Science
AAAS

Supplementary Materials for

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

Haley E. Randolph, Jessica K. Fiege, Beth K. Thielen, Clayton K. Mickelson, Mari Shiratori, João Barroso-Batista, Ryan A. Langlois, Luis B. Barreiro*

*Correspondence to: lbarreiro@uchicago.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S6
Legends for Tables S1 to S10
Table S11

Other supplementary material for this manuscript includes the following:

Tables S1 to S10 as online Excel files:
  Table S1. Sample meta data.
  Table S2. Global infection effects.
  Table S3. Global infection DE enrichments.
  Table S4. Ranked specificity scores and enrichments.
  Table S5. PopDE effects.
  Table S6. PopDE effect enrichments.
  Table S7. PopDR effects.
  Table S8. eQTL effects.
  Table S9. GO enrichments for popDE genes with an eQTL.
  Table S10. WOS effects.

Materials and Methods

Peripheral blood mononuclear cell (PBMC) collections

This study has been approved by the Institutional Review Board at the University of Chicago (protocol #: IRB19-0432). All samples were obtained from BioIVT. Signed, written consent was obtained from each participant. Blood was collected from 90 male donors between the ages of 21 – 69, who identified as either African American (AA) (n = 45) or European American (EA) (n = 45), from the same collection site in Miami, Florida (United States) utilizing a standard protocol with a sodium heparin anticoagulant. Briefly, PBMCs were extracted from whole blood using a density gradient, washed with HBSS, reconstituted in CryoStor CS10 (Sigma-Aldrich, C2874) to a concentration of 10 million (M) cells/ml, and subsequently cryopreserved. Between 6 - 10M cells per individual were frozen per vial. In addition, paired serum for each individual was collected and frozen. We decided to only focus on males to avoid the potentially confounding effects of sex-specific transcriptional differences in the response to infection. The sample size (n = 90 biological replicates) was chosen based on prior empirical data (*36*) showing that this number of individuals provides sufficient power to detect many *cis*-eQTL.

Our sample was restricted to individuals that met the health requirements specified by the US Food and Drug Administration (FDA) and the American Association of Blood Banks (AABB) at the time of collection (based on the "Full-Length Donor History Questionnaire" (DHQ), https://www.aabb.org/docs/default-source/default-document-library/resources/dhq-v2-1/pdfs/dhq-v2-1-prep-pep-art.pdf). BioIVT's standard donor screening practices also prevent biospecimen collection from donors who exhibit flu-like symptoms, fail a temperature check, or are exposed to a COVID-19-infected individual two weeks prior to collection. In addition, all individuals that reported taking a medication on the Medication Deferral List (https://www.aabb.org/docs/default-source/default-document-library/resources/dhq-v2-1/pdfs/dhq-medication-deferral-list-v2-1-prep-pep-art.pdf) were excluded from blood donation. Finally, following blood collection, blood-borne pathogen testing was performed in accordance with FDA regulations. Only donors who were negative for human immunodeficiency virus (HIV)-1/2 antibodies and hepatitis C virus (HCV) antibodies and non-reactive for hepatitis B virus (HBV) antigen, HBV DNA, HIV-1 RNA, HCV RNA, West Nile virus (WNV) RNA, anti-*Trypanosoma cruzi* antibodies, Zika virus RNA, and anti-*Treponema pallidum* antibodies (serological test for syphilis, STS) were retained in the study.

Individuals in this cohort do show a modest correlation between age and genetic ancestry: European American (EA) individuals are slightly older than African American (AA) individuals in the sample (median age EA = 45, AA = 39, t-test p-value = 0.003). However, the age ranges mostly overlap. Additionally, we always take the effect of age on gene expression levels or the response to IAV into account in our statistical models. Notably, if we remove the small set of EA individuals who are over 55 (n = 6), there is no significant difference in age between EA and AA individuals (p = 0.06). Importantly, popDE effect size estimates are highly consistent whether these individuals are included or excluded from the sample (average adj. R^2 across cell types and conditions = 0.987).

Quantification of baseline Cal/04/09-specific serum IgG antibody levels

The amount of circulating human serum antibodies in the media during the *in vitro* IAV infection experiments is expected to be negligible since PBMCs are washed multiple times prior to culture. Regardless, we measured the baseline levels of Cal/04/09-specific serum IgG antibodies for all individuals in our cohort, as the best available proxy for past immunization in the absence of detailed vaccination or infection history records for our donors. 96 well plates were coated with a 1:25 dilution of UV-killed Cal/04/09 diluted in PBS. All antigen coated plates were blocked with 1% BSA in PBS prior to addition of serum. Serial dilutions of human sera (4-fold serial dilutions, a total of eight dilutions per sample) were added to coated and blocked plates, and bound immunoglobulin (Ig) was detected with HRP-anti-human IgG antibody (Southern Biotech, 2040-05) followed by ABTS Peroxidase Substrate (SeraCare, 5120-0043). OD_{405} was detected by a Synergy H1 plate reader (BioTek). Area under the curve (AUC) was calculated, and linear regressions were performed to determine associations with various variables (Figs. S3C-D). All individuals had detectable levels of Cal/04/09-specific serum titers, but no differences in antibody titers were identified between European and African-ancestry individuals (Fig. S3C). Baseline titers were not correlated with IFN response (measured via the IFN score, Fig. S3D), nor did they influence gene expression levels or the response to IAV ("Effect of baseline serum titers on gene expression", below).

Generation of virus and viral titers

Influenza A virus California/04/2009 (Cal/04/09) virus was rescued in 293T cells (ATCC, CRL-3216) by plasmid-based transfection with IAV Cal/04/09 in the pDZ vector using methods previously described (*37–39*). 24 hours following transfection, 7.5×10^5 MDCK cells (ATCC, CCL-34, NBL-2) were added to the culture in Opti-MEM (ThermoFisher Scientific, 31985062) containing TPCK trypsin (1 μg/ml). For the following two days, 500 μl of Opti-MEM containing TPCK trypsin (2 μg/ml) was added to the culture. One day later, the supernatant was harvested, centrifuged to remove cellular debris, and stored at -80°C. Cal/04/09 was amplified on MDCK cells to generate a stock. Uninfected MDCK cells were cultured for 48 - 72 hours and supernatant was harvested to generate the control, mock-conditioned media. Stocks and Cal/04/09-infected PBMC supernatants (collected as described below in "*In vitro* PBMC infections and sample collections") were plaqued on MDCK cells. PBMC supernatants were treated with TPCK for 30 minutes prior to plaquing. MDCKs were infected in infection media (PBS with 10% Ca/Mg, 1% penicillin/streptomycin, 5% BSA) at 37°C for 1 hour. Infection media was replaced with an agar overlay (2X MEM, 1 μg/ml TPCK trypsin, 1% DEAE-dextran, 5% $NaCo_3$, 2% oxoid agar), and cells were cultured at 37°C for 40 hours then fixed with 4% formaldehyde. Blocking and immunostaining were done for 1 hour at 25°C in 5% milk. Primary stain was mouse anti-Cal/04/09 (1:5000), secondary stain was peroxidase sheep anti-mouse-HRP (1:5000, GE Healthcare, 45001275). The polyclonal mouse anti-Cal/04/09 was generated in-house (mice were infected with 5,000 PFU of Cal/04/09, and serum was harvested 30 days post infection). TrueBlue Peroxidase Substrate (Kirkegard & Perry Laboratories, 50-647-28) was used as directed for detection of virus plaques.

Viral titer measurements were corrected for time course experiment batch (detailed below, "*In vitro* PBMC infections and sample collections") within time point by fitting a model to estimate the effect of experiment batch on titers (titers ~ batch) using the lm function in R, taking the residuals of this model, and adding back the model intercept. Individuals with negative corrected viral titer measurements (n = 1 for 24 hpi, n = 2 for 48 hpi) were manually set to zero. These batch-corrected values are reported in Fig. 2G. A nonparametric 2-group Mann-Whitney U test was used to assess significance between group medians because of the relatively small sample size (n = 20 total).

In vitro PBMC infections and sample collections

PBMCs were unfrozen approximately 14 hours prior to infection and cultured overnight in RPMI 1640 supplemented with 10% fetal bovine serum (Corning, MT35015CV), 2 mM L-glutamine (ThermoFisher Scientific, 25-030-081), and 10 ug/ml gentamicin (ThermoFisher Scientific, 15710064). Infection experiments were performed over 15 batches, where each experimental batch was balanced for self-identified ancestry label to avoid introducing a batch effect confounded with genetic ancestry. The morning of the experiment, 1M PBMCs were plated at a concentration of 1M/ml for each condition, and exposed to either mock-conditioned media (negative control) or Cal/04/09 IAV at a multiplicity of infection (MOI) of 0.5. After 30 minutes, the control media or virus was washed from PBMC cultures. Cells were then replated and incubated for 6 hours at 37°C in 5% CO_2 and 20% O_2. Following the 6 hour incubation, approximately 0.5M cells per sample were collected, washed, and prepared for single-cell capture using the 10X workflow, and approximately 0.5M cells per sample were collected for paired bulk RNA-sequencing. All bulk RNA samples were stored in Qiazol (Qiagen, 79306) at -80°C prior to RNA extraction. Immediately prior to single-cell capture, cells from different samples were combined into two pools (6 samples per pool), each balanced for infection status (mock and IAV-infected) and genetic ancestry (Table S1). Multiplexed cell pools were used as input for the single-cell captures. For each cell pool, 10,000 cells were targeted for collection using the Chromium Single Cell 3' Reagent (v2 chemistry) kit (10X Genomics, 120234). Post Gel Bead-in-Emulsion (GEM) generation, the reverse transcription (RT) reaction was performed in a thermal cycler as described (53°C for 45 min, 85°C for 5 min), and post-RT products were stored at -20°C until downstream processing (no longer than 4 days post-RT reaction). For DNA processing, 1M PBMCs were collected, and DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69504) following the "Cultured cells" protocol.

For the time course IAV infection experiment, PBMCs from 10 individuals with the highest and 10 individuals with the lowest IFN responses were unfrozen and cultured as described above in two batches (first batch n = 10, second batch n = 10, each batch balanced for equal numbers of high [n = 5] vs low [n = 5] responders). Low and high responder groups stratified with respect to expected genetic ancestry. The low responder group included mostly majority-African ancestry individuals (n = 8 African, n = 2 European), and the high responder group contained mostly majority-European ancestry individuals (n = 6 European, n = 4 African) (Table S1).

For each individual, 0.5M cells were exposed to mock-conditioned media, and between 3 - 4M cells (depending on the initial number of cells recovered) were infected with Cal/04/09 IAV at an MOI of 0.5. After 30 minutes, the control media or virus was washed from PBMC cultures, and 0.5M cells were replated at a concentration of 1M/ml for each of the following time points: mock, 2 hours, 6 hours, 12 hours, 18 hours, 24 hours, 36 hours, and 48 hours. PBMCs from all 20 individuals were plated for all time points except at 18 hours (3 missing) because a sufficient number of PBMCs were not recovered to plate this time point across individuals. Cells were incubated at 37°C in 5% CO_2 and 20% O_2 until they were collected for downstream processing. At each time point, bulk RNA from PBMCs and culture supernatant were collected for all individuals plated. All bulk RNA samples were stored in Qiazol (Qiagen, 79306) at -80°C prior to RNA extraction. All supernatant samples were stored at -80°C prior to cytokine level and viral titer measurements.

Luminex cytokine assays

Secreted cytokine measurements were assessed using the Bio-Plex Pro Human Inflammation Panel, 37-Plex kit (Bio-Rad, 171AL001M) according to the manufacturer's instructions. Undiluted supernatant samples were used as input, and analyte measurements were detected using the Luminex 200 System. Cytokine measurements were corrected for time course experiment batch by fitting a model to estimate the effect of experiment batch on cytokine levels (cytokine ~ batch) with the lm function in R, taking the residuals of this model, and adding back the model intercept within time point. These batch-corrected values are reported in Fig. 2F. A nonparametric 2-group Mann-Whitney U test was used to assess significance between group medians because of the relatively small sample size ($n = 10$ per group).

Single-cell RNA-sequencing library preparation and sequencing

Post-RT reaction cleanup, cDNA amplification, and sequencing library preparation were performed as described in the Single Cell 3' Reagent Kits v2 User Guide (10X Genomics). Briefly, cDNA was cleaned with DynaBeads MyOne SILANE beads (ThermoFisher Scientific, NC0949127) and amplified in a thermal cycler using the following program: 98°C for 3 min, [98°C for 15 s, 67°C for 20 s, 72°C for 1 min] x 11 cycles, 72°C 1 min. After cleanup with the SPRIselect reagent kit (Beckman Coulter, B23317), the libraries were constructed by performing the following steps: fragmentation, end-repair, A-tailing, SPRIselect cleanup, adaptor ligation, SPRIselect cleanup, sample index PCR (98°C for 45 s, [98°C for 20 s, 54°C for 30 s, 72°C for 20 s] x 14 cycles, 72°C 1 min), and SPRIselect size selection. Batches of four experiments (corresponding to eight multiplexed single-cell captures) were processed at a time. Prior to sequencing, all multiplexed single-cell libraries ($n = 30$) were quantified using the KAPA Library Quantification Kit for Illumina Platforms (Roche, 501965234) and pooled in an equimolar ratio. Libraries were sequenced 100 base pair paired-end (R1: 30 cycles, I1: 10 cycles, R2: 85 cycles) on an Illumina NovaSeq 6000 to an average depth of 45,612 mean reads per cell across all batches (average median genes detected per cell across batches = 689).

Bulk RNA-sequencing library preparation and sequencing

Total RNA was extracted using the miRNeasy Micro Kit (Qiagen, 217084) according to the manufacturer's instructions. RNA libraries were generated using TM3'seq, a 3'-enriched library preparation protocol that generates transcriptomes of the same quality as those generated using standard RNA-seq approaches (*40*). RNA (50 ng) was added to 1 µl of 0.83 uM oligo (Tn5ME-B-30T) and incubated for 3 minutes at 65°C. Following the incubation, the mixture was combined with 1 µl SMARTScribe™ RT (Takara, 639538), 1 µl dNTPs 10mM (NEB, N0447S), 2 µl DTT 0.1M (Takara, 639538), 4 µl 5X First-Strand buffer (Takara, 639538), and 1 µl B-tag-sw oligo and incubated for 1 hour at 42°C for first strand cDNA synthesis and 70°C for reverse transcriptase inactivation. A cDNA amplification mix was prepared by combining 7.5 µl of first strand cDNA product with 7.5 µl of OneTaq HS Quick-load 2X (NEB, M0486L). This reaction was amplified in a thermal cycler set to the following program: 68°C 3 min, 95°C 30 s, [95°C 10 s, 55°C 30 s, 68°C 3 min] x 3 cycles, 68°C 5 min.

Tn5 transposase (gifted from Julien Ayroles' lab, Princeton University) was combined with pre-annealed adaptor A (10 µM), a forward oligo, at a ratio of 11 to 1 and incubated at 37°C for 30 minutes. The Tn5 mix was then diluted at a ratio of 1 to 5 in a solution of 50% glycerol and 50% reassociation buffer (10 mM Tris pH 8.0, 50 mM NaCl, 1 mM EDTA). This mixture was then diluted 1 to 4 in 5X TAPS buffer (50 mM TAPS, 25 mM $MgCl_2$, 50% v/v DMF). The cDNA was tagmented by combining 4 uL of the diluted Tn5 in TAPS buffer with 11 µl of the second strand cDNA and incubated at 55°C for 7 minutes. Tn5 was dissociated from the cDNA by adding 3.5 µL of 0.2% SDS followed by a 7 minute incubation at 55°C. The cDNA was cleaned using Agencourt AMPure XP beads (Beckman Coulter, A63881), and 8 µl of cDNA was combined with 1 µl of i5 primer at 1 µM, 1 µl of i7 primer at 1 µM, and 10 µl of OneTaq HS Quick-load 2X (NEB, M0486L) prior to amplification. Libraries were amplified in a thermal cycler set to the following program: 68°C 3 min, 95°C 30 s, [95°C 10 s, 55°C 30 s, 68°C 3 min] x 18 cycles, 68°C 5 min. PCR products were pooled and cleaned using Agencourt AMPure XP beads (Beckman Coulter, A63881). Library eluate was stored at -20°C until sequencing. All bulk RNA-sequencing libraries were pooled in an equimolar ratio, and this library pool was sequenced 100 base pair single-end on an Illumina NovaSeq 6000 to an average depth of 7.5 million reads per sample.

Low-pass whole genome DNA sequencing and VCF processing

Out of the 90 individuals in the cohort, 89 were successfully genotyped using DNBseq low-pass whole genome sequencing (BGI) at 4X coverage. Variants were called across individuals using the human reference genome (GRCh37), yielding a merged VCF. The ImputeSeq low-pass imputation pipeline (Gencove) was used to perform VCF imputation. The imputed merged VCF was lifted over to GRCh38 with CrossMap (v0.3.9) (*41*) using the GRCh37 to GRCh38 Ensembl chain file downloaded at ftp://ftp.ensembl.org/pub/assembly_mapping/homo_sapiens/ and the GRCh38 FASTA file from ftp://ftp.ensembl.org/pub/release-92/fasta/homo_sapiens/dna/. For each individual, low-quality variants were filtered by retaining those with a maximum genotype probability (GP in FORMAT field) > 0.90 using QCTOOL (v2.0.7, https://www.well.ox.ac.uk/~gav/qctool_v2/). If the max(GP) for a variant was < 0.90, the variant call was automatically set to missing. Only autosomal, biallelic SNPs were kept for downstream analysis using the SelectVariants function (--select-type-to-include SNP) from the Genome Analysis Toolkit (GATK) (v3.7) (*42*).

Whole exome sequencing

DNBseq whole exome sequencing (BGI, average 50X coverage) was performed on the 20 individuals included in the high versus low responder time course experiment. To construct the exome sequencing libraries, genomic DNA was randomly fragmented by sonication (average fragment size between 150 - 250 bp). DNA fragments were end repaired, A-tailed, and ligated with adapters prior to amplification. Size-selected DNA fragments were amplified by ligation-mediated PCR, purified, and hybridized to the exome array for enrichment. Non-hybridized fragments were washed out. Captured products were circularized, and rolling circle amplification was performed to produce DNA nanoballs, which were sequenced on the DNBseq platform.

Low-quality raw reads were removed if: 1) reads contained a sequence adaptor, 2) reads contained a low-quality base ratio (base quality <= 5) greater than 50%, and 3) reads contained an unknown base ("N") ratio greater than 10%. Filtered, high-quality sequencing reads were mapped to the human reference genome (GRCh37) using Burrows-Wheeler Aligner (BWA-MEM v0.7.17) (*43*). Picard tools (v2.5.0) (http://broadinstitute.github.io/picard/) was used to sort the SAM files by coordinate and convert SAM files to BAM files. Duplicate reads were marked using the MarkDuplicates function in Picard tools (v2.5.0). To obtain more accurate base qualities, Base Quality Score Recalibration (BQSR) was performed using the BaseRecalibrator and ApplyBQSR functions from GATK (v4.1.4) (*42*). Variants were called using GATK's Best Practices for variant analysis pipeline. Briefly, SNPs were detected using the HaplotypeCaller function (GATK, v4.1.4) and selected using the SelectVariants function (GATK, v4.1.4) (*42*). To obtain a high-confidence call set, hard-filtering was applied via the VariantFiltration function (--filter-expression "QD<2.0 || FS>60 || MQ<40 || MQRankSum<-12.5 || ReadPosRankSum<-8.0") (GATK, v4.1.4) (*42*). Finally, SnpEff (*44*) was used to perform various annotations, including gene-based annotation (e.g., to identify whether SNPs cause protein coding changes and determine which amino acids are affected) and the addition of pathogenicity prediction scores (e.g., SIFT, Polyphen-2, and CADD).

To evaluate the contribution of rare genetic variants in the type I IFN pathway, we focused our attention on coding variants within genes previously associated with defects in the type I interferon pathway (*14*, *15*), including four genes reported to be mutated in patients with life-threatening influenza (*GATA2*, *IRF7*, *IRF9, TLR3*), six genes in the TLR3-dependent type I IFN induction pathway (*TICAM1*, *IKBKG*, *UNC93B1*, *TRAF3*, *TBK1*, *IRF3*), and four genes in the IRF7- and IRF9-dependent type I IFN amplification pathway (*IFNAR1*, *IFNAR2*, *STAT1*, *STAT2*).

Of these 14 genes, 13 harbored variants that fall within coding regions (n = 51 coding variants total, all of which have previously been reported in dbSNP). Of these 51 variants, 22 were synonymous mutations that are likely non-pathogenic, 27 were missense variants, 1 was a splice donor variant (*IRF3*), and 1 was a nonsense mutation (*TICAM1*). The *IRF3* splice donor variant is found at high frequency across populations (1000 Genomes Project [1000GP] global allele frequency = 0.37), suggesting that it is not deleterious. The stop-gain mutation in *TICAM1* (c.421C>T, rs387907307) is present in one individual in our cohort (a "high responder") in the heterozygous state. This mutation has previously been associated with herpes simplex encephalitis in children with TRIF deficiency (*45*); however, the specific allele harbored by the individual in our cohort is the variant responsible for the autosomal recessive form of TRIF deficiency (*45*), consistent with observed, normal TRIF-dependent signaling in this individual.

Among the missense variants we detected, 25 of 27 are predicted to be benign/tolerated by CADD, SIFT, and/or PolyPhen-2, and/or are common variants in 1000GP (global allele frequency > 0.10). Two rare variants (rs145480303 in *IRF9* [ExAC allele frequency < 0.001] and rs369756552 in *TICAM1* [1000GP global allele frequency < 0.01]) are predicted to be pathogenic by both PolyPhen-2 and SIFT. However, rs145480303 is reported as a benign variant in ClinVar, and rs369756552 is present in the heterozygous state in one "high responder," suggesting that it does not lead to loss of function. Together, the exome sequencing results suggest that rare, loss-of-function mutations do not explain the large differences in IFN response and viral control observed between high versus low responders.

HLA typing

HLA alleles were imputed from whole-genome sequencing data for each individual using the multi-ethnic HLA imputation panel available through the Michigan Imputation Server. A series of linear models were used to evaluate whether common HLA variants (minor allele frequency $[\mathrm{MAF}]_{\text{both populations}} > 0.05$, n = 88) were associated with either the IFN response or IAV transcript levels across individuals for each cell type (e.g., IFN response ~ HLA genotype dosage). Models were fit using the big_univLinReg function from the R package bigsnpr (*46*). P-values were corrected for multiple testing using the Benjamini-Hochberg procedure (p.adjust function in R, method = "BH"), and genotypes were considered significantly associated with response variables if they had an FDR < 0.05. One variant, HLA-DPB1*03:01, showed a significant association (FDR = 0.004) with IAV transcript levels in $CD4^{+}$T cells, and no variants were associated with IFN response. The potential consequences of this association in our viral infection setting are not clear, especially given that CD4$^+$T cells do not perform MHC class II presentation. However, the HLA-DPB1*03:01 genotype explains only 15.1% of the variance in IAV transcript levels in CD4$^+$T cells, and the association between genetic ancestry and IAV transcript levels in PBMCs remains significant when this genotype is included in the model ($p = 0.007$), suggesting that this allele does not substantially contribute to genetic ancestry-associated variation in our data.

Estimation of genome-wide admixture levels

Prior to estimation of genome-wide admixture proportions, samples were merged with CEU (n = 99, Utah Residents [CEPH] with Northern and Western European Ancestry) and YRI (n = 108, Yoruba in Ibadan, Nigeria) samples from the 1000 Genomes Project (1000GP) Phase 3 dataset (*47*) (downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/GRCh38_positions/). The proportion of European and African genetic ancestry for each individual was estimated using the supervised clustering algorithm in ADMIXTURE (v1.3.0) (*48*). A total of 13,518,147 unlinked SNPs (r^2 between all pairs < 0.1) were used for genetic ancestry assignments, assuming k = 2 ancestral clusters. Self-identified African American (AA, n = 45) individuals had a modest, although highly variable, percentage of European ancestry (mean = 11%, range = 0 - 43%), while self-identified European American (EA, n = 44) individuals displayed more limited levels of African ancestry (mean = 1%, range = 0 - 23%) (Fig. S1A, Table S1). These estimated quantitative genetic ancestry proportions were used to assess differences in immune responses between populations.

Bulk RNA-sequencing data processing

Adapter sequences and low-quality score bases were trimmed from reads using Trim Galore (v0.6.2, Cutadapt v2.2) (*49*) in single-end mode (-q 20 --paired --phred33). Trimmed reads were pseudoaligned to a custom transcriptome containing both the *Homo sapiens* reference transcriptome (GRCh38) and the Cal/04/09 transcriptome (downloaded from Ensembl) using the quant function in kallisto (v0.43) (*50*) (average depth of 4.3 million pseudoaligned reads per sample for the true bulk RNA-seq samples). Gene-level expression estimates were calculated using the R (v3.6.3) package tximport (v1.14.2) (*51*). Expression data was filtered for protein-coding genes that were sufficiently expressed across all samples (median logCPM > 1). After removing non-coding and lowly-expressed genes, normalization factors to scale the raw library sizes were calculated using calcNormFactors in edgeR (v3.26.8) (*52*). The voom function in limma (v3.40.6) (*8*) was used to apply these size factors, estimate the mean-variance relationship, and convert counts to logCPM values. Technical effects (e.g., library preparation batch) were regressed using the ComBat function in sva (v3.32.1) (https://bioconductor.org/packages/sva/). A model evaluating the technical effect of experiment batch (~ 0 + batch, where batch corresponds to a factor variable representing the 15 experimental batches) on gene expression was fit using the lmFit and eBayes functions, and model residuals were obtained using the residuals.MArrayLM function in limma (*8*). The average experiment batch effect was then computed by taking the mean of the capture coefficients across all 15 batches per gene, and this average effect was added back to the residuals.

Single-cell RNA-sequencing mapping, demultiplexing, and initial cell filtering

FASTQ files from each multiplexed capture library were mapped to a custom reference containing GRCh38 and the Cal/04/09 IAV reference genome (downloaded from NCBI, created using cellranger mkref) using the cellranger (v3.0.2) (10X Genomics) count function (*53*). souporcell (v2.0, Singularity v3.4.0) (*54*) in --skip_remap mode (-k 6) was used to demultiplex cells into samples based on genotypes from a common variants file (1000GP samples filtered to SNPs with >= 2% allele frequency in the population, downloaded from https://github.com/wheaton5/souporcell). Briefly, souporcell clusters cells based on cell allele counts in common variants, assigning all cells with similar allele counts to a single cluster corresponding to one individual, while also estimating singlet/doublet/negative status for that cell. For each batch, hierarchical clustering of the true genotypes known for each individual (obtained from low-pass whole-genome-sequencing) and the cluster genotypes estimated from souporcell was used to assign individual IDs to souporcell cell clusters. All 89 individuals were successfully assigned to a single cluster.

After demultiplexing cells into samples, Seurat (v3.1.5, R v3.6.3) (*55*) was used to perform quality control filtering of cells. In total, we captured 255,731 cells prior to filtering (range of cells recovered per capture: min. = 5,534, max. = 10,805). Cells were considered "high-quality" and retained for downstream analysis if they had: 1) a "singlet" status called by souporcell, 2) between 200 – 2500 genes detected (nFeature_RNA), and 3) a mitochondrial reads percentage < 10%, leaving 236,993 cells (n = 19,248 genes).

Clustering, cell type assignment, and UMAP analysis

We performed two versions of clustering analysis and cell type assignment: 1) in which IAV genes were kept in the raw count matrix (used as input for pseudobulk calculations), and 2) in which IAV genes were subset out of the raw count matrix (for visualization of the UMAP in Fig. 1B). All other steps of the clustering workflow (implemented in Seurat v3.1.5) remained the same. Pseudobulk expression estimates (see below) between clustering versions for cell type-matched clusters were extremely similar (adj. $R^2 > 0.999$ for comparisons between versions). For both clustering iterations, we split the cells by infection status (mock or IAV) and ran SCTransform (*56*) to normalize and scale the UMI counts within condition. In this step, we simultaneously regressed out variables corresponding to experiment batch and percent mitochondrial reads per cell. The data was then integrated on infection status using the SelectIntegrationFeatures, PrepSCTIntegration, FindIntegrationAnchors, and IntegrateData workflow (*55*). Following integration, dimensionality reduction was performed via UMAP (RunUMAP function, dims = 1:30) and PCA (RunPCA function, npcs = 30). A Shared Nearest Neighbor (SNN) Graph was constructed using the FindNeighbors function (dims = 1:20, all other parameters set to default), and clusters were subsequently called using the FindClusters algorithm (resolution = 0.5, all other parameters set to default) (*55*).

Clusters were annotated based on the expression of canonical immune cell marker genes ($CD4^+$ T: *CD3D*$^+$, *CD3E*$^+$, *CD8A*$^-$; $CD8^+$ T: *CD3D*$^+$, *CD8A*$^+$; NK cells: *CD3D*$^-$, *NKG7*$^+$, *GNLY*$^+$; monocytes: *CD14*$^+$, *LYZ*$^+$; B: *MS4A1*$^+$; granulocytes: *PRSS57*$^+$; dendritic cells (DCs): *HLA-DRA*$^+$, *HLA-DRB1*$^+$, *CCR7*$^+$, *CST3*$^+$, *CD83*$^+$). A small group of cells, which were identified as B cells, clustered with $CD4^+$ T cells in the UMAP (Fig. 1B), and we investigated this further to see whether this subset represented a distinct, rare cell type. Further analysis revealed that these cells express markers typical of NKT cells, including *CD3D*, *NKG7*, *IL2*, *TNF*, and *IFNG*, and thus, these cells were manually annotated as NKT cells. In the UMAP constructed from input data containing IAV genes, we excluded 1,832 cells for which we could not confidently assign a cell type, as they clustered on the basis of high IAV transcript expression, leaving us with 235,161 cells across all individuals and conditions for downstream analysis (n $CD4^+$ T cells = 138,801, $CD8^+$ T cells =

32,446, monocytes = 27,020, B cells = 22,877, NK cells = 13,220, DCs = 374, granulocytes = 301, NKT cells = 122).

Our main analyses focus on the five most common cell types found in PBMCs, including CD4$^+$T cells, CD8$^+$T cells, B cells, monocytes, and NK cells. Because we focus on these five major cell types, this does not allow us to make conclusions about other cell types that are important in the antiviral response, such as plasmacytoid dendritic cells (pDCs), which are key producers of type I IFNs, and neutrophils. We excluded pDCs and neutrophils from our analyses because, across all individuals and conditions, we only identified 374 pDCs and 301 granulocytes. Because these cell types are found at very low numbers within each individual (range = 1 - 17, median = 4 pDCs; range = 1 - 18, median = 4 granulocytes), we do not have the power to detect infection or genetic ancestry effects in these cell types.

Calculation of pseudobulk estimates

Cluster-specific pseudobulk estimates were used to summarize single-cell expression values into bulk-like expression estimates within samples (where, here, a sample is an individual/infection status pair, n = 180). This was performed for all five major cell types (CD4$^+$ T cells, CD8$^+$ T cells, B cells, monocytes, NK cells) and PBMCs, where all high-quality cells from all cell types identified (n = 235,161) were treated as a single aggregate cluster. Within each cluster for each sample, raw UMI counts were summed across all cells assigned to that sample for each gene using the sparse_Sums function in textTinyR (v1.1.3) (https://cran.r-project.org/web/packages/textTinyR/textTinyR.pdf), yielding an n x m expression matrix, where n is the number of samples included in the study (n = 180) and m is the number of genes detected in the single-cell analysis (m = 19,248) for each of the 6 clusters.

Calculation of capture-corrected expression for downstream modeling

From this point forward, pseudobulk estimates were treated as de facto bulk expression data for each cell type considered. As such, calculations of residuals and downstream modeling of infection and genetic ancestry effects (see below) were performed for each cluster independently. For each cell type, lowly-expressed genes were filtered using cell-type specific cutoffs (removed genes with a median logCPM < 1.5 in CD4$^+$ T cells, monocytes, and PBMCs, < 2.5 in B cells and CD8$^+$ T cells, and < 4.0 in NK cells), leaving the following number of genes per cell type: CD4$^+$ T cells = 9,960, CD8$^+$ T cells = 9,335, B cells = 9,291, monocytes = 10,424, NK cells = 7,109, and PBMCs = 10,430. Cell type-specific logCPM cutoffs were used because of the inherent differences in median logCPM distributions for all genes detected in the single-cell data (n = 19,248) across cell types, which likely reflect variation in the total number of transcripts captured per gene per cell type as well as the number of cells used to calculate the pseudobulk sum estimates per cell type.

After removing lowly-expressed genes, normalization factors to scale the raw library sizes were calculated using calcNormFactors in edgeR (v 3.26.8) (*52*). The voom function in limma (v3.40.6) (*8*) was used to apply these size factors, estimate the mean-variance relationship, and convert raw pseudocounts to logCPM values. A model evaluating the technical effect of capture (~ 0 + capture, where capture corresponds to a factor variable representing the 30 experimental capture batches) on gene expression was fit using the lmFit and eBayes functions, and model residuals were obtained using the residuals.MArrayLM function in limma (*8*). The average capture effect was then computed by taking the mean of the capture coefficients across all 30 capture batches per gene, and this average capture effect was added back to the residuals across samples to generate the capture-corrected expression estimates. We corrected for capture because we viewed it as a potentially important source of batch effects in our data. Indeed, although we were careful to balance each capture for infection status (mock and IAV-infected) and genetic ancestry, we could not process all samples in the same day. Therefore, our correction for capture is much like conventional correction for sequencing flow cell, lane, sampling effort, or extraction batch effects in other types of functional genomic data. The inverse variance weights calculated by voom were obtained and included in the respective lmFit call for all downstream models unless otherwise noted (*8*).

While performing quality control checks on our data, we noticed that the density distributions of the capture-corrected expression estimates were bimodal for some samples in certain cell types. We believe this to be a technical artifact associated with the structure of the single-cell data itself, such that the fewer the number of cells used to create the pseudobulk estimate, the greater the bimodal proportion tends to be (Fig. S6A). This correlation is, however, not linear. Once we reach a large enough number of cells summed to estimate pseudobulk expression values, most distributions become unimodal. Using a linear model including both bimodal proportion and cell counts as covariates, we found that, on average across cell types and conditions, the bimodal proportion explains 5.03% of the total variance in gene expression estimates across genes, as compared to only 1.19% for cell counts, suggesting that the bimodal proportion more accurately captures this technical artifact than cell counts themselves. Importantly, our popDE effect size estimates are highly consistent whether we use bimodal proportion or cell count in the model (adj. R^2 range across cell types and conditions = 0.877 – 0.991). Since this technical effect could contribute to noise or bias in our estimates, we decided to correct for it by including the appropriate cell-type specific bimodal proportion vector across samples as a quantitative technical covariate in our downstream models. Of note, our popDE effect sizes are highly consistent whether we include or exclude bimodal proportion from the model (adj. R^2 range = 0.845 – 0.999).

The bimodal proportion in each cell type for each sample was calculated by: i) estimating the local minimum of the density distribution, ii) subsetting the x-axis on a restricted range that was specific to each cell type, iii) using the x value where y equals the estimated local minimum as the bimodal threshold, and iv) calculating the proportion of genes less than this threshold. Kernel density estimation was performed using the density() function from the base R package stats with default bandwidth and kernel parameters (bw = “nrd0”, kernel = “gaussian”). Per-sample bandwidths across cell types are reported in Table S1. Following estimation of the bimodal proportion via the density function, all expression density distributions were visually inspected to ensure that, if a distribution displayed bimodality, a non-zero bimodal proportion was returned and that it was not substantially over or underestimated. If an expression density distribution was indeed unimodal, then the bimodal proportion was set to zero.

Modeling global infection effects

To obtain estimates of the global infection effects, capture-corrected expression levels of samples corresponding to the same individual were compared in a paired design by introducing individuals as additional covariates to the following differential infection effect model that was run per cell type:

$$\mathrm{M}_1\text{: } E(i,j) \sim \begin{cases} \beta_0(i,j) + \beta_{pB}(i) \cdot pB^{mock}(j) + \varepsilon^{mock}(i,j) \text{ if Condition} = mock \\ \beta_0(i,j) + \beta_{pB}(i) \cdot pB^{IAV}(j) + (\beta_{IAV}(i) + \beta^{IAV}_{age}(i) \cdot age(j)) + \varepsilon^{IAV}(i,j) \text{ if Condition} = IAV \end{cases}$$

Here, $E(i,j)$ represents the capture-corrected expression estimate of gene i for individual j and $\beta_0(i,j)$ represents the intercept corresponding to gene i and individual j (i.e., the expectation of gene i's expression level in the mock-infected sample for individual j). When evaluated, this model gives the global estimate of the IAV infection effect per gene, $\beta_{IAV}(i)$, approximated using the within-individual variation in gene expression across conditions, controlled for the effects of age on the response to IAV (captured by the $\beta_{age}^{IAV}(i)$ term, where age represents the mean-centered, scaled [mean = 0, sd = 1] age per individual). Further, pB^{cdt} represents the bimodal proportion estimated per sample for the respective cell type being modeled (where cdt represents either the mock or IAV), with $\beta_{pB}(i)$ being the corresponding effect on gene expression. Finally, ε^{cdt} represents the residuals for each respective condition (mock or IAV) for each gene i, individual j pair.

We did not perform quantile normalization (per gene, across individuals) on the gene expression estimates prior to fitting M_1 (unlike our approach for models M_2/M_3, "Modeling genetic ancestry effects and integration with mashr", below) because, for the large shifts in expression expected in response to infection, quantile normalization compromises the interpretability of the calculated fold changes (Fig. S6B). After IAV infection, we expect that many genes will have completely non-overlapping gene expression distributions between conditions but will also show variation in the strength of upregulation/downregulation across genes (such as in the cases of *IFIT1* and *IRF1*, Fig. S6B, left). In support of the large effects of IAV, PC1 of the PBMC expression data, which is strongly correlated with IAV infection status, explains 43% of the variance in the dataset, as shown in Fig. S1B. If we quantile normalize data in this scenario, the dynamic range of responses to infection will be abbreviated (Fig. S6B, right). We believe this is biologically misleading and would particularly affect our specificity score analyses, which rely on variation in the strength of responses to IAV across cell types.

Of note, when modeling the expression estimates only for the pseudobulk "PBMC" data (e.g., all cell types combined, results shown in gray bar in Fig. 1C), two additional covariates were added to the model, corresponding to the first two principal components of a PCA performed on an n x m cell type proportion matrix (where n = number of samples = 180, m = number of cell types = 10, with the matrix populated by the cell type proportions for each sample [calculated by the number of cells per cell type cluster for a sample divided by the total number of cells assigned to that sample]) to account for the majority of the variance introduced by underlying cell type composition (PC1 percent variance explained (PVE) = 53.8%, PC2 PVE = 23.2%, total = 77.0%).

These models were fit using the lmFit and eBayes functions in limma (*8*), and the estimates of the global infection effect $\beta_{IAV}(i)$ (i.e., the differential expression effects due to IAV infection) were extracted across all genes along with their corresponding p-values. We controlled for false discovery rates (FDR) using an approach analogous to that of Storey and Tibshirani (*25*, *57*), which makes no explicit assumptions regarding the distribution of the null model but instead derives it empirically. To obtain a null, we performed 10 permutations, where infection status label (mock/IAV) was permuted within individual. We considered genes significantly differentially-expressed upon infection if they had a β_{IAV} $|\log_2\mathrm{FC}| > 0.5$ and an FDR < 0.05.

Calculation of specificity score

This metric was obtained for each gene in two steps: first, we calculated the mean $\log_2$ fold-change response to IAV infection per cell type; second, we computed the coefficient of variation of those mean fold changes across cell types. This analysis was limited to genes significantly differentially expressed following IAV infection ($|\log_2$ fold-change$| > 0.5$ and FDR < 0.05) in at least one cell type. High values indicate highly cell type-specific responses to IAV, while low values indicate shared responses to IAV. While we report specificity scores calculated across all individuals in the main text, these scores are also generalizable to European Americans (EA) and African Americans (AA) separately. If calculated within population, specificity scores are highly correlated between the two genetic ancestry groups (Pearson's $r = 0.82$). Specificity scores are also robust to variation in average expression levels across genes: specificity scores adjusted for mean gene expression levels within cell type are highly correlated with unadjusted scores (adj. $R^2 > 0.99$ across all cell type-condition pairs), suggesting that this calculation is largely unaffected by mean-variance trend relationships.

Calculation of IFN score

To construct the IFN score metric, we summarized the expression patterns of genes involved in the type I/II IFN response as a whole, where, within condition, we i) subset on genes belonging to the hallmark IFN gamma and alpha response pathways (*13*), ii) mean-centered and scaled the expression values for each gene across individuals, and iii) computed the average scaled expression across genes per individual. We defined the IFN response score as the difference in IFN score between the IAV-infected and mock conditions per individual.

Correlation between genetic ancestry and IFN score and permutations

To estimate the correlation between genetic ancestry and IFN score in the mock and IAV-infected conditions, we performed linear regressions to obtain Pearson correlation coefficients and p-values within each cell type. To evaluate if the observed Pearson correlation coefficient was more extreme than expected by chance, we performed permutation testing. To perform permutations within each cell type, we i) randomly permuted the genetic ancestry label across individuals, ii) re-calculated the Pearson correlation coefficient using these permuted genetic ancestry values, and iii) asked whether the null correlation coefficient was equal to or less than the observed, negative correlation coefficient (n permutations = 1,000). For four out of the five cell types tested (all except $CD8^+T$ cells: $p = 0.095$), the observed Pearson correlation coefficient is significantly lower than most values obtained from permutation (B p-value = 0.016, $CD4^+T$ = 0.005, NK = 0.002, monocytes = 0.002). Together, this analysis suggests that the association between genetic ancestry and IFN score in the IAV-infected condition is robust, although less so for $CD8^+T$ cells.

Effect of baseline serum titers on gene expression

To determine whether variation in baseline IAV serum titers had a significant impact on gene expression levels, we evaluated the effect of baseline Cal/04/09-specific serum IgG antibody titers: (1) on gene expression levels at both baseline and following IAV infection, and (2) on gene expression responses following IAV infection (i.e., the interaction between serum titer and transcriptional response to IAV infection) using linear models taking into account age and serum batch. The models were fit using the lmFit function in limma (*8*). P-values were extracted and corrected for multiple testing using the Benjamini-Hochberg procedure (p.adjust function in R, method = "BH"). No genes for which expression levels (model 1) or response magnitude (model 2) significantly (FDR < 0.05) correlated with baseline titers were detected in any of the five cell types. Therefore, the levels of Cal/04/09-specific antibody titers do not significantly influence the early transcriptional response to IAV infection, at least in PBMCs.

Modeling genetic ancestry effects and integration with mashr

Prior to modeling genetic ancestry effects, capture-corrected expression estimates were quantile normalized (QN) within condition using qqnorm in R to minimize the risk of identifying spurious associations due to outlier effects. We note, however, that our key findings are robust to our decision to quantile normalize the data: our popDE and popDR ancestry estimates are highly concordant whether we use QN or non-QN gene expression estimates (adj. R^2 range = 0.66 – 0.92 across all cell types and conditions, mean = 0.82). Moreover, with or without quantile normalization, we observe that higher levels of European ancestry are associated with increased type I/II interferon pathway activity across cell types following IAV infection (*without* QN: mean Pearson's r = -0.26, Fisher's meta-p = 2.9×10^{-6}; *with* QN: mean Pearson's r = -0.25, Fisher's meta-p = 4.8×10^{-6}**),** and that genetic ancestry-associated genes are enriched among genes positively correlated with COVID-19 disease severity (*without* QN: 2.0 to 2.1 fold-enrichment, p-values = 2.5×10^{-5} [mock] and 9.2×10^{-6} [IAV]; *with* QN: 2.0 to 2.2 fold-enrichment, p-values = 2.7×10^{-8} [mock] and 3.5×10^{-6} [IAV]).

The following nested linear model was used to identify genes for which expression levels correlated with the proportion of African ancestry across individuals successfully genotyped (n = 89/90 individuals) within condition (i.e., popDE genes):

$$\mathrm{M}_2\text{: } E(i,j) \sim \begin{cases} \beta_0(i) + \beta_{AA}^{mock}(i) \cdot AA(j) + \beta_{pB}(i) \cdot pB^{mock}(j) + \beta_{age}(i) \cdot\ age(j) + \varepsilon^{mock}(i,j) \text{ if Condition} = mock \\ \beta_0(i) + \beta_{IAV}(i) + \beta_{AA}^{IAV}(i) \cdot AA(j) + \beta_{pB}(i) \cdot pB^{IAV}(j) + \beta_{age}(i) \cdot\ age(j) + \varepsilon^{IAV}(i,j) \text{ if Condition} = IAV \end{cases}$$

Here, $E(i,j)$ represents the capture-corrected expression estimate of gene i for individual j, $\beta_0(i)$ is the global intercept accounting for the expected expression of gene i in a 100% European-ancestry mock-infected individual, $\beta_{AA}^{mock}(i)$ and $\beta_{AA}^{IAV}(i)$ indicate the effects of African admixture (mean-centered, scaled African ancestry proportion, $AA(j)$) on gene i within each condition, and $\beta_{IAV}(i)$ represents the intrinsic infection effect of IAV infection. Age represents the mean-centered, scaled (mean = 0, sd = 1) age per individual, with $\beta_{age}(i)$ being the effect of age on expression levels. All other terms in the model are analogous to that described in M_1. Again, the model was fit using limma (*8*), and the estimates $\beta_{AA}^{mock}(i)$ and $\beta_{AA}^{IAV}(i)$ of the genetic ancestry effects were extracted across all genes, along with their corresponding p-values. Each of these estimates represents the genetic ancestry-related differential expression effects within each condition.

Genes for which the response to IAV infection correlate with the proportion of African ancestry (i.e., popDR genes) were detected using the following model:

$$\mathrm{M}_3\text{: } E(i,j) \sim \begin{cases} \beta_0(i,j) + \beta_{pB}(i) \cdot pB^{mock}(j) + \varepsilon^{mock}(i,j) \text{ if Condition} = mock \\ \beta_0(i,j) + \beta_{pB}(i) \cdot pB^{IAV}(j) + (\,\beta_{IAV}(i) + \beta_{age}^{IAV}(i) \cdot age(j) + \beta_{AA}^{IAV}(i) \cdot AA(j)\,) + \varepsilon^{IAV}(i,j) \text{ if Condition} = IAV \end{cases}$$

This model is similar to M_1 (effect of IAV infection), in that it allows us to obtain estimates based on within-individual variability (i.e., estimation of individual effects), with the difference that the IAV infection effect is no longer built in a genetic ancestry-independent manner, as in model M_1, since it is now dependent on genetic ancestry as follows: $\beta_{IAV}(i) + \beta_{age}^{IAV}(i) \cdot age(j) + \beta_{AA}^{IAV}(i) \cdot AA(j)$. In this model, we explicitly correct for age effects on the response to IAV itself (captured by the $\beta_{age}^{IAV}(i)$ term). In this context, $\beta_{AA}^{IAV}(i)$ denotes the genetic ancestry-infection interaction effect induced by IAV infection (i.e., the effect of genetic ancestry on the response to IAV) corrected for age effects, which represents variation in the response to infection that is correlated with the proportion of African ancestry. The key difference between M_2 and M_3 is that M_2 does not control for the effects of individual identity. Including individual-wise intercepts in M_3 allows us to better take into account the paired nature of the data to assess the effect of admixture on within-individual responses to IAV.

To assess sharing of genetic ancestry effects across cell types and to increase our power to detect these effects, we applied Multivariate Adaptive Shrinkage in R (mashr v0.2.28) (*12*) to the outputs of our popDE and popDR cell type-by-cell type models. mashr is able to learn prior patterns of effect size sharing across data sets (here, cell types and infection conditions) using an empirical Bayes approach, combining information across genes to fit flexible prior models. mashr was applied independently to both the popDE and popDR priors, so all of the following methods were performed twice, once for the popDE effects and then again for the popDR effects. Effect size priors were obtained directly from limma (*8*) and merged into matrices including all effect sizes across cell types, only keeping those genes detected in all cell types (i.e., n x m matrices, where for popDE effects: n = 6,847 genes, m = 10 conditions [mock- and IAV-infected popDE effects for each of the 5 main cell types], and for popDR effects: n = 6,847 genes, m = 5 conditions [popDR effects for each of the 5 main cell types]). Standard errors of the effect size priors were calculated per gene by multiplying the square root of the posterior variance (s2.post) of each gene by the unscaled standard deviation for the effect size of interest for that gene (stdev.unscaled) estimated by limma, and these values were similarly formatted into matrices as described above. To account for correlations among measurements across conditions in our data, we used the estimate_null_correlation_simple function implemented in mashr (*12*) to specify a correlation matrix prior to fitting the mash model. We included both the canonical covariance matrices provided by default in mashr and data-driven covariance matrices (defined as the top 5 PCs from a PCA performed on the significant (lfsr < 0.05) signals detected in the condition-by-condition model results) learned from our data in the mash model fit. For both popDE and popDR effects, the mash model was fit to all tests using the mash function. Posterior summaries of the effect sizes, standard deviations, and measures of significance were extracted. We used the estimated local false sign rate (lfsr) to assess significance of our posterior popDE and popDR effects and considered genes significantly population differentially-expressed or differentially-responsive if the lfsr of the posterior mean was < 0.10. We chose to implement mashr to gain power to identify genetic ancestry effects and to obtain improved effect size estimates by leveraging shared information across different cell types. We used mashr for this purpose because it is both computationally tractable and highly flexible in learning patterns of correlation across conditions.

Comparison with Quach et al. data

Raw count files from Quach et al. (*6*) were downloaded from the European Genome-Phenome Archive (https://ega-archive.org/studies/EGAS00001001895). To determine the extent to which our popDE effects in the monocytes coincided with those identified in the Quach et al. data, we modeled genetic ancestry effects in the Quach et al. dataset (197 European-ancestry and African-Ancestry individuals residing in Belgium) utilizing a similar approach to that used for our data ("Modeling genetic ancestry effects and integration with mashr", Methods). Prior effect size and standard error estimates were obtained using a model similar to M_2 but accounting for dataset-specific covariates (e.g., percent GC content, 5' to 3' bias ratio, experiment date, and library date). To assess sharing of genetic ancestry effects across datasets, we integrated our monocyte results and the Quach et al. results using the mashr (*12*) framework. PopDE effect sizes were largely concordant between the two datasets among genes with significant genetic ancestry effects (lfsr < 0.10) in our single-cell IAV data (Pearson's r = 0.662 for the mock condition, Pearson's r = 0.499 for the IAV condition, $p \ll 1\text{x}10^{-10}$ for both conditions, Fig. S2C).

We note that the observed correlation is quite strong, considering that our study is based on individuals living in Miami, Florida, while the cohort in Quach et al. (*6*) includes individuals living in Belgium (likely leading to variation in ancestry effects due to unmeasured environmental and/or social factors that are differentially correlated with genetic ancestry in different countries). Additionally, while Quach et al. evaluated the transcriptomic responses of purified $CD14^{+}$ monocytes to IAV infection, we performed infections on PBMCs as a whole (allowing us to also capture the effects of paracrine signaling between cell types and direct cell-cell interactions) using a different IAV strain (A/California/04/2009 in this study versus A/USSR/90/1977 in Quach et al).

eQTL mapping and integration with mashr

eQTL mapping was performed using the pseudobulk expression data independently in each cell type against the sets of genes retained after lowly-expressed gene filtering (n genes: $CD4^{+}$ T cells = 9,960, $CD8^{+}$ T cells = 9,335, B cells = 9,291, monocytes = 10,424, NK cells = 7,109, PBMCs = 10,430) and was also performed using the paired "true" bulk RNA expression data. A linear regression model was used to examine associations between SNP genotypes and expression levels, in which expression levels were regressed against genotype. Input expression matrices were quantile-normalized within condition prior to running the association. Mock-exposed and IAV-infected eQTL were mapped separately across all cell types. All regressions were performed using the R package MatrixEQTL (v2.3) (*58*). Only SNPs with a minor allele frequency > 5% across all individuals were tested, and SNPs with > 10% of missing data or deviating from Hardy-Weinberg equilibrium at $p < 10^{-5}$ were excluded (--maf 0.05 --geno 0.10 --hwe 0.00001 PLINK v1.9 filters, www.cog-genomics.org/plink/1.9/) (*59*). In total, 6,305,923 SNPs passed our quality-control filters. Local associations (i.e., putative *cis*-eQTL) were tested against all SNPs located within the gene body or 100kb upstream and downstream of the transcription start site (TSS) and transcription end site (TES) for each gene tested. We recorded the minimum p-value (i.e., the strongest association) observed for each gene, which we used as statistical evidence for the presence of at least one eQTL for that gene. To estimate an FDR, we permuted the genotype data ten times, re-performed the linear regressions, and recorded the minimum p-values for the gene for each permutation. These sets of minimum p-values were used as an empirical null distribution and FDRs were calculated using the method described in the section "Modeling global infection effects".

Power to detect *cis*-eQTL can be increased by accounting for unmeasured surrogate confounders. To identify these confounders, we first performed PCA on a correlation matrix based on gene expression for mock and IAV-infected samples. Subsequently, up to 20 principal components (PCs) were regressed out prior to performing the association analysis for each gene. A specific number of PCs to regress in each condition and cell type, corresponding to the number of PCs that empirically led to the detection of the largest number of eQTL in each condition, was then chosen from these results. The exact number of PCs regressed in each of the analyses can be found in Table S11. Of note, while PC corrections increase our power to detect eQTL, they do not affect the underlying structure of the expression data.

Mapping was performed combining both EA and AA individuals to increase power. To avoid spurious associations resulting from population structure, the first two eigenvectors obtained from a PCA on the genotype data using SNPRelate (v1.20.1, gdsfmt v1.22.0) (*60*) were included in the Matrix eQTL model. Other covariates included in the linear model were the following: the condition/cell type-specific bimodal proportion and age (mean-centered, scaled), with two additional covariates included when mapping eQTL using the PBMC expression data, corresponding to the first two PCs from the cell type composition PCA described in "Modeling global infection effects".

Our ability to detect eQTL in the pseudobulk expression data was highly dependent on the number of cells identified in each cell type cluster (correlation between the total number of cells recovered per cell type across all individuals/conditions versus the number of significant eQTL [$\mathrm{FDR} < 0.10$] detected: adj. $R^2 = 0.983$, $p = 1\mathrm{x}10^{-8}$). To gain power to detect *cis*-eQTL effects using sharing information across cell types, we again implemented mashr (*12*). Out of necessity of the method, we only considered shared genes that were tested across all cell types (n = 6,573). For each of these genes, we chose a single, top *cis*-SNP, defined as the SNP with the lowest FDR across all cell types (n = 5) and conditions (n = 2), to input into mashr, yielding a total of 6,573 gene-SNP pairs. We extracted the prior effect sizes (betas) and computed the standard errors (SEs) of these betas (defined as the beta divided by the t-statistic) from the Matrix eQTL outputs for each gene-SNP pair across cell types and conditions. We defined a set of "strong" tests (i.e., the 6,573 top gene-SNP associations) as well as a set of random tests, including both null and non-null tests, which we obtained from randomly sampling 200,000 rows of a matrix containing all gene-SNP pairs tested by Matrix eQTL merged across conditions. Our mashr workflow was as follows: i) the correlation structure among the null tests was learned using the random test subset, ii) the data-driven covariance matrices were learned using the strong test subset, iii) the mash model was fit to the random test subset using canonical and data-driven covariance matrices, with two additional "infection" covariance matrices (i.e., one matrix capturing shared effects in only the mock-exposed samples and another matrix capturing shared effects in only the IAV-infected samples), and iv) the posterior summaries were computed for the strong test subset (*12*). We used the estimated local false sign rate (lfsr) to assess significance of our posterior eQTL effects and considered a gene-SNP pair to have a significant eQTL effect if the lfsr of the posterior mean was < 0.10, which we defined as an eGene. For the effect size comparison and eGene sharing analyses between the pseudobulk PBMC and true bulk PBMC expression data (Figs. S4A and S4B), matrix eQTL effect size estimates for the same top gene-SNP associations as defined above were input into mashr (*12*) to obtain posterior effect size estimates and significance values.

Identification of condition-specific popDE genes and eGenes

Within each cell type, we considered either popDE genes or eGenes as condition-specific (i.e., only showing an effect in either the mock or IAV infection condition) if they had an lfsr < 0.10 in only one condition. Here, we assume that the risk of identifying a true effect in both mock and IAV-infected cells (i.e., a shared popDE gene/eGene) as falsely condition-specific due to lack of power is low, specifically because we employed the multivariate adaptive shrinkage framework, which draws information across conditions to make better-informed posterior estimates regarding the sharing of effects, so we do not expect to see many posterior effects called as "condition-specific" when, in fact, they are not.

Enrichment of eGenes within popDE genes

We tested for an enrichment of eGenes among the genes identified as popDE genes within each cell type and condition. For each cell type, condition pair, we created two vectors: i) a popDE gene vector, where significant popDE genes (lfsr < 0.10) were coded as a 1 and non-significant popDE genes were coded as a 0, and ii) an eGene vector, where significant eGenes (lfsr < 0.10) were coded as a 1 and non-significant eGenes were coded as a 0. The logistic regression was performed on the popDE gene and eGene vectors using glm in R, where the eGene vector was used as the predictor variable and the popDE gene vector was used as the response variable (popDE[0,1] ~ eGene[0,1]). The odds ratios output by glm were converted to $\log_2$ fold enrichments with a 95% confidence interval (plotted along the x-axis in Fig. 3C).

Calculation of predicted and observed population differences in expression

We estimated the predicted *cis*-genetic population differences in gene expression using a method in which we first computed the predicted expression of each gene considering only the posterior effect size of the top *cis*-SNP for that gene and an individual's genotype dosage (a vector of 0, 1, or 2), where, for gene *i,* individual *j*:

$$\text{predicted expression}_{i,j} = \text{eQTL effect size}_{i} * \text{genotype}_{j}$$

We then modeled these predicted expression values using a model analogous to that of M_2 (model evaluating the popDE effects, "Modeling genetic ancestry effects and integration with mashr") to obtain the predicted genetic ancestry effects (plotted on the x-axis for genetically-driven popDE genes in Fig. 3D). The observed population differences in expression were taken directly from the post-mash effect size estimates of M_2 (plotted on the y-axis for genetically-driven popDE genes in Fig. 3D).

Modeling the effect of *cis*-regression on the observed population differences in expression

To assess the impact of *cis*-regression on population-associated expression differences, we used two models evaluating the effect of continuous genetic ancestry (African ancestry proportion) on gene expression: i) a model analogous to M_2 (model evaluating the popDE effects, "Modeling genetic ancestry effects and integration with mashr"), and ii) a model in which, for each gene, the top *cis*-SNP for that gene was regressed by including the genotype dosage for that SNP across individuals as a covariate in the model. The models were fit using limma (*8*) and mashr was applied (as described in the section "Modeling genetic ancestry effects and integration with mashr") (*12*) to the prior effect sizes and standard errors derived from both models independently. The mashr posterior summaries were used to directly obtain the population differences in expression for each gene.

For each significantly enriched GO term (FDR < 0.01, hypergeometric test) identified in Fig. 3E (see "Gene set enrichment analyses" section below), we calculated summaries of the observed population difference in expression among the genes that belong to each term that are also popDE genes with evidence of an eQTL in at least one cell type. To do this, for each cell type for each term, we collected the observed population differences among these term-specific genetically-driven popDE genes and calculated the median and standard error (SE) for these values (plotted on the x-axis in Fig. 3G). This was performed for both the observed ("real") model outputs (model i) as well as the *cis*-regressed model outputs (model ii). For each cell type, we obtained a p-value for the real effects using a permutation method. To obtain a null distribution, we performed 1,000 permutations where, for each iteration, we: 1) sampled the same number of observed term-specific, genetically-driven popDE genes for that cell type from a background set of all genetically-driven popDE for that cell type, 2) obtained the population differences in expression among these genes, and 3) calculated the median for these null values. We then computed a one-sided, empirical p-value, where we considered the number of instances more extreme in the median null difference compared to the median observed difference in the real data given the sign of the difference in the real data (i.e., if the observed difference in the real data was < 0, we counted the number of observations in the null distribution equal to or less than the observed value, and if the observed difference in the real data was > 0, we counted the number of observations in the null distribution equal to or greater than the observed value), where p = number of instances more extreme divided by the number of permutations (n = 1,000). Similarly, we obtained a p-value for the *cis*-regressed effects using the same method, except for that in steps 2 and 3, we considered the *cis*-regressed population differences as opposed to those seen in the real data. To calculate the directional p-value for the *cis*-regressed case, we used the magnitude of the median *cis*-regressed population difference but still considered the sign of the median observed population difference.

Processing of COVID-19 patient single-cell RNA-sequencing data

Raw and normalized count files from Su et al. (*21*) were downloaded from EMBL-EBI ArrayExpress (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9357/). The entire dataset consisted of 549,047 cells across 270 samples (505,616 cells from COVID-19 patients, n = 254 COVID-19 samples). Normalized counts were scaled and the effect of percent mitochondrial reads per cell was regressed using the ScaleData function in Seurat (v3.1.5) (*55*). After scaling, dimensionality reduction was performed via UMAP (RunUMAP function, dims = 1:30) and PCA (RunPCA function, npcs = 30) (*55*). A Shared Nearest Neighbor (SNN) Graph was constructed using the FindNeighbors function (dims = 1:20, all other parameters set to default), and clusters were subsequently called using the FindClusters algorithm (resolution = 0.5, all other parameters set to default) (*55*). Clusters were annotated as described in the section "Clustering, cell type assignment, and UMAP analysis," and any clusters with ambiguous immune cell marker gene expression patterns were excluded. In total, considering the five major PBMC immune cell populations, we identified 493,809 cells ($CD4^+$ T cells = 145,698, $CD8^+$ T cells = 79,626, monocytes = 174,980, B cells = 33,478, and NK cells = 60,027) across samples. Pseudobulk estimates and associated per-sample bimodality proportions were calculated for each cell cluster independently as described in "Calculation of pseudobulk estimates" and "Calculation of capture-corrected expression for downstream modeling". Lowly-expressed genes were filtered using cell-type specific cutoffs (removed genes with a median logCPM < 1.5 in $CD4^+$ T cells and monocytes, < 2.0 in NK cells and $CD8^+$ T cells, and < 3.0 in B cells), leaving the following number of genes per cell type: $CD4^+$ T cells = 10,079, $CD8^+$ T cells = 9,809, B cells = 8,865, monocytes = 10,069, NK cells = 10,301. Batch-corrected expression estimates and inverse variance weights were obtained as described in "Calculation of capture-corrected expression for downstream modeling".

Modeling COVID-19 disease severity-associated genes

Only COVID-19 patient samples from Blood.draw.time.point == "T1" (blood draw performed shortly after the initial clinical diagnosis (*21*)) and with non-missing disease severity scores (n = 129) were included in the downstream analyses. Additionally, two individuals were removed due to extremely skewed pseudobulk expression density distributions, leaving n = 127 individuals for downstream analysis. Severity of COVID-19 was assessed using the 9-point World Health Organization (WHO) Ordinal Scale (WOS) for Clinical Improvement (https://www.who.int/publications/i/item/covid-19-therapeutic-trial-synopsis) that includes the following categories: 0 = uninfected - no evidence of infection; 1 = ambulatory - no limitation of activities; 2 = ambulatory - limitation of activities; 3 = hospitalized, mild - no oxygen therapy; 4 = hospitalized, mild - oxygen by mask or nasal prongs; 5 = hospitalized, severe - non-invasive ventilation or high-flow oxygen; 6 = hospitalized, severe - intubation and mechanical ventilation;

7 = hospitalized, severe - ventilation and additional organ support; and 8 = dead – death. If WOS score for an individual was coded as “1 or 2”, this value was manually set to 1.5. If self-identified ethnicity or race was not reported for an individual (n missing ethnicity = 5, n missing race = 18), this missing data was filled with the factor variable most likely to represent the missing value, i.e., for ethnicity, "Unknown / Not Reported" label changed to "Not Hispanic or Latino" label, and for race, "Unknown / Not Reported" label changed to “White” label.

Prior to modeling WOS effects, batch-corrected expression estimates were quantile-normalized using qqnorm in R. The following linear model was used to identify genes for which expression levels were correlated with WOS across individuals (i.e., WOS-associated genes):

$$E(i,j) \sim \beta_0(i) + \beta^{T1}_{WOS}(i) \cdot WOS(j) + \beta_{pB}(i) \cdot pB^{T1}(j) + \beta_{age}(i) \cdot age(j) + \beta_{sex}(i) \cdot sex(j) + \beta_{ethnicity}(i) \cdot ethnicity(j) + \beta_{race}(i) \cdot race(j) + \varepsilon^{T1}(i,j)$$

Here, $E(i,j)$ represents the batch-corrected expression estimate of gene i for individual j, $\beta_0(i)$ is the global intercept accounting for the expected expression of gene i in a female, non-Hispanic white individual, and $\beta^{T1}_{WOS}(i)$ indicates the effect of WOS ($WOS(j)$) on gene i in blood draw time point T1. Further, pB^{T1} represents the bimodal proportion estimated per sample for the respective cell type being modeled in blood draw T1 samples, age represents the mean-centered, scaled (mean = 0, sd = 1) age per individual, sex represents the self-identified sex for each individual (factor levels = “Male”, “Female”), ethnicity represents the self-identified ethnicity for each individual (factor levels = "Not Hispanic or Latino", "Hispanic or Latino"), and race represents the self-identified race for each individual (factor levels = "White", "Asian", "Black or African American", "Native Hawaiian or Other Pacific Islander", "American Indian/Alaska Native", "More Than One Race"). Finally, ε^{T1} represents the residuals for the T1 time point samples for each gene i, individual j pair. The model was fit using limma (*8*), and the estimate $\beta^{T1}_{WOS}(i)$ was extracted across all genes, along with the corresponding p-values. This estimate represents the WOS-associated differential expression effects within blood draw time point T1.

We applied mashr (v0.2.28) (*12*) to assess cell type sharing of WOS effects as described in “Modeling genetic ancestry effects and integration with mashr”. Only genes detected across all cell types considered were kept (n = 7,866 genes). Posterior summaries of the effect sizes, standard deviations, and measures of significance were extracted. The estimated local false sign rate (lfsr) was used to assess significance of WOS effects, and we considered genes significantly WOS-associated if the lfsr of the posterior mean was < 0.01. To calculate enrichments of popDE genes among WOS-associated genes (Figs. 4B and S5A), only WOS-associated genes in the background set of genes tested for popDE effects across cell types (n = 6,847 genes) were considered.

Gene set enrichment analyses

Gene set enrichment analysis was performed using three independent methods, including fgsea (https://bioconductor.org/packages/release/bioc/html/fgsea.html), GOrilla (*61*), and ClueGO (*62*), depending on the type of data being evaluated. The enrichment program specifications and the data in which they were used to assess enrichments are described below:

The R package fgsea (v1.10.1) was used to perform gene set enrichment analysis for the global infection effects using the C5 gene ontology (GO) biological processes gene sets (Fig. 1D), for the popDE effects using the H hallmark gene sets (Fig. 2C) (*63*), and for the popDE effects using the WOS-associated gene sets (Figs. 4C and S5B). For the infection effects, t-statistics were obtained directly from the topTable function in limma (*8*), and for the popDE effects, t-statistics were calculated from the posterior mashr outputs, where the t-statistic = posterior effect size divided by the posterior standard error for each gene. For infection effects, the background sets were the sets of genes sufficiently expressed (i.e., passed the lowly-expressed gene filter threshold) for each cell type. For popDE effects, the background set was the set of genes detected in all cell types (i.e., the intersection set of genes that were measurably expressed across all cell types, n = 6,847). This set was chosen as the background because we conducted our enrichment analyses using t-statistics calculated from mash-adjusted posterior effect sizes and standard errors (see "Modeling genetic ancestry effects and integration with mashr"), which could only be calculated for genes detectably expressed in all conditions (i.e., all 5 cell types in mock- and IAV-infected conditions). The t-statistics were then ranked, and these pre-ranked t-statistics were used to perform the enrichment using fgsea with the following parameters: minSize = 15, maxSize = 500, nperm = 100000. Enrichments scores (ES) and Benjamini-Hochberg adjusted p-values output by fgsea were collected for each analysis.

We also used fgsea to generate the barcode plots shown in Fig. 1G to visualize where the genes in the highlighted pathways are found in the ranked specificity score list among the set of all infection differentially-expressed genes in at least one cell type. To obtain p-values for the ranked list of specificity scores, we used GOrilla (*61*). GOrilla relies on a statistical framework (the minimum hypergeometric score) that allows the calculation of exact p-values for observed enrichments in ranked lists of genes, taking into account multiple testing without needing to perform simulations, unlike fgsea. Because GOrilla only identifies GO terms that are significantly enriched at the top of the ranked gene list, we performed the enrichments in two ways, once with the list ranked from high to low specificity scores and again with the list ranked from low to high specificity scores. The Benjamini-Hochberg adjusted FDR q-values calculated by GOrilla for the "viral gene expression" and "response to type I interferon" terms are reported in Fig. 1G.

We performed gene set enrichment analysis for our intersection set of popDE genes and eGenes (Fig. 3E) using the ClueGO (v2.5.7) (*62*) Cytoscape (v3.7.1) (*64*) module in functional analysis mode, where the target set of genes was the list of popDE eGenes in the mock or IAV condition and the background set was the list of genes tested across all cell types. Specifically, we tested for the enrichment of GO terms related to biological processes (ontology source: GO_BiologicalProcess-EBI-UniProt-GOA_04.09.2018_00h00) using the following parameters: visual style = Groups, default Network Specificity, no GO Term Fusion, min. GO Tree Interval level = 3, max. GO Tree Interval level = 8, min. number of genes = 3, min. percentage of genes = 4.0, statistical test used = Enrichment/Depletion (two-sided hypergeometric test), p-value correction = Benjamini-Hochberg. For the graphical representation of the enrichment analysis, ClueGO clustering functionality was used (kappa threshold score for considering or rejecting term-to-term links set to 0.4). Only pathways with an FDR < 0.01 were reported.

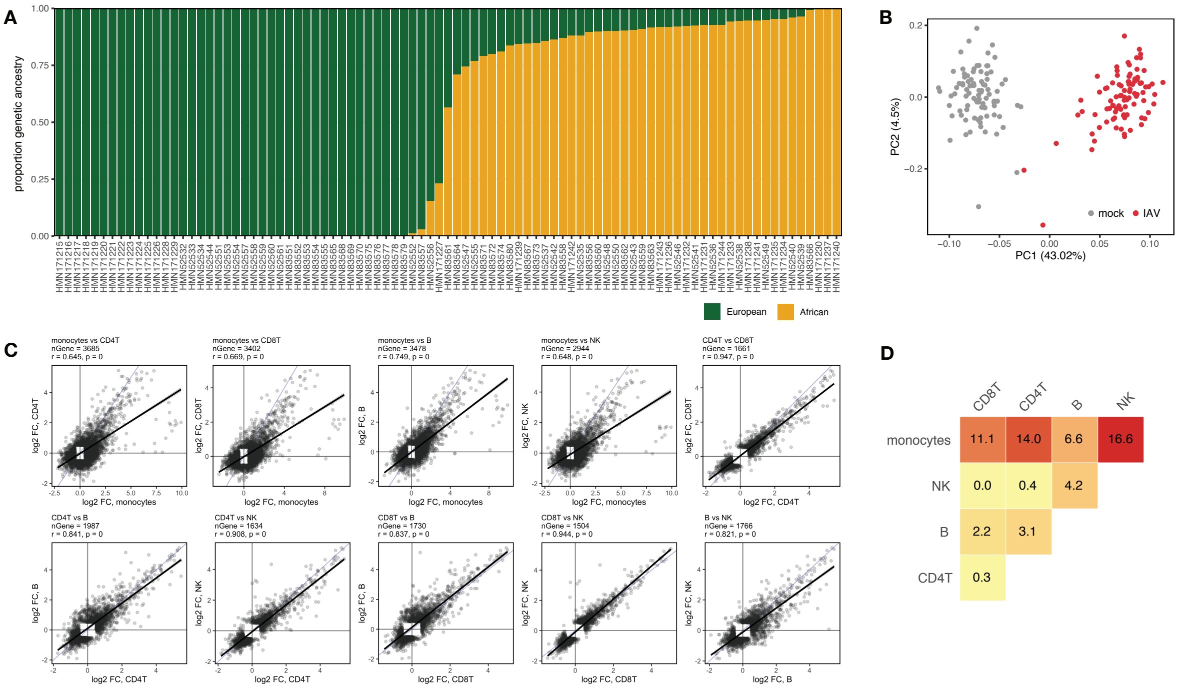

Fig. S1. Overview of samples and global infection effects. **(A)** Quantitative genetic ancestry proportions partitioned into European (green) and African (yellow) components for each individual. **(B)** PCA decomposition of the pseudobulk PBMC expression data in mock-exposed (grey) and IAV-infected (red) samples. PC1 (percent variance explained = 43.02%) separates samples by infection status. **(C)** Pairwise effect size correlations across cell types among genes that are DE ($|\log_2 FC| > 0.5$, FDR < 0.05) upon IAV infection in either of the cell types being compared. Black line shows the best-fit line from a linear model, blue line shows the x = y line. P-values were obtained from linear regression models **(D)** Pairwise comparisons of the percentage of DE ($|\log_2 FC| > 0.5$, FDR < 0.05) genes in both cell types being compared that show discordant effect sizes.

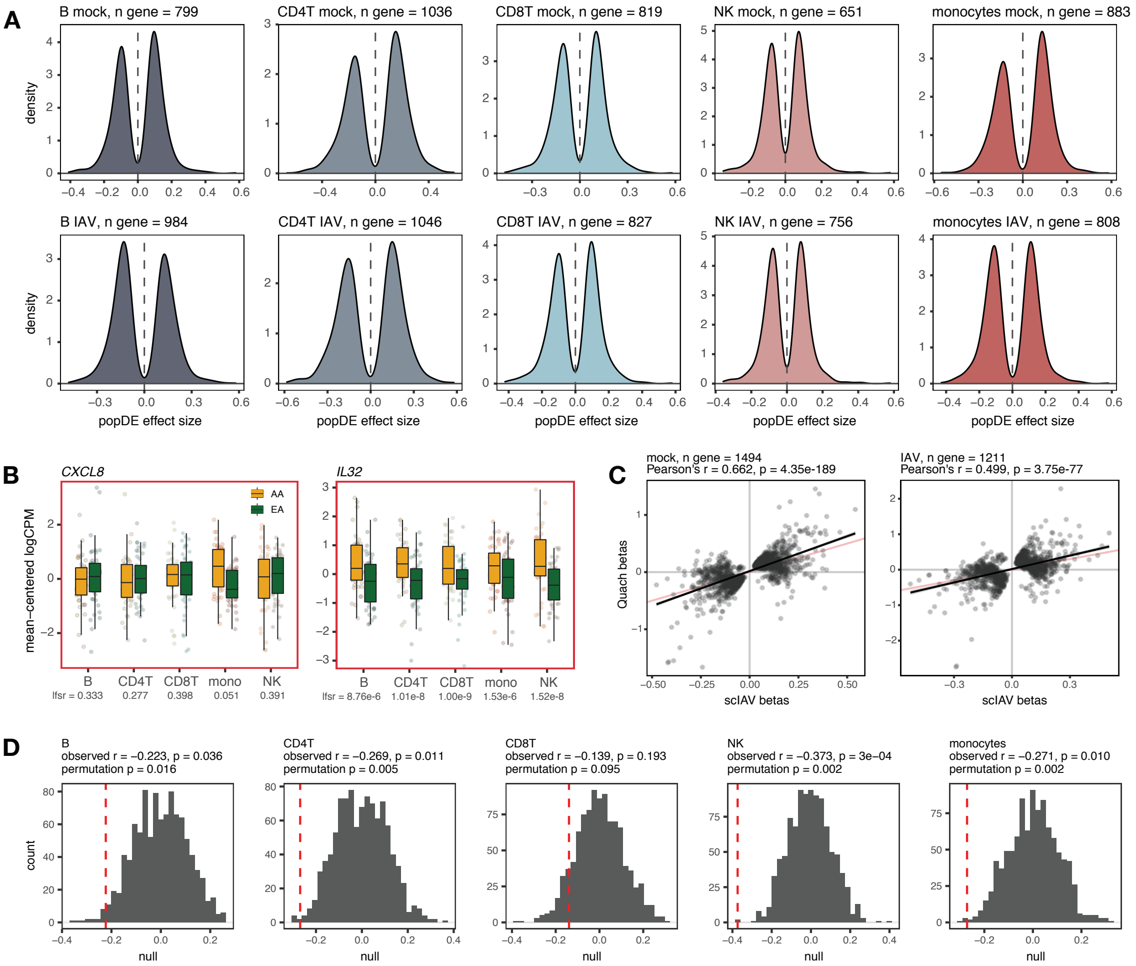

Fig. S2. Population-associated expression patterns. **(A)** PopDE effect size distributions among popDE genes (lfsr < 0.10) detected in each cell type in the mock condition (top row) and following IAV infection (bottom row). Overall, popDE effects are balanced with respect to sign across cell types and conditions. **(B)** Examples of cell type-specific (*CXCL8*, monocytes lfsr = 0.051, lfsr > 0.25 in all other cell types) and shared (*IL32*, lfsr < $8.8\text{x}10^{-6}$ in all cell types) popDE genes (AA in yellow, EA in green) in the IAV-infected condition. **(C)** Correlation between the popDE effect size estimates in monocytes in our single-cell IAV data (x-axis) and the Quach et al. data (*6*) (y-axis) among significant popDE genes (lfsr < 0.10) in our single-cell IAV dataset. Black line shows the best-fit line from a linear model, orange line shows the x = y line. P-values obtained from linear regressions. These correlations are strong, particularly when considering the inherent differences in experimental design between the two studies that are expected to contribute to variation in popDE effect sizes between the datasets (detailed in "Comparison with Quach et al. data"). **(D)** Observed Pearson correlation coefficients between the proportion of African genetic ancestry and IFN score following IAV infection (red dotted line) compared to null expectations when permuting the genetic ancestry vector across individuals (n permutations = 1,000, density distribution in gray) for each cell type. Observed correlation coefficients and p-values from linear regressions are listed at the top of the plot, along with the p-values obtained from permutations. For four out of the five cell types tested (all except CD8^{+}T cells: p = 0.095), the observed Pearson correlation coefficient is significantly lower than most values obtained from permutation (B p-value = 0.016, $CD4^{+}T$ = 0.005, NK = 0.002, monocytes = 0.002). Together, this analysis suggests that the association between genetic ancestry and IFN score in the IAV-infected condition is robust, although less so for $CD8^{+}T$ cells.

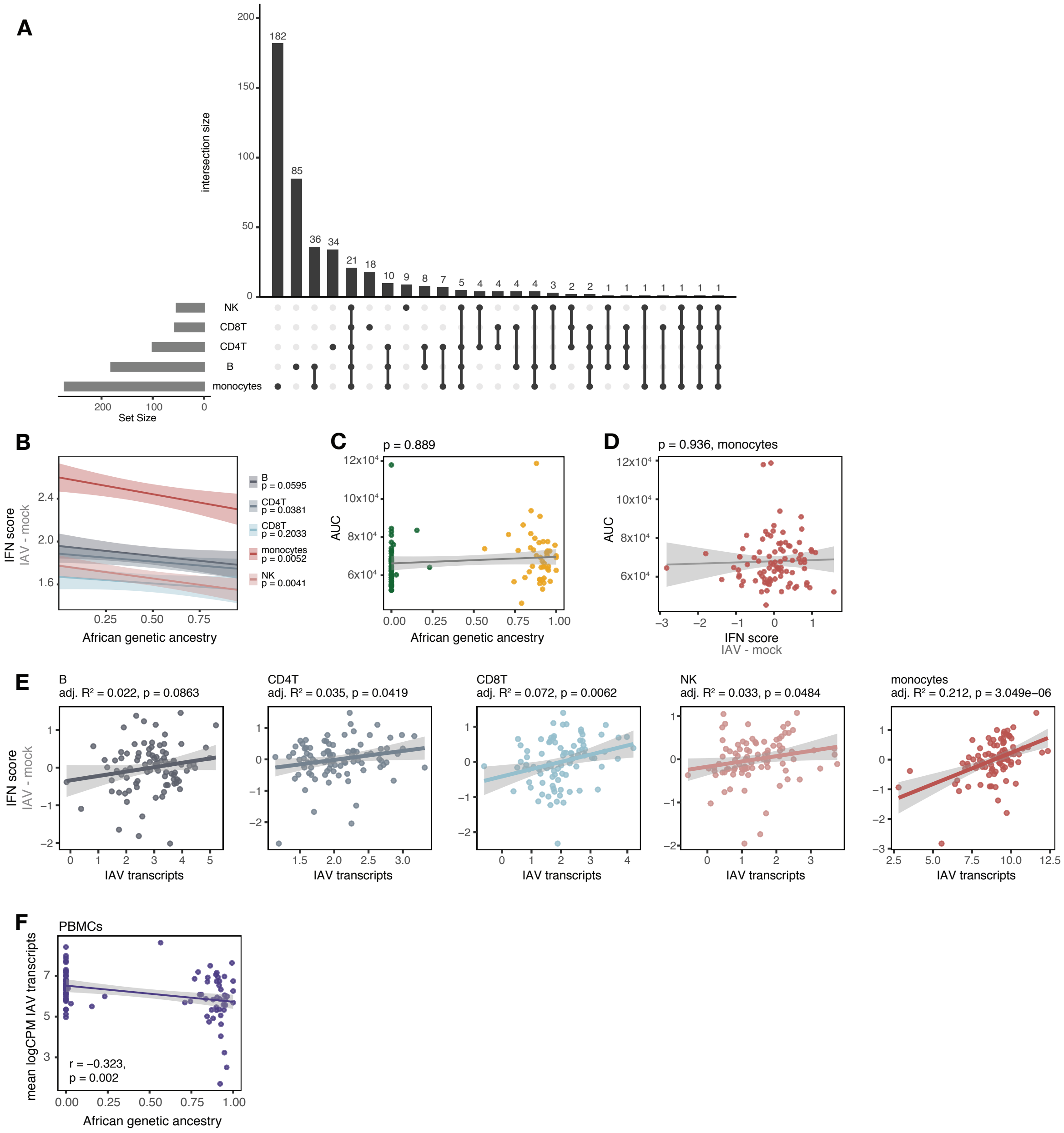

Fig. S3. Population-associated responses to IAV infection. **(A)** Sharing of popDR genes across cell types. **(B)** Correlation between the proportion of African genetic ancestry (x-axis) and IFN response (y-axis) across individuals (mean Pearson's r across cell types = -0.23, Fisher's meta-p = $6x10^{-5}$). **(C)** Correlation between the proportion of African genetic ancestry (x-axis) and baseline levels of IAV-specific serum IgG antibodies (y-axis). Anti-A/Cal/04/09 antibody titers were quantified using 4-fold serial dilutions for each individual's serum (total of eight dilutions per sample). Dilution and ELISA absorbance values were used to generate an area under the curve (AUC, y-axis), which was used to summarize the levels of IAV (A/Cal/04/09)-specific serum IgG antibodies detected in each individual. **(D)** Correlation between IFN response (x-axis) and baseline levels of IAV-specific serum IgG antibodies (y-axis). **(E)** Correlation between IAV transcript expression (x-axis) and IFN response (y-axis) across individuals for each cell type. Higher IAV transcript expression is associated with stronger IFN responses in CD4$^+$ T cells, CD8$^+$ T cells, monocytes, and NK cells, with monocytes showing the strongest relationship (adj. $R^2 = 0.212$, $p = 3.1\times10^{-6}$). **(F)** African genetic ancestry (x-axis) is negatively correlated with IAV transcript expression (y-axis) (Pearson's $r = -0.323$, $p = 0.002$) in PBMCs. To assess the robustness of this association, we tested whether outliers drive the overall correlation between genetic ancestry and intracellular IAV transcripts in the full sample. We 1) excluded individuals with standardized |z-score| values for IAV transcript measurements > 3; and 2) removed individuals with mean logCPM IAV transcripts$_{\text{IAV – mock}}$ < 4. In both cases, the relationship between genetic ancestry and IAV transcript levels was unchanged (z-score approach: Pearson's $r = -0.284$, $p = 0.008$; hard threshold approach: Pearson's $r = -0.258$, $p = 0.016$), suggesting that outliers do not drive the association described in the main text. In (B) - (F), p-values and best-fit slopes were obtained from linear regression models.

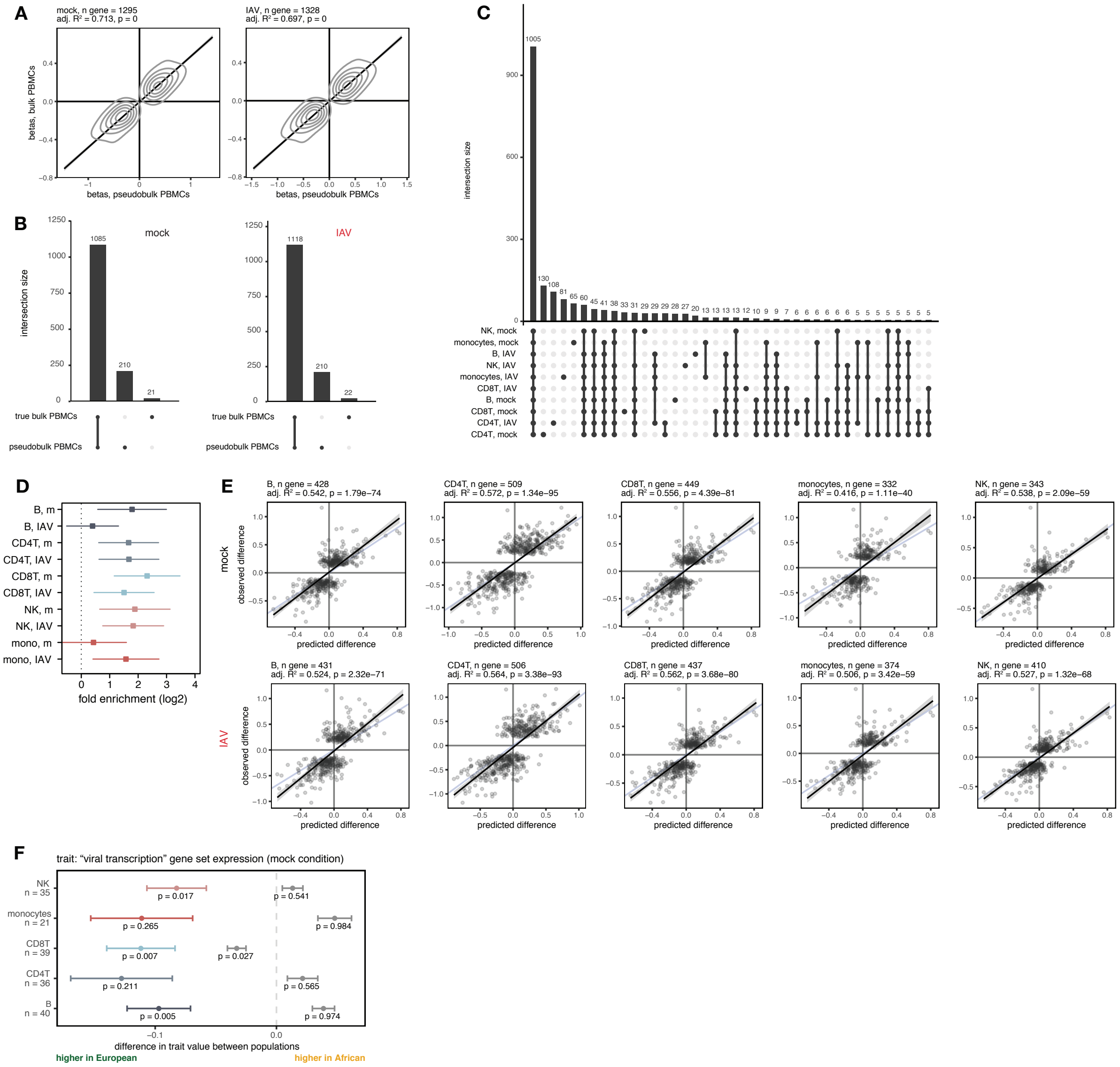

Fig. S4. *Cis*-genetic effects regulate gene expression variation. **(A)** Correlation between the eQTL mapping effect size estimates in the pseudobulk PBMC expression data (x-axis) and in the true bulk PBMC expression data (y-axis) among gene-SNP pairs with an eQTL effect in the pseudobulk data. P-values and best-fit slopes (black line) were obtained from linear regression models. **(B)** Sharing of eGenes in the pseudobulk PBMC expression data and the true bulk PBMC expression data for the mock (left) and IAV (right) conditions. **(C)** Sharing of eGenes across cell types and treatment conditions. **(D)** Enrichment of eGenes (lfsr < 0.10) among IFN-associated popDE genes (lfsr < 0.10 and in the Hallmark IFN-α/IFN-γ response pathway gene sets) identified in each cell type and condition ($\log_2$ fold enrichment with a 95% confidence interval, logistic regression; "m" = mock). **(E)** Correlation of the *cis*-predicted population differences in expression (x-axis) versus the observed population differences in expression (y-axis) among popDE genes with an eQTL across all cell types in the mock-exposed condition (top row) and IAV-infected condition (bottom row). The black line shows the best-fit line from a linear model, and the blue line shows the x = y line. P-values were obtained from linear regression models. **(F)** Example term showing the effect of *cis*-SNP regression. In the mock condition, European-ancestry individuals display higher expression (median observed ancestry-associated difference < 0, colored points ± SE) of the genes belonging to the “viral transcription” term in the observed data. *Cis*-SNP regression (gray points ± SE) reduces this effect.

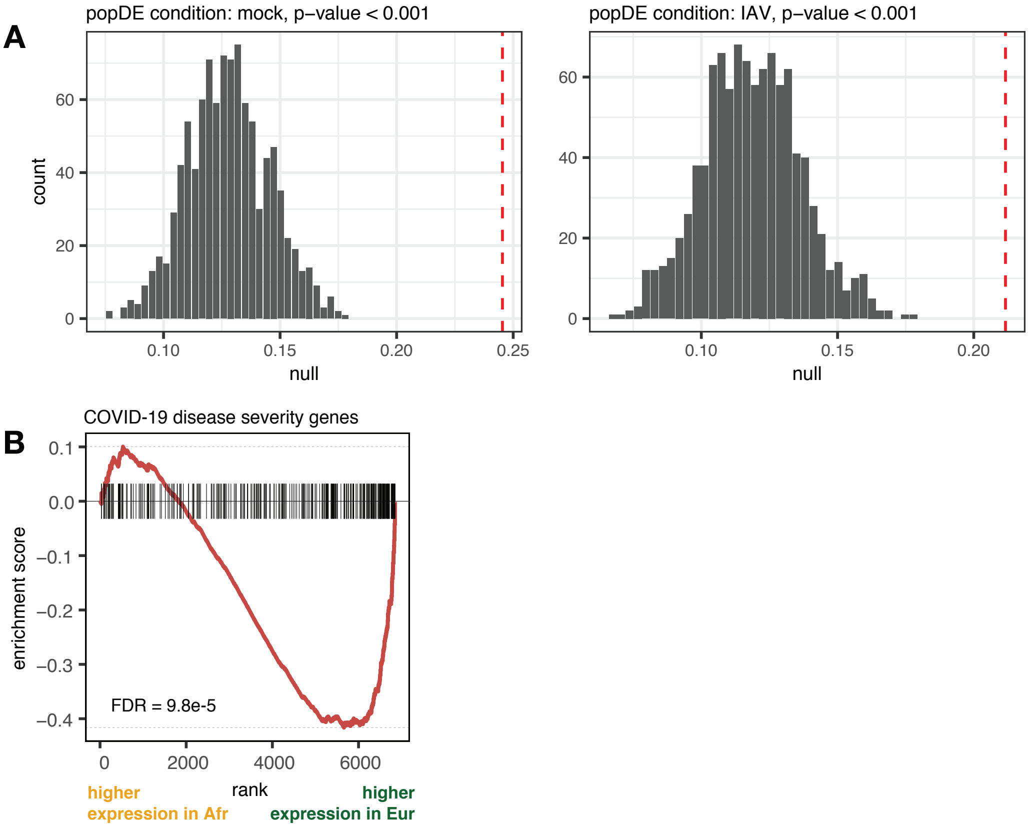

Fig. S5. COVID-19 severity-associated genes are enriched for genes differentially-expressed between populations. **(A)** Proportion of COVID severity-associated genes in monocytes that are popDE in monocytes (red dotted line) in both the mock (left) and IAV-infected (right) conditions compared to random expectation (gray, density distribution) when sampling the same total number of genes 1,000 times from all genes tested in the popDE analysis. **(B)** Barcode enrichment plot of genes positively associated with severity in monocytes, where popDE effect sizes at baseline (mock condition) are ranked from most positive to most negative (x-axis).

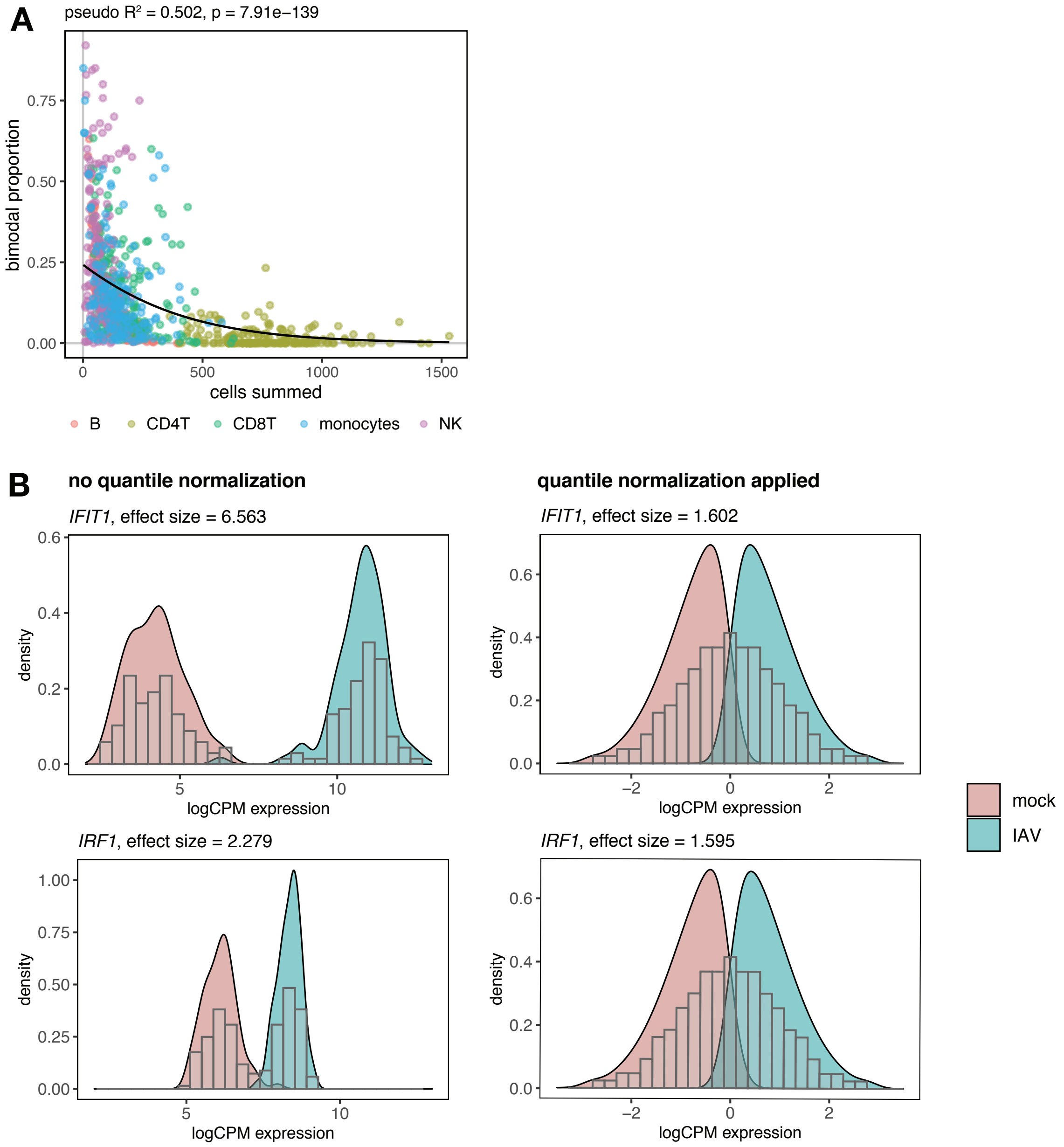

Fig. S6. Bimodal proportion and quantile normalization example. **(A)** Correlation between the number of cells summed to obtain pseudobulk gene expression estimates (x-axis) and the calculated bimodal proportion (y-axis) per sample for each cell type. Best-fit line (black) and p-value were obtained from a beta regression model. **(B)** Both *IFIT1* and *IRF1* display non-overlapping expression distributions between conditions prior to quantile normalization (left), although *IFIT1* shows a much stronger upregulation following IAV infection compared to *IRF1*. After quantile normalization (right), the transformed gene expression distributions for *IFIT1* and *IRF1* are very similar, and therefore, the estimated infection effect sizes are virtually equivalent.

Table S1. Sample meta data. Description of cohort samples, including demographic information about donors, technical batch/variable information, and experimental information. Also included is a summary of the exome-sequencing results and kernel density estimation bandwidth parameters for the bimodal proportion calculations.

Table S2. Global infection effects. IAV infection differential expression effect for each gene in each cell type is reported, including effect size estimates, p-values, FDRs (Benjamini-Hochberg and permutation-based), and t-statistics.

Table S3. Global infection DE enrichments. Gene ontology enrichments for the differential expression effects are reported for each cell type. Term-specific p-values, FDRs (Benjamini-Hochberg-adjusted), enrichment scores, and leading edge gene sets are included.

Table S4. Ranked specificity scores and enrichments. Specificity score calculated for each gene. Also included are the gene ontology enrichments for ranked specificity scores. Term-specific p-values, FDRs (Benjamini-Hochberg-adjusted), and enrichment scores are reported.

Table S5. PopDE effects. Population differential expression effect for each gene in each cell type-condition combination is reported, including local false sign rate (lfsr), effect size estimate, and standard deviation of the effect size estimate.

Table S6. PopDE effect enrichments. Hallmark enrichments for the population differential expression effects are reported for each cell type-condition combination. Term-specific p-values, FDRs (Benjamini-Hochberg-adjusted), enrichment scores, and leading edge gene sets are included.

Table S7. PopDR effects. Population differential response effect for each gene in each cell type is reported, including local false sign rate (lfsr), effect size estimate, and standard deviation of the effect size estimate.

Table S8. eQTL effects. eQTL effect for the lead *cis*-SNP per gene is reported across cell types, including local false sign rate (lfsr), effect size estimate, and standard deviation of the effect size estimate.

Table S9. GO enrichments for popDE genes with an eQTL. Gene ontology enrichments for the popDE genes with evidence of an eQTL. Term-specific p-values and FDRs (Benjamini-Hochberg-adjusted) are included.

Table S10. WOS effects. Results from re-analysis of the Su et al. data (*21*). WOS effect (i.e. COVID-19 severity score association) for each gene is reported across cell types, including local false sign rate (lfsr), effect size estimate, and standard deviation of the effect size estimate.

Tables S1 – S10 are posted online as Excel files.

Table S11. Gene expression principal components (PCs) regressed in the eQTL analysis.

Cell type	**Regressed PCs (mock)**	**Regressed PCs (IAV)**	**No. genes under 0.10 FDR (mock)**	**No. genes under 0.10 FDR (IAV)**
$CD4^+$ T	1 to 4	1 to 2	1377	1176
B	1 to 6	1 to 3	152	196
NK	1 to 2	1 to 2	68	76
monocytes	1 to 10	1 to 7	265	251
$CD8^+$ T	1 to 6	1 to 4	204	178
pseudobulk PBMCs	1 to 6	1 to 3	2095	1809
true bulk PBMCs	1 to 2	1 to 3	100	105