

RESEARCH

Open Access



# Pro-inflammatory feedback loops define immune responses to pathogenic *Lentivirus* infection

Aaron J. Wilk<sup>1,2,3</sup>, Joshua O. Marceau<sup>4</sup>, Samuel W. Kazer<sup>5,6,7,8,9</sup>, Ira Fleming<sup>5,6,7,8,9</sup>, Vincent N. Miao<sup>5,7,8,9,10</sup>, Jennyfer Galvez-Reyes<sup>5,6,7,8,9</sup>, Jason T. Kimata<sup>11</sup>, Alex K. Shalek<sup>5,6,7,8,9</sup>, Susan Holmes<sup>12</sup>, Julie Overbaugh<sup>4</sup> and Catherine A. Blish<sup>1,2,3,13\*</sup> 

## Abstract

**Background** The *Lentivirus* human immunodeficiency virus (HIV) causes chronic inflammation and AIDS in humans, with variable rates of disease progression between individuals driven by both host and viral factors. Similarly, simian lentiviruses vary in their pathogenicity based on characteristics of both the host species and the virus strain, yet the immune underpinnings that drive differential *Lentivirus* pathogenicity remain incompletely understood.

**Methods** We profile immune responses in a unique model of differential lentiviral pathogenicity where pig-tailed macaques are infected with highly genetically similar variants of SIV that differ in virulence. We apply longitudinal single-cell transcriptomics to this cohort, along with single-cell resolution cell-cell communication techniques, to understand the immune mechanisms underlying lentiviral pathogenicity.

**Results** Compared to a minimally pathogenic lentiviral variant, infection with a highly pathogenic variant results in a more delayed, broad, and sustained activation of inflammatory pathways, including an extensive global interferon signature. Conversely, individual cells infected with highly pathogenic *Lentivirus* upregulated fewer interferon-stimulated genes at a lower magnitude, indicating that highly pathogenic *Lentivirus* has evolved to partially escape from interferon responses. Further, we identify *CXCL10* and *CXCL16* as important molecular drivers of inflammatory pathways specifically in response to highly pathogenic *Lentivirus* infection. Immune responses to highly pathogenic *Lentivirus* infection are characterized by amplifying regulatory circuits of pro-inflammatory cytokines with dense longitudinal connectivity.

**Conclusions** Our work presents a model of lentiviral pathogenicity where failures in early viral control mechanisms lead to delayed, sustained, and amplifying pro-inflammatory circuits, which in turn drives disease progression.

**Keywords** Lentiviruses, Single-cell transcriptomics, Cell-cell communication

\*Correspondence:

Catherine A. Blish  
cblish@stanford.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Since the beginning of the HIV/AIDS pandemic, HIV has caused over 36 million deaths worldwide and remains a significant cause of morbidity and mortality globally ([www.who.int](http://www.who.int)). Chronic inflammation and CD4<sup>+</sup> T-cell loss are central to the pathogenesis of HIV [1–5]. In contrast, HIV's simian counterparts, collectively referred to as simian immunodeficiency viruses (SIV), are highly variable in pathogenicity depending on both host genetic and virus strain factors, with many SIV strains being non-pathogenic in their natural hosts [3–5]. One of the most long-standing questions in lentiviral biology concerns the mechanistic basis for these starkly variable outcomes.

Differential immune responses are thought to be a primary driver of divergent outcomes in lentiviral infection. For example, higher levels of C-reactive protein and IL-6 independently predict development of opportunistic infections in chronic HIV [6, 7]. Additionally, increased soluble CD14 levels, which reflect LPS-induced monocyte activation, predict increased mortality in HIV-infected individuals [8, 9]. Analysis of immune responses to HIV-2, which is associated with a longer asymptomatic phase and slower progression to AIDS than HIV-1, has also shown less persistent T-cell activation [10] and reduced type I IFN responses [11] in HIV-2-infected individuals relative to HIV-1. Further clues to differential lentiviral pathogenicity can be found in the features of simian precursors to HIV-1 that are thought to have conferred enhanced pathogenicity relative to SIV. For example, in simian precursors of HIV-1, the viral protein Nef has decreased activity in downregulating expression of CD3 [12]. By promoting a greater degree of T-cell activation upon infection, this modulation in Nef activity is thought to be a mechanism by which HIV-1 can cause chronic immune activation [13, 14].

One useful model for differential lentiviral pathogenicity comes from the study of emerging pathogenic variants of SIV where specific highly related variants were tested and shown to drive different clinical outcomes. It has long been known that genetic variants arise over the course of *Lentivirus* infection that are antigenically and phenotypically distinct from the founder viruses present early in infection [15]. For example, in pig-tailed macaques (*M. nemestrina*), genetic variants that evolve in the late stages of infection are highly cytopathic and rapidly replicating, unlike their precursors. By infecting immunocompetent pig-tailed macaques with a SIV clone that has properties similar to transmitted forms of HIV, SIVmneCL8 (hereafter CL8), or a late-stage genetic variant that evolved from CL8 (SIVmne170, hereafter 170), Kimata et al. showed that 170 was significantly more highly pathogenic than CL8 in pig-tailed macaques,

thereby demonstrating that late-stage genetic variants were capable of directly driving disease progression [16].

These SIV variants that are closely related but show variable virulence thus represent a useful model to explore differential lentiviral pathogenicity within a host to define host virus-host interactions that drive pathogenesis. While the role of viral genetic determinants as the proximal cause of this differential pathogenicity has been described, the distal immune mechanisms resulting in these distinct outcomes remain undefined. Here, we performed longitudinal single-cell transcriptomics on peripheral blood mononuclear cells (PBMCs) from macaques infected either with CL8 or 170. We reveal broad and sustained pro-inflammatory responses as a defining feature of pathogenic lentiviral infection with the highly virulent variant, nominate potential molecular drivers of these pathways, and describe longitudinal cell-cell communication (CCC) networks distinguishing responses between non-pathogenic vs. pathogenic lentiviral infection. Collectively, our work provides a model for understanding beneficial vs. deleterious immune responses to *Lentivirus* infection.

## Methods

### Sample collection and storage

All PBMC samples profiled in this study were processed and collected as a part of a study by Kimata et al. [16]. All animal work was performed as part of the Kimata et al. study [16], where all animals were maintained and cared for in accordance with the guidelines of the AAALAC and the Animal Care and Use Committee of the University of Washington [16]. No additional animal work was performed for this study. Briefly, juvenile pig-tailed macaques (*Macaca nemestrina*) were inoculated intravenously with equal amounts of SIVmneCL8 or SIVmne170. Serial samples of peripheral blood, serum, and plasma were collected prior to and weekly after inoculation to assess animal health, CD4<sup>+</sup> T-cell count, and viral load. Viral load measurements reported in this study were determined by Kimata et al. [16] by Chiron branch DNA (bDNA) assay. PBMCs were isolated by density centrifugation and cryopreserved. Samples from two animals infected with SIVmneCL8 and two animals infected with SIVmne170 were analyzed. From each animal, a pre-infection sample along with eight post-infection samples (from weeks 1, 2, 3, 4, 6, 8, 12, and 20 post-infection) were analyzed.

### Sample thawing and processing for transcriptomic analysis

PBMCs were thawed at 37 °C in complete RPMI-1640 media (supplemented with 10% FBS, L-glutamine, and penicillin-streptomycin-amphotericin; RP10) containing benzonase (EMD Millipore). Cells were then counted and

diluted in RP10 to a concentration of 75,000 live cells/mL. Cell viability was determined by trypan blue staining and is reported in Additional file 1: Fig. S1. Unless otherwise noted, all centrifugations were performed at 300  $\times$  g for 3 min. We observed that samples from both 170-infected animals at 1 week post-infection had extremely low viability post-thaw (18% and 8%, respectively). To attempt to rescue cell viability in these, we applied two strategies for live cell enrichment. First, we performed soft centrifugation on these samples, at 100  $\times$  g for 1 min, and discarded the supernatant. Next, we applied the Dead Cell Removal Kit (catalog no. 130-090-101; Miltenyi Biotec) to the remaining cell pellet. This process improved viability to 88% and 80%, respectively, and we proceeded with downstream transcriptomic profiling. A table listing sample thawing and processing batches is shown as Additional file 1: Table S1.

#### scRNA-seq by Seq-Well

The Seq-Well platform for scRNA-seq was utilized as described previously [17–19]. A total of 200  $\mu$ L of the cell suspension of thawed PBMCs at 75,000 live cells/mL (15,000 cells) was loaded onto Seq-Well arrays preloaded with mRNA capture beads (ChemGenes). Following four washes with serum-free RPMI-1640 to remove serum, the arrays were sealed with a polycarbonate membrane (pore size of 0.01  $\mu$ m) for 30 min at 37 °C. Next, arrays were placed in lysis buffer, transcripts hybridized to the mRNA capture beads, and beads recovered from the arrays and pooled for downstream processing. Immediately after bead recovery, mRNA transcripts were reverse transcribed using Maxima H-RT (Thermo Fisher EPO0753) in a template-switching-based RACE reaction, excess unhybridized bead-conjugated oligonucleotides removed with exonuclease I (NEB M0293L), and second-strand synthesis performed with Klenow fragment (NEB M0212L) to enhance transcript recovery in the event of failed template switching [19]. Whole transcriptome amplification (WTA) was performed with KAPA HiFi PCR Mastermix (Kapa Biosystems KK2602) using approximately 6000 beads per 50- $\mu$ L reaction volume. Resulting libraries were then pooled in sets of 6 (approximately 36,000 beads per pool) and products purified by Agencourt AMPure XP beads (Beckman Coulter, A63881) with a 0.6 $\times$  volume wash followed by a 0.8 $\times$  volume wash. Quality and concentration of WTA products were determined using an Agilent Fragment Analyzer (Stanford Functional Genomics Facility), with a mean product size of > 800 bp and a nonexistent primer peak indicating successful preparation. Library preparation was performed with a Nextera XT DNA library preparation kit (Illumina FC-131-1096) with 1 ng of pooled library using dual-index primers. Tagmented

and amplified libraries were again purified by Agencourt AMPure XP beads with a 0.6 $\times$  volume wash followed by a 1.0 $\times$  volume wash and quality and concentration determined by fragment analysis. Libraries between 400 and 1000 bp with no primer peaks were considered successful and pooled for sequencing. Sequencing was performed on a NovaSeq 6000 instrument (Illumina; Chan Zuckerberg Biohub). The read structure was paired end with read 1 beginning from a custom read 1 primer [17] containing a 12-bp cell barcode and an 8-bp unique molecular identifier (UMI) and with read 2 containing 50 bp of mRNA sequence.

#### Bulk RNA-sequencing

PBMCs were removed from liquid nitrogen and thawed on ice. A subset of cells were counted and removed for Seq-Well analysis, while the remaining cells were pelleted and immediately disrupted with Buffer RLT (Qiagen). Extraction and purification of total RNA from bulk PBMCs were conducted using Qiagen RNeasy plus kit (no. 74034). All extractions included genomic DNA removal using Qiagen gDNA Eliminator Spin Columns. The concentration of RNA was quantified using a NanoDrop 2000 Spectrophotometer (ThermoFisher). RNA quality was assessed using RIN values obtained from Agilent 4200 TapeStation analysis. Reverse transcription and tagmentation were conducted using the SMART-Seq v4 (Takara) and Nextera XT library prep kit (Illumina) with an input of 10 ng of RNA per sample. After cDNA synthesis and library preparation, the samples were sequenced using Illumina NovaSeq SP with paired-end 50-bp read length, yielding approximately 30-M reads per sample.

#### Processing and analysis of bulk RNA-sequencing data

To support RNA-seq analysis, a *Macaca nemestrina* genome reference was prepared by downloading reference sequence and gene annotations from Ensembl Mnem\_1.0 release 98 in October 2019. For each RNA-seq sample, sequences that failed default Illumina RTA quality checks were removed. The remaining “PF passing” reads were aligned to the Mnem\_1\_0\_98 reference using the STAR 2.7.1a splice-aware aligner in single-sample two-pass mode [20]. For each sample, per-gene expression was calculated using featureCounts from the Subread 1.6.5 package [21].

Gene-level counts from all samples were merged, and differential expression was assessed with the edgeR Bioconductor package [22]. Briefly, genes with low expression across all sample groups were excluded using filterByExpr, TMM normalization was applied to account for composition biases among samples, and differential expression was assessed using the quasi-likelihood *F*-test

methods. Differentially expressed genes between CL8- and 170-infected animals were considered significant when  $FDR < 0.05$  and  $|\log_2FC| > 1$ . Volcano plots were generated in the R Programming Environment using the edgeR differential expression summaries.

#### Alignment and quality control of single-cell RNA sequencing data

Sequencing reads were aligned and count matrices assembled using STAR [20] and dropEst [23], respectively. Briefly, the mRNA reads in read 2 demultiplexed FASTQ files were tagged with the cell barcode and UMI for the corresponding read in the read 1 FASTQ file using the dropTag function of dropEst. Next, reads were aligned with STAR using the *M. nemestrina* reference genome Mnem\_1.0 that included the sequences of the SIVMneCL8 and SIVMne170 genomes for alignment. Given the similarity between the SIVMneCL8 and SIVMne170 genomes, any read aligning to either SIV genome sequence was annotated as an SIV viral read. To identify SIV-infected cells, we first used Cellbender [24, 25] to estimate the ambient RNA composition for each sample and generate an ambient-corrected count matrix. Any cell with  $> 0$  UMIs aligned to the SIV genome in the ambient-corrected matrix was then considered SIV<sup>+</sup>. Count matrices were built from resulting BAM files using dropEst [23]. We noted that several Ensembl gene IDs had annotations available that were not present in the GTF from the latest Mnem\_1.0 release (Additional file 1: Fig. S1). We transferred these additional annotations to the count matrices using the R package biomaRt [26]. Additional transferred annotations are listed in Additional file 1: Table S2. Cells that had fewer than 1000 UMIs or greater than 15,000 UMIs, as well as cells that contained greater than 20% of reads from rRNA genes (*RNA18S5* or *RNA28S5*) or greater than 0.1% of reads from one of the 6 annotated mitochondrial genes in the *M. nemestrina* genome, were considered low quality and removed from further analysis. To remove putative multiplets (where more than one cell may have loaded into a given well on an array), cells that expressed more than 75 genes per 100 UMIs were also filtered out. Genes that were expressed in fewer than 10 cells were removed from the final count matrix.

#### Pre-processing of single-cell RNA sequencing data

The R package Seurat [27–29] was used for data scaling, transformation, clustering, dimensionality reduction, differential expression analysis, and most visualizations. Data were scaled and transformed and variable genes identified using the SCTransform() function and linear regression performed to remove unwanted variation due to cell quality (% mitochondrial reads, % rRNA reads).

PCA was performed using the 3000 most highly variable genes. The first 50 principal components (PCs) were used to perform dimensionality reduction by UMAP [30, 31] and to construct a shared nearest neighbor graph (SNN; FindNeighbors()), which was used to cluster the dataset (FindClusters()).

#### Automated annotation of granular cell types by Seurat v4

We used the multimodal (whole transcriptome plus 228 cell surface proteins) PBMC dataset published by Hao et al. [29] as a reference for cell type annotation. We first scaled both the transcriptome and protein assays and ran PCA on both modalities. Next, we found multimodal neighbors between the modalities via weighted nearest neighbor (WNN) analysis, which learns the relative utility of each data modality in each cell. Supervised PCA (SPCA) was then run on the WNN SNN graph, which seeks to capture a linear transformation that maximizes its dependency to the WNN SNN graph. These SPCA-reduced dimensions were then used for identification of anchors between the reference and query datasets as previously described [27]. Finally, these anchors were used to transfer reference cell type annotations to the query dataset. As the Hao et al. reference dataset used for annotation is a human dataset, we confirmed the validity of the transferred annotations by calculating differentially expressed genes for each cell type annotation using Seurat's implementation of the Wilcoxon rank-sum test (FindMarkers()) and comparing those markers to known cell type-specific genes from previous datasets [32–37]. Cell type annotations transferred with this strategy matched biological expectations and overlapped with results from graph-based clustering (Additional file 1: Fig. S1).

#### Calculation of transcriptomic perturbation score

To prioritize analysis of cell types of interest, we calculated a perturbation score for each cell type of each sample as previously described [29, 38]. This perturbation score is motivated by the observation that the statistical significance of per-gene differential expression tests is strongly influenced by the number of cells in each cluster or cell type. To overcome this, we first identified a set of genes for each cell type that showed evidence of differential expression ( $p$ -value  $< 0.1$ ) between a post-infection and pre-infection time point. Next, we defined the global perturbation vector as the average log-fold change of each DEG relative to pre-infection samples normalized to length 1. Finally, we projected the transcriptome of each sample onto this vector and defined the perturbation score as the absolute value of the magnitude of this projection. This approach enables prioritization of cell type perturbations when comparing cell types of different abundances.

### Gene module scoring analysis

The Seurat function `AddModuleScore()` was used to score single cells by expression of a list of genes of interest. This function calculates a module score by comparing the expression level of an individual query gene to other randomly selected control genes expressed at similar levels to the query genes and is therefore robust to scoring modules containing both lowly and highly expressed genes, as well as to scoring cells with different sequencing depth. Gene lists for module scoring were generated through gene module discovery. A previously published list of ISGs was used to define the interferon signature score (Additional file 1: Table S3) [39].

### Differential expression testing

We used Seurat's implementation of differential expression testing in `FindMarkers()` to identify DE genes and DE ligand-receptor pairs. Unless otherwise specified, testing for DE genes was performed via Seurat's implementation of the Wilcoxon rank-sum test, which reports two-sided  $p$ -values corrected for multiple hypothesis testing by Bonferroni's correction. For single-cell resolution CCC analysis with Scriabin, ligand-receptor pair activity values are not independent observations, and therefore, a receiver operating characteristic curve (ROC) test was used for differential expression testing. In this test, a classifier is built on each gene which is used to classify two groups of cells being compared. An AUC value of 1 indicates that the expression values for that gene can perfectly classify the two cell groups.

Differential expression testing of ISGs was performed on a per animal per cell type basis: for each sample, DE genes were calculated between each post-infection time point relative to the respective pre-infection time point. Cumulative  $\log(\text{fold-changes})$  represent the sum of the fold-changes for each post-infection time point at which the gene was significantly ( $p$ -value  $< 0.05$ ) upregulated.

### Gene module discovery with WGCNA

Unless otherwise noted, gene module analysis was performed as previously described [18]. This process was performed separately for each major cell type from each animal. Briefly, principal components (PCs) of biological interest were first identified by examining the variance structure described by each PC and the genes contributing to each PC. PCs used for each WGCNA analysis are shown in Additional file 1: Table S4. Variable genes of biological interest were defined as the top and bottom 50 genes for each PC, and the input to WGCNA was subset to these genes.

Next, a signed genes covariance matrix is calculated for each sample:

$$s_{ij}^{\text{signed}} = 0.5 + 0.5\text{cor}(g_i, g_j),$$

Where  $g_i, g_j$  are individual genes.  $s_{ij}^{\text{signed}}$  is next converted into an adjacency matrix via soft thresholding:

$$a_{ij} = (s_{ij}^{\text{signed}})^{\beta},$$

Where  $\beta$  is the soft power. Soft power is a user-defined parameter that is recommended to be the lowest value that results in a scale-free topology model fit of  $> 0.8$  [40]. Soft power parameters used for analysis are shown in Additional file 1: Table S4. Next, this adjacency matrix is converted into a TOM as described [41]. Next, this TOM was hierarchically clustered and `cutreeDynamic` used to generate modules of highly correlated genes with a minimum module size of 10. Similar modules were merged using a dissimilarity threshold of 0.5. Subsequently, the correlation structure of each module was tested using a permutation test, comparing the correlation of randomly generated modules to the true module. A module was considered significantly correlated if  $< 500$  of 10,000 one-sided Mann-Whitney  $U$ -tests were significant at the  $p < 0.05$  level.

Single cells were then scored for expression of each module as described above. Finally, we tested for modules with significant variation in expression over time. Since testing for differences in distribution is highly sensitive to sample size, we performed iterations of a Kruskal-Wallis test for expression of each module relative to time point on 10,000 equally sized cell subsets. Modules were considered significantly temporally variable if  $< 500$  of 10,000 Kruskal-Wallis tests were significant at the  $p < 0.05$  level.

### Analysis of CCC with Scriabin

Scriabin [42] was used for analysis of CCC at single-cell resolution as previously described. In brief, the center of Scriabin's workflow is the generation of a cell-cell interaction matrix (CCIM) that captures the full structure of CCC phenotypes in the dataset. For a dataset of  $n$  cells and  $k$  possible ligand-receptor pairs that can be used for CCC, the CCIM is a matrix of  $n \times n$  columns of single cell-cell pairs by  $k$  ligand-receptor pairs where each matrix element represents the predicted activity of a given ligand-receptor pair between a cell-cell pair.

### Generation of summarized interaction graphs

Because the CCIM scales exponentially with dataset size, it was impractical to calculate a CCIM for all cell-cell

pairs  $n_i n_j$ . To address this problem, a summarized cell-cell interaction graph  $\mathcal{S}$  was built in lieu of the CCIM where an element of  $\mathcal{S}$  is the sum of all potential ligand-receptor activities between a cell-cell pair. A matrix  $\mathcal{S}$  was built for each time point.

#### Dataset binning for comparative CCC analyses

Conceptually, comparing  $\mathcal{S}$  from multiple samples requires that cells from different samples share a set of labels or annotations denoting what cells represent the same identity. Each identity class to be compared then requires representation from each of the samples to be compared. This inter-dataset alignment was accomplished through Scriabin's binning workflow. The goal of binning is to assign each cell a bin identity so that  $\mathcal{S}$  from multiple samples can be summarized into equidimensional matrices based on shared bin identities.

First, a shared nearest neighbor (SNN) graph was constructed via FindNeighbors() to define connectivity between all cells to be compared. Next, mutual nearest neighbors (MNNs) were identified between all time points to be compared via Seurat's integration workflow (FindIntegrationAnchors()) [27]. Next, bin identities were assigned based on these inter-dataset MNN pairings and iteratively optimized for the SNN connectivity of cells within the same bin and for the number of time points represented within each bin. Finally, incompletely represented bins were merged together based on SNN connectivity. At the end of this process, each cell has a single assigned bin identity, where each bin contains cells from all samples to be compared.

Bins were then tested for statistical significance of connectivity structure using a permutation test. For each

$$V_{N_i N_j} = \left[ \sqrt{N_i^{l_1} \times (N_j^{r_1} + N_j^{a_{l_1}})}, \sqrt{N_i^{l_2} \times (N_j^{r_2} + N_j^{a_{l_2}})}, \dots, \sqrt{N_i^{l_n} \times (N_j^{r_n} + N_j^{a_{l_n}})} \right],$$

bin, random bins of the same size and number of cells per sample were generated iteratively (by default 10,000 times). The connectivity vector of the real bins was tested against each of the random bins by a one-sided Mann-Whitney  $U$ -test. If the bin failed 500 or more of these tests ( $p$ -value 0.05), it was considered non-significant, discarded, and merged with its most similar counterpart.

Finally, bin-bin pairs that were variable in their CCC magnitude over time were identified. For each bin-bin pair, a Kruskal-Wallis test was used to assess differences in the magnitude of CCC between cell-cell pairs from different time points. Cells from bin-bin pairs with a two-sided Kruskal-Wallis  $p$ -value  $< 0.05$  and within the top 75 percentile of Kruskal-Wallis test statistic were considered highly perturbed. Dunn's post hoc test was used to

identify the most frequently perturbed time point among highly variable bin-bin pairs. Cells from highly variable bins in cell type combinations of interest were then used to construct CCIMs.

#### Identification of biologically active ligand-receptor edges

To identify what ligands were predicted to be active in each receiver cell, we first calculated a per-cell gene signature of temporally variable genes using the package CelliD [43]. Next, potential ligands were defined as those ligands that were expressed by at least 2.5% of sender cells being analyzed. The ligand activity for each receiver cell was defined as the Pearson correlation coefficient between each cell's gene signature and the vector of ligand-to-target gene links for that ligand, as provided by the package NicheNet [44]. For each active ligand, target gene weights for each cell were defined as the ligand-target matrix regulatory score for the top 250 targets for each ligand that appear in a given cell's gene signature. We selected a Pearson coefficient threshold (by default 0.075) to define active ligands in each cell.

To identify statistically significant changes in ligand activity over time, we performed iterations of a Kruskal-Wallis test for expression of each ligand activity score relative to time point on 10,000 equally sized cell subsets. Ligand activities were considered significantly temporally variable if  $< 500$  of 10,000 Kruskal-Wallis tests were significant at the  $p < 0.05$  level.

#### CCIM construction

We defined the interaction vector  $V$  between sender cell  $N_i$  and receiver cell  $N_j$  as follows:

Where  $l_n, r_n$  represent a cognate ligand-receptor pair and  $a_{l_n}$  represents the predicted activity of an active ligand  $l_n$  in  $N_j$ . The CCIM was assembled by concatenating interaction vectors  $V_{N_i N_j}$  for all cell-cell pairs. Individual CCIMs were generated for each sample, so that cell-cell pairs were only generated between cells that could be biologically interacting.

#### Downstream analysis of CCIM

The CCIM can be treated analogously to the gene expression matrix and used for downstream analysis tasks like dimensionality reduction. To focus our downstream analysis on ligand-receptor pairs of biological interest, we subset the CCIM to ligand-receptor pairs where either the ligand or receptor appeared

within the top 1000 genes of the parent object. To remove additional sources of noise in downstream visualization, we also discarded cell-cell pairs where fewer than five ligand-receptor pairs were non-zero in activity. Next, we scaled this object by `ScaleData()`, found latent axes by `RunPCA()`, and used the top 15 PCs to embed the dataset in two dimensions using UMAP [31]. Neighbor graphs were constructed by `FindNeighbors()`, which were then clustered via modularity optimization graph-based clustering [45] as implemented by Seurat's `FindClusters()` [29].

### Identification of longitudinal CCC circuits

Longitudinal communication circuits were identified as previously described [42]. Briefly, a longitudinal CCC circuit is composed of  $S_1-L_1-R_1-S_2-L_2-R_2$ , where S are sender cells and R are receiver cells at time points 1 and 2, and where  $L_1$  is expressed by/sensed by  $S_1/R_1$  and  $L_2$  is expressed by/sensed by  $S_2/R_2$ . Ligand activities and target gene linkages used for circuit generation were calculated as described above. Genes defined as pro-inflammatory cytokine-encoding genes are shown in Additional file 1: Table S5.

### Visualization

Wrappers provided by Seurat were used to generate UMAP projection and dot plots. `ComplexHeatmap` was used to generate all heatmaps. Visualization functions available from `Scrublet` were used to visualize most cell-cell communication pathways. For all boxplots, features include the following: minimum whisker, 25th percentile —  $1.5 \times$  inter-quartile range (IQR) or the lowest value within; minimum box, 25th percentile; center, median; maximum box, 75th percentile; and maximum whisker, 75th percentile +  $1.5 \times$  IQR or greatest value within. Feature line plots are boxplots where consecutive time points from the same animal are connected via a straight line. Unless otherwise noted, all feature line plots contain a 95% confidence interval around each point.

## Results

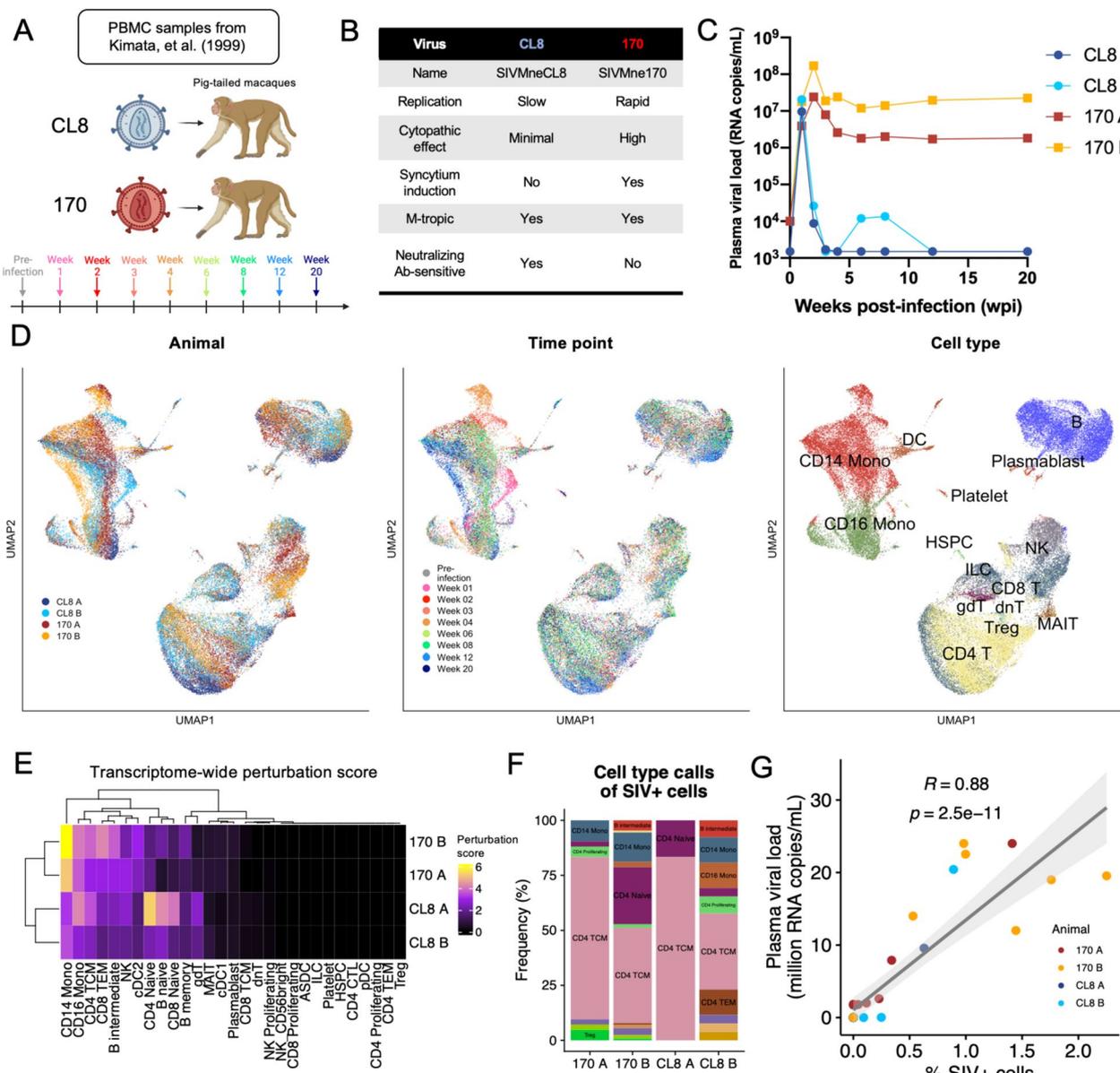
### Single-cell transcriptomics of immune responses to lentiviral pathogenic variants

To investigate the immune mechanisms that distinguish animals infected with variants of SIV that differ in their pathogenic properties, we performed Seq-Well-based [17, 19] massively parallel single-cell transcriptomic profiling on longitudinally sampled PBMCs from pig-tailed macaques infected with either CL8 ( $n = 2$ ; shades of blue) or 170 ( $n = 2$ ; shades of red). As it is known that features of initial pathology in the days and weeks immediately after infection can predict overall disease outcome and tissue damage [46, 47], we profiled PBMCs

from pre-infection and eight post-infection time points, focusing on the hyperacute and acute stages of infection (Fig. 1A). CL8 is a non-syncytium-inducing SIV with slow replication kinetics and minimal cytopathic effect; 170, conversely, is derived from an animal with late-stage infection with CL8 and is rapidly replicating, syncytium inducing, and induces high cytopathic effects (Fig. 1B). The cytopathicity and replication kinetics of 170 have previously been shown to be determined by mutations in Gag and Env [48], with canonical Env mutations typical of very-late-stage SIV [49]. Both viruses are macrophage tropic on infection of macaque monocyte/macrophage cultures *in vitro* (Fig. 1B). Compared to CL8-infected animals, animals infected with 170 have a peak viremia that is on average 6-fold higher ( $\sim 0.8$  log) and 1 week delayed and have a viral set point that is three orders of magnitude higher in subacute and chronic infection (Fig. 1C) [16]. One animal, 170 B, died from simian AIDS-related complications at 31-week post-infection.

In total, we sequenced 56,554 cells with an average of 1570 cells per animal per time point (Additional file 1: Fig. S1) [50], which we aligned to a *M. nemestrina* reference genome that contained the viral genome sequences of CL8 and 170 in order to detect cells infected with SIV. Samples from 170-infected animals at 1-week post-infection had low viability, and few high-quality cells were sequenced from these samples (Additional file 1: Fig. S1; see “Methods”). Next, we created a merged feature matrix of all profiled samples that we subjected to dimensionality reduction by uniform manifold approximation and projection (UMAP), graph-based clustering, and cell type annotation (see “Methods”). UMAP projections of the full dataset indicated regions of the gene expression manifold that were strongly separated between time points and between 170- and CL8-infected animals, particularly within the myeloid compartment (Fig. 1D). For example, we note a transcriptionally unique population of monocytes that is composed almost entirely of cells from both 170-infected animals at 4-week post-infection, as well as distinct populations of monocytes from both CL8-infected animals at 1-week post-infection (Fig. 1D, Additional file 1: Fig. S2; see Fig. 3). It is unlikely that this unique population of monocytes is related to a batch effect because other cell subsets from that time point, including T and B cells, appear well integrated in dimensionality reduction embeddings.

To formally quantify the degree of this transcriptional perturbation, we calculated a genome-wide perturbation score for each cell type relative to each animal's pre-infection sample (see “Methods”). This perturbation score is calculated by first identifying genes that display evidence of differential expression between pre- and post-infection time points, calculating the difference of



**Fig. 1** Single-cell transcriptomics of immune responses to SIV pathogenic variants. **A** Schematic of sample collection strategy. Pig-tailed macaques were infected with CL8 (shades of blue) or 170 (red/orange) and PBMCs sampled at the indicated time points. **B** Virological and pathological characteristics of SIV pathogenic variants. **C** Plasma viral loads of CL8- and 170-infected animals at the profiled time points. Animal 170 B died at 31 weeks post-infection from simian AIDS-related complications. **D** UMAP projections of full scRNA-seq dataset colored by animal (left), time point (middle), and coarse cell type annotation (right). **E** Heatmap of whole transcriptome perturbation score per cell type per animal, as described by [38]. **F** Bar plot depicting cell type proportions of cells with SIV-aligned reads. **G** Scatter plot depicting association between plasma viral load and the percentage of SIV<sup>+</sup> cells in the scRNA-seq dataset.

pseudobulk expression vectors of these genes between pre- and postinfection time points, and finally projecting the whole transcriptome of each animal onto this vector [38]. This score thereby represents the magnitude of longitudinal whole-transcriptome shifts in gene expression and reveals a strong longitudinal perturbation of CD14 monocytes in both 170-infected animals, as well as

perturbation of CD16 monocytes and CD8<sup>+</sup> T-cell subsets (Fig. 1E).

We next examined our ability to detect peripheral immune cells that were infected with SIV. The proportion of SIV<sup>+</sup> cells ranged from 0 to 2.5% per sample and were predominantly found in memory CD4<sup>+</sup> T cells and monocytes, matching biological expectations (Fig. 1F).

SIV<sup>+</sup> cells in CL8-infected animals were detected mainly at weeks 1–2 post-infection, while 170-infected cells were also present at later time points (Additional file 1: Fig. S1). The distribution of SIV UMIs in each infected cell was similar for CL8- and 170-infected cells (Additional file 1: Fig. S1). The proportion of SIV<sup>+</sup> cells correlated strongly with plasma viral load measurements (Pearson's  $r = 0.88$ ), suggesting the accurate identification of SIV<sup>+</sup> cells (Fig. 1G; discussed further in Fig. 2).

To provide complementary validation to our single-cell transcriptional data, we also performed bulk RNA sequencing of all animals from pre-infection through 4 weeks post-infection (Additional file 1: Fig. S3). This analysis revealed strong transcriptional perturbation in 170-infected animals that was most pronounced at 4 weeks post-infection, whereas transcriptional perturbation was largely resolved by 4 weeks post-infection in CL8-infected animals (Additional file 1: Fig. S3). Additionally, these transcriptional responses were largely congruent between animals infected with the same virus. While we had very poor recovery of viable cells at 1 week post-infection in 170-infected animals, we were able to generate high-quality bulk transcriptomic data from one 170-infected animal at 1 week post-infection. Compared to pre-infection, the 1 week post-infection sample showed strong upregulation of interferon (IFN)-stimulated genes (ISGs; including *GBP3*, *MX2*, *DDX58*, and *APOBEC3B*) and pro-inflammatory chemokines *CCL8* and *CXCL10*, as well as platelet markers *PPBP*, *ALAS2*, *CAVIN2*, and *PRR35*. This sample also showed downregulation of cell type-specific marker genes including *CCR7*, *CD8B*, *GZMK*, and *CD14*, suggestive of a composition shift with decreased monocytes and lymphocytes and increased platelets (Additional file 1: Fig. S3). Notably, statistical power to detect DEGs at 1 week post-infection in 170 infection is limited by having bulk RNA-seq data from a single animal at this time point. Future work will be necessary to validate these findings.

#### Differential responses to interferon in minimally pathogenic vs. highly pathogenic SIV infection

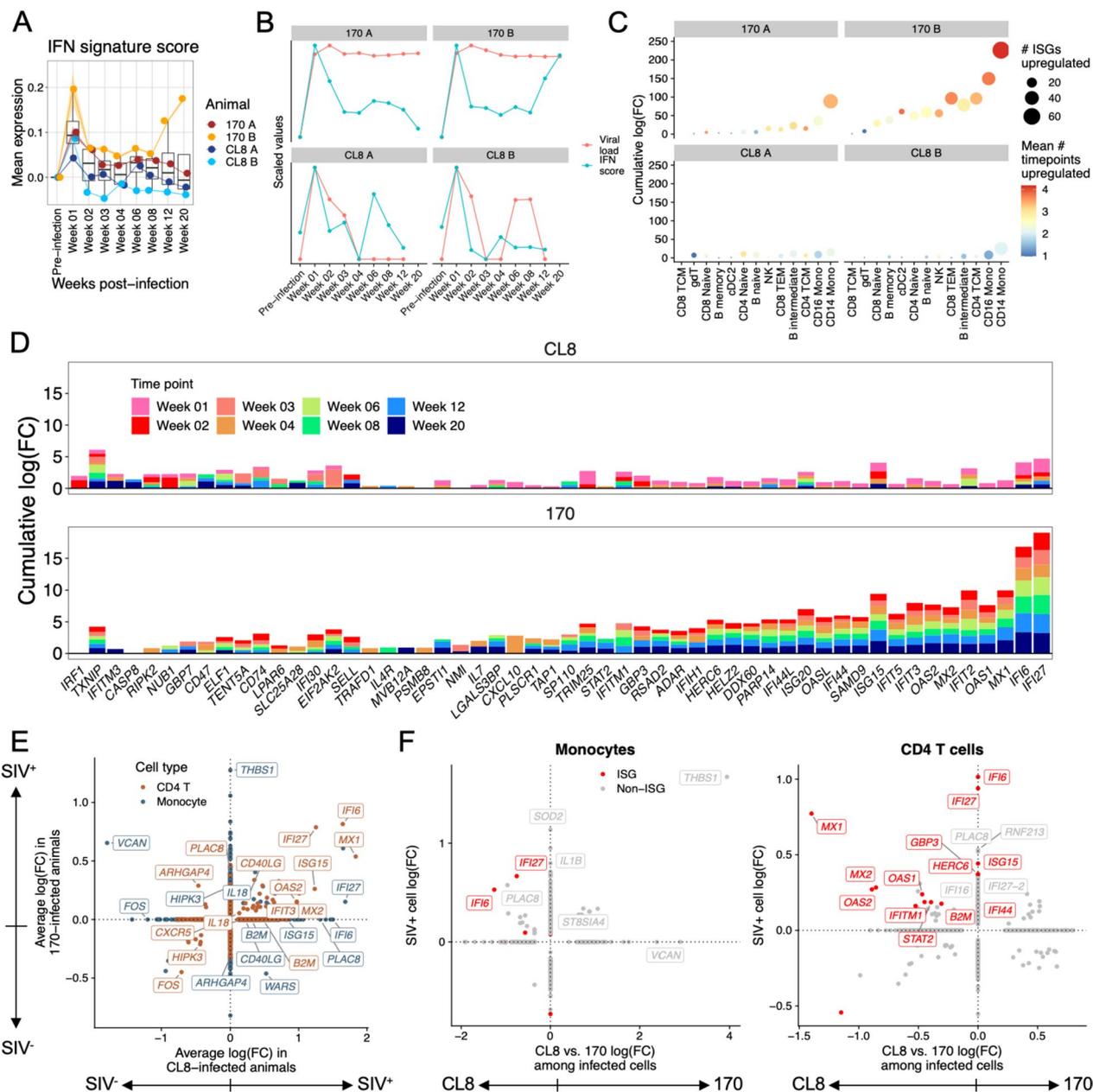
As ISGs act as important restriction factors in acute infection [51–53], we first analyzed how the longitudinal IFN response differed between CL8- and 170-infected animals. We established an IFN signature score composed of known ISGs (see “Methods”) and scored all cells in the dataset by expression of this signature. This revealed a kinetically dynamic IFN response that peaked in all animals at 1 week post-infection but stayed persistently elevated in 170-infected animals relative to CL8-infected animals (Fig. 2A). Additionally, in 170 B, the global IFN signature increased monotonically from 8

weeks post-infection preceding death at 31 weeks post-infection (Fig. 2A). Generally, the IFN signature was most concordant with viral load in both CL8 animals and highly discordant in 170 B, where the late increase in IFN signature is not paralleled by an increase in viral load (Fig. 2B). This may either be the reflection of robust infection in another tissue, such as the intestine, or be the result of non-lentiviral opportunistic infections or ongoing noninfectious inflammatory stimuli.

To examine potential sources of type I IFN, we analyzed the expression of the upstream regulator of IFN, *IRF7*, as type I IFN-encoding genes themselves frequently are not detectable at the RNA level [18, 54]. We found that global expression of *IRF7* generally correlated well with the global IFN signature, and that plasmacytoid dendritic cells (pDCs) were the most prolific source of *IRF7* expression, even when accounting for their relative scarcity (Additional file 1: Fig. S4). Additionally, in animal 170 B, we found that the late increase in IFN signature was accompanied by a shift of *IRF7* expression from pDCs to conventional DCs and monocytes (Additional file 1: Fig. S4), implicating these cells in sustaining IFN responses preceding death in this animal.

Next, we analyzed the breadth and duration of the IFN response. Both 170-infected animals upregulated substantially more ISGs than both CL8-infected animals, and these ISGs were upregulated at greater magnitude and at more time points in 170-infected animals (Fig. 2C). Monocytes were the cell type that upregulated the greatest magnitude of ISGs in 170-infected animals, but cell type distribution of ISG upregulation was more evenly distributed in CL8-infected animals (Fig. 2C). We also observed stark differences in which ISGs were expressed by 170- vs. CL8-infected animals. In 170-infected animals, the most upregulated ISGs were *IFI27*, *IFI6*, *MX1*, and *ISG15*, while in CL8-infected animals, *TXNIP* and *IFI30* were the most upregulated (Fig. 2D). Several ISGs, including *MX1*, *MX2*, and *OAS1*, were upregulated in 170-infected animals at many different time points but were generally only significantly upregulated in CL8 animals at weeks 1 and 2 post-infection (Fig. 2D). There were no cell type-specific patterns in ISG upregulation (Additional file 1: Fig. S5). The ISG response in animal 170 B diversified in the later time points profiled, with ISGs including *IFI44L*, *RSAD2*, and *DDX60* comprising a larger proportion of the IFN response relative to previous time points (Additional file 1: Fig. S5).

As ISGs can act as restriction factors during infection, we next examined how gene expression profiles differed between infected vs. bystander cells in 170- vs. CL8-infected animals. In each animal, we calculated differentially expressed (DE) genes between SIV<sup>-</sup> and SIV<sup>+</sup>



**Fig. 2** Broad and sustained interferon signatures in pathogenic SIV infection. **A** Boxplot depicting mean expression of all ISGs over time across all profiled cells. **B** Correlation between global IFN signature and viral load in each animal. **C–D** For each animal, within each cell type, and at each non-baseline time point, differentially expressed ISGs were calculated relative to pre-infection cells. **C** For each ISG, we calculated the mean number of time points at which that ISG was significantly upregulated (color of points). For each cell type, we calculated the number of ISGs significantly upregulated at any time point (size of points) and the cumulative log(fold-change) in expression of all significantly upregulated ISGs (y-axis). **D** For each ISG, animal class (170 or CL8), and cell type, we calculated the cumulative log(fold-change) of significantly upregulated ISGs across all cell types. Only ISGs with a cumulative log(fold-change) of > 0.75 in at least one animal group are shown. **E** In the two SIV-infected cell types (monocytes and CD4<sup>+</sup> T cells), we calculated DEGs between SIV<sup>+</sup> cells relative to SIV<sup>-</sup> cells in CL8 animals (x-axis) and 170 animals (y-axis). **F** In monocytes (left) and CD4<sup>+</sup> T cells (right), we calculated DEGs between SIV<sup>+</sup> cells relative to SIV<sup>-</sup> cells (y-axis). Then, among the SIV<sup>+</sup> cells, we calculated DEGs between cells from 170-infected animals and cells from CL8-infected animals (x-axis). For **C–F**, all genes shown are significantly differentially expressed along at least one axis by Seurat’s implementation of the Wilcoxon rank-sum test

cells in the two cell types that were the targets of infection, CD4<sup>+</sup> T cells, and monocytes. In SIV<sup>+</sup> cells from all animals, we found upregulation of several ISGs in both cell types, consistent with autocrine IFN signaling (Fig. 2E). Generally, ISGs were more broadly upregulated in infected CD4<sup>+</sup> T cells compared to monocytes. For example, CD4<sup>+</sup> T cells, but not monocytes, upregulated *OAS1*, *OAS2*, and *DDX58* upon infection in all animals (Fig. 2E). We also observed broader ISG upregulation in infected cells from CL8-infected animals: in infected monocytes, *IFI6* and *ISG15* were only upregulated in CL8-infected animals, and *DDX60* was significantly downregulated in infected monocytes from 170-infected animals (Fig. 2E).

While this analysis indicates which genes are upregulated in infected vs. bystander cells, we next sought to analyze how the magnitude of this upregulation differs between 170- vs. CL8-infected cells. To address this question, we calculated DE genes between SIV<sup>+</sup> cells from CL8-infected animals and SIV<sup>+</sup> cells from 170-infected animals. In both monocytes and CD4<sup>+</sup> T cells, CL8-infected cells upregulated ISGs to a greater extent than 170-infected cells (Fig. 2F), suggesting deficient expression of ISGs upon infection with SIVMne170. Since 170-infected animals have more SIV-infected cells at later time points relative to CL8-infected animals, we also assessed if the DEG results between CL8- and 170-infected animals were driven by infection kinetics. However, we found that 170-infected cells from 4 weeks post-infection and later expressed slightly higher magnitudes of some ISGs relative to their hyperacute infection counterparts (Additional file 1: Fig. S5). This indicates that the lower expression of ISGs in 170-infected cells vs. CL8-infected cells (Fig. 2F) is not confounded by infection kinetics. Collectively, this analysis reveals that *global* IFN signatures are broader and more sustained in highly pathogenic lentiviral infection (Figs. 2A–D), but that minimally pathogenic *Lentivirus* induces broader and greater expression of ISGs *within infected cells* (Figs. 2E–F).

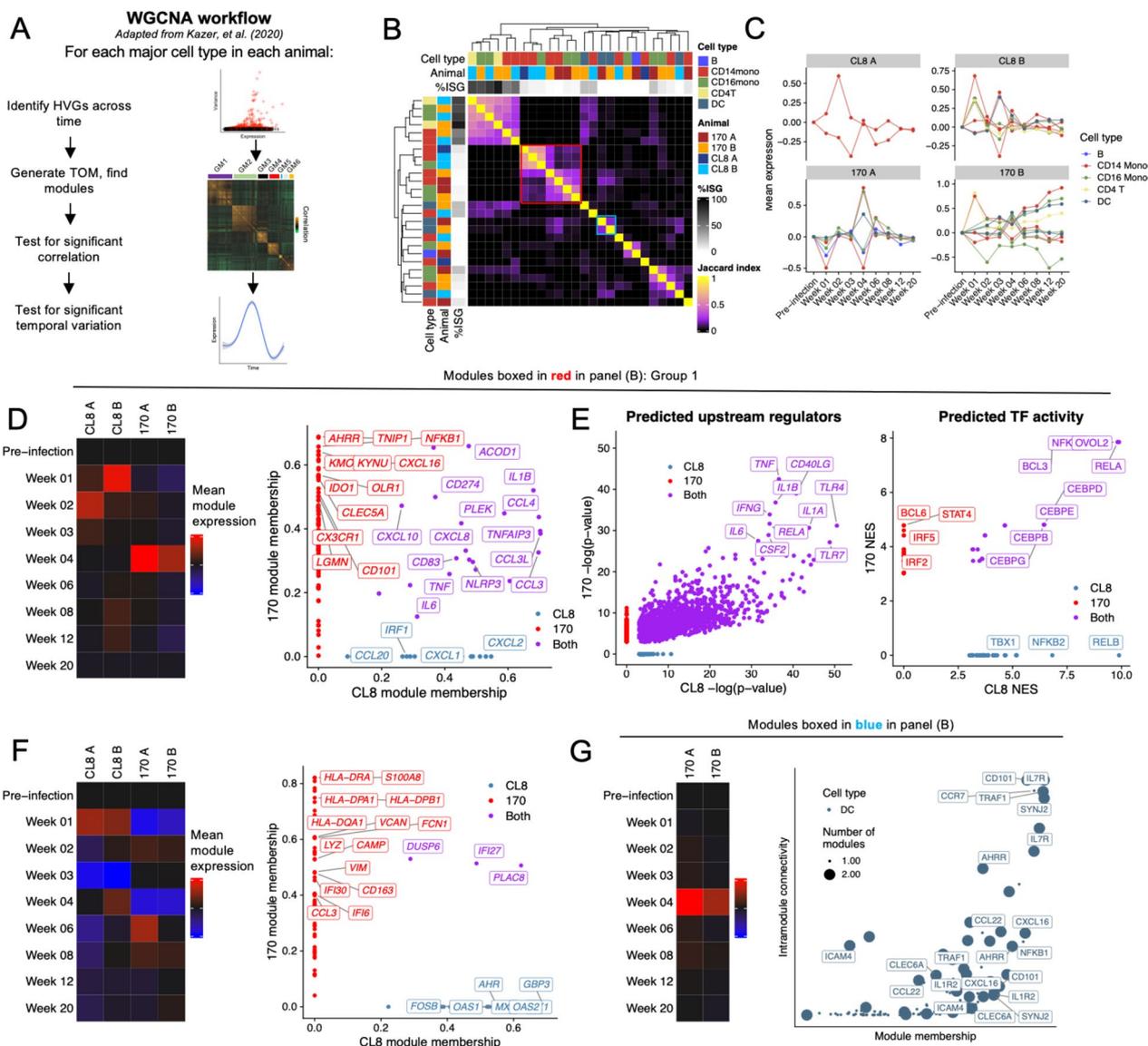
### Gene module discovery reveals structured longitudinal immune responses to infection with lentiviral pathogenic variants

To understand how cell transcriptional phenotype varies across time, we analyzed coordinated shifts in gene expression within each cell type that varied in time. We adapted a weighted gene correlation network analysis (WGCNA)-based [40] approach to identify temporally variable modules of co-expressed genes that was originally described by Kazer et al. [18]. For each major cell type within each animal, we identified variable genes and used the single-cell gene expression matrices to generate

topological overlap matrices (TOM) that capture gene co-expression structure. Modules of co-expressed genes were identified in these TOMs, and the correlation structure of each module was tested for statistical significance. Single cells were then scored for expression of each module and module expression tested for significant variability across time (Fig. 3A; see “Methods”). We perform this analysis (a) for each cell type individually, so that temporally variant gene modules are more likely to represent shifts in cell state rather than changes in cell type proportion, and (b) for each animal individually, in order to be robust to inter-individual heterogeneity. These gene modules thus represent coordinated programs of cellular phenotype and can be used to nominate cell type-specific drivers of synchronized responses to infection.

This process returned 26 distinct, temporally variant gene modules in B cells, monocytes, CD4 T cells, and DCs (Fig. 3B). No significantly correlated, temporally variant gene modules were discovered in CD8 T cells or NK cells. This may indicate a lack of temporally coordinated gene programs in these cell types but may also be due to inadequate sampling times relative to biological shifts or due to sparser RNA content in these cells limiting our ability to detect strong gene-gene correlations. To visualize relationships between the identified gene modules, we calculated the Jaccard index between each set of modules, which revealed groups of partially overlapping modules discovered between distinct animals or cell types (Fig. 3B). For example, gene module discovery revealed a set of six highly overlapping modules that were composed primarily of ISGs (Fig. 3B); expression of two of these modules were highly significantly enriched in SIV<sup>+</sup> cells (Additional file 1: Fig. S6). We next scored the cells of each animal and cell type for expression of the corresponding gene modules to examine how expression of each module varied across time. This revealed shifts in module expression that paralleled global patterns in whole-transcriptome perturbation (Fig. 1D–E), where 170-infected animals display concordant shifts in monocyte phenotype at 4 weeks post-infection (Fig. 3C).

Next, we analyzed a group of seven partially overlapping modules that were discovered within monocytes of all four animals (Fig. 3B; red box, hereafter “group 1”). Expression of the corresponding modules peaked at 1–2 weeks post-infection in CL8-infected animals but at 4 weeks post-infection in both 170-infected animals (Fig. 3D). Genes that were common to all group 1 modules included pro-inflammatory cytokines and chemokines *IL-1B*, *TNF*, *CCL3*, *CCL4*, *IL6*, *CXCL8*, and *CXCL10*, as well as inflammasome component *NLRP3*. Group 1 modules from CL8-infected animals contained the neutrophil-chemoattractants *CXCL1* and *CXCL2*, while modules from 170-infected animals contained



**Fig. 3** Gene module discovery reveals shared and distinct temporal features between immune responses to pathogenic and nonpathogenic lentiviral infection. **A** Overview of gene module discovery by WGCNA, adapted from Kazer et al. [18]. HVG, highly variable gene. **B** Jaccard index heatmap depicting the degree of overlap in gene membership between all discovered modules. **C** Expression of modules over time by the cell type and in the animal in which the modules were discovered. For **D–G**, heatmaps depict average expression of gene modules by the cell type and animal in which the modules were discovered. Scatter plots depict either the module membership of module genes in CL8 and 170 animals (right **D**, **F**, and **G**) or upstream regulators (left **E**) and transcription factor activities (right **E**) predicted to result in expression of the observed gene modules. **D–E** monocyte modules boxed in red in panel (**B**). **F** Modules downregulated by monocytes in at least one time point. **G** DC modules boxed in blue in **B**

the chemokine *CXCL16* (Fig. 3D). Additionally, modules from 170-infected animals contained three separate members of the kynurenine pathway of tryptophan metabolism (*IDO*, *KYNU*, and *KMO*; Fig. 3D); activity of this pathway is known to inhibit T-cell proliferation and IL-17 production and strongly predict HIV mortality

[55, 56]. We performed ingenuity pathway analysis (IPA) to identify predicted upstream regulators of these modules, as well as transcription factor (TF) activity analysis [57] to identify potential regulatory mechanisms underlying module expression. These analyses implicated a similar set of upstream drivers in module expression,

including *IL1B* and *IFNG*, but only partially overlapping TF activities, where TFs from the noncanonical pathway of NF- $\kappa$ B activation (NFKB2 and RELB) were predicted to drive expression of modules from CL8-infected animals but not 170-infected animals (Fig. 3E). This paralleled higher expression of receptors that induce noncanonical NF- $\kappa$ B signaling *TNFRSF13C* (encoding BAFRR) and *TNFRSF11A* (encoding RANK) in CL8-infected animals (Additional file 1: Fig. S6).

We noted there were monocyte-based modules that showed the opposite expression pattern and were predominantly repressed relative to baseline, often concomitant with upregulation of Group 1 modules (Fig. 3C, F). In 170-infected animals, these predominantly repressed modules contain several HLA class II encoding genes as well as the immunoregulatory hemoglobin scavenger *CD163* and antimicrobial agent-encoding genes *LYZ* and *CAMP* (Fig. 3F). Collectively, these results suggest a synchronized and coordinated shift from antigen presentation and antimicrobial function towards pro-inflammatory cytokine and chemokine production in monocytes during highly pathogenic *Lentivirus* infection.

We also identified two overlapping modules discovered in both 170-infected animals in DCs. These modules were expressed in both 170-infected animals at predominantly 4 weeks post-infection and contained genes implicated in DC trafficking to secondary lymphoid tissues, including *CCR7* and *IL7R* [58], pro-inflammatory chemokine *CXCL16*, and molecules involved in communication with T cells, including *CD101* and *CCL22* [59, 60] (Fig. 1G). As *CCR7* and *IL7R* are highly expressed on most T cells, we confirmed that *CCR7*<sup>+</sup>*IL7R*<sup>+</sup> DCs expressed canonical DC markers without co-expression of any canonical T-cell-defining genes (Additional file 1: Fig. S6). Collectively, gene module discovery analysis reveals coordinated pathways that distinguish immune responses between cell types, states, and lentiviral pathogenic variants.

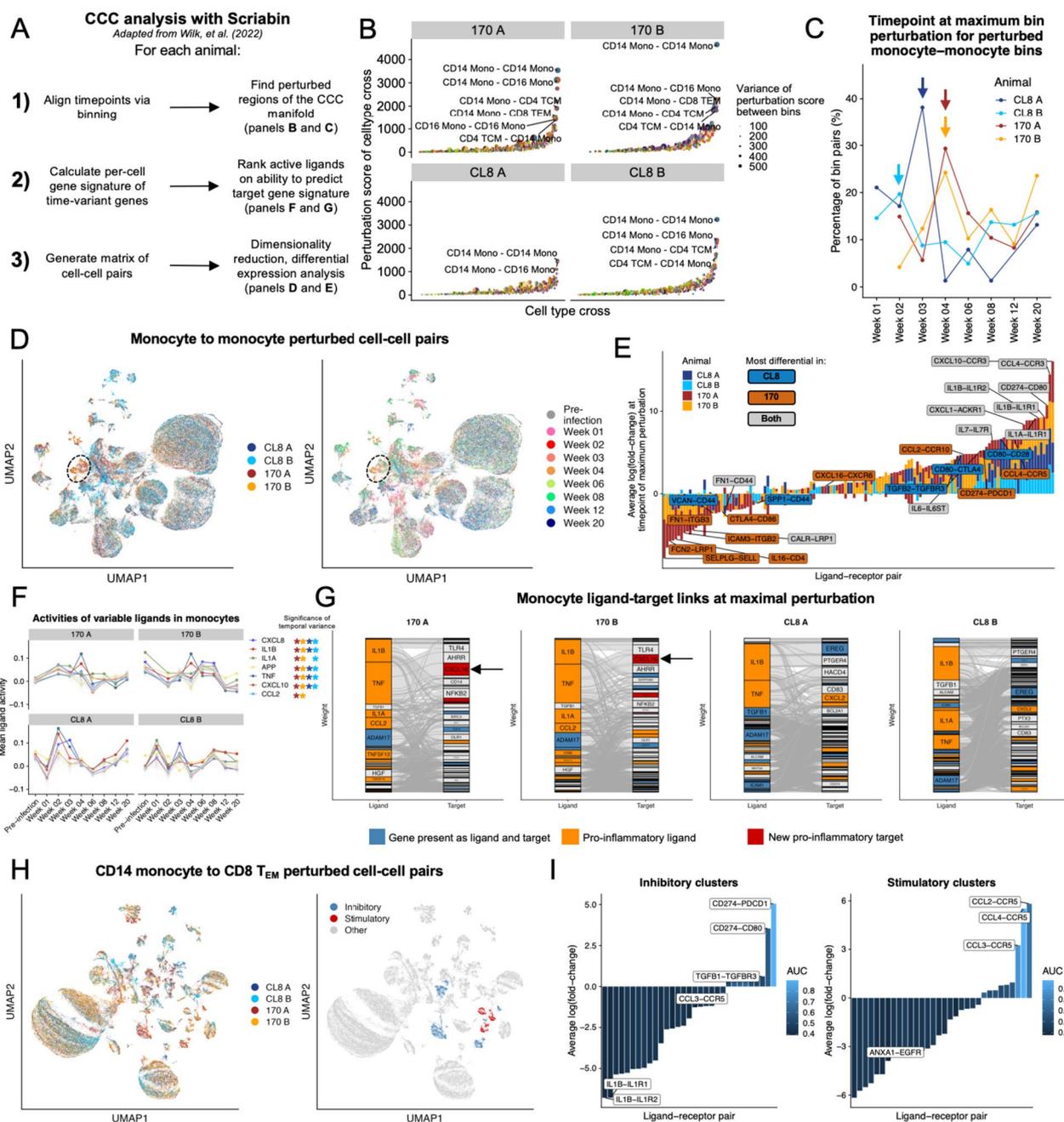
### Single-cell resolution communication analysis reveals inflammatory networks differentiating immune responses between lentiviral pathogenic variants

While our gene module discovery analysis nominates potential drivers of shifts in cell state over time, we wished to formally analyze how communication pathways between cells or cell-cell communication (CCC) influenced downstream transcriptional phenotypes. To address this question, we applied Scriabin, a method for comparative CCC analysis at single-cell resolution that does not require any degree of subsampling or aggregation (Fig. 4A; see “Methods”) [42].

To focus our CCC analysis on the cell types with the highest degree of CCC perturbation over time, we applied Scriabin’s summarized interaction graph workflow: for each sample, we generated a summarized interaction graph representing the total magnitude of potential interactions between every pair of cells. Then, we aligned datasets from all the time points from a given animal using Scriabin’s binning workflow, which assigns each cell a bin identity that maximizes the similarity of cells within each bin while simultaneously maximizing representation of all time points within each bin. These bin identities thus represent high-resolution intra-time point correspondences. We used our reference-based cell type annotations, which were transferred via neighbor graphs that are not used for binning, as orthogonal validation that binning recovered phenotypically similar cells (Additional file 1: Fig. S7). We next assessed variability in total communicative potential across time within each bin, which is summarized in Fig. 4B. This process revealed that monocyte-to-monocyte communication is the most longitudinally perturbed across all four animals and additionally indicated that monocyte communication with memory T cells and NK cells was also significantly perturbed (Fig. 4B). In 170-infected animals, the most highly perturbed monocyte-to-monocyte bins displayed the largest shift in CCC

(See figure on next page.)

**Fig. 4** CCC pathways distinguishing immune responses to pathogenic and nonpathogenic lentiviral infection. **A** Overview of application of Scriabin [42] to analyze single-cell resolution CCC profiles. Briefly, Scriabin’s summarized interaction graph workflow was used to align the time points within an animal and identify cell type combinations with perturbed CCC to prioritize for downstream analysis. Then, Scriabin’s ligand activity ranking process was used to identify biologically active CCC edges. Finally, Scriabin’s CCIM workflow was used to visualize the full structure of CCC phenotypes (see “Methods”). **B** Scatter plots depicting the degree of CCC perturbation predicted by Scriabin’s summarized interaction graph workflow (y-axis). The most highly perturbed cell type combinations for each animal are labeled (sender cell type — receiver cell type). Points are sized by the variability in perturbation score between bins of the same cell type combination. **C** Plot depicting the distribution of time points at which CCC in each monocyte-to-monocyte bin-bin pair is maximally perturbed. **D** UMAP projections of the most highly perturbed monocyte-monocyte cell-cell pairs from all samples, colored by animal (left) and time point (right). **E** Stacked bar plot depicting DE ligand-receptor pairs in the monocytes of each animal at the time point of maximal perturbation. A ligand-receptor pair is considered to be more differentially expressed in an animal class if > 1 standard deviation of that ligand-receptor pair’s log(fold-change) is contributed by a particular animal class. **F** Predicted ligand activities of monocytes over time in each animal. **G** Alluvial plots depicting ligand-target gene connections within monocytes of each animal at the time point of maximal perturbation. **H** UMAP projections of the most highly perturbed CD14 monocyte-CD8 T<sub>EM</sub> sender-receiver cell-cell pairs colored by animal (left) and clusters enriched in 170-infected animals (right). **I** Bar plots depicting DE ligand-receptor pairs in the clusters highlighted in **H**



**Fig. 4** (See legend on previous page.)

magnitude at 4 weeks post-infection, while CL8 animals had the greatest perturbation at 1–2 weeks post-infection (Fig. 4C). For downstream analyses, we defined the time point of maximal perturbation as 4 weeks post-infection for 170-infected animals, 2 weeks post-infection for CL8 A, and 1 week post-infection for CL8 B.

To comprehensively visualize the CCC structure of our dataset, we next applied Scriabin’s cell-cell interaction matrix (CCIM) workflow. This workflow generates a

matrix of cell-cell pairs by ligand-receptor pairs that can be treated analogously to a gene expression matrix for downstream analytical tasks like dimensionality reduction and differential expression testing (Fig. 4A). We also used Scriabin’s implementation of NicheNet [44] to predict which ligand-receptor pairs were biologically active by leveraging information about gene expression changes downstream of specific ligands and incorporated this information into the CCIM (see the “Methods”).

We first used Scriabin's CCIM workflow to analyze the most highly perturbed CCC between cell-cell pairs of monocytes. This analysis revealed groups of monocyte-monocyte pairs that appeared to have distinct communication profiles (Fig. 4D). Differential expression testing indicated that CCC between monocytes at the time point of maximal perturbation was characterized by increased signaling through *IL1B* and *CCL4* (Fig. 4E; Additional file 2: Table S6). 170-infected animals alone had increased signaling with *CXCL16* and *CCL2* while losing communication through the adhesion molecule *SELPLG*. Conversely, monocytes from CL8-infected animals had increased communication via *TGFB2* (Fig. 4E). These shifts in CCC pathways were also reflected in single-cell level ligand activities, which showed increased predicted activity of *IL1B* at the time point of maximal perturbation in all animals but only an increase in *CCL2* activity at 4 weeks post-infection in both 170-infected animals (Fig. 4F).

Next, we examined what downstream gene expression changes were predicted to result from the observed ligand activities. While *IL1B* and *TNF* were the primary predicted drivers of target gene profiles in all animals, in 170-infected animals, these ligands were predicted to result in upregulation of *TLR4*, *NFKB2*, and the aryl hydrocarbon receptor repressor *AHRR* but in CL8-infected animals were predicted to result in the upregulation of neutrophil chemoattractant *CXCL2* and checkpoint molecule *CD83* (Fig. 4G) [61]. Intriguingly, in 170-infected animals alone, we observed upregulation of additional pro-inflammatory cytokines and chemokines like *CXCL16* that had not been predicted to be active as ligands (Fig. 4G), suggesting a broadening of pro-inflammatory signaling in infection with highly pathogenic *Lentivirus*. Collectively, these results demonstrate overlapping yet distinct monocyte-to-monocyte communication mechanisms in response to lentiviral variants of different pathogenicity, with highly pathogenic *Lentivirus* generally inducing a delayed and broader pro-inflammatory communication profile.

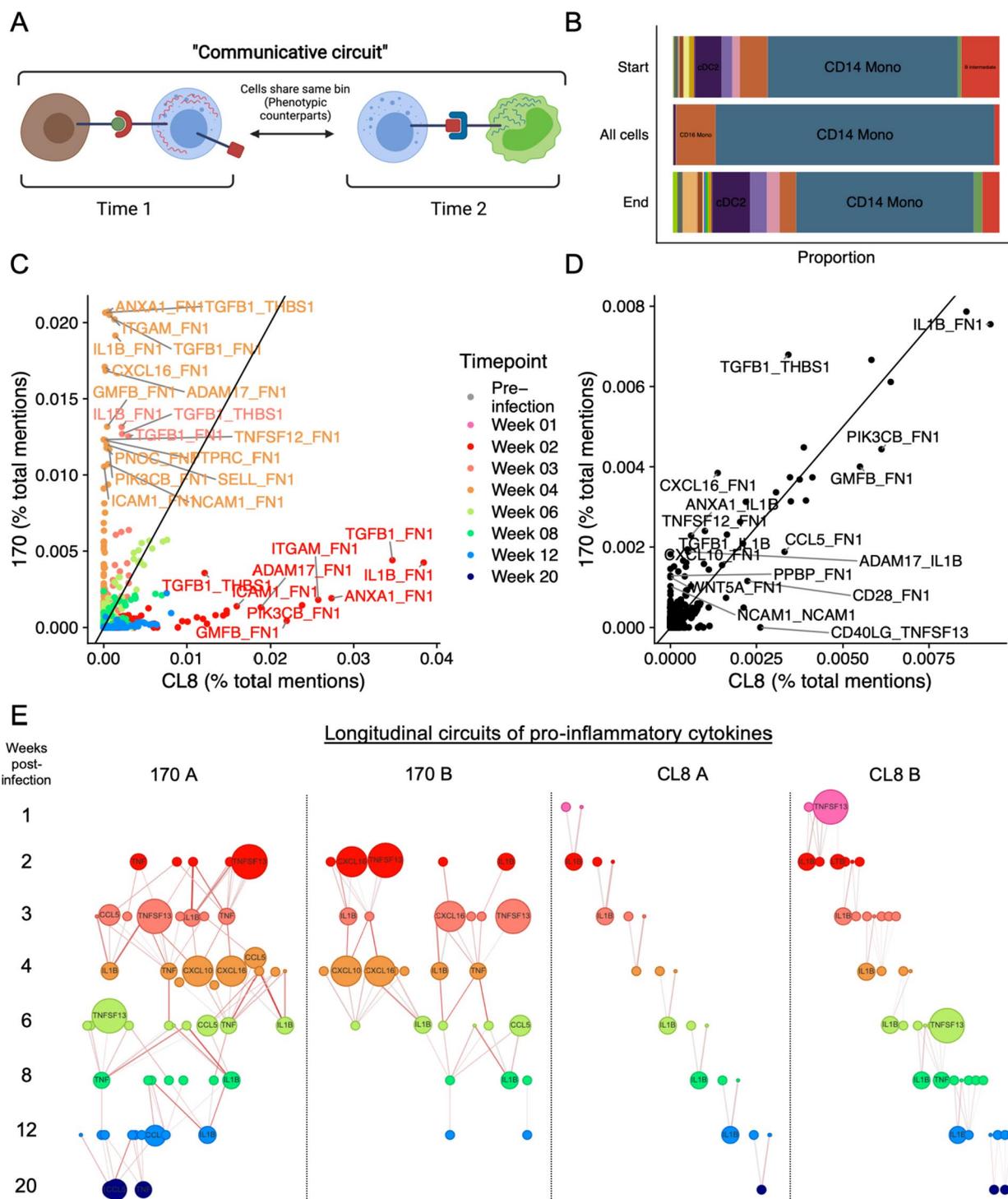
We next analyzed communication from CD14 monocytes to CD8 T<sub>EM</sub> cells, as CCC for this sender-receiver cell type combination was highly perturbed only in 170-infected animals (Fig. 4B). Using Scriabin's CCIM workflow on highly perturbed CD14 monocyte-CD8 T<sub>EM</sub> cell-cell pairs, we found several clusters that appeared highly enriched for cell-cell pairs from 170-infected animals (Fig. 4H). We found that three of these clusters showed increased communication through inhibitory pathways, including the *CD274-PDCD1* (PDL1-PD1) pathway, while the other three clusters showed increased communication through chemokines *CCL2*, *CCL3*, and *CCL4* through their cognate receptor *CCR5* (Fig. 4I). Interestingly, we found that the frequency of both groups of clusters was highest in the 170-infected animals at 4 weeks post-infection (Additional file 1: Fig. S8). Thus, Scriabin revealed the concomitant use of opposing signaling modalities between CD14 monocytes and CD8 T<sub>EM</sub> cells in response to highly pathogenic *Lentivirus* infection, underscoring the utility of using single-cell resolution CCC methods to capture the full structure of CCC phenotypes.

#### Broad longitudinal inflammatory networks in pathogenic lentiviral infection

While our single-cell resolution analysis of CCC pathways enabled us to compare what communication edges were differentially active between different animals and time points, we were next interested in formally analyzing how these communication pathways operate *across* time points in order to understand the dynamic evolution of CCC during infection. To address this question, we applied Scriabin's longitudinal circuitry discovery workflow [42]. A longitudinal communicative circuit is a set of communication edges at two consecutive time points, where communication at the second time point is predicted to be the consequence of communication at the first time point (Fig. 5A). Consider an interacting sender-receiver cell pair  $S_1-R_1$  at time point 1 and interacting sender-receiver cell pair  $S_2-R_2$  at time point 2. If communication between  $S_1-R_1$  results in the expression

(See figure on next page.)

**Fig. 5** Longitudinal CCC circuits during immune response to lentiviral pathogenic variants. **A** Diagram illustrating the definition of a longitudinal communicative circuit. **B** Stacked bar plot depicting the proportion of cell type annotations for cells from animal 170 A participating in longitudinal circuits. Proportions are separated by cells participating at the beginning or the end of the circuit, as well as the total distribution of all cells participating in circuits. **C—D** Scatter plots depicting ligand-target pairs participating in longitudinal communication circuits that differentiate 170-infected animals (y-axis) and CL8-infected animals (x-axis). For **C**, each point represents a ligand-target pair participating in a circuit at a particular time point. For **D**, each point represents the total number of mentions for each ligand-target pair in all circuits, regardless of time point. **E** Longitudinal circuits of pro-inflammatory cytokines in each animal. Circuits were restricted to pro-inflammatory cytokines (see Additional file 1: Table S5) and networks of connected pro-inflammatory cytokine ligand-target pairs plotted for each animal across time. Points are sized by the expression of the respective cytokine at each time point. Edges are colored by the product of the ligand activity score and the target weight for the sequential cytokine (see "Methods")



**Fig. 5** (See legend on previous page.)

of ligand  $L_A$  by  $R_1$ ,  $S_1$ - $R_1$ - $S_2$ - $R_2$  represents a communicative circuit if  $R_1$  and  $S_2$  occupy the same bin (i.e.,  $S_2$  represents the phenotypic counterpart of  $R_1$  at time point 2) and if  $L_A$  is predicted to be an active ligand in the

$S_2$ - $R_2$  interaction (Fig. 5A). Thus, longitudinal circuits represent the single-cell resolution stitching together of two neighboring time points by succeeding communication pathways.

We applied Scriabin's circuitry discovery workflow to identify all temporally consecutive communication circuits across time in each animal. We first noted that the vast majority of cells participating in circuits were monocytes; however, non-monocyte cell types were more highly represented in cells at the beginnings and ends of circuits (Fig. 5B). This indicates that monocytes serve as the hubs and authorities of longitudinal communication networks, integrating and sending signals to other cell types to coordinate downstream responses.

We next examined which longitudinal circuits differed between 170- vs. CL8-infected animals. When analyzing ligand-target pairs participating in circuits at each time point, we observed strong differences between 170- and CL8-infected animals; most 170-infected animal circuits overlapped at the 4 weeks post-infection time point, compared to the 2 weeks post-infection time point for CL8-infected animals (Fig. 5C). However, when summarizing the frequency of ligand-target pairs participating in circuits irrespective of time point, we found more concordant patterns between 170- and CL8-infected animals (Fig. 5D). This indicates that longitudinal communication pathways in response to nonpathogenic vs. pathogenic lentiviral infection differ primarily by the time at which they operate rather than the specific ligands that compose these pathways. Nonetheless, we noted several ligand-target pairs that were more enriched in circuits from 170-infected animals, including *CXCL10* and *CXCL16* acting as ligands for the upregulation of extracellular matrix protein *FNI* (Fig. 5D).

Our single-cell resolution analysis of CCC nominated putative signaling mechanisms for the expansion of the breadth of pro-inflammatory cytokine expression in 170-infected animals (see Fig. 4G), and our analysis of longitudinal communication circuits identified two chemokines that were enriched in circuits from 170-infected animals. We hypothesized that this broader signature of pro-inflammatory cytokine activity in response to pathogenic *Lentivirus* infection may reflect distinct network structures of cytokine regulation. To comprehensively visualize the longitudinal evolution of pro-inflammatory cytokine expression and regulation, we identified all communicative circuits involving sets of pro-inflammatory cytokines that were consecutive in time and regulation. This analysis revealed broad, sustained, and densely connected longitudinal networks of pro-inflammatory cytokine expression exclusively in 170-infected animals (Fig. 5E). Networks from CL8-infected animals, conversely, were limited and sparsely connected across all time points. While circuits from both 170- and CL8-infected animals included *IL1B*, *TNF*, *TNFSF13*, and *CCL5*, only networks from 170-infected animals included the chemokine *CXCL16* (Fig. 5E).

Collectively, these results are suggestive of inflammatory feedback loops that are specific to immune responses to highly pathogenic *Lentivirus* infection.

## Discussion

In this work, we have applied single-cell transcriptomics and novel frameworks for CCC analysis to examine features of the immune response that distinguish infection with minimally vs. highly pathogenic *Lentivirus* infection. A substantial body of work has compared differential lentiviral pathogenicity from the perspective of host genetics by infecting different nonhuman primate (NHP) species with SIVs from different natural hosts [4, 62–66]. However, the role of virus-specific factors remains relatively underexplored. We leveraged unique samples that allowed us to compare responses at weekly intervals after infection to the gene profile at the time of exposure by leveraging a model of differential lentiviral pathogenicity where all hosts were from the same genetic background [16]. Here, we provide the first description of virus-specific immune profiles that distinguish minimally pathogenic vs. highly pathogenic *Lentivirus* infection.

Compared to infection with minimally pathogenic CL8, we find that infection with highly pathogenic 170 induces delayed and sustained signatures of immune activation. This finding is in line with evidence that chronic inflammation and immune activation are a distinguishing feature of pathogenic SIV infection as well as HIV infection relative to SIV infection in natural hosts [18, 51, 67, 68]. Prior work indicates that persistent type I IFN signaling in pathogenic SIV leads to broad dysfunction in both innate and adaptive immunity during chronic infection [67, 69–73]. In line with these observations, here we show that 170-infected animals display a broad and sustained response to type I IFN and demonstrate for the first time that distinct sets of ISGs are upregulated in response to IFN in chronic and pathogenic infection. Similarly, preliminary evidence suggests blockade of type I IFN signaling during chronic infection may suppress other inflammatory pathways and pathology and rescue counts of HIV-specific T cells [69, 72, 74]. The association between delayed/sustained IFN signaling and disease outcome has also been postulated in the setting of acute viral infections like MERS-CoV and SARS-CoV-2 [75–77], indicating this may be a conserved feature of imbalanced or ineffective antiviral immunity.

Although we identified broad and sustained IFN responses in 170-infected animals, we found the opposite pattern when examining *individual* SIV-infected cells: CD4<sup>+</sup> T cells and monocytes infected with mildly pathogenic CL8 upregulated a broader range of ISGs and expressed a higher magnitude of these ISGs than individual 170-infected cells. This finding suggests an

evolved immune evasion mechanism of 170, where 170 may escape detection via decreased PRR sensitivity, IFN production, or autocrine IFN signaling. This hypothesis is also consistent with our observation that peak viremia in 170-infected animals occurs 1 week after the peak of IFN responses, suggesting that IFN responses in these animals may be insufficient to control viral replication. In 170-infected animals, peak viremia occurs at 2 weeks post-infection, 1 week after the peak in ISG signature, further supporting that IFN responses are unable to sufficiently control viral replication of 170. In demonstrating reduced ISG induction in 170-infected cells, we did not rely on the 1 week post-infection time point in 170-infected animals when sample quality was low. Rather, a plurality of 170-infected cells came from 2 weeks post-infection, and a plurality of CL8-infected cells came from 1-week post-infection, matching the time of peak viremia for both viruses

As ISGs can act as critical antiviral restriction factors in infected cells [18, 52], we hypothesize that decreased ISG induction in 170-infected cells represents an evolved immune evasion strategy of 170 and may be an explanatory mechanism for the higher peak viremia and set point in 170-infected animals. Thus, while 170-infected animals have a broad *global* IFN signature, individual cells infected with 170 fail to upregulate the same set of ISGs to the same extent. These findings are specifically made possible by our single-cell approach and ability to accurately distinguish infected from bystander cells (Fig. 1G). Some ISGs that are more highly upregulated by CL8-infected cells, including *MX2* and *IFITM1*, have known roles as HIV-1 restriction factors in cell lines [78–82].

Several of the main immune mediators that are expressed within co-regulated gene modules overlap with those known to be important in the acute phase of SIV and HIV infection, including *CCL2*, *TNF*, and *IL1B* [83–85]. However, we also identify unique chemokine signatures that distinguish responses to lentiviral pathogenic variants. For example, we observe expression of neutrophil chemoattractants *CXCL1* and *CXCL2* primarily by CL8-infected animals. We hypothesize that these chemokines, which are not expressed by 170-infected animals, may play important roles in downstream immune signaling cascades. Such cascades would be expected to exert downstream effects on neutrophils, which have been implicated in modulating HIV-1 disease course [86], but unfortunately, we are unable to analyze neutrophils as we were limited to profiling of cryopreserved mononuclear cells. Additionally, we found that 170-infected animals were the predominant producers of *CXCL10* and *CXCL16*, which both operated in dense longitudinal networks of pro-inflammatory cytokine regulation. These findings implicate *CXCL10* and *CXCL16* in

immune responses to highly pathogenic lentiviral infection and provide an avenue for future study.

Additionally, we used novel longitudinal frameworks for single-cell resolution CCC analysis to reveal that broad, sustained, and densely connected regulatory networks of pro-inflammatory cytokines are a defining feature of responses to highly pathogenic lentiviral infection. While longitudinal cytokine regulatory networks were largely disconnected in both CL8-infected animals, in 170-infected animals, pro-inflammatory cytokines were predicted to upregulate additional cytokines that were active at future time points. This is suggestive of amplifying positive feedback loops that result in the observed breadth of inflammatory activity in highly pathogenic *Lentivirus* infection. While the distinct structure of these regulatory networks may be due to unique immune phenotypes induced by 170-infected cells, it is also possible that this observation is driven primarily by sustained viral loads in these animals or by distinct processes of infection and inflammation occurring in the lymphatics or other tissues. While therapeutic blockade of numerous chemokine and cytokine pathways has been proposed and tested for applications in autoimmunity, these approaches have remained generally unexplored in HIV and other infectious diseases. For example, biologics targeting *CXCL10* are currently being evaluated for treatment of inflammatory bowel disease and multiple sclerosis [87]. By implicating chemokines like *CXCL10* as hubs of putatively pathogenic inflammatory networks in *Lentivirus* infection, we hypothesize that pharmacological targeting of these pathways could represent a novel strategy to modulate lentiviral virulence in HIV. Similar to preliminary work on blockade of type I IFN signaling [69, 72, 74], targeting inflammatory pathways such as *CXCL10* and *CXCL16* may also yield benefits of suppressing inflammatory pathology but with less risk of promoting viral replication. Such therapeutic interventions could also have utility outside of the acute phase of HIV infection, given that even patients with chronic HIV on antiretroviral therapy are known to have residual chronic inflammation that has been linked to comorbidities including cardiovascular disease and neurocognitive impairment [88–91].

There are several potential limitations to our work. First, sample availability limits our sample size, as well as the possible analysis modalities we can apply to understanding longitudinal immune profiles in this cohort. For example, our analyses are limited to transcriptomic readouts; unfortunately, we were unable to perform serum proteomics to validate protein-level changes that we identify at the transcriptional level. Additionally, this work profiled exclusively circulating immune cells; although understanding the peripheral immune

system is critical to understanding aberrant and protective immune responses to SIV infection, it does not capture the immune response at other important sites of infection, including lymph nodes. Further functional experimentation is necessary to validate or refute many of the hypotheses presented here.

## Conclusions

Collectively, we perform the first single-cell resolution dissection of the distal immune mechanisms that distinguish infection with highly related lentiviral variants of different virulence. Our work indicates a model of pathogenic *Lentivirus* infection where cells infected with highly pathogenic *Lentivirus* fail to upregulate a sufficient breadth and magnitude of ISGs, in turn leading to poorly controlled viremia that drives broad and sustained positive feedback loops of inflammatory networks. We nominate key drivers of these inflammatory pathways, including *CXCL10* and *CXCL16*, by using a combination of gene module discovery and novel longitudinal CCC analysis frameworks and reveal distinct regulatory programs of pro-inflammatory chemokines delineating responses to highly pathogenic lentiviral infection. Together, our work provides several mechanistic targets for further investigation that may represent targets for modulating lentiviral pathogenicity.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01290-y>.

**Additional file 1: Table S1.** Sample batch and processing schema. **Table S2.** List of manually-annotated ENSEMBL IDs. **Table S3.** List of ISGs. **Table S4.** Parameters used for WGCNA-based gene module discovery. **Table S5.** List of pro-inflammatory cytokine-encoding genes. **Fig. S1.** Quality control of single-cell transcriptomic dataset. **Fig. S2.** Dimensionality reduction embeddings faceted by time and animal. **Fig. S3.** Additional analyses of bulk transcriptomic profiles. **Fig. S4.** Expression of *IRF7* implicates sources of type 1 IFN. **Fig. S5.** Analysis of IFN response kinetics. **Fig. S6.** Analysis of gene modules correlated with viral load or enriched in SIV<sup>+</sup> cells. **Fig. S7.** Verification of binning fidelity. **Fig. S8.** Abundance of inhibitory and stimulatory CCIM clusters in CD8 T<sub>EM</sub> to monocyte communication.

**Additional file 2: Table S6.** Differential ligand-receptor pairs in monocyte CCC at the timepoint of maximal perturbation relative to baseline.

## Acknowledgements

We thank all current and former members of the Blish laboratory for helpful discussions of this work. We thank S. Sim and the Stanford Functional Genomics Facility for the use of the Fragment Analyzer, as well as A. Detweiler and M. Tan at the Chan Zuckerberg Biohub for assistance with sequencing. We thank M. Fitzgibbon at the Fred Hutchinson Genomics & Bioinformatics core for assistance with bulk RNA-sequencing and analysis. We also thank the Stanford Research Computing Center for providing invaluable computational resources for data analysis.

## Authors' contributions

AJW, AKS, SH, JO, and CAB conceived of the work and designed experiments. AJW, JOM, SWK, IF, VM, and JG-R performed scRNA-seq experiments. JOM

performed bulk RNA sequencing work. AJW performed computational analyses. AJW wrote the manuscript with input from all authors. All authors read and approved the final manuscript.

## Funding

A. J. Wilk is supported by the Stanford Medical Scientist Training Program (T32 GM007365-44) and the Stanford Bio-X Interdisciplinary Graduate Fellowship. This work was also supported by NIH/NIDA DP1 DA04508902 to C. A. B.; NIH HD103571 to J. O.; NIH/NCI 1U54CA217377, U01 28020510, and 1U2CCA23319501 to A. K. S.; NIH/NIDA 1DP1DA053731 to A. K. S.; Bill and Melinda Gates Foundation INV-027498 and OPP1202327 to A. K. S.; the MIT Stem Cell Initiative through Foundation MIT to A. K. S.; and a 2019 Sentinel Pilot Project from the Bill and Melinda Gates Foundation to C. A. B./A. K. S.

## Availability of data and materials

Raw and processed scRNA-seq data generated in this manuscript are publicly available as GEO accession GSE228688 [50], accessible at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228688>.

## Declarations

### Ethics approval and consent to participate

Samples from pig-tailed macaques used in this study were collected as part of a study by Kimata et al. [16], published in 1999. No additional animal work was performed in this study. Our use of these samples was reviewed and approved by Stanford University's Animal Care and Use Committee as APLAC-33350. This research conforms to the principles of the Helsinki Declaration.

### Consent for publication

Not applicable.

### Competing interests

A. K. S. reports compensation for consulting and/or SAB membership from Merck, Honeycomb Biotechnologies, Cellarity, Repertoire Immune Medicines, Ochre Bio, Third Rock Ventures, Hovione, Relation Therapeutics, FL82, FL86, Bio-Rad Laboratories, Empress Therapeutics, IntrECate Biotherapeutics, Senda Biosciences, and Dahlia Biosciences unrelated to this work. C. A. B. reports compensation for consulting and/or SAB membership from Catamaran Bio, DeepCell Inc., Immunebridge, and Revelation Biosciences. J. O. reports SAB membership for Aerium Therapeutics. V. N. M. reports compensation from MPM Capital unrelated to this work. S. W. K. reports compensation for consulting from Radera Biotherapeutics and Monopteros Therapeutics. All other authors declare that they have no competing interests.

### Author details

<sup>1</sup>Stanford Immunology Program, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>2</sup>Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>3</sup>Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA 94305, USA. <sup>4</sup>Division of Human Biology, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. <sup>5</sup>Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>6</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>7</sup>Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>8</sup>Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA 02139, USA. <sup>9</sup>Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. <sup>10</sup>Program in Health Sciences & Technology, Harvard Medical School & MIT, Boston, MA 02115, USA. <sup>11</sup>Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA. <sup>12</sup>Department of Statistics, Stanford University, Stanford, CA 94305, USA. <sup>13</sup>Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

Received: 17 April 2023 Accepted: 19 January 2024

Published online: 05 February 2024

## References

- Giorgi JV, Hultin LE, McKeating JA, Johnson TD, Owens B, Jacobson LP, et al. Shorter survival in advanced human immunodeficiency virus type

- 1 infection is more closely associated with T lymphocyte activation than with plasma virus burden or virus chemokine coreceptor usage. *J Infect Dis.* 1999;179:859–70.
2. Sousa AE, Carneiro J, Meier-Schellersheim M, Grossman Z, Victorino RMM. CD4 T cell depletion is linked directly to immune activation in the pathogenesis of HIV-1 and HIV-2 but only indirectly to the viral load. *J Immunol.* 2002;169:3400–6.
  3. Chahroudi A, Bosinger SE, Vanderford TH, Paiardini M, Silvestri G. Natural SIV hosts: showing AIDS the door. *Science.* 2012;335:1188–93.
  4. Joas S, Parrish EH, Gnanadurai CW, Lump E, Stürzel CM, Parrish NF, et al. Species-specific host factors rather than virus-intrinsic virulence determine primate lentiviral pathogenicity. *Nat Commun.* 2018;9:1371.
  5. Palesch D, Bosinger SE, Tharp GK, Vanderford TH, Paiardini M, Chahroudi A, et al. Sooty mangabey genome sequence provides insight into AIDS resistance in a natural SIV host. *Nature.* 2018;553:77–81.
  6. Rodger AJ, Fox Z, Lundgren JD, Kuller LH, Boesecke C, Gey D, et al. Activation and coagulation biomarkers are independent predictors of the development of opportunistic disease in patients with HIV infection. *J Infect Dis.* 2009;200:973–83.
  7. Kuller LH, Tracy R, Bellosso W, De Wit S, Drummond F, Lane HC, et al. Inflammatory and coagulation biomarkers and mortality in patients with HIV infection. *PLoS Med.* 2008;5:e203.
  8. Sandler NG, Wand H, Roque A, Law M, Nason MC, Nixon DE, et al. Plasma levels of soluble CD14 independently predict mortality in HIV infection. *J Infect Dis.* 2011;203:780–90.
  9. Sandler NG, Douek DC. Microbial translocation in HIV infection: causes, consequences and treatment opportunities. *Nat Rev Microbiol.* 2012;10:655–66.
  10. Salve S, Singh A, Padwal V, Velhal S, Nagar V, Patil P, et al. Immune signatures for HIV-1 and HIV-2 induced CD4+T cell dysregulation in an Indian cohort. *BMC Infect Dis.* 2019;19:135.
  11. Royle CM, Graham DR, Sharma S, Fuchs D, Boasso A. HIV-1 and HIV-2 differentially mature plasmacytoid dendritic cells into IFN-producing cells or APCs. *J Immunol.* 2014;193:3538–48.
  12. Schindler M, Münch J, Kutsch O, Li H, Santiago ML, Bibollet-Ruche F, et al. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. *Cell.* 2006;125:1055–67.
  13. Du Z, Lang SM, Sasseville VG, Lackner AA, Ilyinskii PO, Daniel MD, et al. Identification of a nef allele that causes lymphocyte activation and acute disease in macaque monkeys. *Cell.* 1995;82:665–74.
  14. Khalid M, Yu H, Sauter D, Usmani SM, Schmokel J, Feldman J, et al. Efficient Nef-mediated downmodulation of TCR-CD3 and CD28 is associated with high CD4+ T cell counts in viremic HIV-2 infection. *J Virol.* 2012;86:4906–20.
  15. Kimata JT. HIV-1 fitness and disease progression: insights from the SIV-macaque model. *Curr HIV Res.* 2006;4:65–77.
  16. Kimata JT, Kuller L, Anderson DB, Dailey P, Overbaugh J. Emerging cytopathic and antigenic simian immunodeficiency virus variants influence AIDS progression. *Nat Med.* 1999;5:535–41.
  17. Gierahn TM, Wadsworth MH 2nd, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods.* 2017;14:395–8.
  18. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med.* 2020;26:511–8.
  19. Hughes TK, Wadsworth MH, Gierahn TM, Do T, Weiss D, Andrade PR, et al. Second-strand synthesis-based massively parallel scRNA-Seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity.* 2020;53:878–94.e7.
  20. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
  21. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30:923–30.
  22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.
  23. Petukhov V, Guo J, Baryawno N, Severe N, Scadden DT, Samsonova MG, et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol.* 2018;19:78.
  24. Fleming SJ, Chaffin MD, Arduini A, Akkad A-D, Banks E, Marioni JC, et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using Cell Bender. *Nat Methods.* 2023;20:1323–35.
  25. Ziegler CGK, Miao VN, Owings AH, Navia AW, Tang Y, Bromley JD, et al. Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *Cell.* 2021;184:4713–33.e22.
  26. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc.* 2009;4:1184–91.
  27. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. *Cell.* 2019;177:1888–902.e21.
  28. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36:411–20.
  29. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;<https://doi.org/10.1016/j.cell.2021.04.048>.
  30. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML]. 2018. Available from:<http://arxiv.org/abs/1802.03426>.
  31. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol.* 2018;<https://doi.org/10.1038/nbt.4314>.
  32. Gutierrez-Arcelus M, Teslovich N, Mola AR, Polidoro RB, Nathan A, Kim H, et al. Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nat Commun.* 2019;10:687.
  33. Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science.* 2017;356;<https://doi.org/10.1126/science.aah4573>.
  34. 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;36:89–94.
  35. Nimmerjahn F, Ravetch JV. Fcγ receptors as regulators of immune responses. *Nat Rev Immunol.* 2008;8:34–47.
  36. Nimmerjahn F, Ravetch JV. Fcγ receptors: old friends and new family members. *Immunity.* 2006;24:19–28.
  37. Palmer C, Diehn M, Alizadeh AA, Brown PO. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics.* 2006;7:115.
  38. Papalexi E, Mimitou EP, Butler AW, Foster S, Bracken B, Mauck WM 3rd, et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat Genet.* 2021;53:322–31.
  39. Wilk AJ, Lee MJ, Wei B, Parks B, Pi R, Martínez-Colón GJ, et al. Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J Exp Med.* 2021;218;<https://doi.org/10.1084/jem.20210582>.
  40. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9:559.
  41. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. *BMC Bioinformatics.* 2007;8:22.
  42. Wilk AJ, Shalek AK, Holmes S, Blish CA. Comparative analysis of cell–cell communication at single-cell resolution. *Nat Biotechnol.* 2023;1–14.
  43. Cortal A, Martignetti L, Six E, Rausell A. Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. *Nat Biotechnol.* 2021;39:1095–102.
  44. Browaeys R, Saelens W, Saeyens Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods.* 2020;17:159–62.
  45. Waltman L, van Eck NJ. A smart local moving algorithm for large-scale modularity-based community detection. *Eur Phys J B.* 2013;86:471.

46. Evans DT, Silvestri G. Nonhuman primate models in AIDS research. *Curr Opin HIV AIDS*. 2013;8:255–61.
47. Policicchio BB, Pandrea I, Apetrei C. Animal models for HIV cure research. *Front Immunol*. 2016;7:12.
48. Kimata JT, Overbaugh J. The cytopathicity of a simian immunodeficiency virus Mne variant is determined by mutations in Gag and Env. *J Virol*. 1997;71:7629–39.
49. Overbaugh J, Rudensey LM, Papenhausen MD, Benveniste RE, Morton WR. Variation in simian immunodeficiency virus env is confined to V1 and V4 during progression to simian AIDS. *J Virol*. 1991;65:7025–31.
50. Wilk AJ, Blish CA. GSE228688: pro-inflammatory feedback loops define immune responses to pathogenic Lentivirus infection. *Gene Expression Omnibus*. 2024; Available from: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE228688>
51. Kazer SW, Walker BD, Shalek AK. Evolution and diversity of immune responses during acute HIV infection. *Immunity*. 2020;53:908–24.
52. Chintala K, Mohareer K, Banerjee S. Dodging the host interferon-stimulated gene mediated innate immunity by HIV-1: a brief update on intrinsic mechanisms and counter-mechanisms. *Front Immunol*. 2021;12:716927.
53. Malim MH, Bieniasz PD. HIV restriction factors and mechanisms of evasion. *Cold Spring Harb Perspect Med*. 2012;2:a006940.
54. Bosinger SE, Johnson ZP, Folkner KA, Patel N, Hashempour T, Jochems SP, et al. Intact type I interferon production and IRF7 function in sooty mangabeys. *PLoS Pathog*. 2013;9:e1003597.
55. Byakwaga H, Boum Y 2nd, Huang Y, Muzoora C, Kembabazi A, Weiser SD, et al. The kynurenine pathway of tryptophan catabolism, CD4+ T-cell recovery, and mortality among HIV-infected Ugandans initiating antiretroviral therapy. *J Infect Dis*. 2014;210:383–91.
56. Lee S, Byakwaga H, Boum Y, Burdo TH, Williams KC, Lederman MM, et al. Immunologic pathways that predict mortality in HIV-infected Ugandans initiating antiretroviral therapy. *J Infect Dis*. 2017;215:1270–4.
57. Janky R s, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, et al. iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol*. 2014;10:e1003731.
58. Riol-Blanco L, Sánchez-Sánchez N, Torres A, Tejedor A, Narumiya S, Corbí AL, et al. The chemokine receptor CCR7 activates in dendritic cells two signaling modules that independently regulate chemotaxis and migratory speed. *J Immunol*. 2005;174:4070–80.
59. Bagot M, Martinel I, Charue D, Weill F, Boulland ML, Wechsler J, et al. CD101 is expressed by skin dendritic cells Role in T-lymphocyte activation. *Tissue Antigens*. 1997;50:439–48.
60. Rapp M, Wintergerst MWM, Kunz WG, Vetter VK, Knott MML, Lisowski D, et al. CCL22 controls immunity by promoting regulatory T cell communication with dendritic cells in lymph nodes. *J Exp Med*. 2019;216:1170–81.
61. Grosche L, Knippertz I, König C, Royzman D, Wild AB, Zinser E, et al. The CD83 molecule - an important immune checkpoint. *Front Immunol*. 2020;11:721.
62. Hirsch VM, Dapolito G, Johnson PR, Elkins WR, London WT, Montali RJ, et al. Induction of AIDS by simian immunodeficiency virus from an African green monkey: species-specific variation in pathogenicity correlates with the extent of in vivo replication. *J Virol*. 1995;69:955–67.
63. Kaur A, Grant RM, Means RE, McClure H, Feinberg M, Johnson RP. Diverse host responses and outcomes following simian immunodeficiency virus SIVmac239 infection in sooty mangabeys and rhesus macaques. *J Virol*. 1998;72:9597–611.
64. Silvestri G. Naturally SIV-infected sooty mangabeys: are we closer to understanding why they do not develop AIDS? *J Med Primatol*. 2005;34:243–52.
65. Paiardini M, Cervasi B, Reyes-Aviles E, Micci L, Ortiz AM, Chahroudi A, et al. Low levels of SIV infection in sooty mangabey central memory CD4<sup>+</sup> T cells are associated with limited CCR5 expression. *Nat Med*. 2011;17:830–6.
66. Beaumier CM, Harris LD, Goldstein S, Klatt NR, Whitted S, McGinty J, et al. CD4 downregulation by memory CD4<sup>+</sup> T cells in vivo renders African green monkeys resistant to progressive SIVagm infection. *Nat Med*. 2009;15:879–85.
67. Bosinger SE, Utay NS. Type I interferon: understanding its role in HIV pathogenesis and therapy. *Curr HIV/AIDS Rep*. 2015;12:41–53.
68. Bosinger SE, Jacquelin B, Benecke A, Silvestri G, Müller-Trutwin M. Systems biology of natural simian immunodeficiency virus infections. *Curr Opin HIV AIDS*. 2012;7:71–8.
69. Cheng L, Yu H, Li G, Li F, Ma J, Li J, et al. Type I interferons suppress viral replication but contribute to T cell depletion and dysfunction during chronic HIV-1 infection. *JCI Insight*. 2017;2:<https://doi.org/10.1172/jci.insight.94366>.
70. Soper A, Kimura I, Nagaoka S, Konno Y, Yamamoto K, Koyanagi Y, et al. Type I interferon responses by HIV-1 infection: association with disease progression and control. *Front Immunol*. 2017;8:1823.
71. Wang B, Kang W, Zuo J, Kang W, Sun Y. The significance of type-I interferons in the pathogenesis and therapy of human immunodeficiency virus 1 infection. *Front Immunol*. 2017;8:1431.
72. Zhen A, Rezek V, Youn C, Lam B, Chang N, Rick J, et al. Targeting type I interferon-mediated activation restores immune function in chronic HIV infection. *J Clin Invest*. 2017;127:260–8.
73. Barouch DH, Ghneim K, Bosche WJ, Li Y, Berkemeier B, Hull M, et al. Rapid inflammasome activation following mucosal SIV infection of rhesus monkeys. *Cell*. 2016;165:656–67.
74. Nganou-Makamdop K, Billingsley JM, Yaffe Z, O'Connor G, Sharp GK, Ransier A, et al. Type I IFN signaling blockade by a PASylated antagonist during chronic SIV infection suppresses specific inflammatory pathways but does not alter T cell activation or virus replication. *PLoS Pathog*. 2018;14:e1007246.
75. Park A, Iwasaki A. Type I and type III interferons - induction, signaling, evasion, and application to combat COVID-19. *Cell Host Microbe*. 2020;27:870–8.
76. Channappanavar R, Fehr AR, Zheng J, Wohlford-Lenane C, Abrahante JE, Mack M, et al. IFN-I response timing relative to virus replication determines MERS coronavirus infection outcomes. *J Clin Invest*. 2019;129:3625–39.
77. Zhao J, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao J, et al. Rapid generation of a mouse model for Middle East respiratory syndrome. *Proc Natl Acad Sci U S A*. 2014;111:4970–5.
78. Schoggins JW, Rice CM. Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol*. 2011;1:519–25.
79. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*. 2011;472:481–5.
80. Huang I-C, Bailey CC, Weyer JL, Radoshitzky SR, Becker MM, Chiang JJ, et al. Distinct patterns of IFITM-mediated restriction of filoviruses, SARS coronavirus, and influenza A virus. *PLoS Pathog*. 2011;7:e1001258.
81. Weidner JM, Jiang D, Pan X-B, Chang J, Block TM, Guo J-T. Interferon-induced cell membrane proteins, IFITM3 and tetherin, inhibit vesicular stomatitis virus infection via distinct mechanisms. *J Virol*. 2010;84:12646–57.
82. Siegrist F, Ebeling M, Certa U. The small interferon-induced transmembrane genes and proteins. *J Interferon Cytokine Res*. 2011;31:183–97.
83. Muema DM, Akilimali NA, Ndumnego OC, Rasehlo SS, Durgiah R, Ojwach DBA, et al. Association between the cytokine storm, immune cell dynamics, and viral replicative capacity in hyperacute HIV infection. *BMC Med*. 2020;18:81.
84. Stacey AR, Norris PJ, Qin L, Haygreen EA, Taylor E, Heitman J, et al. Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J Virol*. 2009;83:3719–33.
85. Keating SM, Heitman JW, Wu S, Deng X, Stacey AR, Zahn RC, et al. Magnitude and quality of cytokine and chemokine storm during acute infection distinguish nonprogressive and progressive simian immunodeficiency virus infections of nonhuman primates. *J Virol*. 2016;90:10339–50.
86. Bowers NL, Scott Helton E, Huijbregts RPH, Goepfert PA, Heath SL, Hel Z. Immune suppression by neutrophils in HIV-1 infection: role of PD-L1/PD-1 pathway. *PLoS Pathogens*. 2014; e1003993.<https://doi.org/10.1371/journal.ppat.1003993>.
87. Sandborn WJ, Rutgeerts P, Colombel J-F, Ghosh S, Petryka R, Sands BE, et al. Eldelumab [anti-interferon-γ-inducible protein-10 antibody] induction therapy for active Crohn's disease: a randomised, double-blind, placebo-controlled phase IIa study. *J Crohns Colitis*. 2017;11:811–9.
88. Zicari S, Sessa L, Cotugno N, Ruggiero A, Morrocchi E, Concato C, et al. Immune activation, inflammation, and non-AIDS co-morbidities in HIV-infected patients under long-term ART. *Viruses*. 2019;11:<https://doi.org/10.3390/v11030200>.
89. Zevin AS, McKinnon L, Burgener A, Klatt NR. Microbial translocation and microbiome dysbiosis in HIV-associated immune activation. *Curr Opin HIV AIDS*. 2016;11:182–90.

90. Sokoya T, Steel HC, Nieuwoudt M, Rossouw TM. HIV as a cause of immune activation and immunosenescence. *Mediators Inflamm.* 2017;2017:6825493.
91. Marchetti G, Tincati C, Silvestri G. Microbial translocation in the pathogenesis of HIV infection and AIDS. *Clin Microbiol Rev.* 2013;26:2–18.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.