DOI: 10.1111/imr.13329

INVITED REVIEW

Immunological Reviews WILEY

Shaping immunity: The influence of natural selection on population immune diversity

Haley E. Randolph[1,2] | Katherine A. Aracena[3] | Yen-Lung Lin[4] | Zepeng Mu[1] | Luis B. Barreiro[1,3,4,5]

[1]Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, Illinois, USA

[2]Department of Pediatrics, Columbia University Irving Medical Center, New York, New York, USA

[3]Department of Human Genetics, University of Chicago, Chicago, Illinois, USA

[4]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA

[5]Committee on Immunology, University of Chicago, Chicago, Illinois, USA

Correspondence
Luis B. Barreiro, Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA.
Email: lbarreiro@uchicago.edu

Funding information
National Institutes of Health, Grant/Award Number: AI07647, P30-DK042086 and R01-GM134376; Howard Hughes Medical Institute, Grant/Award Number: James H. Gilliam Fellowship for Advanced Study

Summary

Humans exhibit considerable variability in their immune responses to the same immune challenges. Such variation is widespread and affects individual and population-level susceptibility to infectious diseases and immune disorders. Although the factors influencing immune response diversity are partially understood, what mechanisms lead to the wide range of immune traits in healthy individuals remain largely unexplained. Here, we discuss the role that natural selection has played in driving phenotypic differences in immune responses across populations and present-day susceptibility to immune-related disorders. Further, we touch on future directions in the field of immunogenomics, highlighting the value of expanding this work to human populations globally, the utility of modeling the immune response as a dynamic process, and the importance of considering the potential polygenic nature of natural selection. Identifying loci acted upon by evolution may further pinpoint variants critically involved in disease etiology, and designing studies to capture these effects will enrich our understanding of the genetic contributions to immunity and immune dysregulation.

KEYWORDS
environment, genetics, human immune response variation, natural selection, quantitative trait loci

1 | INTRODUCTION

Pathogens are one of the most powerful selection pressures in human evolutionary history.[1] It is hypothesized that the geographic distribution of pathogens varied significantly between human populations as modern humans migrated out of Africa.[2] In turn, this variation in the magnitude and diversity of pathogen exposure across populations likely drove allele frequencies to diverge at loci influencing the host immune response via natural selection. Such past human evolution is expected to be reflected among individuals living today, and the study of modern human genomes coupled with functional immunological assays has the potential to reveal natural selection's contribution to phenotypic variation in immune responses within and between human populations. Insights from these studies can also inform the biological basis of varying susceptibility to infectious and autoimmune diseases across populations.[3–5]

It is well-documented that the incidence and prevalence of many infectious and chronic complex diseases are unequally distributed between populations, particularly considering common

autoimmune disorders, including Crohn's disease (CD), type 1 diabetes, and psoriasis, among others.[4,6–8] For example, the prevalence of inflammatory bowel disease is considerably higher in non-Hispanic White Americans compared to non-Hispanic Black Americans, Hispanic Americans, and Asian Americans.[6,9] On the other hand, the prevalence of systemic lupus erythematosus is substantially higher in these three populations compared to non-Hispanic White Americans.[10,11]

In line with this, risk alleles associated with complex diseases are found at relatively high frequencies in certain populations compared to others,[12] raising the possibility that these now-deleterious loci may have been advantageous targets of local adaptation in the past. For example, nine putatively causal risk loci for CD located in *NOD2*, the first reported CD risk gene, display a combined allele frequency of 13% in individuals of European descent.[13] In contrast, the combined allele frequency across these same variants is only 0.06% among individuals of East Asian descent, suggesting that these risk alleles contribute little CD risk in East Asian individuals yet considerable CD risk in European individuals.[14,15]

In this Review, we discuss current knowledge concerning phenotypic variation in immune responses within and across human populations, the mechanisms driving such variation, and the impact of such diversity on present-day susceptibility to immune-related disorders. Finally, we explore future directions in this field, emphasizing the importance of extending the study of immune response variation to a broader array of human populations, contending that immune responses should be evaluated dynamically across time, and calling attention to the often-disregarded mechanism of polygenic selection on immune responses.

2 | POPULATION VARIATION IN IMMUNE RESPONSES

In vitro challenge studies of primary immune cells or cell lines obtained from individuals with varying genetic ancestry backgrounds represent a powerful approach to measure population differences in the immune response to infection (Figure 1). Indeed, multiple studies have successfully leveraged in vitro infection models to identify transcriptional variation in the immune responses to pathogens across individuals of different genetic ancestries.[16–20] Two flagship studies focusing on the transcriptional profiles of monocytes and macrophages uncovered thousands of genes displaying a significant divergence in the intensity of the response to various pathogens, including *Listeria monocytogenes*, *Salmonella typhimurium*, and influenza A virus (IAV), between African and European ancestry individuals in independent cohorts.[16,17,21] Of note, a greater proportion of global African ancestry was associated with a stronger pro-inflammatory response to bacterial challenge, and individuals with increased African ancestry were also better able to control intracellular bacterial growth in macrophages compared to individuals with increased European ancestry, highlighting the potential functional consequences of the observed response variation.[16]

More recently, there has been a push to include samples obtained from a broader set of populations as well as immune cell types. In particular, various studies have turned to single-cell RNA-sequencing of peripheral blood mononuclear cells (PBMCs) to measure genetic ancestry effects in the context of viral infections.[18–20,22] Because PBMCs are comprised of multiple, distinct immune cell types, single-cell methods can dissect independent signals from each cell population within the same experiment. Across cell types, genetic ancestry effects on gene expression were found to be highly cell type-specific, with the majority of effects only detected in one or two cell types.[18] The interferon response provided one notable exception: following IAV infection, it was one of the most diverged pathways between European Americans and African Americans, a finding conserved across all cell types tested.[18] Notably, increased European ancestry was associated with higher levels of intracellular IAV transcripts early upon infection and a stronger interferon response, implicating potential population-associated variation in viral control mechanisms.[18,20]

3 | THE ROLE OF ENVIRONMENT IN HUMAN IMMUNE RESPONSE VARIATION

Both genetic and nongenetic factors[23,24] shape immune response heterogeneity and defining the relative contribution of these components to immune response diversity is a principal goal of human immunogenomics research. Many studies define genes differentially expressed between populations as those with expression levels significantly correlated with global genetic ancestry. Because of this, it stands to reason that a significant fraction of the signal identified is driven by environmental confounders that are correlated with quantitative genetic ancestry estimates, not genetic associations themselves. Indeed, environmental factors, such as age, sex, microbiome, previous exposure to pathogens, etc., are responsible for immune response variation across individuals to a large extent. Age-related effects on the immune system have been well characterized, and it is known that immune function declines as a consequence of aging. Specifically, elderly individuals produce fewer B and T cells in primary lymphoid organs and harbor immune cells with reduced functional capacity, leading to overall weaker immune responses compared to younger individuals.[25] Likewise, immunological differences associated with sex have been widely described. In general, adult females mount stronger immune responses compared to adult males, resulting in more rapid pathogen clearance, greater vaccine effectiveness, and increased susceptibility to autoimmune and inflammatory diseases in females.[26] Age and sex have also been shown to directly impact the transcriptional response of immune genes in a wide-ranging but cell type-specific manner, with $CD8^+$ T cells mediating age effects and $CD4^+$ T cells and monocytes mediating sex effects on expression.[23]

In addition, an individual's prior infection history partly determines their subsequent immune responses to previously encountered and novel pathogens owing to adaptive immune memory,

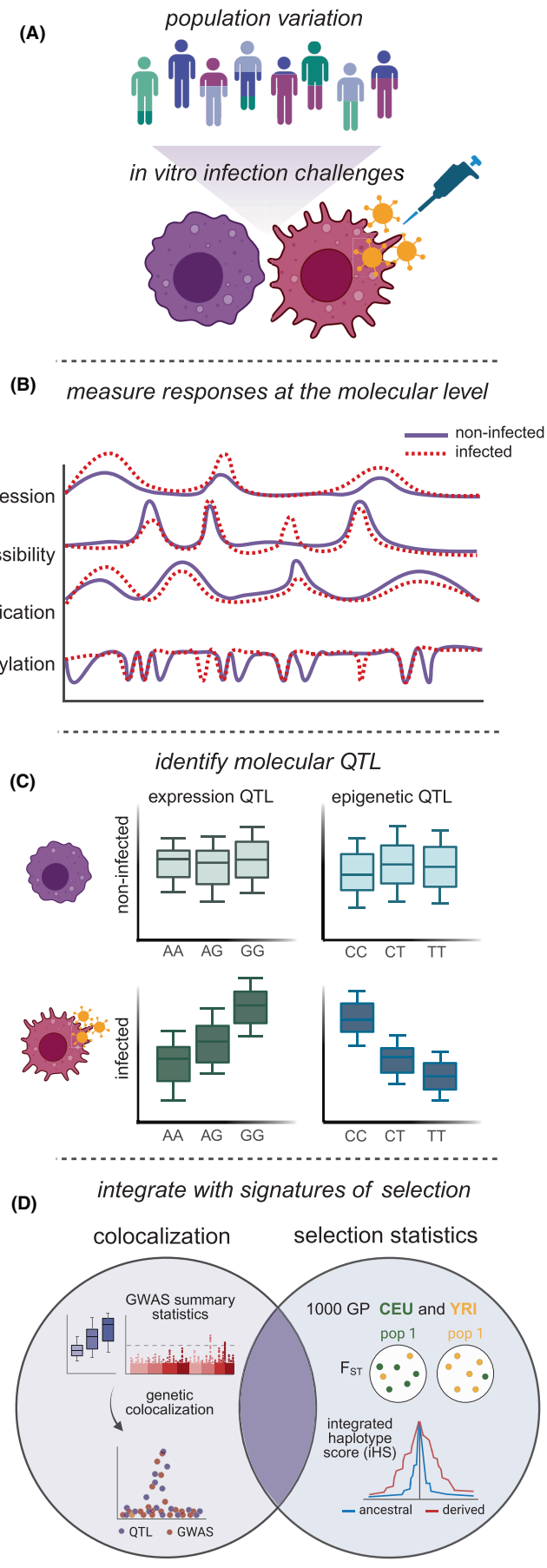

FIGURE 1 Overview of the workflow to identify signatures of selection at *cis*-regulatory regions of the genome linked with phenotypic immune response differences. (A) Samples (i.e., whole blood or PBMCs) from individuals with diverse genetic backgrounds are collected and challenged in vitro with pathogens or immune stimuli. (B) Various molecular traits, including gene expression levels and epigenetic marks (e.g., chromatin accessibility, histone tail modifications, DNA methylation levels) are measured in both non-infected and infected immune cells. (C) Gene expression and epigenetic quantitative trait loci (QTL) are mapped separately for each infection condition. At a QTL (e.g., expression QTL in the infected condition, dark green), individuals display variation in some quantitative trait, with each copy of an allele additively increasing the quantitative trait (here, the G allele increasing expression). Therefore, AA individuals display the lowest levels of the quantitative trait, while GG individuals display the highest. Response QTL occur when a QTL is exclusive to the infection condition and does not appear in the baseline, non-infected state. (D) Colocalization is performed between summary statistics derived from genome-wide association studies (GWAS) of immune-related diseases and results from QTL analyses to determine whether significant GWAS hits and QTL loci share genetic signals. Signatures of recent positive selection (F_{ST} and iHS) are then evaluated among loci that exhibit a genetic colocalization signal, revealing loci that likely were acted upon by evolution and that may contribute to mechanisms of disease.

heterologous immunity, and trained immunity.[27–29] While these mechanisms of immune memory directly influence the immune response itself, pathogens may also exert effects that impact responses indirectly through other means, such as altering cell type composition. For example, latent cytomegalovirus (CMV) infection has been shown to remodel the lymphoid compartment, accounting for up to 73% of population differences in lymphocyte cell type proportions between healthy donors originating from Central Africa and Western Europe.[22] In Europeans, positive CMV serostatus is associated with higher proportions of memory-like NK cells, which are characterized by an exhaustion phenotype, and $CD8^+$ effector memory T cells that re-express CD45RA, which harbor a cytotoxic and pro-inflammatory phenotype.[22] This suggests that differences in cell type composition due to CMV infection partially govern the response to infection.

Other environmental factors closely linked with societal inequalities rather than biological traits, such as socioeconomic status and differences in access to healthcare, also contribute to immune response heterogeneity across individuals and populations. For example, in the United States, racial and ethnic minorities are at much higher risk of significant morbidity and mortality due to IAV infection and COVID-19 compared to non-Hispanic white Americans.[30,31] Given the disproportionate access to healthcare and other health disparities in the United States, much of this imbalance can be attributed to health inequities caused by structural and social determinants. Although these biases likely lead to measurable differences between individuals that are not genetically controlled, it is difficult to tease apart their relative contributions to variation in immune response phenotypes because other nongenetic and genetic factors are often confounded.

Finally, studies of gene–environment interactions, which aim to define how genetic and environmental factors jointly affect response outcome or disease risk, are becoming more common as cohort sample sizes rise.[32] The importance of these non-additive effects in modifying the immune response cannot be discounted, although they are generally less well-characterized in the context of human health at present due to the difficulty of identifying such loci. Combining genomic data with self-reported ancestry labels and electronic health records is a valuable way to measure fine-scale population structure that, in some cases, is linked to shared culture and environment.[33] Using this approach, the prediction of complex disease risk can be improved within fine-scale groups.[33] Therefore, a better understanding of both genetics and environment is needed to push the field of genomic medicine forward, and defining the relative contribution of genetics versus environment to immune response variation is of great interest.

4 | MAPPING THE GENETIC BASIS OF IMMUNE RESPONSE VARIATION

Although a considerable amount of heterogeneity in the response to infection can be ascribed to environmental factors, genetic variation at loci throughout the genome also plays a substantial role in explaining population variation in immune responses.[16,17,23,34] Through the study of human genomes, the genetic underpinnings of complex immune-related diseases can be linked with molecular traits and clinically relevant variables to better define the genetic architecture of these phenotypes. Genome-wide association studies (GWAS) represent one approach to carry out trait mapping, with the principal goal of GWAS being to pinpoint regions of the genome that are associated with complex diseases and traits. These studies rely on sampling many (~thousands to millions) unrelated individuals to assess whether any shared common genetic variants are overrepresented among individuals with a particular disease or trait compared to healthy control individuals.

While GWAS have proven successful in identifying risk alleles for many complex diseases, including autoimmune, cardiovascular, metabolic, and neurodegenerative disorders, relatively few infectious disease GWAS have been performed in comparison.[35] GWAS for various viral (human immunodeficiency virus, IAV, hepatitis B/C virus, etc.) and bacterial (*Staphylococcus aureus*, *Mycobacterium tuberculosis*, etc.) infections exist, although these have attained only modest success, with few variants reaching genome-wide significance and minimal shared signals across studies considering the same pathogen.[35] More recently, the role of host genetics in SARS-CoV-2 infection and COVID-19 severity was investigated in a set of three of the largest infectious disease GWAS meta-analyses conducted to date, consisting of up to ~50,000 patients with COVID-19 across 46 studies globally.[36] Of greater success than previous infectious disease GWAS, these meta-analyses revealed 13 independent variants reaching genome-wide significance, most of which displayed relatively small effect sizes but were shared between two or more COVID-19 phenotypes.[36] These findings indicate that very large sample sizes are needed to reach adequate power to detect infectious disease trait associations genome-wide, which are difficult to obtain except in extraordinary circumstances. Of note, among the significant trait-associated loci identified by complex disease GWAS more generally, the vast majority are located in non-protein coding regions of the genome,[37] pointing towards gene regulatory variation as a crucial factor in modulating disease risk.

In parallel with association studies, quantitative trait loci (QTL) mapping studies have been used to map quantitative traits of interest (e.g., gene expression, protein expression, chromatin accessibility, etc.) to genomic regions. Expression QTL, or eQTL, studies map gene expression phenotypes to particular genomic loci by combining measures of gene expression with genome-wide genotyping data.[38–40] eQTL have been identified in an extensive variety of cell types and environmental contexts, and their study has significantly shaped our understanding of gene regulation and the genetic architecture of gene expression.[41–43] More specifically, eQTL mapping has proven to be uniquely powered to identify genetic factors that explain between-individual and between-population variation in the immune response to pathogens,[16,44–47] identifying thousands of variants associated with the expression levels of infected immune cells or the response to immune challenge.

Prior studies have demonstrated empirically that population differences in the gene expression response to infection are partially under genetic control.[16,17] Both *cis*- and *trans*-acting regulatory variants have been shown to markedly influence population-associated immune response variation. Considering genome-wide patterns, changes in allele frequencies of *cis*-eQTL across populations explain, on average, 30%–50% of ancestry-associated differences in immune responses to bacterial and viral infections in monocytes, macrophages, and other cell types in PBMCs.[16–18] In an in vitro infection model of PBMCs challenged with SARS-CoV-2, response eQTL (i.e., genetic variants affecting the magnitude of the gene expression response to infection) exhibited increased population differentiation specifically in individuals of East Asian descent, with ~5% of these variants displaying signals of local adaptation in East Asians.[22] Altogether, these data point to a considerable genetic component driving variation in immune responses among populations today.

Specific loci have also been implicated in the control of population-associated differences in the immune response. One common missense variant in *ERAP1*, rs27895, has been shown to increase IAV burden in lymphoblastoid cell lines in vitro, which was replicated in a human influenza challenge study in vivo, and is known to exhibit population differentiation globally.[19] The ancestral C allele of rs27895, which is associated with IAV resistance, is nearly fixed in East Asian populations, whereas the derived T allele is more common throughout the rest of the world (minor allele frequency, MAF = 6.4% in European populations and 23.8% in African populations in the 1000 Genomes Project[48]). Indeed, cell lines from populations with a higher frequency of the C allele are more protected from IAV in vitro, suggesting this variant has a functional consequence and may play a role in mediating population differences in susceptibility to viral infection.[19]

5 | POSITIVE SELECTION ON EXPRESSION QUANTITATIVE TRAIT LOCI

Positive selection, a form of natural selection in which advantageous genetic variants sweep to high frequency in a population, has substantially influenced the evolution of the human genome.[49,50] Genes that play a central role in innate immunity and immune defense pathways exhibit clear signatures of positive selection in present-day human populations.[51,52] Several studies have shown that regions targeted by positive selection are enriched for genes known to be involved in susceptibility to infectious diseases,[1,53] further indicating that genetic factors play a role in shaping the response to pathogens. RNA viruses, such as lentiviruses and orthomyxoviruses, have imposed some of the strongest evolutionary pressures on humans, with genomic footprints of these viruses present in modern genomes today.[54,55] Specifically, introgressed segments of the human genome derived from ancient hominid populations, like Neanderthals, are enriched for proteins known to interact with viruses. These regions likely represent adaptively introgressed segments that conferred a selective advantage when first introduced into modern human populations.[54,55]

Past selection imposed by pathogen exposure has been speculated to contribute to variation in the prevalence of infectious and autoimmune diseases across human populations[5,21]; however, it remains unclear the extent to which past positive selection on putatively causal *cis*-regulatory variants underlies variation in immune-related disease risk across populations. To address this question, we performed colocalization analysis using a publicly available single-cell RNA-sequencing dataset derived from healthy donor PBMCs challenged with IAV or a mock negative control.[18] Colocalized eQTL are expected to be strongly enriched for causal drivers of variation in disease susceptibility across individuals. We considered the union set of significant eQTL detected across all cell types ($CD4^+$ T cells, $CD8^+$ T cells, B cells, monocytes, and NK cells) and conditions in the Randolph et al. dataset (local false sign rate <0.10, defining a gene with an eQTL as an "eGene") and 14 publicly available GWAS summary statistics for 11 autoimmune and immune-related diseases (Table S1) as previously described.[56] Across all diseases, we colocalized eQTL in the Randolph et al. study[18] with a total of 95 GWAS variants (Figure 2A, Table S1).

To analyze a broader array of immune-related colocalization signals, we combined the Randolph et al. data with colocalization results for bulk eQTL in 18 immune cell types from three large immune eQTL studies (DICE [$n = 91$],[57] DGN [$n = 922$],[58] and BLUEPRINT [$n = 197$][59]) for the same 14 autoimmune and immune disease GWAS.[56] Using this approach, we identified 1030 colocalized GWAS hits across the 11 traits (mapping to 536 eGenes, Table S1). We subsequently investigated whether these potential causal variants showed signs of natural selection among Northern Europeans from Utah (CEU) and the Yoruba in Ibadan, Nigeria (YRI) based on data from the 1000 Genomes Project.[48] To do this, we utilized two methods: the integrated haplotype score (iHS[60]) and extreme values of population differentiation (F_{ST}). iHS represents a measure used within a population to identify recent positive selection, focusing on the length of haplotypes. On the other hand, F_{ST} is a between-population measure that assesses population differentiation through extreme value analysis. iHS scores were calculated using population-specific genetic maps with hapbin (v.1.3.0),[61] and F_{ST} statistics were calculated using an approach analogous to Weir and Cockerham's method.[62,63]

Far more colocalized loci display high |iHS| scores (values >95th percentile of the genome-wide distribution) in the CEU population than expected by chance ($p = 0.008$, Figure 2B), while no significant enrichment was detected in the YRI population. In our analysis, we only considered unlinked single nucleotide polymorphisms (SNPs) in an eGene ($r^2 < 0.8$ calculated with PLINK [v1.9][64]); for each set of SNPs with an $r^2 > 0.8$, we only kept the SNP with the highest |iHS| value. *p*-values were calculated using a permutation-based approach considering an empirically-derived null distribution that mimicked the MAF distribution and underlying linkage disequilibrium structure of the true data. Our results suggest that natural selection has acted on these *cis*-regulatory autoimmune and

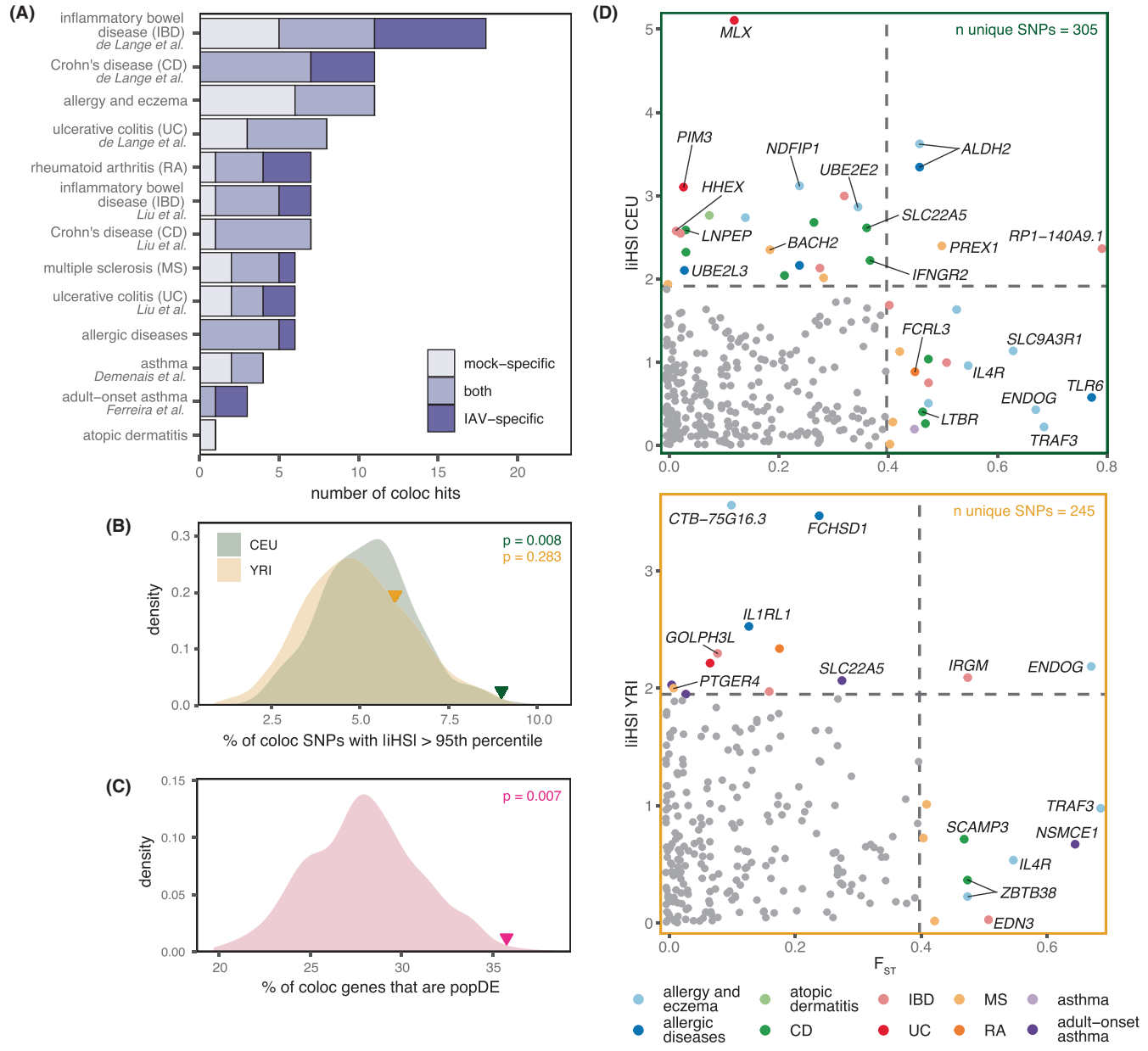

FIGURE 2 Recent positive selection has acted on *cis*-regulatory variants implicated in disease risk. (A) Number of shared and condition-specific colocalization hits identified across cell types (x-axis) in the 11 autoimmune traits tested (y-axis) in Randolph et al.[18] (B) Proportion of independent, colocalized lead genome-wide association studies (GWAS) loci that have |standardized iHS| values >95th percentile of the genome-wide distribution among SNPs with >5% MAF (CEU: green triangle, $p = 0.008$, YRI: yellow triangle, $p = 0.283$) compared to random expectation when sampling the same number of SNPs 1000 times from all variants with a MAF >5% in an LD-matched and MAF-matched manner (density distributions) among all autoimmune traits. (C) Proportion of genes with a colocalization signal that are differentially expressed between populations (pink triangle, $p = 0.007$) compared to random expectation when sampling the same total number of genes 1000 times from all genes tested (density distribution) among all autoimmune traits. (D) F_{ST} and |standardized iHS| values among the colocalized hits shown in (A) as well as those identified in the harmonized bulk eQTL data. F_{ST} values are plotted on the x-axis, while |iHS| values are plotted on the y-axis (top: CEU, bottom: YRI). Dotted lines show the 95th percentile of the genome-wide distribution for the respective selection statistic ($F_{ST} = 0.398$, |iHS| CEU = 1.92, |iHS| YRI = 1.95). eGenes with a selection statistic >95th percentile are represented by a colored point, and colors represent the autoimmune or immune disease-related trait for which a colocalization signal is detected (here, the multiple inflammatory bowel disease, ulcerative colitis, and Crohn's disease GWAS have been collapsed into a single label).

immune disease-related risk variants, particularly in Europeans, with the caveat that the vast majority of GWAS studies to date have focused exclusively on individuals of European ancestry.[65] This bias prevents us from detecting signatures of selection among GWAS loci unique to African-ancestry individuals. Moreover, we observed that colocalized genes are more likely to be differentially

expressed between populations than expected by chance (35.8% are classified as differentially expressed between populations; $p = 0.007$, p-value calculated using a permutation-based approach from an empirically-derived null distribution containing all genes tested in the eQTL analysis) (Figure 2C), pointing to a potential genetic contribution for the differences in the incidence of autoimmune and inflammatory disorders reported between African and European-ancestry individuals.[5]

Within the set of 536 colocalized eGene-SNP pairs, 48 eGenes carried a signature of recent positive selection in either the CEU or YRI populations (|iHS| or F_{ST} >95th percentile of the genome-wide distribution) (Figure 2D). Many of these genes involve crucial immune-related functions. For example, the CD-susceptibility risk variant rs2284553 colocalizes with *IFNGR2*, the gene encoding the beta chain of the IFN-γ receptor, in naïve CD8$^+$ T cells (Figure 3A). This variant is found at much higher frequency in the CEU population (MAF = 0.45) than the YRI population (MAF = 0.05) (Figure 3B) and shows a signature of recent positive selection in the CEU (iHS = 2.22). Another variant detected in the allergic disease GWAS, rs5743618, maps to a non-synonymous SNP located in *TLR1* that is also an eQTL for the nearby gene *TLR6* (Figure 3C). This variant compromises NF-κB signaling and activation to produce an attenuated inflammatory response[66] and is a known *trans*-regulatory hotspot.[17,23] Notably, it is found at low frequency in the YRI population (derived allele frequency [DAF] = 0.04) but at an elevated frequency in the CEU population (DAF = 0.67, Figure 3D).[21] This difference in allele frequency alone explains the positive correlation between African genetic ancestry and the transcriptional response to immune stimulation with antigens that signal through TLR1.[16,17,21]

These results provide evidence that recent, local positive selection has acted on putatively causal regulatory risk variants associated with common immune-related diseases in GWAS, strengthening the link between pathogen-mediated selection and susceptibility to autoimmune disorders.[5,67,68] The connection between infectious diseases and chronic inflammatory disorders is further supported by reports spanning the last two decades that some pathogens contribute to, and possibly cause, the development of certain chronic inflammatory and autoimmune diseases (e.g., Epstein–Barr virus and systemic lupus erythematosus, rheumatoid arthritis, and multiple sclerosis; *Mycobacterium avium* and CD; *Yersinia enterocolica* and inflammatory bowel disease[69–74]). More recently, evidence has emerged suggesting that one variant in *ERAP2* (rs2549794) is protective against infection with *Yersinia pestis*, as it leads to increased control of intracellular *Y. pestis* replication in macrophages ex vivo and shows a signature of recent positive selection.[75] While seemingly protective against *Y. pestis*, this *ERAP2* variant is also a known risk factor for CD and type I diabetes, suggesting that the selective advantage conferred by this allele is likely context-specific.[75,76] In the presence of *Y. pestis* historically, rs2549794 might have provided an evolutionary advantage among human populations; however, this potentially came

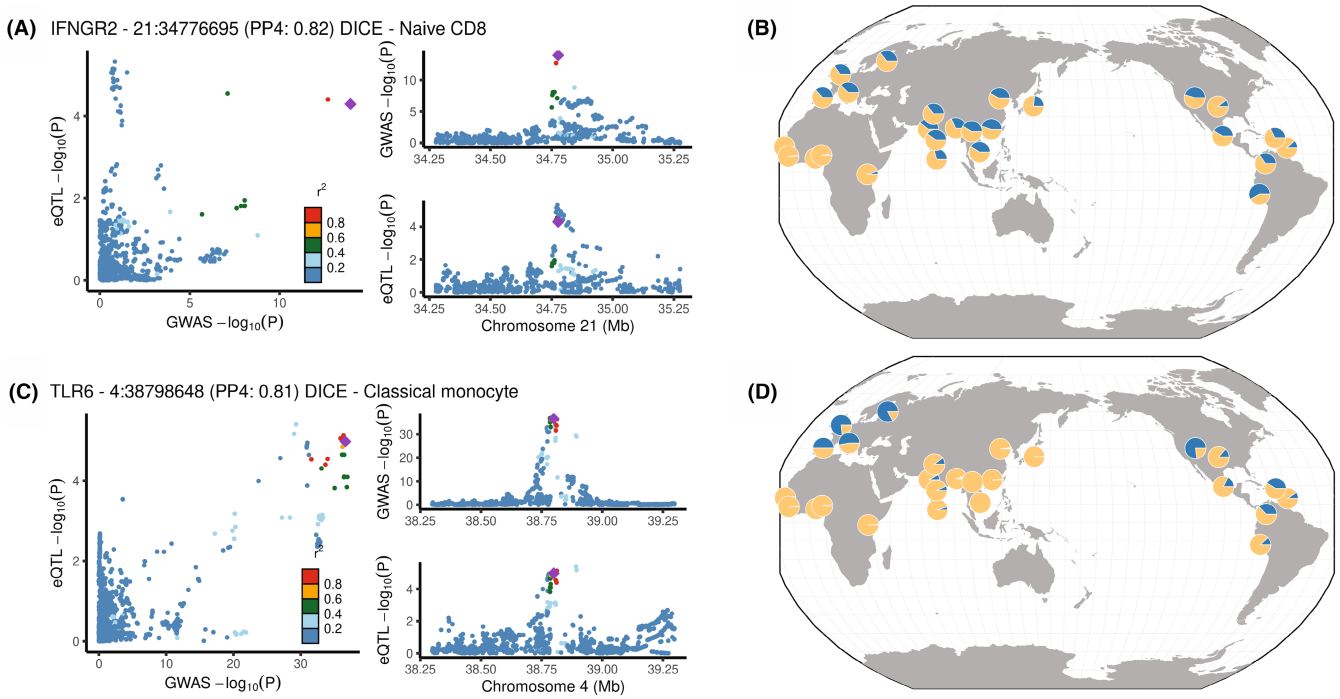

FIGURE 3 Colocalization signals in relevant immune genes. (A) *IFNGR2* colocalizes with rs2284553 in naïve CD8$^+$ T cells in the Crohn's disease Genome-wide association studies (GWAS).[113] (B) Global distribution of the alleles at rs2284553 (A: blue, G: yellow). (C) *TLR6* colocalizes with rs5743618 in classical CD14$^+$ monocytes in the allergic disease GWAS. (D) Global distribution of the alleles at rs5743618 (C: blue, A: yellow). For both (A) and (C), the larger plot on the left shows the correlation between GWAS p-values (x-axis) and eQTL p-values (y-axis). Smaller plots on the right show the Manhattan plots for the GWAS signal (top) and the eQTL signal (bottom). Plots in (B) and (D) were generated with the Geography of Genetic Variants Browser.[114]

at the cost of an overactive immune system in the absence of *Y. pestis*, contributing to disease risk in the modern day. More generally, the study of ancient DNA has revealed that the frequency of risk alleles for inflammatory disorders increased in post-Neolithic Europeans, possibly because of antagonistic pleiotropy following genetic adaptation to pathogens.[77,78] Collectively, these findings shed light on human evolutionary history, suggesting that at least some present-day autoimmune risk loci may have been adaptive and conferred a functional benefit in the past.

6 | POPULATION VARIATION IN THE EPIGENOME

Epigenetic mechanisms have been shown to play a central role in the regulation of immune responses to bacterial and viral pathogens.[79–89] Thus, it is likely that variation in epigenetic profiles across individuals and populations considerably contributes to population variation in innate immune responses and susceptibility to disease. The field of population epigenomics has primarily concentrated on the differences in DNA methylation levels among individuals. Recently, however, the scope has expanded to include a wider range of epigenetic modifications and their complex interactions with gene regulation.[89–92] For example, Aracena et al. performed an in-depth genetic, epigenetic, and transcriptional profiling of primary macrophages derived from European- and African-ancestry individuals before and after IAV infection.[89] These data revealed that baseline epigenetic profiles are strongly predictive of the transcriptional response to IAV across individuals, supporting the concept that changes in gene expression in response to IAV infection are, in part, influenced by the pre-existing epigenetic states of an individual's cells.

In addition, Aracena et al. have shown that ancestry-associated differences in the immune response to IAV infection are tightly linked to changes in enhancer activity, as measured by the level of H3K4 monomethylation (H3K4me1) on histones (Figure 4A).[89] In contrast, genetic ancestry has a limited contribution to population variation in promoter-associated H3K4me3 histone methylation (explaining only 2% of the total variance) and CpG methylation levels (explaining less than 1% of the total variance) (Figure 4A).[89] These data suggest that variation in enhancer activity is the primary epigenetic mechanism underlying differences in immune regulation among populations.

Despite meaning "above the genes", epigenetic variation across individuals is influenced by genetic variation.[59,89–95] In macrophages, for example, ~52% of population variation in chromatin accessibility is explained by differences in allele frequencies of genetic variants associated with chromatin accessibility levels.[89] This fraction rises to up to 65% for DNA methylation differences,[89,92] revealing that genetic ancestry-associated diversity in the epigenetic landscape is strongly genetically controlled. Comparatively, 13%–54% of population variation in gene expression has been explained by genetic factors,[16–18,89] underscoring the fact that, for some cell types, epigenetic variation is even more dependent on underlying genetics than gene expression variation.

7 | EPIGENETIC QTL AND SUSCEPTIBILITY TO IMMUNE DISORDERS

Genetic variants that impact the epigenome of immune cells may provide additional insight into the biological relevance of GWAS variants associated with immune disorders. Supporting this view, Aracena et al. report that epigenetic QTL greatly increase—by approximately 10-fold—the number of variants that colocalize between GWAS variants and regulatory variants using the same 14 autoimmune and immune disease-related GWAS considered above (Table S1).[89] Of all colocalized variants, 93% were captured only when considering epigenetic QTL, revealing that the interplay between genetic and epigenetic variation can provide valuable insight into how GWAS loci act within gene regulatory networks. Evidence of selection on epigenetic variants may further prioritize candidates for disease treatment and therapeutic development. Interestingly, using the data first presented in Aracena et al., we found that significantly more colocalized SNPs showed a signature of natural selection (17%, $p < 2.2e\text{-}16$) than expected by chance, considering either high F_{ST} values between the CEU and the YRI populations or high |iHS| scores in the CEU population (values >95th percentile of the genome-wide distribution) (Figure 4B). Of these, all are epigenetic QTL, suggesting that genetic variants controlling epigenetic modifications have an impact on organismal fitness and can capture evidence of selection on gene regulation.

8 | DISCUSSION

Here, we argue that natural selection has acted on disease-relevant variants and that differences in the frequencies of these variants between populations likely contribute in part to the observed population differences in immune-related phenotypes seen among individuals today. Interrogating positive selection signatures among loci that colocalize between GWAS variants and variants associated with a quantitative trait, whether that trait be gene expression or an epigenetic modification, can point to variants that likely play a functional role in disease. However, our understanding of population variation in immune responses and natural selection's role in shaping this variation is still in its infancy. Most studies assessing population-level immune response heterogeneity have ignored the dynamic nature of the immune response, only sampling quantitative traits at a single time point post-infection, and have greatly restricted the genetic diversity that is surveyed, only considering two or three populations at most. Further, almost all studies have failed to consider natural selection in a polygenic context, instead choosing to focus on the detection of single outlier loci when conducting genomic scans for selection. Going forward, it will be crucial to prioritize immunogenomics studies that (1) cover a greater range of populations globally, considering both genetic and environmental diversity, (2) evaluate population variation in the response to immune challenges in a dynamic manner, and (3) develop tools that allow testing for polygenic selection.

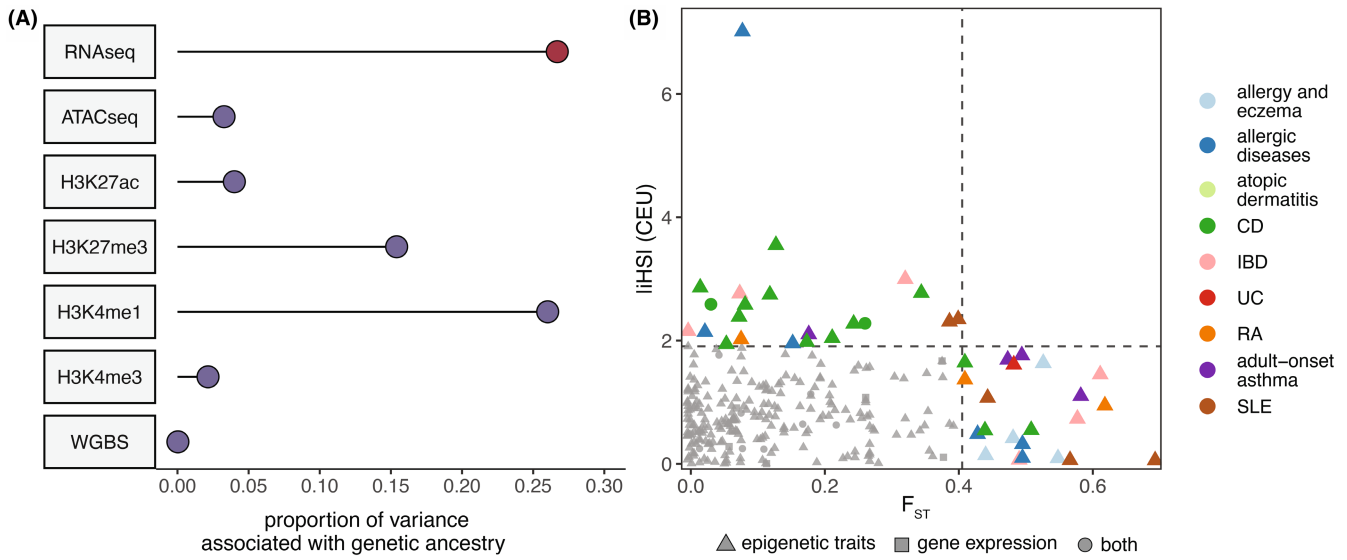

FIGURE 4 Population variation in epigenetic mechanisms. (A) Proportion of variance associated with genetic ancestry across molecular traits (ATAC: assay for transposase-accessible chromatin; H3K27: histone 3 lysine 27; H3K4: histone 3 lysine 4; ac: acetylation; me1: monomethylation; me3: trimethylation; WGBS: whole genome bisulfite sequencing [DNA/CpG methylation]). The mean across the non-infected and IAV-infected conditions is plotted.[89] (B) F_{ST} and |standardized iHS| values among the hits that colocalize between various molecular traits (epigenetic and/or gene expression, designated by the point shape). F_{ST} values are plotted on the x-axis, while |iHS| values (CEU) are plotted on the y-axis. Dotted lines show the 95th percentile of the genome-wide distribution for the respective selection statistic (F_{ST} = 0.398, |iHS| CEU = 1.92). eGenes with a selection statistic >95th percentile are represented by a colored point, and colors represent the autoimmune and immune disease-related trait for which a colocalization signal is detected (here, the multiple inflammatory bowel disease, ulcerative colitis, and Crohn's disease genome-wide association studies have been collapsed into a single label).

8.1 | Expanding the global diversity of immunogenomics studies

One of the most critical gaps in biomedical research today is the lack of non-European ancestry individuals in genomics studies, specifically among cohorts designed to characterize immune variation among healthy individuals in the general population. Most genomics studies to date—approximately 86%—have been conducted solely in individuals of European descent.[96] Increasing representation of global populations in genomics studies that address immune response variation will bolster our awareness of the functional heterogeneity that exists today. Specifically, extending population studies of immune responses to a larger array of genetic backgrounds may reveal differences in the frequencies of disease-relevant alleles and unique population-associated variation in immune gene regulation. Investing in the sampling of diverse cohorts will also likely advance our understanding of disease etiology, with the ultimate goal of making genomics research more broadly applicable to ensure an equitable distribution of the benefits promised by personalized medicine. Of note, precautions must be taken to safeguard individuals within historically underrepresented populations in scientific research from exploitation. Research projects that: (1) clearly define how donor samples will be collected, used, and/or banked through proper informed consent, (2) engage with communities through the sharing of research results and data, and (3) involve local scientists and institutes in the research efforts must be the standard to increase inclusivity and protect against the misuse of data and samples.

8.2 | Mapping dynamic immune response variation

Much of the immune response QTL literature focuses on the early response of innate immune cells (~hours) to experimental challenges. These studies are limited in their conclusions considering variation in the adaptive immune response. Experiments tailored to measure population-associated variation in T and B cell responses are needed to supplement these findings. Specifically, studies that seek to characterize transcriptional variation as well as T-cell and B-cell receptor repertoire diversity across individuals and populations will allow us to discern the extent of adaptive response heterogeneity. Further, studies that exploit the interconnected nature of the immune system will allow us to more comprehensively assess how variation in innate immunity impacts adaptive immunity, how variation in cell–cell interactions and paracrine signaling dictate response differences, and how both arms of the immune system work together to give rise to an overall phenotype.

A natural extension of surveying variation in the immune response at longer time scales involves investigating the dynamics of gene regulation throughout an immune response, or measuring how gene regulatory effects change over time following infection. Most immune response QTL studies to date have failed to account for the dynamic nature of the immune response, as they have only probed gene regulatory patterns at a single time point. To study the dynamics of gene regulation, experiments in which time series gene expression data are generated at detailed temporal resolution are needed. The objective of these studies would be to map

QTL that show an interaction with time (e.g., genetic effects that only appear early or late in the response) or identify QTL associated with some component of the immune response trajectory (e.g., overall magnitude of the response or the global maximum immune response).

Dense time course experiments are difficult to perform with primary human cells and tissues due to irregularities in patient sampling and availability, potentially leading to the introduction of major batch effects. Historically, cell lines have been used to mitigate these issues; however, cell lines do not always faithfully recapitulate features of their primary cell counterparts, especially considering karyotype and cell marker expression,[97] and they cannot match the genetic diversity represented across individuals because they often stem from a single donor. More recently, researchers have turned to induced pluripotent stem cells, or iPSCs, as they represent an alternative source of biological material that overcomes these problems. iPSCs are cells that have been reprogrammed from adult somatic cells (e.g., fibroblasts) into an induced state of pluripotency and self-renewal.[98,99] These cells are advantageous because they can self-renew indefinitely, can feasibly be generated from any individual, and, theoretically, have the capacity to undergo directed differentiation into any cell type present in the three primary germ layers of the human body.

Because of these properties, iPSCs are an attractive model to study tissue types that are difficult to obtain as primary samples, such as tissue-derived immune cells. Further, they allow for the design of more complex studies (e.g., a multiple time point time course in which hundreds of thousands of cells would be needed for each time point) because cell cultures can be scaled up,[100] which is difficult or impossible to do with primary immune cells. They are also an excellent model to study the genetic basis of complex human traits as they preserve phenotypic differences between individuals. Indeed, 5%–46% of the variation in iPSC phenotypes arises from differences between individuals, and many of these phenotypes can be mapped to specific loci, which demonstrates their utility as a powerful model for immune response eQTL studies.[101,102] Notably, immune cells derived from iPSCs are a stable, renewable source. Because of this, large-scale iPSC banks could be established to study immune responses and their genetic determinants reliably and reproducibly. Therefore, studies that seek to examine immune response dynamics across individuals would benefit from relying on iPSCs as a resource.

8.3 | Exploring polygenic selection

Traditional tests of positive selection, such as iHS and F_{ST},[60,103] rely on outlier approaches to pinpoint positively selected regions in the genome that recently swept to high frequency. GWAS performed in human populations have revealed that a large number of traits and common complex diseases are polygenic in nature, with many loci contributing to the overall trait or disease phenotype.[104] It is therefore likely that standing genetic variation directly contributing to these polygenic phenotypes was selected upon as a unit, in turn driving subtle shifts in allele frequencies across many loci.[105,106] Positive selection scans based on detecting selective sweeps at individual loci fail to capture the small changes expected to occur at multiple loci characteristic of polygenic selection, so alternative approaches must be used. Current tests for polygenic selection are subject to confounding factors (e.g., population structure and/or environmental effects in GWAS summary statistics), suffer from effect size misestimation,[106] or rely on a priori information about gene sets involved in particular biological processes.[18,107] Due to these drawbacks, examples of clear polygenic adaptation throughout human evolution are rare. Despite this, a handful of previous studies have described putative instances of polygenic selection using predetermined gene ontology sets to jointly analyze loci in particular pathways,[18,107] demonstrating the utility of tests that integrate empirical data in the search for polygenic adaptation signatures.

The advent of CRISPR/Cas9 as a genome editing tool, along with its many variants that enable the precise editing of regulatory and epigenomic landscapes in a high-throughput manner, has revolutionized the genomics field. CRISPR screens have been widely used to introduce genetically encoded perturbations in pools of target cells, which are then challenged with an experimental pressure (e.g., drug treatment), to reveal mutations that confer resistance or susceptibility to the challenge. In the traditional sense, pooled CRISPR screens in CRISPR-edited iPSC-derived immune cells followed by pathogen challenge unlocks the possibility of probing coding and regulatory elements in disease-relevant cell types to identify and causally test whether perturbation results in significant changes to expression pattens or relevant downstream phenotypes, such as sensitivity to infection.[108,109]

Pooled CRISPR screens can also be used to investigate and empirically define *trans*-regulatory networks,[108] which in turn makes it possible to explore and quantify signals of polygenic selection without the need to rely on known gene sets or possibly confounded GWAS statistics. Historically, identifying *trans*-eQTL has proven to be difficult as very large sample sizes are required to map these loci using traditional methods, partly due to their much smaller effect sizes compared to *cis*-eQTL.[110] Most QTL studies have largely disregarded the existence of any *trans* effects; thus, these regulatory networks are poorly defined. Yet, it is well known that many human traits exhibit polygenic genetic architecture.[111] More recently, an extension of this known polygenic framework was proposed in the "omnigenic model", which posits that gene regulatory networks are highly interconnected and that most genes expressed in disease-relevant cells affect disease risk via network effects.[112] Both polygenic and omnigenic genetic architectures support the idea that *trans* effects are pervasive and, although individually small, are cumulatively relevant and necessary to wholly understand disease risk.

Considering polygenic and omnigenic traits, an alternative method to detect *trans*-regulated genes involves the use of CRISPR editing to experimentally edit genes of interest and subsequently measure changes in expression of other genes. Through genome-wide CRISPR perturbations, both upstream and downstream

regulators of disease genes can be defined in a network discovery step in which *cis*- and *trans*-regulatory relationships are uncovered in a targeted way. Once a *trans* network is defined empirically, the overrepresentation of trait-increasing or trait-decreasing alleles can be tested among individuals with high or low trait values, respectively, within genes in that network. Traits under polygenic selection are expected to show unidirectional allelic bias that is consistent with the trait of interest, e.g., if significantly more trait-increasing alleles are found in populations with high trait values than expected by chance, then that trait may exhibit a signature of polygenic adaptation. In theory, this approach provides a platform to test for polygenic selection among many different molecular quantitative traits using a less biased approach than traditional methods. These studies will extend our understanding of the mechanisms by which natural selection has played and continues to play a role in the diversification of immune responses among human populations.

ACKNOWLEDGMENTS

We thank all members from the Barreiro laboratory for their comments on the paper. Computational resources were provided by the University of Chicago Research Computing Center. This work was supported by National Institutes of Health Research grants R01-GM134376 and P30-DK042086 to L.B.B. H.E.R. was supported by a National Institutes of Health Virus-Host Interactions Public Health Service Institutional Research Training Award AI07647. K.A.A. was supported by a grant from the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program. Figure 1 was created with BioRender.

CONFLICT OF INTEREST STATEMENT

Authors have no competing interests to declare.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Haley E. Randolph https://orcid.org/0000-0003-2044-1180
Katherine A. Aracena https://orcid.org/0000-0002-8830-904X
Yen-Lung Lin https://orcid.org/0000-0003-1596-944X
Zepeng Mu https://orcid.org/0000-0002-7717-3247
Luis B. Barreiro https://orcid.org/0000-0001-9456-367X

REFERENCES

1. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet*. 2014;15(6):379-393. doi:10.1038/nrg3734
2. Stephens PR, Sonia A, Smith KF, et al. The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts. *Ecol Lett*. 2016;19(9):1159-1171. doi:10.1111/ele.12644
3. Duffy D, Rouilly V, Libri V, et al. Functional analysis via standardized whole-blood stimulation systems defines the boundaries of a healthy immune response to complex stimuli. *Immunity*. 2014;40(3):436-450. doi:10.1016/j.immuni.2014.03.002
4. Pennington R, Gatenbee C, Kennedy B, Harpending H, Cochran G. Group differences in proneness to inflammation. *Infect Genet Evol*. 2009;9(6):1371-1380. doi:10.1016/j.meegid.2009.09.017
5. Brinkworth JF, Barreiro LB. The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease. *Curr Opin Immunol*. 2014;31:66-78. doi:10.1016/j.coi.2014.09.008
6. Betteridge JD, Armbruster SP, Maydonovitch C, Veerappan GR. Inflammatory bowel disease prevalence by age, gender, race, and geographic location in the U.S. military health care population. *Inflamm Bowel Dis*. 2013;19(7):1421-1427. doi:10.1097/MIB.0b013e318281334d
7. Gelfand JM, Stern RS, Nijsten T, et al. The prevalence of psoriasis in African Americans: results from a population-based study. *J Am Acad Dermatol*. 2005;52(1):23-26. doi:10.1016/j.jaad.2004.07.045
8. Bell RA, Mayer-Davis EJ, Beyer JW, et al. Diabetes in non-Hispanic white youth: prevalence, incidence, and clinical characteristics: the SEARCH for Diabetes in Youth Study. *Diabetes Care*. 2009;32(Suppl 2):S102-S111. doi:10.2337/dc09-S202
9. Ogunbi SO, Ransom JA, Sullivan K, Schoen BT, Gold BD. Inflammatory bowel disease in African-American children living in Georgia. *J Pediatr*. 1998;133(1):103-107. doi:10.1016/s0022-3476(98)70187-8
10. Feldman CH, Hiraki LT, Liu J, et al. Epidemiology and sociodemographics of systemic lupus erythematosus and lupus nephritis among US adults with Medicaid coverage, 2000-2004. *Arthritis Rheum*. 2013;65(3):753-763. doi:10.1002/art.37795
11. Lim SS, Bayakly AR, Helmick CG, Gordon C, Easley KA, Drenkard C. The incidence and prevalence of systemic lupus erythematosus, 2002-2004: The Georgia Lupus Registry. *Arthritis Rheumatol*. 2014;66(2):357-368. doi:10.1002/art.38239
12. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat Rev Genet*. 2010;11(5):356-366. doi:10.1038/nrg2760
13. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017;547(7662):173-178. doi:10.1038/nature22969
14. Liu JZ, van Sommeren S, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979-986. doi:10.1038/ng.3359
15. Liu Z, Liu R, Gao H, et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet*. 2023;55(5):796-806. doi:10.1038/s41588-023-01384-0
16. Nédélec Y, Sanz J, Baharian G, et al. Genetic ancestry and natural selection drive population differences in immune responses to pathogens. *Cell*. 2016;167(3):657-669.e21. doi:10.1016/j.cell.2016.09.025
17. Quach H, Rotival M, Pothlichet J, et al. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell*. 2016;167(3):643-656.e17. doi:10.1016/j.cell.2016.09.024
18. Randolph HE, Fiege JK, Thielen BK, et al. Genetic ancestry effects on the response to viral infection are pervasive but cell type specific. *Science*. 2021;374(6571):1127-1133. doi:10.1126/science.abg0928
19. Schott BH, Wang L, Zhu X, et al. Single-cell genome-wide association reveals that a nonsynonymous variant in ERAP1 confers increased susceptibility to influenza virus. *Cell Genom*. 2022;2(11):100207. doi:10.1016/j.xgen.2022.100207
20. O'Neill MB, Quach H, Pothlichet J, et al. Single-cell and bulk RNA-sequencing reveal differences in monocyte susceptibility to influenza A virus infection between Africans and Europeans. *Front*

Immunol. 2021;12:12. Accessed March 8, 2023. doi:10.3389/fimmu.2021.768189
21. Sanz J, Randolph HE, Barreiro LB. Genetic and evolutionary determinants of human population variation in immune responses. *Curr Opin Genet Dev*. 2018;53:28-35. doi:10.1016/j.gde.2018.06.009
22. Aquino Y, Bisiaux A, Li Z, et al. Dissecting human population variation in single-cell responses to SARS-CoV-2. *Nature*. 2023;9:1-9. doi:10.1038/s41586-023-06422-9
23. Piasecka B, Duffy D, Urrutia A, et al. Distinctive roles of age, sex, and genetics in shaping transcriptional variation of human immune responses to microbial challenges. *PNAS*. 2018;115(3):E488-E497. doi:10.1073/pnas.1714765115
24. Brodin P, Jojic V, Gao T, et al. Variation in the human immune system is largely driven by non-heritable influences. *Cell*. 2015;160(1-2):37-47. doi:10.1016/j.cell.2014.12.020
25. Montecino-Rodriguez E, Berent-Maoz B, Dorshkind K. Causes, consequences, and reversal of immune system aging. *J Clin Invest*. 2013;123(3):958-965. doi:10.1172/JCI64096
26. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-638. doi:10.1038/nri.2016.90
27. Cooper MD, Alder MN. The evolution of adaptive immune systems. *Cell*. 2006;124(4):815-822. doi:10.1016/j.cell.2006.02.001
28. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. *Nat Immunol*. 2015;16(4):343-353. doi:10.1038/ni.3123
29. Netea MG, Domínguez-Andrés J, Barreiro LB, et al. Defining trained immunity and its role in health and disease. *Nat Rev Immunol*. 2020;20(6):375-388. doi:10.1038/s41577-020-0285-6
30. Ko JY, Danielson ML, Town M, et al. Risk factors for coronavirus disease 2019 (COVID-19)–associated hospitalization: COVID-19—Associated Hospitalization Surveillance Network and Behavioral Risk Factor Surveillance System. *Clin Infect Dis*. 2021;72(11):e695-e703. doi:10.1093/cid/ciaa1419
31. Chandrasekhar R, Sloan C, Mitchel E, et al. Social determinants of influenza hospitalization in the United States. *Influenza Other Respir Viruses*. 2017;11(6):479-488. doi:10.1111/irv.12483
32. Virolainen SJ, VonHandorf A, Viel KCMF, Weirauch MT, Kottyan LC. Gene–environment interactions and their impact on human health. *Genes Immun*. 2022;30:1-11. doi:10.1038/s41435-022-00192-6
33. Belbin GM, Cullina S, Wenric S, et al. Toward a fine-scale population health monitoring system. *Cell*. 2021;184(8):2068-2083.e11. doi:10.1016/j.cell.2021.03.034
34. Bakker OB, Aguirre-Gamboa R, Sanna S, et al. Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses. *Nat Immunol*. 2018;19(7):776-786. doi:10.1038/s41590-018-0121-3
35. Mozzi A, Pontremoli C, Sironi M. Genetic susceptibility to infectious diseases: current status and future perspectives from genome-wide approaches. *Infect Genet Evol*. 2018;66:286-307. doi:10.1016/j.meegid.2017.09.028
36. Niemi MEK, Karjalainen J, Liao RG, et al. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600(7889):472-477. doi:10.1038/s41586-021-03767-x
37. Edwards SL, Beesley J, French JD, Dunning AM. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet*. 2013;93(5):779-797. doi:10.1016/j.ajhg.2013.10.012
38. Cheung VG, Spielman RS, Ewens KG, Weber TM, Morley M, Burdick JT. Mapping determinants of human gene expression by regional and genome-wide association. *Nature*. 2005;437(7063):1365-1369. doi:10.1038/nature04244
39. Gibson G, Weir B. The quantitative genetics of transcription. *Trends Genet*. 2005;21(11):616-623. doi:10.1016/j.tig.2005.08.010
40. Morley M, Molony CM, Weber TM, et al. Genetic analysis of genome-wide variation in human gene expression. *Nature*. 2004;430(7001):743-747. doi:10.1038/nature02797
41. Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010;464(7289):768-772. doi:10.1038/nature08872
42. Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506-511. doi:10.1038/nature12531
43. Aguet F, Brown AA, Castel SE, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213. doi:10.1038/nature24277
44. Barreiro LB, Tailleux L, Pai AA, Gicquel B, Marioni JC, Gilad Y. Deciphering the genetic architecture of variation in the immune response to *Mycobacterium tuberculosis* infection. *PNAS*. 2012;109(4):1204-1209. doi:10.1073/pnas.1115761109
45. Lee MN, Ye C, Villani AC, et al. Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science*. 2014;343(6175):1246980. doi:10.1126/science.1246980
46. Manry J, Nédélec Y, Fava VM, et al. Deciphering the genetic control of gene expression following *Mycobacterium leprae* antigen stimulation. *PLoS Genet*. 2017;13(8):e1006952. doi:10.1371/journal.pgen.1006952
47. Kim-Hellmuth S, Bechheim M, Pütz B, et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun*. 2017;8(1):266. doi:10.1038/s41467-017-00366-1
48. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
49. Booker TR, Jackson BC, Keightley PD. Detecting positive selection in the genome. *BMC Biol*. 2017;15(1):98. doi:10.1186/s12915-017-0434-y
50. Sabeti PC, Varilly P, Fry B, et al. Genome-wide detection and characterization of positive selection in human populations. *Nature*. 2007;449(7164):913-918. doi:10.1038/nature06250
51. Nielsen R, Bustamante C, Clark AG, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*. 2005;3(6):e170. doi:10.1371/journal.pbio.0030170
52. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet*. 2010;11(1):17-30. doi:10.1038/nrg2698
53. Fumagalli M, Sironi M. Human genome variability, natural selection and infectious diseases. *Curr Opin Immunol*. 2014;30:9-16. doi:10.1016/j.coi.2014.05.001
54. Enard D, Petrov DA. Evidence that RNA viruses drove adaptive introgression between Neanderthals and modern humans. *Cell*. 2018;175(2):360-371.e13. doi:10.1016/j.cell.2018.08.034
55. Enard D, Petrov DA. Ancient RNA virus epidemics through the lens of recent adaptation in human genomes. *Philos Trans R Soc Lond Ser B Biol Sci*. 1812;2020(375):20190575. doi:10.1098/rstb.2019.0575
56. Mu Z, Wei W, Fair B, Miao J, Zhu P, Li YI. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol*. 2021;22(1):122. doi:10.1186/s13059-021-02334-x
57. Schmiedel BJ, Singh D, Madrigal A, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell*. 2018;175(6):1701-1715.e16. doi:10.1016/j.cell.2018.10.022
58. Battle A, Mostafavi S, Zhu X, et al. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res*. 2014;24(1):14-24. doi:10.1101/gr.155192.113
59. Chen L, Ge B, Casale FP, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*. 2016;167(5):1398-1414.e24. doi:10.1016/j.cell.2016.10.026
60. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4(3):e72. doi:10.1371/journal.pbio.0040072
61. Maclean CA, Chue Hong NP, Prendergast JGD. Hapbin: An efficient program for performing haplotype-based scans for

positive selection in large genomic datasets. *Mol Biol Evol*. 2015;32(11):3027-3029. doi:10.1093/molbev/msv172
62. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156-2158. doi:10.1093/bioinformatics/btr330
63. Weir BS, Cockerham CC. Estimating *F*-statistics for the analysis of population structure. *Evolution*. 1984;38(6):1358-1370. doi:10.2307/2408641
64. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7. doi:10.1186/s13742-015-0047-8
65. Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. 2011;475(7355):163-165. doi:10.1038/475163a
66. Barreiro LB, Ben-Ali M, Quach H, et al. Evolutionary dynamics of human toll-like receptors and their different contributions to host defense. *PLoS Genet*. 2009;5(7):e1000562. doi:10.1371/journal.pgen.1000562
67. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the peopling of the world through genomics. *Nature*. 2017;541(7637):302-310. doi:10.1038/nature21347
68. Quach H, Quintana-Murci L. Living in an adaptive world: genomic dissection of the genus homo and its immune response. *J Exp Med*. 2017;214(4):877-894. doi:10.1084/jem.20161942
69. Abubakar I, Myhill D, Aliyu SH, Hunter PR. Detection of Mycobacterium avium subspecies paratuberculosis from patients with Crohn's disease using nucleic acid-based techniques: a systematic review and meta-analysis. *Inflamm Bowel Dis*. 2008;14(3):401-410. doi:10.1002/ibd.20276
70. Feller M, Huwiler K, Stephan R, et al. Mycobacterium avium subspecies paratuberculosis and Crohn's disease: a systematic review and meta-analysis. *Lancet Infect Dis*. 2007;7(9):607-613. doi:10.1016/S1473-3099(07)70211-6
71. Ramos-Casals M, Jara LJ, Medina F, et al. Systemic autoimmune diseases co-existing with chronic hepatitis C virus infection (the HISPAMEC registry): patterns of clinical and immunological expression in 180 cases. *J Intern Med*. 2005;257(6):549-557. doi:10.1111/j.1365-2796.2005.01490.x
72. James JA, Neas BR, Moser KL, et al. Systemic lupus erythematosus in adults is associated with previous Epstein–Barr virus exposure. *Arthritis Rheum*. 2001;44(5):1122-1126.
73. Yamazaki M, Kitamura R, Kusano S, et al. Elevated immunoglobulin G antibodies to the proline-rich amino-terminal region of Epstein-Barr virus nuclear antigen-2 in sera from patients with systemic connective tissue diseases and from a subgroup of Sjögren's syndrome patients with pulmonary involvements. *Clin Exp Immunol*. 2005;139(3):558-568. doi:10.1111/j.1365-2249.2004.02704.x
74. Saebo A, Vik E, Lange OJ, Matuszkiewicz L. Inflammatory bowel disease associated with yersinia enterocolitica O:3 infection. *Eur J Intern Med*. 2005;16(3):176-182. doi:10.1016/j.ejim.2004.11.008
75. Klunk J, Vilgalys TP, Demeure CE, et al. Evolution of immune genes is associated with the black death. *Nature*. 2022;611(7935):312-319. doi:10.1038/s41586-022-05349-x
76. Hamilton F, Mentzer AJ, Parks T, et al. Variation in ERAP2 has opposing effects on severe respiratory infection and autoimmune disease. *Am J Hum Genet*. 2023;110(4):691-702. doi:10.1016/j.ajhg.2023.02.008
77. Kerner G, Neehus AL, Philippot Q, et al. Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell Genom*. 2023;3(2):100248. doi:10.1016/j.xgen.2022.100248
78. Barrie W, Yang Y, Irving-Pease EK, et al. Elevated genetic risk for multiple sclerosis emerged in steppe pastoralist populations. *Nature*. 2024;625(7994):321-328. doi:10.1038/s41586-023-06618-z
79. Pacis A, Mailhot-Léonard F, Tailleux L, et al. Gene activation precedes DNA demethylation in response to infection in human dendritic cells. *Proc Natl Acad Sci USA*. 2019;116(14):6938-6943. doi:10.1073/pnas.1814700116
80. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *elife*. 2013;2:e00523. doi:10.7554/eLife.00523
81. Pai AA, Pritchard JK, Gilad Y. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genet*. 2015;11(1):e1004857. doi:10.1371/journal.pgen.1004857
82. Bekkering S, Domínguez-Andrés J, Joosten LAB, Riksen NP, Netea MG. Trained immunity: reprogramming innate immunity in health and disease. *Annu Rev Immunol*. 2021;39:667-693. doi:10.1146/annurev-immunol-102119-073855
83. Zhang Q, Cao X. Epigenetic remodeling in innate immunity and inflammation. *Annu Rev Immunol*. 2021;39(1):279-311. doi:10.1146/annurev-immunol-093019-123619
84. Quintin J, Saeed S, Martens JHA, et al. Candida albicans infection affords protection against reinfection via functional reprogramming of monocytes. *Cell Host Microbe*. 2012;12(2):223-232. doi:10.1016/j.chom.2012.06.006
85. Cheng SC, Quintin J, Cramer RA, et al. mTOR- and HIF-1α-mediated aerobic glycolysis as metabolic basis for trained immunity. *Science*. 2014;345(6204):1250684. doi:10.1126/science.1250684
86. Netea MG, Joosten LAB, Latz E, et al. Trained immunity: a program of innate immune memory in health and disease. *Science*. 2016;352(6284):aaf1098. doi:10.1126/science.aaf1098
87. Pacis A, Tailleux L, Morin AM, et al. Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome Res*. 2015;25(12):1801-1811. doi:10.1101/gr.192005.115
88. Ostuni R, Piccolo V, Barozzi I, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell*. 2013;152(1):157-171. doi:10.1016/j.cell.2012.12.018
89. Aracena KA, Lin YL, Luo K, et al. Epigenetic variation impacts individual differences in the transcriptional response to influenza infection. *Nat Genet*. 2024;56(3):408-419. doi:10.1038/s41588-024-01668-z
90. Degner JF, Pai AA, Pique-Regi R, et al. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*. 2012;482(7385):390-394. doi:10.1038/nature10808
91. Alasoo K, Rodrigues J, Mukhopadhyay S, et al. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet*. 2018;50(3):424-431. doi:10.1038/s41588-018-0046-7
92. Husquin LT, Rotival M, Fagny M, et al. Exploring the genetic basis of human population differences in DNA methylation and their causal impact on immune gene regulation. *Genome Biol*. 2018;19(1):222. doi:10.1186/s13059-018-1601-3
93. McVicker G, van de Geijn B, Degner JF, et al. Identification of genetic variants that affect histone modifications in human cells. *Science*. 2013;342(6159):747-749. doi:10.1126/science.1242429
94. Carja O, MacIsaac JL, Mah SM, et al. Worldwide patterns of human epigenetic variation. *Nat Ecol Evol*. 2017;1(10):1577-1583. doi:10.1038/s41559-017-0299-z
95. Waszak SM, Delaneau O, Gschwind AR, et al. Population variation and genetic control of modular chromatin architecture in humans. *Cell*. 2015;162(5):1039-1050. doi:10.1016/j.cell.2015.08.001
96. Fatumo S, Chikowore T, Choudhury A, Ayub M, Martin AR, Kuchenbaecker K. A roadmap to increase diversity in genomic studies. *Nat Med*. 2022;28(2):243-250. doi:10.1038/s41591-021-01672-4
97. Geraghty RJ, Capes-Davis A, Davis JM, et al. Guidelines for the use of cell lines in biomedical research. *Br J Cancer*. 2014;111(6):1021-1046. doi:10.1038/bjc.2014.166
98. Yu J, Vodyanik MA, Smuga-Otto K, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science*. 2007;318(5858):1917-1920. doi:10.1126/science.1151526

99. Okita K, Matsumura Y, Sato Y, et al. A more efficient method to generate integration-free human iPS cells. *Nat Methods*. 2011;8(5):409-412. doi:10.1038/nmeth.1591
100. Strober BJ, Elorbany R, Rhodes K, et al. Dynamic genetic regulation of gene expression during cellular differentiation. *Science*. 2019;364(6447):1287-1290. doi:10.1126/science.aaw0040
101. Kilpinen H, Goncalves A, Leha A, et al. Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature*. 2017;546(7658):370-375. doi:10.1038/nature22403
102. Banovich NE, Li YI, Raj A, et al. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res*. 2018;28(1):122-131. doi:10.1101/gr.224436.117
103. Sabeti PC, Schaffner SF, Fry B, et al. Positive natural selection in the human lineage. *Science*. 2006;312(5780):1614-1620. doi:10.1126/science.1124309
104. Visscher PM, Yengo L, Cox NJ, Wray NR. Discovery and implications of polygenicity of common disease. *Science*. 2021;373(6562):1468-1473. doi:10.1126/science.abi8206
105. Pritchard JK, Di Rienzo A. Adaptation – not by sweeps alone. *Nat Rev Genet*. 2010;11(10):665-667. doi:10.1038/nrg2880
106. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genet*. 2014;10(8):e1004412. doi:10.1371/journal.pgen.1004412
107. Daub JT, Hofer T, Cutivet E, et al. Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol*. 2013;30(7):1544-1558. doi:10.1093/molbev/mst080
108. Freimer JW, Shaked O, Naqvi S, et al. Systematic discovery and perturbation of regulatory genes in human T cells reveals the architecture of immune networks. *Nat Genet*. 2022;54(8):1133-1144. doi:10.1038/s41588-022-01106-y
109. Mamedov MR, Vedova S, Freimer JW, et al. CRISPR screens decode cancer cell pathways that trigger γδ T cell detection. *Nature*. 2023;621(7977):188-195. doi:10.1038/s41586-023-06482-x
110. Võsa U, Claringbould A, Westra HJ, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300-1310. doi:10.1038/s41588-021-00913-z
111. Stranger BE, Stahl EA, Raj T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*. 2011;187(2):367-383. doi:10.1534/genetics.110.120907
112. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. *Cell*. 2017;169(7):1177-1186. doi:10.1016/j.cell.2017.05.038
113. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256-261. doi:10.1038/ng.3760
114. Marcus JH, Novembre J. Visualizing the geography of genetic variants. *Bioinformatics*. 2017;33(4):594-595. doi:10.1093/bioinformatics/btw643

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Randolph HE, Aracena KA, Lin Y-L, Mu Z, Barreiro LB. Shaping immunity: The influence of natural selection on population immune diversity. *Immunol Rev*. 2024;00:1-14. doi:10.1111/imr.13329