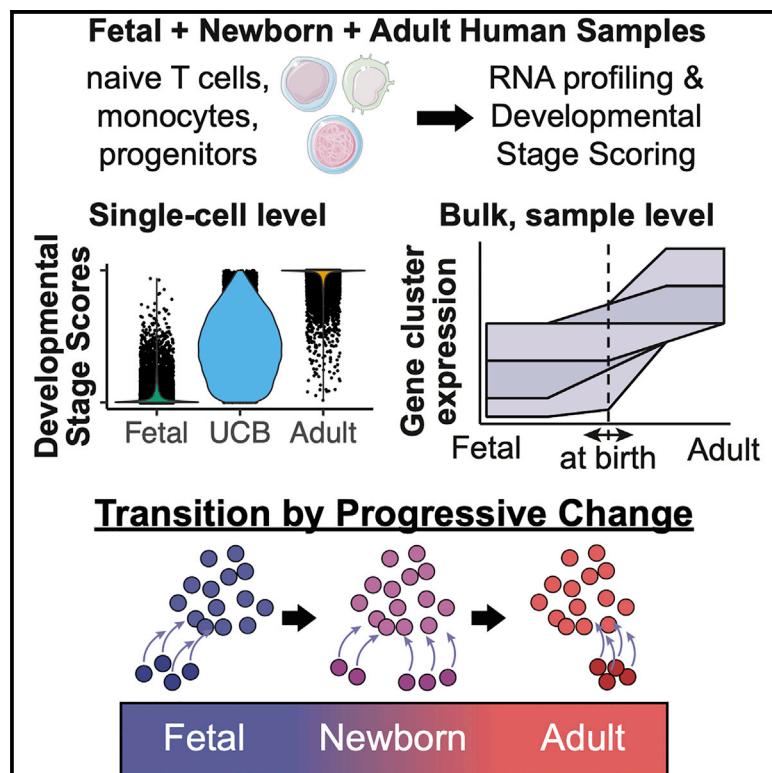


## Single-Cell Mapping of Progressive Fetal-to-Adult Transition in Human Naive T Cells

### Graphical Abstract



### Authors

Daniel G. Bunis, Yelena Bronevetsky, Elisabeth Krow-Lucal, ..., Marina Sirota, Joseph M. McCune, Trevor D. Burt

### Correspondence

marina.sirota@ucsf.edu (M.S.), mike.mccune@gatesfoundation.org (J.M.M.), trevor.burt@duke.edu (T.D.B.)

### In Brief

Through transcriptional profiling of human fetal, perinatal, and adult lymphoid, myeloid, and hematopoietic progenitor cells, Bunis et al. demonstrate that fetal-to-adult transition occurs progressively along a continuum of maturity—with a substantial degree of inter-individual variation at the time of birth—rather than via regime change between discrete waves.

### Highlights

- Human naive T cell fetal-to-adult transition fits a model of progressive change
- Newborn cells appear to be relatively homogeneous, not an admixture from distinct waves
- Transition progress at birth shows a high degree of inter-individual variability
- Hematopoietic progenitors of umbilical cord blood are not fully adult transitioned



## Resource

# Single-Cell Mapping of Progressive Fetal-to-Adult Transition in Human Naive T Cells

Daniel G. Bunis,<sup>1,2</sup> Yelena Bronevetsky,<sup>3,14,17</sup> Elisabeth Krow-Lucal,<sup>3,17</sup> Nirav R. Bhakta,<sup>4</sup> Charles C. Kim,<sup>3,13</sup> Srilaxmi Nerella,<sup>4</sup> Norman Jones,<sup>1,3</sup> Ventura F. Mendoza,<sup>1,3</sup> Yvonne J. Bryson,<sup>5</sup> James E. Gern,<sup>6</sup> Rachel L. Rutishauser,<sup>3</sup> Chun Jimmie Ye,<sup>2,7,8,9,10,11</sup> Marina Sirota,<sup>2,12,18,\*</sup> Joseph M. McCune,<sup>3,15,18,\*</sup> and Trevor D. Burt<sup>1,12,16,18,19,\*</sup>

<sup>1</sup>Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA

<sup>2</sup>Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA, USA

<sup>3</sup>Division of Experimental Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

<sup>4</sup>Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

<sup>5</sup>Division of Pediatric Infectious Diseases, Department of Pediatrics, David Geffen School of Medicine at UCLA, Mattel Children's Hospital UCLA, Los Angeles, CA, USA

<sup>6</sup>Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA

<sup>7</sup>Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA

<sup>8</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA

<sup>9</sup>Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA

<sup>10</sup>Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA

<sup>11</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA

<sup>12</sup>Department of Pediatrics, Division of Neonatology, University of California, San Francisco, San Francisco, CA, USA

<sup>13</sup>Present address: Verily Life Sciences, South San Francisco, CA 94080, USA

<sup>14</sup>Present Address: Berkeley Lights Inc., Emeryville, CA, USA

<sup>15</sup>Present Address: HIV Frontiers/Global Health Innovative Technology Solutions, Bill & Melinda Gates Foundation, Seattle, WA, USA

<sup>16</sup>Present Address: Department of Pediatrics, Division of Neonatology and the Children's Health and Discovery Initiative, Duke University School of Medicine, Durham, NC, USA

<sup>17</sup>These authors contributed equally

<sup>18</sup>Senior author

<sup>19</sup>Lead Contact

\*Correspondence: marina.sirota@ucsf.edu (M.S.), mike.mccune@gatesfoundation.org (J.M.M.), trevor.burt@duke.edu (T.D.B.)  
<https://doi.org/10.1016/j.celrep.2020.108573>

## SUMMARY

Whereas the human fetal immune system is poised to generate immune tolerance and suppress inflammation *in utero*, an adult-like immune system emerges to orchestrate anti-pathogen immune responses in post-natal life. It has been posited that cells of the adult immune system arise as a discrete ontological “layer” of hematopoietic stem-progenitor cells (HSPCs) and their progeny; evidence supporting this model in humans has, however, been inconclusive. Here, we combine bulk and single-cell transcriptional profiling of lymphoid cells, myeloid cells, and HSPCs from fetal, perinatal, and adult developmental stages to demonstrate that the fetal-to-adult transition occurs progressively along a continuum of maturity—with a substantial degree of inter-individual variation at the time of birth—rather than via a transition between discrete waves. These findings have important implications for the design of strategies for prophylaxis against infection in the newborn and for the use of umbilical cord blood (UCB) in the setting of transplantation.

## INTRODUCTION

Cells of the developing fetal immune system are functionally and transcriptionally unique compared with phenotypically similar cells in the adult. The biological processes by which cells that are developmentally restricted to the fetal period are replaced by their counterparts in post-natal life have been the subject of much speculation. Previous studies of B and T lymphocyte lineages in mice and birds suggested the programmed, staged emergence of temporally restricted immune

cell populations throughout gestation and beyond (Havran and Allison, 1988; Hayakawa et al., 1985; Jotereau and Le Douarin, 1982; Kantor et al., 1992; Lalor et al., 1989; Le Douarin and Jotereau, 1975; Montecino-Rodriguez et al., 2006). Three decades ago, several seminal publications proposed a model in which development of the immune system is temporally “layered,” positing that fetal and adult lymphoid cells arise from distinct, developmentally restricted lineages of fetal and adult hematopoietic stem cells (HSCs) (Herzenberg and Herzenberg, 1989; Ikuta et al., 1990).



While additional evidence for layered immune system development has been generated more recently in mice (Montecino-Rodriguez et al., 2016, 2018; Ramond et al., 2014), and many immune cells generated in the human fetus are distinct from those arising from adult HSCs, including CD5<sup>+</sup> B-1 B cells, fetal erythrocytes, microglia, and certain  $\gamma\delta$  T cells (Ginhoux et al., 2013; Hadland and Yoshimoto, 2018; Montecino-Rodriguez and Dorshkind, 2012; Stamatoyannopoulos, 2005; Tieppo et al., 2020), it remains unconfirmed whether a similar developmental transition occurs in humans around the time of birth. In our previous work, we showed that midgestational naïve  $\alpha\beta$  CD4 T cells differ from phenotypically similar adult naïve T cells in that they are predisposed to promote tolerance and preferentially differentiate into FoxP3<sup>+</sup> regulatory T ( $T_{reg}$ ) cells upon activation (Bronevetsky et al., 2016; Mold et al., 2008, 2010; Ng et al., 2019). We also reported that fetal classical (CD14<sup>+</sup>CD16<sup>-</sup>) monocytes exhibit distinct responses to cytokine stimulation and impaired antigen presentation capacity compared with adult classical monocytes (Krow-Lucal et al., 2014). Lastly, we demonstrated that the unique nature of fetal naïve T cells is programmed at the level of their hematopoietic progenitors: when transplanted into humanized severe combined immunodeficiency (SCID)-hu Thy/Liv mice, human fetal, but not adult, CD34<sup>+</sup> HSPCs give rise to naïve CD4 T cells with a transcriptional profile, and a predisposition toward  $T_{reg}$  differentiation, that is similar to that found in primary fetal naïve CD4 T cells (Mold et al., 2010). Thus, T cells and monocytes generated in the second trimester may represent effectors of a fetal immune lineage that is distinct from an adult immune lineage generated later in development. The processes by which fetal cells transition to more mature developmental states remain unknown.

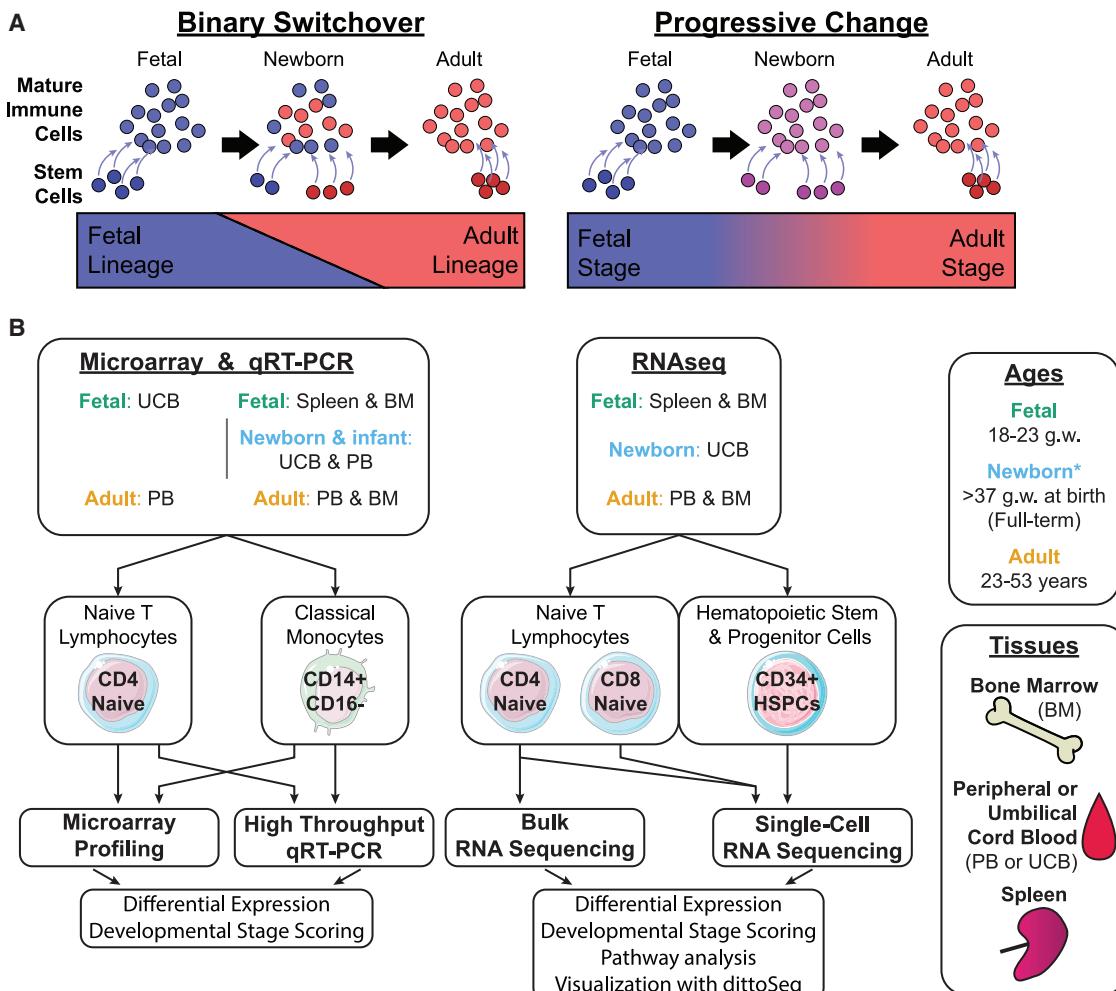
Human newborns have impaired responses to vaccines (Siegrist, 2001) and are more susceptible to serious infections, even by microbes that are generally non-pathogenic in older children and adults (Blanchard et al., 2015; Phares et al., 2008). The persistence of fetal tolerogenic immune responses at birth would predictably contribute to these immunological vulnerabilities by blunting anti-vaccine and anti-microbial immune responses. As the newborn period represents an intermediate time point between fetal and adult immunity, understanding the timing and mechanism by which the fetal-to-adult transition occurs could accordingly have significant implications for newborn health. The layered immune system hypothesis (Herzenberg and Herzenberg, 1989) posits that HSPCs giving rise to a tolerogenic fetal immune system are superseded by distinct HSPCs that instead give rise to a protective adult HSPC-derived immune system and that for a time these immune systems are layered upon each other, exerting opposing influences. Alternatively, the fetal-to-adult transition might instead occur through relatively uniform progressive maturational changes across immune cell populations, leading to a continuous spectrum of intermediate phenotypes (Figure 1A). Depending on which of these two models is operative, naïve T cells at the time of full-term birth would be predicted to either consist of a mixture of fetal- and adult-like T cells or a relatively homogeneous population of cells with an intermediate newborn phenotype. This study was designed to distinguish between these alternatives.

We initially utilized conventional transcriptional profiling approaches (mRNA microarray and quantitative reverse transcriptase polymerase chain reaction [qRT-PCR]) to identify a transcriptional signature of fetal and adult immune cells and to demonstrate that immune cells of newborn humans show intermediate and variable expression of this maturation signature. To better understand the maturational state and composition of T cells at the single-cell level during the newborn period, we then leveraged single-cell RNA sequencing (scRNA-seq) technology, a machine-learning approach within our bioinformatics pipeline, and a newly developed visualization tool (dittoSeq; (Bunis et al., 2020)) built for side-by-side analysis of bulk and scRNA-seq data to directly compare the transcriptomes of fetal, full-term umbilical cord blood (UCB), and adult naïve T cells and CD34<sup>+</sup> HSPCs. By doing so, we were able to ask whether fetal-associated features persist in UCB immune cells at birth and, if so, whether T cells in UCB are composed either of layered waves of cells with either fetal or adult transcriptional programs or of a single population with an intermediate transcriptional program (Figure 1A).

## RESULTS

### The Transition between Fetal and Adult Transcriptional Programs in Human T Cells and Monocytes Is Incomplete at Full-Term Gestation

We first sought to investigate the extent to which the immune system at full-term gestation has transitioned from a fetal to an adult program. Most transcriptional profiling studies of fetal immunity have been carried out utilizing cells associated with lymphoid organs (e.g., thymus, spleen, gut, mesenteric lymph node, or bone marrow [BM]) due to the technical difficulty of sampling fetal peripheral blood (PB) (Bronevetsky et al., 2016; Cupedo et al., 2005; Halkias et al., 2019; Krow-Lucal et al., 2014; Li et al., 2019; Mold et al., 2008, 2010; Ng et al., 2019). Here, a fetal-versus-adult expression signature was first derived using global gene expression analysis in phenotypically similar populations of sort-purified fetal and adult peripheral blood classical monocytes (human leukocyte antigen [HLA]-DR<sup>+</sup>CD14<sup>+</sup>CD16<sup>-</sup>) and naïve, non- $T_{reg}$ , CD4 T cells (CD3<sup>+</sup>CD4<sup>+</sup>CD25<sup>-</sup>CD45RA<sup>+</sup>CD27<sup>+</sup>). Adult and fetal peripheral blood cells were obtained from adult donor apheresis units and from fetal UCB, respectively (sorting strategy shown in Figure S1). To define common hematopoietic developmental programs conserved between myeloid and lymphoid lineages, we identified 159 genes that were significantly differentially expressed in the same direction (false discovery rate [FDR] < 0.05 and expression fold change > 1.5) between fetal and adult cells in both monocytes and naïve T cells (Table S1). These signature genes were then used to build a qRT-PCR-based developmental stage scoring system for use in full-term UCB and other peripheral blood samples. After validation and optimization testing of qRT-PCR primers, a set of 33 genes shared between monocytes and T cells met criteria to represent a fetal versus adult transcriptional signature (Figure 2A). Principal component analysis (PCA) was used to weigh the expression of signature genes by the degree to which they contribute to distinguishing fetal-versus-adult identity for T cells and for monocytes separately (Figures S2A and S2B), and we confirmed that fetal and adult T cell and monocyte



**Figure 1. Theoretical and Experimental Overview**

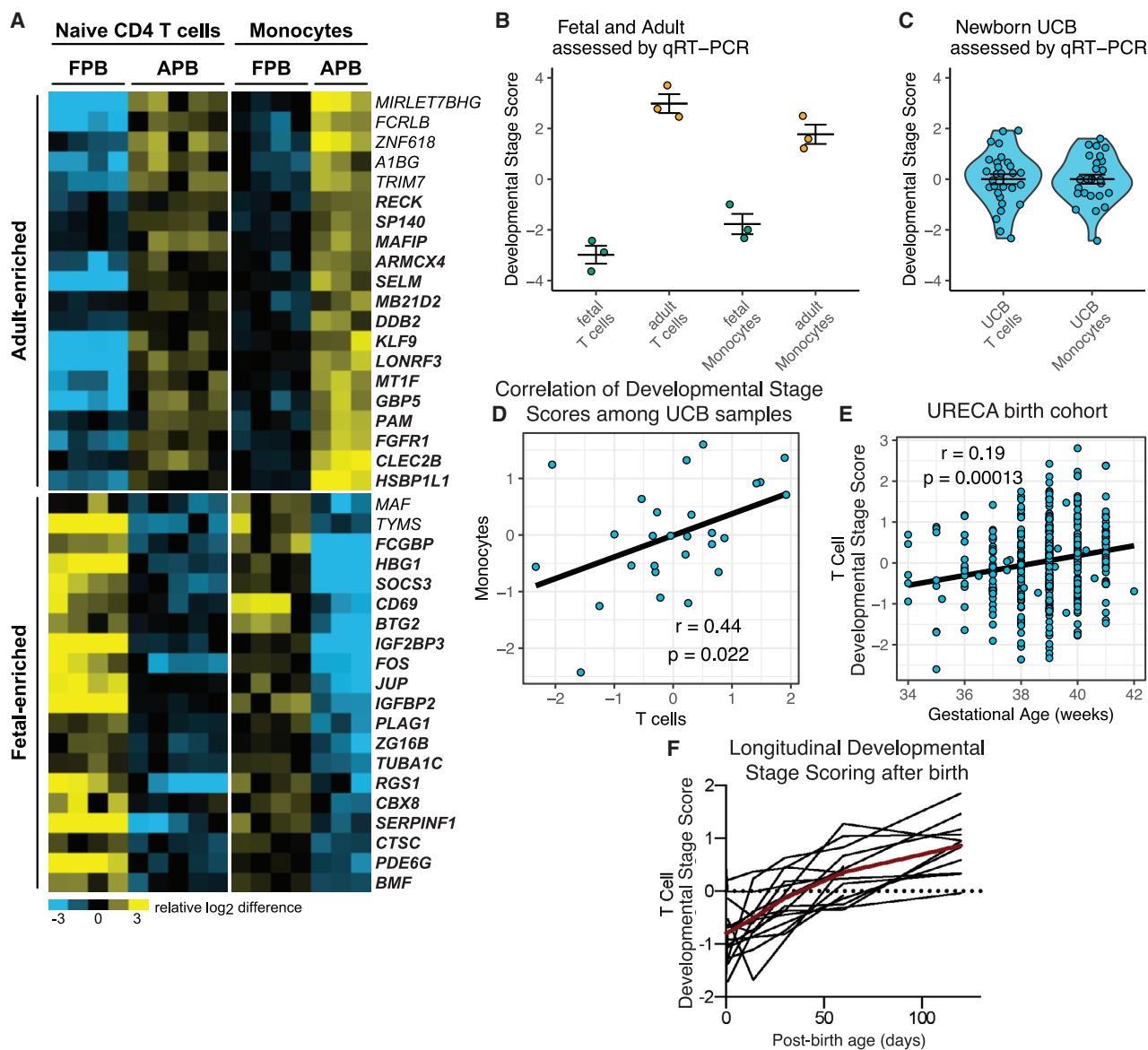
(A) Theoretical models for how the immune cell fetal-to-adult transition may occur. Left: the binary switchover model in which an initial wave of hematopoiesis emanating from fetal HSPCs gives rise to a fetal layer of differentiated immune cells. A subsequent wave arising from distinct adult HSPCs then results in the adult layer of differentiated immune progeny cells. During the transition period, the co-existence of fetal and adult layers would have an intermediate phenotype at the population level, but not at the single-cell level. Right: the progressive change model in which an original population of fetal-phenoype HSPCs undergoes multiple gradual, relatively homogeneous changes that ultimately result in adoption of an adult HSPC phenotype. Throughout this transition, HSPCs that are intermediate between fetal and adult give rise to progeny cells that also have intermediate phenotypic characteristics. In the newborn period, the immune system would have an intermediate phenotype at the population and single-cell level.

(B) Experimental overview showing the tissues collected and cells sorted to generate the various datasets of the current study, and methods used to compare fetal, newborn, and adult cells within these datasets. Asterisk (\*) indicates all newborn samples were collected at the time of birth from deliveries in this time period except for samples of the URECA cohort, which spanned deliveries from 34 to 42 gestational weeks (g.w.).

populations from prior microarray analyses (Krow-Lucal et al., 2014; Mold et al., 2010) scored appropriately as fetal or adult (Figures S2C and S2D). We further demonstrated that naive CD4 and classical monocyte populations sorted from newly processed fetal tissue (spleen or BM) and adult peripheral blood, subjected to qRT-PCR and analyzed for developmental stage score, also scored appropriately as highly fetal or adult (Figure 2B).

To assess the degree of immune maturation (i.e., the degree of transition from fetal-to-adult gene expression patterns) at full-term gestation, we extended our developmental stage score analysis to naive CD4 T cells and classical monocytes that were sort purified from the UCB of a cohort of healthy, vaginally

delivered, full-term newborns (>37 weeks gestational age at birth, n = 29). Although we observed a wide range of inter-individual variability among samples, this variability across UCB CD4 T cell and monocyte populations was less than the variation between fetal and adult samples (Figure 2C). Notably, some samples had monocytes and T cells with more adult-like gene expression scores, while others scored more similarly to fetal samples. We also found that, within a given individual, there was a significant correlation between T cell and monocyte signature scores, suggesting that the fetal-to-adult transition of these hematopoietic lineages may be under control of a shared regulatory mechanism (Figure 2D).



**Figure 2. Population-Level Developmental Stage Scoring Places UCB and Infant Immune Cells Intermediate between Fetal and Adult**

Fetal and adult peripheral blood (PB) naive CD4 T cell and classical monocytes were transcriptionally profiled by microarray. Genes differentially expressed across ages within both cell types were used to build a developmental stage score model that was then used to score fetal-to-adult transition progress of new samples. See also [Figures S1](#) and [S2](#).

(A) Heatmap showing relative expression of top differentially expressed genes within individual samples. Bolded genes were used for developmental stage scoring.

(B and C) Developmental stage scores of naive CD4 T cells and classical monocytes from fetal and adult samples (B; n = 3 each) or from UCB samples (C; T cells n = 29, monocytes n = 27). Points are score from bulk analysis of individual samples. Bars are mean and standard error of the mean (SEM).

(D) Scatterplot showing correlation between developmental stage scores of UCB naive CD4 T cells and monocytes from the same samples (n = 27). Pearson correlation coefficient (r), associated p value, and least-squares regression line are shown.

(E) Scatterplot showing correlation between developmental stage score and gestational age from T cells of a larger birth cohort (n = 405 samples). Points are score from bulk analysis of individual samples. The r, associated p value, and least-squares regression line are shown.

(F) Developmental stage scoring of naive CD4 T cell samples tracked longitudinally. Black lines indicate individuals (n = 15); heavier red line indicates mean value at a given time point.

We then applied the developmental stage scoring system to analyze sort-purified CD4 T cells and classical monocytes from a much larger independent birth cohort of infants born between

34 weeks gestational age (GA; i.e., late pre-term) and 42 weeks GA (i.e., post-term) with a reduced set of markers (see [STAR Methods](#) for details). A wide range of variability at any given

GA was observed, as well as a significant positive correlation between developmental stage score and GA at birth for T cells but not monocytes (Figure 2E). The absence of correlation between developmental stage score and GA in monocytes may be the result of the relatively short lifespan of monocytes resulting in increased heterogeneity in samples' developmental stages at any a given time point compared with T cells.

Given that CD4 T cells displayed graded maturation even in a focused window of GA ranges at birth, we proceeded to investigate the pace of fetal-to-adult transition specifically in CD4 T cells after birth in a separate cohort of HIV-exposed, but uninfected and otherwise healthy neonates. Developmental stage scoring analysis performed on naive CD4 T cells collected longitudinally for the first several months of life revealed a steady transition in developmental stage scores, beginning with a largely fetal signature and transitioning to a more adult-like signature within 1–2 months (Figure 2F).

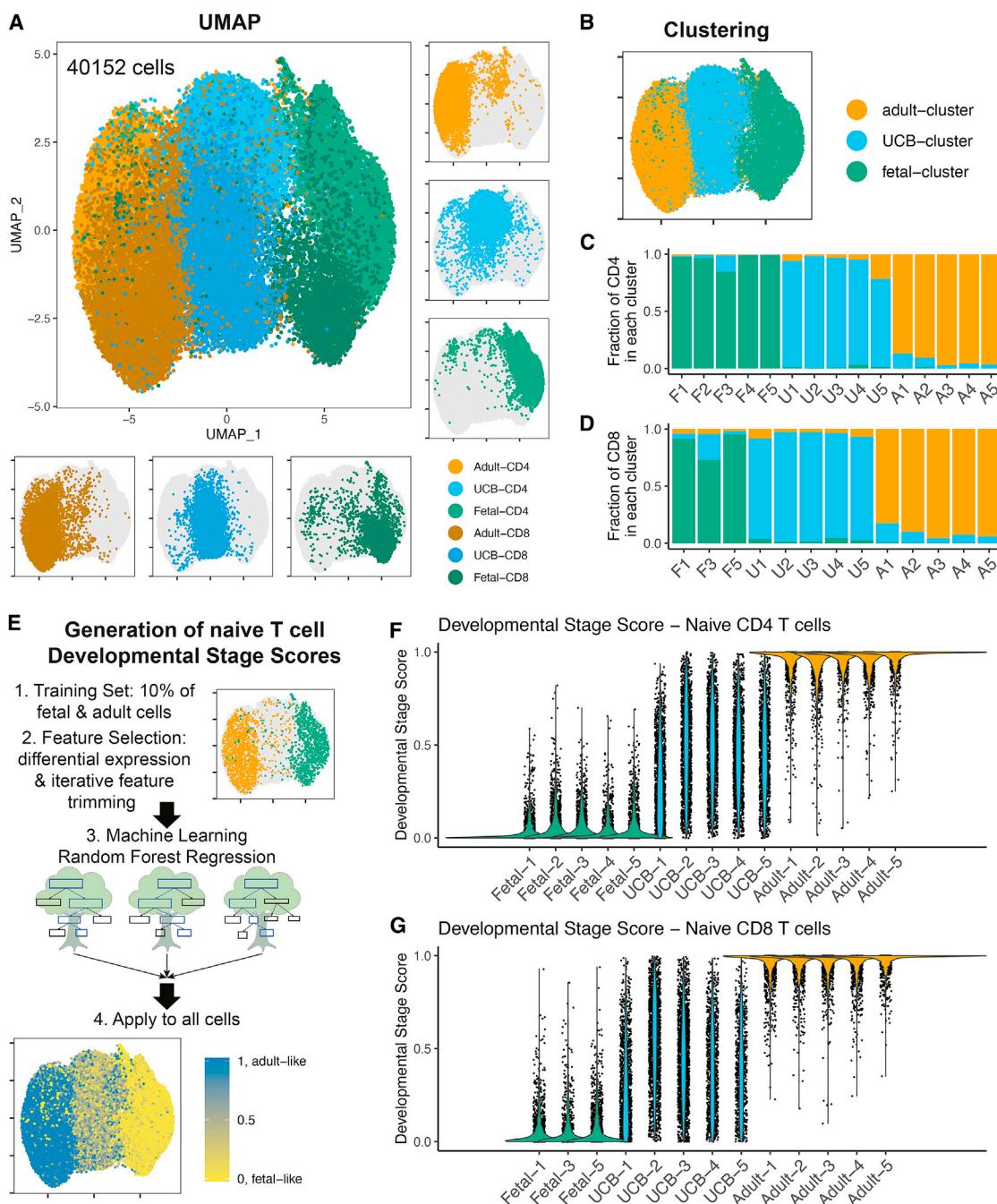
#### Individual UCB T Cells Have Intermediate Transcriptional Profiles and Are Unimodally Distributed between Fetal and Adult Cell Profiles

While bulk sample analysis revealed that UCB cell transcriptional profiles are generally intermediate between those of fetal and adult cells, such an approach does not distinguish between the two proposed models of fetal-to-adult transition (Figure 1A). To determine whether UCB cells are either composed of a mixture of fetal- and adult-like cells (i.e., as predicted by the layering hypothesis) or a relatively homogeneous population with intermediate gene expression (i.e., consistent with a model of progressive change), we turned to scRNA-seq. Fetal splenic, full-term UCB, and adult peripheral blood naive, non-T<sub>reg</sub>, CD4 T cells (CD4<sup>+</sup>CD45RA<sup>+</sup>CD27<sup>+</sup>CD25<sup>-</sup>) and naive, non-stem cell memory, CD8 T cells (CD8<sup>+</sup>CD45RA<sup>+</sup>CD27<sup>+</sup>CCR7<sup>+</sup>CD95<sup>-</sup>) were purified by fluorescence-activated cell sorting (FACS) ( $n = 5$  per age group; Figure S3). Cells from all samples yielding more than 1,000 cells were pooled between three lanes for library generation in such a way that CD4-versus-CD8 T cell identity would be known after sample identification and doublet removal with Demuxlet. Altogether, we obtained 40,152 T cell transcriptomes representing a median of 1,464 CD4 or CD8 cells per sample (range 777–1,829; Figures S4A and S4B).

Dimensionality reduction analysis and clustering were performed using Seurat (Butler et al., 2018) to query general transcriptional relationships. In uniform manifold approximation and projection (UMAP) (McInnes et al., 2018), cells from distinct ages grouped in distinct regions of the UMAP plot, with fetal cells at the positive end, UCB cells in the middle, and adult cells at the negative end of UMAP\_1. Using Seurat's Louvain clustering algorithm to quantify the separation within dimensionality reduction space, clustering at low resolution grouped cells of the same age with 92.5% accuracy overall (Figure 3B). Notably, and in contradistinction to the layering hypothesis, rather than consisting of a mixture of fetal-clustering and adult-clustering cells, CD4 and CD8 naive T cells from UCB samples instead clustered mostly (91.9% mean per sample, 77.3% minimum) as a singular population of cells that aggregated between the adult and fetal clusters (Figures 3C and 3D).

To more directly characterize the fetal-to-adult progression of individual cells, a single-cell developmental stage score was derived. Here, we leveraged a robust machine learning approach, which focused on commonly captured genes that were part of the fetal-to-adult transition and which would therefore be less influenced by noise within sparsely captured single-cell transcriptome data. We applied the relatively assumption-agnostic random forest machine learning regression method to an initial training set of 10% of the fetal and adult naive T cells and an initial set of marker genes that were differentially expressed (log fold change  $\geq 0.585$ , FDR  $< 5\%$ ) between these cells (Figure 3E). To narrow down the number of marker genes, and thereby reduce the chances of the model becoming overfit to the training set, the iterative model building and feature elimination algorithm from the feseR package (Perez-Riverol et al., 2017) was utilized. The final set of marker genes spanned diverse functionalities including transcriptional and post-translational regulation, signal transduction, and cytoskeletal regulation (Table S2, bolded genes). Finally, expression of these 17 definitive markers within the training set cells was used to train a random forest regression model. Confirming this single-cell developmental stage score model's ability to properly score non-transitioned (i.e., fetal) and fully transitioned (i.e., adult) cells, it accurately scored the set of fetal and adult cells which were left out of the training process (Figures 3F and 3G; area under the receiver operating characteristic curve [AUROC] = 0.9996). Importantly, although phenotypic differences, such as those resulting from prior homeostatic proliferation, exist between naive cells within an individual fetal or adult sample (van den Broek et al., 2018; Mold et al., 2019), our system consistently scores most naive T cells of the same fetal or adult sample similarly, indicating that the system is not sensitive to such features.

We next evaluated the scores given to cells from UCB samples, noting four key features. First, both CD4 and CD8 UCB naive T cells generally received intermediate developmental stage scores (Figures 3F and 3G): whereas 84% of fetal cells were scored 0.1 or lower and 89% of adult cells scored 0.9 or greater, 88% of newborn UCB cells scored between 0.1 and 0.9 (mean = 0.43, standard deviation = 0.24). Second, the scores of cells within each individual UCB sample were unimodally distributed, consistent with progressive change, rather than with a binary switch, which would have shown a bimodal distribution (Figures 3F and 3G). Third, variation between samples appears as a shift in the mode of a sample's scores rather than via differences in abundance of very low scoring (i.e., fetal-like) or very high scoring (i.e., adult-like) cells, indicating that variation in progression manifests as relatively homogeneous changes rather than as differences in preponderance of cells with distinct phenotypes. Lastly, the expression pattern among UCB cells for six of the marker genes was more closely aligned with either the fetal cell pattern (*RPS24*, *SOX4*, *RGS1*) or the adult cell pattern (*UBB*, *ACTB*, *HSP90AA1*), suggesting that individual genes that make up the signature undergo transitions in expression with different timing (Figure S4C). These findings demonstrate that the transition from fetal-to-adult layers during human T cell development occurs as a process of gradual, progressive change and not by layering of a fetal lineage alongside a distinct adult lineage.



**Figure 3. Single-Cell-Level Developmental Stage Scoring Places Individual UCB Naive T Cells Intermediate between Fetal and Adult**  
Fetal splenic, full-term UCB, and adult PB naive CD4 ( $n = 15$ ) and CD8 ( $n = 13$ ) T cell samples were profiled by single-cell RNA sequencing (scRNA-seq). See also Figures S3 and S4.

(A) UMAP plot showing distribution of cells of each age and CD4-versus-CD8 lineages. Separate plots along the bottom and right show cells of individual identities.

(B) UMAP plot showing Louvain clustering of single-cell profiles.

(C and D) Quantification of CD4 (C) and CD8 (D) T cell clustering where each column represents a different sample. A, adult; F, fetal; U, UCB.

(E) Overview of single-cell developmental stage score model generation.

(F and G) Developmental stage scoring of individual CD4 (F) and CD8 (G) naive T cells for each sample. Points are individual cells.

### Subsets of Genes Undergo Fetal-to-Adult Transition with Different Timing

Given the evidence that the fetal-to-adult transition occurs via progressive change between fetal and adult transcriptional programs, we hypothesized that the UCB naive CD4 T cell transcriptome would share some properties with fetal naive CD4 T cells and others with adult naive CD4 T cells. We also hypothesized that some fetal and adult genes may be intermediately expressed by UCB cells, while other genes might be unique to UCB. To address these hypotheses, we utilized bulk RNA-seq to obtain greater sequencing depth, and to thus achieve more complete transcriptional profiling, of fetal, UCB, and adult naive CD4 T cells. Both the bulk and single-cell datasets showed similar intermediate placement of UCB cells in dimensionality reduction analysis. In PCA of the bulk RNA-seq dataset, UCB samples clustered at an intermediate point between fetal and adult samples in PC1 ([Figure 4A](#)), and CD4 and CD8 T cells from UCB similarly clustered intermediately in UMAP\_1 between fetal and adult cells in UMAP analysis of the scRNA-seq data ([Figure 3A](#)). Furthermore, when genes differentially expressed between ages in both these datasets were compared using equivalent differential expression criteria, all but seven genes identified in the single-cell dataset were also identified in the bulk dataset ([Figure S5A](#)). Finally, to address the potential confounder of tissue origin in the RNA-seq datasets (i.e., fetal spleen versus blood for newborn and adult samples), we utilized the differentially expressed genes from our prior microarray analysis ([Figure 2](#)) to compare expression between circulating and tissue-derived fetal T cells and found that genes differentially expressed between circulating fetal-versus-adult cells could discriminate expression patterns in fetal spleen- versus newborn UCB- and versus adult peripheral blood-derived T cells analyzed by bulk RNA-seq ([Figures S5B–S5D](#)).

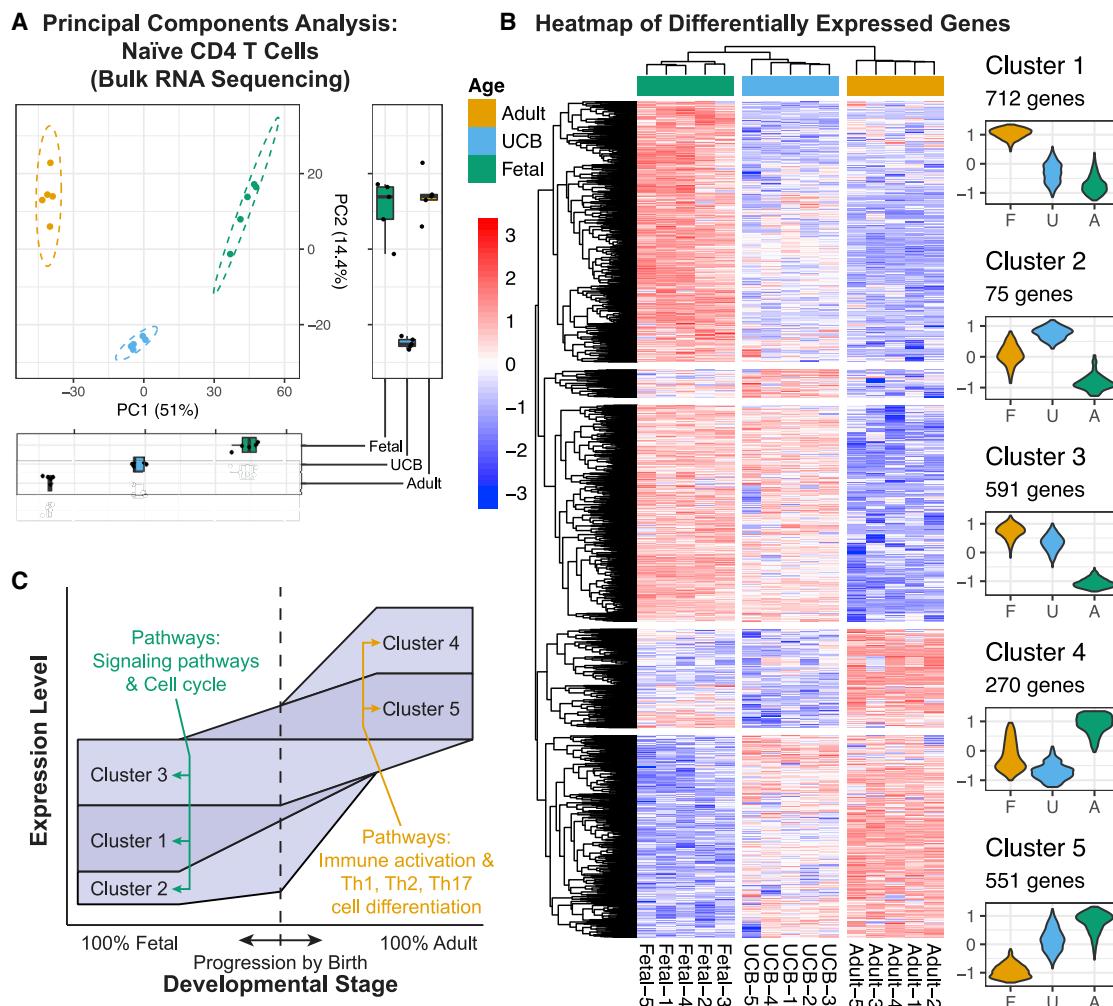
Upon differential expression ( $\log_2$  fold change  $\geq 1.5$ , FDR  $\leq 0.05$ ) analysis of the bulk RNA-seq dataset, 2,201 unique genes were identified as significantly differentially expressed between all pairwise age comparisons: 1,840 genes between fetal and adult naive T cells; 713 genes between fetal and UCB; and 855 genes between UCB and adult. Of these pairwise comparisons, 443 genes were upregulated in both fetal and UCB samples, but not adult samples, while 154 genes were enriched in both UCB and adult samples, but not fetal samples ([Figure S5E](#)). The k-means clustering of all differentially expressed genes ([Figure 4B](#)) revealed a spectrum of gene expression clusters that were distinctively fetal or adult upregulated with varying degrees of expression in UCB cells. One cluster of fetal upregulated genes remained fully upregulated in UCB cells (cluster 3), while, in contrast, another cluster of fetal upregulated genes was partially downregulated (cluster 1) and a third cluster was further upregulated (cluster 2) in UCB T cells. In the other direction, cluster 5 represented adult-upregulated genes that were already partially expressed in UCB T cells, while cluster 4 represented a subset of adult-upregulated genes that showed relatively low expression in UCB T cells. These results strongly suggest that clusters of co-regulated groups of genes, or gene expression “modules,” undergo transition from fetal-like expression patterns to adult-like expression patterns with different timing ([Figure 4C](#)).

### Several Gene Pathways Are Uniquely Enriched within the UCB T Cell Transcriptome

Gene pathway analysis is frequently used to interpret gene expression data in such a way as to better understand the functional programs that are active in a given cell or population of cells. Comparing gene pathways that are activated in fetal, UCB, and adult samples, particularly in comparison with one another, may provide a deeper understanding of the cellular functional adaptations at each developmental stage. To identify specific fetal- and adult-associated genetic programs expressed in UCB cells, we initially applied unbiased pathway enrichment analysis to fetal-associated cluster genes (clusters 1, 2, and 3) and adult-associated cluster genes (clusters 4 and 5) highlighted in [Figures 4B](#) and [4C](#) and listed in [Table S3](#). Immune activation pathways, both general (e.g., antigen processing and presentation, T helper 1 [Th1], Th2, and Th17 differentiation, intestinal immune network for immunoglobulin A [IgA] production, and phagosome pathways) and pathologic (graft-versus-host disease [GVHD], inflammatory bowel disease, rheumatoid arthritis, systemic lupus erythematosus, type 1 diabetes, asthma, allograft rejection, and autoimmune thyroid disease-associated pathways), were enriched in the adult clusters ([Figure 4C](#); [Table S4](#)), consistent with previous findings that fetal CD4 T cells have decreased enrichment for genes that determine other effector differentiation fates due to their bias toward  $T_{reg}$  differentiation ([Mold et al., 2008, 2010](#); [Ng et al., 2019](#)). Also consistent with previous reports that fetal naive T cells are more highly proliferative compared with adult naive T cells ([Halikias et al., 2019](#); [Michaëlsson et al., 2006](#)), the cell-cycle gene pathway was enriched within fetal-associated gene clusters ([Figure 4C](#); [Table S4](#)).

To understand at higher granularity how different pathways are enriched within the distinct stages of the fetal-to-adult transition, unbiased pathway enrichment analysis was next applied to each of the separate pairwise comparison upregulated gene sets ([Figure 5](#); [Table S4](#)). Suggestive that some tolerogenic fetal programs might persist at birth, some immune activation pathways were enriched within the adult relative to UCB as well as fetal samples (e.g., asthma, allograft rejection, autoimmune thyroid disease, and intestinal immune network for IgA production). Gene pathways upregulated only in adult samples, but not in UCB or fetal samples, likely are associated with normal developmental changes from birth to adulthood and/or represent adaptations that occur specifically in response to environmental exposures in the post-natal period. IgA production and intestinal immunity can be associated both with normal colonization by microbiota and with immune responses to mucosal infections. Thus, enrichment of the intestinal immune network for IgA production in adult relative to both fetal and UCB samples is consistent with widespread mucosal responses to microbial exposure after birth ([Dominguez-Bello et al., 2019](#); [Zhuang et al., 2019](#)).

Pathway analysis also revealed differences in several signaling pathways. In particular, multiple signaling pathways (e.g., Notch, oxytocin, phospholipase D, and relaxin signaling pathways) were exclusively enriched within UCB samples relative to adult samples, while others (e.g., Rap1, Ras, and Wnt signaling pathways) were jointly enriched within both fetal and UCB samples relative



**Figure 4. Discrete Subsets of Genes Undergo Fetal-to-Adult Transition with Varied Timing in Naïve CD4 T Cells**

Naïve CD4 T cells from fetal splenic, newborn UCB, and adult PB ( $n = 5$  samples each) were profiled by bulk RNA-seq of 50,000 cells per sample. See also Figures S3 and S5.

(A) Principal components analysis (PCA) plot, with projections into PC1 and PC2.

(B) Left: heatmap showing expression levels of all genes differentially expressed (FDR  $< 0.05$ ,  $\log_2$  fold change  $\geq 1.5$ ) between ages. Colors represent relative, Z score, log-normalized expression across each gene. Clusters obtained from k-means clustering with  $k = 5$ . Right: log-normalized expression values for each gene summarized, separately for each cluster, by mean Z score across each age.

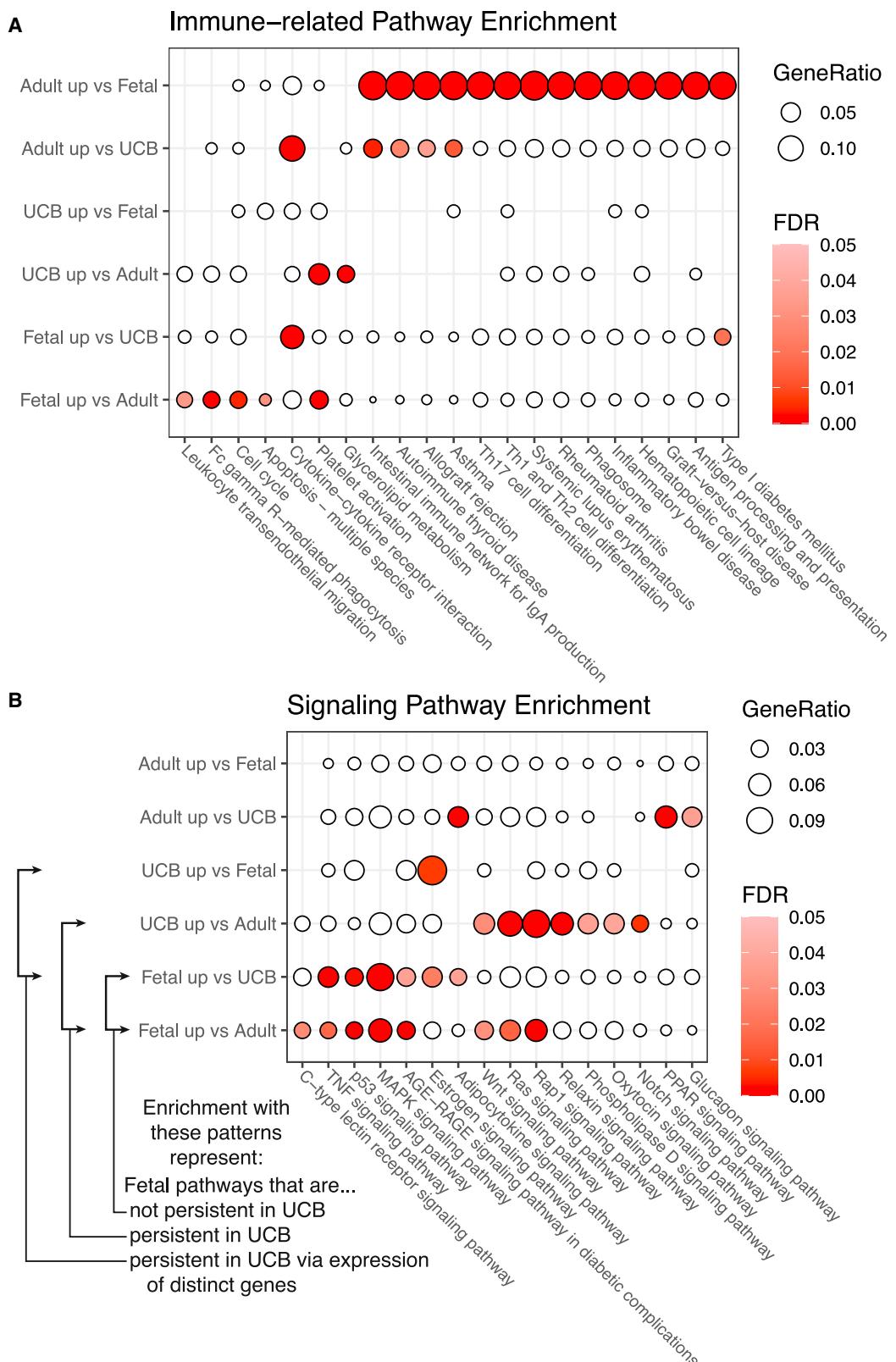
(C) Conceptual overview showing how the expression levels of genes in clusters, from the heatmap in (B), fluctuate over the course of the fetal-to-adult transition with distinct timing. Heights of bars represent relative expression level at a given time. Pathways reflect a summary of the pathways enriched within fetal-associated clusters (1, 2, and 3) and adult-associated clusters (4 and 5) genes. See also Table S4.

to adult samples (Figure 5B). Yet other signaling pathways (e.g., advanced glycation end product-receptor for advanced glycation end product [AGE-RAGE], mitogen-activated protein kinase [MAPK], p53, and tumor necrosis factor [TNF] signaling pathways) were found to be enriched within fetal samples compared with both UCB and adult samples. Interestingly, the estrogen signaling pathway stood out among these pathway enrichment comparison analyses in that it was enriched in both the “Fetal up vs UCB” and “UCB up vs fetal” gene lists, indicating that fetal and UCB naïve T cells differentially express distinct sets of genes within this single pathway. We examined all genes contributing to the estrogen signaling pathway enrichments and found that

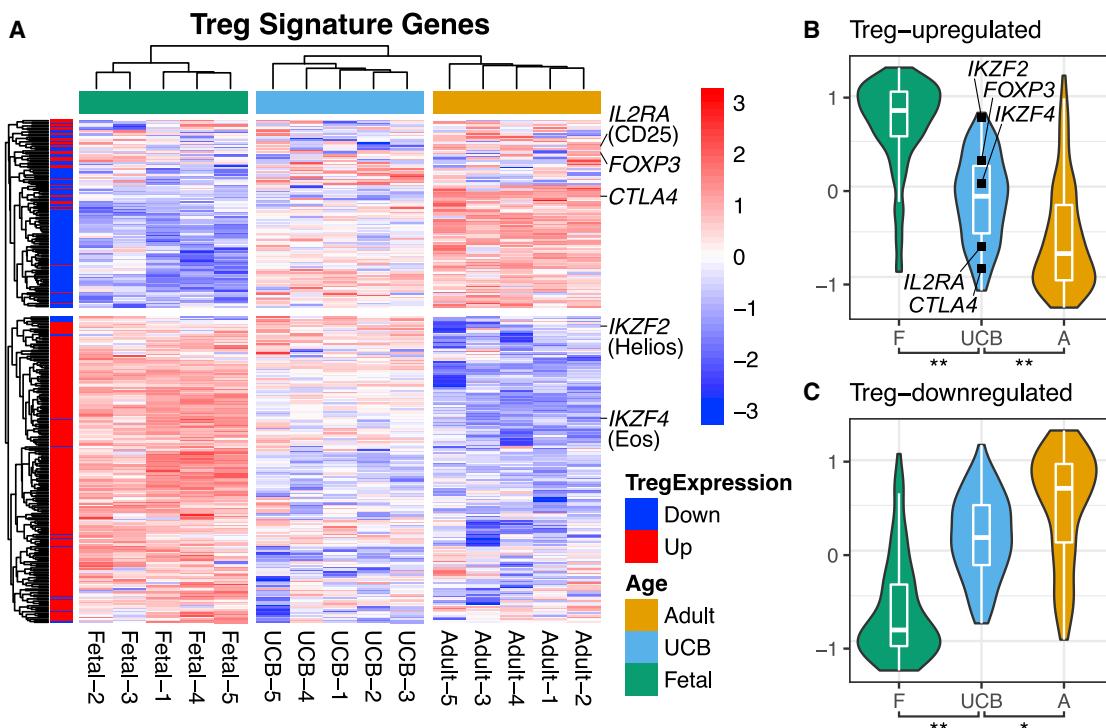
UCB samples showed higher expression of *BCL2*, *ESR1*, *GABBR1*, and *PLCB1*, while the fetal samples showed higher expression of *CREB5*, *FKBP4*, *HBEGF*, *HSPA1A*, *HSPA1B*, *HSPA2*, *HSPA6*, *HSP90AA1*, *RARA*, and *TGFA*. We conclude that, while multiple fetal gene expression pathways persist within UCB at the time of birth, others do not. We also identify multiple signaling pathways that seem to arise exclusively in UCB.

#### UCB Naïve T Cells Retain a Partial Expression of a Fetal-Associated T<sub>reg</sub> Signature

The relative enrichment of several pathways related to inflammatory immune responses in adult samples compared with both



(legend on next page)



**Figure 6. T<sub>reg</sub> Signature Gene Expression Is Partially Maintained in UCB Naive CD4 T Cells**

(A) Heatmap showing expression levels of T<sub>reg</sub> cell signature genes within the fetal, UCB, and adult naive CD4 T cells characterized by bulk RNA-seq. Gene signature derived from comparison of genes differentially expressed between adult naive T cells versus fetal and adult T<sub>regs</sub> (Ng et al., 2019). Location of selected T<sub>reg</sub>-associated genes within the heatmap are highlighted. Colors represent relative, Z score, log-normalized expression. Clusters obtained from k-means clustering with k = 2. Gene annotations: downregulated (blue) and upregulated (red) in T<sub>reg</sub> signature.

(B and C) Log-normalized expression values summarized separately for T<sub>reg</sub>-upregulated (B) and T<sub>reg</sub>-downregulated (C) genes by mean Z score expression, across each age, for each gene. Expression level of particular T<sub>reg</sub>-associated genes are highlighted within UCB samples. p values from Mann-Whitney U test, \*p < 1 × 10<sup>-7</sup> and \*\*p < 1 × 10<sup>-15</sup>.

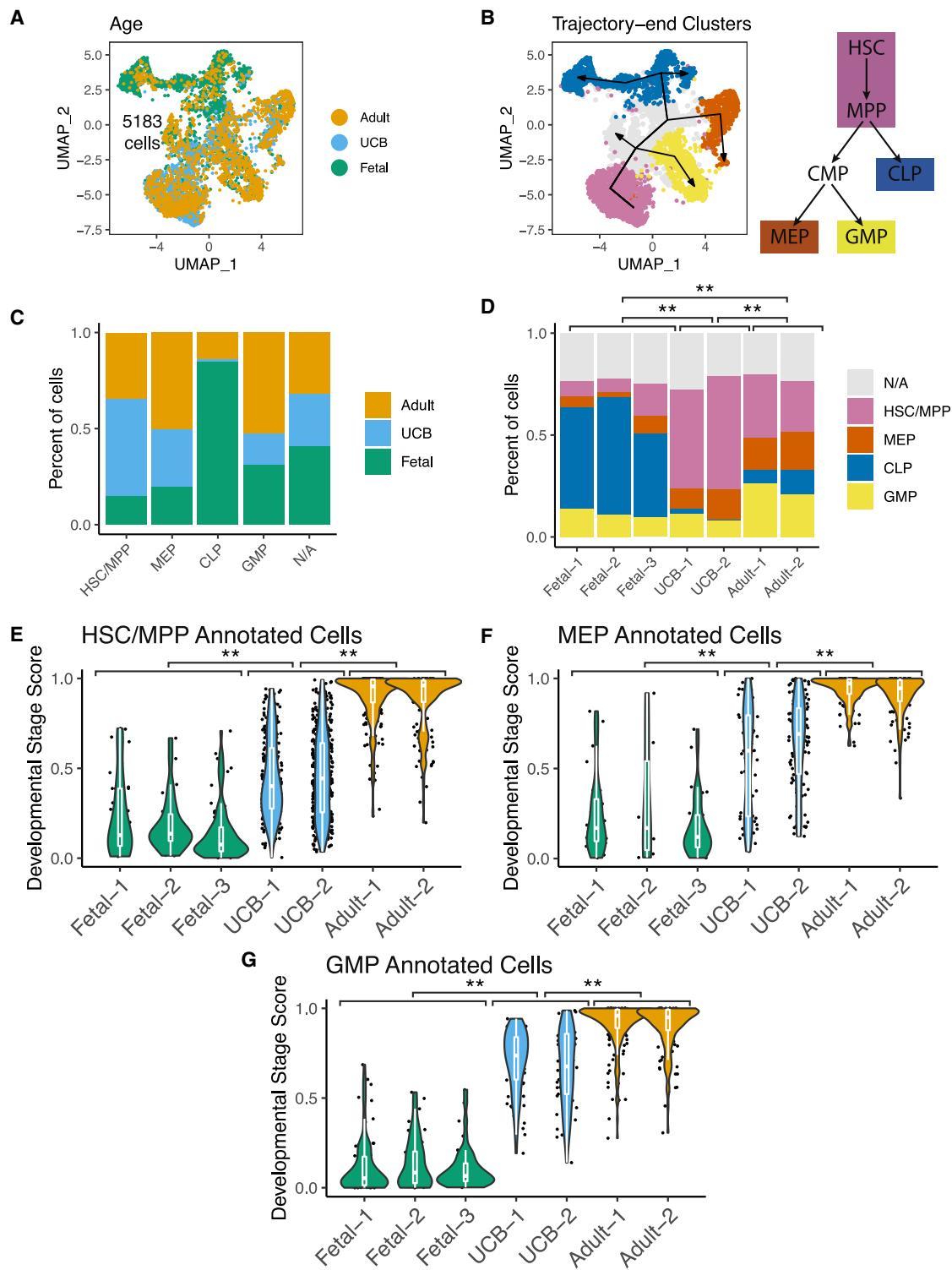
fetal and UCB samples likely reflects the long-standing observation that prenatal immunity is largely dominated by tolerogenic and/or less-inflammatory responses (Burt, 2013; Cupedo et al., 2005; Darrasse-Jéze et al., 2005; Krow-Lucal et al., 2014; Michaësson et al., 2006; Mold et al., 2008, 2010). Fetal naive CD4 T cells have elevated expression of *IKZF2* (Helios) and other T<sub>reg</sub>-associated genes, which are thought to contribute to their T<sub>reg</sub>-differentiation predisposition (Ng et al., 2019). Thus, we next determined whether elevated expression of a published T<sub>reg</sub>-associated transcriptome (Ng et al., 2019) might be retained in UCB cells. Overall, UCB naive T cells expressed T<sub>reg</sub>-upregulated genes more highly, including *IKZF2* and *FOXP3*, and expressed T<sub>reg</sub>-downregulated genes less highly than adult naive T cells (Figure 6). Of note, UCB naive T cells were intermediate in their expression of the overall T<sub>reg</sub>-associated gene signature between fetal and adult samples, suggesting that they have shut down parts of the program predisposing them toward T<sub>reg</sub> differentiation.

#### The Transcriptomes of CD34<sup>+</sup> HSPCs in UCB Are Intermediate between Fetal and Adult HSPCs

In previous work, we demonstrated that human fetal CD34<sup>+</sup> HSPCs give rise to CD4 T cells that are transcriptionally similar to primary fetal T cells and are also predisposed toward T<sub>reg</sub> differentiation (Mold et al., 2010). It has also been reported that both human embryonic stem cells and fetal CD34<sup>+</sup> liver HSPCs, but not CD34<sup>+</sup> HSPCs from peripheral blood of adult donors, differentiate *in vitro* into monocytic cells with functional and gene expression profiles that are characteristic of anti-inflammatory M2-type macrophages (Klimchenko et al., 2011). Given the finding that individual UCB naive T cells express an intermediate transcriptional phenotype, we sought to understand whether CD34<sup>+</sup> HSPCs in UCB might also have a transcriptional profile that is intermediate between fetal and adult HSPCs. The answer to this question may also have significant clinical implications as UCB is commonly used as a source of hematopoietic progenitors for therapeutic transplantation in the settings of

**Figure 5. Distinct Immune-Related and Signaling Pathways Are Enriched within Fetal, UCB, and Adult Naive CD4 T Cells**

(A and B) KEGG Pathway enrichment assessed within naive CD4 T cell bulk RNA-seq differential expression gene sets. Enriched pathways are represented by FDR-corrected p values (color) and the percentage of genes of a given set that are part of the given pathway (GeneRatio, size). Selected non-infection immune-related (A) and signaling (B) pathways are shown. See also Table S4 for the full set of enriched pathways.



**Figure 7. Single-Cell Developmental Stage Scoring Places UCB HSPCs Intermediate between Fetal and Adult**

CD34<sup>+</sup> HSPCs from fetal bone marrow (BM) ( $n = 3$ ), full-term UCB ( $n = 2$ ), and adult BM ( $n = 2$ ) samples were profiled by single-cell RNA-seq. Fetal-to-adult transition phenotype of cells were developmental stage scored, separately for each annotated cell type. See also Figures S3 and S6.

(A) UMAP plot showing cells of each age after batch correction

(legend continued on next page)

both primary immunodeficiency and cancer. Thus, given that the marker CD34 is present on multiple early stages in the developmental ontogeny of human HSPCs and that the entire pool of CD34<sup>+</sup> cells is regularly used in clinical transplantation protocols, we included all CD34<sup>+</sup> cells in our sequencing approach.

CD34<sup>+</sup> HSPCs were sort purified from fetal BM, full-term UCB, and adult BM (Figures 1B and S3B). Sorted cells from seven samples (three fetal BM, two full-term UCB, and two adult BM) were pooled for sequencing, and then Demuxlet was used to annotate the sample identity of each cell and to identify and remove doublets. We obtained 5,183 single-cell transcriptomes, representing 747 median (range 602–898) cells per sample. We then used a combination of bioinformatic techniques to identify cells in similar differentiation states (Figures 7A and 7B; see STAR Methods for details). This approach annotated 1,408 cells putatively in the earliest stages of hematopoiesis (i.e., HSCs and multipotent progenitor cells [MMPs], HSC/MPP cells) and 2,583 progenitors of three distinct branches of hematopoiesis: 591 putative megakaryocyte-erythroid progenitors (MEPs); 789 granulocyte-monocyte progenitors (GMPs); and 1,203 common lymphoid progenitors (CLPs). Supportive of these cell-type assignments, expression patterns of genes canonically associated with these cell types match our annotations (Figure S6). Notably, we found that cell type annotation frequencies varied between ages and tissues (Figures 7C and 7D; chi-square p value < 2.2e-16 for all pairwise age comparisons). In particular, only 19 of 1,203 cells annotated as CLPs originated from UCB samples, and more than twice as many HSC/MPP-annotated cells originated from UCB (n = 2) or adult BM (n = 2) samples as were annotated from fetal BM samples (n = 3). These data provide evidence that the abundance of specific developmental intermediates is different between fetal BM, UCB, and adult BM samples and specifically that CLPs may be of particularly low abundance in UCB.

To compare the transcriptomes of cells across ages, separate developmental stage scoring models were built for each HSPC subtype with sufficient numbers of UCB cells (i.e., HSC/MPP, MEP, GMP). Accuracy of the model for each cell type was confirmed to be high for non-transitioned (i.e., fetal) and fully transitioned (i.e., adult) cells that were not used in training the models (above 0.99 for all cell types: HSC, 0.9901; MEP, 0.9691; GMP, 0.9986). Finally, scores for UCB cells, of all cell types with substantial numbers of cells recovered in UCB, were found to be intermediate between those of fetal and adult (Figures 7E–7G). The intermediate scoring of progenitor cells indicates that HSPCs in full-term UCB are at an intermediate state along the fetal-to-adult transition and are thus transcriptionally unique from fetal and adult HSPCs.

## DISCUSSION

It is clear that several uniquely fetal immune populations can be generated only by progenitors that are present during fetal development, and the interpretation of earlier experiments in mice, birds, and humans favored the possibility that ontogeny of the hematopoietic system is regulated by distinct layered waves of unrelated HSCs and their progeny (Hadland and Yoshimoto, 2018; Herzenberg and Herzenberg, 1989; Ikuta et al., 1990). Here, we have studied the gene expression programs of monocytes, T cells, and HSPCs at varying time points during human ontogeny—fetal (18–23 weeks GA), newborn (34–42 weeks GA), infant (0–7 months), and adult (27–53 years)—to characterize the progress of fetal-to-adult immune transition at the time of birth. While the present study is not the first to use a system that assigns an immunological age score to human samples—Alpert et al. recently developed an immune aging (IMMAGE) score based on immune cell frequencies in adults (20–96 years) (Alpert et al., 2019)—here, we developed cell-type-specific models of samples' and cells' age-associated or developmental-associated differences (termed “developmental stage score”) for directed application to sorted cell populations. Through single-cell developmental stage score characterization of naive T cells and HSPCs, we find evidence that human immune ontogeny follows a progressive, and not a layered, pattern of transition during late fetal development.

Using this approach, we demonstrate that individual UCB T cells received unimodally distributed intermediate developmental stage scores and were not composed of a mixture of cells with fetal or adult transcriptional identities. It is important to note that, while most developmental stage score markers denoting extremes of fetal or adult identity are intermediate in expression at the single-cell level, some are skewed toward fetal (e.g., RPS24) or adult (e.g., HSP90AA1) expression patterns (Figure S4C). As these findings occur with the same pattern in the majority of cells analyzed, such a phenotype cannot be explained by a mixture of fetal-layer and adult-layer cells and instead indicate a transition that includes progression of distinct genetic programs, each with different timing. Furthermore, variation between samples was noted as a shift in the modes of UCB samples' per-cell scores, rather than via differences in the abundance of cells with particular ranges of scores—such as very low scoring (i.e., fetal-like) or very high scoring (i.e., adult-like) cells—as we would have otherwise expected if the fetal-to-adult transition occurred via layering of distinct cell populations. It should be noted that the proportion of maternally derived cells in UCB has been shown to be fewer than 0.4% (Mold et al., 2008; Opstelten et al., 2019); thus, it is unlikely that maternal cells contributed significantly to these analyses. Although these data represent a “snapshot” in time

(B) Left: UMAP plot showing final cell type annotations with inferred cluster-based differentiation trajectories overlaid on top. Right: tree diagram showing accepted hematopoietic differentiation trajectories leading to all cell types annotated in this dataset. CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; HSC, hematopoietic stem cell; MEP, megakaryocyte-erythroid progenitor; MPP, multipotent progenitor.  
 (C) Percentages of age identities of cells annotated as each cell type. N/A refers to cells not annotated in trajectory-end cell type clusters.  
 (D) Percentages of cell type annotations per sample. p values from chi-square test, \*\*p < 1 × 10<sup>-15</sup>.  
 (E–G) Developmental stage scoring of individual HSC/MPP (E) MEP (F), and GMP (G) cells for each sample. Points are individual cells. Fetal and adult cells used to build developmental stage score models are not included. p values from Mann-Whitney U test, \*\*p < 1 × 10<sup>-15</sup>.

along the course of development, they suggest that UCB cells are captured at a single point along a relatively uniform and progressive fetal-to-adult transition. Future analyses including additional intermediate GAs could potentially clarify whether changes truly occur progressively between the time points we have demonstrated here or, conversely, via multiple stepwise transitions between dominant fetal and newborn and then newborn and adult cell populations.

The findings presented here demonstrate that UCB naive T cells represent a transitional population with transcriptional features that are both distinct from, and intermediate between, fetal and adult naive T cells. Thus, the unique nature of immune responses in neonates and young children, including blunted responses to infection and vaccination, may be related to the transcriptional and functional state of these, and possibly other, immune cells. Our bulk developmental stage scoring model additionally revealed that there is a high degree of inter-individual variability at birth, and that variability persists even while developmental stage scores become more adult-like during the first several months of life. Such findings are consistent with clinical observations that several years are required for the acquisition of robust, mature immune responses (Dowling and Levy, 2014) and suggest that certain characteristics of the fetal immune system may persist for a considerable period of time after birth. It is well established that some infants produce less effective responses to vaccination or are more susceptible to infection than others, even in the absence of an identifiable primary immunodeficiency (Borghesi et al., 2017; Newport et al., 2004; Siegrist, 2001). Our findings may suggest that these infants are born with a more fetal-like immune system. Assessing this possibility would require using developmental stage scores as a tool to stratify infants in studies of immune responses, including vaccine responses, in early life. Of note, the UCB samples studied here were obtained from clinical cohorts in which exhaustive demographic data collection and laboratory screening were not always available. Therefore, the observed inter-individual variability in gene expression may also be influenced by uncontrolled factors, including environmental exposures, maternal medication use, and undiagnosed infection.

Further mechanistic studies to understand neonatal transcriptional programs will not only provide insight into early-life immune responses but also may ultimately suggest druggable targets to enhance immune maturation and shift neonatal T cells toward adult-like protective immune responses. Here, unbiased pathway analysis demonstrated that UCB T cells are enriched for both Wnt and Notch signaling pathways compared with adult cells. Notch signaling plays a pivotal regulatory role in multiple aspects of T cell development and lineage differentiation (Vijayaraghavan and Osborne, 2018). The Wnt signaling pathway has been shown to promote differentiation of T follicular helper ( $T_{fh}$ ) cells, which are crucial for effective vaccine responses (van Loosdregt and Coffer, 2018), and pharmacological inhibition of the Wnt/ $\beta$ -catenin signaling pathway modulates fate decisions between self-renewal and differentiation of long-lived central memory and stem cell-like memory CD4 T cells (Mavigner et al., 2019).

In several trials, UCB transplantation has been shown to result in overall lower rates of GVHD compared with adult BM trans-

plantation (Ballen et al., 2013; Merindol et al., 2011). Direct evidence of a salutary influence of  $T_{regs}$  in HSPC transplantation is clearly demonstrated in recent clinical trials where co-transplantation of UCB-derived  $T_{regs}$  resulted in a marked reduction in GVHD (Blazar et al., 2018; Elias and Rudensky, 2019; McKenna et al., 2017). In light of findings presented here that UCB T cells retain a partial  $T_{reg}$ -associated transcriptome (Figure 6), that variable bulk transcriptional signatures were detected in two distinct hematopoietic lineages (lymphoid and myeloid; Figure 2), and that fetal HSPCs give rise to T cells with transcriptional and functional characteristics of fetal T cells (Mold et al., 2010), we analyzed CD34 $^{+}$  HSPC cell populations with cell-type-specific single-cell transcriptional developmental stage score analysis to determine whether progenitor cell populations also undergo gradual progressive transition from fetal to adult identity. We found that the developmental age signatures of HSC/MPP, MEP, and GMP cells from UCB were distributed between fetal and adult cells of the same differentiation stage, suggesting that the intermediate nature of T cell progeny may be reflective of (and potentially programmed by) the intermediate development stage of the HSPCs from which they arise. Along these lines, a recent study comparing naive T cells recovered 2 months after T-replete UCB transplant (UCB-T) versus T-replete BM transplant (BM-T), to naive T cells from control fetal spleen, newborn UCB, and post-natal (7- to 40-year-old) peripheral blood donors found that T cells arising from UCB-T were more similar to those of control UCB, while those arising from BM-T were more similar to those of post-natal peripheral blood (Hiwarkar et al., 2017). It is unknown whether such results were reflective of new T cells arising from transplanted hematopoietic progenitors versus expansion of remnant donor T cells. In either case, the retention of newborn-like and adult-like phenotypes after clinical transplantation into varied recipient hosts demonstrates the crucial influence of cell-intrinsic programs determined by the developmental stage of the donor cells.

We found that the frequency of HSPC developmental lineage intermediates was variable between fetal BM, newborn UCB, and adult BM samples (Figures 7C and 7D). In particular, very few CLP cells were annotated within UCB samples. The relative dearth of intermediate progenitors that we observe in UCB compared with adult BM may provide an explanation for clinical observations of slower engraftment of T cells after UCB transplantation compared with adult BM-Ts (Komanduri et al., 2007; Servais et al., 2017). Our data suggest that, compared with adult BM, UCB has a smaller population of CLP precursors that can migrate directly to the thymus. In that case, generating a robust population of nascent donor-derived T cells from UCB transplants would be delayed by the need for HSC/MPP cells to transit through the CLP stage prior to initiating thymopoiesis.

Although, overall, our data are consistent with a model of gradual, progressive change from fetal to adult immunity at the level of individual cells around the time of birth, multiple limitations associated with human immunological research must be acknowledged. The expected variability inherent in human studies, as revealed in our survey of a healthy full-term birth cohort, raises the concern that a limited number of UCB transcriptomes ( $n = 5$  for T cells,  $n = 2$  for HSPCs) may not be reflective of the greater human population. Although this concern is

considerably mitigated by the consistent intermediacy of UCB samples between fetal and adult samples, it underscores the need to extend these findings to a larger and more diverse population of human subjects in the future. It must be noted that tissue-, environment-, diet-, and hormonal milieu-associated differences unavoidably exist between samples and that such differences would be aggregated, inextricably, alongside age-based differences in our transcriptional data. Because of practical considerations, we primarily studied T cells isolated from spleen, UCB, and peripheral blood as representative of fetal, newborn, and adult time points, respectively. Yet, we were able to compare gene expression in these tissues using a signature derived from microarray profiling of fetal peripheral (umbilical cord) blood to adult peripheral blood (Table S1) and confirmed that expression patterns remained consistent for these genes in RNA-seq comparison of fetal spleen versus adult peripheral blood, and that newborn UCB T cells exhibit an intermediate expression profile for these genes as well (Figures S5B–S5D). Whereas UCB has often been used for assessment of newborn immune cell phenotypes, it has been shown recently that immune system phenotypes can change drastically between those measured in UCB and 1 week later in peripheral blood (Olin et al., 2018). Although the overall increase in mean bulk developmental stage scoring of at-birth UCB versus 2-week-old peripheral blood naive T cells was about the same as the difference between 2- versus 4-week-old peripheral blood (Figure 2F), it will be important to follow up in newborn peripheral blood samples to determine whether UCB-specific signaling pathways remain expressed in the days and weeks after birth, and whether gene expression is altered after delivery by Caesarean section in the absence of labor.

In conclusion, we provide evidence at the level of single cells that, in humans, T cell populations transition along a continuum that is characterized by progressive downregulation of fetal genes and upregulation of adult genes during the course of ontogeny from midgestational fetal development to birth and into adulthood. We find that the transcriptomes of newborn T cells and hematopoietic progenitors possess features that are unique from, and intermediate between, those of their fetal and adult counterparts, and we identify particular pathways enriched in UCB T cells that warrant further study. While our findings do not rule out the concept of “layered” hematopoiesis as invoked in other species (Havran and Allison, 1988; Hayakawa et al., 1985; Herzenberg and Herzenberg, 1989; Ikuta et al., 1990; Jotereau and Le Douarin, 1982; Kantor et al., 1992; Lalor et al., 1989; Le Douarin and Jotereau, 1975; Montecino-Rodríguez et al., 2006, 2016, 2018; Ramond et al., 2014), and at earlier time points in humans (Ginhoux et al., 2013; Hadland and Yoshimoto, 2018; Montecino-Rodríguez and Dorshkind, 2012; Stamatoyannopoulos, 2005; Tieppo et al., 2020), they demonstrate that layering of distinct fetal-like and adult-like cell populations is not present at the time of birth in human  $\alpha\beta$  T cells. Further study of how the unique functional properties of UCB HSPCs may affect the outcomes of hematopoietic transplants is clearly needed. Additionally, further study of the unique properties of immune cells of all types in the newborn and young infant will clarify whether or not a layered fetal-to-adult transition underlies the ontogeny of other immune cells. Of particular interest will be

studies to determine whether the inter-individual variation in the progress of fetal-to-adult transition by the time of birth, noted here in both naive T cells and monocytes, has a significant effect on the nature of immune responses to vaccines and infections in early life. Such studies may lead to more focused stratification of children according to their individual risk for infection and for inadequate vaccine responses. They may also lead to the design of precision therapeutic approaches to boost newborn immunity and to improved public health outcomes in this vulnerable population.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Tissue collection
- **METHOD DETAILS**
  - Cell isolation
  - Fluorescence activated cell sorting (FACS) for microarray and Fluidigm qRT-PCR
  - RNA preparation for microarray analysis
  - Statistical analysis of microarrays
  - Gene signature derivation for bulk developmental stage score
  - Fluidigm qRT-PCR of signature genes
  - Statistical analysis of Fluidigm qPCR for comparison of fetal, newborn, and adult samples
  - Statistical analysis of Fluidigm qPCR for large, URECA, birth cohort
  - Cell enrichment and FACS for RNA sequencing
  - Preparation for single cell RNA-seq library generation
  - Raw sequencing data pre-processing
  - Dimensionality reduction analysis and clustering of RNA-seq data
  - Differential expression and pathway analysis of RNA-seq data
  - Cell type annotation in the hematopoietic stem and progenitor cell single-cell dataset
  - Developmental stage score generation with machine learning through random forest regression
  - Visualization of RNA sequencing data
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108573>.

## ACKNOWLEDGMENTS

Thanks to the NIH/NIAID Infant Immunity Program and Mercy PrabhuDas for critical support of this project and to Melissa Ng; Elze Rackaityte; Joanna

Halkias; Dmitry Rychkov; members of the Burt, McCune, and Sirota labs; Dror Assa; Rebecca Jaszczak; Prescott Woodruff; Jeffrey Mold; Jakob Michaelsson; Susan Fisher; Tippi MacKenzie; Michaela Frascoli; and Cheryl Stoddart for helpful discussion. Thanks to Ekatarina Maidji, Alexander Olson, and Joseph Hiatt for experimental support. Thanks to the UCSF Parnassus Flow Cytometry Core for flow cytometry assistance, the Institute for Human Genetics at UCSF for single-cell next-generation sequencing (NGS) preparation, and the UCSF Center for Advanced Technology for NGS sequencing. We thank all cell donors and clinical staff for their participation. J.M.M. was supported by NIH awards R21 AI094009, R01 AI102951, U01 AI43641, and R37 AI40312 and is the recipient of the NIH Director's Pioneer Award, part of the NIH Roadmap for Medical Research, through grant DPI OD00329. T.D.B. was supported by an award from the Burroughs Wellcome Fund Preterm Birth Initiative and by NIH awards K08 HD067295 and R21 AI120032. Y.B. was supported by T32 AI007334-25. Y.J.B. was supported by R01 HD080474. M.S. was supported by Burroughs Wellcome Fund, PREMIER, a NIH/NIAMS P30 Center for the Advancement of Precision Medicine in Rheumatology at UCSF (P30AR070155), NLM/NIH K01LM012381. Urban Environment and Childhood Asthma (URECA) cord blood sample collection and analysis were supported by UM1 AI114271.

#### AUTHOR CONTRIBUTIONS

D.G.B., Y.B., E.K.-L., M.S., T.D.B., and J.M.M. designed the study and analyzed the data. D.G.B., Y.B., E.K.-L., V.F.M., and N.J. performed the experiments. M.S. advised on the computational analysis; T.D.B. and J.M.M. advised on experiments. N.R.B. and S.N. provided critical bioinformatic assistance. C.C.K. assisted with microarray experimental workflow and data analysis. C.J.Y. contributed to Demuxlet analysis. R.L.R. contributed to data analysis of URECA cohort samples. J.E.G. and Y.J.B. provided technical counseling and samples. D.G.B., Y.B., E.K.-L., M.S., T.D.B., and J.M.M. wrote the paper. All authors reviewed and commented on the manuscript.

#### DECLARATION OF INTERESTS

C.J.Y. declares partnership at Related Sciences and TRex Bio and owns stock in DroPrint Genomics and Related Sciences. C.J.Y. is supported by research grants from CZ Biohub, CZI, and PICL.

Received: March 18, 2020

Revised: June 1, 2020

Accepted: December 8, 2020

Published: January 5, 2021

#### REFERENCES

- Alpert, A., Pickman, Y., Leipold, M., Rosenberg-Hasson, Y., Ji, X., Gaujoux, R., Rabani, H., Starovetsky, E., Kveller, K., Schaffert, S., et al. (2019). A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring. *Nat. Med.* **25**, 487–495.
- Andrews, S. (2018). FastQC: a quality control tool for high throughput sequence data.
- Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172.
- Ballen, K.K., Gluckman, E., and Broxmeyer, H.E. (2013). Umbilical cord blood transplantation: the first 25 years and beyond. *Blood* **122**, 491–498.
- Blanchard, A.C., Quach, C., and Autmizguine, J. (2015). Staphylococcal infections in infants: updates and current challenges. *Clin. Perinatol.* **42**, 119–132, ix.
- Blazar, B.R., MacDonald, K.P.A., and Hill, G.R. (2018). Immune regulatory cell infusion for graft-versus-host disease prevention and therapy. *Blood* **131**, 2651–2660.
- Borghesi, A., Stronati, M., and Fellay, J. (2017). Neonatal Group B Streptococcal Disease in Otherwise Healthy Infants: Failure of Specific Neonatal Immune Responses. *Front. Immunol.* **8**, 215.
- Bronevetsky, Y., Burt, T.D., and McCune, J.M. (2016). Lin28b Regulates Fetal Regulatory T Cell Differentiation through Modulation of TGF- $\beta$  Signaling. *J. Immunol.* **197**, 4344–4350.
- Bunis, D.G., Andrews, J., Fragiadakis, G.K., Burt, T.D., and Sirota-DittoSeq, M. (2020). Universal User-Friendly Single-Cell and Bulk RNA Sequencing Visualization Toolkit. *Bioinformatics* **1011**. <https://doi.org/10.1093/bioinformatics/btaa1011>.
- Burt, T.D. (2013). Fetal regulatory T cells and peripheral immune tolerance in utero: implications for development and disease. *Am. J. Reprod. Immunol.* **69**, 346–358.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420.
- Chicco, D. (2017). Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, 35.
- Cupedo, T., Nagasawa, M., Weijer, K., Blom, B., and Spits, H. (2005). Development and activation of regulatory T cells in the human fetus. *Eur. J. Immunol.* **35**, 383–390.
- Darrasse-Jeze, G., Marodon, G., Salomon, B.L., Catala, M., and Klatzmann, D. (2005). Ontogeny of CD4+CD25+ regulatory/suppressor T cells in human fetuses. *Blood* **105**, 4715–4721.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
- Dominguez-Bello, M.G., Godoy-Vitorino, F., Knight, R., and Blaser, M.J. (2019). Role of the microbiome in human development. *Gut* **68**, 1108–1114.
- Dowling, D.J., and Levy, O. (2014). Ontogeny of early life immunity. *Trends Immunol.* **35**, 299–310.
- Elias, S., and Rudensky, A.Y. (2019). Therapeutic use of regulatory T cells for graft-versus-host disease. *Br. J. Haematol.* **187**, 25–38.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278.
- Gern, J.E., Visness, C.M., Gergen, P.J., Wood, R.A., Bloomberg, G.R., O'Connor, G.T., Kattan, M., Sampson, H.A., Witter, F.R., Sandel, M.T., et al. (2009). The Urban Environment and Childhood Asthma (URECA) birth cohort study: design, methods, and study population. *BMC Pulm. Med.* **9**, 17.
- Ginhoux, F., Lim, S., Hoeffel, G., Low, D., and Huber, T. (2013). Origin and differentiation of microglia. *Front. Cell. Neurosci.* **7**, 45.
- Hadland, B., and Yoshimoto, M. (2018). Many layers of embryonic hematopoiesis: new insights into B-cell ontogeny and the origin of hematopoietic stem cells. *Exp. Hematol.* **60**, 1–9.
- Halkias, J., Rackaityte, E., Hillman, S.L., Aran, D., Mendoza, V.F., Marshall, L.R., MacKenzie, T.C., and Burt, T.D. (2019). CD161 contributes to prenatal immune suppression of IFN $\gamma$ -producing PLZF $^+$  T cells. *J. Clin. Invest.* **129**, 3562–3577.
- Havran, W.L., and Allison, J.P. (1988). Developmentally ordered appearance of thymocytes expressing different T-cell antigen receptors. *Nature* **335**, 443–445.
- Hayakawa, K., Hardy, R.R., Herzenberg, L.A., and Herzenberg, L.A. (1985). Progenitors for Ly-1 B cells are distinct from progenitors for other B cells. *J. Exp. Med.* **161**, 1554–1568.
- Herzenberg, L.A., and Herzenberg, L.A. (1989). Toward a layered immune system. *Cell* **59**, 953–954.
- Hiwarkar, P., Hubank, M., Qasim, W., Chiesa, R., Gilmour, K.C., Saudemont, A., Amrolia, P.J., and Veys, P. (2017). Cord blood transplantation recapitulates

- fetal ontogeny with a distinct molecular signature that supports CD4<sup>+</sup> T-cell reconstitution. *Blood Adv.* 1, 2206–2216.
- Ikuta, K., Kina, T., MacNeil, I., Uchida, N., Peault, B., Chien, Y.H., and Weissman, I.L. (1990). A developmental switch in thymic lymphocyte maturation potential occurs at the level of hematopoietic stem cells. *Cell* 62, 863–874.
- Jotereau, F.V., and Le Douarin, N.M. (1982). Demonstration of a cyclic renewal of the lymphocyte precursor cells in the quail thymus during embryonic and perinatal life. *J. Immunol.* 129, 1869–1877.
- Kang, H.M., Subramaniam, M., Targ, S., Nguyen, M., Maliskova, L., McCarthy, E., Wan, E., Wong, S., Byrnes, L., Lanata, C.M., et al. (2018). Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* 36, 89–94.
- Kantor, A.B., Stall, A.M., Adams, S., Herzenberg, L.A., and Herzenberg, L.A. (1992). Differential development of progenitor activity for three B-cell lineages. *Proc. Natl. Acad. Sci. USA* 89, 3320–3324.
- Klimchenko, O., Di Stefano, A., Geoerger, B., Hamidi, S., Opolon, P., Robert, T., Routhier, M., El-Benna, J., Delezoide, A.-L., Boukour, S., et al. (2011). Monocytic cells derived from human embryonic stem cells and fetal liver share common differentiation pathways and homeostatic functions. *Blood* 117, 3065–3075.
- Komanduri, K.V., St John, L.S., de Lima, M., McMannis, J., Rosinski, S., McNiece, I., Bryan, S.G., Kaur, I., Martin, S., Wieder, E.D., et al. (2007). Delayed immune reconstitution after cord blood transplantation is characterized by impaired thymopoiesis and late memory T-cell skewing. *Blood* 110, 4543–4551.
- Krow-Lucal, E.R., Kim, C.C., Burt, T.D., and McCune, J.M. (2014). Distinct functional programming of human fetal and adult monocytes. *Blood* 123, 1897–1904.
- Krueger, F. (2019). Trim Galore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files (Babraham Bioinformatics).
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* 28, 1–26.
- LaIor, P.A., Stall, A.M., Adams, S., and Herzenberg, L.A. (1989). Permanent alteration of the murine Ly-1 B repertoire due to selective depletion of Ly-1 B cells in neonatal animals. *Eur. J. Immunol.* 19, 501–506.
- Larsson, J. (2019). eulerr: Area-Proportional Euler and Venn Diagrams with Ellipses. R package version 6.1.0. <https://cran.r-project.org/package=eulerr>.
- Le Douarin, N.M., and Jotereau, F.V. (1975). Tracing of cells of the avian thymus through embryonic life in interspecific chimeras. *J. Exp. Med.* 142, 17–40.
- Li, N., van Unen, V., Abdelaal, T., Guo, N., Kasatskaya, S.A., Ladell, K., McLaren, J.E., Egorov, E.S., Izraelson, M., Chuva de Sousa Lopes, S.M., et al. (2019). Memory CD4<sup>+</sup> T cells are generated in the human fetal intestine. *Nat. Immunol.* 20, 301–312.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- 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.
- Martens, J.H.A., and Stunnenberg, H.G. (2013). BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* 98, 1487–1489.
- Mavigner, M., Zanoni, M., Tharp, G.K., Habib, J., Mattingly, C.R., Lichterfeld, M., Nega, M.T., Vanderford, T.H., Bosinger, S.E., and Chahroudi, A. (2019). Pharmacological modulation of the Wnt/β-catenin pathway inhibits proliferation and promotes differentiation of long-lived memory CD4<sup>+</sup> T-cells in ART-suppressed SIV-infected macaques. *J. Virol.* 94, e01094–e010919.
- McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). tSNE: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* 3, 861.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303.
- McKenna, D.H., Jr., Sumstad, D., Kadidlo, D.M., Batdorf, B., Lord, C.J., Merkel, S.C., Koellner, C.M., Curtsinger, J.M., June, C.H., Riley, J.L., et al. (2017). Optimization of cGMP purification and expansion of umbilical cord blood-derived T-regulatory cells in support of first-in-human clinical trials. *Cytotherapy* 19, 250–262.
- Merindol, N., Charrier, E., Duval, M., and Soudeyns, H. (2011). Complementary and contrasting roles of NK cells and T cells in pediatric umbilical cord blood transplantation. *J. Leukoc. Biol.* 90, 49–60.
- Michaëlsson, J., Mold, J.E., McCune, J.M., and Nixon, D.F. (2006). Regulation of T cell responses in the developing human fetus. *J. Immunol.* 176, 5741–5748.
- Mold, J.E., Michaëlsson, J., Burt, T.D., Muench, M.O., Beckerman, K.P., Busch, M.P., Lee, T.-H., Nixon, D.F., and McCune, J.M. (2008). Maternal allo-antigens promote the development of tolerogenic fetal regulatory T cells in utero. *Science* 322, 1562–1565.
- Mold, J.E., Venkatasubrahmanyam, S., Burt, T.D., Michaëlsson, J., Rivera, J.M., Galkina, S.A., Weinberg, K., Stoddart, C.A., and McCune, J.M. (2010). Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. *Science* 330, 1695–1699.
- Mold, J.E., Réu, P., Olin, A., Bernard, S., Michaëlsson, J., Rane, S., Yates, A., Khosravi, A., Salehpour, M., Possnert, G., et al. (2019). Cell generation dynamics underlying naive T-cell homeostasis in adult humans. *PLoS Biol.* 17, e3000383.
- Montecino-Rodriguez, E., and Dorshkind, K. (2012). B-1 B cell development in the fetus and adult. *Immunity* 36, 13–21.
- Montecino-Rodriguez, E., Leathers, H., and Dorshkind, K. (2006). Identification of a B-1 B cell-specified progenitor. *Nat. Immunol.* 7, 293–301.
- Montecino-Rodriguez, E., Fice, M., Casero, D., Berent-Maoz, B., Barber, C.L., and Dorshkind, K. (2016). Distinct Genetic Networks Orchestrate the Emergence of Specific Waves of Fetal and Adult B-1 and B-2 Development. *Immunity* 45, 527–539.
- Montecino-Rodriguez, E., Casero, D., Fice, M., Le, J., and Dorshkind, K. (2018). Differential Expression of PU.1 and Key T Lineage Transcription Factors Distinguishes Fetal and Adult T Cell Development. *J. Immunol.* 200, 2046–2056.
- Newport, M.J., Goetghebuer, T., Weiss, H.A., Whittle, H., Siegrist, C.-A., and Marchant, A.; MRC Gambia Twin Study Group (2004). Genetic regulation of immune responses to vaccines in early life. *Genes Immun.* 5, 122–129.
- Ng, M.S.F., Roth, T.L., Mendoza, V.F., Marson, A., and Burt, T.D. (2019). Helios enhances the preferential differentiation of human fetal CD4<sup>+</sup> naïve T cells into regulatory T cells. *Sci. Immunol.* 4, eaav5947.
- Olin, A., Henckel, E., Chen, Y., Lakshminanth, T., Pou, C., Mikes, J., Gustafsson, A., Bernhardsson, A.K., Zhang, C., Bohlin, K., and Brodin, P. (2018). Stereotypic Immune System Development in Newborn Children. *Cell* 174, 1277–1292.e14.
- Opstelten, R., Slot, M.C., Lardy, N.M., Lankester, A.C., Mulder, A., Claas, F.H.J., van Rood, J.J., and Amsen, D. (2019). Determining the extent of maternal-foetal chimerism in cord blood. *Sci. Rep.* 9, 5247.
- Perez-Riverol, Y., Kuhn, M., Vizcaíno, J.A., Hitz, M.-P., and Audain, E. (2017). Accurate and fast feature selection workflow for high-dimensional omics data. *PLoS ONE* 12, e0189875.
- Phares, C.R., Lynfield, R., Farley, M.M., Mohle-Boetani, J., Harrison, L.H., Petit, S., Craig, A.S., Schaffner, W., Zansky, S.M., Gershman, K., et al.; Active Bacterial Core surveillance/Emerging Infections Program Network (2008). Epidemiology of invasive group B streptococcal disease in the United States, 1999–2005. *JAMA* 299, 2056–2065.
- Ramond, C., Berthault, C., Burlen-Defranoux, O., de Sousa, A.P., Guy-Grand, D., Vieira, P., Pereira, P., and Cumano, A. (2014). Two waves of distinct hematopoietic progenitor cells colonize the fetal thymus. *Nat. Immunol.* 15, 27–35.
- R Core Team (2019). R: A language and environment for statistical computing (R Foundation for Statistical Computing).

- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140.
- RStudio Team (2016). RStudio: Integrated Development Environment for R (RStudio).
- Servais, S., Hannon, M., Peffault de Latour, R., Socie, G., and Beguin, Y. (2017). Reconstitution of adaptive immunity after umbilical cord blood transplantation: impact on infectious complications. *Stem Cell Investig.* 4, 40.
- Siegrist, C.-A. (2001). Neonatal and early life vaccinology. *Vaccine* 19, 3331–3346.
- Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Exp. Hematol.* 33, 259–271.
- Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E., and Dudoit, S. (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* 19, 477.
- Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e21.
- Tieppo, P., Papadopoulou, M., Gatti, D., McGovern, N., Chan, J.K.Y., Gosslin, F., Goetgeluk, G., Weening, K., Ma, L., Dauby, N., et al. (2020). The human fetal thymus generates invariant effector  $\gamma\delta$  T cellsHuman fetal invariant effector  $\gamma\delta$  thymocytes. *J. Exp. Med.* 217, jem.20190580.
- Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* 98, 5116–5121.
- van den Broek, T., Borghans, J.A.M., and van Wijk, F. (2018). The full spectrum of human naive T cells. *Nat. Rev. Immunol.* 18, 363–373.
- van Loosdregt, J., and Coffer, P.J. (2018). The Role of WNT Signaling in Mature T Cells: T Cell Factor Is Coming Home. *J. Immunol.* 201, 2193–2200.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, research0034.1.
- Vijayaraghavan, J., and Osborne, B.A. (2018). Notch and T Cell Function – A Complex Tale. In *Molecular Mechanisms of Notch Signaling*, T. Borggreve and B.D. Giaimo, eds. (Springer International Publishing), pp. 339–354.
- Wright, M.N., and Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* 77, 1–17.
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287.
- Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.
- Zhuang, L., Chen, H., Zhang, S., Zhuang, J., Li, Q., and Feng, Z. (2019). Intestinal Microbiota in Early Life and Its Implications on Childhood Health. *Genomics Proteomics Bioinformatics* 17, 13–25.
- Zota, A.R., Mitro, S.D., Robinson, J.F., Hamilton, E.G., Park, J.-S., Parry, E., Zoeller, R.T., and Woodruff, T.J. (2018). Polybrominated diphenyl ethers (PBDEs) and hydroxylated PBDE metabolites (OH-PBDEs) in maternal and fetal tissues, and associations with fetal cytochrome P450 gene expression. *Environ. Int.* 112, 269–278.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
mouse anti-human CD4-BV650 monoclonal antibody (clone SK3)	BD Horizon	Cat#563875; RRID: AB_2744425
mouse anti-human CCR7-BV785 monoclonal antibody (clone G043H7)	Biolegend	Cat#353230; RRID: AB_2563630
mouse anti-human CD25-FITC monoclonal antibody (clone 2A3)	BD Biosciences	Cat#347643; RRID: AB_400334
mouse anti-human CD45RA-PE monoclonal antibody (clone HI100)	BD Biosciences	Cat#555489; RRID: AB_395880
mouse anti-human CD8-PECy7 monoclonal antibody (clone RPA-T8)	BD Biosciences	Cat#557746; RRID: AB_396852
mouse anti-human CD95-APC monoclonal antibody (clone DX2)	Invitrogen	Cat#17-0959-42; RRID: AB_10807091
mouse anti-human CD27-APC-eFluor780 monoclonal antibody (clone O323)	Invitrogen	Cat#47-0279-42; RRID: AB_1272040
mouse anti-human CD34-PE monoclonal antibody (clone 581)	Biolegend	Cat#343506; RRID: AB_1731862
mouse anti-human CD45-APC monoclonal antibody (clone 8I30)	Tonbo Biosciences	Cat#20-0459-T500; RRID: AB_2621579
mouse anti-human CD3-AlexaFluor700 monoclonal antibody (clone SP34-2)	BD Biosciences	Cat#561805; RRID: AB_10893800
mouse anti-human CD14-Qdot605 monoclonal antibody (clone Tuk4)	Invitrogen	Cat#Q10013; RRID: AB_2556439
mouse anti-human CD16-FITC monoclonal antibody (clone 3G8)	BD PharMingen	Cat#555406; RRID: AB_395806
mouse anti-human HLA-DR-PECy7 monoclonal antibody (clone G46-6)	BD PharMingen	Cat#560651; RRID: AB_1727528
mouse anti-human CD4-Qdot605 monoclonal antibody (clone S3.5)	Life Technologies (now Invitrogen)	Cat#Q10008; RRID: AB_11180611
mouse anti-human CD8-PE monoclonal antibody (clone SK1)	BD Biosciences	CT#345773; RRID: AB_2868801
<b>Biological Samples</b>		
Whole Blood & Whole Bone Marrow, Fresh	AllCells	Custom Order
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
RPMI Medium 1640	GIBCO	Cat#201870-076
DPBS	GIBCO	Cat#14190-144
UltraPure EDTA	Invitrogen	Cat#15575-038
Penicillin Streptomycin	GIBCO	Cat#15140-122
L-Glutamine	GIBCO	Cat#25030-081
Ficoll-Paque	GE Healthcare	Cat#17-5442-03
Percoll	GE Healthcare	Cat#17-0891-01
Fetal Bovine Serum (FBS)	GIBCO	Cat#10437-028
Ghost Dye Violet 510	Tonbo Biosciences	Cat#13-0870
Collagenase Type IV	GIBCO	Cat#17104-019
DNase I	Millipore	Cat#1010459001

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Critical Commercial Assays</b>		
NEBNext Poly(A) mRNA Magnetic Isolation Module	NEB	Cat#E7490
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	NEB	Cat#E7760
NEBNext Library Quantification Kit for Illumina	NEB	Cat#E7630
Agencourt Ampure XP Beads	Beckman Coulter	Cat#A63880
Chromium Single Cell 3' Library & Gel Bead Kit v2	10X Genomics	Cat#PN-120237
EasySep Human CD34 Positive Selection Kit II	STEMCELL Technologies	Cat#17856
EasySep Human T Cell Isolation Kit	STEMCELL Technologies	Cat#17951
QIAshredder	QIAGEN	Cat#79654
RNAqueous-Micro kit	Ambion - Life Technologies	Cat#AM1931
Amino Allyl MessageAmp II aRNA Amplification Kit	Invitrogen	Cat#AM1753
SuperScript® III CellsDirect cDNA Synthesis Kit (Invitrogen)	Invitrogen	Cat#18080200
SsoFast EvaGreen Supermix with Low ROX	Bio-rad	Cat#1725210
20X DNA Binding Dye Sample Loading Reagent	Fluidigm	Cat#100-7609
2X Assay Loading Reagent	Fluidigm	Cat#100-7611
1X DNA Suspension Buffer	Teknova	Cat#T0227
96.96 Dynamic Array IFC	Fluidigm	Cat#BMK-M-96.96
RNAqueous Total RNA Isolation Kit	Invitrogen	Cat#AM1912
SurePrint G3 Human GEv2 8x60K array	Agilent	Cat#G4851B
<b>Deposited Data</b>		
Single-cell mapping of progressive fetal-to-adult transition in human naïve T cells	This Paper	GEO: GSE158493
BlueprintENCODE Reference Data	The ENCODE Project Consortium, 2012; Martens and Stunnenberg, 2013; Aran et al., 2019	Downloadable from within R using the BlueprintEncodeData function of the SingleR package.
Microarray profiling of human fetal mesenteric lymph node and adult peripheral naïve CD4+ T cells	GEO	GEO: GSE25087
Microarray profiling of human fetal and adult classical monocytes from bone marrow	GEO	GEO: GSE54817
<b>Software and Algorithms</b>		
MAST	Finak et al., 2015	1.12.0 (Bioconductor)
ranger	Wright and Ziegler, 2017	0.12.1 (CRAN)
SingleR	Aran et al., 2019	1.0.6 (Bioconductor)
slingshot	Street et al., 2018	1.4.0 (Bioconductor)
eulerr	Larsson, 2019	6.1.0 (CRAN)
clusterProfiler	Yu et al., 2012	3.14.3 (Bioconductor)
DESeq2	Love et al., 2014	1.26.0 (Bioconductor)
limma	Ritchie et al., 2015	3.42.2 (Bioconductor)
Seurat	Butler et al., 2018	3.0.2 (CRAN)
R	R Core Team, 2019	3.6.3 ( <a href="https://cran.r-project.org/">https://cran.r-project.org/</a> )
Rstudio	RStudio Team, 2016	1.3.1056 ( <a href="https://rstudio.com/">https://rstudio.com/</a> )
caret	Kuhn, 2008	6.0-86 (CRAN)
FastQC	Andrews, 2018	0.11.7 (Conda, bioconda channel)
Trim Galore!	Krueger, 2019	0.5.0 (Conda, bioconda channel)
STAR	Dobin et al., 2013	2.6.1b (Conda, bioconda channel)
featureCounts (part of the Subread package)	Liao et al., 2014	1.6.2 (Conda, bioconda channel)

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Genome Analysis Toolkit (GATK)	McKenna et al., 2010	3.8-1-0 (Broad Institute)
CellRanger	10X Genomics	3.0.2 (10X Genomics, <a href="https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest">https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/latest</a> )
Demuxlet	Kang et al., 2018	no version number (Github, <a href="https://github.com/statgen/demuxlet">https://github.com/statgen/demuxlet</a> , commit 85dca0a)
umap-learn	McInnes et al., 2018	0.3.10 (Conda, bioconda channel)
feseR	Perez-Riverol et al., 2017	0.2.0 (Github, <a href="https://github.com/enriquea/feseR">https://github.com/enriquea/feseR</a> , commit 248642c)
geNorm, ctrlGene package	Vandesompele et al., 2002	1.0 (CRAN)
Fluidigm Real Time PCR Analysis	Fluidigm	(within the Biomark & EP1 Software, <a href="https://www.fluidigm.com/software/">https://www.fluidigm.com/software/</a> )
Feature Extraction Software	Agilent	10.7.3.1 ( <a href="https://www.agilent.com/en/product/mirna-microarray-platform/mirna-microarray-software/feature-extraction-software-228496">https://www.agilent.com/en/product/mirna-microarray-platform/mirna-microarray-software/feature-extraction-software-228496</a> )
dittoSeq	(Bunis et al., 2020)	1.1.9 (Github, <a href="https://github.com/dtm2451/dittoSeq">https://github.com/dtm2451/dittoSeq</a> , commit a3bfe2b)

**RESOURCE AVAILABILITY**
**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Trevor D. Burt ([trevor.burt@duke.edu](mailto:trevor.burt@duke.edu)).

**Materials availability**

This study did not generate new unique reagents

**Data and code availability**

Microarray and RNA-seq counts data have been deposited to GEO: GSE158493. To enable exploration of these datasets by other researchers, raw counts and fully processed objects for the bulk and single-cell RNA-seq data are available on figshare ([https://figshare.com/projects/Single-cell\\_mapping\\_of\\_progressive\\_fetal-to-adult\\_transition\\_in\\_human\\_naive\\_T\\_cells/76143](https://figshare.com/projects/Single-cell_mapping_of_progressive_fetal-to-adult_transition_in_human_naive_T_cells/76143)), and all code necessary for recreating the reported analyses and figures within R is available on GitHub (<https://github.com/dtm2451/ProgressiveHematopoiesis>). The dittoSeq visualization software created to aid color blindness-friendly visualization of these, and other, RNA-seq data (Bunis et al., 2020), is part of Bioconductor and is available through Bioconductor at <https://bioconductor.org/packages/release/bioc/html/dittoSeq.html> or GitHub at <https://github.com/dtm2451/dittoSeq>.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**
**Tissue collection**

Fetal bone marrow (BM), spleen, and umbilical cord blood (UCB; 18–23 weeks GA) were obtained from San Francisco General Hospital (San Francisco, California, USA) after elective termination of pregnancy and written informed consent was obtained with IRB-approval from, and under the guidelines of, the UCSF Human Research Protection Program. Samples were excluded in the cases of (a) known maternal infection, (b) known intrauterine fetal demise, and/or (c) known or suspected chromosomal abnormality. Samples were fully de-identified, had no associated Personal Health Information (PHI), and researchers had no access to PHI. Fetal UCB was collected from the umbilical cord after sterile preparation with antiseptic swabs as described previously (Zota et al., 2018). Samples were transported in media on ice, and were processed within 1–2 h of collection. Full-term UCB was obtained in a de-identified manner under the auspices of CHR approved protocols from healthy, full-term infants from San Francisco General Hospital (San Francisco, California, USA) with a gestational age of 37 weeks or greater, and from a subset of children from the Boston metropolitan area who participated in the URECA (Urban Environment and Childhood Asthma) study (Gern et al., 2009). Adult peripheral blood mononuclear cells (PBMCs) were isolated from TRIMA Apheresis collection kit residues that were obtained from the Blood Centers of the Pacific (San Francisco, CA). Matched adult peripheral blood and BM samples were obtained from healthy volunteer donors through AllCells (Alameda, CA). UCB and longitudinal post-natal samples were obtained from a cohort of 15 infants that were followed for up to the first 6 months of life by venipuncture under protocols approved by the UCLA CHR. 11 of these infants

were HIV-exposed, but not infected (mother HIV+). All infants were healthy at the time of collection and none had congenital CMV infection.

## METHOD DETAILS

### Cell isolation

BM cells were isolated by dispersion of BM in R10 (RPMI-1640 supplemented with 10% FBS with 100U/ml penicillin, 100U/ml streptomycin, 2mM L-Glutamine). Splenic tissue was digested with collagenase IV and DNase I in R10 to yield a single cell suspension. Mononuclear cells were isolated from all samples (fetal splenocytes, fetal UCB, newborn UCB and peripheral blood, adult peripheral blood, and fetal and adult BM cells) by density centrifugation of a Ficoll-paque gradient (or Percoll gradient for three fetal splenic T cell samples, which has been previously established to result in identical T cell phenotypes compared with Ficoll ([Ng et al., 2019](#))). All samples were viably cryopreserved prior to use.

### Fluorescence activated cell sorting (FACS) for microarray and Fluidigm qRT-PCR

Mononuclear cell preparations were incubated in FACS staining buffer (PBS with 2% FBS and 2 mM EDTA) with fluorochrome-conjugated anti-human surface monoclonal antibodies (mAbs). Antibodies used included: CD3 Alexa-700 (SP34-2, BD Biosciences), CD14 Qdot605 (Tuk4, Invitrogen), CD16 FITC (3G8, BD PharMingen), HLA-DR PE-Cy7 (G46-6, BD Biosciences), CD4 Qdot655 (S3.5, Life Technologies), CD45RA ECD (3P, Beckman Coulter), CD27 APC-eFluor780 (O323, eBiosciences), CD25 PE(PC61, BD Biosciences), and CD8 PE(SK1, BD Biosciences). All cells were stained with a live/dead marker (Amine-Aqua/AmCyan; Invitrogen) to exclude dead cells from the analysis and sorting. Cells were then sorted by FACS (FACS Aria, BD Biosciences) into PBS. These cells were then re-sorted to a purity of greater than 99% directly into RNAqueous Micro lysis buffer (Ambion – Life Technologies). For Fluidigm qRT-PCR, cells were sorted into Cells Direct 2x Reaction Mix (Life Technologies).

### RNA preparation for microarray analysis

RNA was isolated from FACS-sorted samples using the RNAqueous-Micro kit (Life Technologies) and subjected to two rounds of linear amplification using the Aminoallyl MessageAmp II aRNA Amplification kit (Invitrogen). Cy3-coupled aRNA was fragmented and hybridized overnight to a SurePrint G3 Human GE v2 8x60K microarray, which was washed and scanned per the manufacturer's instructions (Agilent Technologies).

### Statistical analysis of microarrays

Raw intensities were extracted using Feature Extraction software (Agilent) and log<sub>2</sub> transformed. Probes with at least 70% missing data were filtered, then any other missing data were imputed. Data were then quantile normalized using the normalizeBetweenArrays function of the Limma package ([Ritchie et al., 2015](#)), followed by filtering of probes without expression above a 2<sup>7</sup> background for every replicate in at least one sample group, followed by median scaling per gene. Differentially expressed genes were identified using Limma ([Tusher et al., 2001](#)) in R, and data visualized as heatmaps using custom Perl scripts. After unbiased clustering of fetal and adult gene expression, we identified two adult monocyte microarray samples, APB1 and APB5, that clustered with fetal samples. The corresponding T cells for these samples, on the other hand, clustered appropriately with the other adult samples. Based on pathway analysis of genes driving this aberrant cluster, we hypothesized that these two adult samples had a viral infection at the time of blood draw ([Figure S7](#)). We thus removed those two adult samples when performing differential expression analysis for the monocytes. Further, the fetal sample FPB5 was an outlier in clustering and we removed it from further analysis. Thus, for T cell adult versus fetal microarray analysis, we compared 5 adult and 4 fetal samples, while 3 adult and 4 fetal samples were used in monocytes. All 5 adult and 4 fetal samples were used in subsequent PCA analyses to calculate relative weights of signature genes for developmental stage scoring.

### Gene signature derivation for bulk developmental stage score

To derive a broad fetal versus adult transcriptional signature, we used microarray gene expression data to identify genes that were significantly differentially expressed (FDR < 0.05) between fetal and adult samples, and greater than 1.5 fold differentially expressed in the same direction in both the monocytes and the naive T cells. After removing uninformative genes that include "XLOC," "LOC1," "ENST," "ORF," "A\_19," "A\_24," or "A\_33" we identified 169 genes that met these criteria ([Table S1](#)). After optimizing PCR primers, we identified a final list of 33 genes for further analysis.

### Fluidigm qRT-PCR of signature genes

Classical monocytes and naive T cells were sorted directly into 96-well plates containing Cells Direct 2x Reaction Mix, with 200 cells per well in replicates of 6. Lysed cells were subjected to reverse transcription and gene-specific pre-amplification using the SuperScript® III CellsDirect cDNA Synthesis Kit (Invitrogen). Primers were purchased from IDT and the specific targeting amplification (STA) mix was prepared according to the Fluidigm protocol (Gene Expression Using SsoFast EvaGreen SuperMix with Low ROX on the BioMark or BioMark HD System). Unincorporated primers were digested with Exonuclease I (New England Biolabs). Amplified cDNA was diluted and combined with 2X SsoFast EvaGreen Supermix with Low ROX (Bio-rad) and 20X DNA Binding Dye Sample

Loading Reagent (Fluidigm). Primers were combined with 2X Assay Loading Reagent (Fluidigm) and 1X DNA Suspension Buffer (Teknova). Sample and primer mixes were loaded onto 96.96 Dynamic Array IFC (Fluidigm) and qPCR was performed using the BioMark System (Fluidigm).

#### Statistical analysis of Fluidigm qPCR for comparison of fetal, newborn, and adult samples

BioMark qPCR data were analyzed using R software and filtered for quality: Cts greater than 27 (or with amplification quality < 0.5 as determined by Fluidigm Real Time PCR Analysis software) were marked as failed, followed by removal of replicates with greater than 80% of primer assays failed, and of remaining Ct values that were outliers from a normal distribution (two-tailed p < 0.02). Subjects with fewer than three successful replicate wells were then removed. Next, primers were removed which displayed: greater than or equal to 80% of samples failed, followed by removal of primers with an average Ct of greater than 25. Normalization between chips was performed by subtracting the mean Ct value for a gene within a chip from all values for that gene on that chip. Principal component analysis was performed on the processed microarray data to generate PC1 loadings for signature genes. Log2 expression values were averaged over replicates; values were then z-score standardized across samples within a gene and multiplied by the PC1 loading values before summation to generate the signature score.

#### Statistical analysis of Fluidigm qPCR for large, URECA, birth cohort

BioMark qPCR data were analyzed using R software via RStudio ([R Core Team, 2019](#); [RStudio Team, 2016](#)) to mitigate technical variation and filter for quality through removal of reactions with amplification quality < 0.5, followed by inter-chip normalization, normalization across samples by endogenous control genes selected to be stable by geNorm (GAPDH, RPL35A, RPL11 for monocytes; GAPDH, RPL35A, B2M for T cells) ([Vandesompele et al., 2002](#)), and removal of reactions with un-normalized Cts greater than 27. We then required that there be at least three successful replicate wells for each gene associated with every sample. Genes were removed from the signature in order to maximize the number of samples retained. Principal component analysis was performed on the processed microarray data to generate PC1 loadings for these signature genes. Log2 expression values were averaged over replicates; values were then z-score standardized across samples within a gene and multiplied by the PC1 loading values before summation to generate the signature score.

#### Cell enrichment and FACS for RNA sequencing

Mononuclear cell preparations were pre-enriched by immunomagnetic selection with either the EasySep Human CD34 Positive Selection Kit (STEMCELL Technologies) or the EasySep Human T Cell Isolation Kit (STEMCELL Technologies) following manufacturer's protocols. Pre-enriched cells were then incubated in FACS staining buffer (PBS with 2% FBS and 2 mM EDTA) with fluorochrome-conjugated anti-human monoclonal antibodies. Pre-enriched T cells were stained with CD4-BV650 (SK3, BD Biosciences), CCR7-BV785 (G043H7, Biolegend), CD25-FITC (2A3, BD Biosciences), CD45RA-PE (HI100, BD Biosciences), CD8-PECy7 (RPA-T8, BD Biosciences), CD95-APC (DX2, Invitrogen), and CD27-APC-eFluor780 (O323, Invitrogen). Pre-enriched CD34+ hematopoietic stem and progenitor cells were stained with CD34-PE (581, Biolegend) and CD45-APC (8I30, Tonbo Biosciences). All cells were also stained with a live/dead marker (Ghost Dye Violet 510, Tonbo Biosciences) to exclude dead cells from analysis and sorting. Stained, pre-enriched cells were then sorted by FACS (FACS Aria Fusion, BD Biosciences) into R10 and kept on ice until further use. Sort purity was assessed by running a small fraction of the sorted cells from each sample through the FACS Aria Fusion a second time, after all samples had been sorted. Two cord samples yielded fewer than 1000 naive CD8 T cells after sorting, and these cells were left out from further processing.

#### Preparation for single cell RNA-seq library generation

After sort purification, 25 thousand naive CD4 T cells, naive CD8 T cells, and CD34+ hematopoietic stem and progenitor cells (or as many as were sorted), from up to 10 distinct samples, were pooled at equal cell concentrations, encapsulated into droplets, and converted into single-cell transcriptome libraries with 10X Chromium 3' v2 chemistry (10X Genomics) as described previously ([Kang et al., 2018](#); [Zheng et al., 2017](#)). For bulk RNA-seq library generation, after sort purification, RNA from 50 thousand fetal, cord, and adult naive CD4 T cells was extracted and purified using RNAqueous Total RNA Isolation columns (Invitrogen AM1912), with addition of QIAshredder homogenization columns (QIAGEN 79654) at the cell lysate step. mRNA was enriched with poly-dT beads (NEB E7490) reverse transcribed, ligated to adapters, dual barcoded, and amplified using a NEBNext Ultra II Directional RNA Library Prep Kit (NEB E7760) according to the manufacturer's protocol. Completed bulk and single-cell RNA sequencing libraries were assessed for proper sizing via Agilent Bioanalyzer and concentration via qRT-PCR using a NEBNext Library Quantification Kit (NEB E7630). Bulk RNA sequencing libraries were pooled at equal concentration and sequenced on an Illumina HiSeq via paired end 2x100bp sequencing. Single-cell RNA sequencing libraries were pooled at equal concentration and sequenced on an Illumina Novaseq via paired end 2x150bp sequencing.

#### Raw sequencing data pre-processing

Sequencing reads from bulk RNA sequencing were assessed for quality using FastQC ([Andrews, 2018](#)). Low quality reads and ends of reads were trimmed using Trim Galore! ([Krueger, 2019](#)). Reads were aligned to the human genome (hg38) using STAR ([Dobin et al., 2013](#)). Aligned reads were then used for two different purposes. First, reads for each sample overlapping with exons were quantified,

using featureCounts (Liao et al., 2014) for the purpose of differential gene expression and pathways analysis. Second, genotype information was extracted using the Genome Analysis Toolkit (McKenna et al., 2010) for use in Demuxlet deconvolution of single cells' sample identities in single cell RNA sequencing datasets. Sequencing reads from single-cell RNA sequencing were quality controlled, matched by cell barcode, aligned to the genome, and overlapping gene features with unique UMIs quantified using cellranger (10X Genomics). Genetic polymorphisms captured within the single cell RNA sequencing reads were then extracted and compared to genotyping information, previously extracted from bulk RNA sequencing reads, in order to match cells to their original samples using Demuxlet (Kang et al., 2018).

#### Dimensionality reduction analysis and clustering of RNA-seq data

Further processing and analysis of the data was carried out in R using RStudio (R Core Team, 2019; RStudio Team, 2016). For bulk RNA sequencing, regularized log (rlg) normalization of gene counts for all genes with more than 10 reads total was performed using DESeq2 (Love et al., 2014). Rlg values for the 2500 genes, captured in at least 4 of 5 samples of each age, with highest coefficient of variation (standard deviation / mean) were used for principal components analysis (PCA), which was calculated with prcomp, a function built into the base R package (R Core Team, 2019). T cells and hematopoietic stem and progenitor cell single-cell sequencing datasets were analyzed separately. Cells with fewer than 1500 unique molecular identifiers (UMI) captured, fewer than 750 genes captured, or greater than 5 (T cells) or 7 (HSPCs) percent of reads coming from mitochondrial genes were filtered out. Cells determined to be doublets by Demuxlet, as well as any cell assigned to a sample that was not included in the 10X Chromium lane that the cell came from, were also filtered out. After quality control, a standard workflow for dimensionality reduction and clustering analysis in Seurat (Butler et al., 2018) was used for these datasets: Gene counts per cell were log normalized. Top variable genes were selected using the FindVariableGenes function with default settings. For the hematopoietic stem and progenitor cell dataset, cells for the distinct ages were additionally aligned with each other at this step, using the IntegrateData function, in order to enable cell type identification later through joint clustering and trajectory inference analysis (Stuart et al., 2019). Expression (or IntegrateData-adjusted expression) of the identified variable genes was then scaled using the ScaleData function. CellCycle, percentage of mitochondrial reads, and number of UMI were regressed out with the ScaleData function for non-Integrated data. The first 50 principal components (PCs) were calculated based on this scaled expression matrix. Standard deviation of the resulting PCs, and empirical testing by the jack straw method were used to determine how many PCs to carry forward for further dimensionality reduction and clustering. Further dimensionality reduction was carried out with uniform manifold approximation and projection (UMAP) by the umap-learn algorithm (McInnes et al., 2018), as well as clustering by the default Seurat Louvain algorithm with resolution parameter empirically chosen and set to 0.1 for T cells and 1.0 for HSPCs.

#### Differential expression and pathway analysis of RNA-seq data

For the naive CD4 T cell bulk RNA sequencing dataset, DESeq2 was used for calculating genes differentially expressed between ages (FDR < 0.05 and log2 fold change  $\geq 1.5$ , unless otherwise stated in the text). For the naive CD4 and CD8 T cell single-cell dataset and for our hematopoietic stem and progenitor cell dataset, differentially expressed genes were determined using the Seurat adaptation of MAST (Finak et al., 2015) via the FindMarkers function with cutoffs—FDR < 0.05 and relative log fold change  $\geq 0.585$ —applied externally to the function. Venn diagram comparisons of gene sets were generated with the eulerr package (Larsson, 2019). For unbiased pathway analysis, enrichment tests were run, on given gene sets, for all KEGG pathways based on a hypergeometric distribution, using clusterProfiler (Yu et al., 2012) with a FDR cutoff of less than 0.05. In order to facilitate visualization, interpretation, and discussion of data, we focused on signaling pathways by subsetting the full list of enriched pathways to those that included “signal” in their name, or to non-infection immune-related pathways by manual curation based on domain knowledge (Figure 5). All significant pathway enrichments are reported in Table S4.

#### Cell type annotation in the hematopoietic stem and progenitor cell single-cell dataset

To facilitate cell type identification, expression differences likely related to sample age or tissue were first mitigated using the IntegrateData batch correction tool. This was followed by combination of an reference-based cell type assignment, via the Bioconductor version of SingleR (Aran et al., 2019), and a trajectory inference, via slingshot (Street et al., 2018), to identify cell types at the start or end of the differentiation tree. SingleR, along with its included Blueprint-ENCODE (Aran et al., 2019; ENCODE Project Consortium, 2012; Martens and Stunnenberg, 2013) reference dataset, was used to initially score our cells, and identified which cluster contained the highest numbers of hematopoietic stem cells (HSCs). This cluster was then used as the starting point (the start.clus parameter) for slingshot trajectory analysis (Street et al., 2018) based on the principal components of the dataset. Clusters before and after the first and last splits in the trajectory map were then merged and annotated based on most frequent SingleR annotations within the combined clusters. At this point, two separate endpoint clusters branching from the same upstream cluster had been annotated as common lymphoid progenitors (CLP) so these endpoint clusters and their common upstream cluster were merged and all annotated as CLPs.

#### Developmental stage score generation with machine learning through random forest regression

Random forest regression was chosen as the algorithm for generating a developmental stage score of each single-cell transcriptome due to the relative simplicity and assumption-free nature of this algorithm. First, a randomly selected training set of fetal and adult

cells for each target cell type was chosen. For naive T cells, the training set consisted of a random 10% of fetal and adult naive CD4 and CD8 T cells. For individual hematopoietic progenitor cell type annotations, the training set consisted of ~30% of fetal and adult cells of those populations, but the exact percentage of fetal and adult cells included in the training set was adjusted in order to account for the fact that there were unequal numbers of fetal and adult cells for each annotation. Heavily imbalanced training sets can reduce the performance of machine learning algorithms, but applying a 50% correction based on such imbalances is recommended (Chicco, 2017). To apply such a correction, we used the following equations to establish adjusted training cell percentages:

$$\text{Fetal training percentage} = 30\% * \left( 0.5 + \frac{a}{f+a} \right)$$

$$\text{Adult training percentage} = 30\% * \left( 0.5 + \frac{f}{f+a} \right)$$

in which f equals the number of fetal cells and a equals the number of adult cells with each cell type annotation. After training set selection, differentially expressed genes, calculated within the training sets, were used as potential features. These potential features were narrowed down to less than 20 markers using the filter.corr, and rferRF functions of the feseR package (Perez-Riverol et al., 2017) in order to reduce the potential for model over-fitting when too many features are used for training. With this package, after an initial correlation filter of 0.3, iteration of random forest model generation and evaluation, followed by removal of least important feature, was carried out until a maximum area under the receiver operator curve (AUC) was reached. If the number of features leading to maximal AUC was greater than 20, iteration was continued until the final set of variables for which each of three measures of evaluation (accuracy, specificity, and sensitivity) remained at or above 0.99. Next, a final random forest regression model was generated with the caret package using expression of these final markers within training set cells and the ranger method (Kuhn, 2008; Wright and Ziegler, 2017). Accuracy of the models were confirmed based on AUC within the fetal and adult cells that were left out of the training set before being applied to all cells of the target populations. Resulting developmental stage scores were then stored as a metadata within the Seurat objects for further analysis and visualization.

#### Visualization of RNA sequencing data

The R package “dittoSeq: User Friendly Single-Cell and Bulk RNA Sequencing Visualization” was created to complement the analysis of the sequencing data in this paper by powering and simplifying the visualization of both bulk and single-cell RNA-seq data types. Visualization of the dimensionality reduction analyses, cell clustering, developmental stage scores, gene expression levels, differentiation trajectories, and cell type annotations for RNA-seq datasets were all performed with dittoSeq. The package has been described previously (Bunis et al., 2020), but in brief, it is available through Bioconductor and on GitHub at <https://github.com/dtm2451/dittoSeq>. The package works directly with single-cell RNA-seq data stored as either Seurat or SingleCellExperiment data structures, and with bulk data stored as SummarizedExperiment objects. It can additionally import bulk RNA-seq data that have been analyzed by edgeR (Robinson et al., 2010). It contains functions for overlaying gene expression and cellular meta data on top of gene-by-gene or dimensionality reduction scatterplots; for plotting gene expression or other numerical metadata (for example developmental stage scores), grouped by samples or grouped by any other discrete per-cell metric, in violin and/or box or ridge plot format; for plotting discrete per-cell data, such as single-cell clusters, grouped by samples or grouped by any other discrete per-cell metric, as percent composition within the groupings; for generation of gene expression heatmaps; and additionally for import of Demuxlet sample calls, and generation of Demuxlet-associated quality assessment metrics and plots.

#### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses of differential gene expression were performed with packages appropriate to each of our particular types of transcriptional data (Limma for microarray, Seurat and MAST for scRNA-seq, and DESeq2 for bulk RNA-seq), as described within associated Method Details sections. Details of all other statistical tests are provided within associated figure legends. All statistical tests were computed in R.