ENGINEERING

Profiling of mature-stage human breast milk cells identifies six unique lactocyte subpopulations

John P. Gleeson[1], Namit Chaudhary[1], Katherine C. Fein[1], Rose Doerfler[1], Patricia Hredzak-Showalter[1], Kathryn A. Whitehead[1,2]*

Breast milk is chock-full of nutrients, immunological factors, and cells that aid infant development. Maternal cells are the least studied breast milk component, and their unique properties are difficult to identify using traditional techniques. Here, we characterized the cells in mature-stage breast milk from healthy donors at the protein, gene, and transcriptome levels. Holistic analysis of flow cytometry, quantitative polymerase chain reaction, and single-cell RNA sequencing data identified the predominant cell population as epithelial with smaller populations of macrophages and T cells. Two percent of epithelial cells expressed four stem cell markers: SOX2, TRA-1-60, NANOG, and SSEA4. Furthermore, milk contained six distinct epithelial lactocyte subpopulations, including three previously unidentified subpopulations programmed toward mucosal defense and intestinal development. Pseudotime analysis delineated the differentiation pathways of epithelial progenitors. Together, these data define healthy human maternal breast milk cells and provide a basis for their application in maternal and infant medicine.

INTRODUCTION

Although human infants are born immature (*1*, *2*), their diet of breast milk enables continued organ growth and development postpartum, particularly within the intestine (*3*). Human milk contains growth factors, immunomodulatory components, and bacterial and maternal cells. A substantial body of research has delineated the role of many nutritional components of milk, including proteins, fats, and sugars, with a primary goal of creating formula that recapitulates breast milk (*4*). There is also considerable interest in bacterial cells present in the maternal milk microbiome, which help to establish the infant gut microflora (*5*). By contrast, despite their abundance (~10^6 cells/ml), the biology of breast milk cells and their potential role in infant development remain poorly understood.

The first step in determining the physiological relevance of breast milk cells is to identify and characterize them. Various breast milk cell types have been described to some degree in the literature, including lactocytes, immune cells, myoepithelial cells, and stem cells (*6*). Immune cells are the dominant population in early-stage milk (colostrum; <1 week after birth), whereas, in mature-stage milk from healthy mothers (>2 weeks postpartum), most of the cells are lactocytes, which are terminally differentiated milk-producing epithelial cells (*7*). Multipotent stem cells have also been identified in breast milk on the basis of the expression of several stem cell gene markers, including *p63* and *KLF4* (*8*, *9*). These cells were also differentiated into all three germ layers (*10*), a unique characteristic of pluripotent stem cells (*11*).

Although the roles of the maternal cells in breast milk are not yet clear, there is early evidence that they may participate in several vital functions. For example, bulk RNA sequencing (RNA-seq) studies showed that the maternal cells in colostrum express a high level of non-nutritive genes (iron binding, triglyceride catabolism, and protein digestion). Dominant gene expression then shifts in mature-stage milk to the two genes responsible for producing milk

proteins—α-lactalbumin and β-casein (*12*). Establishing baseline concentrations and gene expression profiles for distinct breast milk cell types might also aid the development of diagnostics for the detection of cancer in the breastfeeding person (*13*).

In addition, there is a report that maternal stem and immune cells exit the gastrointestinal tract of the infant and functionally integrate into the infant's organs, similarly to immune cells (*14*). These data open the possibility that maternal cells play a role in organ development and passive immunity. For example, necrotizing enterocolitis, an intestinal disease associated with the use of formula, affects ~7% of premature infants and leads to 1000 to 2000 infant deaths a year in the United States (*15*, *16*).

Given this mounting evidence that maternal breast milk cells are involved in infant development and may have properties amenable to cellular therapies and diagnostics (*17*), it is imperative that milk cells are clearly and comprehensively characterized. To that end, we were motivated to conduct a three-pronged analysis of milk cells with the broadest applicability to nursing people worldwide—fresh, mature-stage milk cells from healthy donors. The experiments reported here were designed to address the limitations of previous milk cell studies.

Specifically, numerous studies relied exclusively on flow cytometry to identify distinct milk cell populations (*10*). This is problematic because of variable fluorescent antibody quality and unavailability of antibodies for previously unidentified cell types in milk. Another limitation is that maternal breast milk cells are sometimes frozen after collection and before analysis (*13*). This can be problematic because some populations of breast milk cells do not survive the freezing process; thus, analysis biases against cell populations that are not sufficiently robust.

Through this work, we have circumvented these challenges by analyzing fresh maternal breast milk cells by flow cytometry, quantitative real-time polymerase chain reaction (RT-qPCR), and single-cell RNA-seq (scRNA-seq) simultaneously. Although there is one other study that characterized breast milk cells using scRNA-seq, it was performed on frozen breast milk samples from females with gestational diabetes (*13*). To our knowledge, our study is the first of its kind, as it holistically describes the complete population of

[1]Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [2]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
*Corresponding author. Email: kawhite@cmu.edu

maternal breast milk cells in mature-stage breast milk from healthy females. Because most lactating mothers are classified as healthy, these findings provide a broadly applicable reference dataset for maternal breast milk cells and lay the foundation for future studies on the role of these cells in infant development and their potential as therapeutic interventions.

RESULTS

Epithelial lactocytes are the most abundant cell population in breast milk

For our initial analysis, we collected freshly expressed breast milk from donors to characterize the maternal cell populations using flow cytometry and RT-qPCR (Fig. 1). Although recruitment was open to all donors regardless of postpartum age, we received donations only of mature-stage milk (≥ 4 weeks, averaging 32 weeks postpartum; table S1) (12). We collected data on the donors' health status (whether they or their infant were recently ill), their infants' age, and gestational age at birth. Following the cell isolation process as previously reported (10), we assessed viability by flow cytometry using a fluorescent live-dead stain and observed a cell viability of $79.2 \pm 6.9\%$ ($n = 10$; fig. S1). The traditional trypan blue cell viability assay reported a high degree of false negatives due to the presence of milk fat globules; this assay should be avoided with breast milk cells. We prioritized sample analysis by flow cytometry, as that required smaller cell pellets to yield data, while RNA extraction required much higher amounts (table S1).

To understand the cell populations present, we selected surface markers for flow cytometry analysis based on previous studies: epithelial cells [epithelial cell adhesion molecule–positive ($EpCAM^+$)], immune cells ($CD45^+$), and mesenchymal cells [Vimentin (VIM^+)] (6, 10, 12). We also assayed for lactocytes, which derive from epithelial cells, and thus are $EpCAM^+$ in addition to being $CK18^+$. The selected histograms indicate that the vast majority of breast milk cells are of epithelial origin (93%), with more modest populations of $CK18^+$, $CD45^+$, and VIM^+ cells from healthy donors ($n = 13$) (Fig. 2A). Immune cell numbers increased from ~7% in milk from healthy donor-infant dyads to ~63 and ~36% when the infant ($n = 1$) or the

mother ($n = 3$), respectively, had been sick (Fig. 2B). These flow cytometry data are shown for all donors in Fig. 2C and agree with previous reports, aside from the low percentage of cells expressing CK18 (10, 13). In the case of illness, the number of donors was too low to run an appropriately powered statistical analysis but shows that, in the case of both recent mastitis and infant illness, there is an increase in immune cells (Fig. 2D). There was no statistically significant shift in the mesenchymal cell population regardless of health status (Fig. 2E). However, mRNA expression analysis shows reasonably high expression of *KRT18*, the gene associated with CK18 and lactocytes (Fig. 2F). This agrees with other reported profiles of lactocyte populations in breast milk regardless of the number of children a person has birthed (13, 18). Accordingly, we concluded that the lactocyte population observed by flow cytometry was artificially low, likely because of poor binding affinity of the antibody. Gene expression of an immune cell gene, *PTPRC*, and the mesenchymal gene marker, *VIM*, were significantly lower in expression, consistent with the flow cytometry data.

Stem-like transcription factors are highly expressed in 5% of breast milk cells

There has been a significant interest in breast milk as a potentially unique source of stem cells. Stem cells can be identified by the transcription factors SOX2, NANOG, and OCT4, which are considered the master regulators in stem cells, and by REX1, SSEA4, TRA-1-60, and KLF4, which are co-regulators of pluripotency (19). Hence, we asked whether the unknown fraction of epithelial cells in breast milk were stem like using flow cytometry and RT-qPCR.

For flow cytometry, we gated on $EpCAM^+$ cells and then measured the percentage of cells expressing SOX2, TRA-1-60, NANOG, and SSEA4 (Fig. 3A). Although we wanted to assess OCT4 as well, we were limited by the constraints of our flow panel. Ultimately, we deemed it more informative to have two master regulator markers and two pluripotency markers in our analysis. These experiments indicated that most epithelial cells were $SOX2^+$ and $TRA-1-60^+$ but significantly fewer were $NANOG^+$ and $SSEA1^+$ (Fig. 3, B and D).

We also conducted gene expression analysis by RT-qPCR for the genes encoding SOX2, TRA-1-60, NANOG, and SSEA4: *SOX2*, *PODXL*,
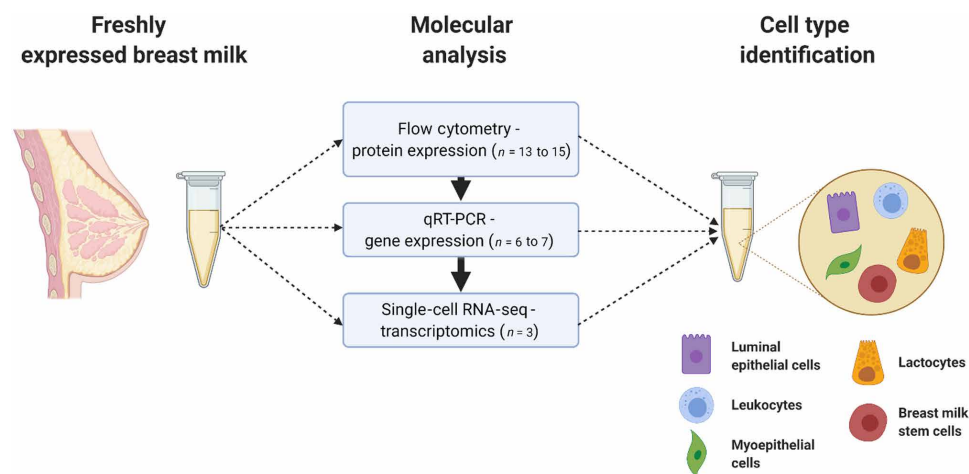

Fig. 1. Maternal cells in breast milk were identified by protein, gene, and transcriptome analyses. Mature-stage breast milk from healthy donors was assessed for expression of cell marker proteins (e.g., EpCAM and CD45) in freshly isolated samples by flow cytometry. Then, RNA was isolated and used to conduct complementary gene expression analysis by RT-qPCR. Single-cell transcriptomics added depth to the analysis and defined the lactocyte population in breast milk.
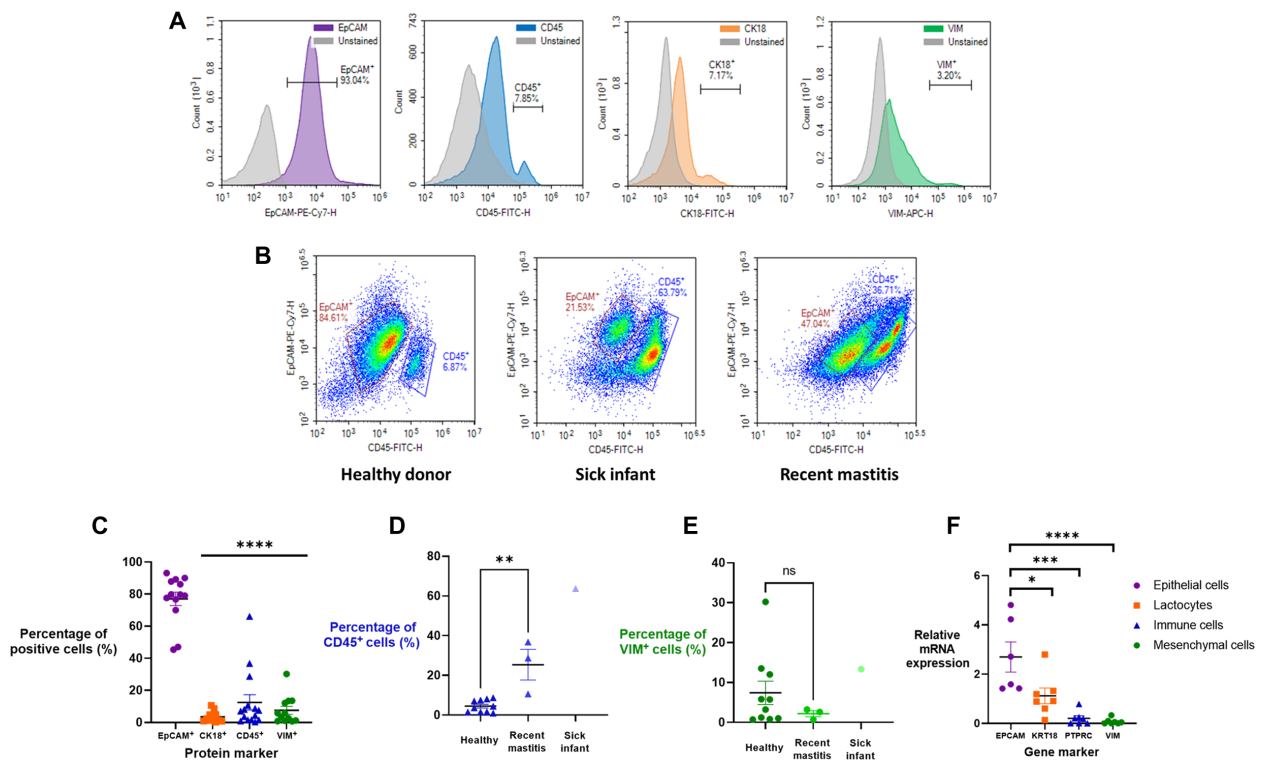
Fig. 2. Mature-stage breast milk from donors contained epithelial lactocytes and a smaller population of immune cells. (**A**) Flow cytometry quantified expression of cell-specific markers: EpCAM+ (epithelial cells), CD45+ (immune cells), CK18+ (lactocytes), and VIM+ (mesenchymal cells). These data suggest that the CK18 antibody bounds poorly to lactocytes and likely underestimates true lactocyte percentages. FITC, fluorescein isothiocyanate. (**B**) The relative cellular composition in breast milk was affected by the health status of the mother and infant. Representative density plots are shown of EpCAM versus CD45 cell populations in milk from healthy donors ($n = 10$), donors with sick infants ($n = 1$), or donors with recent mastitis ($n = 3$). (**C**) Flow cytometry analysis of protein markers in all health statuses ($n = 14$) and the breakdown of (**D**) CD45+ immune cells and (**E**) VIM+ mesenchymal cells in different health statuses. (**F**) mRNA expression of gene markers corresponding to proteins in (C) indicates the presence of epithelial lactocytes ($n = 6$ to 7). Data are shown as means ± SEM. One-way analysis of variance (ANOVA) with Tukey's multiple comparison test; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. ns, not significant.

NANOG, and *FUT4*, respectively. We quantified expression levels relative to the housekeeping gene, *EEF1A*, which was highly and uniformly expressed in all breast milk cells. Epithelial cells expressed 1000-fold greater *FUT4* mRNA than the other assessed genes, with decreasing expression of *PODXL* > *NANOG* > *SOX2* (Fig. 3C). The higher *FUT4* expression is likely due to its role in human milk oligosaccharide synthesis, so expression here may not be solely pluripotency related (*20*). To calculate the stem-like population, we used a gating strategy based on EpCAM+ cells (fig. S2) that subsequently gated on positivity for stem cell markers. To ensure accuracy in analysis, we verified the fluorescence minus one of each antibody (fig. S3) and the stem cell markers were verified in induced pluripotent stem cells (iPSCs) (fig. S4). This allowed us to calculate the percentage of cells that were positive for two or more stem cell markers (Fig. 3D). When considered together, these data indicate that ~5% of the epithelial cells were positive for all four transcription factors assessed.

Single-cell transcriptomic analysis reveals unique lactocyte subtypes
Our flow cytometry and mRNA expression data suggest that breast milk cells are primarily epithelial with a moderate number of lactocytes and a small number of stem-like cells. However, only limited conclusions can be drawn from these two techniques, as the quality

of flow data may have been influenced by antibody quality, and bulk gene expression data cannot delineate gene expression patterns in varied cell populations.

To overcome the latter shortcoming, we performed scRNA-seq to robustly quantify and classify the maternal cell types in milk. For these experiments, we identified six donors with similar cellular profiles by flow cytometry as determined by a first donation. These donors provided samples a second time, and we selected three of those donors (mean, 28 weeks postpartum) to advance to scRNA-seq experiments based on sample consistency and cell yield as determined by flow cytometry (table S1). Upon their third donation, we submitted samples from these three donors within 1 to 2 hours of expression for scRNA-seq analysis. In total, the transcriptomes of 26,922 maternal breast milk cells were sequenced, with a median of 1057 genes per cell.

Gene expression across all three donor samples was dominated by two genes, *LALBA* and *CSN2*, that accounted for >40% of gene counts (Fig. 4A). These two genes encode the proteins α-lactalbumin and β-casein, respectively, which are the two most abundant proteins in human milk (*21*). The genes identified by scRNA-seq were confirmed in a larger donor population by RT-qPCR ($n = 7$); high expression of *CSN2* and *LALBA* with decreasing expression of *CSN1S1* > *FTH1* > *LYZ*; and genes encoding the proteins αS1-casein, ferritin heavy chain 1, and lysozyme, respectively (Fig. 4B).
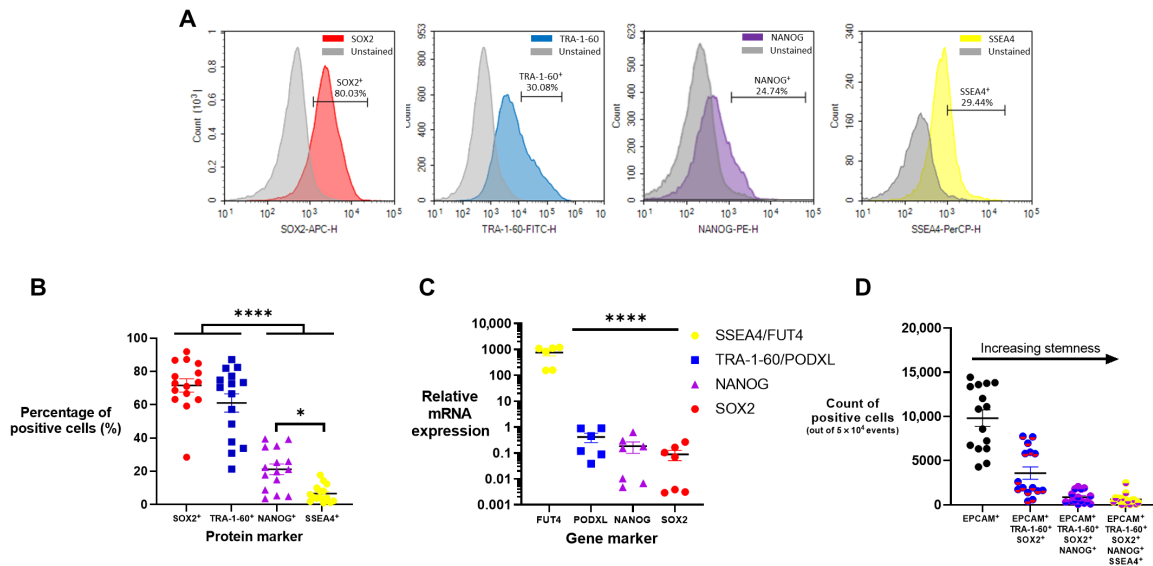
Fig. 3. Stem-like transcription factors were expressed in 5% of breast milk cells. Four stem-like transcription factors were assessed in breast milk epithelial cells. (**A**) From flow cytometry analysis, representative histograms of $SOX2^+$, $TRA\text{-}1\text{-}60^+$, $NANOG^+$, and $SSEA4^+$ stained and unstained samples are shown. (**B**) Flow cytometry analysis of $EpCAM^+$ cells ($n = 15$) indicated high expression of SOX2 and TRA-1-60 with moderate to low expression of NANOG and SSEA4, respectively. (**C**) mRNA expression relative to a housekeeping gene was quantified for the genes corresponding to the proteins in (B) ($n = 7$). (**D**) Flow cytometry analysis of $EpCAM^+$ cells and their increasing stemness based on positive expression of multiple stem-like markers ($n = 15$). Data are means ± SEM. One-way ANOVA with Tukey's multiple comparison test; *$P < 0.05$; ****$P < 0.0001$.

To identify the main cell types within the breast milk cell population that are generalizable across individuals, we performed a combined analysis of the scRNA-seq data using the Seurat and Monocle pipeline (*22*). This analysis identified one large group of cells comprising the contiguous clusters 1 to 7 and two smaller independent clusters (8 and 9) on the basis of uniform manifold approximation and projection (UMAP) clustering (Fig. 4C) (*23*). Comparison across donors showed that their cluster mappings were reasonably homogeneous (Fig. 4D), with notable variations in clusters 8 and 9 (fig. S7).

We then examined the differential expression of genes within each cluster (Fig. 4E), which identified two immune cell clusters, cluster 8 as macrophages ($CD68^+$ and $ITGAX^+$) and cluster 9 as T cells ($PTPRC^+$ and $CD3E^+$). Although the major cluster (1 to 7) was epithelial ($LALBA^+$ and $EPCAM^+$), there were unique subpopulations that we believe to be previously unidentified in human breast milk: $FGFBP1^+$ lactocytes for intestinal development (cluster 2), fatty acid synthesizing and insulin-resistant $PTPRF^+$ lactocytes (cluster 6), and secretory/chemotactic epithelial cells (cluster 7). Cluster 5 contained the cells that had undergone the least differentiation. These included a small population of progenitor-like cells, defined by high expression of $CD55$ and $CLDN4$ (*24*), and a larger population of differentiating cells, which are the precursors of epithelial cells and lactocytes. Last, cluster 4 comprised apoptotic epithelial cells, and it is unclear what cluster these cells belonged to before degradation.

To determine the pathway of differentiation of progenitor-like epithelial cells, we performed pseudotime trajectory analysis (Fig. 5). Figure 5A shows the differentiation pathways of the breast milk cells, originating from the most progenitor-like cells, which are part of cluster 5. The starting node is found in the top, deep purple portion of cluster 5. As pseudotime progresses, the cells differentiate along several distinct pathways, with partially and terminally differentiated cells appearing in reddish orange and yellow, respectively. Figure 5B shows the spatiotemporal gene expression profiles of nine selected cell markers. High expression of genes associated with progenitor cells or stemness (e.g., $CD44$ and $SOD2$) occurred earliest in pseudotime and corresponded with cells in cluster 5 (*25–27*), with $SOD2$ peaking in cluster 1. These progenitor cells briefly expressed $CALML5$, which was exclusively expressed by differentiating epithelial cells, before they diverged into several cell types. Differentiated cells, which appeared in late pseudotime, included mature lactocytes in clusters 2 and 3, lipid and fatty acid synthesizing cells in cluster 6, and chemotactic/secretory epithelial cells in clusters 1 and 7 (Fig. 5C).

Last, we examined the expression homogeneity of selected markers across seven donors (fig. S7), two of which were included in our scRNA-seq (table S1). $SOD2$ and $CALML5$ mRNAs were moderately and homogeneously expressed (fig. S7, A and B), suggesting that progenitor and epithelial cell populations are consistent across donors. By contrast, the chemotactic epithelial marker $FDCSP$ had 30- and 100-fold higher expression in two donors (fig. S7D). These data suggest that lactocytes are homogeneous across donors, while the specialized subtypes may be enriched for a specific need of the feeding infant, such as intestinal stem cell development ($FGFBP1^+$ lactocytes; cluster 2), insulin resistance ($PTPRF^+$ lactocytes; cluster 6), or intestinal immune defense (chemotactic epithelial cells; cluster 7). It should be noted that one of the pseudotime branches terminates in the apoptotic cell cluster 4. This affects the pseudotime expression of genes that are up-regulated in the apoptotic cluster, such as $LALBA$, $FASN$, and $CALML5$. This impact is not apparent in the pseudotime expression of $LALBA$ and $FASN$ because of their overall high expression levels. $CALML5$ peaks in the undifferentiated progenitor-like cells and decreases over time but has a smaller spike in the apoptotic cluster.
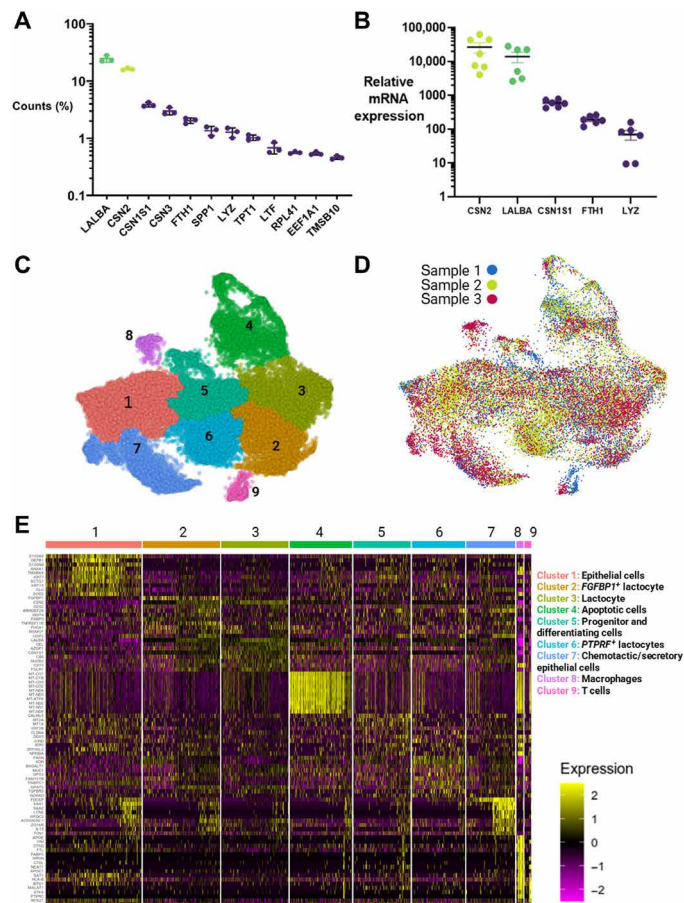
Fig. 4. Single-cell transcriptome analysis identified breast milk cells as primarily epithelial-derived lactocytes, including three novel subpopulations involved in mucosal defense. (**A**) Fifty percent of the transcripts produced by mature breast milk cells from healthy donors corresponded to the milk-producing genes for lactalbumin (*LALBA*) and three types of caseins (*CSN2*, *CSN1S1*, and *CSN3*), $n = 3$. (**B**) RT-qPCR quantified mRNA expression relative to *EEF1A* expression for several top genes ($n = 6$). (**C**) scRNA-seq data generated a UMAP plot that identified one major population of cells (clusters 1 to 7) and two minor populations (clusters 8 and 9). (**D**) The UMAP cluster map is consistent across all donors ($n = 3$). Individual donor UMAP cluster maps are shown in fig. S5. (**E**) Each cluster had unique expression patterns of the top 10 marker genes. These patterns were used to identify specific cell types, including three previously unidentified types of epithelial lactocytes: *FGFBP1*+ lactocytes, *PTPRF*+ lactocytes, and chemotactic epithelial cells. Expression values are relative (unitless) and arbitrary. A higher-resolution image can be found in fig. S6.

DISCUSSION

Through this work, we have provided a detailed characterization of the cellular population in healthy mature-stage breast milk. We confirmed that most cells were epithelial lactocytes and that ~5% of breast milk cells are stem like. Single-cell transcriptomic analysis identified previously unidentified subpopulations of lactocytes programmed toward mucosal defense and intestinal development; however, this requires further in vitro assessment of efficacy.

Most research on human breast milk cells has focused on the characterization of stem-like cells present in the milk (*14*, *17*, *28*, *29*). The most thorough of these studies isolated stem-like cells expressing multiple stem transcription factors (OCT4+, SOX2+, and NANOG+) and differentiated them into all three germ layers (*10*). Unfortunately,

this previous study was unable to determine the percentage of cells positive for more than one marker, because the antibodies used in flow cytometry experiments were all fluorescein isothiocyanate labeled. To our knowledge, we are the first to report that the percentage population of stem-like cells in breast milk is ~5% (SOX2+/TRA-1-60+/NANOG+/SSEA4+) (Fig. 3).

Single-cell transcriptomic analysis did not identify a cluster of stem-like cells within our breast milk cell population, nor were we able to isolate any clusters with high expression of the typical stem-like transcription factors, which is in agreement with recent publications (*30*). These data are not necessarily contradictory to our flow cytometry and qPCR data. One limitation of scRNA-seq is that information about rare genes can be obscured because of several factors, including GC content, secondary RNA structure, RNA sheering process, and RNA-seq library prep. For example, Rachinger and colleagues (*31*) recently demonstrated this issue when analyzing the SOX family of genes. While complementary DNA (cDNA) microarray, RT-PCR, and Western blot data concurred, RNA-seq data did not. In addition, Zhang *et al.* (*32*) have stated that genes identified by RNA-seq must be validated by qPCR, while the opposite is not necessary. With these technique shortcomings in mind along with previously published work (*10*), we therefore conclude that milk contains a stem-like cell population. Future analysis of this population may benefit from combination transcriptomic and proteomic approaches (*33*) or mass cytometry [Cytometry by time of flight (CyTOF)] in combination with our reference dataset (*34*).

Our findings are generally aligned with those reported from the scRNA-seq analysis of two donors with gestational diabetes (*13*) and a longitudinal scRNA-seq lactation study (*30*). Differences observed might be due to the number of cells analyzed, number of donors, and/or our use of fresh rather than frozen samples. The use of fresh cells assures that results are not biased toward more robust milk cells that survive freeze/thaw cycles. Furthermore, freeze/thaw processes have been linked to altered gene expression and translation in human primary cells (*35*, *36*). For example, Lee *et al.* (*37*) demonstrated that fresh peripheral blood mononuclear cells had a higher unique molecular identifier, higher medium genes per cell, and altered scRNA-seq cell clustering compared to cells that were freeze-thawed. Nyquist and colleagues (*30*) found that freezing breast milk cell samples led to a decrease in quality data and recommend processing fresh or storage at 4°C overnight.

These methodological differences enabled our identification of previously unidentified subpopulations: *FGFBP1*+ lactocytes (cluster 2), *PTPRF*+ lactocytes (cluster 6), and chemotactic epithelial cells (cluster 7; Fig. 4E). Regarding cluster 2, *FGFBP1* is a soluble carrier protein that binds to fibroblast growth factor (FGF), which plays a key role in intestinal development and disease regulation (*38*, *39*). Cluster 2 also expressed *TNFRSF11B*, which activates the Wnt/β-catenin pathway (essential for intestinal stem cell proliferation) (*40*) and regulates the differentiation of immune sampling M cells (*41*).

The *PTPRF*+ lactocytes highly expressed *MUC1*, *TGFBR3*, and *PTPRF*. *PTPRF* overexpression in breast milk cells might be linked to insulin-resistant mothers (*12*). *MUC1* secreted in breast milk binds to the lectin domain of dendritic cells found in infant intestines, preventing pathogenic interactions (*42*). *TGFBR3* plays a key role in activation and inhibition of T and B cells, respectively (*43*). In summary, this population may contribute to maternal insulin resistance signaling and intestinal mucosal defense in the infant.
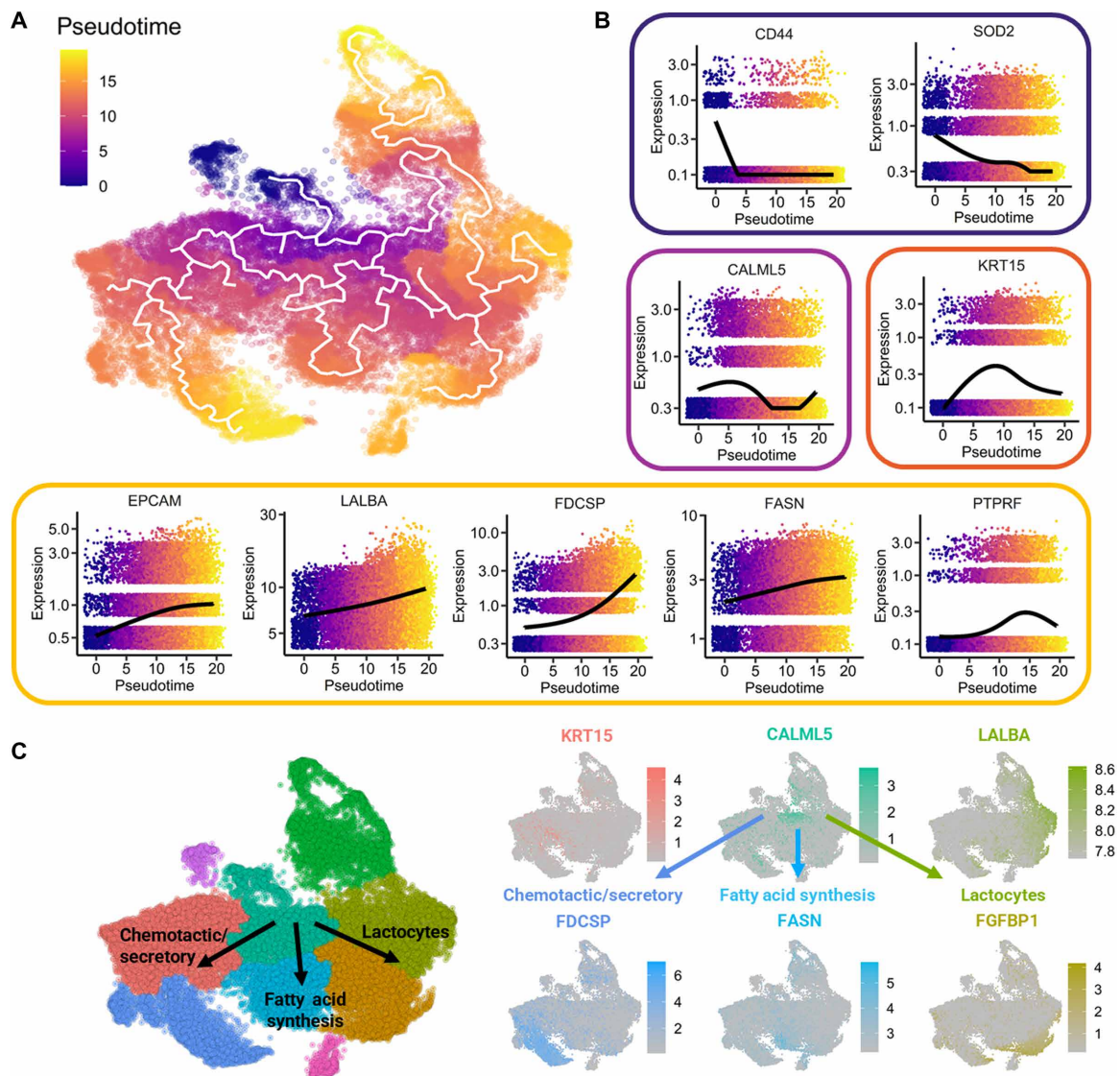
Fig. 5. Pseudotime trajectory analysis identified the pathways by which breast milk epithelial cells differentiate into lactocytes. (**A**) Pseudotime trajectory analysis determined the pathway of differentiation of progenitor-like epithelial cells (starting node is located in the dark purple region). (**B**) As pseudotime progressed, the expression of progenitor genes subsided as cells differentiated into mature phenotypes. (**C**) On the basis of the pseudotime trajectories and gene expression kinetics, milk cells differentiated along three pathways into chemotactic/secretory epithelial cells, fatty acid–synthesizing cells, and mature lactocytes. Each pathway was associated with the overexpression of key genes compared to the other cell types, including *KRT15* (cluster 1), *FDCSP* (cluster 7), *FASN* (cluster 6), *LALBA* (cluster 2), and *FGFBP1* (cluster 3).

Last, chemotactic epithelial cells (cluster 7) had high expression of *FDCSP*, *IL19*, *SAA1*, *WFDC2*, *TCN1*, and *LCN2*. *FDCSP* is considered an ancestral precursor gene to *CSN3*, which produces the milk protein κ-casein (*44*, *45*). The other genes with high expression in this cluster appear to confer mucosal defense, chemokine signaling, and intestinal immune maturation. For example, interleukin-19 expression was increased in the milk cells of lactating females with mastitis that had been treated with oral probiotics (*46*). The bioactive protein product of *SAA1* in human milk confers protection to the neonate in part by intestinal host defense and Th17 T-cells activation (*47*). In addition, *WFDC2* contributes to innate immune defenses of the lung, nasal, and oral cavities (*48*, *49*), and *LCN2* expression occurs in response to small bowel villous disruption (*50*). Assessing the functionality of these cells was beyond the scope

of this study, given the need for custom antibodies to facilitate cell sorting. Once isolated, it would be possible to determine the response of these cells to external stimuli (cytokines, immune cells, etc.) and their interaction to oral mucosa and intestinal epithelium (*51*, *52*). In our study, collection of donor data was blinded to anything beyond health status at the time of milk donation, infant age, and gestational age at birth. Age, ethnicity, and diet can contribute to the nutritional aspects of breast milk, and future studies should delve into the impact of these factors on the breast milk cell population (*53*, *54*).

Bach and colleagues (*55*) suggested that mammary epithelial cells should be conceptualized as being part of a continuous spectrum of differentiation, and our results support this. Overall, our data suggest that the unique lactocyte subtypes identified here by

scRNA-seq are temporally and uniquely suited to the health and development of mother-infant dyads and strengthen the infants' immune system during early development. In addition, we showed (with limited statistical significance; Fig. 2, C and D) that the immune cell population in breast milk is modulated on the basis of both the mother and infant's health status. Hassiotou *et al.* (*7*) reported that up to 94% of the total breast milk cell population consists of leukocytes in the case of infection, and Trend *et al.* (*56*) reported a decrease in basophils in mature breast milk in people with infections. This immune cell population in breast milk could be exploited for therapeutic applications such as chimeric antigen receptor (CAR) T cell therapy (*57*).

These findings have provided a genetic fingerprint for the cells in healthy, mature-stage breast milk and a framework for examining the functionality of these cell types in biologic, prophylactic, and therapeutic contexts. Future work will determine the role of the subpopulations identified here in healthy and sick infants and apply those learnings to the development of novel, oral therapeutic interventions for gastrointestinal and other infantile diseases.

MATERIALS AND METHODS
Breast milk sample collection and cell isolation
This study was approved by Carnegie Mellon University Institutional Review Board under the protocol number STUDY2019_00000084. iPSCs used for stem cell flow cytometry marker validation were obtained from Cedars-Sinai Medical Center's iPSC Core (cell line number CS688iCTR). Participants were recruited via flyers and social media advertisements in mother and infant groups in the greater Pittsburgh area. All participants provided informed written consent and provided data on health status at the time of milk donation, infant age, and gestational age at birth. Pump-expressed mature breast milk was obtained from each participant at lactation suites at Carnegie Mellon University, placed on ice, and transported to the laboratory immediately after expression for analysis. Breast milk samples were processed as previously reported (*10*). In brief, breast milk samples were diluted 1:1 with sterile phosphate-buffered saline (PBS; Life Technologies), centrifuged, washed, and resuspended in PBS before centrifugation three times. The resultant breast milk cell pellet was then processed for scRNA-seq, RT-qPCR, or flow cytometry.

Cell viability and flow cytometry analysis
Cell pellets were resuspended in 100 μl of 1% fetal bovine serum (FBS; VWR) and incubated with either trypan blue (Invitrogen) or LIVE/DEAD Fixable Yellow stain (Invitrogen) for 5 min at room temperature and 30 min at 37°C, respectively. Fixable Yellow samples were subsequently fixed using Flow Cytometry Fixation Buffer (R&D Systems), centrifuged, and resuspended in 1% FBS in PBS. Trypan blue cell viability was measured using Countess II Automated Cell Counter (Thermo Fisher Scientific). For flow cytometry, samples were fixed in fixation buffer for 30 min at 4°C, washed with permeabilization buffer (R&D Systems), centrifuged, and incubated with fluorescent antibodies (table S2) in permeabilization buffer at 4°C for 1 hour. Samples were then washed with 10% FBS in PBS, centrifuged, and resuspended in 10% FBS in PBS. Appropriate negative internal controls were used. Data acquisition was done with a NovoCyte 3000 (ACEA Biosciences) and data analysis with NovoExpress.

Gene expression analysis by RT-qPCR
Cell pellets were lysed in buffer RLT and mixed with an equal volume of 100% ethanol, and the RNA was extracted with an RNeasy mini kit (QIAGEN). cDNA was generated from 1500 ng of RNA using High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). RT-qPCR was performed using SYBR Select Master Mix (Applied Biosystems) on a ViiA 7 Real-Time PCR System (Applied Biosystems). Primer sequences are presented in table S3.

Single-cell RNA-seq
Cells were processed with Chromium Next GEM Single Cell 3′ GEM Library and Gel Bead Kit version 3.1 (10× Genomics) according to the manufacturer's instructions. In brief, cells were diluted in master mix containing reverse transcription reagents and primer and were transferred to the Chromium chip with gel beads and partitioning oil for preparation of nanoliter-scale Gel Bead-In-Emulsions (GEMs). Reverse transcription produced cDNA with a cellular 10× barcode and unique molecular identifier (UMI), and the cDNA was recovered with Dynabeads MyOne SILANE magnetic beads. The cDNA was then subjected to 8 to 14 cycles of amplification before cleanup using SPRIselect Reagent (Beckman Coulter). Total cDNA yield was calculated following quality assessment (High Sensitivity D5000 ScreenTape, Agilent) and fluorometric quantitation (Qubit, Thermo Fisher Scientific). Libraries were constructed by enzymatic fragmentation, end repair, and A-tailing according to the manufacturer's instructions. Two rounds of cleanup were performed: First, a double-sided SPRIselect Reagent (Beckman Coulter) prepared samples for adapter ligation. Then, another round of magnetic bead cleanup was performed before sample indexing PCR for 5 to 16 cycles depending on the cDNA input for library construction. A further double-sided size selection produced libraries of 400 to 600 base pairs (bp), which were quantified for paired-end sequencing (26 bp × 98 bp). The 10× single-cell RNA libraries quality was checked using Fragment Analyzer (NGS Fragment Kit, Agilent), and they were quantified by qPCR (KAPA qPCR quantification kit, Kapa Biosystems). The libraries were then normalized, pooled, and sequenced on a NovaSeq 6000 platform (Illumina) with S1 100-cycles kit (read 1: 28 bp and read 2: 91 bp).

scRNA-seq analysis
All reads were processed using the 10× cellranger pipeline. In brief, reads were demultiplexed, using the cellranger mkfastq function. Demultiplexed reads from all lanes were aligned to the GRCh38 genome, followed by filtering, barcode counting, and UMI counting using the cellranger count function. After constructing the gene expression matrix, cells were processed using Seurat and Monocle packages. Cells were filtered to remove low-quality cells (<400 genes), doublets (>5000 genes), and cells with high mitochondrial fraction (>25%). Cells were assigned cell cycle scores based on the presence of S, G_1, and G_2-M genes. The filtered subset was normalized and scaled using Seurat's SCT pipeline. Mitochondrial percentage, cell cycle scores, gene counts, and UMI counts were regressed out. Samples from all three donors were integrated using Seurat's SCTransform integration pipeline. Highly variable genes from all 26,922 cells were used as features for principal components analysis (PCA) using the RunPCA function. The first 50 principal components were used for UMAP dimensionality reduction using the RunUMAP function. The integrated RNA counts from the Seurat object along with the first 50 principal components and the UMAP

coordinates were transferred to a cell_data_set for clustering and trajectory inference using monocle3. Cells were clustered using the Leiden algorithm in monocle using the cluster_cells function. For trajectory inference, the principal graph was learned using the learn_graph function. Cells were ordered in pseudotime by choosing the root node using the order_cells function. A node in the epithelial cell cluster with high expression of Keratin 7 (KRT7) and Keratin 18 (KRT18) was chosen as the root node. Gene dynamics as a function of pseudotime were analyzed using the plot_genes_in_pseudotime function. Last, the clusters from monocle3 were added to the Seurat object's metadata. Differentially expressed genes were identified using the Model-based Analysis of Single-cell Transcriptomics (MAST) algorithm with the FindAllMarkers function. Gene expression levels in Fig. 4E and fig. S4 are calculated relative to one another and are unitless.

Statistical analysis

All data are represented as means ± SEM. Statistical analysis was carried out using Prism 8 software (GraphPad) using Student's t test and one-way analysis of variance (ANOVA) with Bonferroni's post hoc tests. A significant difference was defined as $P < 0.05$.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abm6865

REFERENCES AND NOTES

1. A. G. Cummins, F. M. Thompson, Postnatal changes in mucosal immune response: A physiological perspective of breast feeding and weaning. *Immunol. Cell Biol.* **75**, 419–429 (1997).
2. J. P. Gleeson, K. C. Fein, N. Chaudhary, R. Doerfler, A. N. Newby, K. A. Whitehead, The enhanced intestinal permeability of infant mice enables oral protein and macromolecular absorption without delivery technology. *Int. J. Pharm.* **593**, 120120 (2021).
3. A. S. Goldman, Modulation of the gastrointestinal tract of infants by human milk. Interfaces and interactions. An evolutionary perspective. *J. Nutr.* **130**, 426S–431S (2000).
4. D. K. Dror, L. H. Allen, Overview of nutrients in human milk. *Adv. Nutr.* **9**, 278S–294S (2018).
5. K. Kordy, T. Gaufin, M. Mwangi, F. Li, C. Cerini, D. J. Lee, H. Adisetiyo, C. Woodward, P. S. Pannaraj, N. H. Tobin, G. M. Aldrovandi, Contributions to human breast milk microbiome and enteromammary transfer of *Bifidobacterium breve*. *PLOS ONE* **15**, e0219633 (2020).
6. M. Witkowska-Zimny, E. Kaminska-El-Hassan, Cells of human breast milk. *Cell. Mol. Biol. Lett.* **22**, 11 (2017).
7. F. Hassiotou, A. R. Hepworth, P. Metzger, C. T. Lai, N. Trengove, P. E. Hartmann, L. Filgueira, Maternal and infant infections stimulate a rapid leukocyte response in breastmilk. *Clin. Transl. Immunol.* **2**, e3 (2013).
8. M. D. Cregan, Y. Fan, A. Appelbee, M. L. Brown, B. Klopcic, J. Koppen, L. R. Mitoulas, K. M. E. Piper, M. A. Choolani, Y.-S. Chong, P. E. Hartmann, Identification of nestin-positive putative mammary stem cells in human breastmilk. *Cell Tissue Res.* **329**, 129–136 (2007).
9. E. Thomas, N. Zeps, M. Cregan, P. Hartmann, T. Martin, 14-3-3σ (sigma) regulates proliferation and differentiation of multipotent p63-positive cells isolated from human breastmilk. *Cell Cycle* **10**, 278–284 (2011).
10. F. Hassiotou, A. Beltran, E. Chetwynd, A. M. Stuebe, A.-J. Twigger, P. Metzger, N. Trengove, C. T. Lai, L. Filgueira, P. Blancafort, P. E. Hartmann, Breastmilk is a novel source of stem cells with multilineage differentiation potential. *Stem Cells* **30**, 2164–2174 (2012).
11. I. International Stem Cell, Assessment of established techniques to determine developmental and malignant potential of human pluripotent stem cells. *Nat. Commun.* **9**, 1925–1925 (2018).
12. D. G. Lemay, O. A. Ballard, M. A. Hughes, A. L. Morrow, N. D. Horseman, L. A. Nommsen-Rivers, RNA sequencing of the human milk fat layer transcriptome reveals distinct gene expression profiles at three stages of lactation. *PLOS ONE* **8**, e67531 (2013).
13. J. F. Martin Carli, G. D. Trahan, K. L. Jones, N. Hirsch, K. P. Rolloff, E. Z. Dunn, J. E. Friedman, L. A. Barbour, T. L. Hernandez, P. S. MacLean, J. Monks, J. L. McManaman, M. C. Rudolph, Single cell RNA sequencing of human milk-derived cells reveals sub-populations of mammary epithelial cells with molecular signatures of progenitor and mature states: A novel, non-invasive framework for investigating human lactation physiology. *Mammary Gland Biol. Neoplasia* **25**, 367–387 (2020).
14. J.-P. Molès, E. Tuaillon, C. Kankasa, A.-S. Bedin, N. Nagot, A. Marchant, J. M. McDermid, P. Van de Perre, Breastmilk cell trafficking induces microchimerism-mediated immune system maturation in the infant. *Pediatr. Allergy Immunol.* **29**, 133–143 (2018).
15. J. R. Robinson, E. J. Rellinger, L. D. Hatch, J.-H. Weitkamp, K. E. Speck, M. Danko, M. L. Blakely, Surgical necrotizing enterocolitis. *Semin. Perinatol.* **41**, 70–79 (2017).
16. S. C. Fitzgibbons, Y. Ching, D. Yu, J. Carpenter, M. Kenny, C. Weldon, C. Lillehei, C. Valim, J. D. Horbar, T. Jaksic, Mortality of necrotizing enterocolitis expressed by birth weight categories. *J. Pediatr. Surg.* **44**, 1072–1075 (2009).
17. M. Witkowska-Zimny, E. K.-El-Hassan, Stem cells in human breast milk. *Hum. Cell* **32**, 223–230 (2019).
18. A.-J. Twigger, L. K. Engelbrecht, K. Bach, I. S.-Pernice, S. Petricca, C. H. Scheel, W. Khaled, Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. bioRxiv 2020.11.06.371443 [Preprint]. 27 November 2020. https://doi.org/10.1101/2020.11.06.371443.
19. T. Seymour, A.-J. Twigger, F. Kakulas, Pluripotency genes and their functions in the normal and aberrant breast and brain. *Int. J. Mol. Sci.* **16**, 27288–27301 (2015).
20. D. S. Newburg, G. M. Ruiz-Palacios, M. Altaye, P. Chaturvedi, J. Meinzen-Derr, M. de L. Guerrero, A. L. Morrow, Innate protection conferred by fucosylated oligosaccharides of human milk against diarrhea in breastfed infants. *Glycobiology* **14**, 253–263 (2003).
21. M. J. C. van Herwijnen, M. I. Zonneveld, S. Goerdayal, E. N. M. Nolte-'t Hoen, J. Garssen, B. Stahl, A. F. M. Altelaar, F. A. Redegeld, M. H. M. Wauben, Comprehensive proteomic analysis of human milk-derived extracellular vesicles unveils a novel functional proteome distinct from other milk components. *Mol. Cell. Proteomics* **15**, 3412–3423 (2016).
22. C. Soneson, M. D. Robinson, Bias, robustness and scalability in single-cell differential expression analysis. *Nat. Methods* **15**, 255–261 (2018).
23. M. D. Luecken, F. J. Theis, Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
24. P. Wang, K. D. McKnight, D. J. Wong, R. T. Rodriguez, T. Sugiyama, X. Gu, A. Ghodasara, K. Qu, H. Y. Chang, S. K. Kim, A molecular signature for purified definitive endoderm guides differentiation and isolation of endoderm from mouse and human embryonic stem cells. *Stem Cells Dev.* **21**, 2273–2287 (2012).
25. C. He, J. M. Danes, P. C. Hart, Y. Zhu, Y. Huang, A. L. de Abreu, J. O'Brien, A. J. Mathison, B. Tang, J. M. Frasor, L. M. Wakefield, D. Ganini, E. Stauder, J. Zielonka, B. N. Gantner, R. A. Urrutia, D. Gius, M. G. Bonini, SOD2 acetylation on lysine 68 promotes stem cell reprogramming in breast cancer. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23534–23541 (2019).
26. C. Solari, C. V. Echegaray, M. S. Cosentino, M. V. Petrone, A. Waisman, C. Luzzani, M. Francia, E. Villodre, G. Lenz, S. Miriuka, L. Barañao, A. Guberman, Manganese superoxide dismutase gene expression is induced by nanog and Oct4, essential pluripotent stem cells' transcription factors. *PLOS ONE* **10**, e0144336 (2015).
27. H. Ghebeh, G. M. Sleiman, P. S. Manogaran, A. Al-Mazrou, E. Barhoush, F. H. Al-Mohanna, A. Tulbah, K. Al-Faqeeh, C. N. Adra, Profiling of normal and malignant breast tissue show CD44high/CD24lowphenotype as a predominant stem/progenitor marker when used in combination with Ep-CAM/CD49f markers. *BMC Cancer* **13**, 289 (2013).
28. M. Ş. Aydın, E. N. Yiğit, E. Vatandaşlar, E. Erdoğan, G. Öztürk, Transfer and integration of breast milk stem cells to the brain of suckling pups. *Sci. Rep.* **8**, 14289 (2018).
29. S. Li, L. Zhang, Q. Zhou, S. Jiang, Y. Yang, Y. Cao, Characterization of stem cells and immune cells in preterm and term mother's milk. *J. Hum. Lact.* **35**, 528–534 (2019).
30. S. K. Nyquist, P. Gao, T. K. J. Haining, M. R. Retchin, Y. Golan, R. S. Drake, K. Kolb, B. E. Mead, N. Ahituv, M. E. Martinez, A. K. Shalek, B. Berger, B. A. Goods, Cellular and transcriptional diversity over the course of human lactation. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2121720119 (2022).
31. N. Rachinger, S. Fischer, I. Böhme, L. Linck-Paulus, S. Kuphal, M. Kappelmann-Fenzl, A. K. Bosserhoff, Loss of gene information: Discrepancies between RNA sequencing, cDNA microarray, and RT-qPCR. *Int. J. Mol. Sci.* **22**, 9349 (2021).
32. M. Zhang, Y.-H. Liu, C.-S. Chang, H. Zhi, S. Wang, W. Xu, C. W. Smith, H.-B. Zhang, Quantification of gene expression while taking into account RNA alternative splicing. *Genomics* **111**, 1517–1528 (2019).
33. A. N. Habowski, J. L. Flesher, J. M. Bates, C.-F. Tsai, K. Martin, R. Zhao, A. K. Ganesan, R. A. Edwards, T. Shi, H. S. Wiley, Y. Shi, K. J. Hertel, M. L. Waterman, Transcriptomic and proteomic signatures of stemness and differentiation in the colon crypt. *Commun. Biol.* **3**, 453 (2020).
34. K. A. Oetjen, K. E. Lindblad, M. Goswami, G. Gui, P. K. Dagur, C. Lai, L. W. Dillon, J. Philip McCoy, C. S. Hourigan, Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight* **3**, e124928 (2018).
35. X. Fu, B. Xu, J. Jiang, X. Du, X. Yu, Y. Yan, S. Li, B. M. Inglis, H. Ma, H. Wang, X. Pei, W. Si, Effects of cryopreservation and long-term culture on biological characteristics and proteomic profiles of human umbilical cord-derived mesenchymal stem cells. *Clin. Proteomics* **17**, 15 (2020).
36. E. Kadić, R. J. Moniz, Y. Huo, A. Chi, I. Kariv, Effect of cryopreservation on delineation of immune cell subpopulations in tumor specimens as determined by multiparametric single cell mass cytometry analysis. *BMC Immunol.* **18**, 6 (2017).

37. J. S. Lee, K. Yi, Y. S. Ju, E.-C. Shin, Effects of cryopreservation and thawing on single-cell transcriptomes of human T cells. *Immune Netw.* **20**, e34 (2020).

38. C. W. Houchen, R. J. George, M. A. Sturmoski, S. M. Cohn, FGF-2 enhances intestinal stem cell survival and its expression is induced after radiation injury. *Am. J. Physiol. Gastrointest. Liver Physiol.* **276**, G249–G258 (1999).

39. S. Kanazawa, T. Tsunoda, E. Onuma, T. Majima, M. Kagiyama, K. Kikuchi, VEGF, basic-FGF, and TGF-beta in Crohn's disease and ulcerative colitis: A novel mechanism of chronic intestinal inflammation. *Am. J. Gastroenterol.* **96**, 822–828 (2001).

40. F. Luan, X. Li, X. Cheng, L. Huangfu, J. Han, T. Guo, H. Du, X. Wen, J. Ji, TNFRSF11B activates Wnt/β-catenin signaling and promotes gastric cancer progression. *Int. J. Biol. Sci.* **16**, 1956–1971 (2020).

41. S. Kimura, Y. Nakamura, N. Kobayashi, K. Shiroguchi, E. Kawakami, M. Mutoh, H. T. -Iwanaga, T. Yamada, M. Hisamoto, M. Nakamura, N. Udagawa, S. Sato, T. Kaisho, T. Iwanaga, K. Hase, Osteoprotegerin-dependent M cell self-regulation balances gut infection and immunity. *Nat. Commun.* **11**, 234 (2020).

42. N. Koning, S. F. M. Kessen, J. P. Van Der Voorn, B. J. Appelmelk, P. V. Jeurink, L. M. J. Knippels, J. Garssen, Y. Van Kooyk, Human milk blocks DC-SIGN-pathogen interaction via MUC1. *Front. Immunol.* **6**, 112 (2015).

43. Y. S. Jang, G. Y. Seo, J. M. Lee, H. Y. Seo, H. J. Han, S. J. Kim, B. R. Jin, H. J. Kim, S. R. Park, K. J. Rhee, W. S. Kim, P. H. Kim, Lactoferrin causes IgA and IgG2b isotype switching through betaglycan binding and activation of canonical TGF-β signaling. *Mucosal Immunol.* **8**, 906–917 (2015).

44. C. Holt, J. A. Carver, Darwinian transformation of a "scarcely nutritious fluid" into milk. *J. Evol. Biol.* **25**, 1253–1263 (2012).

45. K. Kawasaki, A.-G. Lafont, J.-Y. Sire, The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol. Biol. Evol.* **28**, 2053–2061 (2011).

46. J. de Andrés, E. Jiménez, I. Espinosa-Martos, J. M. Rodríguez, M.-T. García-Conesa, An exploratory search for potential molecular targets responsive to the probiotic lactobacillus salivarius PS2 in women with mastitis: Gene expression profiling vs. interindividual variability. *Front. Microbiol.* **9**, 2166 (2018).

47. G. H. Sack Jr., N. Zachara, N. Rosenblum, C. C. Talbot Jr., S. Kreimer, R. Cole, T. L. McDonald, Serum amyloid A1 (SAA1) protein in human colostrum. *FEBS Open Bio* **8**, 435–441 (2018).

48. L. Bingle, S. S. Cross, A. S. High, W. A. Wallace, D. Rassl, G. Yuan, I. Hellstrom, M. A. Campos, C. D. Bingle, WFDC2 (HE4): A potential role in the innate immunity of the oral cavity and respiratory tract and the development of adenocarcinomas of the lung. *Respir. Res.* **7**, 61 (2006).

49. K. M. Morris, D. O'Meally, T. Zaw, X. Song, A. Gillett, M. P. Molloy, A. Polkinghorne, K. Belov, Characterisation of the immune compounds in koala milk using a combined transcriptomic and proteomic approach. *Sci. Rep.* **6**, 35011 (2016).

50. M. de M. G. Prata, A. Havt, D. T. Bolick, R. Pinkerton, A. Lima, R. L. Guerrant, Comparisons between myeloperoxidase, lactoferrin, calprotectin and lipocalin-2, as fecal biomarkers of intestinal inflammation in malnourished children. *J. Transl. Sci.* **2**, 134–139 (2016).

51. E. Driehuis, N. Oosterom, S. G. Heil, I. B. Muller, S. Lin, S. Kolders, G. Jansen, R. de Jonge, R. Pieters, H. Clevers, M. M. van den Heuvel-Eibrink, Patient-derived oral mucosa organoids as an in vitro model for methotrexate induced toxicity in pediatric acute lymphoblastic leukemia. *PLOS ONE* **15**, e0231588 (2020).

52. J. P. Gleeson, H. Q. Estrada, M. Yamashita, C. N. Svendsen, S. R. Targan, R. J. Barrett, Development of physiologically responsive human iPSC-derived intestinal epithelium to study barrier dysfunction in IBD. *Int. J. Mol. Sci.* **21**, 1438 (2020).

53. A.-J. Twigger, A. R. Hepworth, C. T. Lai, E. Chetwynd, A. M. Stuebe, P. Blancafort, P. E. Hartmann, D. T. Geddes, F. Kakulas, Gene expression in breastmilk cells is associated with maternal and infant characteristics. *Sci. Rep.* **5**, 12933 (2015).

54. A. L. Patel, T. J. Johnson, P. P. Meier, Racial and socioeconomic disparities in breast milk feedings in US neonatal intensive care units. *Pediatr. Res.* **89**, 344–352 (2021).

55. K. Bach, S. Pensa, M. Grzelak, J. Hadfield, D. J. Adams, J. C. Marioni, W. T. Khaled, Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nat. Commun.* **8**, 2128 (2017).

56. S. Trend, E. de Jong, M. L. Lloyd, C. H. Kok, P. Richmond, D. A. Doherty, K. Simmer, F. Kakulas, T. Strunk, A. Currie, Leukocyte populations in human preterm and term breast milk identified by multicolour flow cytometry. *PLOS ONE* **10**, e0135580 (2015).

57. A. Cabinian, D. Sinsimer, M. Tang, O. Zumba, H. Mehta, A. Toma, D. Sant'Angelo, Y. Laouar, A. Laouar, Transfer of maternal immune cells by breastfeeding: Maternal cytotoxic t lymphocytes present in breast milk localize in the peyer's patches of the nursed infant. *PLOS ONE* **11**, e0156762 (2016).