Published in final edited form as:

J Autoimmun. 2012 August ; 39(1-2): 57–63. doi:10.1016/j.jaut.2012.01.008.

Genetics of Sjögren's syndrome in the genome-wide association era

John A. Ice, MD¹, He Li, MS^{1,2}, Indra Adrianto, PhD, MS, Paul Chee Lin, MS, Jennifer A. Kelly, MS¹, Courtney G. Montgomery, PhD, Christopher J. Lessard, PhD^{1,2}, and Kathy L. Moser, PhD^{1,2}

¹Arthritis and Clinical Immunology, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, Oklahoma 73104, USA

²Department of Pathology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd, BMSB 451, Oklahoma City, Oklahoma, 73104, USA

Abstract

While Sjögren's syndrome (SS) is more common than related autoimmune disorders, such as systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), scientific and medical research in SS has lagged behind significantly. This is especially true in the field of SS genetics, where efforts to date have relied heavily on candidate gene approaches. Within the last decade, the advent of the genome-wide association (GWA) scan has altered our understanding of disease pathogenesis in hundreds of disorders through the successful identification of novel risk loci. With strong evidence for a genetic component in SS as evidenced by familial aggregation of SS as well as similarities between SS and SLE and RA, the application of GWA approaches would likely yield numerous novel risk loci in SS. Here we review the fundamental scientific principles employed in GWA scans as well as the limitations of this tool, and we discuss the application of GWA scans in determining genetic variants at play in complex disease. We also examine the successful application of GWA scans in SLE, which now has more than 40 confirmed risk loci, and consider the possibility for a similar trajectory of SS genetic discovery in the era of GWA scans. Ultimately, the GWA studies that will be performed in SS have the potential to identify a myriad of novel genetic loci that will allow scientists to begin filling in the gaps in our understanding of the SS pathogenesis.

Keywords

genetics; Sjögren's syndrome; genome-wide association

1.1 Introduction

Sjögren's syndrome (SS) is a chronic, progressive exocrinopathy characterized by lymphocytic infiltration and destruction of glandular tissue, namely the salivary and lacrimal glands, with or without the production of autoantibodies. Primary symptoms include severe

© 2012 Elsevier Ltd. All rights reserved.

Corresponding Author: Kathy L. Moser, PhD, Address: 825 NE 13th Street, Oklahoma City, Oklahoma, 73104, USA, Telephone: +1 (405) 271-2534, Fax: +1 (405) 271-2536, moserk@omrf.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

dry eyes and dry mouth, although clinical signs and symptoms are extremely heterogeneous and can involve virtually any organ system. Even though SS has a relatively high prevalence (0.5–1.0%) compared to similar autoimmune diseases like systemic lupus erythematosus (0.1% for European-Americans and 0.4% for African-Americans), it remains vastly understudied [1, 2]. This disparity in the research of autoimmune diseases becomes clear if one considers the number of genetics publications in the field of autoimmunity across the last four decades (Figure 1.1). Note that there are far few SS genetics publications each year than in comparable autoimmune disorders. The reasons for this lag in SS research are complex and multifaceted, but stem in large part from challenges of assembling large cohorts of patients with clearly defined SS. Nonetheless, with the formation of global partnerships and collaborations, we are beginning to build the framework for advanced genetic research to elucidate SS pathogenesis.

While many diseases have clear-cut diagnostic criteria, the same cannot be said for SS. Since this syndrome was first described in 1933, more than eight sets of classification criteria for SS have been developed and published by different groups from across the world. These variations in criteria have led to extreme heterogeneity and the potential for inconsistencies in the classification process, making the assembly of large patient cohorts with clearly defined SS very difficult. Currently, the most widely accepted classification criteria were published in 2002 by the American-European Consensus Group, which had sought to develop a set of internationally recognized classification criteria, but instead represents a revision to the previous European criteria [3]. These criteria require a battery of multi-disciplinary tests, several of which (including ocular staining and minor salivary gland biopsy) require the skill of an experienced clinician in order to be reliable.

While SS genetic research has continued in spite of these challenges, the field has fallen short of other autoimmune diseases. The successful application of modern genetic tools in autoimmune diseases such as systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) has resulted in discovery of numerous variants associated with disease. To date, genetic studies in SS have relied primarily on candidate gene approaches, focusing on those genes with biological plausibility for a role in SS etiology or evidence of association in other autoimmune diseases (Table 1). Frequently, genetic studies in SS have been confounded by poorly reproducible results due to small cohort sizes, differential distribution of variation between cohorts being studied, and variations in allele frequencies due to ancestral population differences from study to study.

Increasingly, the paucity of reliable genetic studies in SS makes it more challenging for scientists to unravel the complex mechanisms that underlie the immune dysregulation of this syndrome. Recently, however, the application of cutting-edge computational technology and high-throughput genotyping platforms has transformed the way in which we study genetic disease. One tool in particular, known as the genome-wide association (GWA) scan, has proven remarkably adept in identifying the association of common genetic variants and disease.

In this review, we outline recent historical events that have paved the way for use of GWA scans as significant tools for the study of complex disease, discuss the fundamental concepts required to interpret GWA scans, illustrate the use of GWA scans in SLE as an example and the significant findings that have emerged, and discuss near-future directions for the study of SS genetics.

1.2 Epidemiology of SS

The case for the influence of heritable factors in the etiology of SS is promising, despite limited availability of robust genetic studies or more traditional measures of genetic effects,

such as twin concordance estimates or familial aggregation studies. Given the overlap of certain clinical and serologic features of SS and other autoimmune diseases, such as SLE and RA, we can reasonably assume that the heritable contribution to SS is due to multiple genes influencing various immunologic-related pathways. As with most diseases, however, risk for developing SS is likely influenced by several variables, including environmental triggers, epigenetic effects, gene-gene interactions, etc., superimposed on an underlying genetic profile.

Traditionally, with a lack of resounding evidence for a genetic component to disease, geneticists have turned to concordance rates of disease in monozygotic twins and evidence of familial aggregation of disease. Again, in SS we find far fewer of these studies than in other autoimmune diseases. Even though several case reports have been published describing twins with SS, no reliable concordance rates have yet been determined [4–7]. Estimates of twin concordance rates in SS are likely to be similar to those of RA (15%) and SLE (25%). [8, 9] Additionally, several multiplex families with SS have been reported, and interestingly, there exists an increased prevalence for autoimmune diseases amongst family members of patients with SS, including SS (12%), RA (14%), autoimmune thyroid disease (AITD, 14%), and SLE (5–10%) [10]. Again considering SLE and RA for comparison, female sibling or dizygotic twins rates for SS could be estimated at 2–4%, while female sibling concordance (λ_s) could be between 8 and 30 [8, 11].

1.3 Historical events in genetics leading to GWA scans

The emergence of the GWA scan as a major gene discovery tool has been built on progressively advancing technology in human genetics, beginning first and foremost with the sequencing of the human genome. Started in 1990 and sponsored by the Department of Energy and National Institutes of Health, the Human Genome Project (HGP) was designed to sequence the first entire human genome in under 15 years. The HGP, a publicly funded endeavor that made its results freely available, would soon face competition when, in 2000, it was announced that a private endeavor to sequence the human genome, led by Celera Genomics, was also underway. With rapid advances in technology, however, scientists in both the public and private sectors were able to complete the first working drafts of the human genome much earlier than expected, and both versions were published in February 2001 [12, 13]. Immediately, the long and arduous task of scouring the human genome for genes and genomic features that could help elucidate the nature of human disease began.

In 2002, shortly after the drafts of the human genome were published, an international endeavor to study and catalog the variation of the human genome was formed. Known as the International HapMap Project, the goal of this consortium was to map the patterns of DNA sequence variation of individuals with ancestry from parts of Africa, Asia, and Europe, and to make this information available to the public [14]. To do this, the project would genotype at least 1 million DNA sequence variants, determine the frequencies of these variants in four ancestral populations (Northern and Western European, Yoruban, Japanese, and Han Chinese), and determine the degree of association between variants. Some shortcomings of this project have become apparent, however, as it captures only the most common variants present in the human genome and is restricted to only four ancestral populations.

In 2008, the 1000 Genomes Project was initiated, with similar goals of characterizing the extent of human genomic variation. This project, however, had the goal of providing reference sequences across a larger spectrum of the world's ancestral populations. The 1000 Genomes Project began with three pilot projects: a trio project to perform whole-genome sequencing in two families, one of Yoruban origin and another of European ancestry, each consisting of two parents and one daughter, across 3 different platforms; a low-coverage

project in 179 unrelated individuals from 4 ancestral backgrounds (European, Han Chinese, Yoruban, and Japanese); and a high-coverage ($> 50\times$) project restricted to 8,140 exons from 906 selected genes across 7 populations [15]. Interestingly, in the first published account of the findings from the 1000 Genomes Project, this consortium described the allele frequency and haplotype structure for more than 15 million SNPs, 1 million short insertion/deletion events, and 20,000 structural variants [15].

1.4 Genetics in the Era of Genome-wide association studies

The first GWA scan published in 2005 studied the genetic influences of macular degeneration and led to a dramatic shift in study of human genetics [16]. Since this time, GWA scans have revolutionized our understanding of complex disease genetics. Previously, linkage scans would identify large genomic segments often containing hundreds of genes, many of which had obvious biological relevance to disease risk. In addition to linkage scans, candidate gene studies were also performed that were focused only on those regions of biological relevance to a particular disease or those that were identified through animal models. Unlike these methods, GWA scans survey the genome in an unbiased manner while covering an exceedingly large amount of the variation in a given population's genome. These studies have been very successful with more than 2700 reported associations exceeding the accepted genome-wide significance threshold of $P < 5 \times 10^{-8}$ across 358 phenotypes [17].

To test for single SNP-disease association, the allele frequencies are calculated for each group (e.g. cases vs. controls), and these allele frequencies can be compared using a variety of statistical models to determine if the differences in allele frequency between the groups are significant. Depending on the data and the scientific questions being posed, a simple 2×2 comparison, such as Chi-square testing, or more complex models, such as logistic regression using additive, recessive, or dominant genetic models that adjust for potential confounders (i.e. population substructure or gender), can be used.

Central to the success of GWA scans is the extensive genomic variation now known to exist in human genomes. Genomic variation occurs in many forms, including insertion/deletion, duplication, and translocation events. The simplest and most prevalent genetic variant is the single-nucleotide polymorphism (SNP). A SNP is defined as a single base-pair change that is present in a population at a measurable frequency. Early studies comparing two human DNA sequences estimated that a SNP occurs approximately every 1,000–2000 nucleotides of DNA [17]. More recent studies, however, estimate that SNPs occur once every 100–300 nucleotides out of a total of more than 3.2 billion base pairs [15]. This translates to anywhere from 10 to 30 million SNPs in the human genome for variants with relatively common allele frequencies. Of course, rare variants (commonly defined as those having an allele frequency less than 1%) continue to be discussed through in-depth sequencing studies and increase the total number of known variants in the human genome to more than 15 million [15].

GWA scans are advantageous in that they exploit several interesting features of genetic variation within the human genome. Every genetic variant has at least two or more alleles or polymorphisms occurring in various frequencies according to an individual's ancestral origins. A typical GWA scan, compares genotypes at hundreds of thousands of SNPs from samples in two populations: those with disease (cases) and those without disease (healthy controls). Another valuable feature of the human genome is the phenomenon known as linkage disequilibrium (LD) (Figure 1.2). When 2 SNPs are in LD with each other, historical recombination events are less likely to have occurred in the interval between the two of them than would be expected due to random chance alone. As a result, genetic variants may be

inherited together in predictable ways through generations. This leads to the creation of so-called “haplotype blocks,” in which specific alleles are strongly correlated with one another and are transmitted in blocks from one generation to the next.

The pairwise LD between variants within the human genome and the haplotype blocks they form have been characterized principally through the International HapMap Project and the 1000 Genomes Project, but they have also been measured using GWA datasets. It is important to note that the structure of haplotypes often differ between populations. For example, subjects of European ancestry typically have larger haplotype blocks across the genome than do those of African ancestry. This is due to the fact that the African population is much older than the European population and thus has undergone more recombination events. Over time, populations can undergo *de novo* mutations, selection for beneficial alleles, bottlenecks of population growth, and other events that can have significant influences on overall allele frequencies for specific variants. This can lead to vast differences in the frequency of alleles and haplotypes between populations. Yet with knowledge of the underlying haplotype structure within a particular region, geneticists can use genotyped variants to predict *in silico* the genotype of variants that were not directly measured. This procedure is known as imputation, and is extremely beneficial when attempting to localize a potential causal variant for disease.

Imputation allows geneticists to use a reference panel consisting of a very large number of variants that are subsequently phased into haplotypes to predict genotypes *in silico* and to merge these predicted genotypes into existing GWA datasets. The reference panel haplotypes are compared to the haplotypes present in the GWA dataset to predict the most likely missing genotypes. Each imputed variant is given a probability score to indicate likelihood of being correct, given the data available from the GWA scan. Thus, with a limited number of variants, one can use a reference panel to impute and gain many thousands of additional genotypes *in silico*.

1.5 Advantages and Disadvantages to GWA scans

As has been outlined above, GWA studies have several advantages over linkage or candidate gene studies. They also, however, have certain limitations that must be given appropriate consideration before any data is interpreted.

One caveat that must be considered when determining the significance of a *P* value obtained through GWA scan is the effect of multiple testing on significance thresholds. As the number of tests in an experiment increases, the likelihood of type 1 error or false-positive associations increases. Therefore, *P* values must be corrected to account for multiple tests in order to decrease the likelihood of false-positive associations. One type of adjustment for multiple testing commonly used in GWA studies is known as the Bonferroni correction, and is calculated by dividing a significance of 0.05 by the number of independent tests within a given experiment. For a study using GWA data, the most widely accepted *P* value threshold required to achieve genome-wide significance is 5×10^{-8} (for 1 million tests), although the topic is a frequent source of discussion despite having been studied extensively [18, 19]. As a consequence, relatively large sample sizes (often including thousands of subjects) may be required. Furthermore, current standards for GWA studies also include replication of a putative association in an independent cohort before a genetic effect can be considered established.

Another caveat is that while LD can decrease the number of genotyped variants needed to evaluate the majority of variation in a given population's genome, this very advantage can turn into a disadvantage when attempting to pinpoint a causal variant. When a causal variant is in strong LD with a large, highly correlated haplotype block, it becomes impossible to

determine which variant in that haplotype block is in fact the causal variant using statistical tools. To determine the causal variant, scientists must turn to alternative methods, such as functional studies using molecular techniques to localize any possible biological effect to a smaller region of DNA. In addition, intensive resequencing efforts looking for distinct or interesting genomic features that could explain possible biological effects are needed to insure that all variation, including those with allele frequencies of <1%, is captured.

Despite stringent corrections for multiple testing and challenges in determining an exact causal variant, GWA studies have proven extremely valuable to the field of human genetics. From 2005 to December 2010, more than 1200 GWA scans have been performed across 210 distinct phenotypes [17]. Whereas linkage analysis works best to identify rare variants in Mendelian diseases, GWA scans work best to identify common variants (that is, those with allele frequencies > 10%) associated with a phenotype [20].

1.6 SLE as a model for SS

SLE provides an excellent model to hypothesize the possible trajectory of genetic discovery for SS in the era of GWAS. Researchers studying SLE have been able to amass large cohorts of patients for participation in genetic studies and over 40 significant and reproducible genetic associations have been established in SLE. Prior to the application of GWA scans in SLE, several pertinent genetic discoveries using linkage studies and candidate gene approaches were made. As with SS, the first genetic associations were found within the HLA in the 1970s [21, 22]. Following soon thereafter were associations to the complement genes *C2*, *C4*, and *C1q*, which, although rare, are highly penetrant [23–25]. In the 1990s and early 2000s, a handful of genetic associations were discovered, including *FcγR2A*, *FcγR3A*, *PDCD1*, *PTPN22*, *IRF5*, *STAT4*, and *TREX1* [26–36].

The year 2008 marked a dramatic shift in the genetics of SLE. With the publication of the first 4 GWA scans in SLE, a total of 9 novel risk loci were discovered [37–40]. In the years since, number of confirmed risk loci in SLE has jumped to more than 40, with most of them discovered through GWA scans. Interestingly, each GWA study was able to confirm some of the pre-GWA era findings, but more importantly, each scan revealed novel associations. This occurred primarily for two reasons. First, different cohorts of cases and controls were evaluated in study. SLE is a very heterogeneous disease (similar to SS), which can lead to enrichment in different cohorts for subsets of patients and ultimately differential effects between studies. Second, different genotyping arrays were used that varied in SNP content and variably included between 100,000 and 500,000 variants. This limited the coverage of variation captured in the genome in any given array. Thus, each of the four studies revealed genetic effects other studies were not able to identify.

For example, three studies were published simultaneously in early 2008. Kozyrev et al. used the Affymetrix 100K array to identify the association of *BANK1* with SLE [37]. Hom et al. used the Illumina HumanHap550 Genotyping BeadChip and reported the association of *ITGAM-ITGAX* and *C8orf13-BLK* [38]. Harley et al. also found association with *ITGAM-ITGAX* and *C8orf13-BLK* using the Illumina HumanHap 330 BeadChip, but in addition found *PXK*, *KIAA1542/IRF7*, and others in gene deserts [39]. Later in 2008, Graham et al. reported their results from the Affymetrix Human 550 array, which identified *TNFAIP3* [40]. It is important to point out that these studies all were done on subjects of European ancestry. The first non-European GWA scan in SLE was reported in 2009 using the Illumina HumanHap 610 BeadChip [41]. This study identified two novel associations, *EST1* and *SLC15A* [41].

Most of these GWA studies showed suggestive evidence for association for many of the regions that were unique to each study, and follow-up studies continue to confirm and

identify additional novel associations. One such example is a recent SLE study by Lessard et al. examined a putative association at 11p13 near *CD44/PDHX*, which had not achieved genome-wide significance in a prior GWA scan ($p=0.004$), but showed robust association when replicated in an independent cohort ($p=9.03\times10^{-8}$) to yield in meta-analysis a p-value that far exceeded genome-wide significance ($p=2.36\times10^{-13}$) [42]. This and examples in other autoimmune disorders illustrate the importance of carrying out further replication studies and large-scale meta-analyses in the gene-discovery process to realize the full potential of GWA scan results.

Considering the plethora of recent gene discoveries made using large, genome-wide case/control association studies coupled with the genetic foundations built by linkage analysis and candidate-gene approaches, a greater picture of the pathophysiological mechanisms underlying SLE is beginning to take shape. Many diverse immunological pathways are beginning to reshape and refine our understanding of these mechanisms. For example, *IRF5* and *STAT4* have been shown to play important roles in toll-like receptor and interferon signaling, while *TNFAIP3* and *TNIP1* play key roles in tumor-necrosis factor and nuclear factor kappa beta signaling. In addition, gene discoveries continue to elucidate the common mechanisms shared across multiple autoimmune diseases, potentially leading to discoveries that could more generally redefine our understanding of autoimmune dysfunction.

1.7 Future directions for SS genetics

The future of genetic studies in SS is extremely promising. Rapid advances in technology and a better understanding of human genomic variation have made high-throughput genotyping a powerful tool to survey genetic variants for association with disease. While a number of candidate-gene approaches have been carried out in SS, they have lacked sufficiently large sample sizes, have not been replicated in independent cohorts, and have not taken into account ancestral differences inherent to modern populations.

The next logical step, then, is to perform a GWA scan in SS that addresses the limitations faced in prior genetic studies. Currently, research groups from around the globe are forming consortiums to amass the large cohorts of SS patients that will be required to perform GWA scans in SS. These GWA scans must have adequate power and sample diversity to capture the broad spectrum of common variants that are likely involved in SS pathogenesis. Based on the results of other GWA scans published in other complex phenotypes, we can anticipate that a similar study in SS has the potential to identify many variants that convey risk.

In SLE, additional gene discoveries have been made due to diversity in the ancestral composition of cohorts, variation that is inherent in cohorts of different sizes and origins, and variation between the genotyping platforms used to perform the GWA scans. SS will similarly have need for multiple GWA scans performed on different genotyping platforms, with larger numbers of samples, and in a variety of ancestral populations. This is of particular importance since in addition to sample size, effect size, typically expressed as odds ratios (OR), and minor allele frequencies also influence power to detect genetic associations. As sample sizes increase the ability to detect association with smaller effect sizes and lower minor allele frequencies increases (Figure 3). Future studies that take these variables into account will better pinpoint key risk variants that are involved in the immunological dysregulation or other pathways leading to the development of SS.

One essential aspect that must be addressed in future genetic studies is the contribution of rare variants (those with allele frequencies $< 1\%$) to disease risk. Within the human genome, there are far fewer common variants than rare variants, suggesting a large role for rare variants in disease pathogenesis [43]. Rare variants can be identified through intensive next-generation sequencing endeavors that will focus on performing high-coverage sequencing of

the whole genome of SS patients. Additionally, there will be projects focusing on the epigenetic influences on SS risk and efforts to perform whole-transcriptome sequencing to determine key differences in the expression of certain gene products.

Numerous examples of important gene discoveries across hundreds of phenotypes illustrate the tremendous impact GWA scans have had on the study of genetics. Yet as with any experiment, considerable effort is required to design an effective GWA scan while understanding its limitations. Thus, the GWA scan is not the endpoint, but rather the beginning of a journey to elucidating the genetic mechanisms that contribute to human disease. We must, therefore, continue to develop new and innovative solutions to the challenges that arise in our study of human genetics, whether political, economic, or scientific.

Highlights

- Genetics research in Sjögren's syndrome (SS) lags behind other related disorders.
- Genome-wide association (GWA) scans have identified risk loci for many disorders.
- GWA scans in SS promises to identify novel disease loci and pathways.

Acknowledgments

We are grateful to all individuals afflicted with SS as well as volunteers who serve as healthy controls, without whom our ongoing research would not be possible. We thank those clinicians who have volunteered their time to help us evaluate the patients who have visited our research clinic. We would also like to thank all our collaborators, and in particular those collaborators in SGENE, for their continual support, willingness to share resources, and assistance in obtaining vitally necessary samples for our genetic research.

Support for this work was obtained from the US National Institutes of Health grants: P50 AR060804 (K.L.M.), R01 DE018209 (K.L.M.), R01 DE017589 (K.L.M.), R01 AR050782 (K.L.M.), and U19 A1082714 (K.L.M.). This review was also supported by grants from the Sjögren's Syndrome Foundation (K.L.M.), American College of Rheumatology/Research and Education Foundation Professional Graduate Student Preceptorship award (C.J.L.) and the Phileona Foundation (K.L.M.). Additional support was obtained from the Oklahoma Medical Research Foundation Barrett Award Predoctoral Scholarship Fund award (C.J.L. and H.L.).

REFERENCES

1. Binard A, Devauchelle-Pensec V, Fautrel B, Jousse S, Youinou P, Saraux A. Epidemiology of Sjogren's syndrome: where are we now? *Clin Exp Rheumatol*. 2007; 25:1–4. [PubMed: 17417982]
2. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. *Arthritis Rheum*. 2008; 58:15–25. [PubMed: 18163481]
3. Vitali C, Bombardieri S, Jonsson R, Moutsopoulos HM, Alexander EL, Carsons SE, et al. Classification criteria for Sjogren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis*. 2002; 61:554–558. [PubMed: 12006334]
4. Besana C, Salmaggi C, Pellegrino C, Pierro L, Vergani S, Faravelli A, et al. Chronic bilateral dacryo-adenitis in identical twins: a possible incomplete form of Sjogren syndrome. *Eur J Pediatr*. 1991; 150:652–655. [PubMed: 1915519]
5. Bolstad AI, Haga HJ, Wassmuth R, Jonsson R. Monozygotic twins with primary Sjogren's syndrome. *J Rheumatol*. 2000; 27:2264–2266. [PubMed: 10990245]

6. Houghton KM, Cabral DA, Petty RE, Tucker LB. Primary Sjogren's syndrome in dizygotic adolescent twins: one case with lymphocytic interstitial pneumonia. *J Rheumatol.* 2005; 32:1603–1606. [PubMed: 16078341]
7. Scofield RH, Kurien BT, Reichlin M. Immunologically restricted and inhibitory anti-Ro/SSA in monozygotic twins. *Lupus.* 1997; 6:395–398. [PubMed: 9175026]
8. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, et al. A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis Rheum.* 1992; 35:311–318. [PubMed: 1536669]
9. Silman AJ, MacGregor AJ, Thomson W, Holligan S, Carthy D, Farhan A, et al. Twin concordance rates for rheumatoid arthritis: results from a nationwide study. *British journal of rheumatology.* 1993; 32:903–907. [PubMed: 8402000]
10. Cobb BL, Lessard CJ, Harley JB, Moser KL. Genes and Sjogren's syndrome. *Rheum Dis Clin North Am.* 2008; 34:847–868. vii. [PubMed: 18984408]
11. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum.* 2005; 52:1138–1147. [PubMed: 15818688]
12. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409:860–921. [PubMed: 11237011]
13. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science.* 2001; 291:1304–1351. [PubMed: 11181995]
14. The International HapMap Project. *Nature.* 2003; 426:789–796. [PubMed: 14685227]
15. A map of human genome variation from population-scale sequencing. *Nature.* 2010; 467:1061–1073. [PubMed: 20981092]
16. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005; 308:385–389. [PubMed: 15761122]
17. Hindorf, LAJH.; Hall, PN.; Mehta, JP.; Manolio, TA. [Accessed August 29, 2011] A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies
18. Dudbridge F, Gusnanto A. Estimation of significance thresholds for genomewide association scans. *Genetic epidemiology.* 2008; 32:227–234. [PubMed: 18300295]
19. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America.* 2003; 100:9440–9445. [PubMed: 12883005]
20. Risch NJ. Searching for genetic determinants in the new millennium. *Nature.* 2000; 405:847–856. [PubMed: 10866211]
21. Grumet FC, Coukell A, Bodmer JG, Bodmer WF, McDevitt HO. Histocompatibility (HL-A) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease. *The New England journal of medicine.* 1971; 285:193–196. [PubMed: 5087722]
22. Waters H, Konrad P, Walford RL. The distribution of HL-A histocompatibility factors and genes in patients with systemic lupus erythematosus. *Tissue antigens.* 1971; 1:68–73. [PubMed: 4116474]
23. Agnello V, De Bracco MM, Kunkel HG. Hereditary C2 deficiency with some manifestations of systemic lupus erythematosus. *Journal of immunology.* 1972; 108:837–840.
24. Hauptmann G, Grosshans E, Heid E. Lupus erythematosus syndrome and complete deficiency of the fourth component of complement. *Bollettino dell'Istituto sieroterapico milanese.* 1974; 53(suppl):228.
25. Nishino H, Shibuya K, Nishida Y, Mushimoto M. Lupus erythematosus-like syndrome with selective complete deficiency of C1q. *Annals of internal medicine.* 1981; 95:322–324. [PubMed: 7271093]
26. Edberg JC, Langefeld CD, Wu J, Moser KL, Kaufman KM, Kelly J, et al. Genetic linkage and association of Fcγ receptor IIIA (CD16A) on chromosome 1q23 with human systemic lupus erythematosus. *Arthritis Rheum.* 2002; 46:2132–2140. [PubMed: 12209518]

27. Salmon JE, Millard S, Schachter LA, Arnett FC, Ginzler EM, Gourley MF, et al. Fc gamma RIIA alleles are heritable risk factors for lupus nephritis in African Americans. The Journal of clinical investigation. 1996; 97:1348–1354. [PubMed: 8636449]
28. Prokunina L, Castillejo-Lopez C, Oberg F, Gunnarsson I, Berg L, Magnusson V, et al. A regulatory polymorphism in PDCD1 is associated with susceptibility to systemic lupus erythematosus in humans. Nature genetics. 2002; 32:666–669. [PubMed: 12402038]
29. Lee YH, Woo JH, Choi SJ, Ji JD, Song GG. Association of programmed cell death 1 polymorphisms and systemic lupus erythematosus: a meta-analysis. Lupus. 2009; 18:9–15. [PubMed: 19074163]
30. Chung SA, Criswell LA. PTPN22: its role in SLE and autoimmunity. Autoimmunity. 2007; 40:582–590. [PubMed: 18075792]
31. Orru V, Tsai SJ, Rueda B, Fiorillo E, Stanford SM, Dasgupta J, et al. A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. Human molecular genetics. 2009; 18:569–579. [PubMed: 18981062]
32. Cunninghame Graham DS, Manku H, Wagner S, Reid J, Timms K, Gutin A, et al. Association of IRF5 in UK SLE families identifies a variant involved in polyadenylation. Human molecular genetics. 2007; 16:579–591. [PubMed: 17189288]
33. Kozyrev SV, Lewen S, Reddy PM, Pons-Estel B, Witte T, Junker P, et al. Structural insertion/deletion variation in IRF5 is associated with a risk haplotype and defines the precise IRF5 isoforms expressed in systemic lupus erythematosus. Arthritis Rheum. 2007; 56:1234–1241. [PubMed: 17393452]
34. Kelly JA, Kelley JM, Kaufman KM, Kilpatrick J, Bruner GR, Merrill JT, et al. Interferon regulatory factor-5 is genetically associated with systemic lupus erythematosus in African Americans. Genes and immunity. 2008; 9:187–194. [PubMed: 18288123]
35. 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. The New England journal of medicine. 2007; 357:977–986. [PubMed: 17804842]
36. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell. 2008; 134:587–598. [PubMed: 18724932]
37. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40:211–216. [PubMed: 18204447]
38. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. The New England journal of medicine. 2008; 358:900–909. [PubMed: 18204098]
39. 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, PXX, KIAA1542 and other loci. Nature genetics. 2008; 40:204–210. [PubMed: 18204446]
40. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40:1059–1061. [PubMed: 19165918]
41. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nature genetics. 2009; 41:1234–1237. [PubMed: 19838193]
42. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, Adler A, et al. Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. American journal of human genetics. 2011; 88:83–91. [PubMed: 21194677]
43. Gorlov IP, Gorlova OY, Frazier ML, Spitz MR, Amos CI. Evolutionary evidence of the effect of rare variants on disease etiology. Clinical genetics. 2011; 79:199–206. [PubMed: 20831747]

Figure 1.1. Number of genetics publications since 1970 for Sjögren's syndrome and several autoimmune diseases

This figure illustrates the differences in the number of genetics publications generated since 1970 for a variety of autoimmune diseases. Data was generated by searching PubMed for the autoimmune disease and the term "genetics" (e.g. "rheumatoid arthritis [AND] genetics"), while restricting the publications to those written in English and published between January 1 and December 31 for the year in question. Note that Sjögren's syndrome consistently has the fewest genetics publications per year when compared with several other well-studied autoimmune diseases.

Figure 1.2. Linkage disequilibrium (LD) between SNPs within the human genome

LD (or correlation) is present between variants within the genome where portions of sequence are inherited non-randomly as units, and is typically expressed as r^2 values. An example is shown with regions of strong LD and weak LD. Each diamond represents the r^2 values between any 2 SNPs with the shading of each diamond being proportional to the r^2 value (e.g. black is $r^2 = 1.0$). In this example, the r^2 value between SNP1, SNP2, and SNP3 is $r^2 = 0.99$. The r^2 between SNP8 and SNP9 is 0.05.

Figure 1.3. Power to detect association

This graph illustrates 3 different scenarios for studies with sample sizes of 100 cases-100 controls (blue line; majority of manuscripts published to date in SS), 1000 cases-1000 controls (green dashed line; most current GWA scans), and 5000 cases-5000 controls (gold dashed line; future GWA scans and some current meta-analysis between multiple GWA studies). Power to detect genetic association is a function of not only sample size, but also the allele frequency and the odds ratio (or increase in disease risk of an allele) of causal alleles can influence the ability to detect association.

Table 1

Summary of association studies in SS for genetic loci outside the HLA.

Chromosome	Gene of Interest	Phenotype Studied	No. of Cases	No. of Controls	P-value	Odds Ratio	Publication Referenced
1	CHRM3	SS	532	532	0.0033	1.93	Appel et al. Ann Rheum Dis 2011;70:1327
1	FCGR3B	SS with <2 copies of FCGR3B	774	409	0.074	2.01	Mantani et al. Genes Immun 2010;1:155.
1	GSTM1	SS with >2 copies of FCGR3B	774	409	0.048	2.26	Mantani et al. Genes Immun 2010;1:155.
1		SS	106	143	0.035	1.72	Morinobu et al. Arthritis Rheum 1999;42(12):2612.
1	IL10	s-IgG concentration in SS	28	9	0.012	-	Origuchi et al. Ann Rheum Dis 2003;62:1117.
1		SS	129	96	0.036	2.25	Gottenberg et al. Arthritis Rheum 2004;50(2):570.
1	PTPN22	early onset SS	63	150	0.001	-	Font et al. Rheumatology (Oxford) 2002;41(9):1025.
1	TNFSF4	SS	70	308	0.01	2.42	Gomez et al. Genes Immun 2005;6(7):628.
2	CLTA4	SS	540	532	0.00074	1.34	Nordmark et al. Genes Immun 2011;12:100.
2		anti-Ro or anti-La in SS	391	532	0.00076	1.46	Nordmark et al. Genes Immun 2011;12:100.
2		SS	111	156	0.032	1.78	Downie-Doyle et al. Arthritis Rheum 2006;54(8):2432.
2	Ig KM	anti-La in SS	6	56	0.016	-	Pertovaara et al. J Rheumatol 2004;31:2175.
2		LSG histological severity in SS	35	27	0.004	-	Pertovaara et al. J Rheumatol 2004;31:2175.
2		p-IgG in SS	35	27	0.002	-	Pertovaara et al. J Rheumatol 2004;31:2175.
2		s-β2-m concentration in SS	35	27	0.024	-	Pertovaara et al. J Rheumatol 2004;31:2175.
2	IL1RN	SS	36	100	0.04	2.38	Perrier et al. Clin Immunol Immunopathol 1998;87(3):309.
2	STAT4	SS	120	1112	0.01	1.47	Korman et al. Genes Immun 2008;9(3):267.
2		SS	368	711	0.0014	1.41	Nordmark et al. Genes Immun 2009;10:68.
2		SS	540	532	0.0007	1.4	Nordmark et al. Genes Immun 2011;12:100.
2		anti-Ro or anti-La in SS	391	532	0.00069	1.44	Nordmark et al. Genes Immun 2011;12:100.
3	CCR5	SS	39	76	0.043	0.35	Petrek et al. Clin Exp Rheumatol 2002;20(5):701.
4	IL2-IL21	SS	94	368	0.033	0.46	Maiti et al. Arthritis Rheum 2010;62(2):323.
5	EBF1	SS	540	532	0.000099	1.68	Nordmark et al. Genes Immun 2011;12:100.
5		anti-Ro or anti-La in SS	391	532	0.00051	1.65	Nordmark et al. Genes Immun 2011;12:100.
6	C4A	anti-Ro or anti-La in SS	391	532	0.00092	0.45	Nordmark et al. Genes Immun 2011;12:100.
6	TAP2	anti-Ro (+) vs. anti-Ro (-) in SS	51	57	0.001	-	Kumagai et al. Arthritis Rheum 1997;40(9):1685.

Chromosome	Gene of Interest	Phenotype Studied	No. of Cases	No. of Controls	P-value	Odds Ratio	Publication Referenced
6	TNF2	SS	129	96	0.00028 ^c	2.86	Gottenberg et al. Arthritis Rheum 2004;50(2):570.
7	IL6	SS	66	400	<0.0001	-	Hulkkonen et al. Rheumatology (Oxford) 2001;40(6):656.
7	IRF5	SS	210	154	0.01	1.93	Miceli-Richard et al. Arthritis Rheum 2007;56(12):3989.
		SS	368	711	0.000024	1.49	Nordmark et al. Genes Immun 2009;10:68.
		SS	368	711	0.00032	1.57	Nordmark et al. Genes Immun 2009;10:68.
7	IRF-5 / TNPO3	SS	540	532	0.0000055	1.7	Nordmark et al. Genes Immun 2011;12:100.
		anti-Ro or anti-La in SS	391	532	0.0000017	1.81	Nordmark et al. Genes Immun 2011;12:100.
7	TCRBV	SS	61	121	0.018	3	Lawson et al. Ann Rheum Dis 2005;64(3):468.
8	FAM167A-BLK	SS	540	532	0.00047	1.37	Nordmark et al. Genes Immun 2011;12:100.
		anti-Ro or anti-La in SS	391	532	0.00082	1.4	Nordmark et al. Genes Immun 2011;12:100.
10	Fas	anti-Ro/anti-La negative SS	101	108	0.04	-	Mullighan et al. Ann Rheum Dis 2004;63(1):98.
10	MBL	SS	104	143	0.011	-	Wang et al. Ann Rheum Dis 2001;60(5):483.
		SS	104	143	0.024	1.93	Wang et al. Ann Rheum Dis 2001;60(5):483.
		SS	14	129	0.0479	-	Tsutsumi et al. Genes Immun 2001;2(2):99.
11	Ro52	anti-Ro52 SS vs. healthy controls	38	72	0.00003	-	Nakken et al. Arthritis Rheum 2001;44(3):638.
		anti-Ro52 (+) vs. anti-Ro (-) in SS	39	23	0.011	2.67	Imanishi et al. Clin Exp Rheumatol 2005;23(4):521.
13	BAFF	anti-Ro/anti-La in SS	123	136	<0.001	-	Nossent et al. Rheumatology (Oxford) 2008;47:1311.
14	IκBα	SS	98	110	<0.001	16.2	Ou et al. J Clin Immunol 2008;28:440.
		SS	98	110	<0.007	34.14	Ou et al. J Clin Immunol 2008;28:440.
16	IL-4Rα	SS	45	74	0.035	2.6	Youn et al. Immunogenetics 2000;51(8-9):743.
19	ApoE	Early onset SS	63	64	0.0407	-	Pertovaara et al. Rheumatology (Oxford) 2004;43(12):1484.
19	HA-1	SS	88	271	0.003	-	Harangi et al. Eur J Immunol 2005;35(1):305.
19	ILT6	SS	149	749	0.0093	2.65	Kabalak et al. Arthritis Rheum 2009;60(10):2923.
19	TGFβ1	SS with anti-La	129	96	0.0006 ^c	10.2	Gottenberg et al. Arthritis Rheum 2004;50(2):570.
20	CD40	anti-Ro or anti-La in SS	391	532	0.00098	0.73	Nordmark et al. Genes Immun 2011;12:100.
X	MECP2	SS	460	1828	0.0016	1.33	Cobb et al. Ann Rheum Dis 2010;69:1731.