NIH Public Access
Author Manuscript
Autoimmun Rev. Author manuscript; available in PMC 2013 February 1.

Published in final edited form as:
Autoimmun Rev. 2012 February ; 11(4): 267–275. doi:10.1016/j.autrev.2011.10.003.

The Genomics of Autoimmune Disease in the Era of Genome-Wide Association Studies and Beyond

Christopher J Lessard[1,2], **John A Ice**[1], **Indra Adrianto**[1], **Graham Wiley**[1], **Jennifer A Kelly**[1], **Patrick M Gaffney**[1], **Courtney G Montgomery**[1], and **Kathy L Moser**[1]

[1]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73003, USA

Abstract

Recent advances in the field of genetics have dramatically changed our understanding of autoimmune disease. Candidate gene and, more recently, genome-wide association (GWA) studies have led to an explosion in the number of loci and pathways known to contribute to autoimmune phenotypes. Since the 1970s, researchers have known that several alleles in the MHC region play a role in the pathogenesis of many autoimmune diseases. More recent work has identified numerous risk loci involving both the innate and adaptive immune responses. However, much remains to be learned about the heritability of autoimmune conditions. Most regions found through GWA scans have yet to isolate the association to the causal allele(s) responsible for conferring disease risk. A role for rare variants (allele frequencies of <1%) has begun to emerge. Future research will use next generation sequencing (NGS) technology to comprehensively evaluate the human genome for risk variants. Whole transcriptome sequencing is now possible, which will provide much more detailed gene expression data. The dramatic drop in the cost and time required to sequence the entire human genome will ultimately make it possible for this technology to be used as a clinical diagnostic tool.

Keywords

Genetics; Genomics; Genome-wide association study; Autoimmune disease

1. Introduction

By 2006, with a nearly complete draft of the human genome sequence available, the HapMap project had constructed an extensive database of variation within European, Asian, and African genomes [1–3]. These milestones, coupled with scientific and technological advances in microarray technology, now enable scientists to genotype more than 1 million markers in a single experiment, ushering in the ability to perform unbiased surveys of the human genome to identify risk loci in experiments called genome-wide association (GWA) studies. Recently, GWA scans have had a tremendous impact on the study of common, complex diseases in many fields, including the field of autoimmunity. Results of at least one

© 2011 Elsevier B.V. All rights reserved.

[2]Corresponding Author: Christopher J Lessard, Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73003, USA, Tel: +1 (405) 271-1705, Fax: +1 (405) 271-2536, lessardc@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.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

GWA scan have been published for most autoimmune diseases. These GWA studies have often identified novel loci and pathways from which hypothesis-driven research has flourished. In this review, we will discuss the successes of candidate gene and GWA studies, the bottleneck of determining causal variant(s), possible explanations for the majority of heritable disease risk that remains unidentified, and the future of autoimmune genetics in the post-GWA scan era.

2. Pre-GWA Era

Although the precise etiologies of most autoimmune conditions remain elusive, many observational studies have identified families replete with multiple autoimmune diseases [4–6] and high concordance rates have been observed between monozygotic twins [7]. In addition, several studies have found that exposure to environmental triggers, particularly infectious agents, can increase susceptibility to one autoimmune disease while protecting from another [8–10]. These observations have led researchers to hypothesize that the pathophysiology of autoimmune disease likely results from a complex interplay of heritable and environmental factors.

Genetic association with multiple autoimmune diseases was first identified in the 1970s with the major histocompatibility (MHC) region, which encodes the human leukocyte antigens (HLAs). This region was of immediate interest due to its critical role in antigen presentation. Genetic associations with HLA variants and autoimmune phenotypes are typically strong, with relative risks (the ratio of the probability of a particular allele being carried by cases versus controls) greater than 3. Notably, a potent association (relative risk of ~90) exists between the HLA Class I gene, HLA-B27, and ankylosing spondylitis with ~90% of patients carrying the risk allele compared to only 8% of healthy individuals [11]. Reactive arthritis and arthritis in the context of inflammatory bowel disease are also associated with HLA-B27 [12].

Associations with Class II HLA variants have been identified with several autoimmune conditions, including systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), but appear to be weaker than those found with HLA Class I. The associations reported in SLE and RA with Class II are considerably more complex with multiple loci thought to be involved [12]. Interestingly, disease risk associated with particular Class II alleles appears to be increased within particular autoantibody or phenotype subsets of SLE and RA patients [12].

Considerable work prior to the GWA era focused on the identification of associations outside the HLA, particularly in pathways involved in immune system function. Through genome-wide linkage and candidate gene association studies, several loci were identified as risk factors for multiple autoimmune conditions. One such example is PTPN22, which encodes for the protein LYP, a phosphatase specific to lymphoid tissues. This locus was first identified with risk of developing type 1 diabetes (T1D) [13], and was subsequently identified in other autoimmune diseases, including RA, SLE, and autoimmune thyroid disease (Table 1) [14]. The T1D-associated allele within PTPN22 results in a non-synonymous substitution that changes the amino acid at position 620 from an arginine to a tryptophan, and thereby disrupts the association of LYP with CSK (c-sac tyrosine kinase) [13]. Although the exact functional consequence has yet to be determined, studies have shown an increase in phosphatase activity with this amino acid substitution [15].

Interferon regulatory factor 5 (IRF5) was first found to be associated with SLE in 2005 and has since become the most replicated association with SLE outside of the MHC [16]. Three variants have been identified within this locus that, when all present, increase the risk of developing disease [17]. Furthermore, Sigurdsson et al. have found that individuals who

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

carry more than one SLE risk allele for both IRF5 and STAT4 (signal transducer and activator of transcription 4) are at more risk when compared with those individuals who harbor 0 or 1 risk allele [18]. As alleles are added from 2 to 6, the OR increases in a linear fashion. Many autoimmune diseases have now reported association with IRF5 including an autoantibody positive RA cases, multiple sclerosis (MS), and ulcerative colitis among others (Table 1). IRF5 is a transcription factor critical for the mediation of type 1 interferon inflammatory and immune responses and ultimately results in the production of TNF-a, IL-12, and IL-6 after toll-like receptor signaling. Additionally, gene expression profiling studies of SLE, Sjögren's syndrome, MS, RA, and psoriasis have found the overexpression of genes induced by the type 1 interferon pathway [19].

3. Genome-wide Association Era

Though several important associations were identified prior to 2006, copious amounts of data from GWA studies demonstrating genetic associations with disease phenotypes have flooded the literature in recent years. These studies provide unbiased surveys of the genome and give researchers the opportunity to take advantage of linkage disequilibrium (LD; Figure 1). When variants are in high LD with each other (typically with $r^2>0.80$), they are almost always inherited as a unit called a haplotype block. Crossover events break the LD between single nucleotide polymorphisms (SNPs) and cause fragmentation of the haplotype blocks as generations pass. Older populations, such as those of African-decent, tend to have smaller haplotype blocks since crossover events have led to more fragmentation of the LD between variants. SNPs that are in LD are correlated with each other, and thus, they can serve as proxies for one another.

GWA scans exploit LD between SNPs, enabling researchers to assay a manageable number of variants while still capturing the majority of variation in a given population's genome. Since GWA studies test many SNPs for association, a very stringent threshold of statistical significance is needed to decrease the probability of false-positive associations. Therefore, the widely accepted threshold for genome-wide significance is $p < 5 \times 10^{-8}$ based on the Bonferroni correction for multiple testing of 1 million markers (i.e. 0.05 divided by the number of tests performed) [20]. GWA studies have been very successful in the identification of novel loci and pathways contributing to autoimmune disease. Currently, there are approximately 40 ulcerative colitis, 70 Crohn's disease, 35 RA, 35 SLE, 30 MS, and 30 T1D risk loci that have been identified (Table 1; Figure 2). However, the effect sizes reported usually are rather modest, with odds ratios (OR) typically between 1.1 and 1.8. Interestingly, most of the genes identified to date affect more than one autoimmune condition (Table 1; Figure 2).

Although the list of genes identified by GWA scans to date is substantial, few studies have localized the actual causal variant(s). Many genetic associations have been studied in groups of patients with a particular autoimmune condition (e.g. SLE, RA, T1D, etc.) without taking into account the heterogeneity of those particular phenotypes. While it is rare that the loci are specific to a given autoimmune disease, some phenotypes have demonstrated different effects within the same region. This has led to the hypothesis that these regions increase the likelihood of developing autoimmune disease in general rather than a specific phenotype [21]. This highlights the need for clinical subphenotype analysis to further understand how heritable factors might lead to the development of one disease versus another.

Once the initial association signal has been observed, tight LD between markers can impede the refinement of the signal to the variant(s) producing the biologically relevant functional change because of their high correlation with one another. A recent study by Lessard et al. reported an association with SLE at 11p13 and a haplotype approximately 14kb in size of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

strong LD (r^2>0.95) between two genes, PDHX and CD44 [22]. This region contains several putative transcription factor binding sites leading to the hypothesis that the expression of one or both of the neighboring genes could be affected. Since these variants are strongly correlated with each other, statistical analysis alone cannot decipher the effects further to determine the causal allele(s) [22].

The difficulty in localizing observed associations to causal variants as illustrated here has been mirrored time and time again in GWA studies. Despite this obstacle, however, some success stories have emerged. Tumor necrosis factor alpha induced protein 3 (TNFAIP3) was first identified as an RA risk locus [23] and was subsequently reported in SLE [24] and psoriasis [25] among others. Although association with multiple diseases have reported to this locus, the risk alleles appear to be unique. TNFAIP3 encodes the protein A20 that plays a critical role as an attenuator of NF-kB responses. A20 is a dual-functioning enzyme involved in both the de-ubiquitination of TRAF6, RIP1, RIP2, and IKKgamma/NEMO in the NF-kB pathway, as well as an E3 ubiquitin ligase targeting proteins for degradation [26]. Further refinement of the initial association reported in SLE narrowed the risk haplotype to a region of ~109 KB of genomic DNA [26]. Trans-racial mapping, resequencing, and molecular biology have now refined the list of possible causal variants to a di-allelic polymorphism where the non-risk TT genotype is converted to the risk A genotype followed by a deletion (TT/A-) [27]. This region has been found to bind multiple transcription factors, and reduced binding was observed with the risk A- genotype. Adrianto et al. went on to show that reduced binding of transcription factors in the presence of the risk allele in SLE also reduces expression of the mRNA and protein [27]. Reduced attenuation of this pathway could lead to the prolonged activation of the NF-kB pathway, and ultimately prolonged immune responses.

4. Common Versus Rare Variants in the search for causal alleles

Debate has recently erupted in the field of genetics between the common disease/common variant and common disease/multiple rare variant hypotheses. GWA scans are designed to study common variants typically present at an allele frequency of more than 5% and have been exceedingly successful in doing this for autoimmune diseases (Figure 2), offering evidence that the common disease/common variant hypothesis is accurate. However, because GWA studies lack the sensitivity to precisely identify causal variants in loci tagged by common variants and because loci identified by GWA studies account for only a small proportion (typically <10%) of heritability in many diseases [28], researchers now propose that rare variants may contribute significantly to common diseases. Indeed, rare, highly penetrant alleles exist in many common diseases including T1D, blood pressure, colorectal cancer, pancreatitis, Crohn’s disease and heart disease, providing support for the rare variant hypothesis [29–34].

It is also possible that both common and rare variants exist within the same locus and independently influence susceptibility. The interferon induced with helicase C domain 1 (IFIH1) gene was first reported as a risk locus for T1D [35]. Resequencing this region in T1D cases and healthy controls revealed four novel rare variants, each independent from the other, as well as the common variants detected in the GWA scan [30].

Another possibility is that the common variants found in GWA studies may be the result of “synthetic” associations (a genetic association of a genotyped common marker arising from the sum of multiple low-frequency untyped markers) [36]. One example often cited to support the synthetic association hypothesis is the Crohn’s disease risk locus NOD2 (nucleotide-binding olinomerization domain containing 2) that was identified though linkage studies in the pre-GWA era. Linkage was first reported to a region of Chromosome 16

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

encompassing NOD2 and initially referred to as IBD1 [37]. Subsequent fine-mapping and resequencing revealed three rare SNPs within the NOD2 locus as the causal variants [29]. GWA scans also identified an association within this region even though the rare alleles were not typed or tagged well by the genotyping array. The hypothesis is that the total effect sizes of the rare variants (ORs between 3 for a carrier and 38 for a homozygote) are so large that even common variants within the region show association [36]. Certainly, more work needs to be done to elucidate causal variants within regions identified through GWA scans before this debate can be resolved. However, it is likely that both common and rare variants contribute to disease risk.

5. The Future of Genetic Research in Autoimmunity

As described above, tremendous progress has been made in our understanding of the heritable risk factors contributing to the etiology of autoimmune disease. However, a clearer picture of the autoimmune pathophysiology remains elusive despite these efforts. Many in the field of genetics have reported that much of the heritable risk of autoimmune disease and other complex diseases remains to be identified. This is partially because there remains a lack of causal variant identification for those regions found to be associated through GWA approaches or incomplete LD between causal variants and the SNPs in the GWA scan. Moreover, while some Asian GWA scans have been reported (Table 1), the vast majority of phenotypes have focused on subjects of European-decent, leaving African-derived populations largely understudied. It is important that this disparity in particular be resolved since in some diseases, such as SLE, novel loci may contribute to the observed ethnic-specific differences in disease presentation [38, 39]. Studies in African-derived subjects may also help localize causal variants for those regions identified in multiple ethnicities due to the small haplotype bocks.

Few GWA scans have evaluated particular subphenotypes. One reason for this is simply how rare some of these features are in the diseased population, which can decrease power to detect association. Larger sample sizes are needed to address this issue in several phenotypes, including SLE, psoriasis, RA, and others. There also have been examples of genes that are known to interact with each other or are within pathways associated with the same disease (e.g. TNFAIP3 and TNIP1 in SLE and psoriasis). A more comprehensive evaluation of gene-gene interaction is needed to better understand the relationship between these effects.

By far the biggest impact in the future will be from next generation sequencing (NGS). Currently, a large-scale effort, the 1000 Genomes Project, is sequencing multiple populations in search of novel variants in healthy subjects [40]. In addition to the 1000 Genomes Project, studies in many phenotypes are currently utilizing resequencing in regions found though GWA studies to ensure that the majority of variation has been identified before embarking on detailed functional studies. One advantage NGS has over GWA studies is that researchers no longer have to rely on tagging to identify causal polymorphism since all variants are capture simultaneously. NGS can also be used in whole transcriptome sequencing to generate not only more detailed expression data (due to the increased dynamic range as compared to microarrays) from each gene in the genome (annotated or not), but also capture data on all splice variants present within a given sample and microRNAs.

Other genomic features, such as epigenetic status, need to be evaluated in conjunction with the associated regions. A comprehensive assessment of genomic methylation patterns as well as protein and RNA binding to genomic DNA is needed in cases and controls to evaluate the regulation of expression in the disease state. Researchers must also carefully evaluate the impact of environmental influences in combination with genetic predisposition

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to disease to better understand the pathophysiological mechanisms underpinning autoimmune phenotypes.

6. Impact on Clinical Medicine

Revolutions in the study of genetics and genomics will continue to dramatically impact clinical medicine. Certainly, many of the novel loci and pathways identified will lead to the development of therapeutics and will determine those patients who will be the best responders to these new treatments. Once causal variants have been identified, panels of SNPs can be designed to determine if a patient harbors known heritable risk factors, providing information for diagnostic and therapeutic purposes. In addition, with the cost of NGS dropping dramatically as the technology and chemistry improve, it will soon be practical to sequence the entire human genome as a part of a routine diagnostic panel. The use of either of these methods as clinical tools raises many ethical and legal questions that clinicians, researchers, and experts in medical ethics will have to carefully evaluate before they are ever used routinely.

Acknowledgments

Grant Support: PGM (RR020143, AI063274, AR056360, AR058959), CGM (P20 RR020143-06, 1RC2HL101499-01 JIT), KLM (AR043274, AI082714, AI083194, DE018209, AR62277)

We would like to thank Nicholas M Pajewski for providing the R code that helped generate Figure 2 and He Li for his help in preparing Table 1. We also thank all those who are affected by autoimmune disease without whom our work would not be possible.

References

1. 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]
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001; 291:1304–51. [PubMed: 11181995]
3. Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010; 467:52–8. [PubMed: 20811451]
4. Anaya JM, Tobon GJ, Vega P, Castiblanco J. Autoimmune disease aggregation in families with primary Sjogren's syndrome. J Rheumatol. 2006; 33:2227–34. [PubMed: 17086607]
5. Arora-Singh RK, Assassi S, del Junco DJ, Arnett FC, Perry M, Irfan U, et al. Autoimmune diseases and autoantibodies in the first degree relatives of patients with systemic sclerosis. J Autoimmun. 2010; 35:52–7. [PubMed: 20223638]
6. Sestak AL, Shaver TS, Moser KL, Neas BR, Harley JB. Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. J Rheumatol. 1999; 26:1495–9. [PubMed: 10405936]
7. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996; 85:311–8. [PubMed: 8616887]
8. James JA, Harley JB, Scofield RH. Epstein-Barr virus and systemic lupus erythematosus. Curr Opin Rhematol. 2006; 18:462–7.
9. Zandman-Goddard G, Shoenfeld Y. Parasitic infection and autoimmunity. Lupus. 2009; 18:1144–8. [PubMed: 19880560]
10. Blank M, Shoenfeld Y, Perl A. Cross-talk of the environment with the host genome and the immune system through endogenous retroviruses in systemic lupus erythematosus. Lupus. 2009; 18:1136–43. [PubMed: 19880559]
11. Brewerton DA, Hart FD, Nicholls A, Caffrey M, James DC, Sturrock RD. Ankylosing spondylitis and HL-A 27. Lancet. 1973; 1:904–7. [PubMed: 4123836]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

12. Firestein, GS.; Kelley, WN. Kelley's textbook of rheumatology. 8. Philadelphia, PA: Saunders/Elsevier; 2009.
13. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet. 2004; 36:337–8. [PubMed: 15004560]
14. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, et al. Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. AM J Hum Genet. 2005; 76:561–71. [PubMed: 15719322]
15. Vang T, Congia M, Macis MD, Musumeci L, Orru V, Zavattari P, et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet. 2005; 37:1317–9. [PubMed: 16273109]
16. Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G, 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; 76:528–37. [PubMed: 15657875]
17. Graham RR, Kyogoku C, Sigurdsson S, Vlasova IA, Davies LR, Baechler EC, et al. Three functional variants of IFN regulatory factor 5 (IRF5) define risk and protective haplotypes for human lupus. Proc Natl Acad Sci USA. 2007; 104:6758–63. [PubMed: 17412832]
18. Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, Nilsson O, et al. A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. Hum Mol Genet. 2008; 17:2868–76. [PubMed: 18579578]
19. Baechler EC, Batliwalla FM, Reed AM, Peterson EJ, Gaffney PM, Moser KL, et al. Gene expression profiling in human autoimmunity. Immunol Rev. 2006; 210:120–37. [PubMed: 16623768]
20. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273:1516–7. [PubMed: 8801636]
21. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Curr Opin Immunol. 2009; 21:596–605. [PubMed: 19896815]
22. 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. Am J Hum Genet. 2011; 88:83–91. [PubMed: 21194677]
23. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet. 2007; 39:1477–82. [PubMed: 17982456]
24. 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. Nat Genet. 2008; 40:1059–61. [PubMed: 19165918]
25. Nair RP, Duffin KC, Helms C, Ding J, Stuart PE, Goldgar D, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat Genet. 2009; 41:199–204. [PubMed: 19169254]
26. Bates JS, Lessard CJ, Leon JM, Nguyen T, Battiest LJ, Rodgers J, et al. Meta-analysis and imputation identifies a 109 kb risk haplotype spanning TNFAIP3 associated with lupus nephritis and hematologic manifestations. Genes Immun. 2009; 10:470–7. [PubMed: 19387456]
27. Adrianto I, Wen F, Templeton A, Wiley G, King JB, Lessard CJ, et al. Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat Genet. 2011; 43:253–8. [PubMed: 21336280]
28. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. rare allele hypotheses for complex diseases. Curr Opin Genet Dev. 2009; 19:212–9. [PubMed: 19481926]
29. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature. 2001; 411:599–603. [PubMed: 11385576]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

30. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science. 2009; 324:387–9. [PubMed: 19264985]
31. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. Nat Genet. 2008; 40:592–9. [PubMed: 18391953]
32. Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC, et al. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. Proc Natl Acad Sci USA. 2004; 101:15992–7. [PubMed: 15520370]
33. Masson E, Chen JM, Scotet V, Le Marechal C, Ferec C. Association of rare chymotrypsinogen C (CTRC) gene variations in patients with idiopathic chronic pancreatitis. Hum Genet. 2008; 123:83–91. [PubMed: 18172691]
34. Cohen JC, Pertsemlidis A, Fahmi S, Esmail S, Vega GL, Grundy SM, et al. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc Natl Acad Sci USA. 2006; 103:1810–5. [PubMed: 16449388]
35. Wellcome Trust Case Control C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–78. [PubMed: 17554300]
36. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. PLoS Biol. 2010; 8:e1000294. [PubMed: 20126254]
37. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature. 1996; 379:821–3. [PubMed: 8587604]
38. Hart HH, Grigor RR, Caughey DE. Ethnic difference in the prevalence of systemic lupus erythematosus. Ann Rheum Dis. 1983; 42:529–32. [PubMed: 6625702]
39. Petri M. The effect of race on incidence and clinical course in systemic lupus erythematosus: The Hopkins Lupus Cohort. J Am Med Womens Assoc. 1998; 53:9–12. [PubMed: 9458618]
40. Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. [PubMed: 20981092]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Take-Home messages

- The genetic architecture of autoimmune disease is complex with multiple risk loci involved in the pathophysiology. Substantial overlap in susceptibility genes has been identified across multiple autoimmune phenotypes.
- The HLA region contributes to the etiology of multiple autoimmune diseases, but the amount of heritable risk attributed to this region varies widely.
- Genome-wide association studies have been successful in the identification of risk loci and pathways. However, most causal variant(s) remain elusive.
- Despite the extensive progress in identifying autoimmune disease risk loci, most of the heritability has not yet been identified. This is likely due to causal allele(s) and rare variants that remain unidentified.
- Next generation sequencing is rapidly changing our understanding of the human genome though the identification of rare variants, insertion/deletion polymorphisms, and whole transcriptome sequencing.
- In the future, it will be possible to run panels of variants to test patients for risk of developing disease. In addition, the cost and time required to sequence the human genome are dropping rapidly, and therefore, it ultimately may be used as a clinical diagnostic tool.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

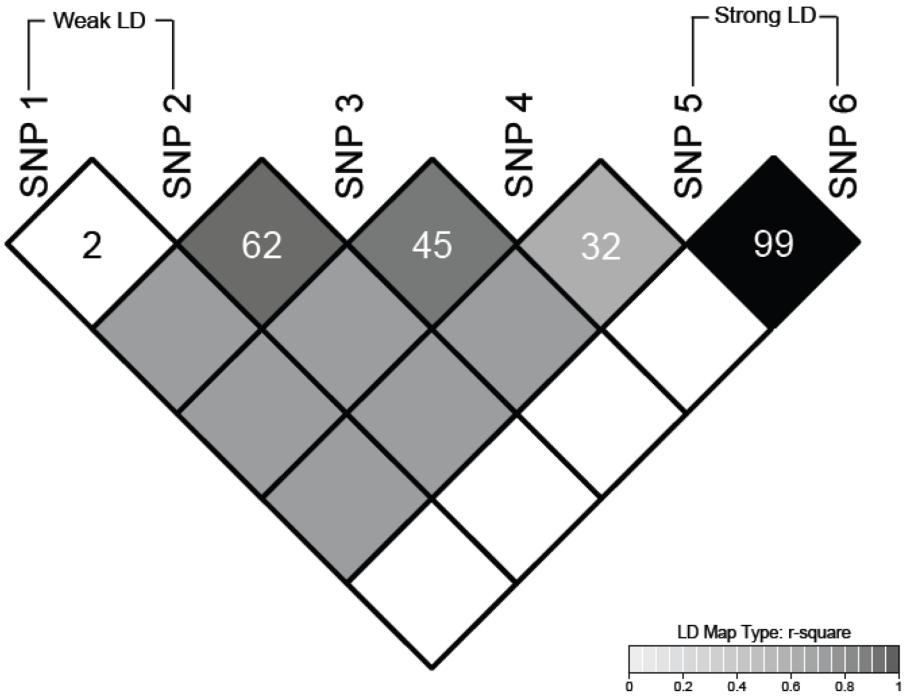

Figure 1. Typical plot of linkage disequilibrium (LD) between variants found in the human genome

LD, or correlation, occurs in the genome when variants are inherited non-randomly as units called haplotype blocks. The degree of LD between variants is typically expressed as r^2 values within the diamonds of the plot above where SNPs 1 and 2 have an r^2=0.2 and SNPs 5 and 6 have an r^2=0.99. The shading of each diamond is also proportional to the r2 value ranging from white (r^2=0) to black (r^2=1.0)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

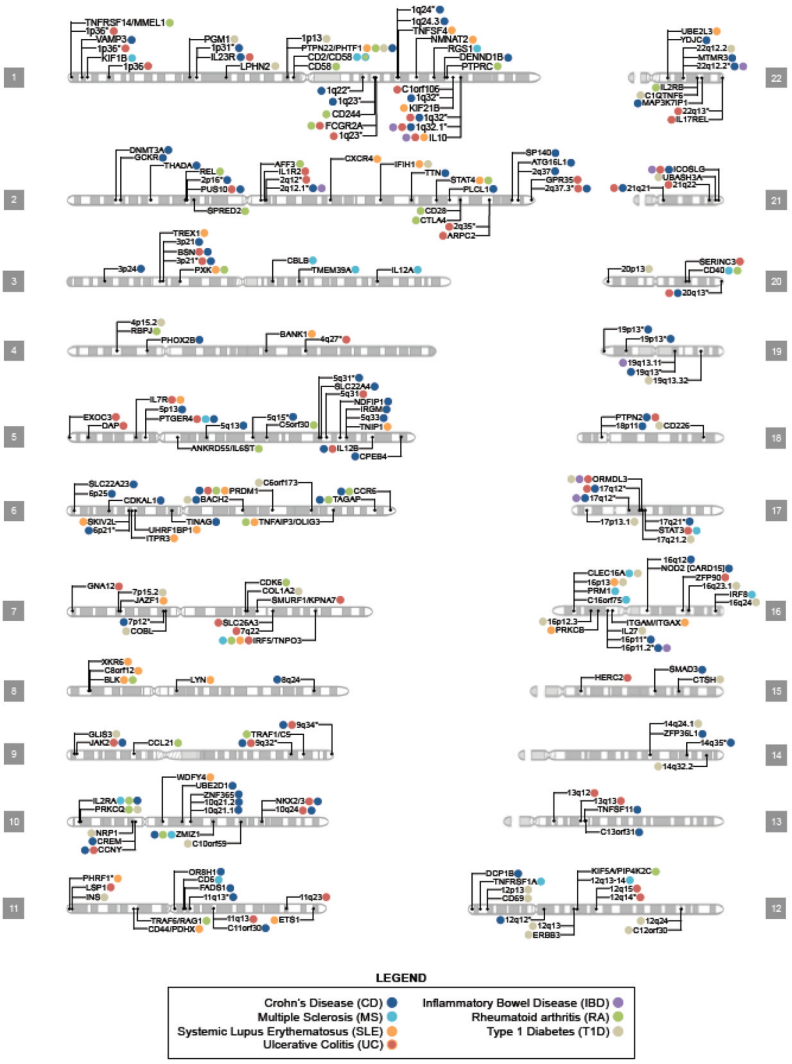

Figure 2. Karyogram of human autosomes depicting loci associated with autoimmune disease

This figure presents select gene/locus associations published since the beginning of the GWAS era for seven autoimmune diseases (AIDs). Genes presented within this figure either exceeded the manuscript's criteria for significance or surpassed genome-wide significance (typically $p < 5 \times 10^{-8}$). Colored dots next to a gene/locus indicate an association with the corresponding AID at that chromosomal location. Regions with asterisks indicate the presence of multiple loci and Table 1 should be consulted to identify the candidate genes suggested by the authors. Some loci lie in close proximity to one another and as a result share the same chromosomal marker even though they represent distinct variants. Note the omission of the MHC region and the sex chromosomes. Like many autoimmune diseases, all seven AIDs presented here have strong associations with the MHC. In our literature search, only two associations were found on chromosome X: (rs2664170, Barrett et al.) with T1D and the association at Xq28 region containing the MECP2/IRAK1 loci with SLE (Webb et al. and Jacob et al.). CD and UC together are commonly referred to as inflammatory bowel disease (IBD); however, some studies perform meta-analysis combining CD and UC results to make an overall IBD combined p-value.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Summary of selected association papers in autoimmune disease

Disease	Loci	First Author	Reference	PubMed ID	Ancestry
Crohn's Disease (CD)	IL23R	Duerr et al.	Science. 314 (5804): 1461–3. 2006.	17068223	European
	APG16L1, CARD15, DCP1B, OR8H1, SLC22A4, TTN, TINAG	Hampe et al.	Nat Genet. 39(2):207–11. 2007.	17200669	European
	10q21.1, ATG16L1, CARD15, IL23R, PHOX2B	Rioux et al.	Na Genet. 39(5):596–604. 2007.	17435756	European
	IL23R, PTGER4 (5p13.1)	Libioulle et al.	PLoS Genet. 3(4):e58. 2007.	17447842	European
	1p31, 2q37, 3p21, 5p13, 5q33, 10q24, 16q12, 18p11, IL23R, NOD2	Welcome Trust Case Control Consortium et al.	Nature. 447(7145):661–78. 2007.	17554300	European
	BSN (3p21.31), CCNY, IL12B, Intergenic (10q21.2), Intergenic (1q24.3), IRGM, NKX2-3, PTPN2, SLC22A23	Franke et al.	Nat Genet. 40(6):713–5. 2007.	18438405	European
	1q32.1 (IL10, IL19. IL20), 2q12.1 (IL18R1, IL18RAP), 16p11 (IL27. SULT1A1, SULT1A2, EIF3C), 17q12 (CCL11, CCL2, CCL7), 22q12.2 (HORMAD2, MTMR3, LIF), ICOSLG1, ZMIZ1	Imielinski et al.	Nat Genet. 41(12):1335–40. 2009.	19915574	European
	1q22 (SCAMP3, MUC1), 1q23 (CD244, ITLN1), 1q24 (C1orf106, KIF21B), 1q24 (TNFSF18, TNFSF4, FASLG), 1q32 (IL10, IL19), 2p16 (C2orf74, REL), 2q12 (IL18RSP, IL12RL2, IL18R1, IL1RL1), 3p21 (MST1, GPX1, BSN), 3p24, 5q13, 5q15, (ERAP2, LRAP), 5q31 (SLC22A4), SLC22A5, IRF1, IL3), 6p21 (LTA, HLA-DQA2, TNF, LST1, LTB), 6p25, 7p12 (IKZF1, ZPBP, FIGNL1), 8q24, 8q24, 9q32, (TNFSF15, TNFSF8), 9q34 (CARD9, SNAPC4), 11q13 (PRDX5, ESRRA), 12q12 (MUC19, LRRK2), 14q35 (GALC, GPR65), 16p11 (IL27, SH2B1, EIF3C, LAT, CD19), 17q12 (CCL2, CCL7), 17q21 (GSMDL, ZPBP2, ORMDL3, IKZF3), 17q21 (MLX, STAT3), 19p13 (GPX4, SBNO2), 19p13 (TYK2, ICAM1, ICAM3), 19q13, 19q13 (FUT2, RASIP1), 20q13 (RTEL, TNFRSF6B, SLC2A4RG), 21q21, ATG16L1, BACH2, C11orf30, C13orf31, CCR6, CDKAL1, CPEB4, CREM, DENND1B, DNMT3A, FADS1, GCKR, ICOSLG, IL12B, IL23R, IL2RA, IRGM, JAK2, MAP3K7IP1, MTMR3, NDFIP1,	Franke et al.	Nat Genet. 42(12):1118–25. 2010.	21102463	European

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Disease	Loci	First Author	Reference	PubMed ID	Ancestry
	NDFIP1, NKX2-3, NOD2, PLCL1, PRDM1, PTGER4, PTPN2, PTPN22, SMAD3, SP140, TAGAP, THADA, TNFSF11, UBE2D1, VAMP3, YDJC, ZFP36L1, ZMIZ1, ZNF365				
	PTPN2, PUS10, TAGAP	Festen et al.	PLoS Genet. 7(1):e1001283. 2011.	21298027	European
Ulcerative Colitis (UC)	3p21.31 (BSN), CCNY, HERC2, NKX2-3, PTPN2, STAT3	Franke et al.	Nat Genet. 40(6):713–5. 2008	18438405	European
	ARPC2, BTNL2, IL10	Franke et al.	Nat Genet. 40(11):1319–23. 2008	18836448	European
	1p36, 12q15, BTNL2/HLA-DQB1 (6p21), IL23R	Silverberg et al.	Na Genet. 41(12):216–20. 2009.	19122664	European
	13q12, FCGR2A, SLC26A3	Asano et al.	Nat Genet. 41(12):1325–9. 2009.	19915573	Asian
	1q32.1 (IL10, IL19, IL20), 2q37.3 (CAPN10, GPR35, KIF1A, RNPEPL1), 17q12 (ORMDL3), 21q22.3 (ICOSLG1)	Imielinski et al.	Nat Genet. 41(12):1335–40. 2009.	19915574	European
	7q22.1 (SMURF1/KPNA7), 22q13.33 (IL17REL)	Franke et al.	Nat Genet. 42(4):292–4. 2010.	20228798	European
	1p36, 1p36 (TNFRSF14, MMEL1, PLH2, C1orf93), 1p36 (TanFRSF9, ERFFI1, UTS2, PARK7), 1q23 (FCGR2A, FCGR2B, HSPA6), 1q32 (IL10, IL19), 2q35 (IL8RA, SLC11A, IL8RB, AAMP, ARPC), 2q37 (GPR35), 3p21 (MST1, UBA7, AMIGO3, GMPPB, BSN), 4q27 (IL21, IL2, ADAD1), 5q31, 6p21, 6p21 (HLA-DRB5, HLA-DQA1, HLA-DRB1, HLA-DRA, BTNL2), 6q23, 7q22, 9q32 (TNFSF15), 9q34 (CARD9, INPP5E, SDCCAG3, SEC16A, SNAPC4), 10q24, 11q13, 11q23, 12q14 (IFNG, IL26), 13q12, 13q13, 16q24, 17q12 (IKZF3, ORMDL3, IKZF3, PNMT, ZPBP2, GSDML), 20q13 (SLC2A4RG, STMN3, ZBTB46, ZGPAT, RTEL1, TNFRSF6B), 21q21, 21q22, 22q13 (PIM3, IL17REL), C1orf106, CCNY, DAP, EXOC3, GNA12, ICOSLG, IL12B, IL1R2, IL23R, IL7R, IRF5/TNPO3, JAK2, LSP1, PDRM1, PTGER4, PUS10, SERINC3, ZFP90	Anderson et al.	Nat Genet. 43(3): 246–52 2011.	21297633	European
Inflammatory Bowel Disease (IBD)	1q32.1 (IL10, IL19, IL20), 2q12.1 (IL18R1, IL18RAP), 16p11 (IL27, SULT1A1, SULT1A2, EIF3C), 17q12 (CCL11, CCL2, CCL7), 19q13.11, 22q12.2 (HORMAD2, MTMR3, LIF), ICOSLG1, ORMDL3, ZMIZ1	Imielinski et al.	Nat Genet. 41(12):1335–40. 2009.	19915574	European
Multiple Sclerosis (MS)	IL2RA, IL7R	Hafler et al.	New Engl J Med. 357 (9):851–62. 2007.	17660530	European

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Disease	Loci	First Author	Reference	PubMed ID	Ancestry
	IL7R	Gregory et al..	Nat Genet. 39(9):1083–91. 2007.	17660817	European
	IRF5	Kristjansdottir et al.	J Med Genet. 45(6):362–369. 2008.	18285424	European
	KIF1B	Aulchenko et al.	Nat Genet. 40(12):1402–3. 2008.	18997785	European
	IL2RA, IL7R, IRF8, MPHOSPH9, PTGER4, RGS1, TNFRSF1A, ZMIZ1	DeJager et al.	Nat Genet. 41(7):776–82. 2009.	19525953	European
	12q13–14, CD40 (20q13), CD58	ANZgene et al.	Nat Genet. 41(7):824–8. 2009.	19525955	European
	CD58, CLEC16A, IL2RA, IL7R	Hoppenbrouwers et al.	J Hum Genet. 54(11):676–80. 2009	19834503	European
	C16orf75, CD58, KIF21B, PRM1, TMEM39A	International Multiple Sclerosis Genetics Consortium et al.	Hum Mol Genet. 19(5):953–62. 2010	20007504	European
	STAT3.	Jakkula et al.	Am J Hum Genet. 86(2):285–91. 2010.	20159113	European
	CBLB	Sanna et al.	Nat Genet. 42(6):495–7. 2010.	20453840	European
Rheumatoid Arthritis (RA)	MHC, PTPN22	Welcome Trust Case Control Consortium et al.	Nature. 447(7145):661–78. 2007.	17554300	European
	C5, TRAF1	Plenge et al.	New Engl J Med. 357(12):1199–209. 2007.	17804836	European
	STAT4	Remmers et al.	New Engl J Med. 357(10):977–986. 2007	17804842	European
	OLIG3/TNFAIP3	Thomson et al.	Nat Genet. 39(12):1431–1433. 2007	17982455	European
	OLIG3, TNFAIP3	Plenge et al.	Nat Genet. 39(12):1477–1482. 2007.	17982456	European
	CCL21, CD40, CDK6, KIF5A/PIP4K2C, MMEL1/TNFRSF14, OLIG3, PRKCQ, PTPN22, TNFAIP3	Raychaudhuri et al.	Nat Genet. 40(10):1216–1223. 2008.	18794853	European
	IL2RA, IL2RB, KIF5A, MMEL1, OLIG3, PPKCQ, TNFAIP3	Barton et al.	Nat Genet. 40(10):1156–1159. 2008	18794857	European
	CD244	Suzuki et al.	Nat Genet. 40(10):1224–1229. 2008	18794858	Asian
	BLK, CTLA4, REL	Gregersen et al.	Nat Genet. 41(7):820–823. 2009	19503088	European
	CD2/CD58, CD28, FCGR2A, PRDM1, PTPRC, TAGP, TRAF6/RAG1	Raychaudhuri et al.	Nat Geriet. 41(12):1313–1318. 2009.	19898481	European
	AFF3, ANKRD55/ILST, C5orf30 CCL21, CCR6, CD40 CTLA4, IL2RA, IRF5, PTPN22, PXK, RBPJ, SPRED2, TNFAIP3	Stahl et al.	Nat Genet. 42(6):509–16. 2010.	20453842	European
Systemic Lupus Erythematosus (SLE)	TYK2, IRF5	Sigurdsson et al.	AmJHumGenet. 76(3):528–537. 2005	15657875	European
	MECP2	Webb et al.	ArthritisRheum. 60(4):1076–84. 2009	19333917	European
	IRAK1	Jacob et al.	Proc Natl Acad Sci USA. 106(15):6256–61. 2009	19329491	European, African, Asian, Hispanic

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Disease	Loci	First Author	Reference	PubMed ID	Ancestry
	C8orf13/BLK, ITGAM/ITGAX	Horn et al.	New Engl J Med. 358(9):900–9. 2008.	18204098	European
	BLK, C8orf12, IRF5/TNPO3, ITGAM, KIAA1542, LYN, NMNAT2, PXK, TNP03, XKR6	International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) et al.	Na Genet. 40(2):204–10. 2008	18204446	European
	BANK1	Kozyrev et al.	Nat Genet. 40(2):211–216. 2008.	18204447	European, Hispanic
	ITGAM	Nath et al.	Nat Genet. 40(2):152–4. 2008.	18204448	European, African
	ITPR3	Oishi et al.	J Hum Genet. 53(2):151–62. 2008.	18219441	Asian
	BLK, STAT4, TNFAIP3	Graham et al.	Nat Genet. 40(9):1059–61. 2008.	19165918	European
	BLK, IFIH1, IL10, IRF5, ITGAM, JAZF1, PHRF1, PRDM1, PTPN22, STAT4, TNFAIP3, TNFSF4, TNIP1, UHRF1BP1	Gateva et al.	Nat Genet. 41(11):1228–33. 2009.	19838195	European
	ETS1, STAT4, WDFY4	Yang et al.	PLoS Genet. 6(2):e1000841. 2010	20169177	Asian
	16p13	Zhang et al.	J Med Genet. 48(1):69–72. 2011.	20805369	Asian
	PRKCB	Sheng et al.	Rheumatology Oxford 50(4):682–8. 2011	21134959	Asian
	CD44/PDHX	Lessard et al.	Am J Hum Genet. 88(1):83–91. 2011	21194677	European, African, Asian
	BANK1, TNFAIP3	Fan et al._	Int J Immuno Genet. 38(2):151–9. 2011	21208380	Asian, European
	TREX1	Namjou et al.	Genes Immun. EpubAOP. 2011	21270825	European
Type 1 Diabetes (T1D)	IFIH1	Smyth et al.	Nat Genet. 38(6):617–9. 2006	16699517	European
	C12orf30, CD226, ERBB3, IFIH1, KIAAD350, NRP1, PHTF1/PTPN22, PTPN22	Todd et al.	Nat Genet. 39(7):857–64. 2007.	17554260	European
	1p13, 12q13, 12q24, 16p13, MHC, PTPN22	Welcome Trust Case Control Consortium et al.	Nature. 447(7145]:661–78. 2007.	17554300	European
	COL1A2, INS, KIAAD350, LPHN2, PTPN22	Hakonarson et al.	Nature. 448(7153):591–4. 2007	17632545	European
	12q13	Hakonarson et al.	Diabetes. 57(4):1143–6. 2008	18198356	European
	INS, UBASH3A	Concannon et al.	Diabetes. 57(4):2858–61. 2008	18647951	Asian, European
	BACH2, C1QTNF6, CTSH, PRKCQ	Cooper et al.	Nat Genet. 40(12):1399–402. 2008.	18978792	European
	4p15.2, 7p15.2, 14q24.1, 14q32.2, 16p12.3, 16q23.1, 17p13.1, 17q21.2, 19q13.32, 20p13, 22q12.2, C10Dorf59, C6orf173, CD69, COBL, GLIS3, IL10, IL27, ORMDL3, PGM1	Barrett et al.	Nat Genet. 41(6):703–10. 2009.	19430480	European