Genetic basis of autoimmunity

Alexander Marson, … , William J. Housley, David A. Hafler


Autoimmune diseases affect up to approximately 10% of the population. While rare Mendelian autoimmunity syndromes can result from monogenic mutations disrupting essential mechanisms of central and peripheral tolerance, more common human autoimmune diseases are complex disorders that arise from the interaction between polygenic risk factors and environmental factors. Although the risk attributable to most individual nucleotide variants is modest, genome-wide association studies (GWAS) have the potential to provide an unbiased view of biological pathways that drive human autoimmune diseases. Interpretation of GWAS requires integration of multiple genomic datasets including dense genotyping, cis-regulatory maps of primary immune cells, and genotyped studies of gene expression in relevant cell types and cellular conditions. Improved understanding of the genetic basis of autoimmunity may lead to a more sophisticated understanding of underlying cellular phenotypes and, eventually, novel diagnostics and targeted therapies.

Find the latest version:
https://jci.me/78086/pdf
Genetic basis of autoimmunity

Alexander Marson,1,2 William J. Housley,3 and David A. Hafler3,4

1Diabetes Center and Division of Infectious Diseases, Department of Medicine, UCSF, San Francisco, California, USA. 2Innovative Genomics Initiative, University of California, Berkeley, California, USA. 3Departments of Neurology and Immunobiology, Yale School of Medicine, New Haven, Connecticut, USA. 4Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Autoimmune diseases affect up to approximately 10% of the population. While rare Mendelian autoimmunity syndromes can result from monogenic mutations disrupting essential mechanisms of central and peripheral tolerance, more common human autoimmune diseases are complex disorders that arise from the interaction between polygenic risk factors and environmental factors. Although the risk attributable to most individual nucleotide variants is modest, genome-wide association studies (GWAS) have the potential to provide an unbiased view of biological pathways that drive human autoimmune diseases. Interpretation of GWAS requires integration of multiple genomic datasets including dense genotyping, cis-regulatory maps of primary immune cells, and genotyped studies of gene expression in relevant cell types and cellular conditions. Improved understanding of the genetic basis of autoimmunity may lead to a more sophisticated understanding of underlying cellular phenotypes and, eventually, novel diagnostics and targeted therapies.

Introduction

Human autoimmune diseases are a major health issue, affecting up to approximately 10% of the population (1). Common human autoimmune diseases are complex disorders that arise from the interactions between polygenic risk factors and environmental factors (2). Intensive investigation of autoimmune disease genetics has the potential to offer an unbiased view of the underlying etiologies of these conditions and, perhaps, identify therapeutic targets. Our early understanding of the disease heritability derived from high rates of autoimmune disease concordance in twins (3, 4) or first-degree family members (5, 6) compared with nonbiological relatives with a shared environment (7). Despite this recognition that autoimmune disease risk is influenced by genetics, it has been challenging to identify the causal nucleotide variants and their functional effects.

The sequencing of the human genome and rapidly emerging genomic technologies are enabling comprehensive interrogation of genetic variants that contribute to autoimmune disease risk. Our understanding of the genetic basis of human autoimmune disease has expanded dramatically in the last 15 years. Here we review biological lessons from genetic studies of human autoimmune diseases. Rare monogenic autoimmune disease syndromes have revealed highly penetrant mutations that disrupt essential mechanisms of central and peripheral immune tolerance. Genome-wide association studies (GWAS) have provided insight into the more subtle immune dysregulation caused by common genetic variants that contribute risk of autoimmunity.

We focus on how new genome sequencing technologies are providing frameworks for the interpretation of GWAS. Human genetic studies of autoimmune diseases benefit from the integration of multiple genomic datasets, including dense genotyping, epigenomic annotation of functional elements in primary immune cells, and large quantitative studies of gene expression in the relevant cell types and cellular conditions. An improved understanding of the genetic basis of autoimmunity will likely lead to a more sophisticated understanding of the cellular phenotypes underlying autoimmune diseases and, eventually, novel diagnostics and targeted therapies.

Monogenic immune dysregulation syndromes

Rare autoimmune disease syndromes have provided insight into biological pathways necessary for the maintenance of immune homeostasis (8). Although far less prevalent than polygenic autoimmune diseases, patients with Mendelian immune dysregulation syndromes caused by monogenic mutations have been identified. Successful linkage analysis and positional cloning combined with mouse genetic models have identified causal mutations. Notably, studies of patients with immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome have highlighted the crucial role of FOXP3 in Treg development and function (9–12). The study of patients with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED; also known as autoimmune polyendocrine syndrome type 1) revealed the essential role of autoimmune regulator (AIRE) in thymic selection and central tolerance (13–18). These highly penetrant mutations in key protein-coding genes suggest pathways, relevant cell types, and mechanisms of tolerance that may undergo more subtle forms of dysregulation as a result of common genetic variation.

Mutations in the gene encoding CTLA4 have recently been discovered in families with Mendelian multi-organ autoimmune disease syndromes (19–21). These rare mutations (including missense, nonsense, and splice variants) cause severe disease in heterozygous patients, albeit with incomplete penetrance. The mutations are associated with impairment of Treg suppressive function and...
immune dysregulation. Common genetic variants in the CTLA4 locus are associated with more modest increases in risk of autoimmune diseases including type 1 diabetes (T1D) and Graves’ disease (22, 23). The functional consequences of a common risk allele (SNP rs3087243 located in the 3′-UTR of the CTLA4; A/G) were investigated using phosphorylated site-specific mAbs targeting components of TCR signaling in naive and memory T cells. The relative responsiveness to TCR stimulation, as assessed by phosphorylation levels of downstream signaling molecules, was altered in naive and memory T cells. The relative responsiveness to TCR stimulation, as assessed by phosphorylation levels of downstream signaling molecules, was altered in naive (CD4+CD45RA+) and memory (CD4+CD45RA-) T cells obtained from individuals with the disease susceptibility allele at CTLA4. This was among the early reports that allelic variation associated with autoimmunity can alter the signaling threshold of CD4+ T cells (24). Taken together, these findings are consistent with a spectrum of allelic variants at the same locus causing differing degrees of immune dysregulation and autoimmune disease risk.

The era of GWAS Identification and characterization of the common variants in autoimmune diseases has been more challenging because of their relatively small individual contributions to disease risk. Common variants have not been purified by negative selection, perhaps a result of their somewhat more modest biological effects as compared with gene disruptions seen in Mendelian diseases. A few notable loci, including MHC on chromosome 6p21 (25–27) and the NOD2 locus (28–30), were initially linked to autoimmunity due to their relatively high odds ratios. However, identifying a larger set of loci that contribute to complex autoimmune disease risk has required the development of new genetic tools.

GWAS are large, case-control studies designed to detect variants that confer a modest risk of common diseases as opposed to rare diseases caused by highly penetrant mutations (31). The basic approach is to identify genetic variants that are preferentially associated with patients with a disease or trait relative to healthy individuals. GWAS using SNPs from the International HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) have allowed an unbiased approach in scanning the whole genome and identifying disease-associated regions, with particular success in identifying a large number of loci associated with human autoimmune diseases.

Thousands of loci have been linked to hundreds of human diseases by GWAS (32, 33). Over 20 autoimmune diseases have been subjected to large GWAS that have successfully implicated risk loci of genome-wide significance. GWAS rapidly revealed shared genetic associations among the different autoimmune diseases (34–36). We recently reported that about two-thirds of loci associated with autoimmunity were shared risk factors for multiple autoimmune diseases (37). These findings are consistent with some shared pathological features of disparate autoimmune diseases and familial clustering of multiple autoimmune conditions.

Perhaps one of the more striking observations in applying genetics to autoimmune diseases is that, while the odds ratio to disease risk for each of the genetic variants is small (most have odds ratios less than 1.2), putative etiological pathways have begun to emerge based on the genes in associated genomic regions (31). Deeper understanding requires more thorough investigation of the specific mechanisms by which these genetic variants influence the biology of human disease.

Biological interpretation of GWAS loci Biological insight from the thousands of loci implicated by GWAS depends on analytic tools for interpretation (Figure 1). A model whereby many common SNPs exert modest effects on a shared biological process predicts that disease-associated SNPs would exert their effect on a limited number of pathogenic cell types. Although common SNPs are inherited in the germ line, they could affect genes that are selectively expressed in particular cell types. Indeed, GWAS loci associated with autoimmune disease are enriched in genes that are preferentially expressed in particular immune subsets (38). A comparison of disease loci with gene expression in 223 cell types of the ImmGen dataset (39) revealed that SLE loci encode genes preferentially expressed in B cell subsets, whereas RA-associated loci are most strongly enriched in genes preferentially expressed in CD4+ effector memory T cells.

Furthermore, knowledge of these pathogenic cell types may suggest biologically relevant genes affected by GWAS variants that do not meet strict genome-wide significance thresholds; shared expression patterns could help to triage GWAS results. Subsequent analysis of human RNAseq data has confirmed enrichment of genes in loci associated with immune-mediated diseases that are preferentially expressed in CD4+ T cell subsets (40).

Not only are candidate genes in GWAS loci coexpressed in pathogenic cell types, but they also encode proteins that participate in a disproportionate number of physical interactions that form biological pathways (41). In addition, integration of multiple annotation tools may help to extract enriched gene sets from loci linked to disease phenotypes (42, 43). Collectively, these efforts have two important goals: triage of candidate genes with linkage disequilibrium blocks and discovery of protein complexes and pathways that are affected by disease variants. Ultimately, discovery of such pathways will highlight the biology underlying complex autoimmune diseases and may direct drug discovery efforts. Protein interaction–based pathway analysis has suggested candidate causal genes within GWAS loci and has implicated key pathogenic pathways, including annotated pathways for JAK/STAT signaling and TCR signaling (44).

The influence of disease variants on cellular phenotype Pathway analysis has begun to reveal patterns among the loci linked to autoimmune diseases, but there is also a need to understand the functional mechanisms by which specific nucleotide variants contribute to disease risk. This has been challenging because the causal SNPs that contribute to autoimmune disease risk tend to be inherited along with neighboring neutral SNPs as a result of linkage disequilibrium. The index SNPs that are genotyped and associated with disease risk in GWAS implicate genomic loci comprised of multiple linked SNPs (Figure 1). Nonetheless, a number of studies have investigated the functional impact of polymorphisms. In one commonly used approach to correlate genotype with cellular phenotype, healthy subjects homozygous for either the risk-associated or protective haplotype are investigated. As the phenotypes are closely related to the genotypes, the numbers of subjects required are significantly less than what is required for elucidation of disease risk. However, in these phenotypic studies on a small number of subjects, the genetic
CD6: genetic modulation of immune receptor isoforms. CD6 is a transmembrane glycoprotein expressed on both B and T cells. The ligand for CD6, activated leukocyte adhesion molecule (ALCAM), is expressed on activated APCs. Interaction of CD6 and ALCAM results in phosphorylation of lymphocyte cytosolic protein 2 (SLP-76) and T cell costimulation (62). A GWAS meta-analysis of MS patients identified a variant (rs17824933) that falls within a broad linkage haplotype containing CD6 (63). Homozygous carriers have diminished expression of the full-length CD6 and an increase in a shortened CD6Δ3 isoform. This isoform lacks the ALCAM binding site, and ligation of CD6 with ALCAM results in decreased proliferation in CD4+ T cells that are homozygous for the risk variant (rs17824933 GG) (64).

PTPN22: shared risk by regulation of multiple immune signals. One of the best-characterized genetic variants linked to multiple autoimmune diseases is a nonsynonymous coding SNP in the PTPN22 gene, which encodes a critical immunoregulatory tyrosine phosphatase. In T cells, PTPN22 negatively regulates TCR signaling by dephosphorylation of the TCR coreceptor CD3 and the tyrosine kinases LCK, FYN, and ZAP70 (45). In myeloid cells, PTPN22 modulates TLR signaling by regulating phosphorylation and ubiquitination of TNF receptor–associated factor 3 and promoting type I IFN production (46). Despite strong evidence implicating this specific nucleotide substitution in PTPN22 as a causal autoimmunity variant, dissecting the specific cellular phenotype or phenotypes has remained challenging.

An 1858C>T polymorphism in PTPN22 is strongly associated with many autoimmune diseases including TID, SLE, Graves' disease, and RA (47–51). Of note, the risk variant for these diseases is not associated with multiple sclerosis (MS) and may be a protective variant in Behcet’s disease and Crohn’s disease (52–56). The 1858C>T variant results in an arginine-to-tryptophan substitution at position 620 (R620W). The impact of the PTPN22 risk variant on TCR signaling remains controversial, with studies suggesting both enhancement and suppression of TCR activation (57, 58). Functional interpretation is further complicated by the possibility of distinct signaling effects in thymic development versus peripheral stimulation and disparate effects on various T cell compartments including naive, memory, and effector subsets and Tregs (59). In B cells, carriers of the risk allele have altered B cell receptor (BCR) signaling and a defect in both central and peripheral B cell tolerance checkpoints, resulting in expansion of autoreactive B cells (60, 61). Through its effects in myeloid cells, the PTPN22 R620W variant may also contribute to autoimmunity by altering TLR signaling and impairing type 1 IFN induction (46).
exon 6 and a premature stop codon. The result is a shortened splice variant in approximately 10% of transcripts. Full-length TNFR1 traffics normally to the cell surface, while the shortened variant accumulates within the cell. The C risk allele results in increased expression of IFN-γ-inducible protein 10 (IP10; also known as CXCL10) from monocytes after stimulation with TNF-α, suggestive of an increased signaling response through TNFR1 (67). Gregory et al. suggested that the risk genotype leads to increased cleavage of TNFR1 from the cell surface, resulting in a soluble decoy receptor capable of blocking TNF-α (68); however, other studies have not observed changes in soluble TNFR1 (67, 69).

The transcription factor NF-kB is a central regulator of inflammation, and a significant number of variants within the NF-kB signaling cascade have been identified in MS and ulcerative colitis. We recently found that MS-associated variants proximal to NFKB1 (rs228614) and in TNFRSFIA (TNFR1, rs1800693) are associated with increased NF-kB signaling after TNF-α stimulation (70). Both variants result in increased degradation of IkBa, a negative regulator of NF-kB and nuclear translocation of p65 NF-kB. The variant proximal to NFKB1 controls signaling responses by altering expression of NF-kB itself, with the GG risk genotype expressing 20-fold more p50 NF-kB. Thus, genetic variants associated with risk of developing MS alter NF-kB signaling pathways and result in enhanced NF-kB activation and greater responsiveness to inflammatory stimuli.

IL-2 receptor α (CD25): complex effects on immune regulation. Mechanistic characterization of causal variants at the IL-2 receptor α (IL2RA; CD25) locus is notably complex, as multiple distinct variants have been associated with different autoimmune diseases. IL2R is a heterotrimeric receptor comprising the IL2Rβ chain (CD122), the common γ chain (CD132), and the high-affinity IL-2Rα chain (CD25). CD25 is upregulated rapidly on T cells after stimulation and is constitutively expressed on Tregs. IL-2 signaling through CD25 stimulates T cell proliferation via phosphorylation of STAT5. Implicated variants at CD25 appear to map to regulatory rather than coding sequences. Recent work suggests that this is a so-called “super-enhancer” locus, where a dense cluster of putative regulatory elements are active in T cells (71).

Multiple studies have attempted to elucidate how genetic variants in this locus affect CD25. For example, the rs2104286 risk allele in CD25 is associated with both MS and T1D, suggesting that alterations in IL-2 signaling may represent a common mechanism across autoimmune diseases. Although the risk variant is associated with increased expression of CD25 on effector T cells and Tregs (72–74), it is also associated with decreased STAT5 signaling downstream of IL-2 (73). Despite diminished IL-2 signaling through STAT5, the MS risk variant was associated with expression of GM-CSF after IL-2 stimulation (72). Recent investigation of five candidate loci within this flanking region and the first intron of NFKB1, sugges-

Genetic fine-mapping

While GWAS have successfully identified broad risk loci, linkage disequilibrium limits their ability to pinpoint causal nucleotide variants within these loci. A risk locus associated with autoimmunity may encode as dozens of proteins as well as noncoding transcripts and regulatory elements. Likely candidate genes within these loci have been identified by systematic efforts to characterize the biological pathways affected by multiple autoimmunity risk variants. Now major efforts are underway to fine map the causal variants based on improved genetic data. Fine mapping requires significantly denser genotyping data than was provided by the initial GWAS studies.

The Immunochip was an international collaborative effort involving investigators of multiple immune-mediated diseases to create a platform to interrogate the autoimmune loci identified to date (76). Whereas initial GWAS genotyping was accomplished with tagging SNPs spread across the whole genome at low density, the Immunochip was an efficient genotyping platform designed to deeply interrogate 186 non-MHC loci with genome-wide significant associations with at least one autoimmune disease (104,425 SNPs, median coverage 486 SNPs/region). It also provided lighter coverage of other genomic regions with suggestive association evidence (49,198 SNPs). GWAS has been replicated for multiple autoimmune diseases, with dense genotyping data generated with large association studies on the Immunochip. This includes MS (66), celiac disease (77), autoimmune thyroid disease (78), primary biliary cirrhosis (PBC) (79), ankylosing spondylitis (80), atopic dermatitis (81), primary sclerosing cholangitis (82), juvenile idiopathic arthritis (JIA) (83), psoriasis (84), inflammatory bowel disease (85), and T1D (23).

Dense genotyping enables statistical analysis to identify a set of candidate causal or “credible” SNPs for a disease phenotype. A Bayesian analysis was pioneered to refine the association signal at 14 loci associated with three different diseases to a set of credible SNPs (86). We recently extended this work to develop a new algorithm for fine-mapping causal variants based on genetic evidence (37). Probabilistic identification of causal SNPs (PICS) is a Bayesian algorithm modeled on dense genotyping data from the recent Immunochip study of MS (66). We were able to assign explicit probabilities of causality to each SNP within GWAS loci based on strength of disease association and haplotype structure. Importantly, we were able to predict causal variants for GWAS data even when dense genotyping data were not available based on imputation to the 1000 Genomes Project (87). PICS is therefore broadly applicable in refining GWAS loci for a wide range of diseases and phenotypes to a tractable set of candidate causal SNPs.

Functional interpretation of noncoding variants

Long-standing knowledge of the genetic code provides a powerful framework to understand deleterious effects of the variants that approach requires identification of the causative SNPs in each haplotype. Use of newer Bayesian approaches to identify causative SNPs (discussed below) in combination with tools such as RNAseq, chromatin mapping, proteomics, and eventually SNP editing in relevant cell types may allow functional testing of the genetic variants found in individual patients to reveal disease-specific pathways for each subject.

Novel approaches to determining the significance of genetic variation

A systematic approach to interrogate the function of genetic variants, including noncoding variants, is needed. This type of
Immunity and provide novel targets for therapeutic intervention. Adapted with permission from might eventually diagnose specific gene regulatory defects caused by autoimmune disease variants binding (circles), chromatin state, target gene regulation, and cellular phenotype. Detailed studies the functional effects of causal autoimmune disease variants in modulating transcription factor (TF) the activity patterns of affected enhancers. Quantitative trait studies are beginning to characterize locations of disease-associated SNPs highlight pathogenic cell types and cell conditions based on dysregulation of gene expression by disease-associated variants. Candidate causal disease variants in well-defined, primary immune cells in both resting and stimulated conditions serve as a powerful resource to identify functional noncoding elements that can be disrupted by disease-associated variants. Candidate causal disease variants (red box) map to regulatory elements (notably enhancers) active in specialized cell types, especially stimulated CD4+ T cell subsets. Enhancers contain binding sites for transcription factors and regulate transcription through long-range interactions with RNA polymerase (RNA Pol) machinery (88). The locations of disease-associated SNPs highlight pathogenic cell types and cell conditions based on the activity patterns of affected enhancers. Quantitative trait studies are beginning to characterize the functional effects of causal autoimmune disease variants in modulating transcription factor (TF) binding (circles), chromatin state, target gene regulation, and cellular phenotype. Detailed studies might eventually diagnose specific gene regulatory defects caused by autoimmune disease variants and provide novel targets for therapeutic intervention. Adapted with permission from Immunity (116).

Figure 2. Integration of genetic and epigenetic fine mapping reveals candidate causal SNPs and pathogenic cell circuits. (A) The goal of genetic fine mapping is to progress from multiple linked SNPs to one (or very few) candidate causal SNPs. Bayesian algorithms have been developed to fine map credible sets of candidate causal SNPs (37, 86) on the basis of dense genotyping data (e.g., ImmunoChip data [ref. 76]) or on the basis of imputation of sparser genotyping data to the 1000 Genomes Project (87). (B) Approximately 90% of causal variants associated with autoimmune diseases are noncoding. Genome-wide chromatin maps of active regulatory elements in primary immune cells in both resting and stimulated conditions serve as a powerful resource to identify functional noncoding elements that can be disrupted by disease-associated variants. Candidate causal disease variants (red box) map to regulatory elements (notably enhancers) active in specialized cell types, especially stimulated CD4+ T cell subsets. Enhancers contain binding sites for transcription factors and regulate transcription through long-range interactions with RNA polymerase (RNA Pol) machinery (88). The locations of disease-associated SNPs highlight pathogenic cell types and cell conditions based on the activity patterns of affected enhancers. Quantitative trait studies are beginning to characterize the functional effects of causal autoimmune disease variants in modulating transcription factor (TF) binding (circles), chromatin state, target gene regulation, and cellular phenotype. Detailed studies might eventually diagnose specific gene regulatory defects caused by autoimmune disease variants and provide novel targets for therapeutic intervention. Adapted with permission from Immunity (116).

alter the amino acid code. However, approximately 90% of causal autoimmune disease variants appear to be noncoding (37). Until recently, the mechanisms by which noncoding DNA variation contributes to autoimmune disease have defied understanding.

Recent studies, including systematic efforts of large consortia, have revealed that intergenic regions are densely populated with hundreds of thousands of cis-regulatory elements, including enhancers that shape cell type–specific gene expression programs; mounting evidence suggests that noncoding variation that contributes to the risk of autoimmune may affect these regulatory elements to cause disease (88). Enhancers are distal cis-regulatory elements (in contrast to promoters, which are cis-regulatory elements proximal to transcriptional start sites) that are essential for proper control of gene expression programs controlling cell identity and condition-specific cellular responses (89). Enhancers contain binding sites for transcription factors and regulate transcription through long-range interactions with RNA polymerase machinery. Their incomplete evolutionary conservation (90) and genomic distance from protein-coding genes have hampered efforts to identify these elements. However, recently stereotyped chromatin patterns of active and poised enhancers have been identified, enabling large-scale sequencing efforts to map regulatory elements throughout the human genome (88, 91, 92).

Genetic variation in humans might dysregulate transcriptional circuits to alter specialized cell functions and contribute to disease risk. This is in accordance with reports that direct targets of key transcriptional regulators in the immune system include genes that have been implicated by GWAS of autoimmune functional noncoding elements in immune cells (refs. 23, 37, and A). To interpret fine-mapped autoimmune disease variants, we generated a large epigenomic resource of active regulatory elements in well-defined, primary immune cells in both resting and stimulated conditions. These epigenomic annotations allowed us to map disease variants to regulatory elements active in specialized cell types, especially stimulated CD4+ T cell subsets (Figure 2B). Although the candidate causal SNPs tend to occur in immune enhancers — often at sites bound by multiple key transcription factors — in most cases, they do not directly disrupt or create recognizable transcription factor binding motifs. Instead, the genomic sites of disease SNPs strongly suggest the importance of noncanonical sequences with crucial roles in immune cell gene regulation (37).

We were able to cluster human diseases based on overlap strength between disease-associated SNPs and cell type–specific enhancer locations. Enhancer activity patterns displayed strict cell type (and even cell condition) restrictions. As a result, the locations of disease-associated SNPs highlighted pathogenic cell types and cell conditions based on the activity patterns of affected enhancers. Genetic data for more diseases and epigenomic data for more cell types under more conditions will continue to fill in a comprehensive map of human diseases by pathogenic cell signatures.

Dysregulation of gene expression by disease-associated variants. Genetic mapping of quantitative traits including gene expression is a complementary approach to epigenomic annotation of functional genomic elements. Considerable efforts have linked expression quantitative trait loci (eQTL) to the loci implicated (93–95). Disease genetics could reveal key regulatory circuits in immune cells that fail in autoimmune diseases as a result of noncoding nucleotide variants, shedding new light on the underlying etiologies of human autoimmune diseases.

Epigenomic annotation of noncoding variants. Initial efforts have suggested that cis-regulatory maps may provide a useful interpretative framework for functional annotation of noncoding disease variants (71, 91, 96–99). However, two major challenges have limited efforts to identify and characterize noncoding variants: (a) GWAS studies do not readily distinguish causal variants from neighboring noncausal variants inherited in linkage disequilibrium and (b) epigenomes and transcriptomes must be mapped in diverse human cell types under multiple physiological conditions to identify the full range of regulatory elements and expression programs that may be affected by noncoding variants to cause disease.

Computational fine mapping approaches based on dense genotyping and imputation to the 1000 Genomes project provide a new opportunity to map causal autoimmune disease variants to
in disease risk (100, 101). Genetic factors affecting immune cell numbers may also contribute to autoimmune disease risk (102). We found that approximately 10% of fine-mapped variants contributing to autoimmune disease risk also contribute to alterations of transcript levels in a population-based study of the transcriptome in peripheral blood cells (37, 103). Despite the statistically significant overlap between disease risk and eQTL, the analysis also suggests that many of the regulatory effects of disease variants will only be observable in restricted cellular contexts. This is consistent with the finding of pervasive response eQTL (reQTL), where expression phenotypes are unmasked in response to selective stimulation conditions (104). Indeed, several recent studies have discovered sets of human reQTL variants that shape transcriptome responses to immune cell stimulation and polarization, demonstrating various degrees of overlap with the loci implicated by GWAS studies of immune-mediated diseases (75, 105–107). eQTL in CD4+ T cells overlap with disease-associated SNPs for RA and MS, whereas SNPs associated with Alzheimer’s and Parkinson’s diseases preferentially overlap with eQTL in monocytes, underscoring the cell type–specific dysregulation of gene expression that likely contributes to disease phenotype (108).

Transcription factor binding studies in inbred mouse strains revealed that genetic variation can significantly alter the genome-wide binding landscape of key immune regulators (109). Quantitative trait studies in human cells suggest that genetic variation has detectable effects of transcription factor binding, gene expression, and chromatin state of regulatory elements (110). Although not statistically significant, recent work suggests that genetic variants associated with asthma risk can be found at Th2 cell enhancers that are differentially active between asthma patients and healthy controls (111). Such studies are beginning to characterize the functional effects of causal autoimmune disease variants in modulating transcription factor binding, chromatin state, gene regulation, and cellular phenotype. Detailed studies might eventually diagnose specific defects of autoimmune disease variants and provide novel targets for therapeutic intervention.

Concluding remarks

Genetic studies of monogenic immune dysregulation syndromes have informed our understanding of essential mechanisms of immune tolerance. New genomic technologies developed over the past decade are now enabling systematic studies of complex autoimmune diseases. Hundreds of relevant loci have been implicated by autoimmune disease GWAS. One approach to understand pathological consequences of GWAS variation is phenotypic analysis of cells from a cohort of genotyped individuals. Here we also discussed systems-scale approaches that have allowed progress from GWAS hits to pathogenic pathways, transcriptional circuits, and immune cell types in specific cellular conditions.

Advances in genomic annotation and integrative analyses should accelerate efforts to connect the human genotypes associated with disease to specific cellular phenotypes. Extensive work has linked specific genetic variations to alterations in amino acid code, transcription, and splice regulation, especially in genes important for cytokine signaling and TCR signaling pathways. There is a growing set of candidate causal SNPs to examine and improved information to inform hypotheses about which biological pathways will be affected. An improved understanding of pathogenic pathways and cell types may eventually help to reveal rare disease-associated variants, epistatic interactions among disease-associated SNPs, and gene/environment interactions, which have been challenging to identify and may contribute to heritability that remains unexplained by GWAS (112). Integrative analyses of autoimmune disease genetics have helped to prioritize likely causal genes in GWAS loci, suggesting not only key biological pathways, but perhaps even drug targets and candidate therapies (113). Looking forward, there is hope that such studies will uncover effects of genetic variants that contribute not only to disease risk, but also to heterogeneity in disease progression (114) and responses to therapy.

Even with rapid progress in genetic analysis, functional studies are essential to advance our understanding of disease genetics (115). Targeted genome editing of disease variants in human cells and animal models will provide opportunities for functional assessment of genetic effects on specific cellular pathways that contribute to autoimmune pathology. These studies will benefit from global analyses of cell types and conditions that are relevant for study; however, considerable work will be required to dissect the subtle and pleiotropic effects of disease variants, which may only be observable in restricted cellular contexts.

Improved analytic and experimental techniques also raise hopes for clinical applications of autoimmune genetics. A refined understanding of the causal variants contributing to autoimmune disease pathology could aid in the development of new diagnostic tests for disease risk and, perhaps, biomarkers for disease progression or therapeutic response. Genetic insight into key pathogenic circuits will focus future drug discovery efforts to correct the specific biochemical and epigenetic pathways that are dysregulated in human autoimmune diseases.

Acknowledgments

We thank C.J. Ye, G.E. Haliburton, and K.K. Farh for suggestions on the manuscript. The review also benefitted from our conversations with all members of the Marson and Hafler laboratories. We are grateful for support from a UCSF Sandler Fellowship (to A. Marson); a gift from Jake Aronov (to A. Marson); a National MS Society Collaborative Research Centre Award (CA1061-A-18, to D.A. Hafler and A. Marson); a Life Sciences Research Foundation Pfizer Fellowship (to W.J. Houssley); NIH grants P01 AI045757, U19 AI046130, U19 AI070352, and P01 AI039671 (to D.A. Hafler); the Penates Foundation (to D.A. Hafler); and the Nancy Taylor Foundation for Chronic Diseases Inc. (to D.A. Hafler).

Address correspondence to: Alexander Marson, University of California, San Francisco, 513 Parnassus Ave., Health Sciences West, Room 1053, San Francisco, California 94143, USA. Phone: 415.502.2611; E-mail: alexander.marson@ucsf.edu. Or to: David Hafler, Yale University School of Medicine, 300 George Street, New Haven, Connecticut 06511, USA. Phone: 203.737.4802; E-mail: david.hafler@yale.edu.
The Journal of Clinical Investigation


