



**HAL**  
open science

## Single-Cell and Bulk RNA-Sequencing Reveal Differences in Monocyte Susceptibility to Influenza A Virus Infection Between Africans and Europeans

Mary B O’neill, H el ene Quach, Julien Pothlichet, Yann Aquino, Aur elie Bisiaux, Nora Zidane, Matthieu Deschamps, Valentina Libri, Milena Hasan, Shen-Ying Zhang, et al.

► **To cite this version:**

Mary B O’neill, H el ene Quach, Julien Pothlichet, Yann Aquino, Aur elie Bisiaux, et al.. Single-Cell and Bulk RNA-Sequencing Reveal Differences in Monocyte Susceptibility to Influenza A Virus Infection Between Africans and Europeans. *Frontiers in Immunology*, 2021, 12, 10.3389/fimmu.2021.768189 . hal-03499159

**HAL Id: hal-03499159**

**<https://hal.sorbonne-universite.fr/hal-03499159>**

Submitted on 21 Dec 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d’enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons Attribution 4.0 International License



# Single-Cell and Bulk RNA-Sequencing Reveal Differences in Monocyte Susceptibility to Influenza A Virus Infection Between Africans and Europeans

## OPEN ACCESS

### Edited by:

Petter Brodin,  
Science for Life Laboratory  
(SciLifeLab), Sweden

### Reviewed by:

Karthik Shekhar,  
University of California, Berkeley,  
United States  
Trine Hyrup Mogensen,  
Aarhus University, Denmark

### \*Correspondence:

Lluis Quintana-Murci  
quintana@pasteur.fr

†These authors have contributed  
equally to this work and  
share senior authorship

### Specialty section:

This article was submitted to  
Systems Immunology,  
a section of the journal  
Frontiers in Immunology

**Received:** 15 September 2021

**Accepted:** 27 October 2021

**Published:** 29 November 2021

### Citation:

O'Neill MB, Quach H, Pothlichet J,  
Aquino Y, Bisiaux A, Zidane N,  
Deschamps M, Libri V, Hasan M,  
Zhang S-Y, Zhang Q, Matuozzo D,  
Cobat A, Abel L, Casanova J-L,  
Naffakh N, Rotival M and Quintana-  
Murci L (2021) Single-Cell and  
Bulk RNA Sequencing Reveal  
Differences in Monocyte Susceptibility  
to Influenza A Virus Infection  
Between Africans and Europeans.  
*Front. Immunol.* 12:768189.  
doi: 10.3389/fimmu.2021.768189

Mary B. O'Neill<sup>1</sup>, Hélène Quach<sup>2</sup>, Julien Pothlichet<sup>3</sup>, Yann Aquino<sup>1,4</sup>, Aurélie Bisiaux<sup>1</sup>, Nora Zidane<sup>5</sup>, Matthieu Deschamps<sup>1</sup>, Valentina Libri<sup>6</sup>, Milena Hasan<sup>6</sup>, Shen-Ying Zhang<sup>7,8,9</sup>, Qian Zhang<sup>7,8,9</sup>, Daniela Matuozzo<sup>8,9</sup>, Aurélie Cobat<sup>8,9</sup>, Laurent Abej<sup>7,8,9</sup>, Jean-Laurent Casanova<sup>7,8,9,10</sup>, Nadia Naffakh<sup>11</sup>, Maxime Rotival<sup>1†</sup> and Lluis Quintana-Murci<sup>1,12\*†</sup>

<sup>1</sup> Human Evolutionary Genetics Unit, Institut Pasteur, UMR 2000, Centre National de la Recherche Scientifique (CNRS), Paris, France, <sup>2</sup> Muséum National d'Histoire Naturelle, UMR7206, Centre National de la Recherche Scientifique (CNRS), Université de Paris, Paris, France, <sup>3</sup> DIACCURATE, Paris, France, <sup>4</sup> Sorbonne Université, Collège doctoral, Paris, France, <sup>5</sup> Biodiversity and Epidemiology of Bacterial Pathogens Unit, Institut Pasteur, Paris, France, <sup>6</sup> Cytometry and Biomarkers UTechS, Institut Pasteur, Paris, France, <sup>7</sup> St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York, NY, United States, <sup>8</sup> Laboratory of Human Genetics of Infectious Diseases, Necker Hospital for Sick Children, INSERM UMR 1163, Necker Hospital for Sick Children, Paris, France, <sup>9</sup> Imagine Institute, Paris University, Paris, France, <sup>10</sup> Howard Hughes Medical Institute, New York, NY, United States, <sup>11</sup> RNA Biology of Influenza Virus Unit, Institut Pasteur, Paris, France, <sup>12</sup> Chair of Human Genomics and Evolution, Collège de France, Paris, France

There is considerable inter-individual and inter-population variability in response to viruses. The potential of monocytes to elicit type-I interferon responses has attracted attention to their role in viral infections. Here, we use single-cell RNA-sequencing to characterize the role of cellular heterogeneity in human variation of monocyte responses to influenza A virus (IAV) exposure. We show widespread inter-individual variability in the percentage of IAV-infected monocytes. Notably, individuals with high cellular susceptibility to IAV are characterized by a lower activation at basal state of an IRF/STAT-induced transcriptional network, which includes antiviral genes such as *IFITM3*, *MX1* and *OAS3*. Upon IAV challenge, we find that cells escaping viral infection display increased mRNA expression of type-I interferon stimulated genes and decreased expression of ribosomal genes, relative to both infected cells and those never exposed to IAV. We also uncover a stronger resistance of *CD16*<sup>+</sup> monocytes to IAV infection, together with *CD16*<sup>+</sup>-specific mRNA expression of *IL6* and *TNF* in response to IAV. Finally, using flow cytometry and bulk RNA-sequencing across 200 individuals of African and European ancestry, we observe a higher number of *CD16*<sup>+</sup> monocytes and lower susceptibility to IAV infection among monocytes from individuals of African-descent. Based on these data, we hypothesize that higher basal monocyte activation, driven by environmental factors and/or weak-effect genetic variants, underlies the lower cellular susceptibility to IAV

infection of individuals of African ancestry relative to those of European ancestry. Further studies are now required to investigate how such cellular differences in IAV susceptibility translate into population differences in clinical outcomes and susceptibility to severe influenza.

**Keywords: Monocytes, single-cell 'omics, transcriptomics, ancestry, population, influenza virus**

## INTRODUCTION

Respiratory viruses with pandemic potential pose enormous health and economic impacts on the human population. In the last century, we have witnessed outbreaks of several coronaviruses, including SARS-CoV-2, SARS-CoV-1 and MERS, and a number of avian and swine influenza A viruses (IAV). A particularly harrowing and shared feature of these pandemics are the sudden deaths of otherwise healthy individuals (1). A hyperinflammatory state characterized by high levels of inflammatory cytokines, often referred to as a 'cytokine storm' (2, 3), has emerged as a hallmark of these severe viral infections. While still controversial, there is increasing evidence to suggest that the mononuclear phagocyte system is an important immunological determinant of this phenotype (4–6). Upon viral infection, sentinel cells such as lung-resident macrophages trigger complex signaling cascades that recruit leukocytes to the site of infection, among them monocytes. These infiltrating monocytes differentiate into monocyte-derived dendritic cells or macrophages, enabling viral clearance through the induction of the adaptive response, and help replenish the pool of tissue-resident alveolar macrophages (4, 7).

In humans, circulating monocytes are divided into classical (~80%), intermediate (~15%), and nonclassical (~5%) subsets, based on surface receptor expression of the cluster-determinant antigens CD14 and CD16 (8). While nonclassical monocytes (CD14<sup>+</sup>CD16<sup>++</sup>) are long-lived and 'patrol' healthy tissues through long-range crawling on the endothelium, classical (CD14<sup>++</sup>CD16<sup>-</sup>) and intermediate (CD14<sup>++</sup>CD16<sup>+</sup>) monocytes are recruited to the lung in response to viral infection, where they secrete inflammatory cytokines and chemokines, as well as type I interferons (IFNs) (7, 9–11). In most individuals, recruited cells help clear infection despite being susceptible to infection themselves (12, 13); yet, in some individuals, a dysfunctional immune response occurs resulting in widespread lung inflammation. Whether monocyte subsets behave differently upon viral exposure, and how direct viral sensing and exposure to secreted cytokines shape monocyte activation and differentiation are not well understood.

Variation in blood composition and cellular proportions have been shown to be one of the main factors underlying transcriptional variation in immune genes across individuals (14), with these proportions being influenced by both genetic and non-heritable factors (15–17). Recently, we characterized the genetic architecture of transcriptional responses of primary monocytes from 200 individuals of African and European ancestry to *ex vivo* challenge with viral stimuli (18). In this model, where we were able to control for viral determinants of disease (i.e. dose and strain), we reported marked inter- and intra-

population differences in transcriptional responses to IAV. While our analyses revealed numerous *cis*-expression quantitative trait loci (18), genetic variants could only account for a small fraction of expression variation, in line with other studies (14, 19).

Here, we implemented single-cell RNA-sequencing (scRNA-seq) on human primary monocytes exposed to IAV to investigate (i) the effects of direct viral infection versus activation by exposure to secreted cytokines, (ii) the subset-specific responses of monocytes to viral challenge, and (iii) the extent of inter-individual and between-population variation in the proportions of monocyte subsets and the degree of monocyte susceptibility to IAV infection. Our study reveals a profound reprogramming of monocyte transcriptomes upon viral infection and shows a proinflammatory role of CD16<sup>+</sup> monocytes following IAV challenge. Furthermore, it highlights that African-ancestry individuals are characterized by both a higher frequency of CD16<sup>+</sup> monocytes and a generally lower susceptibility of their monocytes to IAV infection. Based on these results, we propose that population differences in the composition of circulating monocytes and their susceptibility to infection may contribute to the higher severity of IAV infections reported among African-ancestry individuals.

## RESULTS

### Using scRNA-Seq to Investigate Cellular Heterogeneity

To investigate the role of cellular heterogeneity in driving immune variability across individuals, we performed a time-course experiment where we monitored the CD14<sup>+</sup> fraction of peripheral blood mononuclear cells (PBMCs) from eight donors, both in the presence and absence of viral challenge. To maximize inter-individual variability, we chose individuals from two distinct ancestries whose cells demonstrated extreme responses to viral stimuli in a previous bulk RNA-seq experiment (18). Droplet-based scRNA-seq was performed on monocytes from all eight donors immediately before infection initiation (T<sub>0</sub>), as well as at 2 (T<sub>2</sub>), 4 (T<sub>4</sub>), 6 (T<sub>6</sub>), and 8 (T<sub>8</sub>) hours post challenge with A/USSR/90/1977(H1N1) at a multiplicity of infection (MOI) equal to 1 (IAV-challenged) and mock infection (non-infected). To mitigate batch effects, we pooled IAV-challenged and non-infected cells from distinct donors in each library, assigning cells to their condition in silico *via* genetic barcoding (20). After stringent quality control where we removed low-quality, dying, and contaminants of the CD14<sup>+</sup> monocyte isolation, our final dataset contained 88,559 high-quality cells, among which we

predicted >99% monocyte purity at  $T_0$  (**Figure 1A**; **Supplementary Figures 1** and **2**). At later time points, a substantial fraction of non-infected cells (up to 70% at  $T_8$ ) were predicted to be macrophage-like, indicating monocyte differentiation over the course of the experiment. For clarity, we refer to cells as monocytes at  $T_0$  and as monocyte-derived cells from  $T_2$ - $T_8$ .

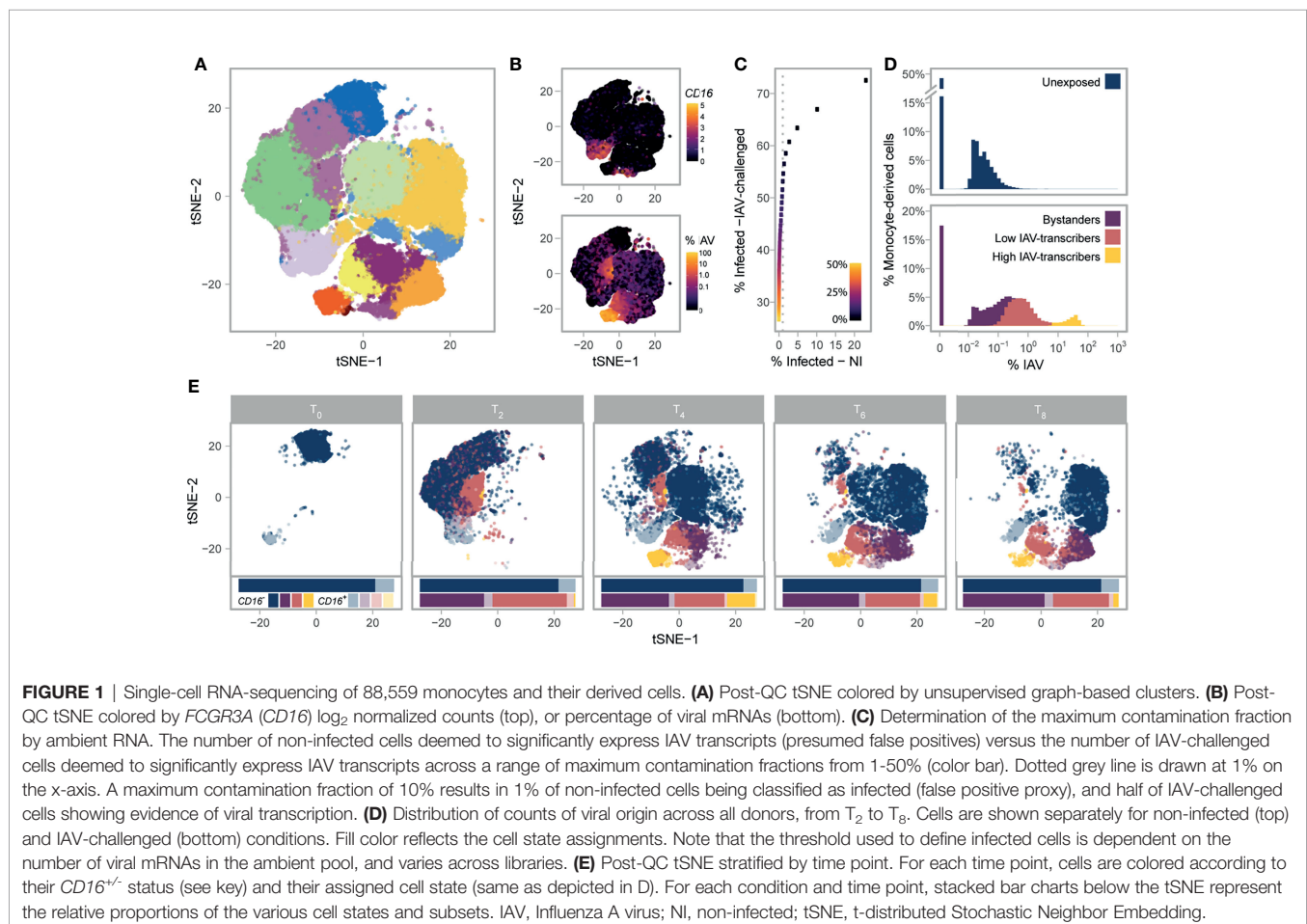
## Stable *FCGR3A* Expression Distinguishes Monocyte Subsets Over Time

We next sought to characterize each cell by its mRNA expression of the canonical monocyte markers, *CD14* and *CD16*, given that much of the structure in our data was associated with *FCGR3A* (aka *CD16*) mRNA expression. In droplet-based scRNA-seq, encapsulation of ambient mRNAs emanating from dying cells can occur during library preparation leading to spurious mRNA detection (21). We thus used a statistical framework to test whether *CD14* and *CD16* were expressed at a level significantly higher than expected when accounting for potential contamination from the ambient pool (*Methods*). Despite having been positively selected for the *CD14* antigen, only 32.4% of monocytes significantly expressed *CD14* at  $T_0$ ; this percentage further decreased at later time points and remained

<15% across all time points and conditions (average 6.4% s.d.: 5.0%, **Supplementary Figures 3A–C**). On the other hand, 12.1% of monocytes significantly expressed *FCGR3A* (*CD16*) (referred to as *CD16*<sup>+</sup>) at  $T_0$ , this marker proving much more stable across conditions and time points (9.3% of *CD16*<sup>+</sup> cells on average, s.d.: 1.8%, **Figure 1B** and **Supplementary Figures 3D–F**). While we deciphered classical, intermediate, and nonclassical monocytes subsets at  $T_0$  (**Supplementary Note 1**; **Supplementary Figure 4** and **Supplementary Data 1**), we focus on the simpler distinction of *CD16*<sup>-</sup> and *CD16*<sup>+</sup> subsets given that positive-selection for monocytes does not capture the entire nonclassical population and that we were unable to distinguish the intermediate and nonclassical subsets after  $T_0$ .

## Functional Features of Monocyte Subsets Are Conserved Upon Manipulation

To assess how transcriptional profiles of *CD16*<sup>-</sup> and *CD16*<sup>+</sup> monocytes and their derived-cells differ, we focused on the 5,681 genes expressed with a normalized log<sub>2</sub> count > 0.1 in at least one condition, time point, and subset (**Supplementary Data 2A**). We found that the log<sub>2</sub> fold change (log<sub>2</sub>FC) in gene expression between *CD16*<sup>+/-</sup> subsets remained relatively stable over the course of the experiment (Pearson r between time points >0.42





and  $>0.52$  for the non-infected and IAV-challenged conditions respectively,  $p$ -values  $<2.2 \times 10^{-16}$ ; **Supplementary Figure 5A**), and differentially expressed genes between  $CD16^{+/-}$  subsets were largely the same across conditions (Pearson  $r = 0.92$ ,  $p$ -value  $< 2.2 \times 10^{-16}$ ; **Supplementary Figure 5B**). We thus searched for genes that were consistently differentially expressed between  $CD16^{+}$  and  $CD16^{-}$  cells across all time points (including  $T_0$ ), conditions, and donors. We identified 266 genes over-expressed ( $\log_2FC > 0.2$ ,  $FDR < 1\%$ ) in  $CD16^{+}$  cells relative to  $CD16^{-}$  cells, and 389 genes that showed the opposite pattern, and performed a GO-term enrichment analysis on these genes (**Supplementary Data 2B**). Consistent with previous reports (22–24),  $CD16^{-}$  subsets were characterized by high expression of several proinflammatory *S100 Calcium Binding Proteins* (*S100A12*, *S100A9*, and *S100A8*), contributing to a sizable GO-term enrichment in the defense response to fungus pathway (GO:0050832: OR=41.3,  $FDR=4.9 \times 10^{-4}$ ), while  $CD16^{+}$  subsets were characterized by high expression of Fc-gamma receptor signaling pathway genes (GO:0038096: OR=8.7,  $FDR=6.2 \times 10^{-6}$ ). Notably,  $CD16^{+}$  subsets over-expressed several type I IFN stimulated genes (ISGs) relative to  $CD16^{-}$  subsets (e.g. GO:0071357: OR=5.3,  $FDR=2.6 \times 10^{-3}$ ), including the well-known viral restriction factors *IFITM3* and *OAS1*. Collectively, these results demonstrate  $CD16$  is a reliable marker at the mRNA level and that  $CD16^{+/-}$  monocyte subsets maintain functional differences upon manipulation.

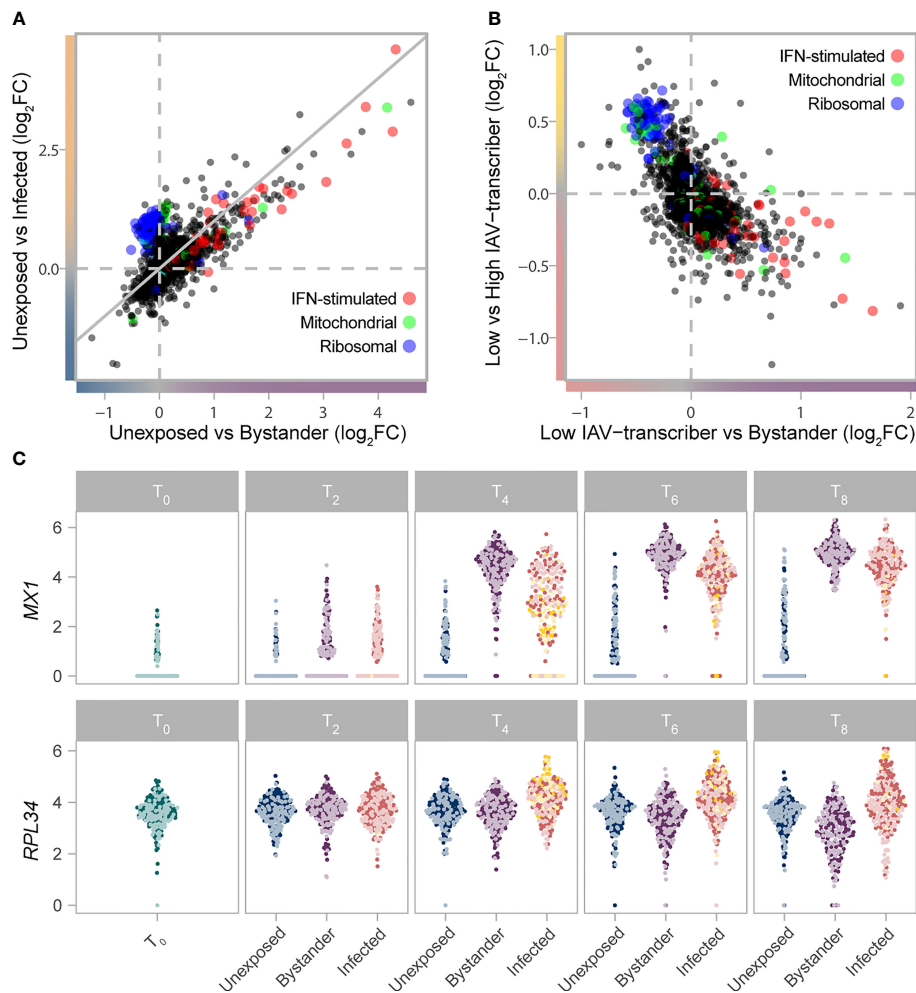
### scRNA-Seq Highlights Heterogeneity in Monocyte Susceptibility and Viral Transcription

Using the presence of IAV transcripts as a proxy for infection (**Figure 1B**), we next sought to distinguish cells that were successfully infected from those that were not. Among monocyte-derived cells that were exposed to IAV, we found that 50.3% expressed IAV transcripts above ambient levels when allowing up to 10% of mRNAs to come from the ambient pool. In contrast, less than 1% of non-infected cells showed evidence of viral transcription, supporting the validity of the threshold used to detect IAV expressing cells (**Figure 1C**). We deemed cells with statistical evidence for expression of IAV transcripts from the IAV-challenged condition as ‘infected’, while the remaining cells from this condition were considered as ‘bystanders’, as these either did not come into contact with the virus or were able to fully repress viral mRNA transcription. When comparing the percentage of infected cells between subsets, we noticed that  $CD16^{+}$  cells were slightly less likely to be infected than  $CD16^{-}$  cells (42.3% sd: 4.0% for  $CD16^{+}$  relative to 49.4% sd: 5.4% for  $CD16^{-}$ , generalized linear model with  $CD16^{+/-}$  status, donor, and time point as covariates,  $p$ -value=0.006), possibly related to the higher expression of ISGs observed in this subset (**Supplementary Data 2A, B**). We further confirmed experimentally that intermediate and nonclassical ( $CD16^{++}$ ) monocytes display increased resistance to IAV challenge by monitoring intracellular IAV nucleoprotein by flow cytometry in PBMCs challenged with another H1N1 strain (**Supplementary Figure 6**).

We observed that the proportions of viral mRNAs among infected cells were bimodally distributed and largely varied between the clusters identified in our unsupervised analysis (**Figure 1D**). We used a Gaussian mixture model to locate the two modes of the distribution and further sub-classify infected cells into those with lower IAV mRNA levels ( $<1$ -6%) and those with higher IAV mRNA levels (6-83%); while viral mRNA levels are dictated by both the rate of transcription and degradation, for simplicity we refer to these infected cell states as ‘low IAV-transcribers’ and ‘high IAV-transcribers’, respectively. The proportions of infected cells among individuals remained largely unchanged over the course of the experiment; however, high IAV-transcribers were virtually absent at 2h ( $<2\%$  of infected cells), peaked to  $\sim 36\%$  of IAV-infected cells at 4h, and decreased to 8.5% by 8h, suggesting that high-IAV transcribers represent a transient state of IAV-infection preceding IAV-induced apoptosis (**Figure 1E**). These results reveal profound heterogeneity in monocyte susceptibility and subsequent viral transcription upon IAV challenge.

### Interplay of Cytokine and Ribosome Networks Drive Cell States Upon Infection

To characterize host transcriptional responses over time, we next subsampled each subset ( $CD16^{-}/CD16^{+}$ ), cell state (unexposed, bystander, infected), and time point in our scRNA-seq data to a uniform number of cells to avoid biases emanating from differences in sample sizes. Limited by the number of  $CD16^{+}$  high IAV-transcribing cells, we randomly sampled 100 cells from each subgroup, while ensuring representation of all donors. We then focused on the 6,669 host genes with average  $\log_2$  normalized count  $>0.1$  in at least one subgroup (**Supplementary Data 3A**). Overall,  $CD16^{-}$  and  $CD16^{+}$  subsets behaved similarly upon stimulation with changes in gene expression between cell states being strongly correlated among subsets (Pearson  $r=0.83$ - $0.95$ ,  $p$ -values  $<2.2 \times 10^{-16}$ ; **Supplementary Figure 7**). GO term enrichment analyses of shared responses ( $FDR < 1\%$  &  $\log_2FC > 0.2$  in same direction in both subsets) uncovered several functional categories interacting to shape the activation state of cells (**Figure 2A; Supplementary Data 3B**). Both bystander and infected cells showed increased mRNA expression of genes involved in antigen processing and presentation *via* class I MHC (GO:0019885, OR=53.7,  $FDR=2.0 \times 10^{-6}$ ) and ISGs (GO:0034340, OR=14.8,  $FDR=3.3 \times 10^{-20}$ ). Yet, bystander cells showed increased mRNA expression of ISGs and defense response to virus pathways relative to infected cells (GO:0034340, OR=13.4,  $FDR=4.4 \times 10^{-7}$ ; GO:0051607, OR=9.0,  $FDR=2.1 \times 10^{-7}$ ), while infected cells displayed higher mRNA expression of mitochondrial (GO:0005743, OR=4.7,  $FDR=3.3 \times 10^{-3}$ ) and ribosomal genes (GO:0005840, OR=117,  $FDR=1.0 \times 10^{-78}$ ). Notably, type-I IFN genes themselves tended to be preferentially expressed by infected cells (e.g.  $\log_2$  normalized count at 6h for *IFNB1*  $\sim 0.12/0.29$  in  $CD16^{-}$  and  $CD16^{+}$  subsets, respectively, vs  $<0.01$  for bystander cells of both subsets), although this difference was only barely significant in our setting ( $FDR=0.03$ ), likely due to the highly transient nature of IFN expression.



**FIGURE 2** | Gradient of mRNA expression from ribosomal and IFN-stimulated genes separates bystander and infected cells. **(A)** Transcriptional responses of cells upon IAV challenge ( $T_2$ – $T_8$ ) highlight the interplay between IFN-stimulated (GO:0034340), ribosomal (GO:0005840), and mitochondrial (GO:0005743) genes. The  $\log_2$ FC change in gene expression between unexposed and bystander cells is plotted on the x-axis, while the  $\log_2$ FC change in gene expression between unexposed and infected cells is plotted on the y-axis. Values are plotted based on a meta-analysis across time points and subsets, of a subsampled dataset with balanced representation of all donors. **(B)** The interplay between IFN-stimulated (GO:0034340), ribosomal (GO:0005840), and mitochondrial (GO:0005743) genes among cells exposed to IAV. The  $\log_2$ FC change in gene expression between low IAV-transcribing infected and bystander cells is plotted on the x-axis, while the  $\log_2$ FC change in gene expression between low IAV-transcribing infected and high IAV-transcribing infected cells is plotted on the y-axis. Values are plotted based on a meta-analysis across monocyte subsets at  $T_4$ . **(C)** mRNA expression levels of representative IFN-stimulated (*MX1*) and ribosomal (*RPL34*) genes across the subsampled dataset. Colors reflect the cell state and subset assignment depicted in **Figures 1D, E**. IFN, Interferon; FC, fold change.

Among infected cells, ribosomal genes showed higher activity among high IAV-transcribing cells relative to low IAV-transcribing cells (**Figure 2B**, comparison only made at  $T_4$  due to sample size constraints, e.g. GO:0019083, OR=137, FDR=6.1x10<sup>-65</sup>). This observation is consistent with the notion that the expression of viral proteins is dependent on cellular ribosomes, with recent data suggesting that IAVs do not induce a global shut-off of cellular translation but rather a reshaping of the translation landscape (25–27). Likewise, among bystander cells, numerous ribosomal genes were downregulated at later time points relative to unexposed cells (**Figures 2A, C**; GO:0019083, OR=5.3, FDR=4.2x10<sup>-6</sup>), suggesting that repression of ribosomal subunits plays an active role in limiting viral replication.

Collectively, these results suggest that expression of ISGs and ribosomal genes interact to shape cell states upon IAV challenge.

### Increased IRF and STAT Activity Drives Stronger Antiviral Response

Despite qualitatively similar responses to infection between  $CD16^-/CD16^+$  subsets (**Supplementary Figure 7**), we hypothesized that subtle differences in the intensity of such responses might contribute to the increased resistance of  $CD16^+$  cells to infection. We thus performed an interaction test on the subsampled scRNA-seq data, and searched for genes for which transcriptional response upon IAV challenge differed between  $CD16^-$  and  $CD16^+$  subsets in either infected and/or bystander

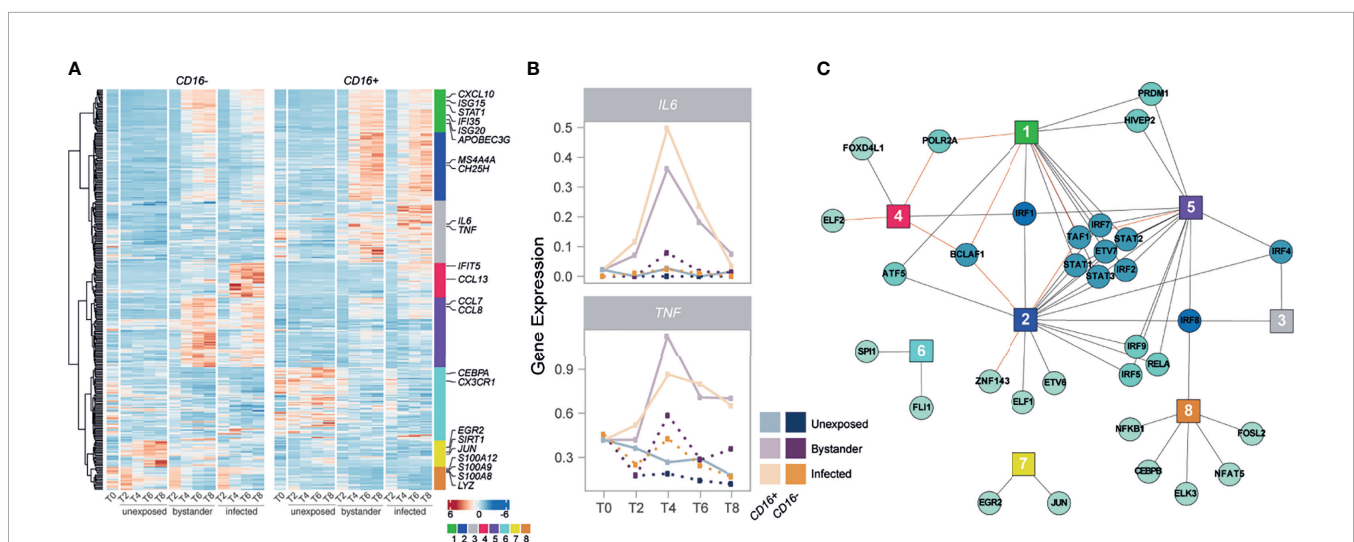
cells (**Supplementary Figures 7A, B; Supplementary Data 3A**). At  $FDR \leq 1\%$ , we identified a total of 335 such genes, of which 98 differed between subsets only in bystander cells, 144 only in infected cells, and 93 in both. Hierarchical clustering highlighted eight major patterns of transcriptional responses (modules) among the 335 genes, several of which were associated with specific biological functions (**Figure 3A; Supplementary Data 3C, D**). Notably, module 1 (green) was enriched for genes in the antiviral response pathway (GO:0051607,  $OR=23.2$ ,  $FDR=5.43 \times 10^{-7}$ ) and displayed a stronger response in infected  $CD16^+$  cells relative to  $CD16^-$  infected cells. Of additional interest was the transient  $CD16^+$ -specific transcription of the inflammatory cytokine genes *IL6* and *TNF*, following viral challenge (**Figure 3B**). We also found that several genes involved in the regulation and production of IL-6 and  $TNF\alpha$  were over-expressed in  $CD16^+$  subsets at all time points and conditions (**Supplementary Data 2B**), but only see active transcription of the cytokines upon viral exposure. These results reveal the strong antiviral and inflammatory potential of  $CD16^+$  relative to  $CD16^-$  monocytes in response to viral infection (28).

We next sought to characterize the regulatory architecture underlying the 335 genes whose transcriptional response to IAV challenge differed between monocyte subsets. Using SCENIC (29), we identified 113 high-confidence gene regulatory networks, or 'regulons', which were active in non-infected and/or IAV-challenged cells, each composed of a transcription factor (TF) and a set of predicted targets (genes). We used these 113 regulons to search for an enrichment/depletion of TF targets among the eight modules of genes displaying subset-specific response to infection (**Supplementary Data 3E**). Among modules associated with an increased expression in cells

exposed to IAV (modules 1-5), we observed a widespread over-representation of targets of IFN regulatory factors (IRFs) and signal transducing and activators of transcription (STATs) (**Figure 3C**), reinforcing the central role of the IFN response upon IAV challenge. Interestingly, several of these factors displayed subset-specific activity themselves in response to IAV (*IRF1/2/7* and *STAT1/2/3*,  $FDR < 1\%$ ), mirroring the expression patterns of module 1 (Pearson  $r > 0.92$ ). These results collectively highlight a  $CD16^+$ -specific inflammatory response upon IAV challenge and suggest stronger activation of IRF and STAT transcription factors as a driver of the increased antiviral response observed in  $CD16^+$  cells upon IAV infection.

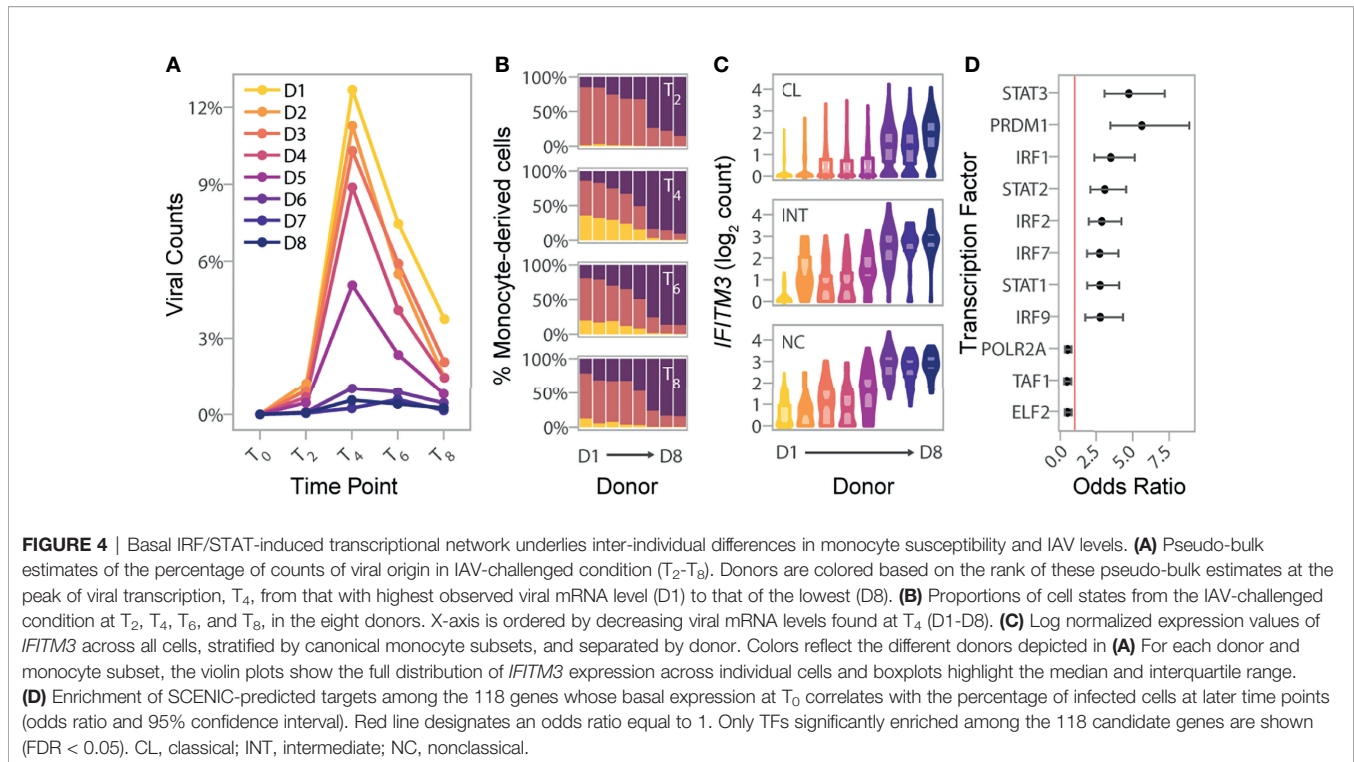
## Basal Activation Differences Correlate With Monocyte Susceptibility

To explore the degree of inter-individual variation upon viral challenge, we next quantified IAV transcripts in the monocyte-derived cells of each individual, and created pseudo-bulk estimates by averaging the percent of viral mRNAs per-cell across all cells from each donor at each time point (**Figure 4A**). While viral mRNAs peaked at the same time for all individuals, we observed extensive variation in the levels of viral mRNAs and percentages of infected cells across individuals (**Figure 4B**). To identify specific genes that might underlie infection potential, we focused on the 4,589 genes that were expressed at  $>0.1 \log_2$  normalized counts in at least one canonical monocyte subset at  $T_0$ . We identified a total of 3,131 genes that differed among our eight donors in either classical, intermediate, and/or nonclassical monocyte subsets (Kruskal-Wallis Rank Test,  $FDR=1\%$ ; **Supplementary Data 4A**). Within each subset, focusing on genes that significantly differed between donors, we



**FIGURE 3** | IRFs and STATs have a central role in the subset-specific responses to IAV infection. **(A)** Heatmap of scaled gene expression from 335 genes displaying a subset-specific response to infection challenge. Genes are grouped into 8 modules based on hierarchical clustering of their expression patterns. Representative genes from each module are labelled. **(B)** Mean expression over time of *IL6* and *TNF*, across the different monocyte subsets and cell states. **(C)** Network of transcription factors (round nodes) associated with each gene expression module (square nodes). Transcription factor nodes are colored according to the number of modules they are associated with. Black lines represent enrichments of the module in TF targets, while red lines represent depletions.





searched for those for which mean expression at basal state was correlated with the percentage of infected cells at  $T_4$  among our eight donors. Despite our limited sample size, we found that cellular susceptibility was strongly correlated with basal expression of the well-known host viral restriction factor *IFITM3*. Although it reached significance only in nonclassical monocytes (FDR~1%), the association remained strong in other subsets ( $p$ -value <  $4.1 \times 10^{-4}$ ; **Figure 4C**).

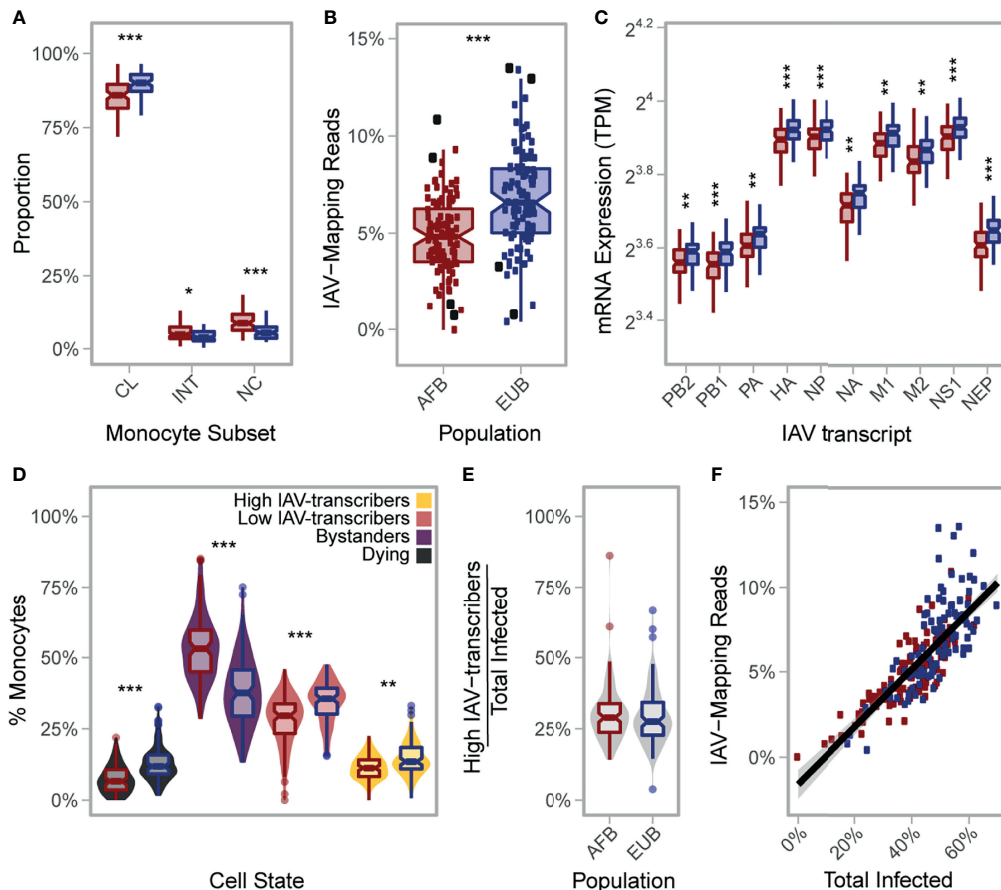
We next relaxed our search to all genes for which basal expression showed nominal correlation ( $p$ -value < 0.01) with the percentage of infected cells at  $T_4$ . Depending on the monocyte subset, between 3.6 to 8.3% of genes matched these criteria, resulting in a set of 118 genes displaying correlation with monocyte susceptibility in at least one subset. These 118 genes were collectively enriched for several related biological processes such as defense response to virus (GO:0051607, OR=15.3, FDR= $9.2 \times 10^{-19}$ ) and ISGs (GO:0034340, OR=19.6 FDR= $8.4 \times 10^{-15}$ ) (**Supplementary Data 4B**). Among genes contributing to this enrichment, we found additional antiviral genes such as *OAS3*, and *MX1*, as well as the critical TF, *IRF7*, involved in the severity of IAV-infection both in mice and humans (30–32). Finally, overlap with the TF targets identified by SCENIC revealed strong enrichments of several IRFs and STATs among the 118 genes, including *IRF7*, as well as *STAT1*, *STAT2* and *IRF9* that form the tripartite IFN-stimulated gene factor 3 (ISGF3) (**Figure 4D**; **Supplementary Data 4C**). Together, our results provide evidence that the basal mRNA expression of genes related to IFN-induced and antiviral responses are indicative of the proportion of cells that will become infected in the first cycle of IAV infection.

## African-Ancestry Monocytes Are More Resistant to Infection

Lastly, we wondered how our findings of inter-individual variation might extrapolate to the population level. In a previous study (18), we challenged the primary monocytes from 200 Belgian individuals of African (AFB) and European (EUB) ancestry with the same IAV strain and MOI used in the present study, and performed bulk RNA-seq at 6 hours post infection (hpi). While basal ( $T_0$ ) expression profiles were not collected, flow cytometry labelling of CD14 and CD16 was performed on the CD14<sup>+</sup>-selected monocytes for the majority of donors. Interestingly, AFB individuals had higher proportions of CD16<sup>+</sup> cells than EUB individuals (**Figure 5A**; **Supplementary Figure 8**). In light of our findings that CD16<sup>+</sup> cells are more resistant to IAV infection, we hypothesized that this might translate to lower infection rates among AFB monocytes relative to EUB monocytes.

To test this hypothesis, we mapped the bulk RNA-seq profiles collected 6hpi challenge with IAV for the 200 individuals to a combined human-IAV reference. Excluding 1 sample with low quality RNAs, we found that 0.02–13.5% of RNA-seq reads from each sample were of viral origin (**Figure 5B**). Reassuringly, these percentages correlated with IAV mRNA levels estimated from the single-cell experiment across all time points for the eight donors used in the present study (Pearson  $r > 0.84$ ,  $p$ -values <  $8.9 \times 10^{-3}$ ), with the strongest correlation being observed at the peak of viral transcription ( $T_4$ ) (Pearson  $r = 0.97$ ,  $p$ -value =  $5.1 \times 10^{-5}$ ). These observations indicate that *ex vivo* cellular susceptibility is highly reproducible among individuals, even across different experimental protocols and technologies.





**FIGURE 5** | African-ancestry individuals display increased number of  $CD16^+$  cells and lower susceptibility to IAV infection. **(A)** Variation in the number of classical ( $CD14^{++}CD16^+$ ), intermediate ( $CD14^{++}CD16^+$ ), and nonclassical ( $CD14^+CD16^{++}$ ) monocytes across African- and European- ancestry individuals following  $CD14^+$  selection from PBMCs ( $n_{AFB} = 89$ ,  $n_{EUB} = 85$ ). Colors reflect population (AFB in red and EUB in blue). All three subsets are significantly different between populations ( $p$ -value $<0.01$ ). **(B)** Inter- and intra-population variation in the percentage of RNA-seq reads mapping to the IAV genome ( $n_{AFB} = 100$ ,  $n_{EUB} = 99$ ). Colors reflect population (AFB in red and EUB in blue). The percentage of RNA-seq reads mapping to the IAV genome is significantly higher in European-ancestry individuals relative to African-ancestry individuals ( $p$ -value =  $5.3 \times 10^{-8}$ ). Donors used in the scRNA-seq experiment ( $n_{AFB} = 4$ ,  $n_{EUB} = 4$ ) are designated with enlarged black squares. **(C)** Inter- and intra-population variation in viral mRNA expression at 6hpi ( $n_{AFB} = 100$ ,  $n_{EUB} = 99$ ). Expression levels for each of the 10 primary transcripts of IAV are plotted. Colors reflect population (AFB in red and EUB in blue). All IAV transcripts are significantly higher in European-ancestry individuals on average ( $p$ -value $<0.001$ ). **(D)** Estimated distribution of the percentage of cells from each cell state in the bulk RNA-seq data ( $n_{AFB} = 100$ ,  $n_{EUB} = 99$ ). Fill colors reflect cell state assignments, while outlines of boxplots reflect population (AFB in red and EUB in blue). **(E)** Distribution of the percentage of high IAV-transcribers among infected cells, stratified by population. One individual with no infected cell was excluded ( $n_{AFB} = 99$ ,  $n_{EUB} = 99$ ). **(F)** Percentage of RNA-seq reads of viral origin as a function of the estimated proportion of infected cells ( $n_{AFB} = 100$ ,  $n_{EUB} = 99$ ), colored by population (AFB in red and EUB in blue). **(A, C)** Outlier points are not displayed. AFB, African-ancestry individuals from Belgium; EUB, European-ancestry individuals from Belgium; TPM, transcripts per million; IAV, influenza A virus; MFI, mean fluorescent intensity; CL, classical; INT, intermediate; NC, nonclassical. \* $p$ -value  $< 0.01$ ; \*\* $p$ -value  $< 0.001$ ; \*\*\* $p$ -value  $< 0.0001$ .

Among the 199 bulk profiles, AFB and EUB samples presented overlapping but significantly shifted distributions of total IAV-mapping reads (**Figure 5B**, 4.9% vs. 6.8% of reads, respectively, Wilcoxon  $p$ -value= $5.3 \times 10^{-8}$ ), and of each of the 10 primary viral transcripts (**Figure 5C**, Wilcoxon  $p$ -values $<5.5 \times 10^{-4}$ ).

Using the transcriptional profiles obtained from the scRNA-seq data at  $T_6$ , we estimated the proportion of reads coming from each inferred cell state in these bulk RNA-seq profiles (**Figures 5D, E**; **Supplementary Note 2** and **Supplementary Figure 9A**). We found that, on average, AFB monocytes were more resistant to IAV infection than EUB monocytes (39.2% vs.

48.9% infected, respectively, Wilcoxon  $p$ -value= $5.3 \times 10^{-10}$ ). Differences in the estimated percentage of infected cells alone explained 63% of the inter-individual variability in viral mRNA levels (**Figure 5F**), and was sufficient to account for the observed difference in viral mRNA levels between AFB and EUB individuals ( $p$ -value=0.16 after adjusting on infected cells, compared to  $p$ -value= $5.3 \times 10^{-8}$  without adjustment). Nonetheless, variation in the percentage of high/low IAV-transcribers among infected cells accounted for an additional 19% of variance in viral mRNA expression (**Supplementary Note 2** and **Supplementary Figure 9B**). Finally, the ratio of

*CD16<sup>+</sup>/CD16<sup>-</sup>* cells negatively correlated with the percentage of infected cells, albeit weakly (-0.27, *p*-value=0.0165 adjusted on population). Altogether, these results show that population differences in viral mRNA levels are primarily driven by the overall proportion of cells that will ultimately become infected, with only a fraction of the differences being attributable to the different proportions of *CD16<sup>+/+</sup>* subsets observed in individuals of African and European ancestry.

## DISCUSSION AND HYPOTHESIS

We performed scRNA-seq on primary monocytes, before and after *ex vivo* IAV challenge, to assess transcriptional differences between monocytes infected by IAV (i.e. infected) versus those activated only by exposure to secreted cytokines (i.e. bystanders), and to identify subset-specific responses of monocytes to viral challenge. We found that bystander cells display increased mRNA expression of ISGs relative to infected cells; yet, we additionally observed both an induction of ribosomal gene mRNA expression in IAV-transcribing cells and a down regulation of these genes in bystander cells at later time points. While the former is likely induced by the virus to enhance mRNA translation (33), the repression of ribosomal expression observed in bystander cells may reflect a host mechanism to contain infection by shutting down the translational machinery of neighboring cells, and we speculate that this may hold true across other cell types and constitute a general cellular defense mechanism against viral infections. Interestingly, the interplay of ribosomal and ISG expression also distinguished infected cells into two distinct states (high and low IAV-transcribers), providing an explanation for the high cell-to-cell variation in IAV replication observed among circulating monocytes, which has also been documented in other cell types and during natural infection (34–41). Notably, type-I IFN genes themselves tended to be preferentially expressed by infected cells in a highly transient manner, suggesting a potential role of monocyte infection in the triggering of the type I IFN response among bystander cells.

While these patterns were generally shared across *CD16<sup>-</sup>* and *CD16<sup>+</sup>* subsets, we found *CD16<sup>+</sup>* cells to be slightly more resistant to infection. This is likely attributable to their higher absolute expression of some ISGs relative to *CD16<sup>-</sup>* cells (independent of viral exposure), as well as their more robust upregulation of antiviral genes upon IAV challenge, which we found to be driven by stronger activity of IRF transcription factors. Interestingly, *CD16<sup>+</sup>* cells displayed transient mRNA expression of *IL6* and *TNF* upon viral exposure (both infected and bystander cells), two cytokines that have been widely implicated in cytokine storms (5). Collectively, these findings highlight the opposing roles of ISG and ribosomal gene mRNA expression on viral transcription, and reveal the stronger antiviral and pro-inflammatory potential of *CD16<sup>+</sup>* monocyte subsets.

At the population level, we found that the ratio of *CD16<sup>+/+</sup>/CD16<sup>-</sup>* at basal state was predictive of the percentage of

monocytes that were susceptible to IAV infection, and observed that African-ancestry individuals, from our sample, harbored more *CD16<sup>+</sup>* monocytes on average than European-ancestry individuals residing in the same city (Ghent, Belgium), consistent with previous observations (42). Independently of monocyte subset proportions, we identified that individuals presenting lower monocyte susceptibility to IAV had a higher basal activation of an IRF/STAT-driven antiviral program. These findings suggest that the fate of a monocyte hinges upon its basal activation state, and that the infection potential differs both within an individual's monocyte population, in part based on the differentiation status of the cell (i.e. *CD16*-positivity), but also between individuals, where a *CD16<sup>-</sup>* cell from one individual may have a higher antiviral state than a *CD16<sup>+</sup>* cell from another individual.

Our finding of a decreased ability of IAV to infect and replicate in monocytes from individuals of African-ancestry was recently replicated in an independent cohort of American individuals with varying levels of African and European ancestry whose PBMCs were challenged with the 2009 pandemic H1N1 strain (43). While the cause of these population differences remains to be determined, we did not find evidence that strong-effect genetic factors, nor evidence of past exposure to H1N1, could explain such an association with the viral replication phenotype. Nevertheless, the observed inter- and intra-population differences are noteworthy in and of themselves, and may reflect the influence of both weak-effect genetic loci, and non-heritable factors, such as stress, nutrition or lifestyle, on transcriptional variation of immune genes (14, 15). Future studies are needed to determine if such population differences hold true across other cell types, such as lung epithelial cells.

Given our finding that *CD16<sup>+</sup>* subsets are the main drivers of inflammatory cytokine gene expression such as *IL6* and *TNF*, and that African-ancestry individuals harbor a larger fraction of these subsets, it is tangible to conceive that monocyte subset composition prior to infection may influence disease outcome. A lower percentage of infected monocytes could also contribute to a faster disease progression, as we find that infected monocytes continue to express antigen-presenting genes. Thus, a higher number of infected cells could lead to a stronger activation of the adaptive immune system. In support of these hypotheses, patients with severe influenza and COVID-19 harbor higher proportions of intermediate monocytes in peripheral blood than patients with mild disease (44, 45), and African Americans are more often hospitalized than other self-defined ethnic groups by both influenza (46, 47) and COVID-19 (48, 49), even when adjusting for age and various social factors such as poverty and vaccination status. In light of these observations, we hypothesize that the higher percentage of *CD16<sup>+</sup>* monocyte observed among African-ancestry individuals may, in conjunction with a stronger basal activation of their monocytes, contribute to poor infectious outcomes. Further studies are now needed to formally establish the clinical relevance of monocyte heterogeneity in the context of viral infections, IAV in particular, and determine its potential use as a biomarker.

## LIMITATIONS OF STUDY

In this Hypothesis and Theory article, we analyze single-cell transcriptional heterogeneity of circulating monocytes before and after *ex vivo* IAV challenge, and propose that differences in basal monocyte activation underlie population disparities in cellular susceptibility to IAV. We acknowledge that our study is based on positively-selected monocytes isolated from PBMCs, and that infection of this cell population *in vivo* would be expected to take place in the lung, after an initial infection has been established. Future studies should investigate how these findings translate to other cell populations - including lung epithelial cells and resident monocyte and macrophage populations - and whether they influence clinical outcomes. This could be achieved, for instance, by using single cell techniques to measure how nasal epithelial cells from healthy patients of various ancestries differ in their expression of viral RNAs and proteins upon IAV challenge. Additionally, sputum extract could be collected from mild and severe influenza patients of both ancestries, to compare the single cell transcriptome of lung epithelial cells and resident macrophage populations, both across ancestries and in relation with disease severity. Another caveat of the study is the lack of detailed lifestyle observations in the cohort used, precluding us from examining in further detail the influence of non-genetic factors. Further studies are now needed to evaluate how non-genetic factors, such as social status, chronic stress levels (and the induced physiological response), previous exposures to pathogens or even the microbiome, could contribute to shape basal monocyte activation and prime the innate immune response to viral infections.

## METHODS

### Experimental Model and Subjects

All individuals from this study were part of the EVOIMMUNOPOP cohort, which has been previously described (18). Human blood was obtained from healthy volunteers who gave informed consent, and the PBMC fraction was isolated and frozen. In brief, 200 healthy male donors living in Belgium of self-reported African descent (AFB) or European descent (EUB) were recruited. Inclusion was restricted to nominally healthy individuals between 19 and 50 years of age at the time of sample collection. The majority of our African-descent individuals originated from West Central Africa, with >90% of our sample being born in either Cameroon or Congo. Serological testing was performed for all donors to exclude those with serological signs of past or ongoing infection with human immunodeficiency virus (HIV), hepatitis B virus (HBV) or hepatitis C virus (HCV).

### Single-Cell Analyses and RNA-Sequencing

For eight selected donors [4 individuals from each ancestry, selected from extremes of the first principal component of gene expression in our previous study of monocyte response to IAV challenge (18)],  $100 \times 10^6$  PBMCs were thawed, washed twice and

resuspended in complete medium: pre-warmed RPMI-1640 Glutamax medium, supplemented with 10% FCS and 1% penicillin/streptomycin (Cat#15140-122, Life Technologies). Monocytes were then positively selected with magnetic CD14 microbeads, according to the manufacturer's instructions (Cat#130-050-201, Miltenyi Biotec). The number of monocytes was determined with the Countless2 automated cell counter system (Cat#AMQAX1000, ThermoFisher Scientific) in the presence of trypan blue. For each donor, monocytes were seeded at  $0.5 \times 10^6$  monocytes per well on 24-well NUNC plates in 500  $\mu$ L of complete media and allowed to rest for one hour at 37°C under 5% CO<sub>2</sub>. Five-hundred microliters of complete media (non-infected) or A/USSR/90/1977(H1N1) at a concentration of  $1 \times 10^6$  pfu/mL in complete media (IAV-challenged, MOI=1) were added to each sample. Following one hour of staging at 4°C, plates were centrifuged at 1300 rpm for 10 minutes at 4°C, media was removed by pipette, and each well was washed with 1mL complete media. The spin was repeated, media removed by pipette, and samples were resuspended in 1mL pre-warmed complete media before being transferred to an incubator at 37°C under 5% CO<sub>2</sub> to initiate infection (T<sub>0</sub>).

At each time point (T<sub>0</sub>, T<sub>2</sub>, T<sub>4</sub>, T<sub>6</sub>, and T<sub>8</sub>), samples were mixed by pipetting and transferred to Eppendorf tubes. Wells were washed with 300 $\mu$ L of PBS + 0.04% BSA and transferred to the same tubes. Collection tubes were centrifuged at 1300 rpm for 10 minutes, media was removed and replaced with 1mL PBS + 0.04% BSA and an aliquot of 10 $\mu$ L was taken to count each sample on a Countless2 automated cell counter system, before repeating the centrifugations. Individual samples were adjusted to  $2 \times 10^6$  live cells/mL.

Samples were multiplexed for running on the 10X Chromium (Cat#120223 & 1000074, 10X Genomics) by mixing equal proportions from 6-8 samples in a manner that balanced conditions and allowed us to assess for batch effects across lanes (**Supplementary Table 1**). Multiplexed samples were counted with the Countless2 automated cell counter system and adjusted to target recovery of 10,000 cells per reaction of the Chromium Single Cell 3' Reagent Kits v3 (Cat#1000092 & 1000078, 10X Genomics) assuming a recovery rate of 50%. GEM Generation & Barcoding, Post GEM-RT Cleanup & cDNA Amplification, and 3' Gene Expression Library Construction were performed as per manufacturer's instructions (50). All 13 libraries were mixed prior to sequencing across 13 different lanes from an Illumina HiSeq X (28bp barcode + 91bp insert - target 400 M reads pairs per lane), leading to a total of 5.3 billion reads.

### Sample Genotyping

Genotyping data [accession EGAS00001001895] were obtained for all 200 individuals from the EvoImmunoPop cohort based on both Illumina HumanOmni5-Quad BeadChips and whole-exome sequencing with the Nextera Rapid Capture Expanded Exome kit (18). The 3,782,260 SNPs obtained after stringent quality control were then used for imputation, based on the 1,000 Genomes Project imputation reference panel (Phase 1 v3.2010/11/23) (51), leading to a final set of 19,619,457 high-quality SNPs, of which 7,766,248 SNPs had a MAF  $\geq 5\%$  in our cohort.



## Processing of scRNA-Seq Data

Basic pre-processing of the sequencing data was performed with Cell Ranger v3.0.2 (52), including the *mkfastq*, *count*, and *aggr* commands. Default parameters and our combined human-IAV reference were used, and batch correction was disabled in the *aggr* command. Cell-containing droplets ( $n=132,130$ ) were traced back to individual donors using two independent methods, Demuxlet and SoupOrCell, which capitalize on genetic variation in the sequencing reads (20, 53). Barcodes with ambiguous and/or non-concordant calls between the two programs were used to establish suitable QC metrics. We found that barcodes deemed as *doublets* (i.e. the droplet contained two or more cells originating from different donors) were more likely to be nearest-neighbors in a *knn*-graph with other *doublets* than assigned *singlets*. We used this feature to identify droplets presumed to contain two or more cells originating from the same donor; barcodes with  $> 5$  *doublets* as nearest-neighbors were excluded from further analysis (**Supplementary Figures 1A, B**). Additionally, droplets containing low-quality cells (i.e. damaged, dying) were excluded using the following thresholds:  $<1500$  total counts,  $<500$  genes, or  $>50\%$  mitochondrial gene content (**Supplementary Figure 1C**). This QC resulted in 96,386 single cells.

Transcriptomes (i.e. counts) were adjusted for the presence of ambient RNA with SoupX, (<https://github.com/constantAmateur/SoupX>, accessed November 28, 2019) (21), using estimated contamination fractions (per 10X library) from SoupOrCell (53). SoupX-adjusted counts were normalized using pool-based size factors followed by deconvolution as implemented in the *scran* R package (54). Feature selection was performed by (i) constructing a mean-variance trend in the log-counts and retaining genes found to exhibit more variation than expected assuming Poisson-distributed technical noise, as implemented in the *makeTechTrend* and *TrendVar* functions from package *scran* (54), and (ii) selecting genes expressed in at least 25 cells ( $n=22,603$ ). The first 10 PCs of the data were retained for data visualization and clustering analyses. Graph-based clustering was performed by building the shared nearest-neighbor graph with the *buildSNNGraph* function from *scran* (54) using a series of  $k$  values, and cell clusters were defined with the *igraph* Walktrap algorithm (55). Similar clustering results were obtained based on the *knn*-graphs generated using  $k=25, 50, 75$ , and  $100$ , and  $k=25$  was used for all downstream analyses (**Supplementary Figure 2A**). Cell types were predicted using SingleR and the built-in *BlueprintEncodeData* reference (56). Based on the clustering and cell-type predictions, we removed cells belonging to clusters associated with lymphoid cell types or low QC metrics from downstream analyses (**Supplementary Figures 2B, C**).

## Accounting for Ambient RNA Contamination in scRNA-Seq Data and Assigning Cell States

Droplet-based scRNA-seq methods capture ambient mRNAs present in the cell suspension in addition to cell specific mRNAs. To estimate which cells in our experiment were genuinely expressing mRNAs for *CD14*, *FCGR3A* (*CD16*), and

those originating from the virus, we implemented a two-step strategy utilizing the *estimateNonExpressionCells* function of the SoupX package (21). This function estimates whether each cell contains significantly more counts of a provided gene-set than would be expected under a Poisson model, given the estimated ambient RNA from its library of origin and the maximum contamination fraction. First, we used the viral genes to estimate the true maximum contamination fraction, based on the assumption that cells from the non-infected state should only contain viral reads from ambient mRNA captured in their droplets. To do so, we modified the *estimateNonExpressionCells* function to return  $p$ -values, and performed the test on each of our 13 libraries with a range of maximum contamination values from 1-50% (step of 1%) using the viral genes. We then computed FDR adjusted  $p$ -values for each maximum contamination value on the 88,559 high-quality, single monocytes. The number of non-simulated cells deemed to significantly express IAV transcripts (FDR $<0.01$ ) was used as a proxy for false positives. In examining the relationship between this number and the number of IAV-challenged cells found to significantly express viral transcripts at FDR $<0.01$  (**Figure 1C**), we found that a maximum contamination fraction of 10% resulted in a 1% false positive rate (defined as the percentage of non-infected cells from T<sub>2</sub>-T<sub>8</sub> that were deemed to significantly express IAV transcripts). This parameter value was then used to correct for contamination from ambient for all genes considered (*CD14*, *FCGR3A* and IAV transcripts).

## Assigning Cell States and Investigating Sources of Variability in IAV Levels

We used a maximum contamination fraction of 10% to test for significant expression of IAV transcripts in each cell (**Figures 1C-E**). IAV-challenged cells that contained a significant amount of IAV transcripts were considered as infected, while the others were deemed bystanders. To distinguish low from high IAV-transcribing cells, a Gaussian mixture model was fitted to the total percentage of viral mRNAs per cell across all infected cells, using the *normalmixEM* function from *mixtools* R package with  $k=2$  (57). Each cell was assigned to the cluster with the highest posterior probability, and the cluster of cells with higher IAV content was annotated as high IAV-transcribing.

## Characterizing Monocyte Subsets and Transcriptional Profiles From scRNA-Seq Data

Principal components analysis of 6,601 cells at T<sub>0</sub> was used to order monocytes along a differentiation axis separating *CD14*<sup>+</sup> cells from *CD16*<sup>+</sup> cells. We then computed the average percentage of classical and nonclassical monocytes obtained by flow cytometry across the eight donors, weighting each individual by the number of high-quality cells in the scRNA-seq data at T<sub>0</sub>. Based on these percentages (87.1% for classical and 7.6% for nonclassical), we annotated the monocytes on each side of the differentiation axis as classical and nonclassical, respectively, with the remaining 5.3% of monocytes being annotated as intermediates. Validity of our approach was



confirmed by correlating the proportion of monocytes assigned to each subset across the eight donors, with the percentage of classical, intermediate and nonclassical monocytes estimated by flow cytometry.

Differential expression between subsets was assessed for the 4,859 genes expressed at a normalized  $\log_2$  count  $> 0.1$  in any of the 3 subsets. Specifically, Wilcoxon rank tests were implemented in the *scran* package (54), using the *findMarkers* function and blocking on donor. We considered genes to be differentially expressed (DE) between monocyte subsets when gene expression was significant at an  $\text{FDR} \leq 1\%$  and  $\log_2\text{FC} > 0.2$ . The 848 genes that differed between classical (CL) and nonclassical (NC) monocyte subsets were classified according to their behavior in intermediate monocytes (INT). They were either deemed 'similar to classical' (DE between INT and NC, but not between INT and CL), 'similar to nonclassical' (DE between INT and CL, but not between INT and NC), or 'intermediate' (all other cases).

At later time points, comparisons between  $CD16^+$  and  $CD16^-$  monocytes subsets were done based on 5,681 genes expressed with a normalized  $\log_2$  count  $> 0.1$  on average in either subset, in at least one condition and time point. For each subset,  $\log_2$  fold change in gene expression relative to  $T_0$  were correlated across times points. Differential expression between  $CD16^+$  and  $CD16^-$  cells was assessed with *findMarkers* (54), based on Wilcoxon rank tests and blocking on donors, time points and condition. Again, an  $\text{FDR} \leq 1\%$  and  $\log_2\text{FC} > 0.2$  were required to define differentially expressed genes. To assess how  $CD16^{+/-}$  status alters the infection of monocytes by IAV, we used logistic regression to model bystander/infected status as a function of  $CD16^{+/-}$  status, while adjusting on donor, and time point (as factors).

## Characterizing Subset-Specific Responses to IAV Challenge

To allow comparison between responses of  $CD16^+$  and  $CD16^-$  monocytes, 100 cells were subsampled from each subset and cell state, and at each time point. When subsampling, we ensured balanced representation of all donors across each monocyte subset and cell state, by using sampling weights that were inversely proportional to each donor representation in the original dataset. After sampling, a total of 6,669 genes with normalized  $\log_2$  counts  $> 0.1$  on average in at least one group (cell-state x subset x time point) was selected for further analyses. For each monocyte subset, differences in expression between cell states (unexposed, bystander, infected) as well as between high- and low-IAV transcribing infected cells were performed using the *findMarkers* function from the *scran* package (54) and blocked on time point. For each comparison, genes were considered to be differentially expressed between cell states when gene expression was significant at an  $\text{FDR} = 1\%$  (Wilcoxon rank tests) and the  $\log_2$  fold change was  $> 0.2$ . In addition, for each comparison between cell states, we tested for differences in response between subsets using a linear model of the form:

$$\text{Expr}_i \sim \text{State}_i + \text{subset}_i + \text{State}_i \cdot \text{subset}_i \quad (1)$$

where  $\text{Expr}_i$  is the expression of the gene being tested in cell  $i$ ,  $\text{State}_i$  is an indicator variable that distinguishes the two cell states being compared (e.g. unexposed and bystander), and  $\text{subset}_i$  is an indicator variable that reflects the  $CD16^{+/-}$  status of cell  $i$ . The 335 genes with significant interactions at a 1% FDR (for unexposed-bystander and unexposed-infected comparisons) were clustered using the *hclust* R function with method 'Ward.D2'. *DynamicTreeCut* algorithm (58) was used to identify eight major patterns of response to IAV.

## Transcription Factor Enrichment Analyses

To estimate Transcription Factor (TF) activity and define TF-targets relationships, we ran the R SCENIC pipeline (29) on the expression matrix (pre-normalization) on a random subsample of 4800 cells (100 cells from each donor at each time point and each condition, pre-exclusion of dying and contaminant cells) with default parameters. For each gene, motif-enrichment was considered for either *cis*-regulatory regions located  $< 10\text{kb}$  from the TSS (distal regulatory elements), or between 500 bp upstream and 100 bp downstream of the promoter (proximal regulatory elements). To do so, motif-enrichment scores for all human genes (hg38 build, refseq\_r80), were retrieved from <https://resources.aertslab.org/cistarget> and used as input for the *Rcistarget* package (29).

Sets of high-confidence targets for the 113 TFs whose activity could be quantified by SCENIC were then extracted and used for enrichment analysis. For each gene module, TF enrichment was assessed using a Fisher's exact test with the 6,669 expressed genes as background (Supplementary Data 3). Resulting  $p$ -values were adjusted using a global Benjamini-Hochberg correction for all eight modules and 113 TFs.

For each TF, with its targets enriched among one of the eight modules, TF activity inferred by SCENIC was used to test for subset-specific activity using a linear model of the form:

$$TF_i \sim \text{State}_i + \text{subset}_i + \text{State}_i \cdot \text{subset}_i \quad (2)$$

where  $TF_i$  is the activity of the TF being tested in cell  $i$ ,  $\text{State}_i$  is an indicator variable that distinguishes the two cell states being compared (e.g. unexposed and bystander), and  $\text{subset}_i$  is an indicator variable that reflects the  $CD16^{+/-}$  status of cell  $i$ . Average TF activity was then computed for each cell state, subset and time point, and correlated with gene expression of the associated module, to assess the link between TF activation and the TF-target enriched modules.

## Association of the Outcome of IAV Infection With Basal Gene Expression

For each of the three monocyte subsets detected at basal state, a Kruskal-wallis test was used to search for genes whose expression levels significantly differ across donors. Within each monocyte subset, we then computed the average expression of each gene for all eight donors and correlated it with the percentage of infected cells at 4hpi. Genes that differed in expression between donors ( $\text{FDR} \leq 1\%$ ), and passed a nominal  $p$ -value threshold of 0.01 for association with IAV levels in any of the 3 subsets, were selected for downstream enrichment analyses.

For genes nominally correlated with viral mRNA levels, TF enrichment was assessed as previously using a Fisher's exact test with all 4,859 genes expressed at  $T_0$  as background (**Supplementary Data 4**), and Benjamini-Hochberg correction for all 113 TFs was applied.

## Gene Ontology Enrichment Analyses

All Gene Ontology (GO) enrichment analyses were performed with the GOSep package using default settings (59). Background gene sets consisted of all genes that had average log-normalized expression values  $> 0.1$  in at least one of the groupings being examined, and are described in the text. Only enrichments significant at  $FDR \leq 5\%$  are reported.

## Pseudo-Bulk Estimates From scRNA-Seq Data

Pseudo bulk estimates of IAV mRNA levels were computed by measuring, for each donor and time point, the mean percentage of reads of viral origin across all cells from the sample. At each time point, we then used a Pearson's correlation test to compare pseudo-bulk estimates for the 8 donors with IAV mRNA levels obtained in bulk data at 6hpi.

## Monocyte Subset Characterization of EVOIMMUNOPOP Samples via Flow Cytometry

For 174 of the 200 EVOIMMUNOPOP donors, proportions of classical, intermediate and nonclassical monocytes were determined based on a fraction of  $10^5$  CD14<sup>+</sup> positively-selected monocytes, stained according to the manufacturer's instructions, with fluorescent APC-conjugated anti-CD14 and PE-conjugated anti-CD16 antibodies (Cat#130-091-243 and Cat #130-091-245, respectively, Miltenyi Biotec). Samples were then analyzed on a MACSQuant Analyzer 10 benchtop flow cytometer (Miltenyi Biotec).

## Quantification of Canonical Monocyte Subsets in EVOIMMUNOPOP Samples

FlowJo v10.6.1 software (60) was used with the gating strategy depicted in **Supplementary Figure 8** to quantify monocyte subsets for 174 EVOIMMUNOPOP donors. Population-level differences in proportion of canonical monocyte subsets were assessed using Wilcoxon Rank tests. Correlation of the ratio of CD16<sup>+</sup> to CD16<sup>-</sup> cells with IAV mRNA levels was assessed using a linear model of the form

$$IAV \sim ratio + Pop, \quad (3)$$

where 'IAV' are IAV mRNA levels, 'ratio' is the percentage of CD16<sup>+</sup> monocytes (nonclassical+intermediates) divided by the percentage of CD16<sup>-</sup> monocytes (classical), and 'Pop' is an indicator variable separating AFB from EUB individuals.

## Analysis of Bulk RNA-Seq Profiles From the EVOIMMUNOPOP Cohort

A combined human-IAV reference was generated by concatenation of the primary human genome assembly

(GRCh38) with the 8 segments of the human influenza A virus (IAV) A/USSR/90/1977(H1N1) genome (accession numbers CY010372-CY010379). Comprehensive human gene annotation was obtained from GENCODE (release 27) and merged with the 12 known transcripts of A/USSR/90/1977 (H1N1). RNA-seq reads (FASTQs) for all 970 samples that passed quality control in our previous study (18) [accession EGAS00001001895] were mapped to the combined reference with the STAR aligner (v.2.5.0a) (61) and assessed for quality with QualiMap 'bamqc' and 'rnaseq' (62, 63). Expression of viral mRNAs was measured as the percentage of uniquely-mapped reads aligning to the IAV genome. Reassuringly, the mean percentage of RNA-seq reads among samples from the IAV-challenged condition was 5.86% versus  $< 0.01\%$  in the other four conditions. Comparison of the percentage of IAV reads between populations was done using a Wilcoxon rank test. StringTie (v.1.3.3) (64) was used to quantify expression levels in transcripts per million mapped reads (TPM) for each annotated transcript. Gene expression data were filtered to remove genes with little evidence of activation (mean zTPM score  $< -3$ ) (65) in any of the 5 conditions, and their quality was checked by principal component analysis (PCA). As GC content, 5'/3' bias, date of the experiment and library batch were previously determined to be the strongest confounding factors on transcript expression (18), we corrected the data for these factors. First, we adjusted the data for GC content and 5'/3' bias using linear models. Then, we imputed missing values by k-nearest neighbor imputation and adjusted for experiment date and library batch by sequentially running ComBat (66) for each batch effect, with condition and population as covariates. After batch effect correction, only IAV-stimulated samples were kept for downstream analyses.

## Cell States Deconvolution From Bulk RNA Sequencing

To assess the percentage of total transcripts that originate from each cell state across the 199 IAV-challenged samples, we pooled cells from  $T_6$  into 3 groups, based on their assigned cell-state (bystander, infected: high and low IAV-transcribing) and to which we added a 4<sup>th</sup> group containing all singlets that were either (i) assigned to cluster numbers 3, 8, 10, and 11 (dying cells) or (ii) discarded based on their high mitochondrial content or low read counts (dead cells). We then estimated pseudo-bulk profiles for each group by summing UMIs across all cells and computing the number of UMIs associated to each gene per million of sequenced UMIs. TPM profiles obtained from bulk data were then normalized to improve comparison with pseudo-bulk. Specifically, we first computed a global pseudo-bulk profile of the entire single cell dataset as the average of the pseudo bulk profiles from the 4 cell states (bystander, infected: high and low IAV-transcribing, or dying/dead), weighted by the percentage of UMIs they contribute to the overall pool of cells. To account for the difference in how gene expression is quantified between the two methods (3' end counts for scRNA-seq and full-length gene coverage for bulk RNA-seq), we computed for each gene  $i$  a normalization factor  $s_i$  given by

$$s_i = \log(\overline{TPM}_i) - \log(PB_i) \quad (4)$$

where  $PB_i$  is the number of UMI per million for gene  $i$  in the global pseudo-bulk profile, and  $\overline{TPM}_i$  is the average expression of the gene  $i$  in the 199 IAV-stimulated samples from the bulk RNA-seq data. For each gene,  $s_i$  was then subtracted from the log transformed TPM to yield normalized TPM profiles. We next applied DeconRNAseq (67) to the normalized log TPM profiles from all individuals, using the log-transformed pseudo bulk profiles from the 4 cell states as a basis for deconvolution. Quality of the deconvolution was assessed using leave-one-out cross validation, based on the eight individuals for whom we had scRNA-seq data. Specifically, for each of these eight individuals, bulk mRNAs were decomposed using pseudo-bulk profiles recomputed based on the seven other individuals. The resulting proportions were then compared with the percentage of UMIs that originate in each cell-state in the scRNA-seq to assess the quality of the deconvolution. Excluding IAV genes from bulk transcriptomic profiles prior to performing the deconvolution had virtually no impact on the estimated proportions (Pearson  $r > 0.98$  with proportions estimated without excluding IAV genes), confirming that our estimates were not driven by IAV expression alone. Comparisons between populations were performed using Wilcoxon rank tests.

The effect of the percentage of infected cells and percentage of high IAV-transcribing cells among infected cells on the total IAV mRNA levels were assessed by modelling

$$IAV \sim INF + POP \quad (5)$$

And

$$IAV \sim HI + POP \quad (6)$$

where  $IAV$  are the IAV mRNA levels across the 199 bulk mRNA samples,  $INF$  and  $HI$  are respectively the percentage of infected cells and the percentage of high IAV-transcribing cells among infected cells that we estimated from the deconvolution, and  $POP$  is a factor variable reflecting the population (EUB or AFB). The fraction  $\eta$  of population differences attributable to difference in rate of infection was estimated by comparing model (5) with model (7) below

$$IAV \sim POP \quad (7)$$

and computing  $\eta = 100 \times (1 - \frac{\beta_i}{\beta_j})$ , where  $\beta_i$  is the effect of population on IAV levels in model (i). To assess how the contribution of the percentage of high IAV transcribing cells to total IAV mRNA levels differed between populations, we used a linear model of the form

$$IAV \sim HI + POP + HI : POP \quad (8)$$

and tested for significant effect of the interaction term  $HI : POP$  on IAV mRNA levels.

## Flow Cytometry Analysis of Monocyte Susceptibility to IAV Infection

Frozen PBMCs from 8 individuals included in the EVOIMMUNOPOP cohort were thawed and allowed to rest

overnight at 37°C, 5% CO<sub>2</sub> in 25cm<sup>2</sup> flasks. Cells were then seeded at 2×10<sup>6</sup>/ml in untreated 96-well plates in RPMI-1640 GlutaMAX supplemented with 10% FCS in the presence of A/PR/8/34 (H1N1) (Charles River Laboratories) at a MOI=1 or media alone for 6h at 37°C, 5% CO<sub>2</sub>. At the end of the incubation, cells were washed in FACS buffer (1X PBS supplemented with 2% FCS and 1mM EDTA) and stained with the LIVE/DEAD fixable violet dead cell stain kit (Cat#L34955, Life Technologies) and human Fc block for 15 min at 4°C, protected from light. Cells were then washed and stained with a mix of 6 surface antibodies for 20 min at 4°C, protected from light (anti-human CD19 BV510 Cat#562947, CD3 APC Cat#561811, CD16 PerCp-Cy5.5 Cat#560717, CD69 PE-Cy7 Cat#557745 from BD Biosciences, anti-human CD14 Cat#301806 from Biolegend, anti-human CD56 APC-Vio770 Cat#130-114-739 from Miltenyi Biotec). After centrifugation at 300g for 5 min, cells were incubated with the Fixation/Permeabilization solution from the Cytofix/Cytoperm kit (BD Biosciences) for 15min at 4°C, followed by intracellular staining with FITC conjugated anti-NP (Cat#MA1-7322, ThermoFisher Scientific) in BD Perm/Wash buffer (1X) for 30min at 4°C. Cells were washed and acquired using a MACSQuant (Miltenyi Biotec), and data were analyzed with FlowJo v10 with the gating strategy depicted in **Supplementary Figure 6A**.

## DATA AVAILABILITY STATEMENT

The bulk RNA-seq data used in this study are available at the European Genome Phenome archive under accession number [EGAS00001001895]. The single cell RNA-seq data generated during this study are available at the European Genome Phenome archive under accession number [EGAS00001005000]. Code generated as part of this study is available on Github ([https://github.com/h-e-g/PopDiff\\_MonocyteIAV](https://github.com/h-e-g/PopDiff_MonocyteIAV)).

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Board of Institut Pasteur (EVOIMMUNO POP-281297) and the relevant French authorities (CPP, CCITRS, and CNIL). All experimental methods were conducted in accordance with the Declaration of Helsinki principles. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MO'N, MR, and LQ-M conceived and designed the study. MO'N, HQ, JP, YA, AB, NZ, and MD conducted the experiments at the Human Evolutionary Genetics Unit. MO'N, MR, YA, and DM developed computational methods and performed bioinformatic analyses. JP, VL, MH, S-ZY, QZ, AC,



LA, J-LC, and NN provided resources, expertise and feedback. MR and LQ-M supervised the study. LQ-M secured funding. MO'N, MR and LQ-M wrote the manuscript with input from all authors. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by the Institut Pasteur, the Collège de France, the French Government's Investissement d'Avenir program, Laboratoires d'Excellence "Integrative Biology of Emerging Infectious Diseases" (ANR-10-LABX-62-IBEID) and "Milieu Intérieur" (ANR-10-LABX-69-01), the Fondation de France (n°00106080), and the Fondation pour la Recherche Médicale (Equipe FRM DEQ20180339214). MO'N was

supported by a European Molecular Biology Organization long-term fellowship (ALTF 229-2017).

## ACKNOWLEDGMENTS

We wish to thank Tzachi Hagai and John Marioni for input and feedback, and Sylvie van Der Werf and Vincent Enouf for providing resources and protocols relating to influenza viruses.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.768189/full#supplementary-material>

## REFERENCES

- Ryabkova VA, Churilov LP, Shoenfeld Y. Influenza Infection, SARS, MERS and COVID-19: Cytokine Storm - The Common Denominator and the Lessons to be Learned. *Clin Immunol (Orlando Fla)* (2021) 223:108652. doi: 10.1016/j.clim.2020.108652
- Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. Influenza. *Nat Rev Dis Primers* (2018) 4(1):3. doi: 10.1038/s41572-018-0002-y
- Zhang Q, Bastard P, Bolze A, Jouanguy E, Zhang SY, Effort CHG, et al. Life-Threatening COVID-19: Defective Interferons Unleash Excessive Inflammation. *Med (N Y)* (2020) 1(1):14–20. doi: 10.1016/j.medj.2020.12.001
- Stegelmeyer AA, van Vloten JP, Mould RC, Klafuric EM, Minott JA, Wootton SK, et al. Myeloid Cells During Viral Infections and Inflammation. *Viruses* (2019) 11(2):168. doi: 10.3390/v11020168
- Fajgenbaum DC, June CH. Cytokine Storm. *N Engl J Med* (2020) 383(23):2255–73. doi: 10.1056/NEJMra2026131
- Guo C, Li B, Ma H, Wang X, Cai P, Yu Q, et al. Single-Cell Analysis of Two Severe COVID-19 Patients Reveals a Monocyte-Associated and Tocilizumab-Responding Cytokine Storm. *Nat Commun* (2020) 11(1):3924. doi: 10.1038/s41467-020-17834-w
- Alon R, Sportiello M, Kozlovski S, Kumar A, Reilly EC, Zarbock A, et al. Leukocyte Trafficking to the Lungs and Beyond: Lessons From Influenza for COVID-19. *Nat Rev Immunol* (2021) 21(1):49–64. doi: 10.1038/s41577-020-00470-2
- Ziegler-Heitbrock L, Ancuta P, Crowe S, Dalod M, Grau V, Hart DN, et al. Nomenclature of Monocytes and Dendritic Cells in Blood. *Blood* (2010) 116(16):e74–80. doi: 10.1182/blood-2010-02-258558
- Geissmann F, Jung S, Littman DR. Blood Monocytes Consist of Two Principal Subsets With Distinct Migratory Properties. *Immunity* (2003) 19(1):71–82. doi: 10.1016/S1074-7613(03)00174-2
- Auffray C, Fogg D, Garfa M, Elaine G, Join-Lambert O, Kayal S, et al. Monitoring of Blood Vessels and Tissues by a Population of Monocytes With Patrolling Behavior. *Science* (2007) 317(5838):666–70. doi: 10.1126/science.1142883
- Ziegler-Heitbrock L. The CD14+ CD16+ Blood Monocytes: Their Role in Infection and Inflammation. *J Leukoc Biol* (2007) 81(3):584–92. doi: 10.1189/jlb.0806510
- Hoeve MA, Nash AA, Jackson D, Randall RE, Dransfield I. Influenza Virus A Infection of Human Monocyte and Macrophage Subpopulations Reveals Increased Susceptibility Associated With Cell Differentiation. *PLoS One* (2012) 7(1):e29443. doi: 10.1371/journal.pone.0029443
- Hou W, Gibbs JS, Lu X, Brooke CB, Roy D, Modlin RL, et al. Viral Infection Triggers Rapid Differentiation of Human Blood Monocytes Into Dendritic Cells. *Blood* (2012) 119(13):3128–31. doi: 10.1182/blood-2011-09-379479
- Piasecka B, Duffy D, Urrutia A, Quach H, Patin E, Posseme C, et al. Distinctive Roles of Age, Sex, and Genetics in Shaping Transcriptional Variation of Human Immune Responses to Microbial Challenges. *Proc Natl Acad Sci USA* (2018) 115(3):E488–97. doi: 10.1073/pnas.1714765115
- Brodin P, Jovic V, Gao T, Bhattacharya S, Angel CJ, Furman D, et al. Variation in the Human Immune System Is Largely Driven by Non-Heritable Influences. *Cell* (2015) 160(1-2):37–47. doi: 10.1016/j.cell.2014.12.020
- Astle WJ, Elding H, Jiang T, Allen D, Ruklija D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* (2016) 167(5):1415–29.e19. doi: 10.1016/j.cell.2016.10.042
- Patin E, Hasan M, Bergstedt J, Rouilly V, Libri V, Urrutia A, et al. Natural Variation in the Parameters of Innate Immune Cells Is Preferentially Driven by Genetic Factors. *Nat Immunol* (2018) 19(3):302–14. doi: 10.1038/s41590-018-0049-7
- Quach H, Rotival M, Pothlichet J, Loh YE, Dannemann M, Zidane N, et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* (2016) 167(3):643–56.e17. doi: 10.1016/j.cell.2016.09.024
- Ouwens KG, Jansen R, Nivard MG, van Dongen J, Frieser MJ, Hottenga JJ, et al. A Characterization of Cis- and Trans-Heritability of RNA-Seq-Based Gene Expression. *Eur J Hum Genet* (2020) 28(2):253–63. doi: 10.1038/s41431-019-0511-5
- Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed Droplet Single-Cell RNA-Sequencing Using Natural Genetic Variation. *Nat Biotechnol* (2018) 36(1):89–94. doi: 10.1038/nbt.4042
- Young MD, Behjati S. SoupX Removes Ambient RNA Contamination From Droplet-Based Single-Cell RNA Sequencing Data. *Gigascience* (2020) 9(12):giaa151. doi: 10.1093/gigascience/giaa151
- Wong KL, Tai JJ, Wong WC, Han H, Sem X, Yeap WH, et al. Gene Expression Profiling Reveals the Defining Features of the Classical, Intermediate, and Nonclassical Human Monocyte Subsets. *Blood* (2011) 118(5):e16–31. doi: 10.1182/blood-2010-12-326355
- Segura V, Valero ML, Cantero L, Muñoz J, Zarzuela E, García F, et al. In-Depth Proteomic Characterization of Classical and Non-Classical Monocyte Subsets. *Proteomes* (2018) 6(1):8. doi: 10.3390/proteomes6010008
- Schmidl C, Renner K, Peter K, Eder R, Lassmann T, Balwierz PJ, et al. Transcription and Enhancer Profiling in Human Monocyte Subsets. *Blood* (2014) 123(17):e90–9. doi: 10.1182/blood-2013-02-484188
- Bercovich-Kinori A, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, et al. A Systematic View on Influenza Induced Host Shutoff. *eLife* (2016) 5:e18311. doi: 10.7554/eLife.18311
- Machkovech HM, Bloom JD, Subramaniam AR. Comprehensive Profiling of Translation Initiation in Influenza Virus Infected Cells. *PLoS Pathog* (2019) 15(1):e1007518. doi: 10.1371/journal.ppat.1007518
- Li S. Regulation of Ribosomal Proteins on Viral Infection. *Cells* (2019) 8(5):508. doi: 10.3390/cells8050508



28. Cros J, Cagnard N, Woollard K, Patey N, Zhang SY, Senechal B, et al. Human CD14dim Monocytes Patrol and Sense Nucleic Acids and Viruses via TLR7 and TLR8 Receptors. *Immunity* (2010) 33(3):375–86. doi: 10.1016/j.immuni.2010.08.012
29. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: Single-Cell Regulatory Network Inference and Clustering. *Nat Methods* (2017) 14(11):1083–6. doi: 10.1038/nmeth.4463
30. Allen EK, Randolph AG, Bhangale T, Dogra P, Ohlson M, Oshansky CM, et al. SNP-Mediated Disruption of CTCF Binding at the IFITM3 Promoter Is Associated With Risk of Severe Influenza in Humans. *Nat Med* (2017) 23(8):975. doi: 10.1038/nm.4370
31. Zhang Q. Human Genetics of Life-Threatening Influenza Pneumonitis. *Hum Genet* (2020) 139(6-7):941–8. doi: 10.1007/s00439-019-02108-3
32. Ciancanelli MJ, Huang SX, Luthra P, Garner H, Itan Y, Volpi S, et al. Infectious Disease. Life-Threatening Influenza and Impaired Interferon Amplification in Human IRF7 Deficiency. *Science* (2015) 348(6233):448–53. doi: 10.1126/science.aaa1578
33. Panthu B, Terrier O, Carron C, Traversier A, Corbin A, Balvay L, et al. The NS1 Protein From Influenza Virus Stimulates Translation Initiation by Enhancing Ribosome Recruitment to mRNAs. *J Mol Biol* (2017) 429(21):3334–52. doi: 10.1016/j.jmb.2017.04.007
34. Wang C, Forst CV, Chou TW, Geber A, Wang M, Hamou W, et al. Cell-To-Cell Variation in Defective Virus Expression and Effects on Host Responses During Influenza Virus Infection. *mBio* (2020) 11(1):e02880–19. doi: 10.1128/mBio.02880-19
35. Steurman Y, Cohen M, Peshes-Yaloz N, Valadarsky L, Cohn O, David E, et al. Dissection of Influenza Infection In Vivo by Single-Cell RNA Sequencing. *Cell Syst* (2018) 6(6):679–91.e4. doi: 10.1016/j.cels.2018.05.008
36. Russell AB, Trapnell C, Bloom JD. Extreme Heterogeneity of Influenza Virus Infection in Single Cells. *eLife* (2018) 7:e32303. doi: 10.7554/eLife.32303
37. Sun J, Vera JC, Drnevich J, Lin YT, Ke R, Brooke CB. Single Cell Heterogeneity in Influenza A Virus Gene Expression Shapes the Innate Antiviral Response to Infection. *PLoS Pathog* (2020) 16(7):e1008671. doi: 10.1371/journal.ppat.1008671
38. Ramos I, Smith G, Ruf-Zamojski F, Martínez-Romero C, Fribourg M, Carbajal EA, et al. Innate Immune Response to Influenza Virus at Single-Cell Resolution in Human Epithelial Cells Revealed Paracrine Induction of Interferon Lambda 1. *J Virol* (2019) 93(20):e00559–19. doi: 10.1128/JVI.00559-19
39. Russell AB, Elshina E, Kowalsky JR, Te Velthuis AJW, Bloom JD. Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity. *J Virol* (2019) 93(14):e00500–19. doi: 10.1128/JVI.00500-19
40. Kudo E, Song E, Yockey LJ, Rakib T, Wong PW, Homer RJ, et al. Low Ambient Humidity Impairs Barrier Function and Innate Resistance Against Influenza Infection. *Proc Natl Acad Sci USA* (2019) 116(22):10905–10. doi: 10.1073/pnas.1902840116
41. Cao Y, Guo Z, Vangala P, Donnard E, Liu P, McDonel P, et al. Single-Cell Analysis of Upper Airway Cells Reveals Host-Viral Dynamics in Influenza Infected Adults. *bioRxiv* (2020) 2020.04.15.042978. doi: 10.1101/2020.04.15.042978
42. Appleby LJ, Nausch N, Midzi N, Mduluzi T, Allen JE, Mutapi F. Sources of Heterogeneity in Human Monocyte Subsets. *Immunol Lett* (2013) 152(1):32–41. doi: 10.1016/j.imlet.2013.03.004
43. Randolph H, Mu Z, Fiege J, Thielen B, Grenier J, Cobb M, et al. Single-Cell RNA-Sequencing Reveals Pervasive But Highly Cell Type-Specific Genetic Ancestry Effects on the Response to Viral Infection. *bioRxiv* (2020), 2020.12.21.423830. doi: 10.1101/2020.12.21.423830
44. Cole SL, Dunning J, Kok WL, Benam KH, Benlahrech A, Repapi E, et al. M1-Like Monocytes Are a Major Immunological Determinant of Severity in Previously Healthy Adults With Life-Threatening Influenza. *JCI Insight* (2017) 2(7):e91868. doi: 10.1172/jci.insight.91868
45. Zhou YH, Fu B, Zheng X, Wang D, Zhao C, Qi Y, et al. Pathogenic T Cells and Inflammatory Monocytes Incite Inflammatory Storm in Severe COVID-19 Patients. *Natl Sci Rev* (2020) 7(6):998–1002. doi: 10.1093/nsr/nwaa041
46. Chandrasekhar R, Sloan C, Mitchell E, Ndi D, Alden N, Thomas A, et al. Social Determinants of Influenza Hospitalization in the United States. *Influenza Other Respir Viruses* (2017) 11(6):479–88. doi: 10.1111/irv.12483
47. Hadler JL, Yousey-Hindes K, Perez A, Anderson EJ, Bargsten M, Bohm SR, et al. Influenza-Related Hospitalizations and Poverty Levels - United States, 2010–2012. *MMWR Morb Mortal Wkly Rep* (2016) 65(5):101–5. doi: 10.15585/mmwr.mm6505a1
48. Shelton JF, Shastri AJ, Ye C, Weldon CH, Filshie-Somnez T, Coker D, et al. Trans-Ethnic Analysis Reveals Genetic and non-Genetic Associations With COVID-19 Susceptibility and Severity. *medRxiv* (2020) 2020.09.04.20188318. doi: 10.1101/2020.09.04.20188318
49. Center for Disease Control and Prevention. *Risk for COVID-19 Infection, Hospitalization, and Death By Race/Ethnicity* (2021). Available at: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html> (Accessed August 30, 2021).
50. 10xgenomics. *Chromium Single Cell 3' Reagent Kits User Guide (V3 Chemistry)* (2020). Available at: <https://support.10xgenomics.com/permalink/OAXXHQmeWa60IWK66uqSo> (Accessed August 30, 2021).
51. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An Integrated Map of Genetic Variation From 1,092 Human Genomes. *Nature* (2012) 491(7422):56–65. doi: 10.1038/nature11632
52. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat Commun* (2017) 8:14049. doi: 10.1038/ncomms14049
53. Heaton H, Talman AM, Knights A, Imaz M, Gaffney DJ, Durbin R, et al. Souporecell: Robust Clustering of Single-Cell RNA-Seq Data by Genotype Without Reference Genotypes. *Nat Methods* (2020) 17(6):615–20. doi: 10.1038/s41592-020-0820-1
54. Lun AT, McCarthy DJ, Marioni JC. A Step-by-Step Workflow for Low-Level Analysis of Single-Cell RNA-Seq Data With Bioconductor. *F1000Res* (2016) 5:2122. doi: 10.12688/f1000research.9501.2
55. P Pons and M Latapy eds. *Computing Communities in Large Networks Using Random Walks. Computer and Information Sciences - ISICIS 2005* Vol. 2005. Berlin, Heidelberg: Springer Berlin Heidelberg (2005).
56. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-Based Analysis of Lung Single-Cell Sequencing Reveals a Transitional Proinflammatory Macrophage. *Nat Immunol* (2019) 20(2):163–72. doi: 10.1038/s41590-018-0276-y
57. Benaglia T, Chauveau D, Hunter DR, Young DS. Mixtools: An R Package for Analyzing Mixture Models. *J Stat Software* (2010) 1(6):2009. doi: 10.18637/jss.v032.i0
58. Langfelder P, Zhang B, Horvath S. Defining Clusters From a Hierarchical Cluster Tree: The Dynamic Tree Cut Package for R. *Bioinformatics* (2008) 24(5):719–20. doi: 10.1093/bioinformatics/btm563
59. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene Ontology Analysis for RNA-Seq: Accounting for Selection Bias. *Genome Biol* (2010) 11(2):R14. doi: 10.1186/gb-2010-11-2-r14
60. *FlowJo™ Software, Version 10.6.1*. Ashland, OR: Becton, Dickinson and Company (2019).
61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* (2013) 29(1):15–21. doi: 10.1093/bioinformatics/bts635
62. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, et al. Qualimap: Evaluating Next-Generation Sequencing Alignment Data. *Bioinformatics* (2012) 28(20):2678–9. doi: 10.1093/bioinformatics/bts503
63. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data. *Bioinformatics* (2016) 32(2):292–4. doi: 10.1093/bioinformatics/btv566
64. Perteau M, Perteau GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie Enables Improved Reconstruction of a Transcriptome From RNA-Seq Reads. *Nat Biotechnol* (2015) 33(3):290–5. doi: 10.1038/nbt.3122
65. Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. Finding the Active Genes in Deep RNA-Seq Gene Expression Studies. *BMC Genomics* (2013) 14:778. doi: 10.1186/1471-2164-14-778
66. Johnson WE, Li C, Rabinovic A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics (Oxford England)* (2007) 8(1):118–27. doi: 10.1093/biostatistics/kxj037
67. Gong T, Szustakowski JD. DeconRNASeq: A Statistical Framework for Deconvolution of Heterogeneous Tissue Samples Based on mRNA-Seq

Data. *Bioinformatics* (2013) 29(8):1083–5. doi: 10.1093/bioinformatics/btt090

**Conflict of Interest:** JP was employed by DIACCURATE.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 O'Neill, Quach, Pothlichet, Aquino, Bisiaux, Zidane, Deschamps, Libri, Hasan, Zhang, Zhang, Matuozzo, Cobat, Abel, Casanova, Naffakh, Rotival and Quintana-Murci. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.