



Reference Transcriptomes of Porcine Peripheral Immune Cells Created Through Bulk and Single-Cell RNA Sequencing

Juber Herrera-Uribe^{1†}, Jayne E. Wiarda^{2,3,4†}, Sathesh K. Sivasankaran^{2,5}, Lance Daharsh¹, Haibo Liu¹, Kristen A. Byrne², Timothy P. L. Smith⁶, Joan K. Lunney⁷, Crystal L. Loving^{2**} and Christopher K. Tuggle^{1**}

OPEN ACCESS

Edited by:

Angela Cánovas,
University of Guelph, Canada

Reviewed by:

Paolo Zambonelli,
University of Bologna, Italy
Sergey V. Kozyrev,
Uppsala University, Sweden

*Correspondence:

Crystal L. Loving
crystal.loving@usda.gov
Christopher K. Tuggle
cktuggle@iastate.edu

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

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

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 31 March 2021

Accepted: 18 May 2021

Published: 23 June 2021

Citation:

Herrera-Uribe J, Wiarda JE,
Sivasankaran SK, Daharsh L, Liu H,
Byrne KA, Smith TPL, Lunney JK,
Loving CL and Tuggle CK (2021)
Reference Transcriptomes of Porcine
Peripheral Immune Cells Created
Through Bulk and Single-Cell RNA
Sequencing.
Front. Genet. 12:689406.
doi: 10.3389/fgene.2021.689406

¹ Department of Animal Science, Iowa State University, Ames, IA, United States, ² Food Safety and Enteric Pathogens Research Unit, National Animal Disease Center, Agricultural Research Service, United States Department of Agriculture, Ames, IA, United States, ³ Immunobiology Graduate Program, Iowa State University, Ames, IA, United States, ⁴ Oak Ridge Institute for Science and Education, Agricultural Research Service Participation Program, Oak Ridge, TN, United States, ⁵ Genome Informatics Facility, Iowa State University, Ames, IA, United States, ⁶ USDA, ARS, U.S. Meat Animal Research Center, Clay Center, NE, United States, ⁷ USDA-ARS, Beltsville Agricultural Research Center, Animal Parasitic Diseases Laboratory, Beltsville, MD, United States

Pigs are a valuable human biomedical model and an important protein source supporting global food security. The transcriptomes of peripheral blood immune cells in pigs were defined at the bulk cell-type and single cell levels. First, eight cell types were isolated in bulk from peripheral blood mononuclear cells (PBMCs) by cell sorting, representing Myeloid, NK cells and specific populations of T and B-cells. Transcriptomes for each bulk population of cells were generated by RNA-seq with 10,974 expressed genes detected. Pairwise comparisons between cell types revealed specific expression, while enrichment analysis identified 1,885 to 3,591 significantly enriched genes across all 8 cell types. Gene Ontology analysis for the top 25% of significantly enriched genes (SEG) showed high enrichment of biological processes related to the nature of each cell type. Comparison of gene expression indicated highly significant correlations between pig cells and corresponding human PBMC bulk RNA-seq data available in Haemopedia. Second, higher resolution of distinct cell populations was obtained by single-cell RNA-sequencing (scRNA-seq) of PBMC. Seven PBMC samples were partitioned and sequenced that produced 28,810 single cell transcriptomes distributed across 36 clusters and classified into 13 general cell types including plasmacytoid dendritic cells (DC), conventional DCs, monocytes, B-cell, conventional CD4 and CD8 $\alpha\beta$ T-cells, NK cells, and $\gamma\delta$ T-cells. Signature gene sets from the human Haemopedia data were assessed for relative enrichment in genes expressed in pig cells and integration of pig scRNA-seq with a public human scRNA-seq dataset provided further validation for similarity between human and pig data. The sorted porcine bulk RNAseq dataset informed classification of scRNA-seq PBMC populations; specifically, an integration of the datasets showed that the pig bulk RNAseq data helped define the CD4CD8 double-positive T-cell populations in the scRNA-seq data. Overall, the data provides deep and well-validated transcriptomic data from sorted PBMC populations and the first single-cell

transcriptomic data for porcine PBMCs. This resource will be invaluable for annotation of pig genes controlling immunogenetic traits as part of the porcine Functional Annotation of Animal Genomes (FAANG) project, as well as further study of, and development of new reagents for, porcine immunology.

Keywords: pig, immune cells, transcriptome, single-cell RNA-seq, bulkRNA-seq, FAANG

INTRODUCTION

A major goal of biological research is using genomic information to predict complex phenotypes of individuals or individual cells with specific genotypes. Predicting complex phenotypes is an important component of broad Genome-to-Phenome (G2P) understanding (Koltjes et al., 2019), and investing in sequencing of multiple animal genomes, including pigs (*Sus scrofa*), for improved genome and cell functional annotation is key in solving the G2P question (Andersson et al., 2015; Giuffra et al., 2019). In addition to their major role in the world supply of dietary protein, pigs have anatomic, physiologic, and genetic similarities to humans and serve as biomedical models for human disease and regenerative medicine (reviewed in Swindle et al., 2012; Kobayashi et al., 2018). Thus, deep annotation of porcine genome function would be a major milestone for addressing the G2P question. A highly contiguous porcine genome assembly with gene model-level annotation was recently published (Warr et al., 2020). However, this annotation is based primarily on RNA sequencing (RNA-seq) data from solid tissues, with few sample types representative of immune cells, with the exception of alveolar macrophages and dendritic cells (Auray et al., 2016). Given the interaction of animal health and growth, any functional annotation of the porcine genome will be incomplete without deep analysis of expression patterns and regulatory elements controlling the immune system.

The transcriptomes of circulating immune cells serve as a window into porcine immune physiology and traits (Chaussabel et al., 2010; Mach et al., 2013; Schroyen and Tuggle, 2015; Auray et al., 2020). Blood RNA profiling has been used to understand variation in porcine immune responses (Huang et al., 2011; Arceo et al., 2013; Knetter et al., 2015; Munyaka et al., 2019) and genetic control of gene expression (Maroille et al., 2017). One goal of such research is to develop gene signatures predictive of disease states (Berry et al., 2010) and predict responses to immunizations

and/or infections (Chaussabel and Baldwin, 2014; Tsang et al., 2014), as has been demonstrated in humans. Whole blood is easily collected from live animals, but represents an extremely complex mixture of cell types. Estimates of gene expression in mixed samples are inherently inaccurate as cell composition differences are difficult to adjust for, complicating the interpretation of RNA differences across samples and treatments. Thus, starting from whole blood transcriptomic data, it is nearly impossible to link gene expression and regulation to a specific cell or cell type. To determine direct regulatory interactions, we must analyze specific cell populations and even individual cells. A cell type-specific understanding of peripheral immune cell gene expression patterns will thus enhance biological understanding of porcine immunity, reveal targets for phenotyping, and provide a comparison to other species.

Predominant immune cell populations in porcine peripheral blood mononuclear cell (PBMC) preparations are comprised mainly of monocytes, B-cells, and T-cells, with minor fractions of dendritic cells (DCs), natural killer (NK) cells, and NKT-cells also present. Porcine peripheral T-cell populations (reviewed in Gerner et al., 2009, 2015) and DCs (Summerfield et al., 2015; Auray et al., 2016) are readily described based on phenotype, though deeper characterization of porcine immune cells could improve identification of valuable reagent targets and biological understanding of porcine immunity. T-cell populations are commonly grouped as $\alpha\beta$ or $\gamma\delta$ T-cells according to T-cell receptor (TCR) chain expression and further divided based on CD2, CD4, CD8 α , and/or CD8 β expression. Pigs have a unique CD2⁻ $\gamma\delta$ T-cell lineage contributing to higher percentages of circulating $\gamma\delta$ T-cells (Takamatsu et al., 2006) and unique $\alpha\beta$ T-cells expressing both CD4 and CD8 α (Zuckermann, 1999). Relatively little is known about different circulating B-cell populations in pigs, as reagents for phenotyping are limited.

Various technical approaches can be used to enrich or isolate specific cell populations, improving resolution of cell types for deeper interrogation of gene expression. Flow cytometry is used to characterize cells based on expression of cell type-specific protein markers, and live cells can be sorted by magnetic- and/or fluorescence-activated cell sorting (MACS/FACS) for use in subsequent assays. MACS/FACS enrichment followed by transcriptomic analysis can provide additional insight of gene expression in specific cell types, but cells expressing the same combination of markers are often still a heterogeneous mixture (Sutermaster and Darling, 2019). Some major subtypes of porcine immune cell populations can be labeled for cell sorting by existing antibody reagents (Gerner et al., 2009), but some subtypes such as B-cells lack these resources.

An exciting alternative to sorting specific cell types for transcriptomic analysis is single-cell RNA-seq (scRNA-seq).

Abbreviations: AUC, area under the curve; ASC, antibody-secreting cell; B, B-cell; bulkRNA-seq, bulk RNA sequencing; cDC, conventional dendritic cell; DC, dendritic cell; DEGs, differentially expressed genes; DGE, differential gene expression; Exp, experiment; FAANG, Functional Annotation of Animal Genomes; FACS, Fluorescent activated cell sorting; G2P, Genome-to-Phenome; GO, gene ontology; GSEA, gene set enrichment analysis/analyses; HBSS, Hank's balanced salt solution; HEGs, highly enriched genes; MACS, Magnetic activated cell sorting; mDC/myDC, myeloid dendritic cell; n, negative; NK, natural killer; p, positive; PBMC, peripheral blood mononuclear cell; PC, principal component; PCA, principal component analysis; pDC, plasmacytoid dendritic cell; RF, random forest; RIN, RNA integrity number; RNA-seq, RNA sequencing; scRNA-seq, single-cell RNA sequencing; scREF-matrix, single-cell reference matrix; SEG, significantly enriched genes; sPCA, supervised principal component analysis; SWC6, swine workshop cluster 6; T, T-cell; TCR, T-cell receptor; TPM, transcripts per million; t-SNE, t-distributed stochastic neighbor embedding; UMAP, uniform manifold approximation and projection; UMI, unique molecular identifier; $\gamma\delta$, Gamma-delta; $\alpha\beta$, alpha beta.

Many scRNA-seq approaches do not require prior phenotypic/functional information or antibody reagents but instead rely on physical partitioning of cells to uniquely tag transcripts from individual cells and sharpen resolution of subsequent transcriptomic analysis to single cells (Liu and Trapnell, 2016; Vieira Braga et al., 2016; Zheng et al., 2017). scRNA-seq methods have been applied to human PBMCs (Zheng et al., 2017) and provide more accurate and detailed analyses of transcriptional landscapes that can identify new cell types (Villani et al., 2017) when compared to other transcriptomic approaches. There are limitations to scRNA-seq, with tradeoffs in total genes detected per cell vs. total cells captured for analysis, depending on the approach used (Wilson and Göttgens, 2018).

To deeply annotate the porcine genome for peripheral mononuclear immune cell gene expression and further inform phenotype and function of the heterogeneous pool of immune cells in PBMC preparations, two approaches were used to isolate peripheral immune cells for RNA-seq. MACS followed by FACS was used to enrich for eight PBMC populations using population-specific cell surface markers, and RNA isolated from enriched populations was used for bulk RNA-seq (bulkRNA-seq) or a NanoString assay to evaluate gene expression. PBMCs were also subjected to droplet-based partitioning for scRNA-seq. Gene expression patterns of porcine immune cells using different approaches were compared to each other and to multiple human datasets. Complementary methods provided an improved annotation and deeper understanding of porcine PBMCs, as well as explicated datasets for further query by the research community.

MATERIALS AND METHODS

Animals and PBMC Isolation

Four separate PBMC isolations were performed, with different animals used in each experiment. Cells were used for bulkRNA-seq, targeted RNA detection (NanoString), or scRNA-seq. PBMCs from experiments were used as follows: Experiment A (ExpA) for bulkRNA-seq of sorted populations from two ~6-month-old pigs (A1, A2); Experiment B (Exp B) for NanoString and scRNA-seq from three ~12-month-old pigs (B1, B2, B3); Experiment C (ExpC) for scRNA-seq from three ~12-month-old pigs (C1, C2, C3); Experiment D (ExpD) for scRNA-seq from two ~7-week-old pigs (D1, D2). All pigs were crossbred, predominantly Large White and Landrace heritage. All pigs from experiments A, B, and C were male. In experiment D, animal D1 was female, and animal D2 was male. All animal procedures were performed in compliance with and approval by NADC Animal Care and Use Committee. PBMCs were isolated, enumerated, and viability assessed as previously described (Byrne et al., 2020).

Enrichment and Sorting Eight Leukocyte Populations by MACS/FACS

Peripheral blood mononuclear cells were labeled with biotin labeled anti-porcine CD3 ϵ (PPT3, Washington State University Monoclonal Antibody Center) for 15 min at 4°C, mixing continuously. Cells were washed with Hank's Balanced Salt

Solution (HBSS), incubated with anti-biotin microbeads (Miltenyi Biotec), placed on LS columns, and separated into CD3 ϵ^+ and CD3 ϵ^- fractions according to manufacturer's directions (Miltenyi Biotec). CD3 ϵ^+ and CD3 ϵ^- fractions were each fluorescently-sorted into four subpopulations based on surface marker expression shown in **Figure 1** and **Table 1**. For NanoString assays, B-cells were sorted as CD3 ϵ^- CD172 α^- CD8 α^- ; CD21 was not used for sorting. Each fraction for FACS was confirmed CD3 ϵ^+ or CD3 ϵ^- by labeling with anti-mouse IgG1-PE-Cy7 to detect anti-CD3 ϵ antibody used for MACS. Cells were sorted into supplemented HBSS using a BD FACSaria II with 70 mm nozzle. After sorting, cells were pelleted and enumerated as described above. Sorted cell purity was >85% for each population. Cells were stained, sorted, and further processed within 10 h of collection keeping cells on ice between processing steps.

RNA Isolation for BulkRNA-Seq/NanoString

BulkRNA-seq: after FACS, cells were pelleted, enumerated, and immediately lysed in RLT Plus buffer. RNA extractions were performed using the AllPrep DNA/RNA MiniKit (QIAGEN) following manufacturer's instructions. Eluted RNA was treated with RNase-free DNase (QIAGEN). RNA quantity/integrity were assessed with an Agilent 2200 TapeStation system (Agilent Technologies). Samples used had RNA integrity numbers (RINs) ≥ 7.9 . From ExpA, only one RNA sample for NK cells was used.

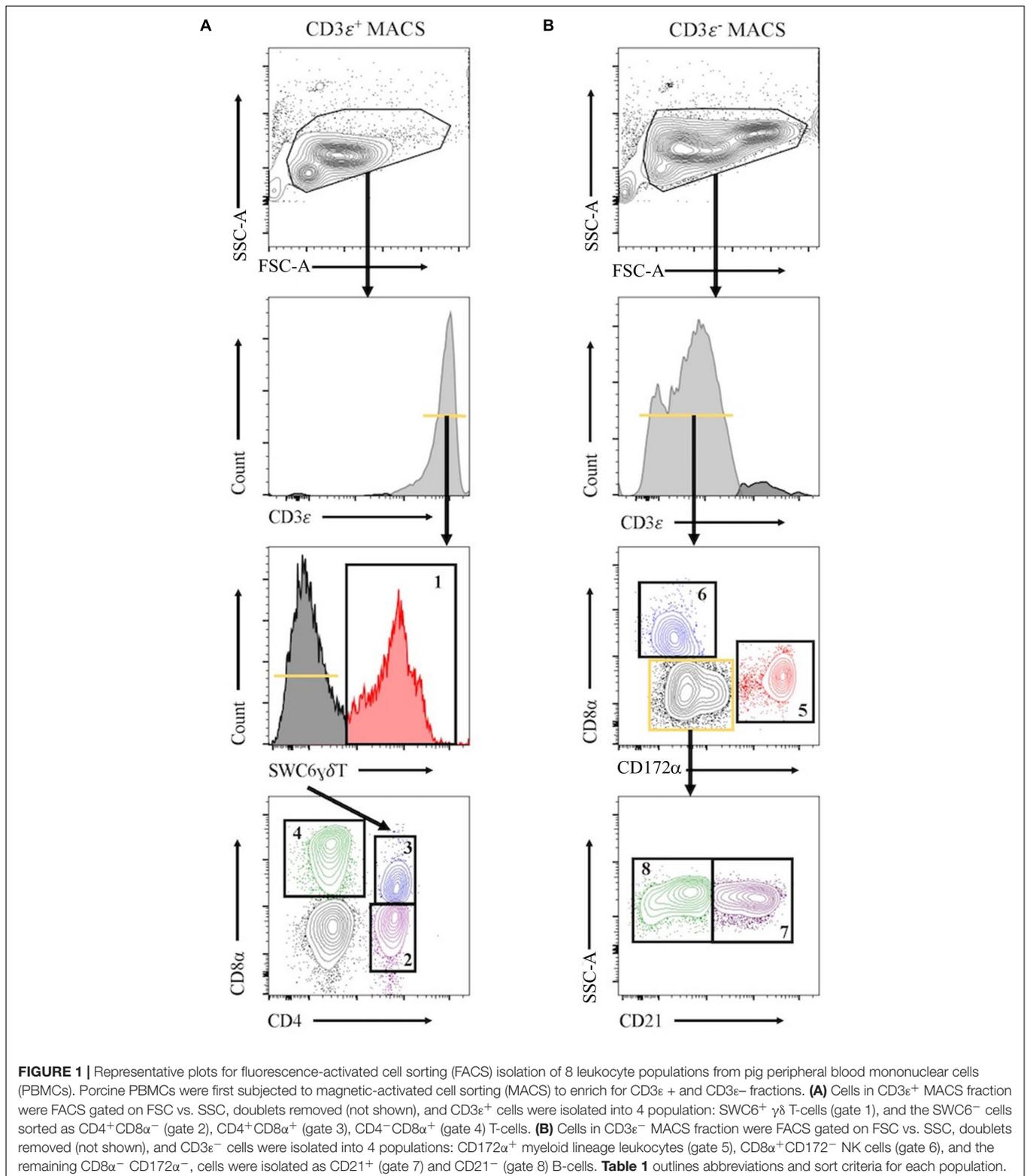
For NanoString assay: after FACS, cells were pelleted, enumerated, and immediately stored in Trizol. RNA extraction was performed using the Direct-zol RNA MicroPrep Kit (Zymo) with on-column DNase treatment following manufacturer's instructions. RNA quantity and integrity were assessed as described above, with RINs ≥ 6.9 . RNA was preserved at -80°C until further use.

BulkRNA-Seq Library Preparation and Data Analysis

RNA was fragmented and 15 libraries prepared using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina). Libraries were diluted and pooled in approximately equimolar amounts. Pooled libraries were sequenced in paired-end mode (2×150 -bp reads) using an Illumina NextSeq 500 (300 cycle kit).

Preprocessing, Mapping, Alignment, Quality Control

Data processing was performed as previously reported (Herrera-Uribe et al., 2020) using R v4.0.3. Sscrofa 11.1 genome and annotation v11.1.97 were used. Counts per gene of each sample in the two count tables were added together to get the final count table. Given that different types of immune cells have different transcriptome profiles (Hicks and Irizarry, 2015), YARN (Paulson et al., 2017), a tissue type-aware RNA-seq data normalization tool, was used to filter and normalize the count table. Genes with extremely low expression levels (<4 counts in at least one cell type) were filtered out using *filterLowGenes()*. The final count table contained 12,261 genes across 15 samples, which was then normalized using *normalizeTissueAware()*, which leverages the smooth quantile normalization method (Hicks et al., 2018).



Data quality control was performed using DESeq2 (v1.24.0) (Love et al., 2014) within RStudio s (v1.2.1335). Regularized log-transformation was applied to the normalized count table

with the `rld` function. Then principal component analysis (PCA) and sample similarity analyses were carried out and visualized using `plotPCA()` and `distancePlot()`, respectively.

Heatmaps to display enriched genes were created using pheatmap (v1.0.12) within RStudio.

Cell Type-Enriched and Cell Type-Specific Gene Identification

The normalized count table was used for differential gene expression (DGE) analysis with DESeq2 by setting the size factor for each sample to 1. A generalized linear model was fitted for each gene in the count table, with negative binomial response and log link function of the effect of cell types and pig subjects. *nbinomWaldTest()* was used to estimate and test the significance of regression coefficients with the following explicit parameter settings: *betaPrior = FALSE, maxit = 50000, useOptim = TRUE, useT = FALSE, useQR = TRUE*. Cell type-enriched genes and cell type-specific genes were identified using the results function separately. A gene was labeled as cell-type enriched if the expression level (averaged across replicates) in one cell type was at least 2× higher than the average across all remaining cell types and adjusted *p*-value < 0.05. A gene was labeled as cell type-specific if the averaged expression level in one cell type was at least 2× higher in pairwise comparison to the average in each other cell type and adjusted *p*-value < 0.05 (Benjamini and Hochberg, 1995). Heatmaps to display specific genes were created as mentioned above.

For cross-species comparison, human hematopoietic cell (Haemopedia) RNA-seq expression data (Hilton Laboratory at the Walter and Eliza Hall Institute¹) was used. Only orthologous genes with one-to-one matches between human and pig [orthologous gene list obtained from BioMart (Durinck et al., 2009)] were compared. Orthologous gene transcript per million (TPM) values from naive and memory B-cells, myeloid dendritic cells (myDC), myeloid dendritic cells CD123 + (CD123PmDC), plasmacytoid DC (pDC), monocytes, NK cells, CD4T and CD8T cells from healthy donors were used (Choi et al., 2019). Spearman rank correlation analyses was performed to identify correlation between orthologous gene expression levels (absolute TPM) in pig and human sorted populations. Significance level was set at

¹<https://www.haemosphere.org/datasets/show>

p < 0.05 and level of Spearman's rank correlation coefficient (rho) was defined as low (<0.29), moderate (0.3–0.49), and strong (0.5–1) correlation.

Gene Ontology (GO) Enrichment Analysis

Metascape analysis (Zhou et al., 2019) was performed for GO analysis of the top 25% enriched genes and specific genes identified as described above, with threshold *p*-value < 0.01. Several terms were clustered into the most enriched GO term. Term pairs with Kappa similarity score > 0.3 were displayed as a network to show relationship among enriched terms. Terms associated with more genes tended to have lower *p*-values. All networks displayed were visualized using Cytoscape. All Ensembl Gene IDs with detectable expression level in each cell type were used as the background reference.

NanoString Assay and Data Analysis

A total of 230 test genes with nine housekeeping genes, eight positive and nine negative control genes were chosen for gene expression quantification on the NanoString nCounter analysis system (NanoString Technologies) using custom-made probes. The custom designed CodeSet was selected from genes and pathways associated with porcine blood, lung, lymph node, endometrium, placenta or macrophage response to infection with a porcine virus (Van Goor et al., 2020). RNA samples were diluted to 25–100 ng/ul in RNase-free water, and 5 ul of each sample was used in the assay using manufacturer's instructions with the nCounter Master kit.

The nCounter analysis system produces discrete count data for each gene assayed within each sample. We used the NanoString software nSolver Analysis Software (v3.0, NanoString Technologies), following manufacturer's instructions. The nSolver corrected for background based on negative control samples, performed within-sample normalization based on positive control probes, and performed normalization across samples using the median expression values of housekeeping genes (*GAPDH*, *HMBS*, *HPRT1*, *RPL32*, *RPL4*, *SDHA*, *TBP*, *TOP2B*, *YWHAZ*), providing confidence in our normalization method.

TABLE 1 | Abbreviations and phenotype information of pig sorted immune cells.

Gate ^a	Population Abbreviation	Marker	Clone	Fluorophore	Company (Catalog #) ^b	FACS Sort Criteria
CD3e⁺ MACS Fraction						
–	–	Anti mouse IgG1	RMG1-1	PE-Cy7	BioLegend(406614)	–
1	SWC6gdT	SWC6gdT	MAC320	APC	BD(561482)	CD3e ⁺ SWC6 ⁺
2	CD4T	CD4	74-12-4	FITC	BD(559585)	CD3e ⁺ SWC6 [–] CD4 ⁺ CD8α [–]
3	CD4CD8T	CD4CD8α	74-12-4/76-2-11	FITC/PE	BD(559585)/BD(559584)	CD3e ⁺ SWC6 [–] CD4 ⁺ CD8α ⁺
4	CD8T	CD8α	76-2-11	PE	BD(559584)	CD3e ⁺ SWC6 [–] CD4 [–] CD8α ⁺
CD3e[–] MACS Fraction						
5	Myeloid	CD172	74-22-15A	FITC	BD(561498)	CD3e [–] CD172α ⁺ CD8α [–]
6	NK	CD8α	76-2-11	PE	BD(559584)	CD3e [–] CD172α [–] CD8α ⁺
7	CD21pB	CD21	BB6-11C9.6	AF647	Southern Biotech(4530-31)	CD3e [–] CD172α [–] CD8α [–] CD21 ⁺
8	CD21nB	–	–	–	–	CD3e [–] CD172α [–] CD8α [–] CD21 [–]

^aRefers to gate in **Figure 1**. ^bReagents listed in **Materials and Methods**.

All statistical analyses were performed using the statistical programming language R v3.5. Raw count data were normalized using *normalizationFactors()* and *NanoStringDataNormalization()* from NanoStringDiff (v1.1.2.0) (Wang et al., 2016). One gene (*ISG20*) without detected expression in any samples was removed. Hierarchical clustering and PCA suggested there were substantial hidden variations among the expression data. Surrogate variable analysis has been shown to be a powerful method to detect and adjust for hidden variations in high throughput gene expression data (Li et al., 2014; Qian Liu, 2016), so surrogate variable analysis was applied to remove further hidden variations in the gene expression data using *svaseq()* from sva (v3.30.1) (Leek et al., 2012). A full model with cell subpopulations and RINs as independent variables, and a reduced model with RINs as the only independent variable were used. Three surrogate variables were estimated and used to adjust for the hidden variations.

Gene expression values were transformed to $\log_2(\text{TPM})$ using *voom()* from limma (Law et al., 2014). Linear mixed effect models were used to fit the transformed gene expression data by using *lmer()* in lme4 (Bates et al., 2015). The model included fixed effect for cell subpopulation, RIN, the three surrogate variables, and random effect for each animal. One minus Spearman correlation coefficient was used as distance measure for gene clustering, and Euclidian distances was used for sample clustering.

Additionally, Spearman correlation analysis was performed to assess the correlation between bulkRNA-seq and NanoString results. The significant level was set at $p < 0.05$, and the level of Spearman's rank correlation coefficient (ρ) was defined as described above.

scRNA-Seq Library Preparation

Peripheral blood mononuclear cell isolation experiments were performed at different times and samples sequenced in different runs. For ExpB, 1×10^7 viable PBMCs per animal were cryopreserved according to 10 \times Genomics Sample Preparation Demonstrated Protocol, shipped on dry ice to University of Minnesota's Core Sequencing Facility, and thawed, partitioned, and scRNA-seq libraries prepared. For ExpC/ExpD, freshly isolated PBMCs were transported on ice to Iowa State University Core Sequencing Facility for partitioning and library preparation. Partitioning and library preparation were performed according to Chromium Single Cell 3' Reagent Kits v2 User Guide (10 \times Genomics). For all experiments, 100 base paired-end reads were sequenced on an Illumina HiSeq3000 at ISU Core Sequencing Facility. One sample from ExpB was omitted from further analyses due to poor sequence performance.

scRNA-Seq Data Analysis

Read Alignment/Gene Quantification

Raw read quality was checked with FASTQC². Reads 2 (R2) were corrected for errors using Rcorrector (Song and Florea, 2015), and 3' polyA tails > 10 bases were trimmed. After trimming, R2 > 25 bases were re-paired using BMAP³. *Sus*

scrofa genome Sscrofa 11.1 and annotation GTF (v11.1.97) from Ensembl were used to build the reference genome index (Yates et al., 2020). The annotation file was modified to include both gene symbol (if available) and Ensembl ID as gene reference (e.g., GZMA_ENSSSCG00000016903) using custom Perl scripts. Processed paired-end reads were aligned and gene expression count matrices generated using Cell Ranger (v4.0; 10 \times Genomics) with default parameters. Only reads that were confidently mapped (MAPQ = 255), non-PCR duplicates with valid barcodes, and unique molecular identifiers (UMIs) were used to generate gene expression count matrices. Reads with same cell barcodes, same UMIs, and/or mapped to the same gene feature were collapsed into a single read.

Quality Control/Filtering

All quality control/filtering steps were performed using R v3.6.2. Cell Ranger output files were used to remove ambient RNA from each sample with SoupX (Young and Behjati, 2020) function *autoEstCont()*. Corrected non-integer gene count matrices were outputted in Cell Ranger file format using DropletUtils (Lun et al., 2019) function *write10xCounts()* and used for further analyses. Non-expressed genes (sum zero across all samples) and poor quality cells (>10% mitochondrial genes, < 500 genes, or < 1,000 UMIs per cell) were removed using custom R scripts and Seurat (Stuart et al., 2019). Filtered count matrices were generated using *write10xCounts()* and used for further analyses. High probability doublets were removed using Scrublet (Wolock et al., 2019), specifying 0.07 expected doublet rate and doublet score threshold of 0.25.

Integration, Visualization, and Clustering

Integration, visualization, and clustering were performed with R v3.6.2 and Seurat v3.2.0. Post-quality control/filtering gene counts/cells from each sample were loaded into a Seurat object and transformed individually using *SCTransform()*. Data were integrated with *SelectIntegrationFeatures()*, *PrepSCTIntegration()*, *FindIntegrationAnchors()*, and *IntegrateData()* with default parameters. PCA was conducted with *RunPCA()*, and the first 14 principal components (PCs) were selected as significant based on <0.1% variation of successive PCs. Significant PCs were used to generate two-dimensional t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) coordinates for visualization with *RunTSNE()* and *RunUMAP()*, respectively, identify nearest neighbors and clusters with *FindNeighbors()* and *FindClusters()* (clustering resolution = 1.85), respectively, and perform hierarchical clustering with *BuildClusterTree()*. Counts in the RNA assay were further normalized and scaled using *NormalizeData()* and *ScaleData()*.

Differential Gene Expression (DGE) Analyses

Differential gene expression analyses were performed with R v3.6.2 and Seurat v3.2.0. Normalized counts from the RNA assay were used for DGE analyses. Differentially-expressed genes (DEGs) between pairwise cluster combinations were calculated using *FindMarkers()*. DEGs in one cluster relative to the

²<https://www.bioinformatics.babraham.ac.uk/projects/fastqc>

³<https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/>

average of all other cells in the dataset were calculated using *FindAllMarkers()*. The default Wilcoxon Rank Sum test was used for DGE analyses. Genes expressed in >20% of cells within one of the cell populations being compared, with $|\log_{2}FC| > 0.25$, and adjusted *p*-value < 0.05 were considered DEGs.

Gene Set Enrichment Analyses (GSEA)

Enrichment of gene sets within our porcine scRNA-seq dataset were performed using AUCCell (v1.10.0) (Aibar et al., 2017). Enriched genes in sorted porcine bulkRNA-seq populations were identified as described in preceding methods. Log₂FC values were used to curate gene sets of genes enriched in the top 25, 20, 15, 10, 5, or 1% of bulkRNA-seq populations. Gene sets from human bulkRNA-seq cell populations (Choi et al., 2019) were recovered by performing a High Expression Search on the Haemosphere website⁴, setting Dataset = Haemopedia-Human-RNASeq and Sample group = celltype. Gene sets for CD4: + T-cell; CD8: + T-cell; Memory B-cell; Monocyte; Myeloid Dendritic Cell; Myeloid Dendritic Cell CD123 + ; Naïve B-cell; Natural Killer Cell; and Plasmacytoid Dendritic Cell options corresponded to CD4T, CD8T, MemoryB, Monocyte, mDC, CD123PmDC, NaïveB, NK, and pDC designations, respectively. Genes with high expression scores >0.5 (lower enrichment level) or > 1.0 (higher enrichment level) were selected and filtered to include only one-to-one gene orthologs as described in preceding methods. Human gene identifiers were converted to corresponding porcine gene identifiers or gene names used for scRNA-seq analyses.

Within each cell of the finalized scRNA-seq dataset, gene expression was ranked from raw gene counts. Area under the curve (AUC) scores were calculated from the top 5% of expressed genes in a cell and the generated gene sets. Higher AUC scores indicated a higher percentage of genes from a gene set were found amongst the top expressed genes for a cell. For overlay of AUC scores onto UMAP coordinates of the scRNA-seq dataset, a threshold value was manually set for each gene set based on AUC score distributions. For visualization by heatmap, AUC scores were calculated for each cell, scaled relative to all other cells in the dataset, and average scaled AUC scores were calculated for each cluster. R v4.0.2 was used.

Deconvolution Analysis (CIBERSORTx)

To deconvolve cluster-specific cell subsets from bulkRNA-seq of sorted populations, CIBERSORTx (Newman et al., 2019) was used to derive a signature matrix from scRNA-seq data. 114 cells were taken from each cluster using the Seurat *subset()* function and labeled with corresponding cluster identities. Cluster-labeled cells were used to obtain a single-cell reference matrix (scREF-matrix) that was used as input and run on CIBERSORTx online server using “Custom” option. Default values for replicates (5), sampling (0.5), and fraction (0.0) were used. Additional options for kappa (999), *q*-value (0.01), and No. Barcode Genes (300–500) were kept at default values. CIBERSORTx scREF-matrix was used to impute cell fractions from the bulkRNA-seq of sorted cell population “mixtures.” The mixture file (TPM values) was used as an input and run on CIBERSORTx online server using

the “Impute Cell Fractions” analysis with the “Custom” option selected, and S-mode batch-correction was applied. Cell fractions were run in relative mode to normalize results to 100%. The number of permutations to test for significance were kept at default (100). Resulting output provided estimated percentages of what scRNA-seq clusters defined each bulkRNA-seq sorted cell population.

Reference-Based Label Transfer/Mapping and *de novo* Integration/Visualization

R v3.6.3 and Seurat v3.9.9.9010 were used for the analyses described in this section. A CITE-seq dataset of human PBMCs (Hao et al., 2020) was used to transfer cell type annotations onto our porcine scRNA-seq dataset. Due to the cross-species comparison, we distilled human reference and pig query datasets to only include 1:1 orthologous gene, and human reference dataset was re-normalized and integrated mirroring previous methods (Hao et al., 2020). Each sample of the porcine query dataset was separately normalized using SCTransform. Anchors were found between the human reference and each pig query sample using FindTransferAnchors. Identified anchors were used to calculate mapping scores for each cell using MappingScore. The mapping scores provided a 0–1 confidence value of how well a porcine cell was represented by the human reference dataset. Prediction scores were calculated using available level 2 cell types from the human reference dataset. Prediction scores provided a 0–1 percentage value for an individual cell type prediction, based on how many nearby human cells shared the same cell type annotation that was predicted. Predicted cell annotations were projected back onto original UMAP of the porcine dataset. Cluster-averaged prediction and mapping scores were also calculated.

In order to identify cells from the porcine dataset that were not well represented by the human reference dataset the two datasets were integrated to perform *de novo* visualization by merging the two datasets and their respective sPCAs to create a new UMAP. From two-dimensional *de novo* UMAP, porcine cells that did not overlap with human cells were identified.

Cluster Subsetting

For deeper analyses of only subsets of clusters in the scRNA-seq dataset, cells belonging to only selected clusters were placed in a new Seurat object using *subset()*. Genes with zero overall expression in the new data subset were removed using *DietSeurat()*, and counts were re-scaled with *ScaleData()*. Original cluster designations and PCs were left intact. UMAP/t-SNE visualization, hierarchical clustering, and DGE analyses were re-performed as described in the original analyses. Pairwise DGE analyses were not re-performed. R v4.0.2 was used.

Random Forest (RF) Modeling

Random forest modeling was performed with R version 4.0.4. The RF models provided an estimate of cluster similarity based on error rates. The R packages *caret*⁵ and *ranger*⁶ were used to create

⁴<https://www.haemosphere.org/searches>

⁵<https://cran.r-project.org/web/packages/caret/caret.pdf>

⁶<https://cran.r-project.org/web/packages/ranger/index.html>

RF models trained on cluster identities of cells. A normalized count matrix was used as input data for RF models. Each cell was labeled by its previously defined cluster. Two different types of models were created: (1) pairwise models where training data included only cells from two different clusters (ex. Clusters 0 and 3); (2) models where training data included cells from all clusters of a specified dataset (ex. all $\gamma\delta$ T-cell clusters). Each model was trained on the cluster identity of each cell, with trees created = 500, target node size = 1, variables = 14,386, variables to sample at each split (Mtry) = 119. Each tree in the model is grown from a bootstrap resampling process that calculates an out-of-bag (OOB) error that provides an efficient and reasonable approximation of the test error. Variable importance was used to find genes or sets of genes that can be used to identify certain types of cells or discriminate groups of cells from one another. RF models are advantageous because they can provide ranked lists of genes most important for discriminating cells between different clusters. This method was used to identify groups of important genes to supplement single DGE analyses. Variable importance was assigned by measuring node impurity (Impurity) and using permutations (Permutation). Features that reduced error in predictive accuracy are ranked as more important. High error rate in the model suggests cells from the groups being compared are more similar to each other, whereas low error rate suggests cells from each cluster are unique.

Gene Name Replacement

Several gene names/Ensembl IDs used for data analysis were replaced in main text/figures for the following reasons: gene symbol was not available in the annotation file but was available under the gene description on Ensembl, gene symbol was updated in future Ensembl releases, or multiple Ensembl IDs corresponded to a single gene symbol. Affected genes included: *ABI3* = ENSSSCG00000035224, *ABRACL* = ENSSSCG0000004145, *ANP32E* = ENSSSCG00000035209, *AP3S1* = ENSSSCG00000037595, *AURKA* = ENSSSCG00000007493, *BANF1* = ENSSSCG00000012961, *BUB1B* = ENSSSCG00000004782, *CBX3* = ENSSSCG00000016711, *CCDC12* = ENSSSCG00000011329, *CCL23* = ENSSSCG00000033457, *CD163L1* = ENSSSCG00000034914, *CDC2* = ENSSSCG00000010214, *CDNF* = ENSSSCG00000039658, *CEP57* = ENSSSCG00000014969, *CLIC1* = ENSSSCG00000039071, *CMC2* = ENSSSCG00000032060, *CR2* = ENSSSCG00000028674, *CRIP1* = ENSSSCG00000037142, *CRK* = ENSSSCG00000038989, *DBF4* = ENSSSCG00000020870, *DEK* = ENSSSCG00000001075, *DHFR* = ENSSSCG00000031117, *EEF1A1* = ENSSSCG00000004489, *KIF23* = ENSSSCG00000004969, *FAM72A* = ENSSSCG00000039370, *FCGR3A* = ENSSSCG00000036618, *GBP1* = ENSSSCG00000024973, *GBP7* = ENSSSCG00000006919, *GFER* = ENSSSCG00000008035, *GGCT* = ENSSSCG00000016679, *GIMAP4* = ENSSSCG00000027826, *GINS1* = ENSSSCG00000034913, *GTSE1* = ENSSSCG00000000002, *GZMA* = ENSSSCG00000016903, *H1-2* = ENSSSCG00000037565, *HMGB1* = ENSSSCG00000009327, *HMGB3* = ENSSSCG00000035908, *HMGNI* = ENSSSCG00000033733, *HMGNI5* = ENSSSCG00000032946, *HNRNPA2B1* = ENSSSCG00000036350, *HNRNPAB* = ENSSSCG00000014031, *HO PX* = ENSSSCG00000008898, *IDH1* = ENSSSCG00000029066, *IFITM1* = ENSSSCG00000014565, *IGLL5* = ENSSSCG0000000

0077, *JPT1* = ENSSSCG00000017213, *KLRB1B* = ENSSSCG00000034555, *KLRC1* = ENSSSCG00000000640, *KLRD1* = ENSSSCG00000026217, *KNL1* = ENSSSCG00000039107, *LSM4* = ENSSSCG00000034314, *LSM5* = ENSSSCG00000026064, *MAGOHB* = ENSSSCG00000000635, *MAL* = ENSSSCG00000040098, *MAN2B1* = ENSSSCG00000013720, *MDK* = ENSSSCG00000013260, *MKI67* = ENSSSCG00000026302, *MYL12A* = ENSSSCG0000003691, *NT5C3A* = ENSSSCG00000022912, *NUSAP1* = ENSSSCG00000035544, *NUTF2* = ENSSSCG00000030295, *PPIA* = ENSSSCG00000016737, *PRIM1* = ENSSSCG00000026055, *PRKCH* = ENSSSCG00000005095, *PTTG1* = ENSSSCG00000017032, *RPL14* = ENSSSCG00000011272, *RPL22L1* = ENSSSCG00000036114, *RPL23A* = ENSSSCG00000035080, *RPL35A* = ENSSSCG00000040273, *RPS15A* = ENSSSCG00000035768, *RPS19* = ENSSSCG0000003042, *RPS27A* = ENSSSCG00000034617, *RPS3* = ENSSSCG00000014855, *RPS8* = ENSSSCG00000003930, *RRM2* = ENSSSCG00000031741, *S100B* = ENSSSCG00000026140, *SEPHS1* = ENSSSCG00000031659, *SIRPA* = ENSSSCG00000028461, *SKA1* = ENSSSCG00000004518, *SLA-DQA1* = ENSSSCG00000001456, *SLA-DRA* = ENSSSCG00000001453 (listed as HLA-DRA in the gene annotation used), *SLA-DRB1* = ENSSSCG00000001455, *SLPI* = ENSSSCG00000022258, *SNRPG* = ENSSSCG00000024776, *SPIB* = ENSSSCG00000034211, *TACC3* = ENSSSCG00000008677, *TMSB15B* = ENSSSCG00000012517, *TMSB4X* = ENSSSCG00000012119, *TUBA1C* = ENSSSCG00000000194, *TXN* = ENSSSCG00000005453, *WEE1* = ENSSSCG00000013411, *WIPF1* = ENSSSCG00000027348, *YBX1* = ENSSSCG00000028485.

RESULTS

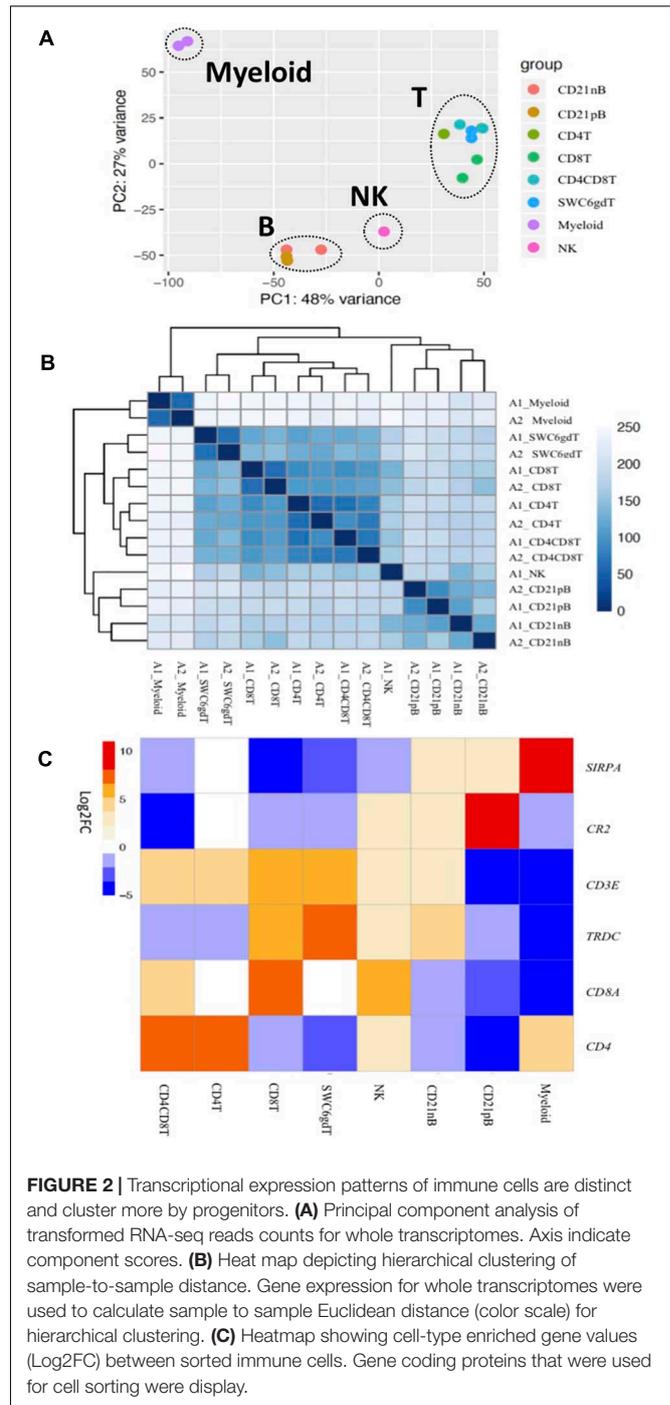
BulkRNA-Seq Revealed Common and Distinct Transcriptomes in Circulating Immune Cells

Eight immune cell populations (Table 1) were sorted by cell-surface marker phenotypes for transcriptomic profiling by bulkRNA-seq (Figure 1) using primarily criteria previously outlined (Gerner et al., 2009), with some modifications. Our protocol utilized an antibody reactive to swine workshop cluster 6 (SWC6) protein to identify $\gamma\delta$ T-cells, but the antibody only labels CD2⁻ $\gamma\delta$ T-cells (Yang and Parkhouse, 1996; Davis et al., 1998; Stepanova and Sinkora, 2013; Sedlak et al., 2014). CD2⁺ $\gamma\delta$ T-cells were likely sorted into the CD3e⁺CD4⁻CD8 α ⁻ fraction that was not retained or the CD8T (CD3e⁺CD4⁻CD8 α ⁺) population (Davis et al., 1998; Stepanova and Sinkora, 2013; Sedlak et al., 2014). A pan-B-cell marker for pigs is not currently available, so B-cells are often characterized through a series of negative gates. Cells in the CD3e⁻ fraction were considered B-cells if they also lacked expression of CD172 α and CD8 α . B-cells characterized in this manner were further terminally sorted into B-cell populations with or without CD21 (complement receptor 2) expression (CD21pB and CD21nB, respectively; Figure 1, gates 7 and 8 respectively). We acknowledge that the CD21nB gate likely contained other circulating cell types that were not sorted through positive gating approaches.

Transcriptomic profiles of sorted cell populations were constructed by bulk RNA-seq, and relationships among porcine immune cell transcriptomes were assessed and visualized through dimensionality reduction and hierarchical clustering (Figures 2A,B and Supplementary File 1). Specifically, T-cell populations (SWC6gdT, CD4T, CD4CD8T, CD8T), B-cell populations (CD21pB, CD21nB), myeloid leukocyte populations (Myeloid), and a single NK cell population (NK) were well separated from each other (Figure 2A) by PCA. Replicates of specific sorted cell populations clustered most closely together, while within T-cell populations or B-cell populations, considerable transcriptional similarity was observed (Figure 2B).

The total number of expressed genes in each sorted population was similar (Supplementary File 1). Significantly enriched genes (SEGs) with expression significantly different and at least $2\times$ greater than the average of all other cell populations (see Methods) were identified for each sorted population (Supplementary File 2). Notably, around 11–23% of SEGs are not fully annotated (no symbol/gene name) in the Sscrofa 11.1 genome and annotation v11.1.97. However, there is evidence of 1:1 orthology within Ensembl for 5–14% of these unannotated genes (Table 2). The SWC6gdT population had the highest number of SEGs (3,591), while the NK population had the fewest (1,885) (Table 2). SEG lists were queried for corresponding protein targets used to sort cells, if known, to confirm enrichment of expression of genes corresponding to protein phenotypes (Figure 2C). Expression of *SIRPA** (encoding CD172 α) had the highest fold-change in the Myeloid population, and *CR2* (encoding CD21, *ENSSSCG0000028674*), was highest in the CD21pB population, as would be predicted based on protein phenotypes. The two CD4⁺ T-cell populations (CD4T and CD4CD8T) had the highest fold-change for *CD4*. The CD8T population had the highest fold-change for *CD8A*, with CD4CD8T and NK populations also having near a \log_2FC enrichment value of 5, in line with these populations also expressing CD8 α . The SWC6gdT population had the highest fold-change for *TRDC*, though CD8T and CD21nB populations also had enrichment for *TRDC*. As noted previously, it's unlikely our sorting for $\gamma\delta$ T-cells based on SWC6 captured all $\gamma\delta$ T-cells, thus some $\gamma\delta$ T-cells may be represented in other sorted populations. Thus, the CD8T population is likely comprised not only CD8 α^+ $\alpha\beta$ T-cells, but also potentially SWC6⁻ $\gamma\delta$ T-cells expressing CD8 α .

A subset of SEGs (25% highest \log_2FC values) for each sorted population, referred to as highly enriched genes (HEGs) that distinguish different circulating pig immune cell populations, were used for data visualization and GO analysis. The \log_2FC values for HEGs were clustered and visualized in Figure 3 (four CD3 ϵ^- populations) and Supplementary Figure 2 (four CD3 ϵ^+ populations). GO analyses using HEG lists for each cell population indicated enrichment for biological processes characteristic of each respective cell population, depicted as networks of similar terms (Figures 3E–H, Supplementary File 3, and Supplementary Figures 2E–H). Terms for Myeloid HEGs included Myeloid leukocyte activation and response to bacterium (Figure 3E), and terms for NK HEGs included positive regulation of cell killing and natural killer cell



mediated cytotoxicity (Figure 3F). Many terms enriched for CD21pB HEGs overlapped with those for CD21nB HEGs, as 38% of HEGs were shared between these populations (Figures 3C,D). Thus, top GO terms for B-cells, including adaptive immune response and B-cell proliferation were present in both populations (Figures 3D,G). However, some GO terms were unique to either B-cell population. GO related to B-cell activation, such as positive regulation of B-cell activation/proliferation processes associated with B-cell receptor

signaling, were identified exclusively for CD21pB HEGs. For CD21nB HEGs, processes associated with humoral immunity and red blood cell processes such as coagulation or platelet activation were noted, which could indicate contamination of different cell types given the non-specific cell sorting approach used for CD21nB cells (Figure 1). For all sorted T-cell populations (CD8T, CD4T, CD4CD8T and SWC6gdT), HEG lists showed overlap (Supplementary Figures 2A–D). GO terms included T-cell activation, T-cell receptor signaling pathway, cytokine-cytokine receptor interaction and biological processes related to cytotoxicity activity (Supplementary Figures 2E,F). Overall, GO exploration of HEGs for sorted populations provided evidence that sorted immune cells represented expected immune cell functions.

The TPM values of expressed genes in sorted porcine cells were compared with orthologous human genes expressed in sorted human naïve hematopoietic cells from the Haemopedia (Choi et al., 2019) in order to identify cell-specific transcriptome similarities across species. Gene expression correlations assessed by Spearman's rank correlation indicated highly significant and moderately strong correlations ($\rho = 0.30\text{--}0.43$, $p < 2.2e\text{--}16$) between porcine and anticipated corresponding human immune cell populations (Supplementary Figure 3 and Supplementary File 4). A closer evaluation of genes reported as canonical cell markers for different mouse and human peripheral immune cell populations and expression of those genes in each of the sorted porcine populations revealed several commonalities. Specifically, genes such as *EBF1*, *CD19*, *MS4A1*, *CD79B*, *PAX5*, *HLA-DOB* (in CD21nB, CD21pB); *CD28* (in CD8T, CD4T, CD4CD8T); *CD5* (in CD8T, CD4T, CD4CD8T, SWC6gdT); *GZMA*, *GZMB*, *GZML*, *KLRK1*, *KLRB1*, *CD244* (in NK, CD8T); and *VLDLR*, *NLRP3*, *CD14*, *STEAP4*, *CD163*, *DEFB1* (in Myeloid) for human cells showed specific enrichment in respective porcine populations (Supplementary Figure 4). Thus, additional query confirmed sorted porcine immune cell populations were equivalent to human counterparts in many ways.

High Homogeneity Amongst Sorted T-Cell and B-Cell Populations and Transcriptomic Distinctions in Myeloid and NK Populations

Pairwise DGE analyses between the cell populations identified genes with transcript abundance at least $2\times$ higher in one population than in all other populations (adjusted p -value < 0.05 , see Methods) which we define as cell type-specific. Consistent with PCA (Figure 2A), more cell type-specific genes were identified in the Myeloid population than in NK, T or B-cells. In total, we identified 2, 5, 8, 29, and 397 cell type-specific genes for CD8T, CD21pB, SWC6gdT, NK, and Myeloid populations, respectively (Table 2 and Supplementary Figure 5). GO analyses using cell type-specific genes for the Myeloid population resulted in enrichment of terms such as Myeloid leukocyte, cytokine-cytokine receptor interaction, and pattern recognition receptor activity (Supplementary Figure 5 and Supplementary File 3). Next, we determined if the cell type-specific genes identified were present in the list of HEGs for each population. In total, 2, 2, 5, 14, and 271 cell type-specific genes were identified in respective HEG lists for CD8T, CD21pB, SWC6gdT, NK, and Myeloid populations, respectively (Table 3), indicating the most highly-enriched cell type-specific genes were present in NK and Myeloid populations. Cell type-specific genes could not be identified for the remaining three sorted populations (CD4T, CD4CD8T, and CD21nB) using the criteria described above, indicating between-population transcriptional heterogeneity even for these enriched populations.

We then explored immune cell transcriptomic patterns to identify genes that could expand our knowledge of pathways active in specific cell populations, as well as predict new genes suitable to use for molecular analyses in immunology studies. Of interest, we found a remarkably high number of HEGs in our Myeloid population (Table 3), including immune-related genes involved in TLR signaling (*CD14*, *CD36*, *TLR2/3/4/8/9*, *NOD2*) and cytokine activity (*CSF1R*, *CSF2RA*, *CSF3R*, *IFNGR1*, *IL1B*, *IL1RAP*, *CXCR2*, *CCL21*, *CCL23*, *TNFRSF1B*, *IL1R2*, *TNFRSF13*, *TNFRSF13B*, *TNFRSF21*, *CXCL16*, *CCR2*). In NK cells fewer specific genes were detected than the Myeloid population (Table 3), with genes such as *OTOP2*, *OTOP3*, *OSPBL3*, *LY6D*, *RET* related to cytotoxic activity, a typical characteristic of NK cells (Rusmini et al., 2013; Rusmini et al., 2014; Belizário et al., 2018; Costanzo et al., 2018; Tu et al., 2018; Upadhyay, 2019), although their function in porcine NK cells is unexplored. In CD21pB cells, the gene for CD21 (*CR2*^{*}) used for sorting the B-cell populations was predicted to be a HEG. The SWC6gdT population showed specific expression of *AVCR2A*, which is a Th17 cell specific gene in mice (Ihn et al., 2011) and regulates the proliferation of $\gamma\delta$ T-cells in murine skin (Antsiferova et al., 2011). The CD8T population specifically expressed *TMIGD2* (a CD28 family member) and *JAML*, which encode T-cell transmembrane proteins (Zhu et al., 2013; Alvarez et al., 2015; Krueger et al., 2017).

TABLE 2 | Cell type-enriched and cell type specific genes identified in pig sorted immune cells.

Cell type	Enriched genes (SEG)	Top 25% SEG	Number of genes without gene name in the top 25% SEG list	Number of top 25% SEG without gene name that have orthologs in human	Specific genes
SWC6gdT	3591	898	141	15	8
CD8T	3318	830	150	19	2
CD4CD8T	2271	568	85	7	0
CD4T	2606	533	95	13	0
NK	1855	464	100	9	29
Myeloid	3440	860	102	15	397
CD21pB	2383	596	124	9	5
CD21nB	2456	614	146	7	0

^{*}Refer to gene name replacement in Materials and Methods section

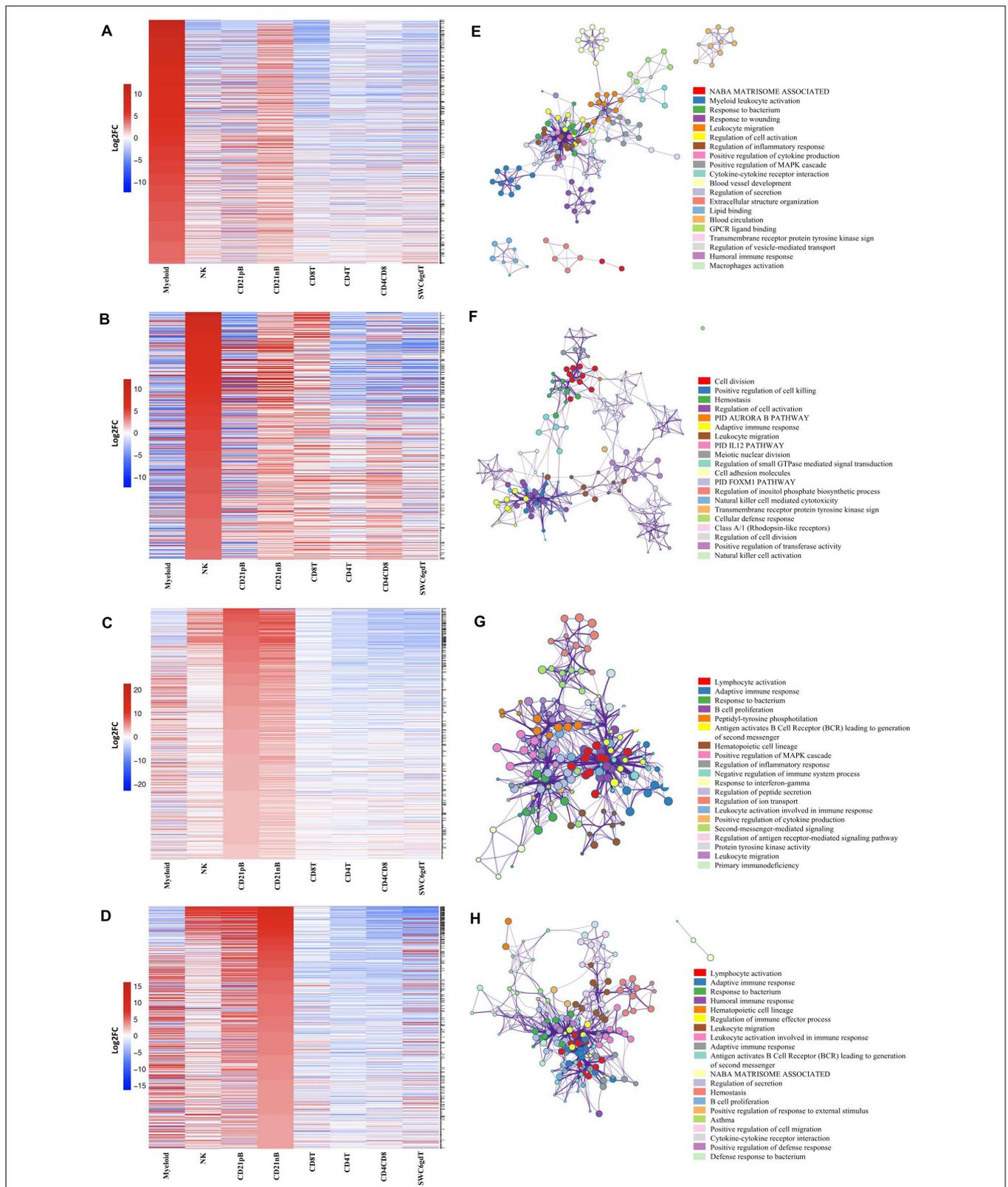


FIGURE 3 | Top 25% highly enriched genes in CD3⁻ sorted cells. Heatmap showing in decreasing order the top 25% of highly enriched genes in **(A)** myeloid, **(B)** NK, **(C)** CD21pB, and **(D)** CD21nB-cells. Ontology enrichment clusters of the top 25% highly enriched genes of **(E)** myeloid, **(F)** NK, **(G)** CD21pB, and **(H)** CD21nB cells. The most statistically significant term within similar term cluster was chosen to represent the cluster. Term color is given by cluster ID and the size of the terms is given by $-\log_{10} p$ -value. The stronger the similarity among terms, the thicker the edges between them.

TABLE 3 | Specific highly enriched genes in myeloid, NK, CD21pB, SWC6gdT, and CD4CD8T-cells.

Group	Total genes	Gene names
Specific Myeloid + Top 25% myeloid	271	<i>SLC18A1, ENSSSCG00000025687, KLHL13, PAK1, C1RL, MITF, SIRPB2, ENSSSCG00000014997, HNMT, C5AR1, A2M, TEK, SEL1L3, TSPAN13, ENSSSCG00000035960, ENSSSCG00000039214, ENSSSCG00000003226, APOE, CHST15, DNMI1, GAS2L1, SERPING1, COL18A1, CDS1, ENSSSCG00000016184, CRHBP, KCNE3, NCAM1, ABHD12, ENSSSCG00000001850, ENSSSCG00000023479, ASAH1, FN1, ENSSSCG00000003554, ENSSSCG00000038429, GAA, ECE1, SLC46A2, UBTD1, CEBPD, CTSB, ENSSSCG00000031640, ENSSSCG00000037466, PLAC9, CCDC60, DOPEY2, TALDO1, ADAMTSL4, ENSSSCG00000034555, STK3, ENSSSCG00000021675, FAM129B, SIGLEC1, SULF2, TRPM2, MGP, CMKLR1, TNFRSF19, DOCK4, ENSSSCG00000027991, ULBP1, SLC11A1, SFXN3, TNFSF13, ENSSSCG00000013380, CD68, KCNQ1, RPS6KA2, CD14, MCF2L, ENSSSCG000000037541, ENSSSCG00000015839, PAM, SERPINB8, TSPAN12, F13A1, SASH1, C9orf72, PLCB4, SH3PXD2B, BLVRA, CXCL2, ADAM28, GPBAR1, CHI3L2, SNX9, LGALS3, SLC2A6, ENSSSCG000000035675, EHD4, ENSSSCG00000039758, UNC13A, ENSSSCG000000038418, C2, PLA2G7, FUCA1, ENSSSCG000000037426, ENSSSCG00000025271, ABCA9, RASGRP4, SLC7A7, VCAN, SLC39A8, ADAP2, SMIM5, DAGLA, RAB11FIP5, ZNF768, ENSSSCG00000007644, CTNND1, ENSSSCG000000022258, ENSSSCG00000017754, STXBP1, ENSSSCG000000027665, MANSC1, RND2, IGSF6, BMX, NLRP12, TPST1, NOD2, TREM1, SEMA6B, JDP2, FAM111B, CIDEB, ENSSSCG00000033457, MMP19, SGK3, CTTNBP2NL, MAPK4, PLAUR, INSIG1, RNASE4, FLVCR2, SCARF1, BCL2L14, ENSSSCG00000026196, MCTP1, WLS, ENSSSCG000000017920, PLOD1, CHPT1, PRCP, ENSSSCG00000013842, SH2D6, CA13, PLCB2, CAPN3, PRAM1, ENSSSCG00000038616, ALOX5, GPNMB, ACVRL1, SMIM3, GPR137B, LAMP1, NR1H3, ARL11, ITGB4, CYSLTR2, CCSE1, NCF2, GPCPD1, PDXK, NACC2, FOLR1, ADGRL2, MERTK, OLFM1, PLXNC1, ECM1, LRRC25, IFIT2, CORO1B, ASAP3, SLC43A3, STEAP4, CAMKK1, CTSS, TMEM47, TTL7, AKR7A2, ENSSSCG00000036342, VIM, TLR8, LIN7A, MPP1, TBXAS1, LIPA, DRAM1, MRC2, TGM3, HEXB, GALM, EREG, JPH4, ANG, QPCT, PPT1, ARRDC4, RAB31, ABHD17C, NFAM1, TLR3, LTB4R, HSD3B7, VDR, ENSSSCG00000010497, CD163, OSCAR, DSC3, LRP6, ENSSSCG00000031951, ENSSSCG00000028635, PSAP, SCPEP1, EPB41L2, ZDHHC9, IL1R2, EXPH5, ENSSSCG00000023264, IFIT5, AGPAT2, NKD2, GUCY1B1, GLUL, COL14A1, TNFRSF1B, SLC16A3, GRN, ENSSSCG00000013100, CEBPA, OLFML2B, TLR4, XG, CCL21, ATF6, SLC49A3, HFE, ACVR1B, IFNGR1, ENSSSCG00000022925, SERPINB10, TCF7L2, ENSSSCG00000008769, ENSSSCG00000016093, UNC93B1, TIMP2, RAMP2, F11R, LGALS8, ENSSSCG00000032723, CFP, ZNF385A, CLIC2, TDRD1, HIP1, ENSSSCG00000026653, GSDMD, CSF1R, NAGK, GAB1, PGD, ENSSSCG00000034639, LRPAP1, DAPK1, ENSSSCG00000039956, GPAT3, GALNTL5, ENSSSCG000000029414</i>
Specific NK + Top 25% NK	14	<i>OTOP2, B3GNT7, OSBPL3, NR4A3, IGF2BP2, OTOP3, ENSSSCG00000010703, LY6D, RET, TUBB6, ENSSSCG00000033385, ENSSSCG00000036743, PTH1R, SUSD1</i>
Specific CD21pB + Top 25% CD21pB	2	<i>GP2, CR2</i>
Specific SWC6gdT + Top 25% SWC6gdT	5	<i>TMEM87B, ACVR2A, ENSSSCG00000028443, SLC4A4, CASS4</i>
Specific CD8T + Top 25% CD8T	2	<i>TMIGD2, JAML</i>

Finally, we compared pair-wise transcriptome differences between our porcine sorted CD4T and CD8T populations (**Supplementary File 2**) with the comparable populations from a previous study (Foissac et al., 2019). Even though the sorting approaches were different, 85% of the genes more highly expressed in CD4T compared to CD8T, respectively, were detected by Foissac and colleagues in their respective CD4⁺ to CD8⁺ comparison. Similar overlap was found (87%) for the genes more abundant in the “CD8 + high” list, while little overlap was found in the inverse comparisons (2.5 and 1%), strongly indicating these cell type gene expression patterns were similar between studies. However, given the lack of identification of cell-type specific genes for CD4T and CD8T populations, shared gene expression patterns may not be surprising.

NanoString Assay Validated BulkRNA-Seq

RNA abundance of each gene target (**Supplementary File 5**) in each sample was used to perform a hierarchical clustering analysis (**Supplementary Figure 6**). Similar to relationships observed in the bulkRNA-seq dataset, biological replicates clustered

most closely together. T-cell populations (SWC6gdT, CD4T, CD4CD8T, CD8T) were more similar to each other than to other populations, with the exception of NK cells. RNA abundance for the genes encoding the marker proteins used for sorting cell populations confirmed cell identity in NanoString assays (**Supplementary Figure 7**). RNA abundance for each tested gene and cell population is included in **Supplementary File 5**. To validate gene expression levels calculated by bulkRNA-seq, a Spearman rank correlation analysis was performed between expression values determined by bulkRNA-seq and NanoString (**Supplementary Figure 8**). Highly significant and strong correlation ($\rho = 0.62-0.88$, p -value $< 2.2e-16$) was observed for all sorted cell types (**Supplementary File 4**). Overall, gene expression estimates in the bulkRNA-seq dataset were confirmed by using the NanoString assay.

Defining the Transcriptomic Landscape of Porcine PBMCs at Single-Cell Resolution

Single-cells from PBMCs of seven conventional pigs were partitioned, sequenced, clustered, and visualized

(**Supplementary File 6**). In total, the final dataset included 28,810 cells, and each cell was assigned to one of 36 transcriptionally distinct clusters, with 9,176–12,683 genes detected within each cluster (**Figure 4A**, **Supplementary Figures 9A–C**; and **Supplementary File 6**). For identification of general cell types in each cluster, expression levels of genes known to be active in distinct porcine immune cell populations were mapped across single-cell clusters (**Figures 4B,C**). The 36 clusters were deduced to 13 general cell types (**Figure 4D**) as described below.

Monocyte clusters (13, 19, 20, 25, 27) expressed *CSF1R* and genes associated with microbial recognition (*CD14*, *CD163*, *NLRP3*, *TLR4*), reported as highly expressed by porcine monocytes (Auray et al., 2016). DC clusters (30, 32) expressed porcine pan-DC marker *FLT3* and were further classified as conventional DCs (cDCs; cluster 30) by elevated expression of *FCER1A* and MHCII-encoding genes (*SLA-DRB1**, *SLA-DRA**) and pDCs (cluster 32) by elevated expression of *TCF4*, *XBPI*, *CLEC12A*, *CD93*, *IRF8*, *CD4*, and *CD8B* (Auray et al., 2016). Co-stimulatory gene *CD86* was expressed by all monocyte and DC clusters as reported (Auray et al., 2016). *SIRPA**, encoding *CD172a* is expressed by porcine monocytes/DCs (Piriou-Guzylack and Salmon, 2008; Auray et al., 2016) and used to sort myeloid leukocytes for bulkRNA-seq above, was minimally expressed in DC clusters.

B-cell clusters (2, 7, 8, 10, 11, 15, 16, 23, 26, 33) expressed *CD79A*, *CD19*, and *PAX5* (Faldyna et al., 2007; Piriou-Guzylack and Salmon, 2008; Bordet et al., 2019). Antibody-secreting cells (ASCs; cluster 29) expressed *IRF4* and *PRDM*, genes ascribed to immunoglobulin secretion (Shi et al., 2015; Liu et al., 2020). Detection of *CR2**, the gene encoding CD21 protein, was very low in any cluster.

Expression of *CD3E*, which encodes pan-T-cell CD3ε protein, identified T-cell clusters (0, 3, 4, 5, 6, 9, 12, 14, 17, 18, 21, 22, 24, 28, 31) (Gerner et al., 2009). Cluster 1 cells largely lacked *CD3E*, *CD5*, and *CD6* expression, while expressing *CD2*, *CD8A*, *PRF1*, NK receptor-encoding genes *KLRB1* (*CD161*) and *KLRK1* (*NRKG2D*), and NK receptor signaling adaptor molecules *HCST* (*DAP10*) and *TYROBP* (*DAP12*), corresponding to a NK cell designation (Denyer et al., 2006; Piriou-Guzylack and Salmon, 2008; Gerner et al., 2009; Toka et al., 2009). $\gamma\delta$ T-cells were identified by *TRDC* expression, encoding the $\gamma\delta$ TCR δ chain, and were subdivided into two major subtypes based on presence/absence of *CD2* expression (Piriou-Guzylack and Salmon, 2008; Gerner et al., 2009; Stepanova and Sinkora, 2013; Sedlak et al., 2014). Clusters 6 and 21 were identified as $CD2^-$ $\gamma\delta$ T-cells and clusters 24 and 31 as $CD2^+$ $\gamma\delta$ T-cells. Clusters expressing *CD3E* but not *TRDC* were considered $\alpha\beta$ T-cells and were further subdivided based on *CD4* expression (0, 3, 4, 28 classified as $CD4^+$ $\alpha\beta$ T-cells) or *CD8A* and *CD8B* expression (9, 12, 14, 18, 22 classified as $CD8\alpha\beta^+$ $\alpha\beta$ T-cells) (Piriou-Guzylack and Salmon, 2008; Gerner et al., 2009). Clusters 5 and 17 were more difficult to fully classify and likely represented a mixture of cells, with some but not all cells expressing *CD3E*. Cells in clusters 5 and 17 largely lacked expression of *CD5*, *CD6*, *TRDC*, *CD4*, and *CD8B* but did largely express *CD2*, *CD8A*, *KLRB1*, and *KLRK1* and were therefore characterized as a mixture of $CD8\alpha^+$ $\alpha\beta$ T- and NK cells.

Cells in cluster 34 could not be characterized well enough to broadly classify as myeloid, B, T, or NK lineage leukocytes based on the porcine cell markers described and remained unclassified. Cluster 35 expressed *HBM* and *AHSP*, indicating erythrocytes. Clusters 34 and 35 were still included in further scRNA-seq analyses; however, results pertaining to these clusters were not discussed.

Gene Signatures of BulkRNA-Seq Populations Had Limitations in Resolving Single-Cell Identities

Gene set enrichment analyses (GSEA) using SEG lists defined at different levels of enrichment for each sorted bulkRNA-seq population (**Supplementary File 3**, see Materials and Methods) was performed to identify which scRNA-seq clusters were likely represented (**Figures 5A,B**, **Supplementary Figure 10A**, and **Supplementary File 8**). Some gene sets had high relative enrichment in anticipated corresponding scRNA-seq clusters, such as Myeloid gene sets to monocyte/DC clusters, *CD21nB/CD21pB* gene sets to B-cell clusters, and *SWC6gdT* gene sets to $CD2^-$ $\gamma\delta$ T-cell clusters. Interestingly, highest relative enrichment (2.51) for the top 1% of *CD21nB* SEGs was noted for ASCs in cluster 29, followed by erythrocytes in cluster 35 (1.68). Within sorted NK and T-cell populations, some gene sets showed high relative enrichment for their anticipated corresponding clusters in the scRNA-seq dataset. We also noted off-target relative enrichment for gene sets in clusters not anticipated to be included in specific sorted cell populations. Cluster 28 had lower relative enrichment for *CD4T* and *CD4CD8T* SEG lists at top 5–25% SEG levels (−0.02 to 0.73) than did several non- $CD4^+$ $\alpha\beta$ T-cell clusters. Similar phenomena were observed for *CD8T* top 5–25% SEG lists, whereby clusters 1, 24, and 31 had higher relative enrichment for *CD8T* SEG lists (0.69 to 1.56) than did clusters 14 or 18 (−0.04 to 0.95 relative enrichment) that were anticipated to be included in the *CD8T* population. Clusters 24 and/or 31 showed off-target relative enrichment for all T/NK gene sets to various degrees, though these cells would not be expected to make up a sizeable portion of any of those sorted cell populations.

Further comparison of porcine bulk and scRNA-seq data by CIBERSORTx deconvolution analysis largely supported our single-cell cluster designations by predicting which clusters proportionally represented the bulk RNA-seq data (**Supplementary Figure 10B** and **Supplementary File 7**). Several clusters with poor AUCell enrichment for anticipated bulkRNA-seq gene sets in **Figures 5A,B**, such as cluster 28, were predicted to constitute considerable proportions of their anticipated cell populations by CIBERSORTx deconvolution analysis. Additionally, clusters that demonstrated off-target enrichment by AUCell analysis, such as clusters 1, 9, 22, 24, and 31, were not predicted to be largely present in those off-target populations using CIBERSORTx. However, CIBERSORTx failed to predict many single-cell clusters to have notable abundances in any bulkRNA-seq populations, such as clusters 8, 19, 26, 32, and 34 having <3.33% predicted abundance for any one bulkRNA-seq sample.

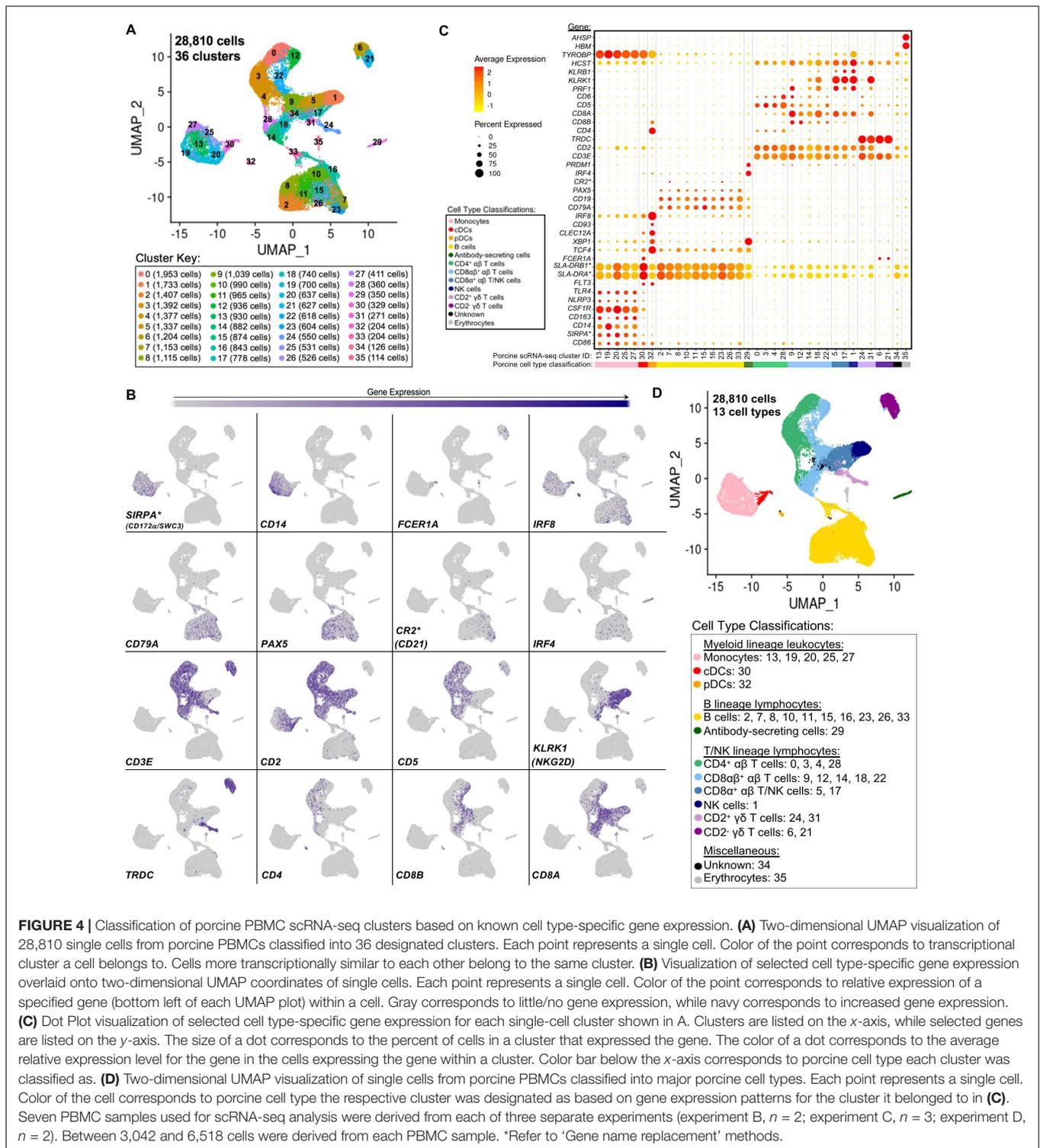


FIGURE 4 | Classification of porcine PBMC scRNA-seq clusters based on known cell type-specific gene expression. **(A)** Two-dimensional UMAP visualization of 28,810 single cells from porcine PBMCs classified into 36 designated clusters. Each point represents a single cell. Color of the point corresponds to transcriptional cluster a cell belongs to. Cells more transcriptionally similar to each other belong to the same cluster. **(B)** Visualization of selected cell type-specific gene expression overlaid onto two-dimensional UMAP coordinates of single cells. Each point represents a single cell. Color of the point corresponds to relative expression of a specified gene (bottom left of each UMAP plot) within a cell. Gray corresponds to little/no gene expression, while navy corresponds to increased gene expression. **(C)** Dot Plot visualization of selected cell type-specific gene expression for each single-cell cluster shown in A. Clusters are listed on the x-axis, while selected genes are listed on the y-axis. The size of a dot corresponds to the percent of cells in a cluster that expressed the gene. The color of a dot corresponds to the average relative expression level for the gene in the cells expressing the gene within a cluster. Color bar below the x-axis corresponds to porcine cell type each cluster was classified as. **(D)** Two-dimensional UMAP visualization of single cells from porcine PBMCs classified into major porcine cell types. Each point represents a single cell. Color of the cell corresponds to porcine cell type the respective cluster was designated as based on gene expression patterns for the cluster it belonged to in **(C)**. Seven PBMC samples used for scRNA-seq analysis were derived from each of three separate experiments (experiment B, $n = 2$; experiment C, $n = 3$; experiment D, $n = 2$). Between 3,042 and 6,518 cells were derived from each PBMC sample. *Refer to 'Gene name replacement' methods.

Additional GSEA comparing gene sets derived from public bulkRNA-seq data of sorted human PBMC populations with porcine single-cell gene expression profiles informed cluster identity as it relates to human immune cells (Figure 5C, Supplementary Figures 10C,D, and Supplementary File 9).

High relative enrichment for human monocyte gene sets in porcine monocyte populations, human CD123PmDC gene sets in porcine cDCs, and human pDC gene sets in porcine pDCs was observed, in general consensus with gene expression profiles of anticipated corresponding porcine single-cell clusters. NaiveB

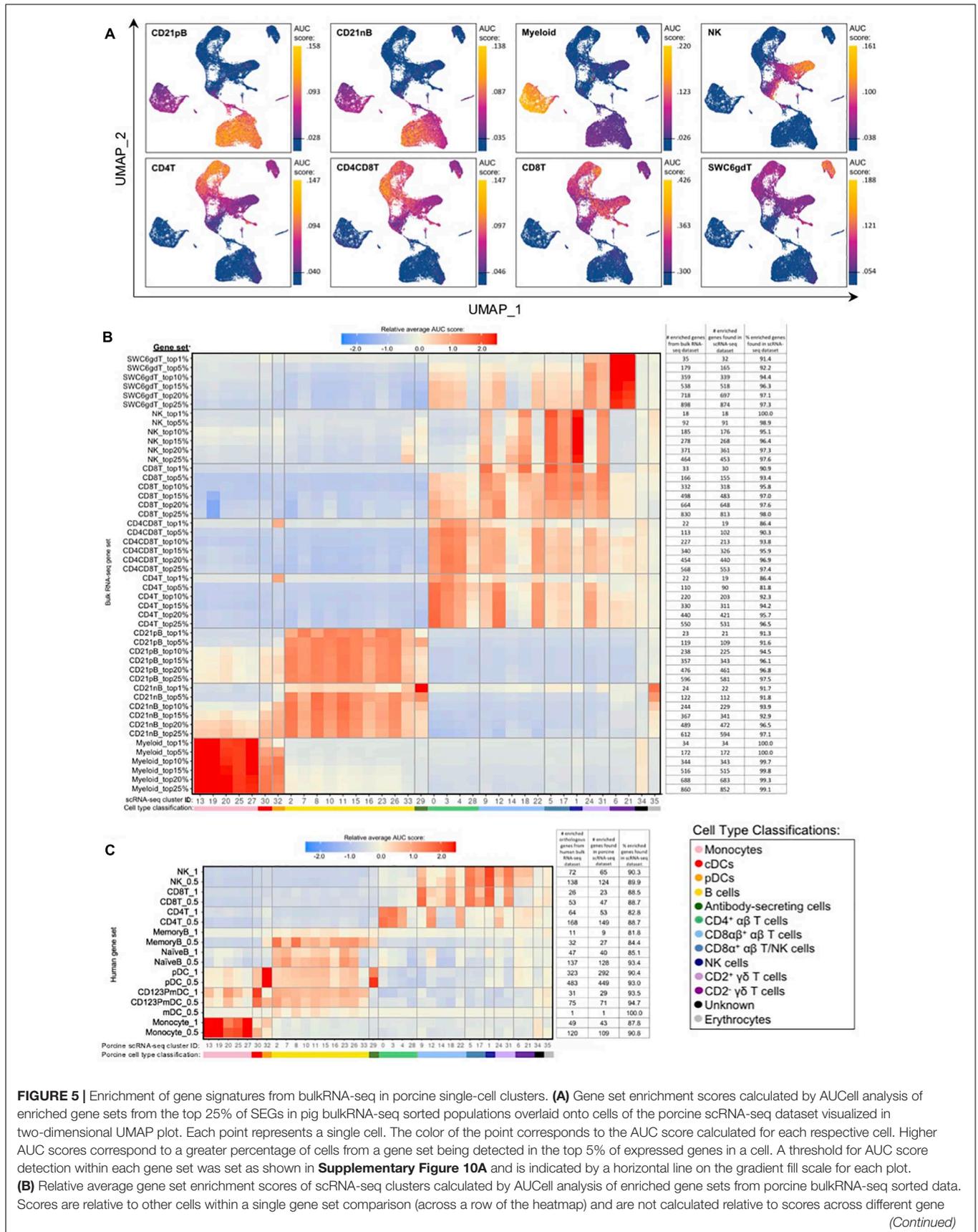


FIGURE 5 | Enrichment of gene signatures from bulkRNA-seq in porcine single-cell clusters. **(A)** Gene set enrichment scores calculated by AUCell analysis of enriched gene sets from the top 25% of SEGs in pig bulkRNA-seq sorted populations overlaid onto cells of the porcine scRNA-seq dataset visualized in two-dimensional UMAP plot. Each point represents a single cell. The color of the point corresponds to the AUC score calculated for each respective cell. Higher AUC scores correspond to a greater percentage of cells from a gene set being detected in the top 5% of expressed genes in a cell. A threshold for AUC score detection within each gene set was set as shown in **Supplementary Figure 10A** and is indicated by a horizontal line on the gradient fill scale for each plot. **(B)** Relative average gene set enrichment scores of scRNA-seq clusters calculated by AUCell analysis of enriched gene sets from porcine bulkRNA-seq sorted data. Scores are relative to other cells within a single gene set comparison (across a row of the heatmap) and are not calculated relative to scores across different gene sets. (Continued)

FIGURE 5 | Continued

sets (across columns in the heatmap). Gene sets were created from the top 1, 5, 10, 15, 20, or 25% of SEGs from sorted populations, as determined by highest log₂FC values in the porcine bulkRNA-seq data. The number of genes included from the bulkRNA-seq dataset and the number and percent of genes detected in the scRNA-seq dataset is listed on the right of the heatmap. A color bar under scRNA-seq cluster IDs indicates the cell type classification, as according to **Figure 4D**.

(C) Relative average gene set enrichment scores of scRNA-seq clusters calculated by AUCell analysis of enriched gene sets from human bulkRNA-seq sorted data. Scores are relative to other cells within a single gene set comparison (across a row of the heatmap) and are not calculated relative to scores across different gene sets (across columns in the heatmap). Gene sets were created from genes with high expression scores >0.5 or >1 for each respective sorted population of cells, with a greater high expression score indicating greater enrichment. The number of genes included from the bulkRNA-seq dataset and the number and percent of genes detected in the scRNA-seq dataset is listed on the right of the heatmap. A color bar under scRNA-seq cluster IDs indicates the cell type classification, as according to **Figure 4D**.

cell gene signatures had positive relative enrichment in all porcine B-cell clusters except cluster 33 at both the 0.5 and 1.0 resolution level, while the MemoryB cell signature had highest relative enrichment scores for B and ASC clusters at the 0.5 level, with little relative enrichment at the 1.0 level (likely due to a limited number of genes in the gene set). Human T/NK gene sets had off-target enrichment very similar to patterns observed in GSEA with porcine gene sets. Overall, GSEA between human bulkRNA-seq gene signatures and gene expression profiles of porcine scRNA-seq data supported many of the same findings when comparing between porcine bulkRNA-seq gene sets and gene expression profiles of porcine scRNA-seq data. Results indicated limitations of gene profiles obtained from sorted bulkRNA-seq populations in accurately describing/accounting for transcriptional heterogeneity resolved by scRNA-seq.

Integration of Porcine and Human scRNA-seq Datasets to Further Annotate Porcine Cells

We examined porcine single-cell identities by comparing the porcine scRNA-seq data to a highly annotated scRNA-seq dataset of human PBMCs, providing a higher level of resolution than available with bulkRNA-seq. Transfer of more highly specified human cell type labels onto porcine cells could reveal the most likely human counterparts for these porcine populations. Mapping scores were further calculated to determine how well porcine cells were truly represented by the human dataset (**Figure 6A**, **Supplementary Figures 11A,B**, and **Supplementary File 10**).

Many porcine clusters had >95% of cells mapping to a specific human cell type, with average mapping scores >0.9, including monocyte, pDC, cDC, and ASC clusters, suggesting high congruency between pig and human for these cell types (**Figure 6B**). All porcine B-cell clusters, omitting cluster 33, mapped primarily to human B-cell clusters, but average mapping scores were slightly lower (0.80–0.87), indicating less ideal representation in the human data. In addition, every porcine B-cell cluster had overlap with all three human B-cell types (**Figure 6A**). Of the porcine CD4⁺ αβ T-cells, most cluster 0 cells were predicted as human CD4 naïve cells, clusters 3 and 4 cells as human CD4 T central memory (TCM) cells, and cluster 28 cells as human CD4 proliferating cells. From porcine CD8αβ⁺ αβ T-cells, clusters 14 and 18 were largely assigned as human replicating cell types, while 90% of cluster 9 cells were predicted as human CD8 T effector memory (TEM) cells. Highest cluster

12 predictions were mainly to human CD4/CD8 naïve T-cells, and cluster 22 cells predicted to match a range of human cell populations, with the largest percentage predicted as human CD8 TEMs. Porcine CD8α⁺ αβ T/NK and NK clusters had predictions split primarily across human CD8 TEM and NK designations. Porcine CD2⁺ γδ T-cell clusters 24 and 31 had 74 and 98%, respectively, of cells predicted as human CD8 TEM, NK, or γδ T-cells. Porcine CD2⁻ γδ T-cell clusters 6 and 21 had the majority of cells predicted as human CD4 TCM, innate lymphoid cell (ILC), or γδ T-cells, though the average mapping scores were lower for those assigned as CD4 TCM (0.73–0.74) or gdT (0.74–0.78) than those assigned as ILCs (0.82–0.83) (**Supplementary File 10**). Overall, cross-species comparison to a well-annotated human scRNA-seq dataset helped elucidate porcine cell type identities at a higher resolution than porcine or human bulkRNA-seq datasets (**Figure 5**), though some discordance was clearly still present.

Several porcine clusters had low mapping scores to a human cell type, indicating the porcine cells may not be well represented by the human reference dataset (**Figure 6B** and **Supplementary File 10**). Therefore, *de novo* visualization was performed on the combined human and porcine data, to identify cells in the pig dataset not well represented in the human data (**Figures 6C,D**). Porcine clusters could be identified that had low similarity to human cells, and vice versa (**Figure 6C**). Specifically, porcine clusters 6, 16, 21, and 33 weakly overlapped human cells in the two-dimensional *de novo* visualization (compare **6C** and **6D**) and had lower average mapping scores to any human cell type (**Figure 6B**). Further inspection revealed clusters 6 and 21 to be CD2⁻ γδ T-cells (identified in **Figures 4B,D**) and their limited representation in human dataset is discussed further below. In contrast, clusters 16 and 33 were B-cells, and to further understand their limited representation compared to other porcine B-cells, clusters 16 and 33 were compared by pairwise comparisons to all remaining B-cell clusters (2, 7, 8, 10, 11, 15, 23, 26; **Supplementary File 6**). Pairwise comparisons revealed significantly increased expression of 33 genes in cluster 16 and 282 genes in cluster 33 relative to every other B-cell cluster (**Table 4**). Compared to other porcine B-cell clusters, cluster 16 had significantly greater expression of several genes associated with B-cell activation (such as *BHL*, *ITGB7*, *JCHAIN*, *ZBTB38*) (Castro and Flajnik, 2014; Kreslavsky et al., 2017; Delecluse et al., 2019; Wong and Bhattacharya, 2020), while many genes with significantly increased expression in cluster 33 were associated with cellular replication and/or division, such as *HIST1H2AB*,

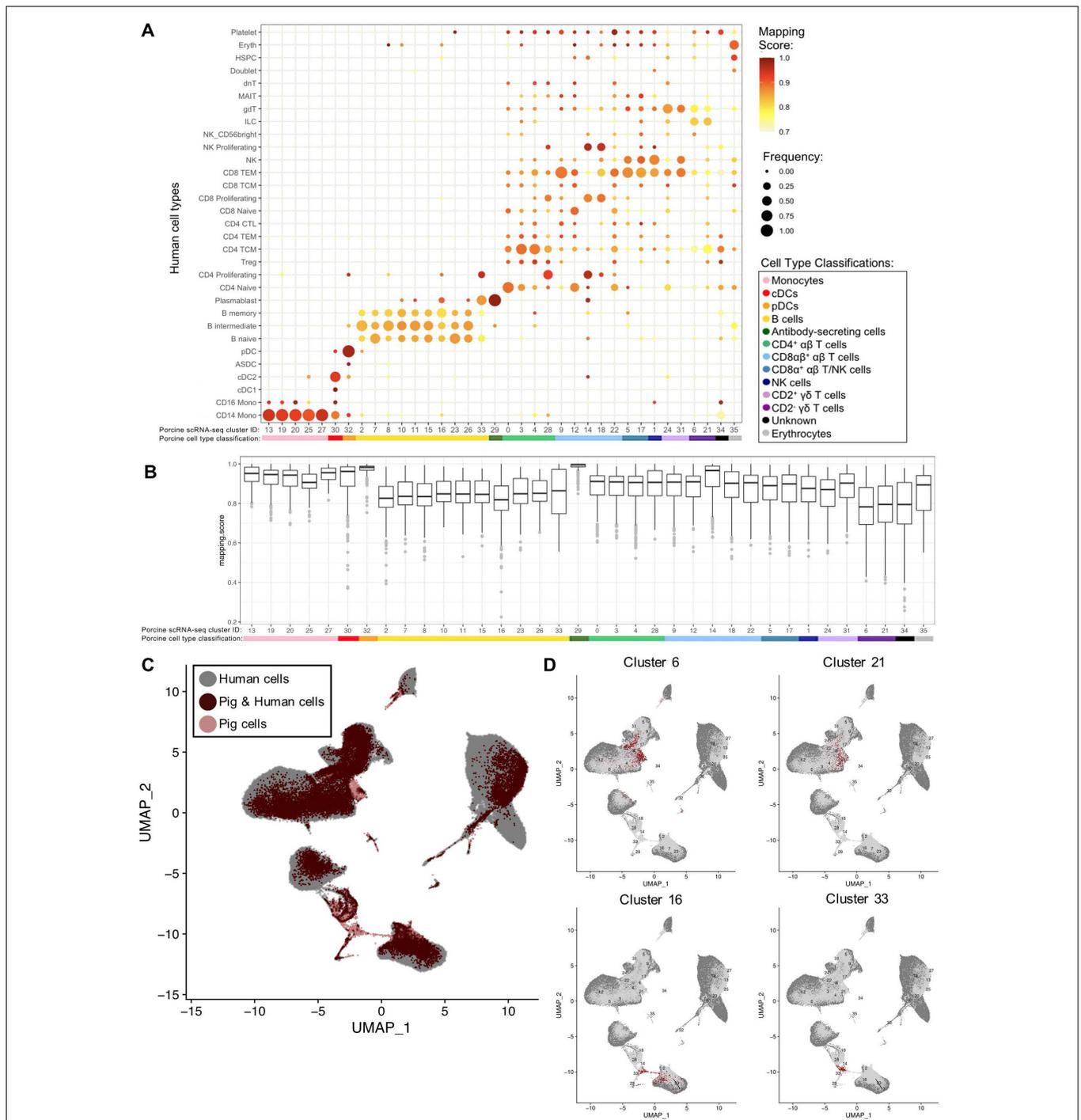


FIGURE 6 | Integration of porcine and human scRNA-seq datasets to further annotate porcine cells. **(A)** Mapping scores calculated to determine how well porcine cells were represented by the human dataset. The human cell type specific frequency (size of the circle) and mapping score for that human cell type (color) are shown for each porcine scRNA-seq cluster. Porcine cell type classifications (color) are shown below the porcine scRNA-seq cluster IDs. **(B)** Mapping scores calculated to determine how well porcine cells were represented by the human dataset. The mapping scores for each porcine scRNA-seq cluster is represented by a box and whiskers plot. Porcine cell type classifications (color) are shown below the porcine scRNA-seq cluster IDs. **(C)** To identify cells in the porcine dataset that were not well represented in the human dataset, a *de novo* visualization of the merged porcine and human data was performed. The porcine (pink) and human (gray) were plotted together using UMAP. An overlap of both porcine and human cells is shown as (dark red). Clusters of porcine cells that are not well represented in the human data can be observed by pink regions in the plot. **(D)** Two primary regions of porcine cells that were not well represented in the human data were identified in **(C)**. In order to clarify which porcine scRNA-seq clusters were represented in these regions, the porcine cluster IDs were projected onto the UMAP and cells from four clusters overlapping the identified regions were colored as dark red.

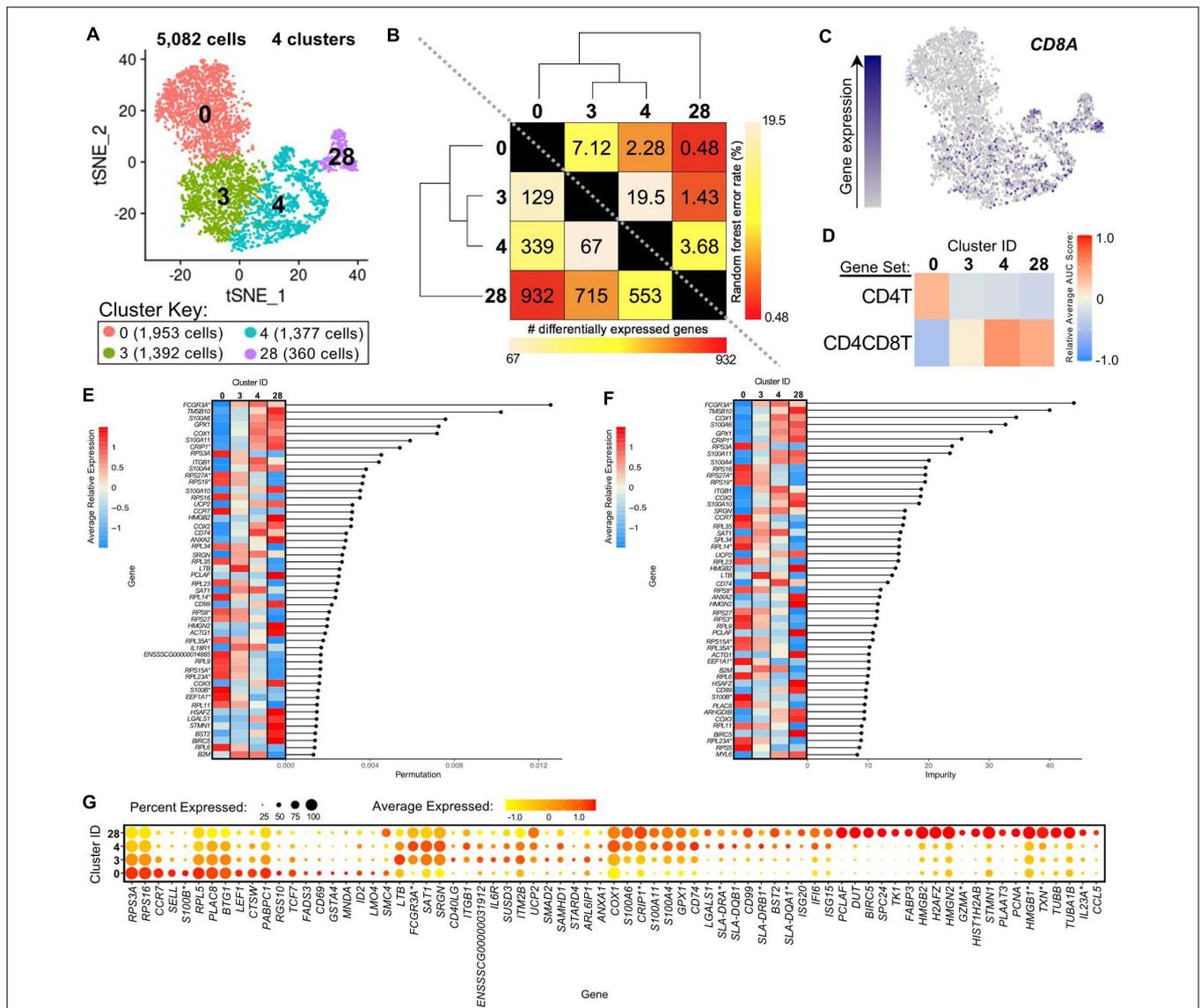


FIGURE 7 | Transcriptional heterogeneity of porcine CD4⁺ αβ T-cells at single-cell resolution. **(A)** Two-dimensional t-SNE plot of 5,082 cells belonging to clusters designated as CD4⁺ αβ T-cells (clusters 0, 3, 4, and 28) in **Figure 4D**. Each point represents a single cell. Color of the cell corresponds to transcriptional cluster a cell belongs to. Cells more transcriptionally similar to each other belong to the same cluster. **(B)** Transcriptomic relationship amongst CD4⁺ αβ T-cell clusters as calculated by three methods: hierarchical clustering (as seen by hierarchical trees on both axes), pairwise random forest analyses (as seen on top right diagonal); and pairwise DGE analyses (as seen on bottom left diagonal). Longer branches on the hierarchical tree corresponds to greater hierarchical distance. Lower numbers of DEGs by DGE analysis and higher out-of-bag (OOB) error rates from random forest analyses indicate greater pairwise transcriptional similarity. **(C)** Visualization of *CD8A* expression overlaid onto t-SNE coordinates of single CD4⁺ αβ T-cells. Each point represents a single cell. Color of the point corresponds to relative expression of *CD8A* within a cell. Gray corresponds to little/no gene expression, while navy corresponds to increased gene expression. **(D)** Relative average gene set enrichment scores of CD4⁺ αβ T-cell clusters calculated by AUCell analysis of DEG sets from pairwise DGE analysis of the CD4T and CD4CD8T populations from porcine bulkRNA-seq. Scores are relative to other cells within a single gene set comparison (across a row of the heatmap) and are not calculated relative to scores across gene set (across columns in the heatmap). **(E,F)** Genes with the largest effects in discriminating CD4⁺ αβ T-cells by cluster identities were determined, as indicated by high permutation **(E)** and/or impurity scores **(F)** calculated from a trained random forest model. Average relative expression for each of these genes within clusters is also depicted by a heatmap. **(G)** Dot plot of up to the top 20 DEGs having logFC > 0 from overall DGE analysis of only CD4 + ab T-cell clusters. Clusters are listed on the y-axis, while selected DEGs are listed on the x-axis. The size of a dot corresponds to the percent of cells in a cluster that expressed the gene. The color of a dot corresponds to the average relative expression level for the gene in the cells expressing the gene within a cluster. *Refer to ‘Gene name replacement’ methods.

activation, such as *ITGB1*, *CD40LG*, *IL6R*, and MHC II-associated genes (*CD74*, *SLA-DRA*, *SLA-DQB1*, *SLA-DRB1**, *SLA-DQAI**) (Grewal and Flavell, 1996; Gerner et al., 2009; Zemmour et al., 2018; Zhu et al., 2020) had significantly greater

expression in clusters 3, 4, and/or 28, and cluster 28 expressed many genes specific for cellular replication and division (*PCLAF*, *BIRC5*, *TK1*, *PCNA*) (Dabydeen et al., 2019; Giotti et al., 2019; **Figure 7G** and **Supplementary Files 9, 13**). Overall, we leveraged

single-cell gene expression profiles to confirm likely identity of cluster 0 as naïve $CD4^+CD8\alpha^-$ $\alpha\beta$ T-cells and clusters 3, 4, and 28 as potentially previously activated $CD4^+CD8\alpha^+$ $\alpha\beta$ T-cells.

Heterogeneity Between/Amongst $CD2^+$ and $CD2^-$ $\gamma\delta$ T-Cells

Clusters predicted to be porcine $\gamma\delta$ T-cells were examined to reveal transcriptional distinctions within this cell type. Four clusters containing 2,652 cells were previously identified as $CD2^-$ $\gamma\delta$ T-cells (clusters 6, 21) or $CD2^+$ $\gamma\delta$ T-cells (clusters 24, 31) (**Figure 8A**). We could further segregate these clusters by $CD2$ and $CD8A$ expression into $CD2^-CD8\alpha^-$ (clusters 6, 21), $CD2^+CD8\alpha^-$ (cluster 24), and $CD2^+CD8\alpha^+$ (cluster 31) designations used to functionally define porcine $\gamma\delta$ T-cells previously (Stepanova and Sinkora, 2013; Sedlak et al., 2014; **Figure 8B**).

$CD2^-$ $\gamma\delta$ T-cell clusters 6 and 21 were most closely related to one another by hierarchical clustering, had the fewest pairwise DEGs (30), and had the highest pairwise RF analysis error rate (23.5), indicating clusters 6 and 21 to be the most transcriptionally similar $\gamma\delta$ T-cell clusters of the four clusters (**Figure 8C** and **Supplementary Files 7, 14**). $CD2^+$ $\gamma\delta$ T-cell clusters 24 and 31 were most similar to each other by hierarchical clustering, had the second fewest pairwise DEGs (236), and had the second highest pairwise RF error rate (5.12), indicating clusters 24 and 31 to be most similar to each other. When performing pairwise comparison between any $CD2^-$ and $CD2^+$ $\gamma\delta$ T-cell clusters, the number of DEGs increased and RF error rates decreased, indicating greater transcriptional differences between cells of the $CD2^-$ and $CD2^+$ $\gamma\delta$ T-cell lineages than amongst them (**Figure 8C** and **Supplementary Files 7, 14**).

The top genes contributing to overall transcriptional heterogeneity amongst $\gamma\delta$ T-cell clusters, as determined by RF analysis (**Figures 8D,E** and **Supplementary File 14**), overlapped with genes identified with significant and highest logFC expression in overall DGE analysis (**Figure 8F** and **Supplementary File 14**). Six of the top seven genes with mutual highest impurity (the best features that correctly split the data) and permutation scores from RF analysis (**Figures 8D,E**) were also DEGs between both $CD2^-$ compared to both $CD2^+$ $\gamma\delta$ T-cell clusters by pairwise DGE analysis (**Supplementary File 7**), again indicating large transcriptional differences between $CD2^-$ and $CD2^+$ $\gamma\delta$ T-cells. In total, 31 genes had significantly greater expression in both $CD2^-$ $\gamma\delta$ T-cell clusters compared to both $CD2^+$ $\gamma\delta$ T-cell clusters, and 49 genes had significantly greater expression in both $CD2^+$ $\gamma\delta$ T-cell clusters compared to both $CD2^-$ $\gamma\delta$ T-cell clusters (**Table 5**), as determined using the pairwise DGE analyses (**Supplementary File 7**).

Intra-lineage heterogeneity of $CD2^-$ $\gamma\delta$ T-cells (between clusters 6 and 21) and $CD2^+$ $\gamma\delta$ T-cells (between clusters 24 and 31) demonstrated additional complexity beyond the inter-lineage heterogeneity between $CD2^-$ and $CD2^+$ $\gamma\delta$ T-cells. Pairwise comparison between clusters 24 and 31 (**Supplementary Data 8**) revealed 80 genes with significantly greater expression in cluster 24 ($CD2^+CD8\alpha^-$ $\gamma\delta$ T-cells) and 156 genes with significantly greater expression in cluster 31 ($CD2^+CD8\alpha^+$ $\gamma\delta$ T-cells). Genes with the greatest logFC expression (logFC > 1.5)

in cluster 31 compared to cluster 24 were related to cellular activation and/or effector functions (*CCL5*, *GZMA*, *FCGR3A*, *KLRK1*, *GZMA*, *NKG7*, *FCER1G*, *GZMB*) (Rincon-Orozco et al., 2005; Pizzolato et al., 2019; Szabo et al., 2019). Of the 30 DEGs between clusters 6 and 21 (**Supplementary Data 8**), three genes had significantly greater expression in cluster 6, while 27 genes had significantly greater expression in cluster 21. Several genes with greater expression in cluster 21 encoded for activation- or stress-induced molecules, including *GPX1*, *LGALS1*, *ITGB1*, *LTB*, several genes encoding for S100 proteins (*S100A4*, *S100A6*, *S100A10*, *S100A11*), and genes related to MHCII presentation (*CD74*, *SLA-DRA*) (Blaser et al., 1998; Ware, 2005; Gerner et al., 2009; Steiner et al., 2011; Kesarwani et al., 2013; Siegers, 2018). Genes encoding transcriptional regulators playing important roles in cell fate determination, including *ID3* and *GATA3*, had greater expression in cluster 6, while *ID2* expression was significantly greater in cluster 21 (Blom et al., 1999; Zhang et al., 2014; Rodríguez-Gómez et al., 2019).⁷

DISCUSSION

We present the first comprehensive annotation of the global transcriptome of all major circulating porcine blood mononuclear cells. We applied bulkRNA-seq to determine transcriptomes of eight sorted PBMC populations and scRNA-seq to annotate transcriptomic diversity of PBMCS into transcriptionally distinct clusters. Deep RNA sequencing detected significant heterogeneity between sorted populations except for T-cell populations, while further heterogeneity was unmasked by scRNA-seq. Collectively, the data sets revealed specific immune functional expression patterns and highlighted substantial diversity in some subsets, such as T-cells. The combined approach helps to unite porcine transcriptomics and cellular immunology, as transcriptional differences and functional relationships of porcine immune cells have remained unclear due to lack of sufficient reagents to label distinct porcine immune cell populations. While cross-species comparisons have been done with many RNA-seq datasets of partially purified cell populations (Kapetanovic et al., 2013; Herrera-Uribe et al., 2020), our new porcine data demonstrates global similarity to human bulkRNA-seq and scRNA-seq transcriptomes that can be used to further unravel porcine cell function and extend comparative immune investigation.

Gene expression patterns from the bulkRNA-seq datasets revealed distinct transcript profiles enriched in biological pathways characteristic of each respective cell population, based on previous findings in pig and other species (Alter et al., 2004; Palmer et al., 2006; Wang et al., 2008; Foissac et al., 2019; Monaco et al., 2019; Summers et al., 2020). However, bulkRNA-seq data from the porcine sorted populations had limited ability to identify genes with specific transcriptional patterns for some sorted lymphocyte populations. The transcriptomes of eight different cell types we provide include three types of transcriptomes that have not reported before in pig, including NK, CD21pB and CD21nB. Lists of SEGs, pairwise DGE between all populations

⁷<http://biocc.hrbmu.edu.cn/CellMarker/#>

TABLE 5 | Genes differentially expressed between both CD2⁻ $\gamma\delta$ T-cell clusters (clusters 6 and 21) and both CD2⁺ $\gamma\delta$ T-cell clusters (clusters 24 and 31).

Population with greater gene expression	Genes
CD2 ⁻ $\gamma\delta$ T-cells (clusters 6, 21)	<i>AP3S1*</i> , <i>ANXA1</i> , <i>BLK</i> , <i>CAPG</i> , <i>CDNF*</i> , <i>CD163L1*</i> , <i>EMP3</i> , <i>ENSSSCG00000032017</i> , <i>ENSSSCG00000033734</i> , <i>FCER1A</i> , <i>GATA3</i> , <i>IL6R</i> , <i>ITM2B</i> , <i>LGALS1</i> , <i>LTB</i> , <i>MAN2B1*</i> , <i>MYL12A*</i> , <i>PARK7</i> , <i>PIK3AP1</i> , <i>PLEKHF2</i> , <i>PPP1CC</i> , <i>RCAN3</i> , <i>RHEX</i> , <i>RPS19*</i> , <i>SAMSN1</i> , <i>SELL</i> , <i>SLC25A24</i> , <i>SRGN</i> , <i>TIMP1</i> , <i>VIM</i> , <i>YBX3</i>
CD2 ⁺ $\gamma\delta$ T-cells (clusters 24, 31)	<i>ABI3*</i> , <i>ACTG1</i> , <i>ARPC1B</i> , <i>ARPC5L</i> , <i>BIN1</i> , <i>CAMK4</i> , <i>CCDC12*</i> , <i>CD2</i> , <i>COTL1</i> , <i>CTSD</i> , <i>DYNLRB1</i> , <i>ENSSSCG00000023584</i> , <i>ENSSSCG00000027196</i> , <i>ENSSSCG00000029596</i> , <i>ENSSSCG00000038825</i> , <i>FAM49B</i> , <i>FSCN1</i> , <i>FYB1</i> , <i>GBP7*</i> , <i>GIMAP4*</i> , <i>H2AFV</i> , <i>IFITM1*</i> , <i>IFI6</i> , <i>IKZF2</i> , <i>IKZF3</i> , <i>IL2RB</i> , <i>ISG15</i> , <i>ITGA4</i> , <i>ITGB2</i> , <i>ITM2C</i> , <i>KRAS</i> , <i>LCK</i> , <i>MAGOHB*</i> , <i>NT5C3A*</i> , <i>PIK3R1</i> , <i>PRKCH*</i> , <i>PSIP1</i> , <i>PTPRC</i> , <i>RESF1</i> , <i>S100A1</i> , <i>SLC9A3R1</i> , <i>SMC4</i> , <i>SNRK</i> , <i>STK17B</i> , <i>STMN3</i> , <i>TRAT1</i> , <i>UBAC2</i> , <i>WCR1</i> , <i>WIPF1*</i>

*Refer to gene name replacement in Materials and Methods section.

identification of cell type-specific genes was likely caused by the lack of further sub-setting during sorting to separate functionally distinct cells. However, we were able to find several specific transcriptional patterns in B- and T-cells using bulkRNA-seq, and some of the identified genes encode for transmembrane proteins. Beyond further description of well-annotated genes, we also demonstrated that up to 18% of our predicted cell-type specific and enriched genes are currently poorly annotated, i.e., genes with no recognized human ortholog. These data thus increase the functional annotation of these genes, as co-expression patterns linking such genes with known genes can be an important component for Gene Ontology classifications and disease-association gene prediction (van Dam et al., 2018), and is an important proposed outcome of the FAANG project (Giuffra et al., 2019).

Comparison of our sorted population expression patterns to a similar human RNA-seq dataset revealed both similarities and differences between species. While we compared the transcriptomes of the sorted cells with human populations that were isolated using similar cell markers, we cannot exclude that we are biasing this comparison due to different immunoreagent markers used across species. However, we did find similar transcriptional patterns across immune cell populations that are intrinsic to a lineage, such as the porcine Myeloid population correlating with the human myDC123 population, in agreement with other studies (Auray et al., 2016).

Previous global gene expression studies using either porcine whole blood or specific immune cell types have failed to thoroughly describe all major PBMC populations (Freeman et al., 2012; Dawson et al., 2013; Mach et al., 2013; Auray et al., 2016; Foissac et al., 2019). Providing the transcriptomes of bulk sorted cell populations will be readily useful to the majority of porcine immunology research labs that use sorting techniques to analyze porcine immune cell function and RNA expression patterns, as new lists of co-expressed genes in these cell populations are now available. However, our combined analysis of such bulkRNAseq data with the scRNAseq data demonstrated that the former approach has significant heterogeneity, limiting the ability to resolve specific cell types for deeper transcriptional interrogation. A combined analysis provided evidence confirming our hypothesis that scRNA-seq would lead to identification of more specific and novel transcriptional signatures to improve annotation and understanding of circulating porcine immune cells.

Single-cell RNA-sequencing provides many noted benefits in transcriptomic analysis, however there are limitations to the

approach. Of benefit, scRNA-seq captured transcriptomes of cells excluded from our bulkRNA-seq analysis, as scRNA-seq approach did not rely on protein marker expression and selection of sorting criteria based on specific marker phenotypes. As mentioned above, scRNA-seq also established that greater levels of cellular heterogeneity exist, since sequencing was resolved to the level of individual cells rather than a sorted population. We recognize the scRNAseq-predicted clusters may contain transitory cell states that may be very challenging to further study for the relationship between cellular function and transcriptional patterns (Bassler et al., 2019). Further, we assumed single-cell gene expression profiles would be indicative of protein expression for cell type-specific markers; however, gene expression for many such markers, including *SIRPA** and *CR2** that encode proteins used for bulk RNA-seq cell sorting, was sparse. Sparsity of data is a known limitation of the scRNA-seq approach utilized herein, while methods such as imputation have been proposed to improve sensitivity (Andrews et al., 2021). We chose not to use imputation due to our current inability to estimate effects on cell patterns through comparison to an external reference (Andrews et al., 2021). Thus, these limitations made it difficult to decipher between low- and non-expression for some genes of interest, including canonical markers used for identifying cell types in the immunology literature. Instead, reliance on gene expression profiles of multiple markers was used. For example, *SIRPA** expression was observed at low levels in monocyte clusters but was virtually absent in DC clusters, though both porcine monocytes and DCs express CD172 α protein. Because DCs express CD172 α at lower levels than monocytes (Piriou-Guzylack and Salmon, 2008; Auray et al., 2016), *SIRPA** expression in DCs may have been below our limit of detection using scRNA-seq, as it was insufficiently expressed in DCs but not in monocytes. We utilized a droplet-based partitioning method for scRNA-seq that can detect a large number of cells but a lower number of transcripts per cell. By this method, we could retain a large number of cells (>25,000 cells from seven samples) at the expense of limited sequencing depth per cell (minimum of 500 unique genes and 1,000 unique transcripts per cell). Utilizing higher sequencing depth per cell or different partitioning platforms for scRNA-seq that have more efficient transcript capture per cell will be beneficial for deeper analysis of specific cells/genes of interest. It is likely some gene expression profiles are not predictive of protein expression, due to post-transcriptional regulation mechanisms. Using newly available co-expression lists to formulate more refined cell sorting regimens and scRNAseq analysis of such sorted populations will also increase the ability

to define transcriptomes of such cell types (Nestorowa et al., 2016). It was notable that the lists of genes predicted to be significantly enriched in the 36 scRNA-seq clusters had overall a very similar fraction of poorly annotated genes (average of 18%; cite in **Supplementary File 6**) to those predicted for bulkRNA-seq, indicating that even the genes with expression patterns predicted to be more discriminatory contribute a similar level of genome annotation improvement.

We used multiple methods to compare these high-dimensional expression datasets to further interpret genes predicted to be different between sorted cell populations, between clusters, or between human and pig. GSEA and/or deconvolution analyses of bulkRNA-seq to scRNA-seq datasets was only partially effective in correlating sorted populations with assumed corresponding clusters in the scRNA-seq dataset (regardless of inter-species or intra-species comparison). At a higher level of resolution, both methods were able to assign most corresponding cell-type designations between scRNA-seq and bulkRNA-seq data. However, several different scRNA-seq clusters were not predicted to make up a large portion of any bulkRNA-seq sample. While methodology could account for these differences, it is more likely that CIBERSORTx was unable to discriminate between certain clusters due to their high similarity. For example, cells that could have been predicted to be assigned to cluster 8, which makes up a large proportion of the scRNA-seq data, may have been assigned to other similar B-cell clusters. The ability to discriminate between similar clusters may have been impacted by down sampling each cluster to include the same number of cells for the analysis. Overall, deconvolution was useful in assigning cell type level data but in some instances, it could not fully deconvolute bulk RNA-seq to the cluster specific level.

Integration of porcine PBMC scRNA-seq with a human PBMC scRNA-seq dataset did allow further resolution of porcine cluster annotations and yielded high confidence of homology between many porcine and human single cell populations. While we cannot completely discount the potential for recognized cell types in our scRNA-seq dataset not being present in sorted populations used for bulkRNA-seq (or vice-versa), it seems more likely this is similar evidence to that described above indicating that the same level of resolution simply was not captured by bulkRNA-seq and could not well represent all cell types found in the scRNA-seq data. Integration with another scRNA-seq dataset, even when accounting for cross-species comparison, was in many ways more informative for further annotating porcine single cells, highlighting the enhanced ability of scRNA-seq to define cellular landscapes. Moreover, cross-species integration extended our knowledge of comparative immunology between humans and pigs, as we could identify most similar human counterparts by reference-based prediction. Conversely, we also identified clusters of CD2⁻ $\gamma\delta$ T-cells (clusters 6 and 21) and B-cells enriched for activation or cycling-specific genes (clusters 16 and 33) that were more prevalent in porcine data by *de novo* visualization of single cells using the combined human and porcine scRNA-seq data. CD2⁻ $\gamma\delta$ T-cells are frequent in porcine circulation but are reported absent in humans and mice (Stepanova and Sinkora, 2013), and our analyses supported the presence in pigs but not humans. On the other hand, B-cells with

transcriptional profiles characteristic of activated or cycling cells, similar to porcine clusters 16 and 33, likely still occur in humans, albeit with low prevalence in circulation. B-cell ontogeny and activation are less fully understood in pigs than in humans, and it's possible peripheral B-cells in clusters 16 and 33 arise from a developmental, activation, or circulation process specific to pigs. In pigs, the majority of leukocytes exit lymph nodes through the vasculature and directly re-enter the blood rather than efferent lymph, as observed in humans (Sasaki et al., 1994; Saalmüller and Gerner, 2016). Thus, it's possible the different patterns of egress for activated cells leaving sites of immune induction might contribute to a higher frequency of activated B-cells entering circulation in pigs compared to humans.

While we did not perform deeper biological query of all cell types identified in our scRNA-seq dataset, we did attempt to deduce biological significance for the different CD4⁺ $\alpha\beta$ T-cell populations that have unique aspects in pigs. Deeper query of CD4⁺ $\alpha\beta$ T-cells was performed, as there is functional interest in determining activation states of porcine CD4⁺ $\alpha\beta$ T-cells based on CD8 α expression, which may be gained upon activation and retained in a memory state (Summerfield et al., 1996; Zuckermann, 1999; Saalmüller et al., 2002; Gerner et al., 2009). We found it difficult to identify CD4⁺ $\alpha\beta$ T-cell clusters as CD8 α ⁺ or CD8 α ⁻ due to sparsity in *CD8A* expression but could leverage comparison of CD4T and CD4CD8T populations from bulkRNA-seq to formulate gene sets enriched in each CD4 expressing T-cell population. GSEA helped identify one cluster of CD4⁺CD8 α ⁻ $\alpha\beta$ T-cells that corresponded mostly to human naïve CD4 T-cells, while three clusters of CD4⁺CD8 α ⁺ $\alpha\beta$ T-cells corresponded to human memory or proliferating CD4 T-cells. Collectively, these data reinforce previous porcine literature, elucidate parallels to human cells, and provide greater insight into the spectrum of activation states present in CD4⁺CD8 α ⁺ $\alpha\beta$ T-cells. Future analysis of activated T-cells or trajectory analysis may provide even further insight on the transition of activation states in porcine peripheral T-cells.

Pigs are a ' $\gamma\delta$ high' species, named as such because they have a higher proportions of $\gamma\delta$ T-cells in circulation, largely attributed to the presence of CD2⁻ $\gamma\delta$ T-cells that are absent in humans and mice (Stepanova and Sinkora, 2013). Three major $\gamma\delta$ T-cell populations are characterized in pigs: CD2⁻CD8 α ⁻ $\gamma\delta$ T-cells that express SWC6 and CD2⁺CD8 α ^{-/+} $\gamma\delta$ T-cells that do not express SWC6, where CD2⁻CD8 α ⁻ $\gamma\delta$ T-cells become CD2⁺CD8 α ⁺ upon activation (Stepanova and Sinkora, 2013; Sedlak et al., 2014). As our sorting strategy for bulkRNA-seq utilized an anti-SWC6 antibody rather than a pan- $\gamma\delta$ T-cell-specific antibody; thus, $\gamma\delta$ T-cells for bulk RNA-seq included CD2⁻CD8 α ⁻ $\gamma\delta$ T-cells in the SWC6gdT population or CD2⁺CD8 α ⁺ $\gamma\delta$ T-cells found in combination with CD4⁻CD8 α ⁺ $\alpha\beta$ T-cells in the CD8T population. CD2⁺CD8 α ⁻ $\gamma\delta$ T-cells were expected to be excluded in cell sorting. In future sorting strategies, it may be beneficial to utilize a pan- $\gamma\delta$ T-cell reactive antibody and/or identify CD4⁻CD8 α ⁺ $\alpha\beta$ T-cells with anti-CD8 β antibody, which should not label with CD2⁺CD8 α ⁺ $\gamma\delta$ T-cells (Gerner et al., 2009). though this may still exclude potential CD4⁻CD8 α ⁺CD8 β ⁻ $\alpha\beta$ T-cells, such as we observed in clusters 5 and 17. Despite limitations in sorting, the bulkRNA-seq profiles were still informative when comparing to

scRNA-seq data. The highest relative enrichment of SWC6gdT gene signatures was detected in CD2⁻ $\gamma\delta$ T-cell clusters, while CD2⁺ $\gamma\delta$ T-cell clusters showed relative enrichment to a lesser level, indicating some conserved gene expression between CD2⁻ and CD2⁺ $\gamma\delta$ T-cells. Comparison between CD2⁺ $\gamma\delta$ T-cell clusters further supported previous biological understanding, where CD2⁺CD8 α ⁺ $\gamma\delta$ T-cells had greater expression of genes related to cellular activation and cytotoxicity relative to CD2⁺CD8 α ⁻ $\gamma\delta$ T-cells (Yang and Parkhouse, 1997; Stepanova and Sinkora, 2013; Sedlak et al., 2014). On the other hand, CD2⁻ $\gamma\delta$ T-cells are less well described than CD2⁺ $\gamma\delta$ T-cells, largely due to lack of comparable populations in humans or mice that may be used for biological inference. Integration with human scRNA-seq data supported previous observations of the absence of CD2⁻ $\gamma\delta$ T-cells in humans, as close counterparts for CD2⁻ $\gamma\delta$ T-cell clusters could not be found by *de novo* visualization, and reference-based integration indicated closest human counterparts to be a mixture of primarily $\gamma\delta$ T-cells, ILCs, and CD4 T_{CMs}, and mapping scores were highest for human ILCs rather than $\gamma\delta$ T-cells, indicating human ILCs to be the closest, albeit still poor, human match. Nonetheless, we were able to highlight transcriptional distinctions that better annotate CD2⁻ $\gamma\delta$ T-cells, including DEGs between CD2⁻ and CD2⁺ $\gamma\delta$ T-cells that defined the two $\gamma\delta$ T-cell lineages and between two clusters of CD2⁻ $\gamma\delta$ T-cells that have not yet been described.

CONCLUSION

This study provides a first-generation atlas annotating circulating porcine immune cell transcriptomes at both the cell surface marker-sorted population and single-cell levels. These findings illuminate the landscape of immune cell molecular signatures useful for porcine immunology and a deeper annotation of the genome, a goal of the FAANG project. These results also provide useful resources to identify new porcine cell biomarkers for discrimination and isolation of specific cell types, urgently needed in the field.

DATA AVAILABILITY STATEMENT

Raw sequencing data from bulkRNA-seq and scRNA-seq are available through the European Nucleotide Archive (project: PRJEB43826), <https://www.ebi.ac.uk/ena/browser/view/PRJEB43826>. Processed and finalized data for single-cell datasets are available at <https://data.nal.usda.gov/dataset/data-reference-transcriptomics-porcine-peripheral-immune-cells-created-through-bulk-and-single-cell-rna-sequencing>. Scripts used for data analyses are available at https://github.com/USDA-FSEPRU/PorcinePBMCs_bulkRNAseq_scRNAseq.

ETHICS STATEMENT

The animal study was reviewed and approved by NADC Animal Care and Use Committee.

AUTHOR CONTRIBUTIONS

CL and CT conceptualized and supervised research. JH-U, JW, KB, and CL collected samples and isolated cells. KB performed cell staining and FACS. JH-U performed bulk RNA isolations. TS performed bulkRNA-seq library preparation and sequencing. JL supervised the NanoString assay data collection. JH-U, LD, and HL performed bulkRNA-seq analyses. HL performed NanoString analyses. JW, LD, SS, and HL performed scRNA-seq analyses. JH-U, JW, CL, and CT interpreted the data and drafted the manuscript. All authors contributed to the writing of the materials and methods, edited the manuscript, and approved the final version.

FUNDING

This work was supported by (1) the National Institute of Food and Agriculture (NIFA) Project 2018-67015-2701, (2) the NRSP-8 Swine Genome Coordination project, (3) appropriated funds from USDA-ARS CRIS project 5030-31320-004-00D and 3040-31000-100-00D, and (4) an appointment to the Agricultural Research Service (ARS) Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the U.S. Department of Agriculture (USDA). ORISE is managed by ORAU under DOE contract number DE-SC0014664, (5) USDA ARS CRIS Project 8042-32000-102. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDA, ARS, DOE, or ORAU/ORISE. Mention of trade names or products is for information purposes only and does not imply endorsement by the USDA. USDA is an equal opportunity employer and provider.

ACKNOWLEDGMENTS

We acknowledge the DNA facility of the Iowa State University for provision of technical support and sequencing platforms utilized in this study. We are grateful to the NADC animal care staff for their efforts. We thank Dr. Catherine Ernst at Michigan State University for providing the Yorkshire pigs used in the bulkRNA-seq study. We also thank Samuel Humphrey for cell sorting technical assistance. We also thank Kristen Walker for performing the NanoString data collection and initial analyses. We also thank Zahra Bond for technical assistance with cell isolation and FACS, and Dr. Julian Trachsel for early data visualization assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.689406/full#supplementary-material>

REFERENCES

- Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* 14, 1083–1086. doi: 10.1038/nmeth.4463
- Alter, G., Malenfant, J. M., and Altfeld, M. (2004). CD107a as a functional marker for the identification of natural killer cell activity. *J. Immunol. Methods* 294, 15–22. doi: 10.1016/j.jim.2004.08.008
- Alvarez, J. I., Kébir, H., Cheslow, L., Charabati, Chabarati, M., Larochelle, C., et al. (2015). JAML mediates monocyte and CD8 T cell migration across the brain endothelium. *Ann. Clin. Transl. Neurol.* 2, 1032–1037. doi: 10.1002/acn3.255
- Andersson, L., Archibald, A. L., Bottema, C. D., Brauning, R., Burgess, S. C., Burt, D. W., et al. (2015). Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol.* 16:57. doi: 10.1186/s13059-015-0622-4
- Andrews, T. S., Kiselev, V. Y., McCarthy, D., and Hemberg, M. (2021). Tutorial: guidelines for the computational analysis of single-cell RNA sequencing data. *Nat. Protoc.* 16, 1–9. doi: 10.1038/s41596-020-00409-w
- Antsiferova, M., Huber, M., Meyer, M., Piwko-Czuchra, A., Ramadan, T., MacLeod, A. S., et al. (2011). Activin enhances skin tumorigenesis and malignant progression by inducing a pro-tumorigenic immune cell response. *Nat. Commun.* 2:576. doi: 10.1038/ncomms1585
- Arceo, M., Ernst, C. W., Lunney, J. K., Choi, I., Raney, N. E., Huang, T., et al. (2013). Characterizing differential individual response to Porcine Reproductive and Respiratory Syndrome Virus infection through statistical and functional analysis of gene expression. *Front. Livestock Genomics* 3:321. doi: 10.3389/fgene.2012.00321
- Auray, G., Keller, I., Python, S., Gerber, M., Bruggmann, R., Ruggli, N., et al. (2016). Characterization and Transcriptomic Analysis of Porcine Blood Conventional and Plasmacytoid Dendritic Cells Reveals Striking Species-Specific Differences. *J. Immunol.* 197, 4791–4806. doi: 10.4049/jimmunol.1600672
- Auray, G., Talker, S. C., Keller, I., Python, S., Gerber, M., Liniger, M., et al. (2020). High-Resolution Profiling of Innate Immune Responses by Porcine Dendritic Cell Subsets. *Front. Immunol.* 11:1429. doi: 10.3389/fimmu.2020.01429
- Bassler, K., Schulte-Schrepping, J., Warnat-Herresthal, S., Aschenbrenner, A. C., and Schultze, J. L. (2019). The Myeloid Cell Compartment-Cell by Cell. *Annu. Rev. Immunol.* 37, 269–293. doi: 10.1146/annurev-immunol-042718-041728
- Bates, D., Machlet, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Statist. Softw.* 67, 1–48.
- Belizário, J. E., Neyra, J. M., Setúbal Destro, and Rodrigues, M. F. (2018). When and how NK cell-induced programmed cell death benefits immunological protection against intracellular pathogen infection. *Innate Immun.* 24, 452–465. doi: 10.1177/1753425918800200
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* 57, 289–300. doi: 10.2307/2346101
- Berry, M. P., Graham, C. M., McNab, F. W., Xu, Z., Bloch, S. A., Oni, T., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. *Nature* 466, 973–977. doi: 10.1038/nature09247
- Blaser, C., Kaufmann, M., Müller, C., Zimmermann, C., Wells, V., Mallucci, L., et al. (1998). Beta-galactoside-binding protein secreted by activated T cells inhibits antigen-induced proliferation of T cells. *Eur. J. Immunol.* 28, 2311–2319. doi: 10.1002/(SICI)1521-4141(199808)28:08<2311::AID-IMMU2311>3.0.CO;2-G
- Blom, B., Heemskerk, M. H., Verschuren, M. C., van Dongen, J. J., Stegmann, A. P., Bakker, A. Q., et al. (1999). Disruption of alpha beta but not of gamma delta T cell development by overexpression of the helix-loop-helix protein Id3 in committed T cell progenitors. *EMBO J.* 18, 2793–2802. doi: 10.1093/emboj/18.10.2793
- Bordet, E., Fréaud, M., Crisci, E., Bouguyon, E., Rault, S., Pezant, J., et al. (2019). Macrophage-B Cell Interactions in the Inverted Porcine Lymph Node and Their Response to Porcine Reproductive and Respiratory Syndrome Virus. *Front. Immunol.* 10:953. doi: 10.3389/fimmu.2019.00953
- Byrne, K. A., Tuggle, C. K., and Loving, C. L. (2020). Differential induction of innate memory in porcine monocytes by β -glucan or bacillus Calmette-Guérin. *Innate Immun.* 2020:1753425920951607. doi: 10.1177/1753425920951607
- Castro, C. D., and Flajnik, M. F. (2014). Putting J chain back on the map: how might its expression define plasma cell development? *J. Immunol.* 193, 3248–3255. doi: 10.4049/jimmunol.1400531
- Chaussabel, D., and Baldwin, N. (2014). Democratizing systems immunology with modular transcriptional repertoire analyses. *Nat. Rev. Immunol.* 14, 271–280.
- Chaussabel, D., Pascual, V., and Banchereau, J. (2010). Assessing the human immune system through blood transcriptomics. *BMC Biol.* 8:84. doi: 10.1186/1741-7007-8-84
- Choi, J., Baldwin, T. M., Wong, M., Bolden, J. E., Fairfax, K. A., Lucas, E. C., et al. (2019). Haemopedia RNA-seq: a database of gene expression during haematopoiesis in mice and humans. *Nucleic Acids Res.* 47, D780–D785. doi: 10.1093/nar/gky1020
- Costanzo, M. C., Kim, D., Creegan, M., Lal, K. G., Ake, J. A., Currier, J. R., et al. (2018). Transcriptomic signatures of NK cells suggest impaired responsiveness in HIV-1 infection and increased activity post-vaccination. *Nat. Commun.* 9:1212. doi: 10.1038/s41467-018-03618-w
- Dabydeen, S. A., Desai, A., and Sahoo, D. (2019). Unbiased Boolean analysis of public gene expression data for cell cycle gene identification. *Mol. Biol. Cell* 30, 1770–1779. doi: 10.1091/mbc.E19-01-0013
- Davis, W. C., Zuckermann, F. A., Hamilton, M. J., Barbosa, J. I., Saalmüller, A., Binns, R. M., et al. (1998). Analysis of monoclonal antibodies that recognize gamma delta T/null cells. *Vet. Immunol. Immunopathol.* 60, 305–316. doi: 10.1016/s0165-2427(97)00107-4
- Dawson, H. D., Loveland, J. E., Pascal, G., Gilbert, J. G., Uenishi, H., Mann, K. M., et al. (2013). Structural and functional annotation of the porcine immunome. *BMC Genomics* 14:332. doi: 10.1186/1471-2164-14-332
- Delecluse, S., Tsai, M., Shumilov, A., Bencun, M., Arrow, S., Beshirova, A., et al. (2019). Epstein-Barr Virus induces expression of the LPAM-1 integrin in B cells *in vitro* and *in vivo*. *J. Virol.* 93, e1618–e1618. doi: 10.1128/JVI.01618-18
- Denyer, M. S., Wileman, T. E., Stirling, C. M., Zuber, B., and Takamatsu, H. H. (2006). Perforin expression can define CD8 positive lymphocyte subsets in pigs allowing phenotypic and functional analysis of natural killer, cytotoxic T, natural killer T and MHC un-restricted cytotoxic T-cells. *Vet. Immunol. Immunopathol.* 110, 279–292. doi: 10.1016/j.vetimm.2005.10.005
- Dong, Q., Lunney, J. K., Lim, K. S., Nguyen, Y., Hess, A. S., Beiki, H., et al. (2021). Gene expression in tonsils in swine following infection with porcine reproductive and respiratory syndrome virus. *BMC Vet. Res.* 17:88. doi: 10.1186/s12917-021-02785-1
- Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. doi: 10.1038/nprot.2009.97
- Fairbairn, L., Kapetanovic, R., Beraldi, D., Sester, D. P., Tuggle, C. K., Archibald, A. L., et al. (2013). Comparative analysis of monocyte subsets in the pig. *J. Immunol.* 190, 6389–6396. doi: 10.4049/jimmunol.1300365
- Faldyna, M., Samankova, P., Leva, L., Cerny, J., Oujezdska, J., Rehakova, Z., et al. (2007). Cross-reactive anti-human monoclonal antibodies as a tool for B-cell identification in dogs and pigs. *Vet. Immunol. Immunopathol.* 119, 56–62. doi: 10.1016/j.vetimm.2007.06.022
- Foissac, S., Djebali, S., Munyard, K., Vialaneix, N., Rau, A., Muret, K., et al. (2019). Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol.* 17:108. doi: 10.1186/s12915-019-0726-5
- Freeman, T. C., Ivens, A., Baillie, J. K., Beraldi, D., Barnett, M. W., Dorward, D., et al. (2012). A gene expression atlas of the domestic pig. *BMC Biol.* 10:90. doi: 10.1186/1741-7007-10-90
- Gerner, W., Käser, T., and Saalmüller, A. (2009). Porcine T lymphocytes and NK cells—an update. *Dev. Comp. Immunol.* 33, 310–320. doi: 10.1016/j.dci.2008.06.003
- Gerner, W., Talker, S. C., Koinig, H. C., Sedlak, C., Mair, K. H., and Saalmüller, A. (2015). Phenotypic and functional differentiation of porcine alphabeta T cells: current knowledge and available tools. *Mol. Immunol.* 66, 3–13. doi: 10.1016/j.molimm.2014.10.025
- Giotti, B., Chen, S. H., Barnett, M. W., Regan, T., Ly, T., Wiemann, S., et al. (2019). Assembly of a parts list of the human mitotic cell cycle machinery. *J. Mol. Cell Biol.* 11, 703–718. doi: 10.1093/jmcb/mjy063
- Giuffra, E., Tuggle, C. K., and Consortium, F. (2019). Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. *Annu. Rev. Anim. Biosci.* 7, 65–88. doi: 10.1146/annurev-animal-020518-114913
- Grewal, I. S., and Flavell, R. A. (1996). The role of CD40 ligand in costimulation and T-cell activation. *Immunol. Rev.* 153, 85–106. doi: 10.1111/j.1600-065x.1996.tb00921.x
- Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W., Zheng, S., Butler, A., et al. (2020). Integrated analysis of multimodal single-cell data. *bioRxiv*. [preprint].

- Herrera-Urbe, J., Liu, H., Byrne, K. A., Bond, Z. F., Loving, C. L., and Tuggle, C. K. (2020). Changes in H3K27ac at Gene Regulatory Regions in Porcine Alveolar Macrophages Following LPS or PolyIC Exposure. *Front. Genet.* 11:817. doi: 10.3389/fgene.2020.00817
- Hicks, S. C., and Irizarry, R. A. (2015). quantro: a data-driven approach to guide the choice of an appropriate normalization method. *Genome Biol.* 16, 117–117. doi: 10.1186/s13059-015-0679-0
- Hicks, S. C., Okrah, K., Paulson, J. N., Quackenbush, J., Irizarry, R. A., and Bravo, H. C. (2018). Smooth quantile normalization. *Biostatistics* 19, 185–198. doi: 10.1093/biostatistics/kxx028
- Huang, T. H., Uthe, J. J., Bearson, S. M., Demirkale, C. Y., Nettleton, D., Knetter, S., et al. (2011). Distinct peripheral blood RNA responses to *Salmonella* in pigs differing in *Salmonella* shedding levels: intersection of IFNG, TLR and miRNA pathways. *PLoS One* 6:e28768. doi: 10.1371/journal.pone.0028768
- Ihn, H. J., Kim, D. H., Oh, S. S., Moon, C., Chung, J. W., Song, H., et al. (2011). Identification of *Acvr2a* as a Th17 cell-specific gene induced during Th17 differentiation. *Biosci. Biotechnol. Biochem.* 75, 2138–2141. doi: 10.1271/bbb.110436
- Kapetanovic, R., Fairbairn, L., Downing, A., Beraldi, D., Sester, D. P., Freeman, T. C., et al. (2013). The impact of breed and tissue compartment on the response of pig macrophages to lipopolysaccharide. *BMC Genomics* 14:581. doi: 10.1186/1471-2164-14-581
- Kesarwani, P., Murali, A. K., Al-Khami, A. A., and Mehrotra, S. (2013). Redox regulation of T-cell function: from molecular mechanisms to significance in human health and disease. *Antioxid Redox. Signal* 18, 1497–1534. doi: 10.1089/ars.2011.4073
- Kim, C., Jin, J., Weyand, C. M., and Goronzy, J. J. (2020). The Transcription Factor TCF1 in T Cell Differentiation and Aging. *Int. J. Mol. Sci.* 21:21186497. doi: 10.3390/ijms21186497
- Knetter, S. M., Bearson, S. M., Huang, T. H., Kurkiewicz, D., Schroyen, M., Nettleton, D., et al. (2015). *Salmonella enterica* serovar Typhimurium-infected pigs with different shedding levels exhibit distinct clinical, peripheral cytokine and transcriptomic immune response phenotypes. *Innate Immunol.* 21, 227–241.
- Kobayashi, E., Hanazono, Y., and Kunita, S. (2018). Swine used in the medical university: overview of 20 years of experience. *Exp. Anim.* 67, 7–13. doi: 10.1538/expanim.17-0086
- Koltes, J. E., Cole, J. B., Clemmens, R., Dilger, R. N., Kramer, L. M., Lunney, J. K., et al. (2019). A Vision for Development and Utilization of High-Throughput Phenotyping and Big Data Analytics in Livestock. *Front. Genet.* 10:1197. doi: 10.3389/fgene.2019.01197
- Kreslavsky, T., Vilagos, B., Tagoh, H., Poliakova, D. K., Schwickert, T., Wohner, M., et al. (2017). Essential role of the transcription factor Bhlhe41 in regulating the development, self-renewal and BCR repertoire of B-1a cells. *Nat. Immunol.* 18, 442–455. doi: 10.1038/ni.3694
- Krueger, J., Jules, F., Rieder, S. A., and Rudd, C. E. (2017). CD28 family of receptors inter-connect in the regulation of T-cells. *Recept. Clin. Investig.* 4:e1581.
- Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15:r29. doi: 10.1186/gb-2014-15-2-r29
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28, 882–883. doi: 10.1093/bioinformatics/bts034
- Li, S., Labaj, P. P., Zumbo, P., Sykacek, P., Shi, W., Shi, L., et al. (2014). Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat. Biotech.* 32, 888–895. doi: 10.1038/nbt.3000
- Liu, G., Wang, B., Chen, Q., Li, Y., Li, B., Yang, N., et al. (2020). Interleukin (IL)-21 Promotes the Differentiation of IgA-Producing Plasma Cells in Porcine Peyer's Patches via the JAK-STAT Signaling Pathway. *Front. Immunol.* 11:1303. doi: 10.3389/fimmu.2020.01303
- Liu, S., and Trapnell, C. (2016). Single-cell transcriptome sequencing: recent advances and remaining challenges. *F1000Res* 5:F1000FacultyRev-182. doi: 10.12688/f1000research.7223.1
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550. doi: 10.1186/s13059-014-0550-8
- Lun, A. T. L., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T., Marioni, J. C., et al. (2019). EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biol.* 20:63. doi: 10.1186/s13059-019-1662-y
- Mach, N., Gao, Y., Lemonnier, G., Lecardonnell, J., Oswald, I. P., Estellé, J., et al. (2013). The peripheral blood transcriptome reflects variations in immunity traits in swine: towards the identification of biomarkers. *BMC Genomics* 14:894. doi: 10.1186/1471-2164-14-894
- Maroille, T., Lemonnier, G., Lecardonnell, J., Esquerré, D., Ramayo-Caldas, Y., Mercat, M. J., et al. (2017). Deciphering the genetic regulation of peripheral blood transcriptome in pigs through expression genome-wide association study and allele-specific expression analysis. *BMC Genomics* 18:967. doi: 10.1186/s12864-017-4354-6
- Monaco, G., Lee, B., Xu, W., Mustafah, S., Hwang, Y. Y., Carré, C., et al. (2019). RNA-Seq Signatures Normalized by mRNA Abundance Allow Absolute Deconvolution of Human Immune Cell Types. *Cell Rep.* 26, 1627.e–1640.e. doi: 10.1016/j.celrep.2019.01.041
- Munyaka, P. M., Kommadath, A., Fohuse, J., Wilkinson, J., Diether, N., Stothard, P., et al. (2019). Characterization of whole blood transcriptome and early-life fecal microbiota in high and low responder pigs before, and after vaccination for Mycoplasma hyopneumoniae. *Vaccine* 37, 1743–1755. doi: 10.1016/j.vaccine.2019.02.016
- Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., et al. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* 128, e20–e31. doi: 10.1182/blood-2016-05-716480
- Newman, A. M., Steen, C. B., Liu, C. L., Gentles, A. J., Chaudhuri, A. A., Scherer, F., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* 37, 773–782. doi: 10.1038/s41587-019-0114-2
- Palmer, C., Diehn, M., Alizadeh, A. A., and Brown, P. O. (2006). Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* 7:115. doi: 10.1186/1471-2164-7-115
- Paulson, J. N., Chen, C. Y., Lopes-Ramos, C. M., Kuijjer, M. L., Platig, J., Sonawane, A. R., et al. (2017). Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC Bioinformatics* 18:437. doi: 10.1186/s12859-017-1847-x
- Piriou-Guzylack, L., and Salmon, H. (2008). Membrane markers of the immune cells in swine: an update. *Vet. Res.* 39:54. doi: 10.1051/vetres:2008030
- Pizzolato, G., Kaminski, H., Tosolini, M., Franchini, D. M., Pont, F., Martins, F., et al. (2019). Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRV δ 1 and TCRV δ 2 $\gamma\delta$ T lymphocytes. *Proc. Natl. Acad. Sci. U.S.A.* 116, 11906–11915. doi: 10.1073/pnas.1818488116
- Qian Liu, M. M. (2016). Evaluation of methods in removing batch effects on RNA-seq data. *Infect. Dis. Transl. Med.* 2, 3–9. doi: 10.11979/idtm.201601002
- Rincon-Orozco, B., Kunzmann, V., Wrobel, P., Kabelitz, D., Steinle, A., and Herrmann, T. (2005). Activation of V gamma 9V delta 2 T cells by NKG2D. *J. Immunol.* 175, 2144–2151. doi: 10.4049/jimmunol.175.4.2144
- Rodríguez-Gómez, I. M., Talker, S. C., Käser, T., Stadler, M., Reiter, L., Ladinig, A., et al. (2019). Expression of T-Bet, Eomesodermin, and GATA-3 Correlates With Distinct Phenotypes and Functional Properties in Porcine $\gamma\delta$ T Cells. *Front. Immunol.* 10:396. doi: 10.3389/fimmu.2019.00396
- Rusmini, M., Griseri, P., Lantieri, F., Matera, I., Hudspeth, K. L., Roberto, A., et al. (2013). Induction of RET dependent and independent pro-inflammatory programs in human peripheral blood mononuclear cells from Hirschsprung patients. *PLoS One* 8:e59066. doi: 10.1371/journal.pone.0059066
- Rusmini, M., Griseri, P., Matera, I., Pontarini, E., Ravazzolo, R., Mavilio, D., et al. (2014). Expression variability and function of the RET gene in adult peripheral blood mononuclear cells. *J. Cell Physiol.* 229, 2027–2037. doi: 10.1002/jcp.24660
- Saalmüller, A., and Gerner, W. (2016). The immune system of swine. *Encyclop. Immunobiol.* 1, 538–548. doi: 10.1016/B978-0-12-374279-7.12019-3
- Saalmüller, A., Werner, T., and Fachinger, V. (2002). T-helper cells from naive to committed. *Vet. Immunol. Immunopathol.* 87, 137–145. doi: 10.1016/s0165-2427(02)00045-4
- Sasaki, K., Pabst, R., and Rothkott, H. J. (1994). The unique ultrastructure of high-endothelial venules in inguinal lymph nodes of the pig. *Cell Tissue Res.* 276, 85–90. doi: 10.1007/BF00354787
- Schroyen, M., and Tuggle, C. K. (2015). Current transcriptomics in pig immunity research. *Mamm. Genome* 26, 1–20. doi: 10.1007/s00335-014-9549-4

- Sedlak, C., Patzl, M., Saalmüller, A., and Gerner, W. (2014). CD2 and CD8 α define porcine $\gamma\delta$ T cells with distinct cytokine production profiles. *Dev. Comp. Immunol.* 45, 97–106. doi: 10.1016/j.dci.2014.02.008
- Shi, W., Liao, Y., Willis, S. N., Taubenheim, N., Inouye, M., Tarlinton, D. M., et al. (2015). Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. *Nat. Immunol.* 16, 663–673. doi: 10.1038/ni.3154
- Siegers, G. M. (2018). Integral Roles for Integrins in $\gamma\delta$ T Cell Function. *Front. Immunol.* 9:521. doi: 10.3389/fimmu.2018.00521
- Song, L., and Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* 4:48. doi: 10.1186/s13742-015-0089-y
- Steiner, J., Marquardt, N., Pauls, I., Schiltz, K., Rahmoune, H., Bahn, S., et al. (2011). Human CD8(+) T cells and NK cells express and secrete S100B upon stimulation. *Brain Behav. Immun.* 25, 1233–1241. doi: 10.1016/j.bbi.2011.03.015
- Stepanova, K., and Sinkora, M. (2013). Porcine $\gamma\delta$ T lymphocytes can be categorized into two functionally and developmentally distinct subsets according to expression of CD2 and level of TCR. *J. Immunol.* 190, 2111–2120. doi: 10.4049/jimmunol.1202890
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888.e–1902.e. doi: 10.1016/j.cell.2019.05.031
- Summerfield, A., Meurens, F., and Ricklin, M. E. (2015). The immunology of the porcine skin and its value as a model for human skin. *Mol. Immunol.* 66, 14–21. doi: 10.1016/j.molimm.2014.10.023
- Summerfield, A., Rziha, H. J., and Saalmüller, A. (1996). Functional characterization of porcine CD4+CD8+ extrathymic T lymphocytes. *Cell Immunol.* 168, 291–296. doi: 10.1006/cimm.1996.0078
- Summers, K. M., Bush, S. J., and Hume, D. A. (2020). Network analysis of transcriptomic diversity amongst resident tissue macrophages and dendritic cells in the mouse mononuclear phagocyte system. *PLoS Biol.* 18:e3000859. doi: 10.1371/journal.pbio.3000859
- Sutermaster, B. A., and Darling, E. M. (2019). Considerations for high-yield, high-throughput cell enrichment: fluorescence versus magnetic sorting. *Sci. Rep.* 9:227. doi: 10.1038/s41598-018-36698-1
- Swindle, M. M., Makin, A., Herron, A. J., Clubb, F. J. Jr., and Frazier, K. S. (2012). Swine as models in biomedical research and toxicology testing. *Vet. Pathol.* 49, 344–356. doi: 10.1177/0300985811402846
- Szabo, P. A., Levitin, H. M., Miron, M., Snyder, M. E., Senda, T., Yuan, J., et al. (2019). Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* 10:4706. doi: 10.1038/s41467-019-12464-3
- Takamatsu, H. H., Denyer, M. S., Stirling, C., Cox, S., Aggarwal, N., Dash, P., et al. (2006). Porcine gammadelta T cells: possible roles on the innate and adaptive immune responses following virus infection. *Vet. Immunol. Immunopathol.* 112, 49–61. doi: 10.1016/j.vetimm.2006.03.011
- Toka, F. N., Nfon, C. K., Dawson, H., and Golde, W. T. (2009). Accessory-cell-mediated activation of porcine NK cells by toll-like receptor 7 (TLR7) and TLR8 agonists. *Clin. Vaccine Immunol.* 16, 866–878. doi: 10.1128/CVI.00035-09
- Tsang, J. S., Schwartzberg, P. L., Kotliarov, Y., Biancotto, A., Xie, Z., Germain, R. N., et al. (2014). Global analyses of human immune variation reveal baseline predictors of postvaccination responses. *Cell* 157, 499–513.
- Tu, Y. H., Cooper, A. J., Teng, B., Chang, R. B., Artiga, D. J., Turner, H. N., et al. (2018). An evolutionarily conserved gene family encodes proton-selective ion channels. *Science* 359, 1047–1050. doi: 10.1126/science.aao3264
- Upadhyay, G. (2019). Emerging Role of Lymphocyte Antigen-6 Family of Genes in Cancer and Immune Cells. *Front. Immunol.* 10:819.
- van Dam, S., Vösa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2018). Gene co-expression analysis for functional classification and gene-disease predictions. *Brief. Bioinform.* 19, 575–592. doi: 10.1093/bib/bbw139
- Van Goor, A., Pasternak, A., Walker, K., Hong, L., Malgarin, C., MacPhee, D. J., et al. (2020). Differential responses in placenta and fetal thymus at 12?days post infection elucidate mechanisms of viral level and fetal compromise following PRRSV2 infection. *BMC Genomics* 21:763. doi: 10.1186/s12864-020-07154-0
- Vieira Braga, F. A., Teichmann, S. A., and Chen, X. (2016). Genetics and immunity in the era of single-cell genomics. *Hum. Mol. Genet.* 25, R141–R148. doi: 10.1093/hmg/ddw192
- Villani, A. C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356:aah4573. doi: 10.1126/science.aah4573
- Wang, H., Horbinski, C., Wu, H., Liu, Y., Sheng, S., Liu, J., et al. (2016). NanoStringDiff: a novel statistical method for differential expression analysis based on NanoString nCounter data. *Nucleic Acids Res.* 44:e151. doi: 10.1093/nar/gkw677
- Wang, M., Windgassen, D., and Papoutsakis, E. T. (2008). Comparative analysis of transcriptional profiling of CD3+, CD4+ and CD8+ T cells identifies novel immune response players in T-cell activation. *BMC Genomics* 9:225. doi: 10.1186/1471-2164-9-225
- Ware, C. F. (2005). Network communications: lymphotoxins, LIGHT, and TNF. *Annu. Rev. Immunol.* 23, 787–819. doi: 10.1146/annurev.immunol.23.021704.115719
- Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., et al. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* 9:giaa051. doi: 10.1093/gigascience/giaa051
- Wilson, N. K., and Göttgens, B. (2018). Single-Cell Sequencing in Normal and Malignant Hematopoiesis. *Hemasphere* 2:e34. doi: 10.1097/HS9.000000000000034
- Wolock, S. L., Lopez, R., and Klein, A. M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* 8, 281.e–291.e. doi: 10.1016/j.cels.2018.11.005
- Wong, R., and Bhattacharya, D. (2020). ZBTB38 is dispensable for antibody responses. *PLoS One* 15:e0235183. doi: 10.1371/journal.pone.0235183
- Yang, H., and Parkhouse, R. M. (1996). Phenotypic classification of porcine lymphocyte subpopulations in blood and lymphoid tissues. *Immunology* 89, 76–83. doi: 10.1046/j.1365-2567.1996.d01-705.x
- Yang, H., and Parkhouse, R. M. (1997). Differential expression of CD8 epitopes amongst porcine CD8-positive functional lymphocyte subsets. *Immunology* 92, 45–52. doi: 10.1046/j.1365-2567.1997.00308.x
- Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Alvarez-Jarreta, J., Amode, M. R., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688. doi: 10.1093/nar/gkz966
- Young, M. D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* 9:giaa151. doi: 10.1093/gigascience/giaa151
- Zemmour, D., Zilionis, R., Kiner, E., Klein, A. M., Mathis, D., and Benoist, C. (2018). Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* 19, 291–301. doi: 10.1038/s41590-018-0051-0
- Zhang, B., Lin, Y. Y., Dai, M., and Zhuang, Y. (2014). Id3 and Id2 act as a dual safety mechanism in regulating the development and population size of innate-like $\gamma\delta$ T cells. *J. Immunol.* 192, 1055–1063. doi: 10.4049/jimmunol.1302694
- Zheng, G. X., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8:14049. doi: 10.1038/ncomms14049
- Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., et al. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10:1523. doi: 10.1038/s41467-019-09234-6
- Zhu, L., Yang, P., Zhao, Y., Zhuang, Z., Wang, Z., Song, R., et al. (2020). Single-Cell Sequencing of Peripheral Mononuclear Cells Reveals Distinct Immune Response Landscapes of COVID-19 and Influenza Patients. *Immunity* 53, 685.e–696.e. doi: 10.1016/j.immuni.2020.07.009
- Zhu, Y., Yao, S., Iliopoulou, B. P., Han, X., Augustine, M. M., Xu, H., et al. (2013). B7-H5 costimulates human T cells via CD28H. *Nat. Commun.* 4:2043. doi: 10.1038/ncomms3043
- Zuckermann, F. A. (1999). Extrathymic CD4/CD8 double positive T cells. *Vet. Immunol. Immunopathol.* 72, 55–66.

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

Copyright © 2021 Herrera-Uribe, Wiarda, Sivasankaran, Daharsh, Liu, Byrne, Smith, Lunney, Loving and Tuggle. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.