IMMUNOGENOMICS

Genetic ancestry effects on the response to viral infection are pervasive but cell type specific

Haley E. Randolph[1], Jessica K. Fiege[2,3], Beth K. Thielen[4], Clayton K. Mickelson[2,3], Mari Shiratori[5], João Barroso-Batista[5], Ryan A. Langlois[2,3], Luis B. Barreiro[1,5,6]*

Humans differ in their susceptibility to infectious disease, partly owing to variation in the immune response after infection. We used single-cell RNA sequencing to quantify variation in the response to influenza infection in peripheral blood mononuclear cells from European- and African-ancestry males. Genetic ancestry effects are common but highly cell type specific. Higher levels of European ancestry are associated with increased type I interferon pathway activity in early infection, which predicts reduced viral titers at later time points. Substantial population-associated variation is explained by cis-expression quantitative trait loci that are differentiated by genetic ancestry. Furthermore, genetic ancestry–associated genes are enriched among genes correlated with COVID-19 disease severity, suggesting that the early immune response contributes to ancestry-associated differences for multiple viral infection outcomes.

Pathogenic viruses are among the strongest sources of selection pressure in human evolution (*1*, *2*). Before the modern era, however, global pandemics on the scale of the 1918 Spanish influenza or the SARS-CoV-2 pandemic were probably rare owing to the restricted potential for long-distance exchange (*3*). If past viral epidemics were geographically stratified, they may have driven population divergence in the frequencies of polymorphisms that mediate the immune response to viral infection. Testing this hypothesis is therefore valuable both for understanding human evolutionary history and for explaining differential susceptibility to viral epidemics in the present day.

Indeed, genetic effects on the response to viruses are well known in human populations (*4*). For example, >120 genetic variants that predict the gene regulatory response to influenza A virus (IAV) in dendritic cells have been identified in humans (*5*). Variation in the transcriptional response to IAV in vitro is also correlated with genetic ancestry in monocytes derived from individuals of African and European descent (*6*). These results suggest that genetic divergence between human populations, especially at loci that are moderately differentiated by genetic ancestry, plays an important role in shaping the immune response to viral infection. However, because studies to date focus on isolated cell types (*5*, *6*), they fail to capture interactions between immune cell types necessary to mount an efficient antiviral response. They also leave unclear whether genetic ancestry effects are specific to, or generalize across, distinct immune cell types.

To address these limitations, we combined single-cell RNA sequencing (scRNA-seq) with in vitro IAV-infection assays in peripheral blood mononuclear cells (PBMCs) from study participants with varying degrees of European versus African genetic ancestry.

Single-cell profiling of the transcriptional response to influenza infection

We exposed PBMCs from a diverse panel of humans (table S1) to either a mock treatment or the pandemic H1N1 Cal/04/09 IAV strain (multiplicity of infection 0.5) (n = 180 samples, paired mock-exposed and IAV-infected samples from 90 males each). We focused on males to avoid potential effects of sex-specific differences in expression (*7*), which would reduce the power of our study. After 6 hours of viral exposure, we performed scRNA-seq on all samples (Fig. 1A). In total, we captured 255,731 single-cell transcriptomes across all individuals and conditions (n = 235,161 high-quality cells) (table S1). We also performed whole-genome sequencing to estimate the proportion of African and European ancestry for each individual (n = 89 individuals who were successfully genotyped) (fig. S1A and table S1). Clustering revealed eight distinct immune cell types (Fig. 1B), with five major cell clusters corresponding to the main PBMC cell types [$CD4^+$ T cells, $CD8^+$ T cells, B cells, natural killer (NK) cells, and monocytes].

We investigated the overall signature of IAV infection by collapsing the single-cell gene expression values for each of the five main clusters and all cells together (i.e., PBMCs) to generate pseudobulk estimates for each sample. Principal component analysis (PCA) of the PBMC pseudobulk data revealed a marked separation of mock and IAV-infected samples on PC1, which explains 43% of the variance in the dataset (paired t test, $P < 1 \times 10^{-10}$) (fig. S1B). Monocytes were the most responsive to IAV infection [n = 3996 differentially expressed (DE) genes identified by using limma (38.3% of those tested compared with 12.4 to 19.6% in other cell types), $|\log_2$ fold change$| > 0.5$, false discovery rate (FDR) < 0.05] (Fig. 1, C and D, and tables S2 and S3) (*8*). Monocytes also exhibited the highest levels of intracellular IAV transcripts (i.e., influenza-derived transcripts generated and processed by infected host cells; three- to sixfold increase in IAV transcript levels in monocytes relative to all other cell types, all t test P values $< 1 \times 10^{-10}$) (Fig. 1E). This observation is consistent with previous work showing that, among blood mononuclear cells, monocytes are particularly susceptible to viral infections (*9*).

We then explored the extent to which the infection response was shared across cell types. Overall, responses were strongly correlated (Pearson's r range, 0.65 to 0.95 for pairwise effect size correlations across cell types among DE genes after IAV infection) (fig. S1C). However, discordant responses were also observed. For example, among DE genes shared by monocytes and NK cells (n = 815), 135 genes (16.6%) (fig. S1D) responded to IAV infection in opposite directions (Fig. 1F). These findings underscore the importance of considering immune responses in a cell type–specific manner. Not only does this approach better capture the biological origins of variation in the response to viral infection, but it also avoids false negative or potentially misleading results that can emerge from bulk analysis.

To further dissect cell type–specific versus shared responses, we generated a specificity score based on variation in the strength of responses across cell types for all genes significantly differentially expressed in at least one cell type (table S4) [see (*10*) for details]. Genes with highly cell type–specific response patterns were enriched for roles in translational initiation and viral gene expression (FDR $< 1 \times 10^{-10}$ for both terms) (Fig. 1G, left, and table S4). By contrast, genes with low specificity scores were enriched for pathways related to type I interferon (IFN) signaling (FDR $< 1 \times 10^{-10}$) and response to type I IFN (FDR $= 2.9 \times 10^{-3}$) (Fig. 1G, right, and table S4). Thus, concordant with previous work in mice (*11*), our data show that the induction of IFN-related genes is a fundamental component of the antiviral response shared across immune cell types (Fig. 1H).

Increased European genetic ancestry predicts a stronger type I and II IFN response after IAV infection

We next identified genes for which expression levels are correlated with quantitative genetic ancestry estimates (i.e., proportion of estimated

[1]Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL, USA. [2]Center for Immunology, University of Minnesota, Minneapolis, MN, USA. [3]Department of Microbiology and Immunology, University of Minnesota, Minneapolis, MN, USA. [4]Division of Pediatric Infectious Diseases and Immunology, Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA. [5]Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL, USA. [6]Committee on Immunology, University of Chicago, Chicago, IL, USA.
*Corresponding author. Email: lbarreiro@uchicago.edu

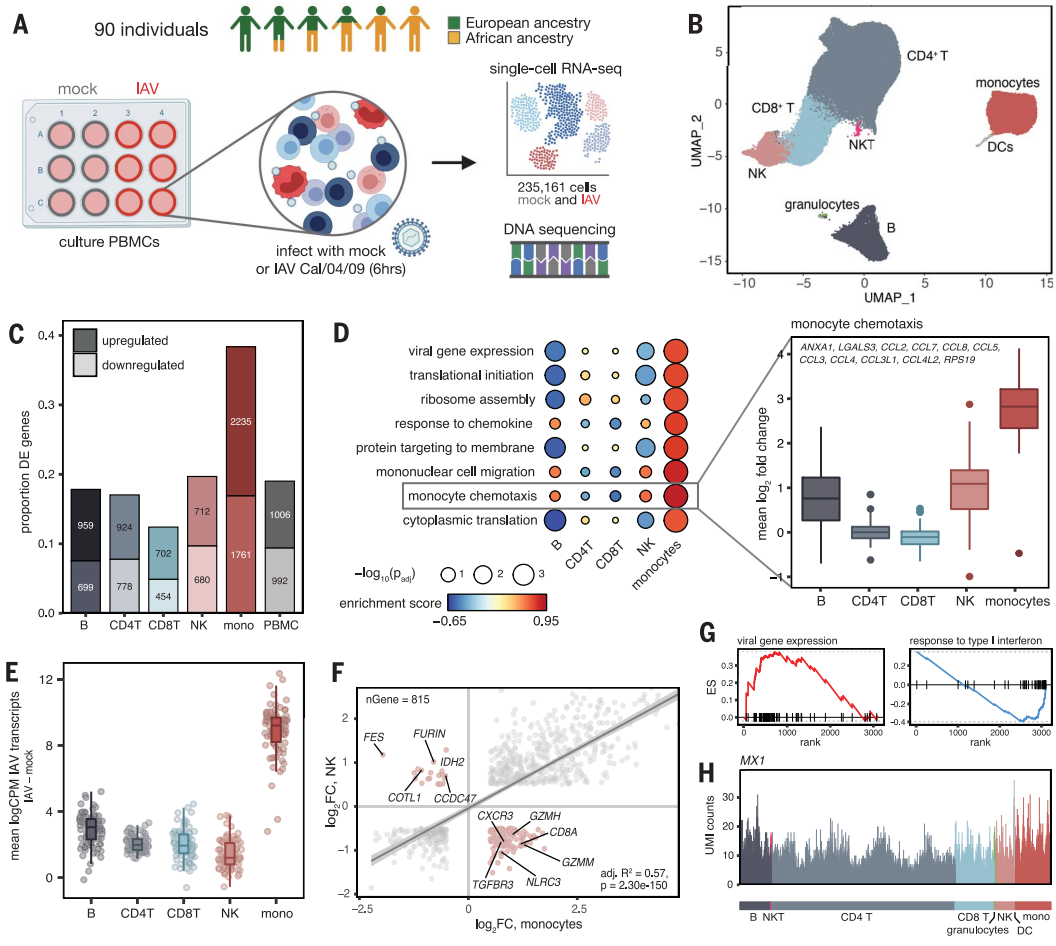

Fig. 1. Shared and cell type–specific responses to IAV infection. (**A**) Study design. (**B**) Uniform Manifold Approximation and Projection (UMAP) of 235,161 mock and IAV-infected cells across individuals. DCs, dendritic cells. (**C**) Numbers and proportions of DE genes upon infection. (**D**) Up-regulated (FDR < 0.10) monocyte-specific GO pathways after infection (table S3). "Monocyte chemotaxis" genes display greater up-regulation in monocytes (plotted means for each individual across genes in IAV minus mock condition, *t* tests, all *P* values < 1×10^{-10} compared with each other cell type). (**E**) Distribution of IAV transcripts across cell types. CPM, counts per million. (**F**) Correlation between global infection effect sizes in monocytes and NK cells among DE genes in both cell types ($n = 815$). *P* value and best-fit slope were obtained from a linear regression model. Highlighted genes (pink) display discordant responses. FC, fold change. (**G**) Example pathways enriched among genes with high (viral gene expression) and low (response to type I IFN) specificity scores. Genes are rank-ordered by specificity score (*x* axis, highest to lowest). ES, enrichment score. (**H**) Unique molecular identifier (UMI) counts per cell in the IAV-infected condition for an example IFN-inducible gene (*MX1*) with a ubiquitous expression pattern.

African ancestry) at baseline, after infection, or both (controlling for age, batch, and other technical covariates). To increase power and improve our effect size estimates for these "population differentially expressed" (popDE) genes, we applied a multivariate adaptive shrinkage method (mash) (*12*), which leverages the correlation structure of genetic ancestry effect sizes across cell types [see (*10*) for details of statistical models]. Across conditions and cell types, we identified 1949 distinct popDE genes [local false sign rate (lfsr) < 0.10], ranging from 830 in NK cells to 1235 genes in $CD4^+$ T cells (Fig. 2A and for distribution of effect sizes, see fig. S2A; table S5). Within each cell type, most popDE genes were shared between conditions (52.9% in monocytes to 77.4% in $CD8^+$ T cells) (Fig. 2A). By contrast, across cell types, genetic ancestry effects on gene expression were highly cell type specific, such that the majority of popDE genes were identified in only one or two cell types (52.2% in mock and 51.4% in IAV infected) (Fig. 2B and fig. S2B, left). Only 17.8% (mock) and 24.7% (IAV infected) of popDE genes exhibited shared genetic ancestry effects across all five cell types (Fig. 2B and fig. S2B, right). Notably, despite differences in study participants' country of origin, IAV strain, and experimental design, our popDE effect size estimates for monocytes were largely concordant with those derived from an independent bulk RNA-seq dataset of IAV-infected monocytes from European- and African-ancestry individuals [Pearson's $r = 0.662$ (mock), Pearson's $r = 0.499$ (IAV); $P < 1.0 \times 10^{-10}$ in both conditions] (fig. S2C) (*6*).

To identify the functional pathways most closely associated with genetic ancestry, we performed gene set enrichment analysis for the MSigDB Hallmark gene sets (Fig. 2C and table S6) (*13*). In monocytes, we identified significant enrichments for multiple immune pathways before infection, including IFN-α response (FDR = 1.9×10^{-3}), IFN-γ response (FDR = 5.4×10^{-4}), TNF-α signaling via NF-κB (FDR = 6.1×10^{-4}), IL-2/STAT5 signaling (FDR = 2.1×10^{-3}), and inflammatory response (FDR = 0.012) (Fig. 2C). In these cases, the enrichments were identified for genes more highly expressed at baseline in individuals with a greater proportion of African ancestry. In IAV-infected monocytes, this pattern reversed: After infection, we observed an enrichment of type I and II IFN pathways (IFN-α response FDR = 0.014 and IFN-γ response FDR = 0.040 in monocytes) in genes more highly expressed with increasing European ancestry (Fig. 2C). Notably, this enrichment of type I and II IFN pathways among genes more highly expressed with greater European ancestry after infection was even clearer in the other four cell types (FDR range, 0.03 to 4.1×10^{-4}) (table S6). To further characterize genetic ancestry–associated differences in the IFN response, we constructed a per-sample score of IFN signaling activity, the "IFN score," which provides an estimate of the average expression of genes belonging to the hallmark IFN-α and IFN-γ gene sets for each individual (*10*). Again, increased European ancestry was strongly correlated with increased IFN score, but only after infection (IAV infected,

mean Pearson's r across cell types = −0.26, Fisher's meta-P = 2.9 × 10^{-6}; mock, mean Pearson's r = −0.0045, Fisher's meta-P = 0.746) (cell type–specific associations are in Fig. 2D and fig. S2D).

These findings suggest that genetic ancestry may also predict the magnitude of the response to IAV infection. In support of this idea, we identified 445 genes for which genetic ancestry was associated with the magnitude of the response to infection [i.e., "population differentially responsive" (popDR) genes, lfsr < 0.10]. PopDR genes were found for all five cell types but were most common in monocytes (popDR genes: $n_{monocytes}$ = 272 versus range of 53 to 181 in other cell types). A core set of 21 popDR genes was shared across all cell types (fig. S3A and table S7). Increased European genetic ancestry predicted a stronger type I and II IFN response (measured as the difference in IFN score between the IAV-infected and mock conditions per individual) across cell types (mean Pearson's r = −0.23, Fisher's meta-P = 6.0 × 10^{-5}) (fig. S3B). This observation was not explained by baseline levels of Cal/04/09–specific serum immunoglobulin G antibodies (a proxy for prior exposure to IAV), which were uncorrelated with genetic ancestry (fig. S3C), the transcriptional response to IAV (fig. S3D), and human leukocyte antigen genotype (*10*). However, stronger type I and II IFN responses predicted increased intracellular IAV transcript levels in PBMCs [adjusted (adj.) R^2 = 0.55, P = 2.8 × 10^{-17}] (Fig. 2E and for cell type–specific effects, see fig. S3E). IAV transcript levels were also significantly higher in individuals with increased European ancestry (Pearson's r = −0.32, P = 0.002) (fig. S3F).

An early-induced type I IFN response is associated with decreased viral titers at later time points

To functionally validate our findings, we infected PBMCs from the 20 individuals with the strongest (n = 10, "high responders") and weakest (n = 10, "low responders") transcriptional type I and II IFN responses at 6 hours post infection (hpi) with IAV. We collected secreted cytokine measurements across eight time points over 48 hours and viral titer measurements at 24 and 48 hpi. High responders produced significantly more secreted IFN-α2 (Fig. 2F, top) and IFN-β (Fig. 2F, bottom) than low responders beginning at 12 hpi, an effect that was exacerbated over time to 4-fold (IFN-α2) and 11.6-fold (IFN-β) more by 48 hpi (P < 0.007 for both cytokines, Mann-Whitney U tests). Viral titers quantified from supernatant at 24 and 48 hpi were also reduced in high responders compared with low responders (Mann-Whiney U tests, P = 0.001 for 24 hpi, P = 0.004 for 48 hpi) (Fig. 2G). None of the 20 study participants in this experiment harbored predicted loss-of-function mutations among genes associated with defects in type I IFN signaling (*14*, *15*), which suggests that these results are not driven by rare genetic variants (table S1). Taken together, these results indicate

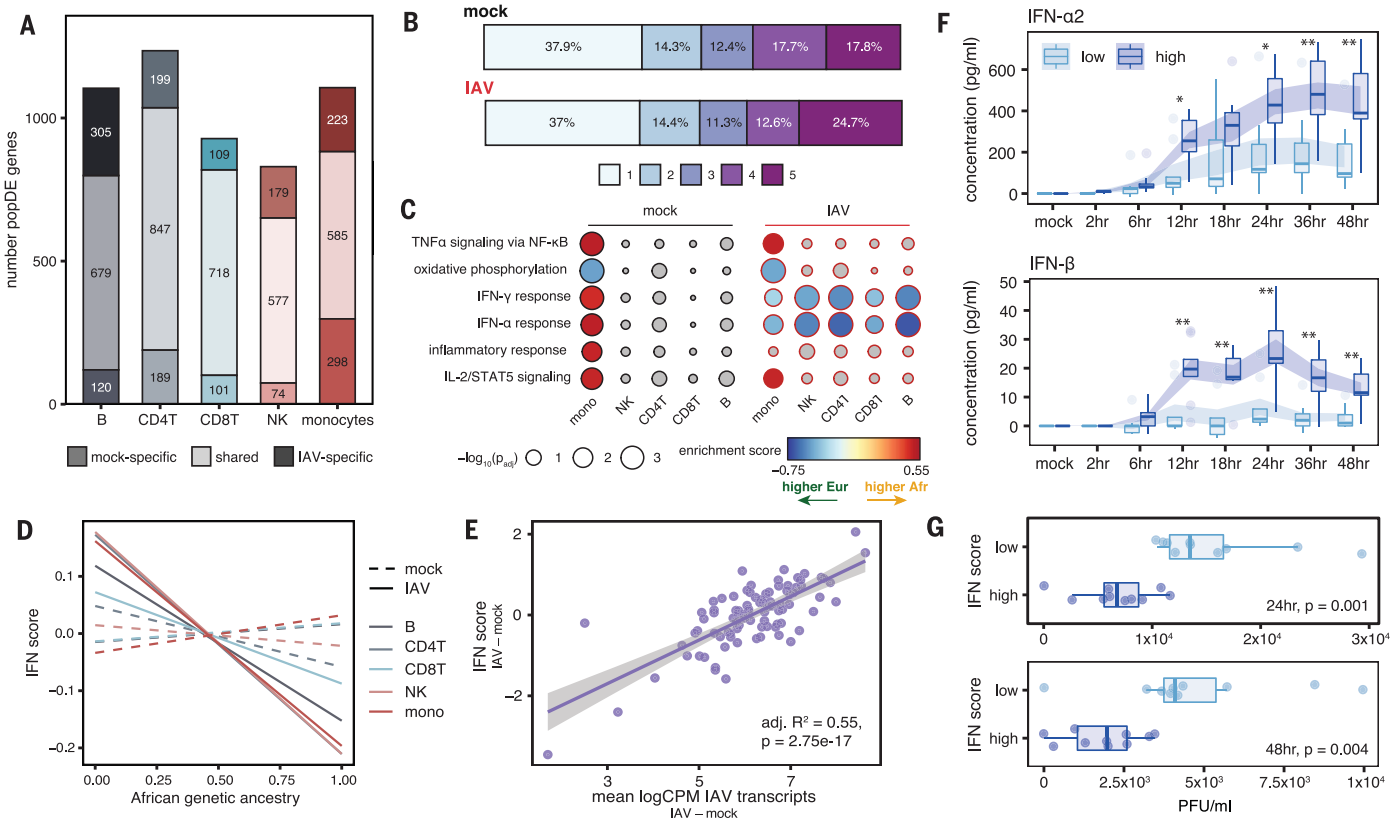

Fig. 2. Genetic ancestry influences the immune response to IAV infection. (**A**) Number of shared and condition-specific popDE genes. (**B**) Cell type sharing of popDE effects (1, detected in a single cell type; 5, detected across all cell types). (**C**) GO enrichments for popDE effects in the mock- and IAV-infected conditions. Colored circles represent pathways with FDR < 0.10. IFN pathways are among the most divergent between European- and African-ancestry individuals in monocytes, with 26% (42 out of 163) of all IFN genes tested classified as popDE after infection. (**D**) Correlation between African genetic ancestry proportion and IFN score in mock (dotted lines) and IAV-infected conditions (solid lines). (**E**) IAV transcript levels are associated with IFN response in PBMCs. (**F**) Secreted IFN-α2 and IFN-β levels in low versus high IFN responders over a 48-hour time course. Shaded area represents the mean ± SE. *P < 0.02, **P < 0.009 (Mann-Whitney U tests). (**G**) Viral titers [plaque-forming units (PFU) per milliliter] detected in supernatant 24 and 48 hpi. In (D) and (E), P values and best-fit slopes were obtained from linear regression models.

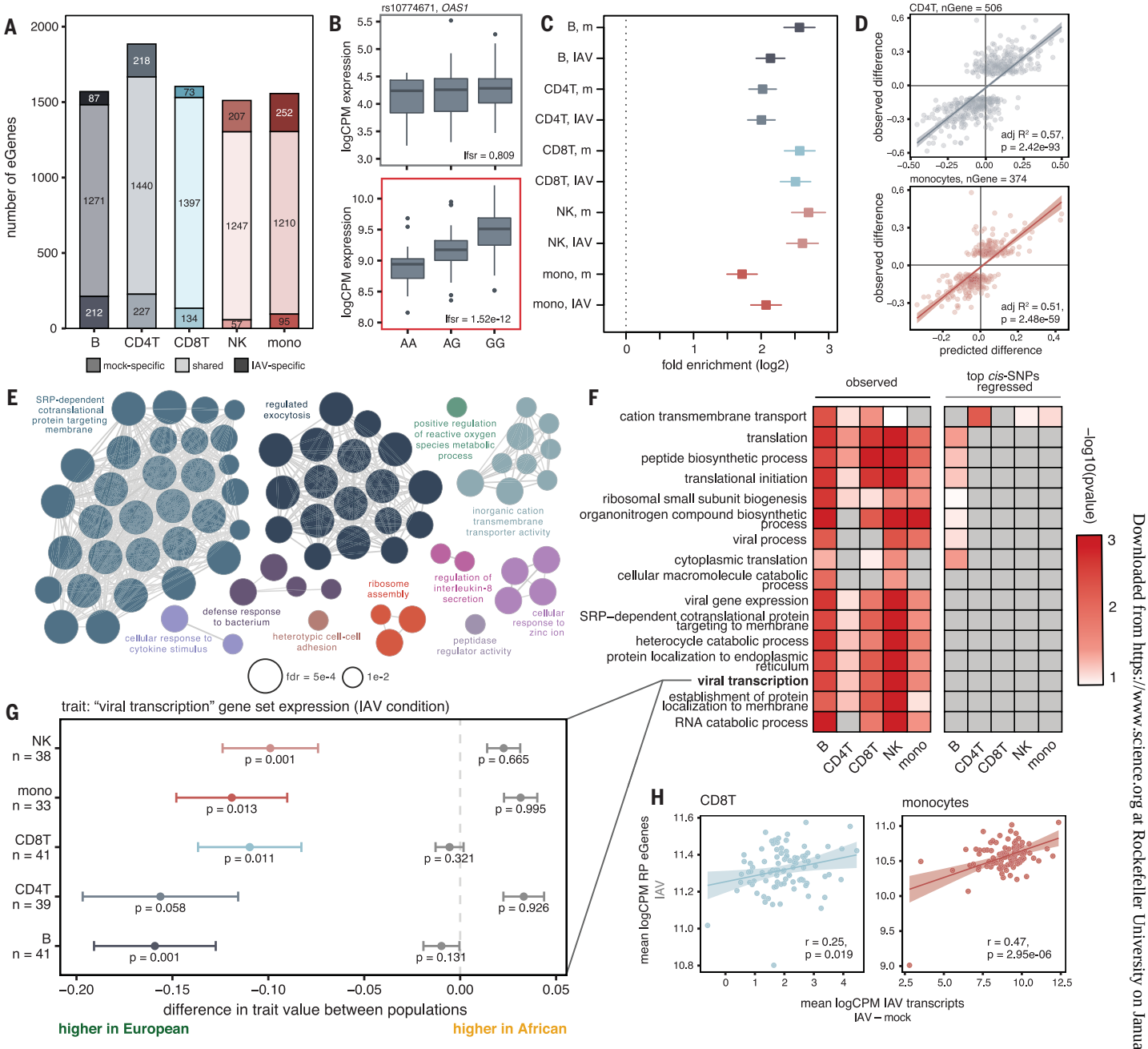

Fig. 3. Cis-regulatory variation drives differences in the antiviral response. (**A**) Number of shared and condition-specific eGenes. (**B**) Condition-specific eQTL example (rs10774671) in CD4$^+$ T cells (top, mock; bottom, IAV infected). (**C**) Enrichment of eGenes among popDE genes in each cell type and condition determined by using logistic regression (log_2-fold enrichment with 95% confidence interval). m, mock. (**D**) Correlation of cis-predicted (*x* axis) versus observed (*y* axis) population differences in expression among popDE genes with an eQTL in CD4$^+$ T cells and monocytes. (**E**) Significant ClueGO enrichments (hypergeometric test, FDR < 0.01) for popDE eGenes across cell types in the IAV-infected condition. SRP, signal-recognition particle. (**F**) Heatmap of $-log_{10}$ *P* values in support of median ancestry-associated differences in gene expression among a subset of enriched GO terms (left) and a model estimating this effect after regressing out the effects of the top cis SNPs for all genes contained in the term (right). (**G**) Example of a GO term for which patterns of population variation are compatible with polygenic selection. PopDE genes with an eQTL that belong to the GO term “viral transcription” (*n* range = 33 to 41 genes) show consistently higher expression levels in European-ancestry individuals [colored points ± SE, median observed ancestry-associated difference (*x* axis) < 0]. After cis-SNP regression (gray points ± SE), the overall trend for higher expression of viral transcription genes in European- compared with African-ancestry individuals is no longer significant. Empirical *P* values were calculated by using a permutation-based approach for (F) and (G) [details in (*10*)]. (**H**) Correlation between IAV transcripts and ribosomal protein eGene expression in CD8$^+$ T cells and monocytes. In (D) and (H), *P* values and best-fit slopes were obtained from linear regression models.

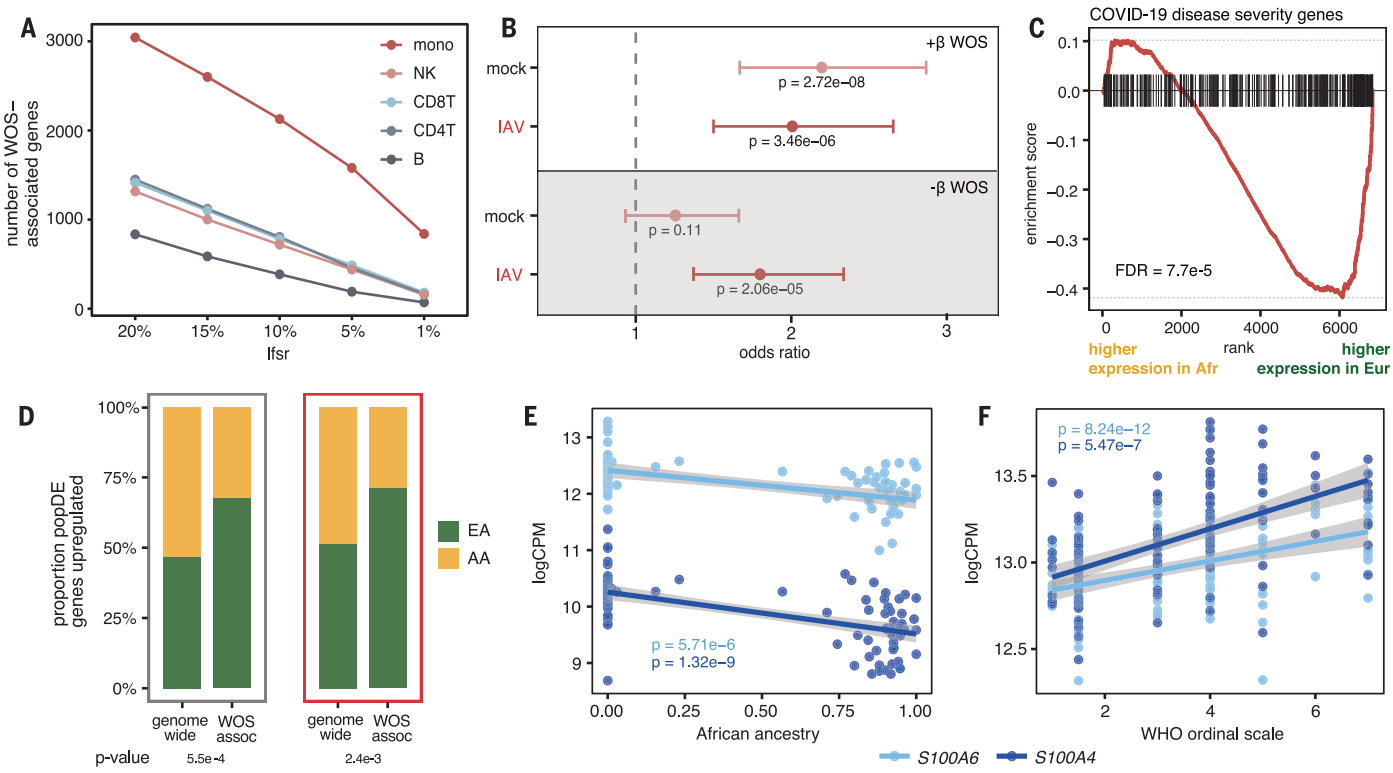

Fig. 4. Genes associated with COVID-19 severity display population-associated variation in expression. (**A**) Number of COVID-19 severity–associated genes by cell type for different significance thresholds (*x* axis). (**B**) Enrichment of popDE genes identified in mock and IAV-infected conditions among genes positively (white) and negatively (gray) associated with severity in monocytes (odds ratio with 95% confidence interval). (**C**) Enrichment plot for genes positively associated with COVID-19 severity in monocytes among the IAV-infection popDE effect sizes in monocytes (*x* axis). (**D**) Proportion of genome-wide popDE and severity-associated popDE genes up-regulated by individuals with a higher level of European (green, EA) or African (yellow, AA) genetic ancestry in mock (gray) and IAV-infected conditions (red). (**E**) Correlation between African genetic ancestry proportion and *S100A4* and *S100A6* expression in monocytes after IAV infection. (**F**) Correlation between WOS and *S100A4* and *S100A6* expression in COVID-19 patients. In (E) and (F), *P* values and best-fit slopes were obtained from linear regression models.

that individuals better able to mount type I IFN responses shortly after infection also displayed a greater capacity to limit productive viral replication later in infection and/or at later time points. These observations are consistent with the finding that individuals with rare immunodeficiencies that lead to defects in type I IFN signaling restrict viral replication poorly and, subsequently, are at increased risk for severe influenza (*16*, *17*).

Cis-regulatory genetic variation explains ancestry-associated differences in gene regulation

To assess the contribution of genetic variation to genetic ancestry–associated differences in the transcriptional response to IAV infection, we mapped expression quantitative trait loci (eQTLs) in the mock and IAV-infected samples. We focused on cis eQTLs, which we defined as single-nucleotide polymorphisms (SNPs) located either within or flanking (±100 kilobases) each gene tested. We identified at least one cis eQTL for 2234 genes (lfsr < 0.10, hereafter referred to as eGenes) across all cell types and conditions (Fig. 3A and table S8). Independent bulk RNA-seq generated from the same samples validated our eGene discovery in the scRNA-seq data (average adj. R^2 = 0.71 for eGene effect sizes in the pseudobulk scRNA-seq and bulk RNA-seq datasets) (fig. S4, A and B).

Although many variants are shared across cell types and conditions (45%, fig. S4C), 13 to 24% of the eGenes identified within each cell type were detected in only one condition, even after probing shared effects with mash (*12*). A small set of 29 eGenes were also only detectable after infection across all cell types, including the key IFN-inducible genes *OAS1* (Fig. 3B), *IFI44L*, *IFIT1*, *IRF1*, and *ISG15* (fig. S4C).

We next tested whether eGenes were likely to be differentially expressed by genetic ancestry. Across cell types and conditions, eGenes (lfsr < 0.10) were 3.2- to 6.5-fold more likely to be classified as popDE (lfsr < 0.10) than expected by chance (Fig. 3C) and 1.3- to 5.0-fold more likely to specifically belong to the set of IFN-associated popDE genes (fig. S4D). These enrichments suggest that ancestry-associated differences in gene expression are likely to have a substantial genetic component, perhaps owing to divergence in allele frequencies at the causal eQTLs. To test this hypothesis, we calculated the correlation between (i) the estimated genetic ancestry effect from our popDE analysis and (ii) the predicted genetic ancestry effect from the effect size of the top eQTL per eGene and the dosage genotype for this SNP across individuals (restricted to popDE genes that were also eGenes in at least one cell type, n = 835 genes) [see (*10*) for details]. The genotype and eQTL effect size for the top eQTLs alone explained an average of 52.5% (mock) and 53.6% (IAV infected) of the variance in genetic ancestry effect sizes across cell types (Fig. 3D and fig. S4E). Thus, among popDE genes with an eQTL, >50% of population differences are explained by differences in the frequency of cis-regulatory variants.

Polygenic selection on ribosomal protein gene expression

We next sought to evaluate whether the intersection set of popDE genes and eGenes clustered into specific biological pathways. Among popDE

genes in which we also observed eQTLs, we identified a strong enrichment for many Gene Ontology (GO) terms related to transcriptional and translational processes, including ribosomal small subunit biogenesis and viral transcription (FDR < 3×10^{-10} in mock and IAV infected) (Fig. 3E and table S9). Consistent population differences in the expression of genes within the same pathway or gene set could be explained by two hypotheses. First, genes in a given gene set may have evolved under relaxed evolutionary constraint, allowing cis-regulatory variants for these genes to diverge in frequency across populations due to genetic drift. Alternatively, if variants within a given pathway have been a repeated target of selection, they may have experienced directionally concordant shifts in allele frequencies across populations, which is a pattern consistent with polygenic selection.

We tested for such a pattern in each of the popDE eGene–enriched pathways in all cell type–condition combinations (n = 10: five cell types in the mock and IAV-infected conditions). To do so, we calculated the median genetic ancestry–associated effect on gene expression (i.e., popDE effect size) across all popDE genes associated with an eQTL for each of the gene sets. Under the hypothesis of neutrality, we expect the direction of ancestry-associated effects to be randomly distributed: Some genes will be more highly expressed in European-ancestry individuals, whereas others will be more highly expressed in African-ancestry individuals. By contrast, under polygenic selection, we expect to find a directional effect, such that most genes for a given pathway show higher expression in one ancestry group than in the other (*18*). Consistent with a history of polygenic selection, most of the GO terms for ribosomal protein (RP)–related pathways (e.g., ribosomal biogenesis, viral transcription) show gene expression levels that are consistently higher in individuals with increased European ancestry across cell types [Fig. 3, F ("observed") and G (colored bars)]. This pattern holds in both mock-exposed (fig. S4F) and IAV-infected cells (Fig. 3, F and G).

An alternative explanation for this observation is that global ancestry is correlated with consistent, directionally biased environmental effects on the expression of genes in RP-related pathways. If so, controlling for local genetic effects on gene expression (e.g., cis eQTL in which allele frequencies are not strongly correlated with ancestry) should not affect the ancestry–gene expression relationship. However, we find the opposite pattern. Specifically, when the effect of the top cis eQTL for each gene is regressed out, the directional bias toward higher expression with increased European ancestry disappears for all RP-related enriched pathways [Fig. 3, F ("top cis-SNPs regressed") and G (gray bars)]. Thus, our results suggest that the higher expression of RP-related pathways in European-ancestry individuals is driven by the cumulative effect of cis-regulatory variants that affect the regulation of genes within these pathways. This shift may in turn be explained by viral infection–induced selection pressures. In support of this possibility, we observed a strong correlation between the average expression of RP eGenes and IAV transcript expression in both $CD8^+$ T cells (Pearson's r = 0.32, P = 0.002) and monocytes (Pearson's r = 0.58, $P < 1 \times 10^{-10}$) (Fig. 3H).

Genes differentially expressed between African- and European-ancestry individuals are enriched among genes associated with COVID-19 severity

The immune pathways activated in response to IAV largely overlap those triggered by other single-stranded RNA viruses (*19*). Thus, our dataset provides an opportunity to evaluate whether differences in COVID-19 susceptibility (caused by SARS-CoV-2, another single-stranded RNA virus) between African Americans and non-Hispanic white Americans (*20*) could be partially explained by differences in population genetic history. We reasoned that if the genetic ancestry–associated differences in gene expression identified in our in vitro infection model also affect susceptibility to COVID-19, those genes should be enriched among genes associated with COVID-19 disease severity in vivo. To test this hypothesis, we reanalyzed a publicly available scRNA-seq dataset consisting of 505,616 PBMCs across 129 COVID-19 patients with varying degrees of disease severity (*21*) based on the World Health Organization Ordinal Scale (WOS) for Clinical Improvement [see (*10*) for details]. Using a model adjusting for age, sex, and self-identified race and ethnicity, we identified genes for which expression levels correlated with severity ("COVID-19 severity–associated genes") within each of the five PBMC cell types included in the IAV dataset. Monocytes, by far, displayed the largest number of genes associated with severity (n = 839, lfsr < 0.01) (Fig. 4A and table S10).

Genes for which higher expression was associated with COVID-19 severity in monocytes (lfsr < 0.01) were 2.0 to 2.2 times more likely to be identified as popDE genes in our single-cell IAV dataset (lfsr < 0.10) compared with genome-wide expectations [Fisher's exact test and permutations, P values = 2.7×10^{-8} (mock) and 3.5×10^{-6} (IAV infected)] (Fig. 4B and fig. S5A). These genes tended to be more highly expressed in monocytes from individuals with more European ancestry [FDRs = 9.8×10^{-5} (mock) and 7.7×10^{-5} (IAV)] (Fig. 4C and fig. S5B). Consequently, 69% of COVID-19 severity–associated genes in monocytes showed increased expression with greater European ancestry, a significantly higher proportion than the 49% observed among all popDE genes [chi-square test, P values = 5.5×10^{-4} (mock) and 2.4×10^{-3} (IAV infected)] (Fig. 4D). Finally, we identified several *S100* family genes among those most strongly associated with both genetic ancestry (Fig. 4E) and COVID-19 disease severity (Fig. 4F). Members of this gene family encode proteins that regulate inflammation and can endogenously activate and amplify inflammatory responses in phagocytes (*22*). *S100A4*, *S100A6*, and *S100A8* expression has been associated with patient improvement when up-regulated early in the course of COVID-19 infection (*21*), and *S100A8* and *S100A9* are systemically up-regulated in immune cells, particularly monocytes, in patients with severe, late-stage COVID-19 (*23*). In our data, *S100A4*, *S100A6*, and *S100A8* are all significantly more highly expressed early after IAV infection in individuals with a greater proportion of European ancestry (Fig. 4E and table S5), consistent with a potential contribution of genetic ancestry to the observed differences in COVID-19 susceptibility between African Americans and European Americans.

Discussion

Together, our results provide a detailed characterization of the genetic determinants that shape interindividual variation in the early response to viral infection in immune cells. Our findings expand on previous work measuring genetic ancestry effects in isolated cell types (*5*, *6*) by showing that most of the ancestry effects on the immune response to IAV are cell type specific. One clear exception to this overall pattern was genetic ancestry–associated differences in the IFN response. Our analysis reveals that, across all cell types, increased European ancestry is associated with a stronger type I IFN response shortly after influenza infection, which in turn predicts reduced viral titers at later time points. Given the central role played by IFNs in conferring antiviral activity to host cells (*16*), our findings have potential clinical implications not only for influenza infection but also for other viruses, including SARS-CoV-2, for which the timing and magnitude of IFN-mediated antiviral responses are associated with disease progression and severity (*24*).

Many of the genetic ancestry–associated differences in immune regulation that we observe are driven by allele frequency differences at cis-regulatory variants. Among popDE genes in which we identify at least one cis eQTL across cell types and conditions, we estimate that, on average, cis eQTLs explain ~53% of the variance in the observed ancestry-associated differences. Our results stress the key role played by genetics in shaping population differences in immune responses, including that these

differences are overwhelmingly due to variants found across populations but that segregate at different frequencies (*6*, *25*). We note, however, that for about half of popDE genes, we were not able to identify an eQTL, pointing to additional, coacting drivers of genetic ancestry–correlated gene expression. These may include other genetic effects [either cis-acting effects or trans-acting effects that we are underpowered to map (*26*, *27*)] or unmeasured environmental factors that are stratified by genetic ancestry.

Viruses have been shown to be among the strongest sources of selection pressure in human evolution (*1*, *2*). Among the different forms of natural selection in humans, polygenic selection is thought to be the most pervasive (*18*), but specific examples of polygenic selection in humans remain rare. Our results provide evidence for ancestry-associated directional shifts in molecular traits (i.e., gene expression phenotypes related to specific biological pathways) that are under cis-regulatory genetic control, which highlights the potential role of polygenic selection in the history of these phenotypes. The best candidate for polygenic selection was observed for RP genes, in which we consistently found that alleles associated with higher expression are also more prevalent in individuals with more European ancestry. This observation represents one of the few instances of polygenic selection in humans that is supported by functional genomic data. The signature of selection at RP genes is particularly interesting in the context of viral infections, as RPs facilitate translation initiation of viral transcripts (*28*) and directly interact with viral mRNA and proteins to enable viral protein synthesis (*29*). Further, a subset of ribosomes, known as immunoribosomes, has been hypothesized to preferentially synthesize antigenically relevant cellular and viral peptides for immunosurveillance by the major histocompatibility complex class I system, which may allow immune cells to more quickly recognize and eliminate infected cells (*30*). Together, these observations raise the possibility that polygenic selection on ribosomal pathways, acting heterogeneously on different human populations, has contributed to present-day variation in viral control.

Our results show that genes differentially expressed by genetic ancestry are enriched among genes associated with COVID-19 disease severity. Our findings suggest that variation in the immune response may therefore interact with or exacerbate environmentally driven health disparities in viral susceptibility and morbidity, which occur for both influenza and COVID-19 (*20*, *31*). An important goal for future work is to evaluate whether the variation we observe early in the viral response translates to differences in COVID-19 patient outcomes. Indeed, time-course studies (*32*, *33*) highlight the importance of temporal dynamics in the immune response to infection, which can include time-dependent reversals of effects. For example, the early up-regulation of antiviral and proinflammatory genes shortly after initial infection has been associated with protection, but their delayed induction is a hallmark of severe illness (*34*). Our results motivate further studies that investigate whether genetic ancestry–linked effects on innate immunity extend to influence the adaptive immune response as well and, ultimately, viral clearance and disease severity over the course of viral infections in vivo.

REFERENCES AND NOTES

1. M. Fumagalli *et al*., *PLOS Genet.* **7**, e1002355 (2011).
2. D. Enard, D. A. Petrov, *Cell* **175**, 360–371.e13 (2018).
3. D. Enard, D. A. Petrov, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375**, 20190575 (2020).
4. A. D. Kenney *et al*., *Annu. Rev. Genet.* **51**, 241–263 (2017).
5. M. N. Lee *et al*., *Science* **343**, 1246980 (2014).
6. H. Quach *et al*., *Cell* **167**, 643–656.e17 (2016).
7. M. Oliva *et al*., *Science* **369**, eaba3066 (2020).
8. M. E. Ritchie *et al*., *Nucleic Acids Res.* **43**, e47–e47 (2015).
9. W. Hou *et al*., *Blood* **119**, 3128–3131 (2012).
10. Materials and methods are available as supplementary materials.
11. Y. Steuerman *et al*., *Cell Syst.* **6**, 679–691.e4 (2018).
12. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, *Nat. Genet.* **51**, 187–195 (2019).
13. A. Liberzon *et al*., *Cell Syst.* **1**, 417–425 (2015).
14. Q. Zhang, *Hum. Genet.* **139**, 941–948 (2020).
15. Q. Zhang *et al*., *Science* **370**, eabd4570 (2020).
16. M. J. Ciancanelli, L. Abel, S.-Y. Zhang, J.-L. Casanova, *Curr. Opin. Immunol.* **38**, 109–120 (2016).
17. M. M. Thomsen *et al*., *Eur. J. Immunol.* **49**, 2111–2114 (2019).
18. J. K. Pritchard, J. K. Pickrell, G. Coop, *Curr. Biol.* **20**, R208–R215 (2010).
19. S. Jensen, A. R. Thomsen, *J. Virol.* **86**, 2900–2910 (2012).
20. J. Y. Ko *et al*., *Clin. Infect. Dis.* **72**, e695–e703 (2021).
21. Y. Su *et al*., *Cell* **183**, 1479–1495.e20 (2020).
22. C. Xia, Z. Braunstein, A. C. Toomey, J. Zhong, X. Rao, *Front. Immunol.* **8**, 1908 (2018).
23. X. Ren *et al*., *Cell* **184**, 1895–1913.e19 (2021).
24. J. S. Lee, E.-C. Shin, *Nat. Rev. Immunol.* **20**, 585–586 (2020).
25. Y. Nédélec *et al*., *Cell* **167**, 657–669.e21 (2016).
26. T. Lappalainen *et al*., *Nature* **501**, 506–511 (2013).
27. H.-J. Westra *et al*., *Nat. Genet.* **45**, 1238–1243 (2013).
28. J.-Y. Huang, W.-C. Su, K.-S. Jeng, T.-H. Chang, M. M. C. Lai, *PLOS Pathog.* **8**, e1002766 (2012).
29. S. Li, *Cells* **8**, 508 (2019).
30. J. Wei, J. W. Yewdell, *Mol. Immunol.* **113**, 38–42 (2019).
31. R. Chandrasekhar *et al*., *Influenza Other Respir. Viruses* **11**, 479–488 (2017).
32. C. Liu *et al*., *Cell* **184**, 1836–1857.e22 (2021).
33. J. P. Bernardes *et al*., *Immunity* **53**, 1296–1314.e9 (2020).
34. S. Zhou *et al*., *Nat. Med.* **27**, 659–667 (2021).
35. H. E. Randolph, *Zenodo* (2021); doi: 10.5281/zenodo.4273999.

ACKNOWLEDGMENTS

We thank J. Tung, B. Mittleman, G. Harrison, and members of the Barreiro laboratory for their constructive comments and feedback. We thank P. Carbonetto and M. Stephens for advice regarding the mash analyses. We thank J. Sanz for guidance with statistics and modeling. We thank J. Ayroles for providing us with the Tn5 transposase used to generate the TM3'seq libraries. Computational resources were provided by the University of Chicago Research Computing Center. We thank the University of Chicago Cytometry Antibody Technologies Facility (RRID: SCR_017760), particularly D. Leclerc and L. Johnston, for their assistance with the Luminex cytokine assays, and the University of Chicago Genomics Facility (RRID: SCR_019196), especially P. Faber, for their assistance with RNA-seq. Figure 1A was created with BioRender.com. **Funding:** This work was supported by grant R01-GM134376 and P30-DK042086 to L.B.B.; H.E.R. was supported by a National Science Foundation Graduate Research Fellowship (DGE-1746045) and a Ruth L. Kirschstein National Research Service Award (F31-HL156419). **Author contributions:** L.B.B. directed the study. H.E.R. and L.B.B. designed the experiments. J.K.F. and B.K.T. generated the influenza A Cal/04/09 strain used for infections, and J.K.F. and C.K.M. performed the plaque assays and baseline antibody titer measurements under the supervision of R.A.L.; H.E.R. performed all in vitro PBMC infection experiments and sample collections. H.E.R. and M.S. performed RNA-seq library preparations. H.E.R. led the computational analyses, with contributions from J.B.B.; H.E.R. and L.B.B. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Fastq and RNA-seq count files are available at GEO under accession number GSE162632. Genome-sequencing data are available at SRA under accession number PRJNA736483. Processed data files, scripts, and associated documentation are available at (*35*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg0928
Materials and Methods
Figs. S1 to S6
Tables S1 to S11
References (*36*–*64*)
MDAR Reproducibility Checklist

10 December 2020; resubmitted 15 March 2021
Accepted 5 October 2021
10.1126/science.abg0928

Genetic ancestry effects on the response to viral infection are pervasive but cell type specific

Haley E. RandolphJessica K. FiegeBeth K. ThielenClayton K. MickelsonMari ShiratoriJoão Barroso-BatistaRyan A. LangloisLuis B. Barreiro

Science, 374 (6571), • DOI: 10.1126/science.abg0928

Ancestry shapes genetic immune responses

Selection for genes affecting the immune system can vary among populations because of selection for local environments. In humans, ancestry has been associated with different responses to infection. Randolph *et al*. examined the molecular determinants of these observations using single-cell RNA sequencing of immune cells from individuals of European and African descent who were infected with influenza in vitro. The experiments showed that infection-induced gene signatures diverged in a cell-type-specific manner that was correlated with ancestry, and that these observed ancestry-related differences were caused by changes in gene regulation and processes involved in transcription and translation. —LMZ