 2008;**2**:204-210.
- 25 Jarvinen TM, Hellquist A, Zucchelli M, Koskenmies S, Panelius J, Hasan T, Julkunen H, D'Amato M, Kere J: Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of *IRF5* in disease susceptibility in a Northern European population. *Rheumatology (Oxford)* 2012;**1**:87-92.
- 26 Kawasaki A, Kyogoku C, Ohashi J, Miyashita R, Hikami K, Kusaoi M, Tokunaga K, Takasaki Y, Hashimoto H, Behrens TW *et al*: Association of *IRF5* polymorphisms with systemic lupus erythematosus in a Japanese population: support for a crucial role of intron 1 polymorphisms. *Arthritis Rheum* 2008;**3**:826-834.
- 27 Kelly JA, Kelley JM, Kaufman KM, Kilpatrick J, Bruner GR, Merrill JT, James JA, Frank SG, Reams E, Brown EE *et al*: Interferon regulatory factor-5 is genetically associated with systemic lupus erythematosus in African Americans. *Genes Immun* 2008;**3**:187-194.
- 28 Nordang GB, Viken MK, Amundsen SS, Sanchez ES, Flato B, Forre OT, Martin J, Kvien TK, Lie BA: Interferon regulatory factor 5 gene polymorphism confers risk to several rheumatic diseases and correlates with expression of alternative thymic transcripts. *Rheumatology (Oxford)* 2012;**4**:619-626.
- 29 Qin L, Lv J, Zhou X, Hou P, Yang H, Zhang H: Association of *IRF5* gene polymorphisms and lupus nephritis in a Chinese population. *Nephrology (Carlton)* 2010;**7**:710-713.
- 30 Reddy MV, Velazquez-Cruz R, Baca V, Lima G, Granados J, Orozco L, Alarcon-Riquelme ME: Genetic association of *IRF5* with SLE in Mexicans: higher frequency of the risk haplotype and its homozygosity than Europeans. *Human genetics* 2007;**6**:721-727.

- 31 Shin HD, Sung YK, Choi CB, Lee SO, Lee HW, Bae SC: Replication of the genetic effects of IFN regulatory factor 5 (*IRF5*) on systemic lupus erythematosus in a Korean population. *Arthritis Res Ther* 2007;**2**:R32.
- 32 Sigurdsson S, Goring HH, Kristjansdottir G, Milani L, Nordmark G, Sandling JK, Eloranta ML, Feng D, Sangster-Guity N, Gunnarsson I *et al*: Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (*IRF5*) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus. *Hum Mol Genet* 2008;**6**:872-881.
- 33 Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G, Jonsen A, Rantapaa-Dahlqvist S, Moller B, Kere J *et al*: Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 2005;**3**:528-537.
- 34 Siu HO, Yang W, Lau CS, Chan TM, Wong RW, Wong WH, Lau YL, Alarcon-Riquelme ME: Association of a haplotype of *IRF5* gene with systemic lupus erythematosus in Chinese. *The Journal of rheumatology* 2008;**2**:360-362.
- 35 Tsuchiya N, Ito I, Kawasaki A: Association of *IRF5*, *STAT4* and *BLK* with systemic lupus erythematosus and other rheumatic diseases. *Nihon Rinsho Men'eki Gakkai kaishi = Japanese journal of clinical immunology* 2010;**2**:57-65.
- 36 Hirschfield GM, Liu X, Han Y, Gorlov IP, Lu Y, Xu C, Chen W, Juran BD, Coltescu C, Mason AL *et al*: Variants at *IRF5-TNPO3*, 17q12-21 and *MMEL1* are associated with primary biliary cirrhosis. *Nat Genet* 2010;**8**:655-657.
- 37 Kottyan LC, Zoller E, Kaufman K, Harley J: The *IRF5-TNPO3* association has two components in systemic lupus erythematosus, which are shared with other autoimmune disorders. *The Journal of Immunology* 2013;**190**:175.5.
- 38 McKeigue PM, Carpenter JR, Parra EJ, Shriver MD: Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Annals of human genetics* 2000;**2**:171-186.
- 39 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006;**8**:904-909.
- 40 Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM: Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003;**6**:1492-1504.
- 41 Hoggart CJ, Shriver MD, Kittles RA, Clayton DG, McKeigue PM: Design and analysis of admixture mapping studies. *Am J Hum Genet* 2004;**5**:965-978.
- 42 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**3**:559-575.