

GENETICS

Genomic characterization of normal and aberrant human milk production

Yarden Golan^{1*}, Sarah K. Nyquist², Zhe Liu^{3,4}, Dena Ennis⁵, Jingjing Zhao^{3,4}, Emily Blair⁶, Abdur Rahim Khan⁷, Mary Prah^{6,8}, Stephanie L. Gaw^{9,10}, Moran Yassour^{5,11}, Barbara E. Engelhardt^{2,12}, Valerie J. Flaherman⁶, Nadav Ahituv^{3,4}

Breastfeeding is essential for reducing infant morbidity and mortality, yet exclusive breastfeeding rates remain low, often because of insufficient milk production. The molecular causes of low milk production are not well understood. Fresh milk samples from 30 lactating individuals, classified by milk production levels across postpartum stages, were analyzed using genomic and microbiome techniques. Bulk RNA sequencing of milk fat globules (MFGs), milk cells, and breast tissue revealed that MFG-derived RNA closely mirrors luminal milk cells. Transcriptomic and single-cell RNA analyses identified changes in gene expression and cellular composition, highlighting key genes (*GLP1R*, *PLIN4*, and *KLF10*) and cell-type differences between low and high producers. Infant microbiome diversity was influenced by feeding type but not maternal milk production. This study provides a comprehensive human milk transcriptomic catalog and highlights that MFG could serve as a useful biomarker for milk transcriptome analysis, offering insights into the genetic factors influencing milk production.

INTRODUCTION

Breastfeeding is associated with reduced risk for morbidity and mortality for the infant early and later in life (1). The World Health Organization recommends exclusively breastfeeding for the first 6 months of life, and a continuation of breastfeeding for up to 2 years or longer with the introduction of complimentary foods (2). However, the rates and duration of exclusive breastfeeding are low, with 48% of infants worldwide (3) and 24% of infants in the United States (4) being exclusively breastfed at 6 months of age. One of the main reasons for early weaning, reported by approximately 35% of individuals who wean their infants earlier than recommended, is perceived insufficient milk supply (PIMS) (5). In some cases, early intervention and support for the mother can improve milk production and help to continue breastfeeding (6). However, little is known about the molecular mechanism leading to PIMS and about the molecular changes in the mammary gland in cases of perceived or measured low milk production. This gap in knowledge leads to very limited treatment options in these cases (7, 8). Previous studies found an association between maternal obesity and low milk production (9, 10). Overweight and obese women were less likely to initiate and maintain breastfeeding and more likely to report that their infant is not satisfied with breast milk alone, in comparison to

women with normal weight (11, 12). In addition, high levels of systemic tumor necrosis factor- α (TNF- α) and inflammation were suggested to be associated with low milk production in obesity (13, 14), as well as insulin dysregulation (15). Nevertheless, PIMS and low milk production is also common among women who are not overweight (5, 16). Larger and controlled clinical studies are needed to better understand these relationships and to examine improved intervention protocols.

On the opposite end of the milk production spectrum, some women suffer from over production (hyperlactation) that may be idiopathic or may be caused by over pumping or use of galactagogues (i.e., substances thought to increase the rate of human milk synthesis but are not approved by the Food and Drug Administration) (17). Hyperlactation is also at risk for early weaning because of higher frequencies of breast pain, plugged ducts, and mastitis (17). Global characterization of the cellular signaling, genes, regulatory elements, and pathways associated with low or high milk production may help to better diagnose, treat, and support these individuals.

Genetic studies on human breastfeeding complications are sparse in part because these conditions are not well documented in medical records, and we lack a gold standard method to diagnose hypo or hyperlactation. Almost all genome-wide association studies on milk yield have been performed in cows and other dairy animals, finding associations between single-nucleotide polymorphisms (SNPs) and milk production traits (18). One example is an SNP associated with the gene *DGAT1*, a key enzyme that catalyzes the final step of triglyceride synthesis in mammary gland cells, and is associated with milk yield and fat and lactose production traits (19–23). To date, only one human study has been published showing a single genetic variant associated with milk production and breastfeeding duration (24). In this candidate gene study, a single SNP in the milk fat globule (MFG) epidermal growth factor and factor V/VIII domain containing gene (*MFGE8*) was found to be associated with PIMS, breastfeeding exclusivity, and duration. In addition, there are a few studies on the relationship between human maternal genetics and milk composition, which have focused on human milk oligosaccharides, zinc

Copyright © 2025 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).

¹Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA. ²Gladstone Institutes, San Francisco, CA, USA. ³Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA, USA. ⁴Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. ⁵Microbiology and Molecular Genetics Department, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. ⁶Department of Pediatrics, University of California, San Francisco, San Francisco, CA, USA. ⁷Benioff Center for Microbiome Medicine, Division of Gastroenterology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. ⁸Division of Pediatric Infectious Diseases and Global Health, University of California, San Francisco, San Francisco, CA, USA. ⁹Division of Maternal-Fetal Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, USA. ¹⁰Center for Reproductive Sciences, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, San Francisco, CA, USA. ¹¹The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel. ¹²Department of Biomedical Data Science, Stanford University, Stanford, CA, USA.

*Corresponding author. Email: yarden.golan@cornell.edu

transport, and fatty acids (25–31). While these studies have not tested for associations with milk production, they nevertheless provide support for the role of various genes in the milk production process to develop more effective diagnostics and interventions. Moreover, one previous study on milk gene expression found higher expression of the *TARDBP* gene in exclusive breastfeeding individuals compared to those supplemented with formula 3 to 5 days postpartum, often an indication of issues with milk production (32). Together, these studies suggest that human milk production or milk yield might also be regulated at the gene expression level.

Recent work used functional genomics tools to characterize the human milk transcriptome (33–35) and have found that MFGs can serve as an information-rich, noninvasive biospecimen for learning about mammary gland function (33–37). MFG contains high amounts of RNA that is easy to extract compared to the relatively low amount of RNA extracted from milk cells, and it does not require sorting for specific cell populations as they are secreted from epithelial cells (38, 39). During this process, MFGs are coated by the epithelial cell membrane, and some of the cell's cytosol is secreted into the milk in the crescent structure between the MFG membrane and the cell membrane (40). However, to date, no direct comparison between MFG RNA and human milk cells was published, and it is unclear whether specific RNA is targeted to be secreted in these compartments. In addition, we and others used single-cell RNA sequencing (scRNA-seq) to characterize various milk cell type populations, at different breastfeeding time points, finding two to six different lactocyte cell type populations (41–46) and cross-talk between these cell types and immune cells (44). The two main epithelial cell subtypes that are found in milk are referred to as luminal cells 1 and 2 (LC1s and LC2s) (42–45). It is unclear whether these epithelial subtypes play different roles in milk production or MFG secretion. A better understanding of the MFG transcriptome and the cells that produce them could dramatically enhance studies that use these accessible milk fractions as a proxy biomarker to understand the mammary gland function during lactation.

Here, we used RNA-seq and scRNA-seq to better understand the human milk transcriptome under normal and aberrant milk production conditions. We show that the MFG transcriptome resembles milk luminal cells with more similarity to the LC2 cells compared to LC1 cells and can be used as a milk biomarker studying the function of the mammary gland luminal cells during lactation. We also provide an important database of genes that are up-regulated during lactation compared to nonlactating breast tissue. Using samples from individuals that differ in their milk production, stratified into low, normal, and high production groups, we identify changes in the cellular transcriptome and milk cell type composition between these groups. In addition, we tested whether maternal low milk production affects the infant microbiome and whether infant formula supplementation, which occurs more frequently in infants nursing from mothers with low milk production, affects the infant microbiome. Together, our findings shed light on molecular and cellular changes under different levels of milk production.

RESULTS

Collection of human milk samples with differences in milk production

To identify the molecular factors associated with milk production, we collected fresh milk samples from 30 lactating individuals during

various lactation stages (Fig. 1A). Participants included 9 individuals with low milk production 7 individuals with high milk production, all referred by lactation consultants after a breast exam and infant latching examination (further details in table S1), and 14 individuals who self-reported normal milk production [recruited from social media and distributed flyers at University of California, San Francisco (UCSF) clinics]. In addition to breastfeeding consultants' categorization (table S1), mothers also reported perception of their milk production, which differed across the study groups ($\chi^2 P \leq 0.001$; Table 1). One mother who self-reported low milk production but was categorized by the lactation consultant as a normal supplier was excluded from further analysis. Furthermore, we assessed maternal reports about their infants' satisfaction with the amount of breast milk the infants received ("Is the infant satisfied with mother's own milk?"); we found this measure to also differ across the study milk-production groups ($\chi^2 P \leq 0.001$; Table 1). Mothers with low milk production were more likely to supplement their infant's diet with baby formula or donor milk; seven of the nine mothers supplemented with infant formula in the first 8 days postpartum (Fig. 1; additional information on formula supplementation and lactation consultant summary for each participant in table S1). Using data collected at enrollment, we found that individuals with normal milk production tended to report that they were breastfed as an infant, compared to the low and high groups that report less frequently that they were breastfed as infants [analysis of variance (ANOVA), $P \leq 0.042$; Table 1]. In addition, there was a trend of self-report of delay in the day they felt that the milk "came in" (corresponding to lactogenesis II/secretory activation) in the low production group (ANOVA, $P \leq 0.072$; Table 1), which was previously shown to be associated with unintended breastfeeding reduction and cessation (47). There was no statistically significant difference in maternal body mass index (BMI) between the groups in our cohort (ANOVA, $P > 0.1$; Table 1). Participants with low, normal, and high milk production showed relevant differences between perceived milk production but no difference in other known risk factors for low milk production, including maternal age, and delivery mode (Table 1). Each individual contributed a mean of 2 ± 1.6 samples to the study.

The MFG RNA signature recapitulates the lactocyte transcriptome

MFG RNA is readily accessible and contains high amounts of RNA compared to the cell pellet in each milk sample. However, the origin of this RNA and direct comparison to milk cells RNA from the same samples was never performed, leaving a gap in knowledge about the usability of MFG RNA as a biomarker for the different cell types in the mammary gland. To address this, bulk RNA-seq was performed on a subset of four paired human milk cells and MFG samples obtained from the same milk sample of three individuals (days 139 to 481 postpartum), followed by transcriptomic comparisons (sample information in Table 2). We found 1696 genes to be differentially expressed (DE) between the milk cells and MFG; of those, 1454 genes showed higher expression in human milk cells and only 242 genes showed higher expression in MFG (Fig. 2A). We next performed gene ontology (GO) enrichment analysis using clusterProfiler [adjusted P values calculated using Benjamini-Hochberg (BH) correction (48)]. We found that genes with higher expression in milk cells are associated with GO terms related to leukocyte and lymphocyte differentiation and cytokines, capturing the presence

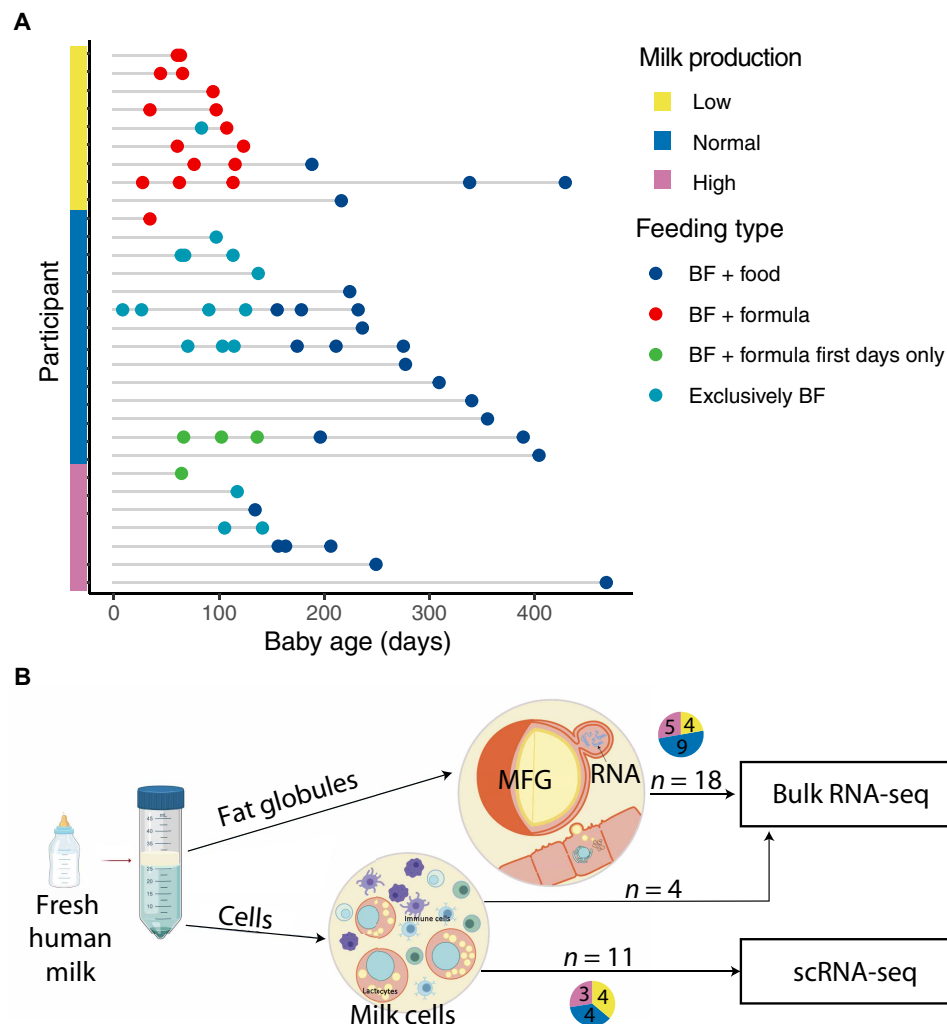


Fig. 1. Samples and milk fractions used in this study. (A) Graph showing the samples used in the study, their categorization into milk production groups, lactation stage, and infant feeding type (BF, breastfeeding) at the time of sample collection. (B) Schematic showing the milk fraction observed after centrifugation of fresh milk samples, MFG at the top and milk cells at the bottom and the experiments carried out on them.

versus absence of immune cells in milk cells versus the MFG fraction, respectively (Fig. 2B). For the genes with higher expression in MFG, we observed no significant GO enrichment. The overexpressed genes in the MFG fraction overlapped several GO term lists (adjusted P values >0.05) including positive regulation of lipid localization, the regulation of fatty acid transport, and the acetyl-coenzyme A metabolic process (Fig. 2C). This analysis showed that most of the MFG transcriptome is also present in the milk cells fraction. In contrast, if one is interested in immune function during lactation, the cell pellet should be used, not MFG RNA.

MFGs, secreted from mature and active milk-producing cells, primarily contain milk protein-related transcripts (33). To better understand aberrant lactation, our first goal was to identify genes and pathways up-regulated during normal lactation. We compared MFG and milk cell transcriptomes to the nonlactating mammary gland transcriptome to determine whether similar pathways are up-regulated in both, providing insight into the representativeness of MFG RNA as a model for the milk-producing cells' transcriptome. For this analysis, we compared publicly available bulk RNA-seq data

from nonlactational breast tissue (49) to bulk RNA-seq from milk cells and MFG in our samples. We identified 11,810 DE genes between nonlactational breast tissue and milk cell or MFG fractions (table S2). Comparing the milk cell transcriptome to nonlactational breast tissue, we found 3462 genes to be highly expressed in milk cells, many of them known to be related to milk production (including the *casein* family genes), and 5796 genes highly expressed in nonlactational breast tissue compared to milk cells (Fig. 2D and table S2). The genes up-regulated in milk cells were enriched for pathways associated with myeloid leukocyte activation (BH-adjusted $P \leq 9.8 \times 10^{-11}$) and positive regulation of cytokine production (BH-adjusted $P \leq 1.3 \times 10^{-09}$), as well as regulation of protein transport (BH-adjusted $P \leq 1.4 \times 10^{-07}$) capturing the higher activity of immune cells during lactation than in nonlactational breast tissue (Fig. 2E and table S2) (50). The up-regulated pathways in the MFG transcriptome compared to the breast tissue transcriptome were associated with cytoplasmic translation (BH-adjusted $P \leq 5.3 \times 10^{-07}$), small molecule catabolic processes (BH-adjusted $P \leq 1.1 \times 10^{-06}$), and fatty acids metabolic process (BH-adjusted $P \leq 1.4 \times 10^{-06}$), which are

Table 1. Study participants' characteristics. ns, not significant; * <i>P</i> < 0.1; ** <i>P</i> < 0.05; *** <i>P</i> < 0.01.						
Category	Overall	Low	Normal	High	<i>P</i> value	Test
n	30	9	14	7		
Age (years) [mean (SD)]	35.85 (3.26)	36.45 (3.15)	35.93 (3.08)	34.93 (4.00)	ns	ANOVA
Ethnic group (%)					ns	χ^2
White	20 (66.7)	7 (77.8)	9 (64.3)	4 (57.1)		
Asian Indian	1 (3.3)	0 (0.0)	1 (7.1)	0 (0.0)		
Chinese	3 (10.0)	1 (11.1)	1 (7.1)	1 (14.3)		
Hispanic	2 (6.7)	0 (0.0)	1 (7.1)	1 (14.3)		
Other	3 (10.0)	1 (11.1)	1 (7.1)	1 (14.3)		
Did not answer	1 (3.3)	0 (0.0)	1 (7.1)	0 (0.0)		
BMI [mean (SD)]	24.37 (4.06)	26.58 (4.02)	22.94 (2.51)	24.40 (5.97)	ns	ANOVA
Number of pregnancies [mean (SD)]	2.22 (1.00)	2.00 (1.07)	2.00 (0.71)	2.83 (1.17)	ns	ANOVA
Number of live birth [mean (SD)]	1.50 (0.58)	1.12 (0.35)	1.57 (0.65)	1.83 (0.41)	*	ANOVA
Age at first pregnancy [mean (SD)]	33.48 (2.74)	34.75 (1.91)	33.44 (2.40)	31.83 (3.60)	ns	ANOVA
Breastfeeding experience (%)					ns	χ^2
Do not have other kids	13 (44.8)	7 (77.8)	7 (50.0)	1 (16.7)		
Breastfed 1 child before	1 (3.4)	2 (22.2)	6 (42.9)	5 (83.3)		
Breastfed 2 children before	15 (51.7)	0 (0.0)	1 (7.1)	0 (0.0)		
Weight before pregnancy (kilograms) [mean (SD)]	63.74 (12.07)	69.92 (14.72)	60.71 (7.62)	61.86 (14.39)	ns	ANOVA
Breastfed as a child = Yes (%)	21 (72.4)	5 (55.6)	13 (92.9)	3 (50.0)	**	χ^2
Age of first menstrual [mean (SD)]	13.09 (1.38)	13.50 (1.60)	13.00 (1.00)	12.67 (1.63)	ns	ANOVA
Baby gender = Male (%)	17 (58.6)	5 (62.5)	8 (57.1)	4 (57.1)	ns	χ^2
Gestational age at delivery [mean (SD)]	39.10 (1.27)	39.33 (1.41)	39.29 (1.27)	38.43 (0.98)	ns	ANOVA
Mode of delivery (%)					ns	χ^2
Vaginal	27 (90.0)	8 (88.9)	14 (100.0)	5 (71.4)		
C-section after failed trail of labor	2 (6.7)	1 (11.1)	0 (0.0)	1 (14.3)		
Intended C-section	1 (3.3)	0 (0.0)	0 (0.0)	1 (14.3)		
Day postpartum that milk came in (Lactogenesis stage II) [mean (SD)]	3.45 (1.75)	4.62 (2.26)	3.04 (1.48)	2.86 (0.90)	*	ANOVA
First try breastfeeding (%)					ns	χ^2
Immediately	14 (46.7)	2 (22.2)	7 (50.0)	5 (71.4)		
<30 min	9 (30.0)	4 (44.4)	4 (28.6)	1 (14.3)		
30 min–1 hour after birth	4 (13.3)	2 (22.2)	1 (7.1)	1 (14.3)		
(Continued)						

(Continued)

Category	Overall	Low	Normal	High	P value	Test
2–6 hours after birth	1 (3.3)	0 (0.0)	1 (7.1)	0 (0.0)	***	χ^2
>24 hours after birth		1 (11.1)	1 (7.1)	0 (0.0)		
Milk production assessment (%)						
I feel that I produce not enough breast milk to meet my baby's needs.	9 (30.0)	9 (100.0)	0 (0.0)	0 (0.0)	***	χ^2
I feel that I produce normal levels of breast milk to meet my baby's needs.	14 (46.7)	0 (0.0)	14 (100.0)	0 (0.0)		
I feel that I produce too much breast milk to meet my baby's needs.	7 (23.3)	0 (0.0)	0 (0.0)	7 (100.0)		
My infant is satisfied with my own milk. (%)					***	χ^2
Agree strongly	16 (53.3)	0 (0.0)	10 (71.4)	6 (85.7)		
Agree moderately	3 (10.0)	0 (0.0)	3 (21.4)	0 (0.0)		
Agree slightly	2 (6.7)	1 (11.1)	1 (7.1)	0 (0.0)		
Disagree moderately	3 (10.0)	3 (33.3)	0 (0.0)	0 (0.0)		
Disagree strongly	5 (16.7)	5 (55.6)	0 (0.0)	0 (0.0)		
Neither agree nor disagree	1 (3.3)	0 (0.0)	0 (0.0)	1 (14.3)		

all activated during lactation (Fig. 2F and table S2). Comparing the MFG transcriptome to nonlactational breast tissue, we found 2935 genes to be highly expressed in MFG and 6122 genes highly expressed in nonlactational breast tissue compared to MFG (Fig. 2G and table S2). Moreover, we found a large overlap in DE genes [DESeq2 (51) adjusted $P \leq 0.05$] between milk cells and MFG compared to nonlactational breast tissue, including known milk production genes such as *CEL*, *OLAH*, *SPP1*, and *CSN3*, with 2148 jointly up-regulated in milk cells and MFG, and 4177 genes jointly down-regulated versus nonlactational breast tissue (Fig. 2H and table S2). In summary, our results show that the MFG transcriptome closely resembles milk cells and may be used as an accessible and homogeneous proxy marker to analyze epithelial cell gene expression during lactation. This analysis also provides an important database for genes that are up-regulating during lactation and can facilitate future studies focusing on lactation specific pathways.

The MFG transcriptome is more similar to the lactocyte LC2 subtype compare to LC1

Previous studies have identified two main epithelial subtypes in human breast milk—LC1s and LC2s (42–45). We next set out to characterize whether the MFG RNA is more similar to one of these subtypes, which might help to better understand the origin of MFG RNA and the specific roles of these sub-cell types in milk production.

To achieve this, we analyzed scRNA-seq data generated using the 10X Genomics platform from 11 milk cells samples in our cohort

(Fig. 1B, Table 2, and table S3). After filtering, we had 17,000 cells that were clustered into nine broad cell-type categories matching previous milk scRNA-seq reports (42–45). These cell-type categories included: (i) six immune cell clusters consisting of T cells, natural killer (NK) cells, B cells, plasma cells, dendritic cells (DCs), and macrophages; (ii) two large luminal epithelial cell clusters LC1 and LC2; and (iii) a small dividing epithelial cell (DEC) cluster (Fig. 3A, fig. S1, and Table 3). We then used the BisqueRNA package (52), which uses scRNA-seq signatures to deconvolve bulk transcriptomes into component cell type proportions as a proxy for cell type of origin for the MFG transcripts. First, we generated a reference signature of general single cell types found in our scRNA-seq samples (Fig. 3A; LC1s, LC2s, macrophages and DCs, B cells, T cells, NK cells, and plasma cells). Then, we applied this reference signature to identify which cell population is more highly represented in the MFG bulk RNA-seq data. We found that LC2 cells had the highest proportion in 12 of the 14 MFG bulk RNA-seq samples that we examined (Fig. 3B). In the two samples for which LC2 did not have the highest proportion, the macrophage and DC signatures had the highest proportions and LC1 and LC2 had the second highest proportions.

We next generated a signature of genes highly expressed in MFG relative to milk cells (242 genes and MFG signature), from our bulk RNA-seq (Fig. 3C and table S2). We used this signature to characterize which cell types could be the source of the unique RNA secreted in the MFG. We found that the MFG signature was higher in epithelial

Table 2. Samples sequences for different analysis in the study. "y" represents samples that were sequenced in the different assays performed in this study.

Participant ID	Milk production group	Days postpartum	MFG RNA-seq	Cell bulk RNA-seq	Cell scRNA-seq	Infant stool microbiome	maternal stool microbiome
ID1024	High	208	y				
ID1025	High	106	y		y	y*	y*
ID1026	High	84	y		y		
ID1012	High	469	y	y			
ID1012	High	481	y	y			
ID1024	High	158			y	y*	y*
ID1018	Low	46				y	y
ID1018	Low	67	y			y	
ID1023	Low	28				y	y
ID1023	Low	64	y		y	y	y
ID1027	Low	95				y	y
ID1028	Low	85	y		y	y	y
ID1028	Low	106				y	
ID1029	Low	62	y		y		
ID1030	Low	36				y	
ID1030	Low	99			y	y	y
ID1036	Low	62				y	y
ID1036	Low	65				y	y
ID1037	Low	84				y	y
ID1016	Normal	68	y		y	y	y
ID1016	Normal	94	y			y	y
ID1016	Normal	138				y	y
ID1020	Normal	70	y		y	y	y
ID1020	Normal	94	y			y	y
ID1020	Normal	116	y			y	y
ID1020	Normal	213	y			y*	y*
VPL131	Normal	68	y				
ID1013	Normal	386	y	y			
ID1014	Normal	139	y	y			
ID1020	Normal	165			y		
ID1033	Normal	61			y	y	y
ID1033	Normal	114				y	

*Samples that were not included in the microbiome analysis due to late time postpartum or low number of participants from high production group.

cells and highest in the LC2 cell subset (Fig. 3D). In addition, the LC2 subpopulation showed high expression levels for genes associated with MFG membrane formation and budding (*FABP3*, *BTN1A1*, *XDH*, and *CIDEA*; Fig. 3E) (53) supporting the hypothesis that MFG may form primarily in LC2 cells. These results further suggest that the MFG transcriptome is more similar to LC2 cells, as compared to LC1 cells, but further studies are needed to delineate the different functions or contribution of each cell type to milk production.

MFG bulk RNA-seq identifies transcriptional changes associated with milk production

We next generated additional bulk RNA-seq from four, seven, and three selected samples collected from low, normal, and high producers (respectively) during postpartum days 62 to 213 (mature milk) to analyze whether different levels of milk production were associated

with MFG gene expression differences. Samples from low and high producers were selected to match as much as possible the days postpartum of normal producers, and baby age was included as a covariate in the differential expression model (Table 2). Following RNA-seq, we used DESeq2 for DE analysis between milk production groups, sequentially comparing low versus normal and high versus normal samples (51). We found that 15 and 65 genes were down-regulated between low versus normal and high versus normal, respectively; 7 and 45 genes were up-regulated between low versus normal and high versus normal, respectively (Fig. 4, A and B, and table S4).

One gene that was up-regulated in high compared to normal producers, *KLF10*, has been linked to the regulation of transforming growth factor-β signaling (54); disruptions in this pathway could affect mammary gland development and lactation (55, 56). *GLP1R* and *PPP1R1A* were down-regulated in high compared to normal producers (Fig. 4A) and have been previously associated with glucose

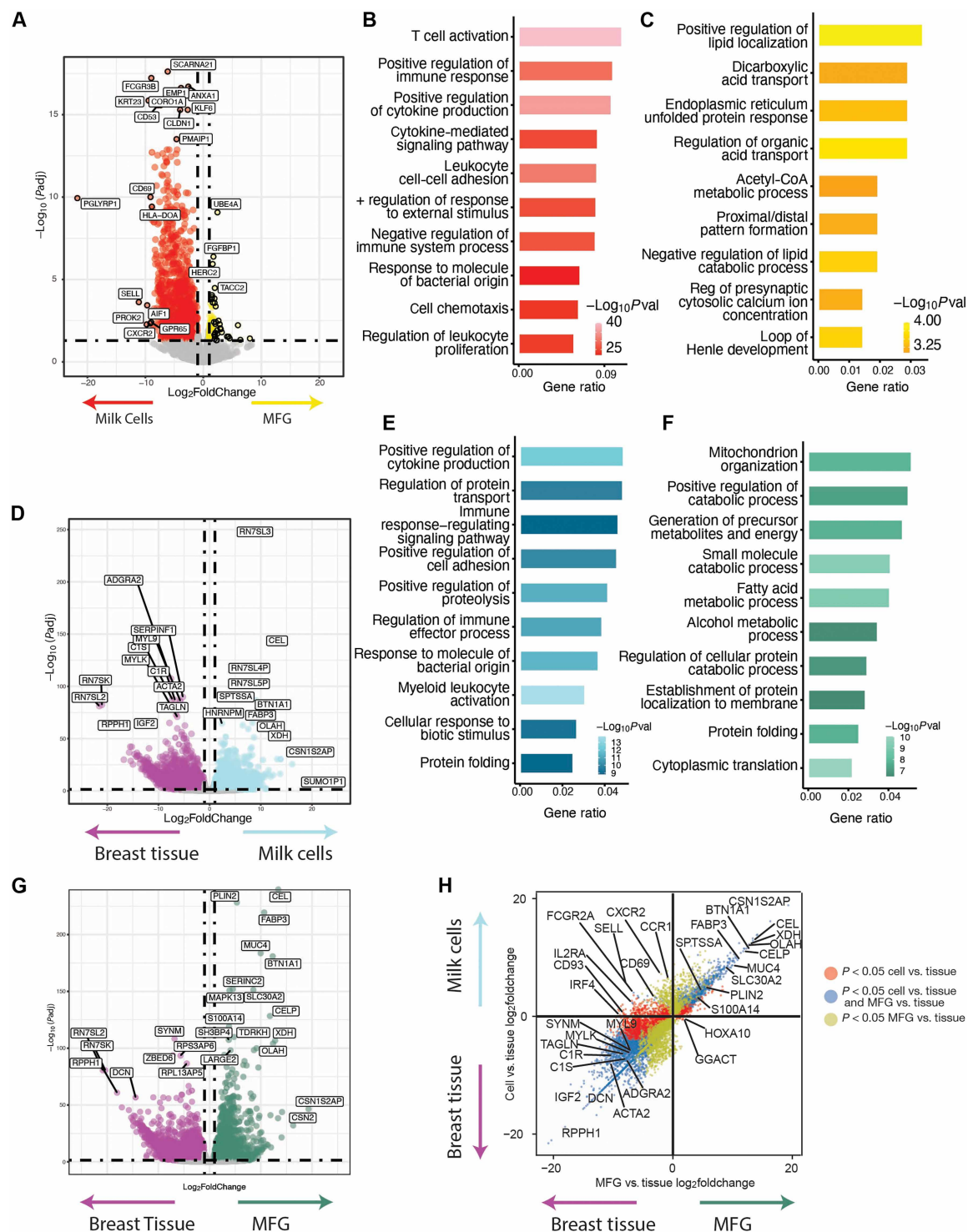


Fig. 2. Transcriptomic comparison between MFG, milk cells, and nonlactating mammary gland tissue. (A) Volcano plot showing RNA-seq DE genes between milk cells (left side) and MFG (right side) as detected by DESeq2 (51). A log-fold change >1 cutoff and BH-adjusted P value < 0.05 was used. (B and C) GO Biological Process (GOBP) pathway analysis (48) of DE genes from (A) including pathways up-regulated in milk cells (B) and in MFG (C). (D) Volcano plot of DE genes comparing milk cells (right side) and nonlactational breast tissue (left side). (E) Corresponding GOBP pathway analysis of DE genes up in milk cells from (D) (right side). (F) Corresponding GOBP pathway analysis of DE genes up in MFG from (G) (right side). (G) Volcano plot of DE genes between MFG (right side) and nonlactational breast tissue (left side). (H) Dot plot of the fold change of the DE genes from (D) and (G) showing similarity in gene expression between milk cells and MFG compared to breast tissue (blue) and DE genes in cells (red) or MFG (green). Full gene and GOBP lists in table S2. Gene ratio is the ratio of input genes that are annotated in a term (pathway of genes).

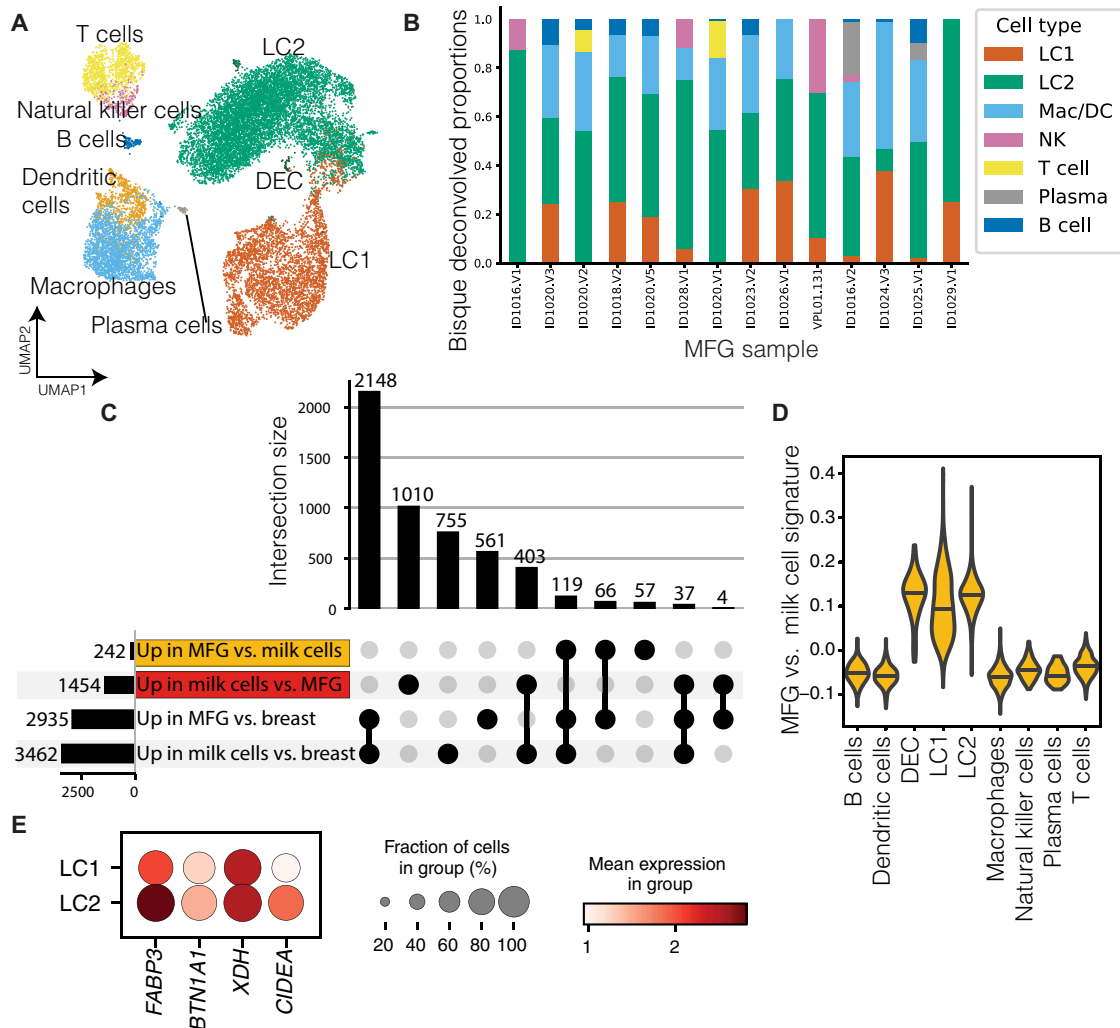


Fig. 3. The MFG transcriptome is more similar to the epithelial subcluster LC2 compared to LC1. (A) UMAP low-dimensional visualization of scRNA-seq data colored by cell type cluster. (B) BisqueRNA package (52) was used to deconvolve bulk transcriptomes from MFG into component cell type proportions as a proxy for cell type. Cells proportion as represent in MFG RNA transcript (y axis) was calculated for each participant (x axis). (C) Bar chart showing the numbers and the overlap of DE genes between milk cells, MFG, and nonlactational breast tissue. Top bars indicate intersection size between sets indicated by dark dots below; left horizontal bars indicate size of each set of genes. (D) Violin plot of the gene set signature score (see Materials and Methods), calculated from genes up-regulated in MFG compared to milk cells in bulk RNA-seq [highlighted in yellow in (C)]. The score is plotted for each cell across the cell types identified in scRNA-seq data (x axis). (E) Genes associated with MFG formation in the LC1 and LC2 epithelial clusters. Dot size represents the percentage of cells in cluster (y axis) expressing the gene at a level > 0, and color indicates the mean log₂-normalized expression of that gene in the cells of that cluster.

homeostasis and insulin secretion, and their dysregulation may potentially affect lactation as insulin signaling plays an important role in secretory differentiation in the mammary gland and in milk production (15, 57). In addition, duodenal *GLP1R* gene expression was negatively correlated with milk production efficiency traits in dairy cattle (58). Currently, little is known about the role of *GLP1R* and its signaling pathway in milk production. In addition, Perilipin 4 (*PLIN4*), a gene associated with lipid accumulation in other tissues and with breast cancer (59, 60), was found in higher levels in low milk producers (Fig. 4B). Additional genes of interest were cytokine-inducible Src homology 2-containing (*CISH*) protein that negatively regulates the Janus kinase–signal transducers and activators of transcription 5 signaling pathway (61, 62), and Lysozyme (*LYZ*) (Fig. 4, A and C).

To validate our findings, we extracted MFG RNA from additional milk samples (total $n = 39$ samples; 7 low, 24 normal, and 8 high), and performed quantitative polymerase chain reaction (qPCR) analysis on five DE genes between the milk production groups in our bulk-RNA analysis (*CISH*, *GLP1R*, *LYZ*, *KLF10*, and *PLIN4*). We calculated the fold change of each sample from the average of the normal group expression levels. We then used the mixed effect model to test whether the expression of these genes is significantly different between the milk production groups, accounting for multiple samples from the same individual and the days postpartum (different times of sample collection). When controlling for time postpartum, *PLIN4* and *GLP1R* gene expression are expected to be higher in the low production group compared with normal and to high producers (P value < 0.01) (Fig. 4D and table S5). In contrast,

Table 3. scRNA-seq cell type proportions per human milk sample.							
Milk production	Low		Normal		High		Cluster marker genes
Cell type	Proportion per sample	SD	Proportion per sample	SD	Proportion per sample	SD	
T cells	0.052	0.032	0.095	0.069	0.056	0.072	<i>ETS1 CD3E</i>
Natural killer cells	0.008	0.006	0.025	0.021	0.01	0.014	<i>NR4A2</i>
B cells	0.013	0.011	0.005	0.002	0.012	0.015	<i>MS4A1, CD79A, BANK1</i>
Plasma cells	0.003	0.004	0.002	0.002	0.0	0.0	<i>JCHAIN, IRF4</i>
Dendritic cells	0.047	0.018	0.041	0.034	0.08	0.116	<i>CSF2RA</i>
Macrophages	0.194	0.124	0.155	0.145	0.066	0.069	<i>CD68</i>
LC1	0.167	0.071	0.094	0.04	0.288	0.12	<i>CLDN4, CLDN3, NECTIN4</i>
LC2	0.513	0.117	0.571	0.236	0.486	0.208	<i>CIDEA, FOLR1</i>
Dividing epithelial cells	0.003	0.002	0.013	0.013	0.001	0.001	<i>TUBB4B, STMN1</i>

KLF10 gene expression is expected to be higher in high producers compared to low producers and tends to be higher compared to normal producers. We also found that *CISH* and *LYZ* expression were not significantly different between the groups over time (table S5 and fig. S2A).

scRNA-seq identifies changes in cell type proportions that are associated with milk production and transcriptional changes in specific cell populations

We next assessed whether milk production is associated with transcriptional changes in other milk cell types that are not present in the MFG transcriptome using scRNA-seq. To determine whether milk cell type composition reflects milk production levels, we used our scRNA-seq data from 11 cryopreserved milk cell samples collected from our cohort (four low, four normal, and three high production). Samples were chosen to match across time postpartum and to match the samples used for bulk RNA-seq when possible (Table 2). Using the nine clusters as previously shown (Fig. 3A), we found high variability in the composition of immune and epithelial cells between samples (Fig. 5A and Table 3); thus, we compared the composition of immune and epithelial cells separately. Using scCODA (63), a method for differential abundance of cell types, we found that the ratio of LC1s to LC2s differs between samples from high production and normal production groups, with a larger proportion of LC1 cells in high production [credible effect with false discovery rate (FDR) = 0.05; Fig. 5B]. While evaluated as credible using this analysis, the small sample size in this comparison leaves this comparison as a topic for further exploration in larger datasets. The low production group also exhibited a higher LC1/LC2 ratio compared to the normal group, although this difference was not statistically significant. Our analysis suggests that the proportion of LC2 relative to LC1 in milk does not directly correlate with the milk production phenotype, as both aberrant production groups displayed elevated LC1 levels. Further studies with larger sample sizes are necessary to determine whether the LC1/LC2 ratio could serve as a potential indicator of milk production dysregulation.

Among the epithelial cells, we identified eight subclusters, including four LC1 subtypes (named LC1, A to D), three LC2 subtypes (named LC2, A to C), and a small population of dividing cells (DEC) (fig. S3). When comparing the proportions of these subpopulations, we found that the proportion of LC2-C cells was higher in normal suppliers compared to the low or high groups (credible effect with FDR = 0.05; Fig. 5C). No other changes were observed for the epithelial cell type proportions. Since the functions of the different milk cell subtypes are unknown, we decided to further characterize whether the changes in milk cell LC2-C proportions are reflected in changes in milk composition. On the basis of our previous finding of similarities between MFG and LC2 cells, we measured total milk fat content using the creatatocrit device (64, 65) in the 40 milk samples from our cohort, including the samples used for the scRNA-seq assay. There was no significant difference in fat content (gram per liter) in samples provided by normal producers compared to low producers (Fig. 5D). Among samples with matched scRNA-seq data, fat content was positively correlated with LC2-C cell proportions (Spearman $P \leq 0.05$; Fig. 5E). Our results identified higher proportions of LC2-C epithelial cells in normal milk suppliers, which correlated with increased milk fat content in these samples. In addition, we found that LC1-Golgi/long noncoding RNA (lncRNA) was negatively associated with fat content (Spearman $P < 0.02$). No other cells types were correlated with milk fat content. To better understand the function of milk epithelial cell types, this correlative relationship between LC2-C and that LC1-Golgi/lncRNA cells and milk fat content should be further studied in future work.

Differential expression analysis between production groups within each of the epithelial cell clusters found several genes consistent with those identified in the bulk RNA-seq (table S6). For example, we identified up-regulation in LC2 cells in high producers of the *LYZ* gene that encodes lysozyme, which is an antibacterial bioactive component found in human milk (Fig. 5F; all BH-adjusted $P \leq 0.05$) (66). Several genes were increased in high suppliers in the LC2 cluster but not in the MFG transcriptome comparisons, including milk component synthesis-related genes (*AZGP1*: zinc

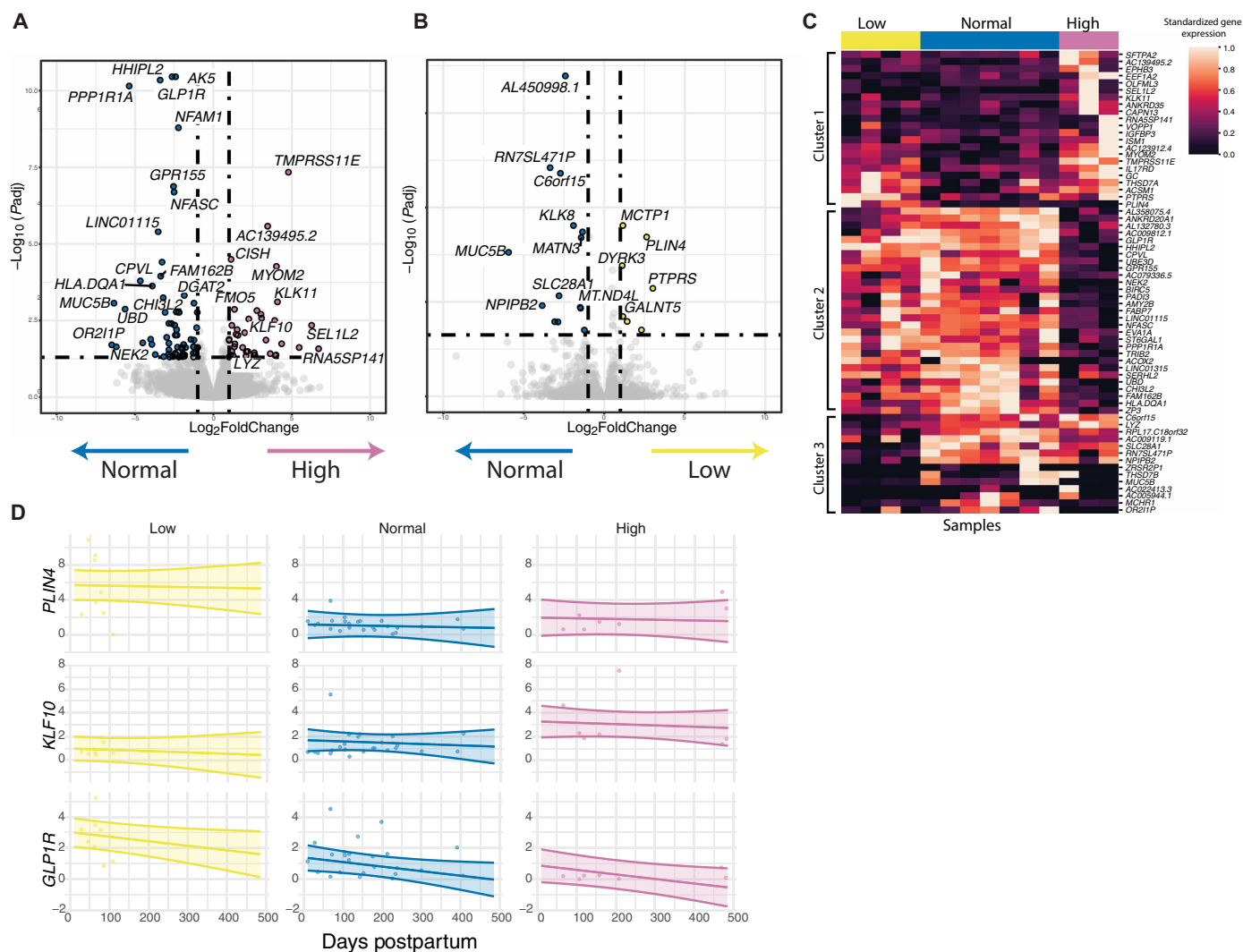


Fig. 4. Transcriptomic changes in low and high milk production. (A) Volcano plot of DE genes between individuals with high ($n = 3$) and normal ($n = 7$) milk production ($\text{log-fold change } \text{logFC} > 1$ and BH-adjusted $P < 0.05$). (B) Volcano plot of DE genes between individuals with low ($n = 4$) and normal ($n = 7$) milk production ($\text{logFC} > 1$ and adjusted $P < 0.05$). (C) Heatmap showing DE genes with $\text{log}_2\text{foldchange} > 2.5$ between the different milk production groups colored by column-standardized gene expression. (D) qRT-PCR results for 39 MFG RNA samples at different time points postpartum. The y axis represents the fold change in gene expression from the mean of normal producers normalized to *GAPDH* expression. A linear mixed effect model was used to determine differences in gene expression between milk production groups controlling for days postpartum. Lines represent the fixed-effect regression with confidence intervals.

binding and lipid metabolism; *B4GALT1*: lactose synthesis; *NUCB2*: calcium level maintenance; *KLK6*: serine protease; and *GC*: vitamin D binding protein) as well as genes related to mammary gland structure (*PERP*: desmosomes; *ODAM*: epithelial cell proliferation and wound healing; and *CEL*: milk synthesis related glycoprotein; Fig. 5F). In the same LC2 cluster, several genes were increased in both high and low suppliers compared to normal suppliers, including genes known to be related to milk synthesis (*LPO*, *NAAA*, and *NUCA1*; Fig. 5F).

Immune cell subclustering identified nine clusters including three myeloid and six lymphoid cell clusters (fig. S4). Most of the immune cells were myeloid cells in each sample (fig. S4B). Of the myeloid cells, we identified a cluster of macrophages, DCs, and milk macrophages, a macrophage subset that also expresses milk production genes such as *CSN3*, and has been described in previous human

milk scRNA-seq studies (43) and murine mammary gland studies (67, 68). In the lymphoid cells, we identified clusters of NK cells, gamma delta (GD) T cells, CD8^+ T cells, CD4^+ T cells, B cells, and plasma cells using canonical marker genes (fig. S4).

We next looked for differences in immune cells related to milk production, focusing on differences within myeloid cells as they are the most abundant immune cells type present in human milk. In immune cells, we found an increase in DCs in individuals with high milk production compared to normal or low suppliers (credible effect with $\text{FDR} = 0.05$; Fig. 6A and fig. S4). In addition, we found a trend of increased macrophage proportion in low producers. These cells also showed increased expression of genes related to $\text{TNF-}\alpha$ signaling via *NFKB* and genes related to the production of cytokines and inflammation, such as *IL1B*, *CCL4*, *CXCL3*, and *CXCL8* (Fig. 6B and table S6). We scored the macrophage cells on the full $\text{TNF-}\alpha$

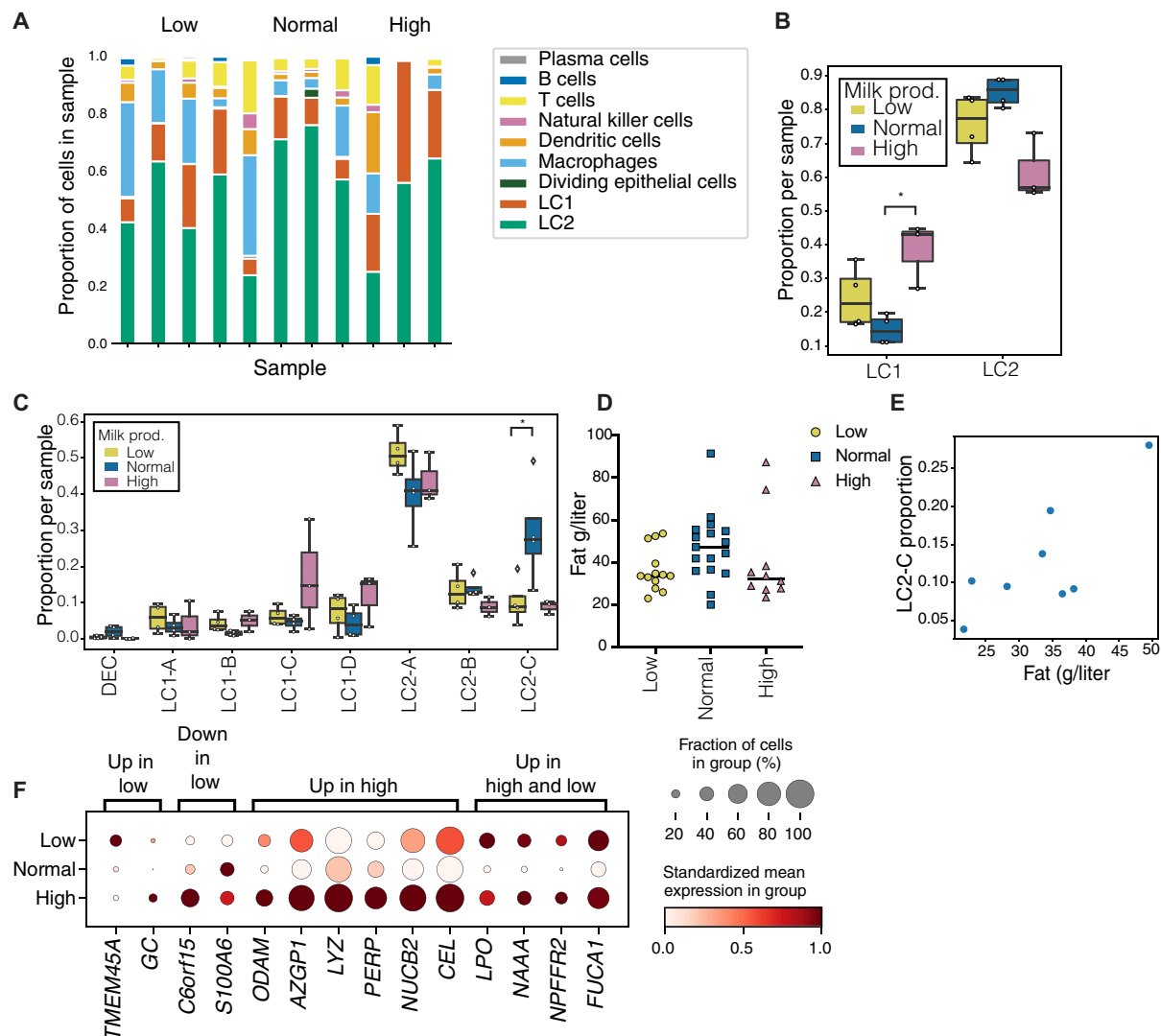


Fig. 5. Cellular composition changes in the different milk production groups. (A) Proportion of cells in each sample as revealed from scRNA-seq. (B) Ratio of LC1 and LC2 subtype proportion for each sample divided into groups based on milk production. (C) Proportion of epithelial subclusters in each sample. (D) Milk fat content in samples from each milk production level. (E) Scatter plot of fat content in each sample compared to the proportion of LC2-C cells among other epithelial cells. Spearman correlation $r = 0.775$, $P \leq 0.03$. (F) DE genes in the LC2 clustered between low, normal, and high milk producers. Dot size represents percentage of cells in cluster (y axis) expressing the gene at a level > 0 and color indicates the mean $\log_2$ -normalized expression of that gene in the cells in that cluster. Asterisks (*) represent a credible effect of difference using scCODA differential abundance.

signaling via the nuclear factor κ B (NF- κ B) hallmark pathway and observed a trend of increased gene set score in three of four individuals with low milk production compared to normal- or high-milk producers (ANOVA, $P \leq 0.1$; Fig. 6C).

Effect of milk production on the infant's microbiome

To investigate whether maternal milk production is affected by the composition of the maternal microbiome, or affects the infant gut microbiome, we collected stool samples from mothers and their infants. We analyzed 20 stool samples from 10 infants of participants with low and normal milk production and 16 samples from their 10 mothers collected at the time of milk collection. Samples were collected at multiple time points from 28 to 150 days postpartum (Table 2), and before the introduction of complementary foods to the

infant diet. We conducted metagenomic sequencing on all samples and analyzed the data using a unique MetaPhlAn database (69). As only two individuals with high milk production provided stool samples, these samples were not included in the group-wise comparisons.

We next set out to characterize the infant microbiome for differences between the milk production groups. We used principal coordinate analysis (PcoA) and k-means clustering ($k = 3$) to characterize the dominant species in the infant gut microbial population in our cohort. We found three main clusters: (i) samples dominated by *Bifidobacterium breve* (*B. breve*), (ii) samples dominated by *Bifidobacterium longum* subsp. *infantis* (*BL. infantis*), and (iii) samples dominated by other bacteria that are typically less common in infants. Samples dominated by *B. breve* or *BL. infantis* clustered separately from samples dominated by other bacteria, revealing a distinct

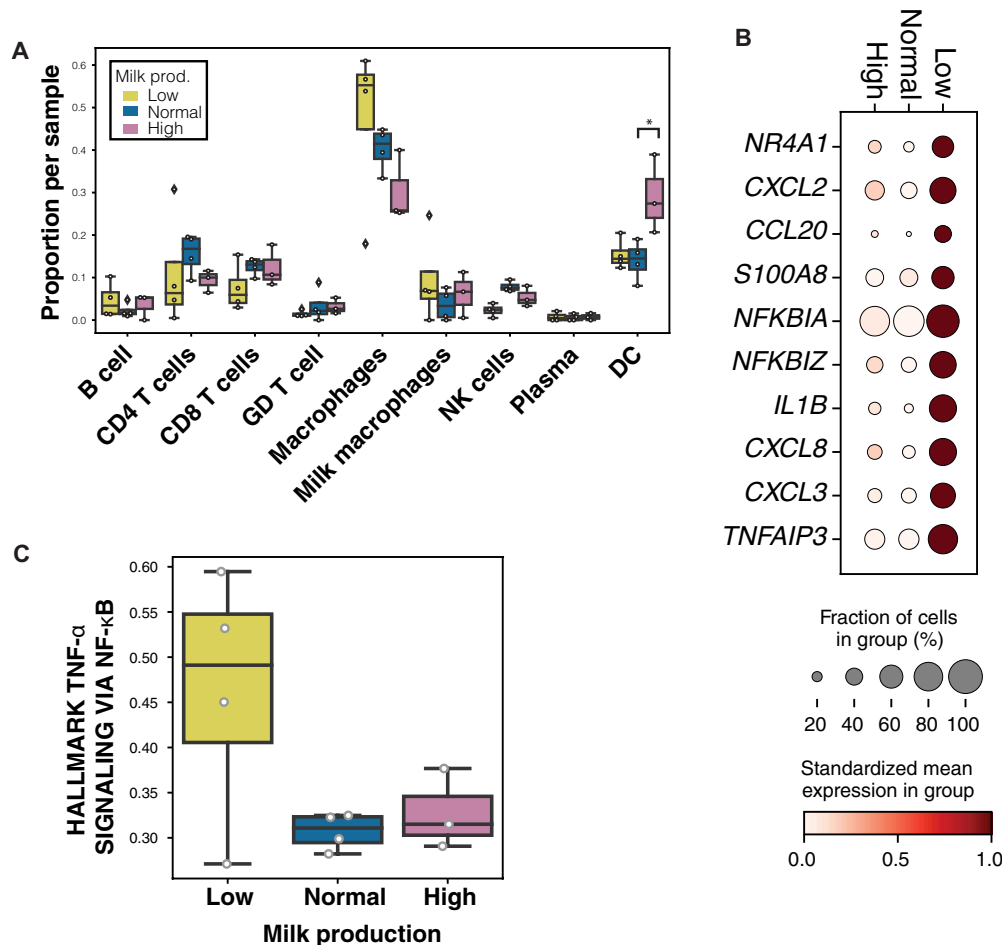


Fig. 6. Transcriptional changes in production groups originate primarily from LC2 cells and macrophages. (A) Proportion of immune cells in each sample. (B) DE genes in the macrophages cluster between low, normal, and high milk producers. Dot size represents percentage of cells in cluster (y axis) expressing the gene at a level > 0 and color indicates the mean log₂-normalized expression of that gene in the cells in that cluster. (C) Mean gene set scores of Hallmark TNF-α signaling pathway in macrophage cells per sample in the different milk production groups computed using Scanpy gene scoring. Asterisks (*) represent a credible effect of difference using scCODA differential abundance.

microbial population (Fig. 7A). This separation can be quantified by a lower Shannon diversity (Fig. 7A; point size) in the samples dominated by *B. breve* and *BL. infantis*. One possible driver of this separation is that, when *B. breve* and *BL. infantis* are dominant, they inhibit or interfere with successful growth of other bacteria (69).

Samples from infants of mothers with low and normal milk production were found in all clusters and there was no significant difference in the abundance of any bacteria species between those groups (fig. S5). As expected, infant Shannon diversity index was lower than the maternal index (Fig. 7B). We found no difference in the maternal microbiome diversity or genus-level composition between individuals in different milk production groups in our cohort (using linear association models, while controlling for multiple samples from the same individual; fig. S5). The Shannon diversity index of infants nursed by individuals with low milk production was slightly higher than that of those with normal milk production, but this difference was not statistically significant (P value = 0.084) (Fig. 7B).

Since infants of low milk producers were more frequently supplemented with infant formula compared to those of normal producers

(Fig. 7B; indicated by point color), we tested whether infant formula supplementation affected the microbiome diversity. We found that exclusively breastfed infants had lower Shannon diversity compared to those who were breastfed but also supplemented with infant formula (P value = 0.023; fig. S5). In summary, this analysis showed a trend of higher microbiome diversity in infants nursed by low compared to normal milk producers. This was likely driven by infants receiving formula supplementation (70), who constitute most in the low producers' group, which was correlated with a more diverse microbiome compared to exclusively breastfed infants.

DISCUSSION

In this study, we present a comprehensive catalog of the transcriptomic and cellular changes occurring during lactation. Our findings highlight the potential of MFG RNA as a biomarker for epithelial cells, as it largely reflects the transcriptomic profile of these cells. Notably, we observed that MFG RNA contains very few unique transcripts when compared to the RNA profile of milk cells, which are commonly used as a representative of the mammary gland

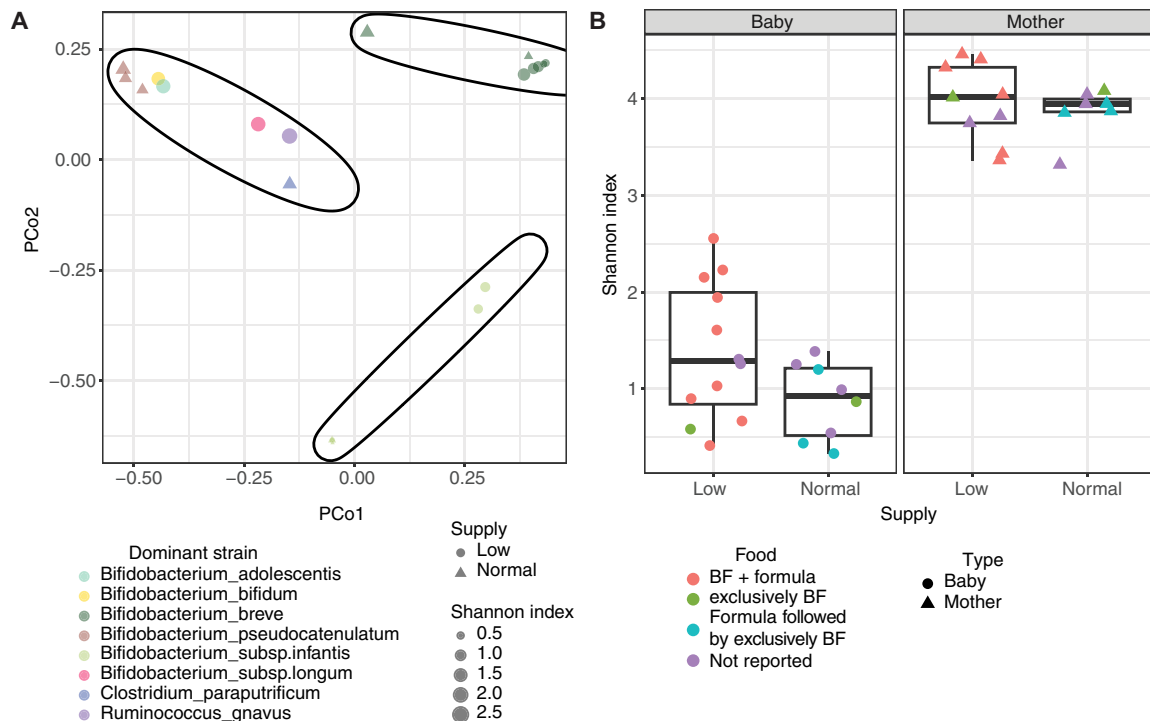


Fig. 7. Characterization of the maternal and infant gut microbiome from different milk production groups. (A) Infant gut microbiome samples were grouped using *k*-means clustering with *k* = 3. The samples are color-coded by their dominant bacteria, with point shapes indicating maternal milk supply (circles for low and triangles for normal milk production) and point sizes representing the Shannon diversity of each sample. (B) Shannon diversity in infant and maternal stool samples from this cohort (including only samples up to 150 days) is shown. Points are colored according to the infant feeding type at the time of sampling.

during lactation. This suggests that MFG RNA primarily originates from mammary epithelial cells and may serve as a useful tool for studying epithelial-specific gene expression in lactation research. Our analysis also revealed that MFG RNA is not enriched in immune-related transcripts, indicating a low contribution from immune cells. This distinction is particularly important for studies focusing on immune function during lactation, as MFG RNA does not provide a reliable representation of immune cell activity. Instead, researchers investigating immunological aspects of lactation may find milk cells to be a more suitable source of RNA.

In addition, this study analyzes changes in transcriptomic in MFG and milk cells collected from fresh milk samples from individuals with low, normal, and high human milk production. Previous studies that examined changes in the milk cell transcriptome of low milk producers were limited by confounding from a higher BMI among individuals with low milk production compared to the normal production group (13, 14). High BMI may lead to greater changes in milk production that are dependent on systemic inflammatory state as well as local inflammation in the mammary gland. We observed a trend of increased TNF- α signaling in macrophage cells in low producers in our scRNA-seq analysis, but this pathway was not different in our bulk RNA-seq analysis in MFG. In our cohort, there was no significant difference in BMI between the study groups, which may explain the smaller differences in the milk inflammatory markers between our study groups compared to previous studies (71). Another study that used frozen milk samples from eight participants found no differences in gene expression between individuals with low and high milk production (72). This might be

due to the use of frozen milk samples, which is the cause fast degradation of RNA and reduce significantly the RNA integrity score. The same study reported that individuals with low milk production ($n = 4$) have more depression and anxiety compared to individuals with high production ($n = 4$) (72); however, depression did not differ between the groups in our study and we did not assess anxiety. Since breastfeeding is a multifactorial process that depends on many environmental and internal factors (73), it will be essential in the future to quantify differences in these confounders within and between the study groups.

Our analysis provides important insights into changes in the transcriptional signature and cell type composition of the mammary gland under aberrant milk production. We found that most genes that are required for milk production, including milk proteins (caseins and lactalbumin) and milk fat secretion, are not DE between the milk production groups and that the conserved cellular milk production machinery functions equivalently across all groups. Our findings also identify gene targets that should be further studied to understand their role in human milk production.

Our study examined milk cell population proportions in samples collected from individuals with low, normal, and high milk production. In terms of total cell proportion, we did not observe any significant differences between our study groups, possibly because of the high variability of immune cells in milk samples and the small sample size in our study. Focusing on epithelial cells alone, the LC2-C subcluster (high *KRT*) was depleted in low and high milk production groups. Along with the correlation between milk fat concentrations and LC2-C cells percentages, these results suggest a

role for LC2-C cells in the fat content in milk and should be further characterized. In immune cells, DCs were increased in the high production group. The DC population was previously shown to increase during involution in mice (74); our results suggest that high milk production might lead to similar cellular changes as in involution due to the constant attempt to reduce milk production in these individuals.

Individuals with low milk production are usually using breast milk pumps to try and increase their milk supply and are often recommended to add infant formula to their infant diet. These different feeding behaviors might also affect the infant microbiome (75), and we therefore aimed to identify whether impaired milk production is associated with changes in the infant microbiome. Our findings revealed no differences in microbiome diversity or species composition between mothers and infants with low versus normal milk production. Our results align with earlier studies from larger cohorts that demonstrated that combining formula feeding with breastfeeding tends to increase microbial diversity compared to exclusive breastfeeding; however, partial and exclusively breastfeeding reduces microbial diversity compared to exclusively formula feeding (70). These findings further support the messaging that individuals with low milk production should be encouraged to continue partial breastfeeding to support healthy infant microbiome development.

Our study has some important limitations. This is a small prospective study, and samples were collected at a relatively late stage in lactation because of the need for lactation consultants to examine each participant before recruitment. Furthermore, larger studies that follow up on lactation in the participants from childbirth through the mature milk stage are needed to better understand the changes occurring just after birth and their effect on establishing lactation and breastfeeding outcomes. In addition to the changes detected in the milk cell populations, it is possible that there are other changes in cell populations that are not shedding into milk (basal cells and stromal cells) (44), which affect the ability individuals to produce or secrete milk. Moreover, hormonal regulation and other upstream factors may also play a role in milk production (62); these factors were not examined in our analysis and cannot be detected using milk cells. Our microbiome data are limited because they do not contain samples of individuals with high milk production, and by the small sample size of the low and normal producers.

In summary, our study provides a unique database of gene expression during lactation and that can be used to find specific genes that might promote or inhibit milk production in cases of aberrant milk production. However, the specific role that these genes play in human milk production needs further research. Our study will pave the way for more research in the area of milk production using genomic characterization, and future studies will assist in our understanding, diagnosis, and treatment of breastfeeding difficulties.

MATERIALS AND METHODS

Cohort

Participant cohort and data collection

Milk samples for this study were collected in two independent cohorts and the institutional review board of the UCSF, approved these studies (UCSF Milk production study #19-29297 and COVID-19 Vaccine in Pregnancy and Lactation (COVIPAL) cohort study #20-32077). Written informed consent was obtained from all study

volunteers. The COVIPAL samples were collected and stored as previously described (76) and were used as additional samples for fat layer RNA-seq.

Clinical data

Data collection

Data were collected through an online questionnaire that was sent to participants via email using REDCap. To assess PIMS, mothers in both cohorts were asked to report whether they feel that they produce too much/normal/not enough breast milk to meet their infant's needs. In the UCSF Milk supply study, mothers were also assessed by lactation consultants to address any obvious reason for low/high milk production (table S1) and were also asked whether their infants are satisfied with the volume of milk they produced.

Statistical analysis

Survey data were exported from RedCap and analyzed on R version 4.3.2 (2023-10-31) using the CompareGroups (77) function to generate Table 1.

Milk sample collection and processing

Fresh human milk samples were self-collected by participants into sterile containers using Ameda Dual HygieniKit Milk Collection System provided to the participants. For each collection, the mothers were instructed to use a new/autoclaved sterile kit, wash hands before pumping, and empty both breasts. Ten to thirty milliliters of milk were collected from each breast for analysis. Milk from each breast was processed separately. Mothers who prefer using different pumps were asked to sterilize the pump parts at home and document it. Samples were transported on ice from the participant's home to the laboratory for processing. Milk samples were collected and processed as soon as possible and no longer than 2 hours after expression by the study staff. In the laboratory, whole milk was aliquoted, and the rest was centrifuged at 800g for 20 min in 4°C to pellet the cells and separate the fat layer. Fat layer was removed using a sterile spoon and mixed with RLT lysis buffer or TRIzol for RNA extraction using the RNeasy kit (Qiagen). Supernatant was aspirated and aliquoted, and milk cells pellet was resuspended in phosphate-buffered saline (PBS) with 0.04 to 1% bovine serum albumin (BSA). Cells were washed once or twice and were analyzed immediately or cryopreserved using CryoStor CS10 freezing media (STEMCELL Technologies). For milk fat content, frozen milk samples were thawed to room temperature and were gently mixed to homogenize before measured. Milk fat content was measured using Creamatocrit Plus (EKF Diagnostics). One outlier sample that had fat levels above the normal range expected for mature human milk (8 g/dl) (78) was excluded from further analysis related to fat composition.

Bulk RNA-seq

Total RNA samples from milk fat layer of milk cells were sequenced at Novogene Co. Ltd. Libraries were generated using the SMARTer V2 or V3 kits. The sequencing was performed using a paired end 150-base pair (bp) strategy on the Illumina platform (Illumina NovaSeq 6000). Sequencing reads were aligned to reference human genome hg38 using STAR aligner. Normal breast RNA-seq data were downloaded from GEO (PRJNA292118) (49). Read counts were normalized to transcripts per kilobase million. Quality control matrixes and details about sequence reads/samples are in table S8.

Bulk differential analysis

Differential expression between paired milk cell and MFG transcriptional data was performed using DESeq2 (51) with designs “~ subject+layer” for milk cells versus MFGs, “~tissue_origin” for cells or MFGs versus breast tissue, and “~baby age + milk_supply” for milk supply. For milk cells versus MFGs, genes with adjusted *P* value less than 0.01 and log₂foldchange greater than 2 or less than -2 were included in enrichment analyses. For milk supply comparisons, genes with adjusted *P* value less than 0.05 and log₂foldchange greater than 1 or less than -1 were included in enrichment analyses.

Bulk gene set enrichment analyses

GO enrichment was run using the clusterProfiler package in R with function enrichGO using biological processes ontologies with BH *P* value correction for multiple tests and genes expressed in dataset as background followed by the simplify function to remove redundant GO hits (48).

qPCR for gene expression

MFG were preserve in RLT-buffer or RNAlater and used for RNA isolation using RNeasy kit (Qiagen, #74134). Reverse transcription was performed with 1 µg of total RNA using qScript cDNA Synthesis Kit (Quantabio, 95047) following the manufacturer’s protocol. Quantitative reverse transcription PCR (qRT-PCR) was performed on QuantStudio 6 Real Time PCR system (Thermo Fisher Scientific) using PerfeCTa SYBR Green FastMix, Low ROX (QuantaBio, 95074). Statistical analysis was performed using ddct method with *Gapdh* primers as control (see primer sequences in table S7). Gene expression results were generated using mean of all participants with normal milk production as the reference value.

Linear mixed effect model was generated in R using the lmerTest package and the following formula: gene, ~ Supply_group + Days_pp + (1 | “Study ID”). The emmeans function was used for post hoc analysis to compare the means of the difference milk production groups using Tukey methods for comparing a family of three estimates, with degree-of-freedom method: Kenward-roger (results in table S5).

Single-cell RNA-seq

Cryopreserved cells were thawed and washed with 10 ml of Mammary Epithelial Cell Growth Medium (PormoCell) and later with 2 ml of PBS + 1 to 2% BSA. Live cells were counted using acridine orange/propidium iodide staining using DeNovix CelloDrop cell counter. Twenty-five thousand cells from each sample were loaded on the 10X single cell chip. Chromium Next GEM Single Cell 3’ Kit v3.1 was used following the manufacturers’ protocol.

scRNA-seq analysis

Alignment

Samples were aligned to hg38 using 10x Genomics CellRanger v6.1.2.

Preprocessing

Using the Scanpy package v1.9.6, data were filtered to remove cells with fewer than 200 genes and genes expressed in fewer than 10 genes (79). Expression was normalized to 1×10^4 total counts per cell, and the log base 2 + 1 was taken.

Clustering analysis

Following a standard Scanpy pipeline (79), highly variable genes were selected using the scanpy highly_variable_genes function with batch key = sample. Twenty-six principal components and 10 neighbors

were used to construct the neighborhood graph. Clustering was performed iteratively beginning with identification of broad immune and epithelial cell clusters using Leiden clustering followed by reselection of variable genes on epithelial and immune subsets then followed by further subclustering on each broad cell type: T cells, B cells, myeloid cells, LC1s, and LC2s. In each subclustering stage, clusters were identified as doublet cells and removed if they contained marker genes from distinct lineages. During subclustering, Harmony was used for batch integration (80). Cell type labeling was done on the basis of comparison of marker genes to those identified in prior human milk cell datasets (42–44).

Differential expression

To define marker genes for each subcluster as genes that identify cells in that cluster regardless of their sample of origin as well as genes specific to milk production level groups that are reproducible across individual samples, we used pseudobulk differential expression analysis as previously described (43, 81–83). Briefly, we generated pseudobulk counts for the cells in each subcluster in each sample by summing the raw counts for each group and used these to identify DE genes between subclusters and between conditions in a sample aware manner. We removed subcluster/sample pools with fewer than 10 cells from these comparisons. Using the DESeq2 package (51), we performed differential expression analyses using the Wald statistical test with the design formula “~ donor + celltype” to identify cell type marker genes and “~ donor + baby age + Milk supply” to identify milk supply differential genes for each cluster (51). Marker genes were then filtered for adjusted *P* value <0.05, log₂foldchange > 0.4, and proportion of cells in cluster expressing the gene >0.4. Milk production differential genes were filtered for adjusted *P* value <0.05 and proportion of cells in cluster expressing the gene >0.1 full differential gene lists are available in tables S1 to S4.

Gene set enrichment analysis

Functional enrichment analysis on these differential genes was performed using Enrichr using the gseapy package with the gene set GO_Biological_Processes_2023 (84, 85). For the macrophage cluster, TNF-α signaling via NF-κB score was calculated using the full list of genes with the scanpy function “score_genes.”

Statistical analysis of scRNA-seq results

The scCODA package was used to test statistically differentially abundant cell type proportions in the scRNA-seq data between low, normal, and high production groups (63). This approach compares the ratio of cell type abundance within a sample to a reference cell type. Since this study did not include a clear cell type whose abundance should be expected to stay stable, the scCODA test was run iteratively with each cell type as the reference cell type, and a cell type was considered differentially abundant if the scCODA test identified it as significantly differential with more than half of the other cell types as the reference. This test was run on the cell type composition across all cell clusters as well as within just the immune cells and just the epithelial cell subclusters separately. All RNA quality control matrixes for the bulk and scRNA sequence analysis are in table S8.

Stool samples collection

Maternal and infant stool samples were collected on the day before, on the day of, or the day after milk sample collection. Maternal stool samples were collected using a Feces Catcher (Zymo Research, R1101-1-10) following the manufacturer’s instructions and then immediately scooped into DNA/RNA Shield fecal collection tubes

(Zymo Research, R1101-E). Infant stool samples were collected directly from the diaper (with mothers instructed not to apply cream at the time of sample collection) immediately after bowel movement into DNA/RNA Shield fecal collection tubes (Zymo Research, R1101-E). Samples were brought to the laboratory within 3 hours of collection or kept frozen in a household freezer if collection was scheduled for a later time point. Samples were shipped to the laboratory on ice and stored at -80°C until analysis.

DNA extraction

Stool samples underwent DNA extraction performed by the Microbial Genomics Core at the UCSF, using a modified cetyltrimethylammonium bromide (CTAB) buffer-based protocol as described in a published article (86). Briefly, each frozen stool pellet was suspended in 500 μl of 5% CTAB extraction buffer in a Lysing Matrix E tube (MP Biomedicals) by vortex and incubating at 65°C for 15 min. Then, 500 μl of phenol:chloroform:isoamyl alcohol (25:24:1) was added, followed by bead-beating at 5.5 m/s for 30 s and centrifugation at 16,000g for 5 min at 4°C . The aqueous phase (~ 400 μl) was transferred to a new 2-ml Eppendorf tube, and the extraction was repeated with an additional 400 μl of 5% CTAB buffer, yielding ~ 800 μl from repeated extractions. Chloroform was then added in equal volume, mixed, and centrifuged (16,000g for 5 min). The resulting aqueous phase (~ 500 μl) was combined with 2 volumes of 30% polyethylene glycol/NaCl solution and stored at 4°C overnight to precipitate DNA. Samples were then centrifuged (3000g for 60 min), washed twice with ice-cold 70% ethanol, and resuspended in 100 μl of sterile water. DNA from each sample was quantified using the Qubit 2.0 Fluorometer with the double-stranded DNA (dsDNA) BR Assay Kit (Life Technologies, #Q32853).

Shotgun metagenomic library preparation

Shotgun metagenomic DNA library preparation was performed using the Illumina DNA Prep Kit (Illumina, #20060059) according to the manufacturer's instructions. A total of 150 ng of input DNA from each sample was used for library preparation, which involved enzymatic fragmentation (tagmentation), index-adaptor ligation, and amplification. The Illumina libraries were quantified using the Qubit 2.0 Fluorometer with the dsDNA High Sensitivity Assay Kit (Life Technologies, #Q32854) and pooled at equal molar concentrations. The final pooled libraries were submitted for sequencing at the Center for Advanced Technology at UCSF, where they were sequenced using the Illumina NovaSeq 6000 in a 2×150 -bp paired-end run protocol.

Metagenomic analysis

Host reads were removed using an in-house pipeline by aligning reads to the human genome by Bowtie2 (2.4.5-1) (87). Samples were filtered and trimmed for Nextera adaptors using fastq-mcf, ea-utils (1.05) (88). Taxonomic profiling was done using MetaPhlAn4 (89) with a custom database that allows quantification of *B. longum* subspecies (69). Further analysis was done using an in-house R (4.2.2) script using dplyr (1.1.2) (90), tidyr (1.3.0) (91), and tidyverse (2.0.0) (92). Plots were created using ggplot2 (3.4.2) (93), colors were used from RColorBrewer (1.1-3) (94) and pals (95) (1.7). Alpha and beta diversity were calculated using "diversity" (Shannon index) and "vegdist" (Bray-Curtis dissimilarity) from the vegan (2.6-4) (96) package, and the PcoA was created using the ape (5.7-1) (97) package.

Independent *t* test was performed to test between groups when mentioned using the R function "t.test." In addition, the "Maaslin2" (98) R package was used to perform linear models to find associations between breast milk supply and species in the infant gut. The individual was set as a random factor to account for the effect of each mother-infant pair.

Supplementary Materials

The PDF file includes:

Figs. S1 to S5

Legends for tables S1 to S8

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S8

REFERENCES AND NOTES

1. J. Y. Meek, L. Noble, Section on Breastfeeding, Policy statement: Breastfeeding and the use of human milk. *Pediatrics* **150**, e2022057988 (2022).
2. World Health Organization, "Breastfeeding." <https://who.int/health-topics/breastfeeding>.
3. UNICEF Data, "Breastfeeding," (2021); <https://data.unicef.org/topic/nutrition/breastfeeding/>.
4. CDC, Results: Breastfeeding rates, *Centers for Disease Control and Prevention* (2023). https://cdc.gov/breastfeeding/data/nis_data/results.html.
5. L. Gatti, Maternal perceptions of insufficient milk supply in breastfeeding. *J. Nurs. Scholarsh.* **40**, 355–363 (2008).
6. R. Galipeau, A. Baillot, A. Trottier, L. Lemire, Effectiveness of interventions on breastfeeding self-efficacy and perceived insufficient milk supply: A systematic review and meta-analysis. *Matern. Child Nutr.* **14**, e12607 (2018).
7. A. J. Kennedy, Breast or bottle - The illusion of choice. *N. Engl. J. Med.* **388**, 1447–1449 (2023).
8. W. Brodribb, ABM clinical protocol #9: Use of galactagogues in initiating or augmenting maternal milk production, second revision 2018. *Breastfeed. Med.* **13**, 307–314 (2018).
9. K. M. Rasmussen, Association of maternal obesity before conception with poor lactation performance. *Annu. Rev. Nutr.* **27**, 103–121 (2007).
10. K. M. Rasmussen, C. L. Kjolhede, Prepregnant overweight and obesity diminish the prolactin response to suckling in the first week postpartum. *Pediatrics* **113**, e465–e471 (2004).
11. L. R. Kair, T. T. Colaizy, When breast milk alone is not enough: Barriers to breastfeeding continuation among overweight and obese mothers. *J. Hum. Lact.* **32**, 250–257 (2016).
12. S. Lyons, S. Currie, D. M. Smith, Learning from women with a body mass index (Bmi) ≥ 30 kg/m² who have breastfed and/or are breastfeeding: A qualitative interview study. *Matern. Child Health J.* **23**, 648–656 (2019).
13. R. E. Walker, K. J. Harvathine, A. C. Ross, E. A. Wagner, S. W. Riddle, A. D. Gernand, L. A. Nommsen-Rivers, Fatty acid transfer from blood to milk is disrupted in mothers with low milk production, obesity, and inflammation. *J. Nutr.* **152**, 2716–2726 (2023).
14. R. Walker, A. Gernand, F. Schozer, P. Tso, E. Wagner, S. Riddle, L. Ward, A. Thompson, L. Nommsen-Rivers, TNF- α concentrations in serum and human milk of mothers with insufficient versus sufficient milk production. *Curr. Dev. Nutr.* **5**, 826 (2021).
15. L. A. Nommsen-Rivers, Does insulin explain the relation between maternal obesity and poor lactation outcomes? An overview of the literature. *Adv. Nutr.* **7**, 407–414 (2016).
16. J. C. Kent, E. Ashton, C. M. Hardwick, A. Rea, K. Murray, D. T. Geddes, Causes of perception of insufficient milk supply in Western Australian mothers. *Matern. Child Nutr.* **17**, e13080 (2021).
17. H. M. Johnson, A. Eglash, K. B. Mitchell, K. Leeper, C. M. Smillie, L. Moore-Ostby, N. Manson, L. Simon, Academy of Breastfeeding Medicine, ABM clinical protocol #32: Management of hyperlactation. *Breastfeed. Med.* **15**, 129–134 (2020).
18. Y. Ma, M. Z. Khan, J. Xiao, G. M. Alugongo, X. Chen, T. Chen, S. Liu, Z. He, J. Wang, M. K. Shah, Z. Cao, Genetic markers associated with milk production traits in dairy cattle. *Agriculture* **11**, 1018 (2021).
19. B. Grisart, W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, R. Snell, Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res.* **12**, 222–231 (2002).
20. P. Wang, X. Li, Y. Zhu, J. Wei, C. Zhang, Q. Kong, X. Nie, Q. Zhang, Z. Wang, Genome-wide association analysis of milk production, somatic cell score, and body conformation traits in Holstein cows. *Front. Vet. Sci.* **9**, 932034 (2022).

21. B. Samuel, D. Mengistie, E. Assefa, M. Kang, C. Park, H. Dadi, H. Dinka, Genetic diversity of DGAT1 gene linked to milk production in cattle populations of Ethiopia. *BMC Genom. Data* **23**, 64 (2022).
22. L. Taherkhani, M. H. Banabazi, N. EmamJomeh-Kashan, A. Noshary, I. Imumorin, The candidate chromosomal regions responsible for milk yield of cow: A GWAS meta-analysis. *Animals (Basel)* **12**, 582 (2022).
23. M. Scholtens, A. Jiang, A. Smith, M. Littlejohn, K. Lehnert, R. Snell, N. Lopez-Villalobos, D. Garrick, H. Blair, Genome-wide association studies of lactation yields of milk, fat, protein and somatic cell score in New Zealand dairy goats. *J. Anim. Sci. Biotechnol.* **11**, 55 (2020).
24. D. Chandran, A. Confair, K. Warren, Y. I. Kawasawa, S. D. Hicks, Maternal variants in the *MFG8* gene are associated with perceived breast milk supply. *Breastfeed. Med.* **17**, 331–340 (2022).
25. L. Bode, Human milk oligosaccharides: Every baby needs a sugar mama. *Glycobiology* **22**, 1147–1162 (2012).
26. J. E. Williams, M. K. McGuire, C. L. Meehan, M. A. McGuire, S. L. Brooker, E. W. Kamau-Mbuthia, E. W. Kamundia, S. Mbugua, S. E. Moore, A. M. Prentice, G. E. Otoo, J. M. Rodríguez, R. G. Pareja, J. A. Foster, D. W. Sellen, D. G. Kita, H. L. Neibergs, B. M. Murdoch, Key genetic variants associated with variation of milk oligosaccharides from diverse human populations. *Genomics* **113**, 1867–1875 (2021).
27. Z. T. Lewis, S. M. Totten, J. T. Smilowitz, M. Popovic, E. Parker, D. G. Lemay, M. L. Van Tassel, M. J. Miller, Y.-S. Jin, J. B. German, C. B. Lebrilla, D. A. Mills, Maternal fucosyltransferase 2 status affects the gut bifidobacterial communities of breastfed infants. *Microbiome* **3**, 13 (2015).
28. Y. Golan, T. Kambe, Y. G. Assaraf, The role of the zinc transporter *SLC30A2/ZnT2* in transient neonatal zinc deficiency. *Metalomics* **9**, 1352–1366 (2017).
29. W. Chohanaisai, B. Lönnnerdal, S. L. Kelleher, Identification of a mutation in *SLC30A2* (*ZnT-2*) in women with low milk zinc concentration that results in transient neonatal zinc deficiency. *J. Biol. Chem.* **281**, 39699–39707 (2006).
30. Y. Golan, A. Lehy, G. Horev, Y. G. Assaraf, High proportion of transient neonatal zinc deficiency causing alleles in the general population. *J. Cell. Mol. Med.* **23**, 828–840 (2019).
31. J. C. Mychaleckyj, U. Nayak, E. R. Colgate, D. Zhang, T. Carstensen, S. Ahmed, T. Ahmed, A. J. Mentzer, M. Alam, B. D. Kirkpatrick, R. Haque, A. S. G. Faruque, W. A. Petri Jr., Multiplex genome-wide association analysis of breast milk fatty acid composition extends the phenotypic association and potential selection of *FADS1* variants to arachidonic acid, a critical infant micronutrient. *J. Med. Genet.* **55**, 459–468 (2018).
32. L. Zhao, H. Ke, H. Xu, G.-D. Wang, H. Zhang, L. Zou, S. Xiang, M. Li, L. Peng, M. Zhou, L. Li, L. Ao, Q. Yang, C.-K. J. Shen, P. Yi, L. Wang, B. Jiao, TDP-43 facilitates milk lipid secretion by post-transcriptional regulation of *Btn1a1* and *Xdh*. *Nat. Commun.* **11**, 341 (2020).
33. 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).
34. M. A. Mohammad, D. L. Hadsell, M. W. Haymond, Gene regulation of UDP-galactose synthesis and transport: Potential rate-limiting processes in initiation of milk production in humans. *Am. J. Physiol. Endocrinol. Metab.* **303**, E365–E376 (2012).
35. M. A. Mohammad, M. W. Haymond, Regulation of lipid synthesis genes and milk fat production in human mammary epithelial cells during secretory activation. *Am. J. Physiol. Endocrinol. Metab.* **305**, E700–E716 (2013).
36. A. Cánovas, G. Rincón, C. Bevilacqua, A. Islas-Trejo, P. Brenaut, R. C. Hovey, M. Boutinaud, C. Morgenthaler, M. K. VanKlombenberg, P. Martin, J. F. Medrano, Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. *Sci. Rep.* **4**, 5297 (2014).
37. S. M. Shaban, R. A. Hassan, A. A. I. Hassanin, A. Fathy, A. A. S. El Nabtiti, Mammary fat globules as a source of mRNA to model alterations in the expression of some milk component genes during lactation in bovines. *BMC Vet. Res.* **20**, 286 (2024).
38. S. Gallier, K. Vocking, J. A. Post, B. Van De Heijning, D. Acton, E. M. Van Der Beek, T. Van Baalen, A novel infant milk formula concept: Mimicking the human milk fat globule structure. *Colloids Surf. B Biointerfaces* **136**, 329–339 (2015).
39. S. Gallier, D. Gragson, R. Jiménez-Flores, K. Everett, Using confocal laser scanning microscopy to probe the milk fat globule membrane and associated proteins. *J. Agric. Food Chem.* **58**, 4250–4257 (2010).
40. C. Lopez, O. Ménard, Human milk fat globules: Polar lipid composition and in situ structural investigations revealing the heterogeneous distribution of proteins and the lateral segregation of sphingomyelin in the biological membrane. *Colloids Surf. B Biointerfaces* **83**, 29–41 (2011).
41. J. A. Sharp, C. Lefèvre, A. Watt, K. R. Nicholas, Analysis of human breast milk cells: Gene expression profiles during pregnancy, lactation, involution, and mastitic infection. *Funct. Integr. Genomics* **16**, 297–321 (2016).
42. 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. *J. Mammary Gland Biol. Neoplasia* **25**, 367–387 (2020).
43. 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).
44. A.-J. Twigger, L. K. Engelbrecht, K. Bach, I. Schultz-Pernice, S. Pensa, J. Stenning, S. Petricca, C. H. Scheel, W. T. Khaled, Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. *Nat. Commun.* **13**, 562 (2022).
45. J. P. Gleeson, N. Chaudhary, K. C. Fein, R. Doerfler, P. Hrdzak-Showalter, K. A. Whitehead, Profiling of mature-stage human breast milk cells identifies six unique lactocyte subpopulations. *Sci. Adv.* **8**, eabm6865 (2022).
46. C. LeMaster, S. H. Pierce, E. S. Geanes, S. Khanal, S. S. Elliott, A. B. Scott, D. A. Louiselle, R. McLennan, D. Maulik, T. Lewis, T. Pastinen, T. Bradley, The cellular and immunological dynamics of early and transitional human milk. *Commun. Biol.* **6**, 539 (2023).
47. E. Brownell, C. R. Howard, R. A. Lawrence, A. M. Dozier, Delayed onset lactogenesis II predicts the cessation of any or exclusive breastfeeding. *J. Pediatr.* **161**, 608–614 (2012).
48. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
49. B. Pang, Q. Wang, S. Ning, J. Wu, X. Zhang, Y. Chen, S. Xu, Landscape of tumor suppressor long noncoding RNAs in breast cancer. *J. Exp. Clin. Cancer Res.* **38**, 79 (2019).
50. D. Corral, E. Ansaldo, J. Delaleu, A. C. Pichler, J. Kabat, C. Oguz, D. Yong, M. Abid, C. A. Rivera, V. M. Link, K. Yang, L. Chi, J. Nie, O. Kamenyeva, Y. Fan, J. K. Y. Chan, F. Ginhoux, R. Bosselut, Y. Belkaid, Mammary intraepithelial lymphocytes promote lactogenesis and offspring fitness. *Cell* **188**, 1662–1680.e24 (2025).
51. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
52. B. Jew, M. Alvarez, E. Rahmani, Z. Miao, A. Ko, K. M. Garske, J. H. Sul, K. H. Pietiläinen, P. Pajukanta, E. Halperin, Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
53. J. F. Martin Carli, M. Dzieciatkowska, T. L. Hernandez, J. Monks, J. L. McManaman, Comparative proteomic analysis of human milk fat globules and paired membranes and mouse milk fat globules identifies core cellular systems contributing to mammary lipid trafficking and secretion. *Front. Mol. Biosci.* **10**, 1259047 (2023).
54. A. Memon, W. K. Lee, KLF10 as a tumor suppressor gene and its TGF- β signaling. *Cancer* **10**, 161 (2018).
55. E. Cocolakis, M. Dai, L. Drevet, J. Ho, E. Haines, S. Ali, J.-J. Lebrun, Smad signaling antagonizes STAT5-mediated gene transcription and mammary epithelial cell differentiation. *J. Biol. Chem.* **283**, 1293–1307 (2008).
56. K. Kahata, V. Maturi, A. Moustakas, TGF- β family signaling in ductal differentiation and branching morphogenesis. *Cold Spring Harb. Perspect. Biol.* **10**, a031997 (2018).
57. M. C. Neville, P. Webb, P. Ramanathan, M. P. Mannino, C. Pecorini, J. Monks, S. M. Anderson, P. MacLean, The insulin receptor plays an important role in secretory differentiation in the mammary gland. *Am. J. Physiol. Endocrinol. Metab.* **305**, E1103–E1114 (2013).
58. T. Alam, D. A. Kenny, T. Sweeney, F. Buckley, R. Prendiville, M. McGee, S. M. Waters, Expression of genes involved in energy homeostasis in the duodenum and liver of Holstein-Friesian and Jersey cows and their F₁ hybrid. *Physiol. Genomics* **44**, 198–209 (2012).
59. W. Chen, B. Chang, X. Wu, L. Li, M. Sleeman, L. Chan, Inactivation of *Plin4* downregulates *Plin5* and reduces cardiac lipid accumulation in mice. *Am. J. Physiol. Endocrinol. Metab.* **304**, E770–E779 (2013).
60. X. Zhang, L. Su, K. Sun, Expression status and prognostic value of the perilipin family of genes in breast cancer. *Am. J. Transl. Res.* **13**, 4450–4463 (2021).
61. S. Tonko-Geymayer, O. Goupille, M. Tonko, C. Soratroi, A. Yoshimura, C. Streuli, A. Ziemiecki, R. Kofler, W. Doppler, Regulation and function of the cytokine-inducible SH-2 domain proteins, CIS and SOCS3, in mammary epithelial cells. *Mol. Endocrinol.* **16**, 1680–1695 (2002).
62. F. M. Hannan, T. Elajnaif, L. N. Vandenberg, S. H. Kennedy, R. V. Thakker, Hormonal regulation of mammary gland development and lactation. *Nat. Rev. Endocrinol.* **19**, 46–61 (2023).
63. M. Büttner, J. Ostner, C. L. Müller, F. J. Theis, B. Schubert, scCODA is a Bayesian model for compositional single-cell data analysis. *Nat. Commun.* **12**, 6876 (2021).
64. P. P. Meier, J. L. Engstrom, J. L. Zuleger, J. E. Motykowski, U. Vasan, W. A. Meier, P. E. Hartmann, T. M. Williams, Accuracy of a user-friendly centrifuge for measuring creatinotocrits on mothers' milk in the clinical setting. *Breastfeed. Med.* **1**, 79–87 (2006).
65. A. Lucas, J. A. Gibbs, R. L. Lyster, J. D. Baum, Creatinotocrit: Simple clinical technique for estimating fat concentration and energy value of human milk. *Br. Med. J.* **1**, 1018–1020 (1978).
66. O. H. Braun, H. Sandkühler, Relationships between lysozyme concentration of human milk, bacteriologic content, and weight gain of premature infants. *J. Pediatr. Gastroenterol. Nutr.* **4**, 583–586 (1985).

67. C. Hassel, B. Gausserès, L. Guzylack-Piriou, G. Foucras, Ductal macrophages predominate in the immune landscape of the lactating mammary gland. *Front. Immunol.* **12**, 754661 (2021).
68. C. A. Dawson, B. Pal, F. Vaillant, L. C. Gandolfo, Z. Liu, C. Blierot, F. Ginhoux, G. K. Smyth, G. J. Lindeman, S. N. Mueller, A. C. Rios, J. E. Visvader, Tissue-resident ductal macrophages survey the mammary epithelium and facilitate tissue remodelling. *Nat. Cell Biol.* **22**, 546–558 (2020).
69. D. Ennis, S. Shmorak, E. Jantscher-Krenn, M. Yassour, Longitudinal quantification of *Bifidobacterium longum* subsp. *infantis* reveals late colonization in the infant gut independent of maternal milk HMO composition. *Nat. Commun.* **15**, 894 (2024).
70. J. D. Forbes, M. B. Azad, L. Vehling, H. M. Tun, T. B. Konya, D. S. Guttman, C. J. Field, D. Lefebvre, M. R. Sears, A. B. Becker, P. J. Mandhane, S. E. Turvey, T. J. Moraes, P. Subbarao, J. A. Scott, A. L. Kozyrskyj, Canadian Healthy Infant Longitudinal Development (CHILD) Study Investigators, Association of exposure to formula in the hospital and subsequent infant feeding practices with gut microbiota and risk of overweight in the first year of life. *JAMA Pediatr.* **172**, e181161 (2018).
71. F. Schozer, P. Tso, R. Karns, E. Wagner, R. Walker, R. Hovey, J. Trott, S. Riddle, A. Thompson, L. Ward, L. Nommensen-Rivers, Chronic elevation of tumor necrosis factor- α suppresses lactation via downregulation of lipoprotein lipase. *Curr. Dev. Nutr.* **5**, 814 (2021).
72. S. Canale, R. Ramanathan, M. Pelligrini, N. C. Rochette, B. B. Nadel, M. Gee, Analysis of gene expression from human breastmilk cells: A comparison between low and high producers, and the influence of anxiety and depression on milk production, gene expression and bacterial production. *Heliyon* **7**, e08335 (2021).
73. M. C. Neville, E. W. Demerath, J. Hahn-Holbrook, R. C. Hovey, J. Martin-Carli, M. A. McGuire, E. R. Newton, K. M. Rasmussen, M. C. Rudolph, D. J. Raiten, Parental factors that impact the ecology of human mammary development, milk secretion, and milk composition—a report from “Breastmilk Ecology: Genesis of Infant Nutrition (BEGIN)” Working Group 1. *Am. J. Clin. Nutr.* **117**, S11–S27 (2023).
74. C. B. Betts, N. D. Pennock, B. P. Caruso, B. Ruffell, V. F. Borges, P. Schedin, Mucosal immunity in the female murine mammary gland. *J. Immunol.* **201**, 734–746 (2018).
75. K. Fehr, S. Moossavi, H. Sbihi, R. C. T. Boutin, L. Bode, B. Robertson, C. Yonemitsu, C. J. Field, A. B. Becker, P. J. Mandhane, M. R. Sears, E. Khafipour, T. J. Moraes, P. Subbarao, B. B. Finlay, S. E. Turvey, M. B. Azad, Breastmilk feeding practices are associated with the co-occurrence of bacteria in mothers’ milk and the infant gut: The CHILD cohort study. *Cell Host Microbe* **28**, 285–297.e4 (2020).
76. Y. Golan, M. Prael, A. G. Cassidy, C. Gay, A. H. B. Wu, U. Jigmeddagva, C. Y. Lin, V. J. Gonzalez, E. Basilio, M. A. Chidboy, L. Warrier, S. Buarpung, L. Li, A. P. Murtha, I. V. Asiodu, N. Ahituv, V. J. Flaherman, S. L. Gaw, COVID-19 mRNA vaccination in lactation: Assessment of adverse events and vaccine related antibodies in mother-infant dyads. *Front. Immunol.* **12**, 777103 (2021).
77. I. Subirana, J. Vila, H. Sanz, compareGroups 4.0: Descriptives by groups (2024); https://cran.r-project.org/web/packages/compareGroups/vignettes/compareGroups_vignette.html.
78. R. Jenness, The composition of human milk. *Semin. Perinatol.* **3**, 225–239 (1979).
79. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
80. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
81. R. A. Amezcua, A. T. L. Lun, E. Becht, V. J. Carey, L. N. Carpp, L. Geistlinger, F. Marini, K. Rue-Albrecht, D. Risso, C. Sonesson, L. Waldron, H. Pagès, M. L. Smith, W. Huber, M. Morgan, R. Gottardo, S. C. Hicks, Orchestrating single-cell analysis with bioconductor. *Nat. Methods* **17**, 137–145 (2020).
82. H. L. Crowell, C. Sonesson, P.-L. Germain, D. Calini, L. Collin, C. Raposo, D. Malhotra, M. D. Robinson, Muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
83. P.-Y. Tung, J. D. Blischak, C. J. Hsiao, D. A. Knowles, J. E. Burnett, J. K. Pritchard, Y. Gilad, Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* **7**, 39921 (2017).
84. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene Ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
85. Z. Fang, X. Liu, G. Peltz, GSEapy: A comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
86. K. E. Fujimura, A. R. Sitarik, S. Havstad, D. L. Lin, S. Levan, D. Fadrosch, A. R. Panzer, B. LaMere, E. Rackaityte, N. W. Lukacs, G. Wegienka, H. A. Boushey, D. R. Ownby, E. M. Zoratti, A. M. Levin, C. C. Johnson, S. V. Lynch, Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat. Med.* **22**, 1187–1191 (2016).
87. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
88. E. A. Genomics, Ea-Utills: Automatically Exported from Code.google.com/p/ea-utills (Github); <https://github.com/ExpressionAnalysis/ea-utills>.
89. A. Blanco-Miguez, F. Beghini, F. Cumbo, L. J. McIver, K. N. Thompson, M. Zolfo, P. Manghi, L. Dubois, K. D. Huang, A. M. Thomas, W. A. Nickols, G. Piccinno, E. Piperni, M. Punčochář, M. Valles-Colomer, A. Tett, F. Giordano, R. Davies, J. Wolf, S. E. Berry, T. D. Spector, E. A. Franzosa, E. Pasoli, F. Asnicar, C. Huttenhower, N. Segata, Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
90. H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan, dplyr: A grammar of data manipulation. R package version 1.1.4 (2025); <https://dplyr.tidyverse.org>.
91. H. Wickham, D. Vaughan, M. Girlich, tidy: Tidy messy data. R package version 1.3.1 (2025); <https://tidyr.tidyverse.org>.
92. H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemond, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
93. H. Wickham, *Ggplot2* (Springer International Publishing, 2016).
94. E. Neuwirth, RColorBrewer: ColorBrewer palettes. R package version 1.1-3 (2025); <https://CRAN.R-project.org/package=RColorBrewer>.
95. K. Wright, pals: Color palettes, colormaps, and tools to evaluate them. R package version 1.10 (2025); <https://kwstat.github.io/pals/>.
96. J. Oksanen, G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O’Hara, P. Solymos, M. H. H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, T. Borman, G. Carvalho, M. Chirico, M. De Cáceres, S. Durand, H. B. A. Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M. O. Hill, L. Lahti, C. Martino, D. McGlinn, M.-H. Ouellette, E. Ribeiro Cunha, T. Smith, A. Stier, C. J. F. Ter Braak, J. Weedon, vegan: Community ecology package. R package version 2.8-0 (2025); <https://vegandevs.github.io/vegan/>.
97. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
98. H. Mallick, A. Rahnavard, L. J. McIver, S. Ma, Y. Zhang, L. H. Nguyen, T. L. Tickle, G. Weingart, B. Ren, E. H. Schwager, S. Chatterjee, K. N. Thompson, J. E. Wilkinson, A. Subramanian, Y. Lu, L. Waldron, J. N. Paulson, E. A. Franzosa, H. C. Bravo, C. Huttenhower, Multivariable association discovery in population-scale meta-omics studies. *PLOS Comput. Biol.* **17**, e1009442 (2021).

Acknowledgments: We thank the donors and their families for participation in this study. We thank the lactation consultants at UCSF clinic for supporting this work. This work is supported by funding from The Benioff Center for Microbiome Medicine at UCSF. We thank the UCSF Benioff Center for Microbiome Medicine Microbial Genomics CoLab Plug-in for assistance in data generation. We thank M. Thomas and the Cornell Statistical Consulting Unit for assistance with statistical analysis. **Funding:** These studies were supported by the Weizmann Institute of Science - National Postdoctoral Award Program for Advancing Women in Science, the International Society for Research in Human Milk and Lactation (ISRHL) Trainee Bridge Fund, the Human Frontier Science Program and the National Institute of Child Health and Human Development (NICHD) 1K99HD113880 (to Y.G.), National Institute of Allergy and Immunology K08AI141728 and a gift from the Krzyzewski Family (to S.L.G.), NIH - National Human Genome Research Institute R01 HG012967 and NIH National Cancer Institute 5U2CCA233195 (to B.E.E.), NIH F32 (F32 HD114427) (to S.K.N.), and Benioff Center for Microbiome Medicine UCSF (to N.A.). **Author contributions:** Conceptualization: Y.G., N.A., and D.E. Methodology: Y.G., M.Y., N.A., S.K.N., B.E.E., D.E., and V.J.F. Formal analysis: Y.G., M.Y., B.E.E., J.Z., D.E., and S.K.N. Resources: Y.G., N.A., M.P., S.L.G., and V.J.F. Software: M.Y., J.Z., D.E., and S.K.N. Investigation: Y.G., Z.L., D.E., E.B., A.R.K., M.P., and V.J.F. Visualization: S.K.N., D.E., Y.G., M.Y., and J.Z. Funding acquisition: Y.G., N.A., M.P., and S.L.G. Data curation: Y.G., N.A., J.Z., D.E., V.J.F., and S.K.N. Validation: Y.G., N.A., D.E., and S.K.N. Supervision: Y.G., N.A., V.J.F., B.E.E., M.P., and S.L.G. Project administration: Y.G., N.A., and V.J.F. Writing—original draft: Y.G., S.K.N., N.A., and D.E. Writing—review and editing: Y.G., N.A., S.K.N., V.J.F., B.E.E., M.P., S.L.G., and D.E. **Competing interests:** B.E.E. is on the Scientific Advisory Board for ArrePath Inc., Crayon Bio, and Freenome; she consults for Neumora. B.E.E. is also a CIFAR Fellow in the Multiscale Human Program. S.K.N. reports compensation for consulting services with Radera Biosciences. All other authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Sequencing data are available at (i) <https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE275626> and (ii) <https://ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE275627>.

Submitted 14 July 2024

Accepted 8 August 2025

Published 10 September 2025

10.1126/sciadv.adr7174