REVIEW

Common genetic factors among autoimmune diseases

Adil Harroud[1,2,3]* and David A. Hafler[4,5,6]*

Autoimmune diseases display a high degree of comorbidity within individuals and families, suggesting shared risk factors. Over the past 15 years, genome-wide association studies have established the polygenic basis of these common conditions and revealed widespread sharing of genetic effects, indicative of a shared immunopathology. Despite ongoing challenges in determining the precise genes and molecular consequences of these risk variants, functional experiments and integration with multimodal genomic data are providing valuable insights into key immune cells and pathways driving these diseases, with potential therapeutic implications. Moreover, genetic studies of ancient populations are shedding light on the contribution of pathogen-driven selection pressures to the increased prevalence of autoimmune disease. This Review summarizes the current understanding of autoimmune disease genetics, including shared effects, mechanisms, and evolutionary origins.

Autoimmune diseases account for considerable population morbidity, particularly among young adults. Collectively, their lifetime prevalence reaches up to 9.4% in the United States (*1*) and up to 10.6% in a recent study of 78 autoimmune diseases in Catalonia (*2*). This heterogeneous group of more than 80 clinically systemic (e.g., systemic lupus erythematosus) or organ-specific (e.g., multiple sclerosis) diseases share common features, such as T cell and antibody reactivity to self-antigens and common association with certain human leukocyte antigen (*HLA*) genetic variants. It is generally thought that autoimmune diseases develop as a result of a breakdown in immune tolerance and activation of autoreactive T cells by autologous or cross-reactive microbial antigens in genetically susceptible individuals. The underlying genetic pathogenesis is evidenced by the high disease concordance among identical twins, and associations with genes in the major histocompatibility complex (MHC) locus have been recognized for over half a century. More recently, the sequencing of the human genome, followed by an understanding of its haplotype structure and the advent of technologies allowing whole-genome analysis of overrepresented haplotypes in disease populations, have led to the elucidation of a substantial degree of disease heritability. Individual haplotypes associated with disease risk have relatively low effect sizes, with MHC odds ratios in the range of 2 to 3 (with notable exceptions, such as *HLA-B*27* in ankylosing spondylitis) and other loci with lesser odds ratios. However, as will be discussed below, the molecular phenotypes associated with these genotypes are notable, and risk genes have, in a nonbiased fashion, implicated the immune system as driving these diseases. Moreover, the common pathophysiology among these diseases is indicated by their shared heritability and genetic overlap.

[1]Department of Neurology and Neurosurgery, McGill University, Montréal, Quebec, Canada. [2]Department of Human Genetics, McGill University, Montréal, Quebec, Canada. [3]The Neuro (Montreal Neurological Institute and Hospital), McGill University, Montréal, Quebec, Canada. [4]Department of Neurology, Yale School of Medicine, New Haven, CT, USA. [5]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. [6]Broad Institute of MIT and Harvard University, Cambridge, MA, USA.
*Email: david.hafler@yale.edu (D.A.H.); adil.harroud@mcgill.ca (A.H.)

"The overwhelming majority (>90%) of putative causal variants associated with autoimmune diseases are in noncoding regions of the genome..."

It can be broadly stated that the underlying causes of autoimmune disease are the unfortunate outcome of gene-environment interactions. Indeed, many variants tied to autoimmune diseases evolved to protect against infectious diseases but have come to be viewed as detrimental in current low-infectious environments, as further described below. Other environmental factors that increase the risk of developing autoimmune disease include diet (*3*), smoking (*4*), Epstein-Barr virus infection (*5*), and other, as-yet-unknown environmental influences. Although the identification of a meaningful proportion of genetic variants associated with autoimmune disease risk has been accomplished, elucidating the environmental factors and their interactions with gene function is considerably more difficult. However, this is necessary to determine the reasons for the reported rise in the prevalence of autoimmunity over the past decades (*6*). This model of autoimmunity driven by autoreactive T cells has led to highly effective therapies. Nevertheless, a paradox remains in that an immunomodulatory therapy that is highly effective in one disease may trigger another autoimmune disease. The potential underlying genetic bases of this observation will also be discussed.

Genetic insights into autoimmune diseases

Autoimmune diseases consistently demonstrate a higher concordance rate among monozygotic twins compared with dizygotic twins (35% and 6%, respectively, for multiple sclerosis) (*7*). In addition, the strongest risk factor for many autoimmune diseases is a positive family history, yet most individuals with the disease have no affected relatives. These observations provided evidence for a substantial genetic component in the etiology of these conditions while posing questions about the expected number of contributing causal variants and their effect sizes. Early studies conducted linkage analysis in families of individuals with autoimmune diseases under the assumption that variants in a small number of genes would harbor large effect sizes and drive genetic risk, similar to single-gene Mendelian disorders. These approaches allowed the discovery of some of the first and strongest susceptibility loci for autoimmune diseases, such as the MHC for type 1 diabetes (*8*) and the nucleotide-binding oligomerization domain-containing protein 2 (*NOD2*) locus in Crohn's disease (*9*). However, these experimental designs were largely unsuccessful in elucidating the genetic architecture of complex diseases.

The development of genotyping array technology and the shift from family-based studies to genome-wide association studies (GWASs) involving large collections of cases and controls has revealed the polygenic nature of autoimmune diseases (*10*). Thousands of robust and replicable genetic associations with autoimmune diseases have been identified by comparing allele frequency between affected and unaffected individuals (Fig. 1). The number of detected loci increases linearly with sample size, whereas effect sizes become smaller, with an estimation that millions of individuals (10 million for inflammatory bowel disease) would be required to fully map the associated genomic regions (*11*). The additive effects of these predominantly common variants (present in 5% or more of the population) typically account for most of the heritability, that is, the fraction of phenotypic variation caused by genetic variation (*12*). Large-scale sequencing studies are also pinpointing rare (often coding) variants associated with autoimmune diseases (*13*). These rare variants typically have larger effect sizes and often converge on the same genes as common variants (*12*, *13*).

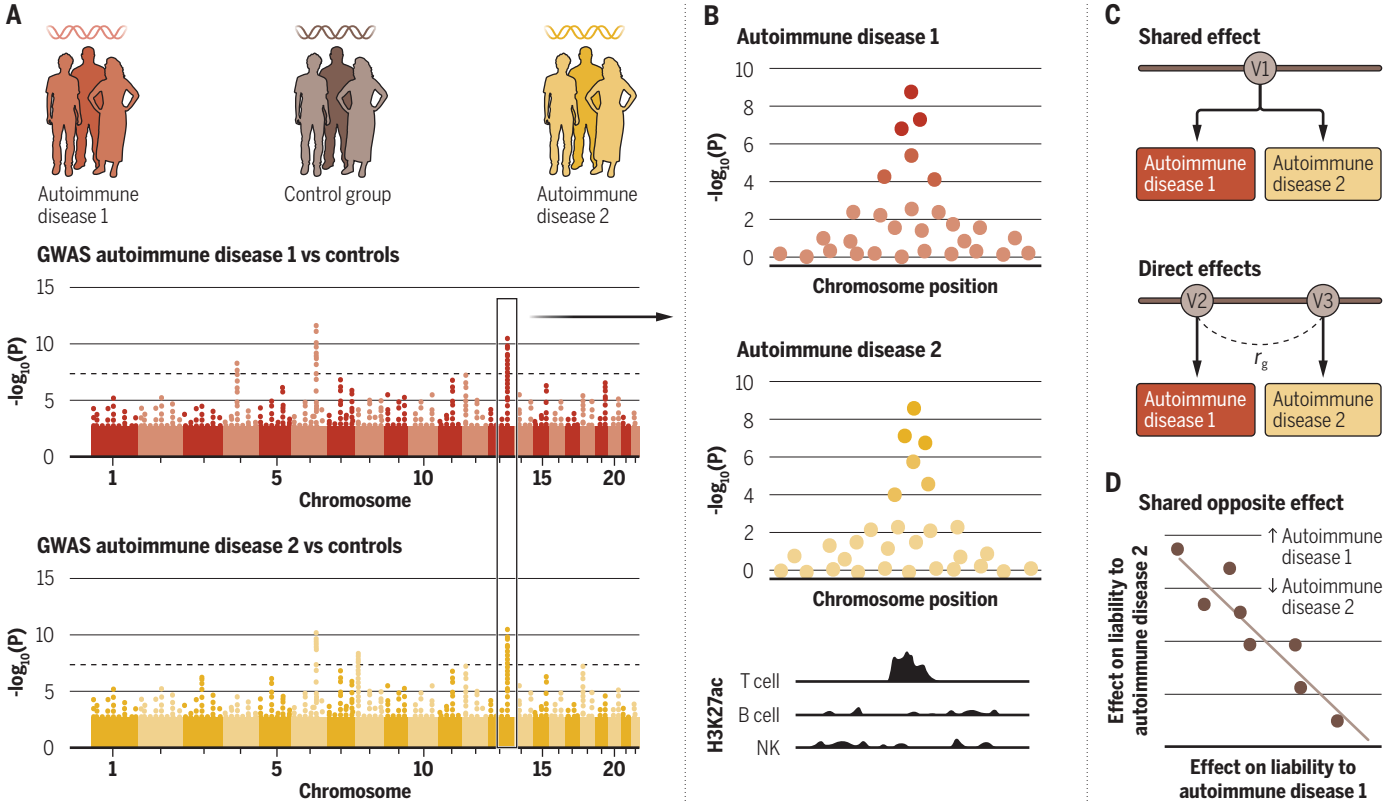

Fig. 1. Exploring shared genetic effects among autoimmune diseases. (**A**) Representation of case-control GWAS designs investigating the genetic susceptibility to two autoimmune diseases by comparing allelic frequencies between affected and unaffected individuals. Genotypes are obtained using microarrays and, increasingly, whole-exome or whole-genome sequencing. The Manhattan plots (bottom) show each variant represented by a point, with its genomic position on the *x* axis and the strength of its association on the *y* axis. The dotted line marks genome-wide significance (typically $P < 5 \times 10^{-8}$). The gray box highlights a locus with overlapping association signals between the two diseases. (**B**) Statistical methods for fine-mapping and colocalization can be applied to pinpoint likely causal variants and assess whether those are shared between the two diseases. These causal variants often localize to enhancers in noncoding regions marked by H3K27ac. These chromatin features (among other approaches) can also be used to implicate disease-causing cell types, such as the T cells shown in this illustration. (**C**) It is important to differentiate between instances in which a single variant (V1) influences multiple diseases, indicating shared molecular mechanisms, and cases in which distinct variants (V2 and V3) at the same locus affect disease risk separately. (**D**) Shared variants often increase or decrease disease risk similarly, although they may occasionally have opposing effects, as represented by the negative correlation in variant coefficients at the shared locus. r_g, genetic correlation.

Evidence from whole-genome sequencing studies in a few complex traits suggests that rare variants (allele frequency <1%) are an important source of heritability not captured by common variants (*14*). Other potential contributors to heritability include gene-environment interactions [such as with the microbiome (*15*)] and gene-gene interactions (*16*).

The overwhelming majority (>90%) of putative causal variants associated with autoimmune diseases are in noncoding regions of the genome, consistent with the observation that <2% of the human genome encodes proteins (*17*). Elucidating the biological consequences of these noncoding variants is necessary for translating genetic discoveries into the clinic, but this presents a number of challenges. The presence of many correlated variants (linkage disequilibrium) at loci implicated by GWASs complicates the identification of causal variants, which can be multiple in a single locus (*18*). Furthermore, even when causal variants are identified with reasonable confidence, such as through fine-mapping (*19*), the functional mechanisms, target genes, and cell types often remain elusive.

Over the past decade, the integration of variant-disease associations with gene regulation traits (such as gene expression, splicing, and chromatin phenotypes) has begun to shed light on the activity of these variants and their cellular contexts. For instance, a study of 21 autoimmune diseases revealed that ~60% of candidate causal variants mapped to enhancers and clusters of enhancer elements (so-called superenhancers) marked by H3 lysine 27 acetylation (H3K27ac) in $CD4^+$ T cell populations and B lymphoblastoid cells (*17*). Candidate causal variants were further enriched within stimulus-dependent enhancers, including those producing regulatory noncoding enhancer RNA (*17*). These observations suggest that most autoimmune disease–associated variants act through transcriptional regulatory mechanisms [such as through altered nuclear factor κB (NF-κB) transcription factor binding] that modulate immune cell fate and function. To illustrate this point, the Crohn's disease variant rs61839660 was mapped by CRISPR activation to a stimulation-dependent intronic interleukin-2 (IL-2) receptor alpha (*IL2RA*) enhancer, and the risk allele was found to impair the timing of *IL2RA* expression in naïve T cells upon stimulation (*20*). IL-2 signaling is essential for the maintenance of forkhead box P3–expressing ($FOXP3^+$) regulatory T cells, which suppress autoimmune responses. The *IL2RA* locus has long been known to harbor variants tied to multiple autoimmune diseases, some with opposite effects on different diseases (*21*).

Overlaying disease-associated variants onto cell type–specific regulatory elements and regions of open chromatin can also be used to

implicate relevant cell types in a disease of interest (Fig. 1). When applied to autoimmune diseases, nearly all variants preferentially mapped to $CD4^+$ T cell subpopulations (*17*, *22*). Some diseases, including multiple sclerosis and systemic lupus erythematosus (*22*), showed additional specificity for B cells, which is consistent with the use of B cell–targeted therapy for the treatment of these conditions (*7*, *23*). Additional commonly identified cell types include natural killer cells, dendritic cells, and mononuclear phagocytes. Other approaches using genome-wide summary statistics (instead of disease-associated loci) annotated on the basis of relative gene expression across different tissues or cell types have yielded similar findings (*24*). Additionally, recent genetic studies have implicated disease-specific cell types, such as mesenchymal cells in Crohn's disease (*13*) and microglia in multiple sclerosis (*25*).

To inform the biological interpretation of disease-linked variants, a common approach is to assess their effect on intermediate gene regulation traits from quantitative trait locus (QTL) studies, which examine the effect of genetic variants on gene expression, splicing, methylation, and chromatin phenotypes, among other molecular phenotypes in specific tissues and cell types. For instance, the integration of GWASs with expression QTLs (eQTLs) can prioritize causal genes in a cell type–specific manner, although only a minority (~25%) of autoimmune disease–associated loci demonstrate strong evidence of colocalization with eQTL effects (*26*). This gap is partly, but not entirely (*27*), explained by data limitations, including the lack of various immune cell types (particularly those less abundant or less well characterized), the limited level of subset resolution, and the lack of dynamic variation (*28*), which is critical given the enrichment of autoimmune disease variants in stimulus-dependent enhancers. This is consistent with a recent study that identified eQTLs active exclusively in immune cells from autoimmune patients, potentially as a result of in vivo stimulation, leading to a high degree of colocalization (63%) with systemic lupus erythematosus risk variants (*29*).

The use of multiplexed single-cell RNA sequencing for eQTL mapping holds promise for further improved cell-type resolution and identification of dynamic regulatory effects. This approach was recently used to characterize the transcriptome and genetic variation across a total of 1,267,758 peripheral blood mononuclear cells from 982 healthy individuals (*30*). The study found that most cis-eQTLs were specific to one of 14 immune cell types. Integration with GWAS risk variants for seven autoimmune diseases also identified 117 non-MHC loci where a disease-associated variant exerted a causal effect through gene expression, most of which (65%) were also cell-type specific. However, because of the differences in power across cell types driven by varying cell proportions and the number of identifiable cells per individual, it is likely that with larger sample sizes, a proportion of eQTLs currently identified as being cell-type specific will be detected in other cell types, potentially with different effect magnitudes. Within a given cell type, single-cell analysis of cell states across differentiation and activation further revealed dynamic eQTLs enriched for colocalization with autoimmune disease loci (*28*, *30*, *31*).

GWASs have identified thousands of genetic variants associated with susceptibility to various autoimmune diseases, most of which are noncoding and reside in stimulation-dependent enhancer elements. These variants may directly disrupt enhancer function, as evidenced by the *IL2RA* example, although their potential consequences are myriad. The integration of disease loci with growing maps of regulatory annotations is rapidly improving the interpretability of those noncoding variants and enabling the generation of precise hypotheses for functional follow-up (*32*). The insights that have emerged so far underscore the importance of cell-type and context specificity and the necessity to further expand the characterization of regulatory elements to encompass rare yet immunologically meaningful cell types and diverse physiological and pathological cell conditions. In light of these challenges and the large number of autoimmune disease–associated variants, it is perhaps unsurprising that downstream functional mechanisms have only been uncovered for a small fraction (Box 1). Finally, the recent development of massively parallel regulatory assays that can test tens of thousands of synthetic noncoding regulatory sequences for functional effects in relevant cell types has the potential to substantially accelerate the discovery of functional mechanisms for autoimmune disease variants (*18*, *33*).

Box 1. From genotype to molecular phenotype.

Understanding how genetic haplotypes dictate biology is the major goal of immunogenetics. Although risk haplotypes have small effects on disease risk, the phenotypes associated with these haplotypes can be substantial. An example is the transcription factor NF-κB, the central regulator of inflammation. GWASs in many autoimmune diseases have identified variants in genes encoding members of the NF-κB signaling cascade, and variants associated with increased risk for multiple sclerosis and ulcerative colitis are strongly enriched within binding sites for NF-κB (*58*, *59*). A multiple sclerosis–associated variant (rs228614) proximal to *NFKB1* was associated with increased NF-κB signaling after TNF-α stimulation and increased degradation of the NF-κB inhibitor. This variant controls signaling responses by altering the expression of NF-κB itself, with homozygous risk allele carriers expressing 20-fold more p50 NF-κB than noncarriers (*58*). NF-κB activation has also been implicated in systemic lupus erythematosus. Two associated variants (rs148314165 and rs200820567) located downstream of TNF alpha-induced protein 3 (*TNFAIP3*), which encodes a negative regulator of NF-κB, were mapped to an enhancer element and found to impair looping interaction with the *TNFAIP3* promoter, resulting in its reduced expression (*59*). The shared consequence of these genetic effects is enhanced NF-κB pathway activity and predisposition to autoimmune disease.

More recently, the immunomodulatory role of a protective variant for multiple sclerosis [rs148755202, which encodes an Arg^{166}→His missense mutation (R166H)] in histone deacetylase 7 (HDAC7) was examined (*60*). Transcriptomic analyses demonstrated that wild-type HDAC7 regulates genes essential for the function of $FOXP3^+$ regulatory T cells (T_{regs}), an immunosuppressive subset of $CD4^+$ T cells that are dysfunctional in patients with multiple sclerosis (*7*). T_{regs} transduced with the protective HDAC7 R166H variant exhibited higher suppressive capacity in in vitro functional assays, mirroring phenotypes previously observed in patients. Moreover, in vivo modeling of the human HDAC7 R166H substitution by the generation of a knock-in mouse model bearing an orthologous R150H mutant demonstrated decreased experimental autoimmune encephalitis severity linked to transcriptomic alterations of brain-infiltrating T_{regs}. Thus, genetic alterations in epigenetic modifiers, a molecular class suitable for therapeutic interventions, can mediate protection from autoimmunity.

Shared heritability and common mechanisms

Different autoimmune diseases co-occur in individuals at higher rates than would be expected given their individual prevalence, and multiple diseases potentially affecting different organ systems cluster in families (*34*). For instance, individuals having a parent with celiac disease harbored a 2.7 relative risk for systemic lupus erythematosus compared with those with unaffected parents in a nationwide Swedish registry (*34*). These observations, initially clinical and then supported by epidemiological evidence, suggested shared risk factors. As hundreds of loci associated with autoimmune diseases were discovered and their genetic architecture better understood, genetic overlaps between disorders indeed became evident.

Clustering on the basis of genetic risk loci highlights a rich network of correlations among

autoimmune diseases, more so than among diseases that affect the same organ systems or that are based on clinical similarity (*17*, *35*). A genome-wide examination of genetic effects, as opposed to statistically significant risk loci, reveals widespread evidence of shared heritability among autoimmune diseases; the pairwise genetic correlation (r_g) averaged 0.39 across a set of five chronic autoimmune disorders and was highest between Crohn's disease and ulcerative colitis (r_g = 0.78) (*36*). These estimates exclude the MHC, in which overlapping haplotypes are among the strongest risk factors for multiple autoimmune diseases. Moreover, the presence of shared genetic loci with opposite effects, increasing risk for one condition while decreasing risk for another (Box 2), can drive negative local genetic correlation and attenuate the overall genome-wide genetic correlation (*37*).

Fig. 2. Shared risk variants in autoimmune diseases. This chord diagram shows the pairwise shared genetic associations driven by the same underlying allele in an Immunochip analysis of six autoimmune diseases. Data are from (*32*). The number of pairwise shared effects for each disease is indicated in parentheses. CeD, celiac disease; IBD, inflammatory bowel disease; MS, multiple sclerosis; T1D, type 1 diabetes; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

Identifying shared genetic effects and resolving their functional implications can uncover common immunopathogenic mechanisms and inform the development or repurposing of rational therapies. It is therefore noteworthy that two-thirds of the variants associated with each of 21 autoimmune diseases were in overlapping loci (*17*). However, proximity within a locus does not necessarily indicate shared underlying mechanisms (Fig. 1). The locus may instead harbor distinct causal variants associated with different diseases. This complexity is compounded by the frequent occurrence of multiple conditionally independent causal variants at a single locus for a given disease. To address these challenges, various colocalization methods have been used to identify instances of true shared molecular effects in which a single variant influences the risk of two or more diseases (*26*, *36*, *38*–*40*). These studies frequently leveraged cohorts genotyped on the Immunochip, a custom array with 196,000 variants that densely covers 186 regions associated with at least one autoimmune condition. Their main findings have been consistent: Approximately half (41 to 60%) of the genetic loci that overlap between two or more autoimmune diseases are attributable to the same underlying genetic effect (Fig. 2). This includes loci with multiple conditionally independent associations, such as signal transducer and activator of transcription 4 (*STAT4*), which harbors two independent genetic effects common to both rheumatoid arthritis and systemic lupus erythematosus (*38*). Most of these shared genetic effects are observed between pairs of diseases, with <10% occurring across four or more autoimmune conditions (*38*). A notable example is a single low-frequency missense coding variant (rs34536443) in *TYK2*, which encodes a nonreceptor tyrosine kinase that is constitutively expressed across immune cell subsets. This variant confers strong uniform protection against the 10 autoimmune conditions in which it was examined (*41*). Homozygosity for the minor allele resulted in disease odds ratios between 0.1 and 0.3 and in a near-complete loss of protein function. This in turn led to impaired signaling for type 1 interferon, IL-12, and IL-23 and consequently perturbation of $CD4^+$ T helper (T_H) type 1 and T_H17 cytokine production (*41*). This study suggested that a drug capable of similarly inhibiting TYK2 function may be beneficial across a range of autoimmune conditions. Indeed, the first selective tyrosine kinase 2 (TYK2) inhibitor, deucravacitinib, was approved by the US Food and Drug Administration (FDA) in September 2022 for the treatment of plaque psoriasis, with ongoing trials in psoriatic arthritis, systemic lupus erythematosus, Crohn's disease, and ulcerative colitis. On the basis of this genetic evidence, TYK2 inhibitors may yet be extended to additional autoimmune conditions. Homozygosity for the same missense variant (rs34536443) also confers a higher risk of mycobacterial disease, including primary tuberculosis, driven by IL-23 disruption (*42*). Therefore, evaluation for active and latent tuberculosis is recommended before initiating deucravacitinib.

These data reveal the existence of a substantial genetic overlap between autoimmune diseases, as evidenced by the presence of pleiotropic variants shared across multiple diseases, occasionally with opposing effects. By leveraging this genetic overlap, cross-disease analysis of existing autoimmune disease cohorts has

Box 2. Genetic basis of treatment outcomes and differential effects in autoimmune diseases.

Inhibitors of TNF-α are widely used and highly effective treatments for a variety of autoimmune diseases, including rheumatoid arthritis, inflammatory bowel disease, and ankylosing spondylitis, with the notable exception of multiple sclerosis. Preclinical studies in an experimental autoimmune encephalomyelitis model suggested that TNF-α depletion would be beneficial for multiple sclerosis. However, a phase 2 randomized trial of TNF-α capture using lenercept, a recombinant TNFα receptor p55-immunoglobulin fusion protein, resulted in dose-dependent disease worsening, leading to early trial discontinuation (*61*). A subsequent GWAS identified a variant (rs1800693) intronic to TNF receptor superfamily member 1A (*TNFRSF1A*). Functional experiments revealed that the risk variant led to exon 6 skipping and premature transcription termination in ~10% of mRNAs, resulting in loss of the transmembrane and intracellular domains and increased expression of a soluble TNF receptor isoform that neutralizes circulating TNF-α, similar to lenercept (*62*). Diseases that benefit from TNF inhibitor treatment either have no association with variants in *TNFRSF1A* (such as rheumatoid arthritis and inflammatory bowel disease) or an opposite association; the risk allele in multiple sclerosis correlates with a protective effect for ankylosing spondylitis of similar but opposite magnitude (*63*). Recently, epidemiological studies have also demonstrated that treatment with TNF-α inhibitors is associated with an increased incidence of multiple sclerosis (*64*).

This example highlights the potential of genetics to inform and predict treatment response and toxicity, including differential effects across diseases. It also illustrates that even though genetic effects may appear small (rs1800693 increases the odds of multiple sclerosis by 15%), pharmacological interventions targeting the same mechanisms can have a substantial impact on outcomes (lenercept increases the relapse rate in multiple sclerosis by up to 68%) (*61*).

been successful in identifying new genetic associations and improving causal variant identification (*36*, *38*). Compelling evidence suggests that some of these shared genetic factors, and consequently shared immunopathology, have evolutionary roots shaped by adaptation to pathogens.

Evolutionary origins of autoimmune diseases

Genetic diversity and variation are shaped by evolutionary forces, including negative selection, a process by which alleles with detrimental effects on fitness are removed from the population. Therefore, the presence of common genetic variants that substantially increase the risk of autoimmune diseases implies that these variants may have provided a beneficial evolutionary trade-off. Considering the immune system's role in combating infections and the strong impact of host genetics on susceptibility to infectious diseases, it has been proposed that autoimmune risk alleles have been preserved at high frequency in the population because of their role in improving resistance to infections (*43*). The canonical example of this antagonistic pleiotropy, genotypes with opposing effects on different traits, is variation of the hemoglobin subunit-β (*HBB*) locus that protects against malaria but causes sickle cell anemia when inherited in a recessive manner. Similarly, genetic variations in the MHC region, which is responsible for a large proportion of the inherited risk for autoimmune diseases, are expectedly also linked with susceptibility to various infections because of the roles of the encoded proteins in antigen presentation and T cell receptor composition (*44*). Recently, a systematic analysis of genetic effects on infectious and autoimmune disorders confirmed that variants associated with both trait categories were significantly more prevalent than expected [by >100-fold (*43*)] (Fig. 3).

Another example is the tumor necrosis factor (TNF) ligand superfamily member 13B (*TNFSF13B*) locus, which was found to be associated with the risk of multiple sclerosis and systemic lupus erythematosus in people from Sardinia (*45*). The causal variant at the locus (rs200748895) results in higher expression of B cell activating factor (BAFF, encoded by *TNFSF13B*) and in turn higher humoral immunity. This variant was several-fold more frequent in people from Sardinia compared with those in mainland Europe (26.5% in Sardinia versus 1.8% in the United Kingdom and Sweden), and the haplotype displayed signatures consistent with positive selection, as opposed to higher frequency arising from random chance (i.e., genetic drift). Because high BAFF expression and antibody production may protect against severe malarial disease, the authors proposed that this positive selection was driven by adaptation to malaria, which was endemic in Sardinia until the 1950s and more prevalent than in mainland Europe. Even though malaria has been eliminated from Sardinia, its genetic imprint may help to explain why the island population currently has some of the highest rates of multiple sclerosis and systemic lupus erythematosus. This example also highlights the potential benefits of studying genetic variation in isolated and diverse populations with varying haplotype frequencies and structures in informing disease biology (*46*).

Although previous evidence of the selective pressures imposed on human populations by specific pathogens has been largely circumstantial, recent advances in ancient DNA preparation and sequencing have provided a direct means of assessing human adaptation (*47*). For instance, a recent study examined the genetic adaptations to the Black Death, a devastating pandemic of plague caused by *Yersinia pestis* that killed 30 to 50% of the populations of Europe, the Middle East, and North Africa in the Middle Ages (*48*). To identify genetic loci under selection from the pandemic, the authors analyzed ancient DNA samples from individuals buried in London cemeteries before (850 to 1250 AD), during (1348 and 1349 AD), and after (1350 to 1539 AD) the Black Death. A cohort of individuals before and after the plague in Denmark provided replication. Overall, 206 individuals were analyzed for 356 immune-related genes and 496 immune disorder GWAS loci. The results showed large changes in allele frequency at immune loci, but not in regions under neutral evolution. The strongest evidence was found for a variant (rs2549794) in endoplasmic reticulum aminopeptidase 2 (*ERAP2*), which conferred an estimated 40% reduction in Black Death mortality in individuals homozygous for the protective allele. The same allele has been associated with increased risk of Crohn's disease, and the corresponding haplotype also increased the risk of other autoimmune diseases, including ankylosing spondylitis, psoriasis, and juvenile idiopathic arthritis. *ERAP2* encodes an aminopeptidase that adapts antigens for binding to MHC class I molecules and presentation to $CD8^+$ T lymphocytes. Likewise, another Black Death–protective allele near cytotoxic T lymphocyte–associated protein 4 (*CTLA4*) increases the risk for celiac disease, rheumatoid arthritis, and systemic lupus erythematosus. It therefore appears that the selection of advantageous alleles in survivors of the plague, and the preservation of their frequencies in descendants, have led to an increased risk of various autoimmune diseases in modern populations.

Just as past selection favored host resistance alleles, it has also selected against variants that weakened immune responses and increased the risk of infection, even when they offered protection against autoimmune diseases. Recent studies have revealed that the *TYK2* missense variant (rs34536443), which reduces the risk of various autoimmune conditions,

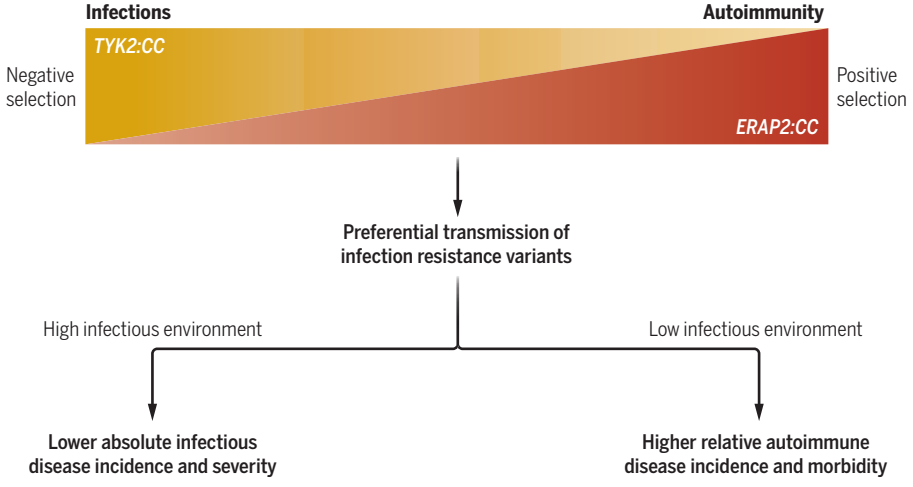

Fig. 3. Antagonistic pleiotropy between risk of autoimmunity and infection. Many genetic variants associated with autoimmune diseases also influence infectious disease risk and severity. Selective pressures from pathogens over millennia have favored variants, such as in *ERAP2*, that provide infection resistance even at the expense of higher autoimmunity (positive selection). Conversely, the same pressures have resulted in decreased frequency of variants, such as in *TYK2*, that protect against autoimmune diseases but increase infectious risk (negative selection). This suggests that the benefits of infection resistance outweigh the risk of autoimmune diseases, especially in environments with high rates of infection (most of human history). These evolutionary pressures have resulted in a higher genetic predisposition to autoimmune diseases, contributing to their high disease burden in modern societies relative to historically low infection rates.

underwent negative selection over the past 2000 years because of an increased predisposition to tuberculosis, resulting in a lower frequency in the population (*43*). This variant also increased the risk of severe COVID-19, consistent with the role of *TYK2* in the type 1 interferon pathway implicated in this form of the disease (*49*).

Throughout history, human immune systems have adapted in response to environmental pressures, perhaps none greater than those imposed by infectious agents. Their genetic imprint favoring more robust immune responses came at the cost of a higher shared genetic risk of various autoimmune diseases. The transition from high levels of infection to lower pathogen exposure in recent times has led to a mismatch between this history of selection and our current environments, contributing to the high disease burden of chronic autoimmune conditions. Furthermore, ancient DNA has also been used to trace the historical origins of autoimmune disease–associated variants (*50*) and revealed that loci from extinct hominins such as Neanderthals and Denisovans are enriched for innate and adaptive immunity genes (*51*).

Concluding remarks

Over the past 15 years, GWASs have substantially advanced our understanding of the genetic architecture of autoimmune diseases. The discovery of hundreds of disease-associated variants, including many overlapping across autoimmune conditions, confirmed their shared genetic and immunopathological etiology. Some of these shared variants have opposing effects on distinct autoimmune diseases, mirroring and offering insight into the differential outcomes observed with TNF-α inhibitors (Box 2). Ancient DNA sequencing analyses are revealing how selection pressure from pathogens over millennia has shaped the human immune genetic repertoire, contributing to higher frequencies for some common genetic factors and consequently to a higher prevalence of autoimmunity. The interpretation of these mostly noncoding risk variants has been challenging, although this is being addressed through multidimensional integration with gene regulation traits at cell-type– and context-specific resolution using single-cell sequencing and data from individuals with autoimmune diseases. Functional studies have begun to shed light on the molecular consequences of several autoimmunity variants, and the development of high-throughput approaches to functional characterization will accelerate this further. As evidenced by the observation that two-thirds of drugs approved by the FDA in 2021 were supported by genetic evidence (*52*), unraveling the biological mechanisms behind these variants has the potential to improve prevention and treatment options for these diseases. Furthermore, the use of polygenic predictors from GWAS results can help to identify healthy individuals at elevated risk of disease, potentially enhancing screening and allowing for presymptomatic interventions, such as in type 1 diabetes prevention trials (*53*, *54*).

Many autoimmune diseases affect population groups differently (*1*), and examining the role of genetic variation in these differences may yield valuable insights, as exemplified by the study of *BAFF* in Sardinia (*45*). However, most genomic studies thus far have been in participants of European ancestry. Going forward, major efforts will be required to expand genetic and epigenetic data collections to diverse ancestries around the world. Moreover, large collections of longitudinally followed and deeply phenotyped cohorts are required to identify genetic variants that contribute, not only to disease risk, but also to heterogeneity in disease progression, because these may differ (*55*, *56*). Finally, although this Review has focused on classic autoimmune diseases, genetic studies have revealed a role for autoimmunity in other conditions. For instance, GWASs in Parkinson's disease have identified HLA associations, and subsequent research reported frequent T cell responses to α-synuclein, which forms pathogenic protein aggregations in this disease (*57*). These and other findings in Alzheimer's disease, atherosclerosis, and cancer are broadening the view of inflammation and autoimmunity across human diseases.

REFERENCES AND NOTES

1. G. S. Cooper, M. L. K. Bynum, E. C. Somers, *J. Autoimmun.* **33**, 197–207 (2009).
2. A. Sisó-Almirall *et al.*, *Autoimmun. Rev.* **19**, 102448 (2020).
3. G. S. Temba *et al.*, *Nat. Immunol.* **22**, 287–300 (2021).
4. K. H. Costenbader, E. W. Karlson, *Lupus* **15**, 737–745 (2006).
5. K. Bjornevik *et al.*, *Science* **375**, 296–301 (2022).
6. G. E. Dinse *et al.*, *Arthritis Rheumatol.* **74**, 2032–2041 (2022).
7. M. Filippi *et al.*, *Nat. Rev. Dis. Primers* **4**, 43 (2018).
8. S. S. Rich, L. R. Weitkamp, J. Barbosa, *Am. J. Hum. Genet.* **36**, 1015–1023 (1984).
9. J. P. Hugot *et al.*, *Nature* **379**, 821–823 (1996).
10. E. Uffelmann *et al.*, *Nat. Rev. Methods Primers* **1**, 59 (2021).
11. L. Yengo *et al.*, *Nature* **610**, 704–712 (2022).
12. D. J. Weiner *et al.*, *Nature* **614**, 492–499 (2023).
13. A. Sazonovs *et al.*, *Nat. Genet.* **54**, 1275–1283 (2022).
14. P. Wainschtein *et al.*, *Nat. Genet.* **54**, 263–273 (2022).
15. D. R. Plichta, D. B. Graham, S. Subramanian, R. J. Xavier, *Cell* **178**, 1041–1056 (2019).
16. J. Domingo, P. Baeza-Centurion, B. Lehner, *Annu. Rev. Genomics Hum. Genet.* **20**, 433–460 (2019).
17. K. K.-H. Farh *et al.*, *Nature* **518**, 337–343 (2015).
18. N. S. Abell *et al.*, *Science* **375**, 1247–1254 (2022).
19. D. J. Schaid, W. Chen, N. B. Larson, *Nat. Rev. Genet.* **19**, 491–504 (2018).
20. D. R. Simeonov *et al.*, *Nature* **549**, 111–115 (2017).
21. L. M. Maier *et al.*, *PLOS Genet.* **5**, e1000322 (2009).
22. A. T. Satpathy *et al.*, *Nat. Biotechnol.* **37**, 925–936 (2019).
23. J. T. Merrill *et al.*, *N. Engl. J. Med.* **386**, 1034–1045 (2022).
24. H. K. Finucane *et al.*, *Nat. Genet.* **50**, 621–629 (2018).
25. N. A. Patsopoulos *et al.*, *Science* **365**, eaav7188 (2019).
26. S. Chun *et al.*, *Nat. Genet.* **49**, 600–605 (2017).
27. H. Mostafavi, J. P. Spence, S. Naqvi, J. K. Pritchard, bioRxiv 491045 [Preprint] (2022); https://doi.org/10.1101/2022.05.07.491045.
28. B. Soskic *et al.*, *Nat. Genet.* **54**, 817–826 (2022).
29. M. Ota *et al.*, *Cell* **184**, 3006–3021.e17 (2021).
30. S. Yazar *et al.*, *Science* **376**, eabf3041 (2022).
31. A. Nathan *et al.*, *Nature* **606**, 120–128 (2022).
32. K. Kundu *et al.*, *Nat. Genet.* **54**, 251–262 (2022).
33. K. Mouri *et al.*, *Nat. Genet.* **54**, 603–612 (2022).
34. K. Hemminki, X. Li, K. Sundquist, J. Sundquist, *Arthritis Rheum.* **60**, 2845–2847 (2009).
35. A. Verma *et al.*, *Am. J. Hum. Genet.* **104**, 55–64 (2019).
36. D. Ellinghaus *et al.*, *Nat. Genet.* **48**, 510–518 (2016).
37. W. van Rheenen, W. J. Peyrot, A. J. Schork, S. H. Lee, N. R. Wray, *Nat. Rev. Genet.* **20**, 567–581 (2019).
38. M. R. Lincoln *et al.*, bioRxiv 2021.05.13.21257044 [Preprint] (2021); https://doi.org/10.1101/2021.05.13.21257044.
39. M. D. Fortune *et al.*, *Nat. Genet.* **47**, 839–846 (2015).
40. H.-J. Westra *et al.*, *Nat. Genet.* **50**, 1366–1374 (2018).
41. C. A. Dendrou *et al.*, *Sci. Transl. Med.* **8**, 363ra149 (2016).
42. S. Boisson-Dupuis *et al.*, *Sci. Immunol.* **3**, eaau8714 (2018).
43. G. Kerner *et al.*, *Cell Genomics* **3**, 100248 (2023).
44. K. Ishigaki *et al.*, *Nat. Genet.* **54**, 393–402 (2022).
45. M. Steri *et al.*, *N. Engl. J. Med.* **376**, 1615–1626 (2017).
46. G. Sirugo, S. M. Williams, S. A. Tishkoff, *Cell* **177**, 1080 (2019).
47. Y. Liu, X. Mao, J. Krause, Q. Fu, *Science* **373**, 1479–1484 (2021).
48. J. Klunk *et al.*, *Nature* **611**, 312–319 (2022).
49. M. E. K. Niemi, M. J. Daly, A. Ganna, *Nat. Rev. Genet.* **23**, 533–546 (2022).
50. W. Barrie *et al.*, bioRxiv 2022.09.23.509097 [Preprint] (2022); https://doi.org/10.1101/2022.09.23.509097.
51. A. Liston, S. Humblet-Baron, D. Duffy, A. Goris, *Nat. Immunol.* **22**, 1479–1489 (2021).
52. D. Ochoa *et al.*, *Nat. Rev. Drug Discov.* **21**, 551 (2022).
53. L. A. Ferrat *et al.*, *Nat. Med.* **26**, 1247–1255 (2020).
54. K. C. Herold *et al.*, *N. Engl. J. Med.* **381**, 603–613 (2019).
55. J. C. Lee *et al.*, *Nat. Genet.* **49**, 262–268 (2017).
56. S. Baranzini, S. Sawcer, International Multiple Sclerosis Genetics Consortium, MultipleMS Consortium, Research Square rs.3.rs-1723574/v1 [Preprint] (2023).
57. D. Sulzer *et al.*, *Nature* **546**, 656–661 (2017).
58. W. J. Housley *et al.*, *Sci. Transl. Med.* **7**, 291ra93 (2015).
59. S. Wang, F. Wen, G. B. Wiley, M. T. Kinter, P. M. Gaffney, *PLOS Genet.* **9**, e1003750 (2013).
60. P.-P. Axisa *et al.*, *Sci. Transl. Med.* **14**, eabl3651 (2022).
61. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group, *Neurology* **53**, 457–465 (1999).
62. A. P. Gregory *et al.*, *Nature* **488**, 508–511 (2012).
63. A. Cortes *et al.*, *Nat. Genet.* **45**, 730–738 (2013).
64. L. Li *et al.*, *Neurology* **100**, e558–e567 (2023).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Institutes of Health (grants P01 AI073748, U24 AI11867, R01 AI22220, UM 1HG009390, P01 AI039671, P50 CA121974, and R01 CA227473 to D.A.H.). **Competing interests:** D.A.H. has received research funding from Bristol-Myers Squibb, Novartis, Sanofi, and Genentech and has been a consultant for Bayer Pharmaceuticals, Bristol Myers Squibb, Compass Therapeutics, EMD Serono, Genentech, Juno Therapeutics, Novartis Pharmaceuticals, Proclara Biosciences, Sage Therapeutics, and Sanofi Genzyme. A.H. has been a consultant for Biogen.

Submitted 13 February 2023; accepted 29 March 2023
10.1126/science.adg2992

Common genetic factors among autoimmune diseases

Adil Harroud and David A. Hafler

Science, **380** (6644), .
DOI: 10.1126/science.adg2992