Milk cell transcriptome opens a new dimension in the mammary gland biology research

M. Zorc and P. Dovc

University of Ljubljana, Biotechnical Faculty, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
Corresponding Author: peter.dovc@bf.uni-lj.si

Introduction

The mammary gland is a highly regenerative organ that undergoes most of its development after birth (Inman *et al*. 2015). The cyclical phases of growth, differentiation, lactation and involution of the mammary gland are regulated by hormones and growth factors (Neville *et al*. 2002). A consequence of the complex function of the mammary gland and intense secretion of milk, which differs significantly among different species, is also the presence of somatic cells in milk. The main fractions of somatic cells in milk are epithelial cells, lymphocytes, polymorphonuclear neutrophils (PMN), and macrophages. The majority of exfoliated epithelial cells present in milk are viable and exhibit characteristics of fully differentiated alveolar cells (Boutinaud and Jammes 2002). The somatic cell count (SCC) in milk, widely used as a marker for udder health, only provides the cumulative number of somatic cells in milk, whereas the differential somatic cell count (DSCC) allows differentiation between two groups of cells: PMN and lymphocytes versus macrophages (Wall *et al*. 2018) and represents therefore a significant step forward in understanding the dynamics of the somatic cell population in the mammary gland during lactation and at infection. In cattle and sheep, the epithelial cell fraction represents only a relatively small part of somatic cells in milk, whereas, in porcine and goat milk, as well as in human milk, epithelial cells are the predominant cell type in milk (Boutinaud and Jammes 2002). In different organs, adult stem cells are present with their primary role of maintaining tissue homeostasis (Biteau *et al*. 2011). However, stem cells in the adult mammary gland serve both, development and homeostasis. Mammary stem cells (MaSCs) can self-renew and differentiate into different cell types during the mammary gland's developing cycles (Visvader and Stingl 2014). In 2006, mouse MaSCs were identified and isolated for the first time (Shackleton *et al*. 2006). Since then, plenty of strategies have been used to identify and characterize MaSC and to delineate the mammary epithelial hierarchy (Inman *et al.* 2015).

Considerable efforts have been made to find a noninvasive way to obtain biological material for molecular analyses of mammary gland cells. The comparison of five different sources of RNA (biopsies of the mammary gland tissue, laser microdissected mammary epithelial cells, milk somatic cells, milk fat globules and antibody-captured milk mammary epithelial cells) for analysis of the bovine mammary gland transcriptome, showed that isolation of a total RNA directly from somatic cells released into milk during lactation, is an effective alternative to mammary gland tissue biopsies and laser microdissection of mammary tissue (Canovas *et al*. 2014).

The organ-specific gene expression studies in the mammary gland were performed using expression microarrays a decade ago (Maningat *et al*. 2009) and allowed a comparative approach between species but were limited with the selection of genes on the chip. The next important step represented bulk RNA sequencing from mammary gland isolates (Medrano 2010). Sequencing of bulk RNA isolated from milk cells in three different lactation stages in Holstein cows revealed expression of more than 19,000 genes as a cumulative number of genes expressed in different cell types

present in cow's milk. Regardless of the lactation stage, approximately 9,000 genes showed ubiquitous expression, however, genes encoding lactoproteins and enzymes in the lactose synthesis pathway showed higher expression in early lactation and the majority of genes in the fat metabolism pathway had high expression in transition and peak lactation (Wickramasinghe *et al*. 2012). Several recent studies examined the distinct gene expression profiles of different mammary epithelial cell lineages at the single-cell level in human and mice (Cristea and Polyak 2018). In mouse, the analysis revealed 11 luminal and four basal clusters (Bach *et al*. 2017). The main advantage of single-cell RNA sequencing over the bulk mammary RNA sequencing is that single-cell RNA sequencing provides a reliable information about gene expression differences among different cell types and allows reliable assignment of transcripts to different cell types. The single-cell RNA sequencing opens a new horizon for documentation of cell type specific expression profiles in the mammary gland and even the possibility to determine different cell types based on cell type specific transcriptomic profile (Nguyen *et al*. 2018). In four studies a complete murine (Han *et al*. 2018; Schaum *et al*. 2018) and bovine (Becker *et al*. 2021; Zorc *et al*. 2024) mammary gland cell population was sequenced at the single-cell level, revealing a number of distinct cell types which exceeds the initially expected number. This approach also allows the identification of cellular sources for several milk components, which did not have defined origin before (Dallas *et al*. 2015).

The amount of data in a typical single-cell sequencing experiment is much larger than in bulk RNA sequencing experiments. The increased amount of data represents a computational challenge and an opportunity to apply advanced approaches such as machine learning. Machine learning concepts are applied in computational pipelines for scRNA-seq data analyses (Hwang *et al*. 2018). From a mathematical point of view, identification of cell-populations in scRNA-Seq data is unsupervised clustering, a problem widely studied in the field of machine learning (Andrews and Hemberg 2018). Dimension reduction is needed before clustering because scRNA-Seq data is high-dimensional (~104 dimensions for mammalian samples) and suffers from the curse of dimensionality (Wagner *et al*. 2016). Methods used for dimension reduction are either Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (tSNE) or diffusion maps (DM).

Here we report the application of scRNA-seq to elucidate the cell type repertoire in bovine milk based on the transcriptomic differences among different cell clusters. Milk contains mammary epithelial cells and immune system cells (lymphocyte, macrophage, neutrophils), which reflect the activity of the mammary gland and illustrate the response of the mammary gland to environmental challenges.

Material and methods

Milk samples were collected from two healthy Holstein cows (less than 50.000 SSC) in mid lactation. The cells were pelleted and washed in cold PBS. Single-cell library was generated using 10X Genomics technology and Chromium Single Cell 3' Reagent Kit. Library samples were diluted to a concentration of 10 nM and loaded onto NovaSeq 6000 (Illumina) instrument. Sample demultiplexing, barcode processing, read alignment to the bovine reference genome (ARS-UCD1.2.108), quantification and initial quality control of the paired-end sequencing data were performed for each sample using Cell Ranger software (version 7.1.0, 10X Genomics). Genes expressed in less than three cells were removed from the gene expression matrix.

We applied "anchor-based" Seurat's workflow (Stuart *et al*. 2018) to integrate two datasets. After filtering, we log-normalized the raw counts with LogNormalize and identified highly variable genes with FindVariableFeatures for each batch at default settings. We then ran FindIntegrationAnchors with dims = 1:30. The resulting anchors

were used for IntegrateData with the 30 dimensions. The most variable genes based on their expression in the entire population were determined using the FindVariableGenes function with default parameters (selection.method = "vst", nfeatures = 2000). Clusters were identified using the FindClusters function with a resolution of 0.8 and then visualized using the RunTSN and RunUMAP functions (reduction = "pca"). For fully automated cell-type identification we used ScType with ScType's marker database (https://www.nature.com/articles/s41467-022-28803-w).

Results

A total of ~361M reads were obtained with 36,315 mean reads per cell for the first cow and ~257M reads with 17,459 mean reads per cell for the second. The efficiency of read mapping was between 94.1 and 96.6%. In total, 15,630 and 16,497 genes were identified. We performed an anchor-based integration analysis to explore all cells in both samples simultaneously. After UMAP reduction, a clear cell clustering highlighting 22 distinct cell populations was obtained (Figure 1).

The identified clusters can be grouped into two larger categories, including immune and epithelial cells. We identified classical and intermediate monocytes, naïve, effector and memory CD8+ T cells, naïve and effector CD4+ T cells, ISG expression immune cells, neutrophils, macrophages, mast cells, platelets, neutrophils, naïve B cells, progenitor cells, dendritic cells, luminal, ductal and alveolar cells. Identification of milk producing cells was based on expression of casein (CSN1S1, CSN1S2, CSN2, CSN3) and whey protein (PAEP, LALBA) genes. The cell cluster with significantly higher levels of caseins and whey proteins was annotated as alveolar cell cluster. The expression of casein (CSN1S1, CSN1S2, CSN2, CSN3), whey protein (PAEP, LALBA), MUC15 and BTNA1 genes allows identification of cell-type specific profiles indicating differences among bovine somatic milk cell clusters.

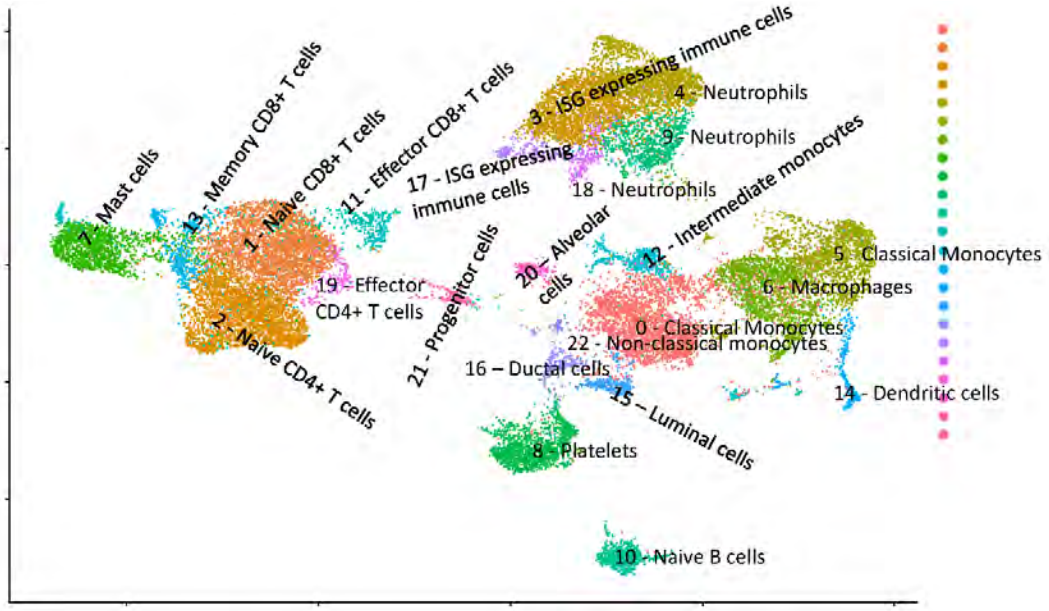

Figure 1. Cell type clusters in bovine milk somatic cells. Cluster 20 (alveolar cells) represents a typical milk producing cluster.

Discussion

Traditionally, somatic cells in the milk are expected to belong to myo/epithelial mammary gland cells, different types of immune cells (lymphocytes, neutrophils, macrophages) and stromal cells (Wickramasinghe *et al*. 2011). However, since precise markers for sub differentiation of cell types in the mammary gland are not present in all mammalian species (agricultural species are there not very well covered), the number of different cell types in the somatic cell fraction was normally underestimated. The analysis of bulk RNA transcripts from milk somatic cells revealed a very wide range of expressed genes and consequently indicated a wider range of cell-types in the milk somatic cell fraction. Single-cell sequencing of human and mouse mammary somatic cells revealed a much wider range of cell types, which are present in the milk (Nguyen *et al.* 2018; Schaum *et al.* 2018). Our single-cell RNA sequencing analysis of bovine milk has unveiled a cellular landscape of bovine milk somatic cells, highlighting a rich diversity of cell types pivotal for lactation, immune response and tissue homeostasis. Similar to the findings of Becker *et al*. (2021) (Becker *et al.* 2021), our study confirms the mammary gland's complexity, revealing a broad array of immune and epithelial cells.

A recent analysis of single-cell transcriptomes in mice has revealed important differences in gene expression between different cell types, which can significantly vary during the development of the mammary gland as well as in the course of lactation (Giraddi *et al*. 2018). The identification of a considerably higher number of cell-types in the milk somatic cell fraction compared to the traditional expectation opens a new horizon for more complex interpretation of the biological processes in the mammary gland.

List of references

Andrews T.S. and Hemberg M. (2018) Identifying cell populations with scRNASeq. Mol Aspects Med 59, 114-22.

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

Becker D., Weikard R., Hadlich F. and Kühn C. (2021) Single-cell RNA sequencing of freshly isolated bovine milk cells and cultured primary mammary epithelial cells. Scientific Data 8, 177.

Biteau B., Hochmuth C.E. and Jasper H. (2011) Maintaining tissue homeostasis: dynamic control of somatic stem cell activity. Cell stem cell 9, 402-11.

Boutinaud M. and Jammes H. (2002) Potential uses of milk epithelial cells: a review. Reprod Nutr Dev 42, 133-47.

Canovas A., Rincon G., Bevilacqua C., Islas-Trejo A., Brenaut P., Hovey R.C., Boutinaud M., Morgenthaler C., VanKlompenberg M.K., Martin P. and Medrano J.F. (2014) Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Sci Rep 4, 5297.

Cristea S. and Polyak K. (2018) Dissecting the mammary gland one cell at a time. Nat Commun 9, 2473.

Dallas D.C., Murray N.M. and Gan J. (2015) Proteolytic Systems in Milk: Perspectives on the Evolutionary Function within the Mammary Gland and the Infant. Journal of mammary gland biology and neoplasia 20, 133-47.

Giraddi R.R., Chung C.Y., Heinz R.E., Balcioglu O., Novotny M., Trejo C.L., Dravis C., Hagos B.M., Mehrabad E.M., Rodewald L.W., Hwang J.Y.,

Fan C., Lasken R., Varley K.E., Perou C.M., Wahl G.M. and Spike B.T. (2018) Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. Cell Rep 24, 1653-66.e7.

Han X., Wang R., Zhou Y., Fei L., Sun H., Lai S., Saadatpour A., Zhou Z., Chen H., Ye F., Huang D., Xu Y., Huang W., Jiang M., Jiang X., Mao J., Chen Y., Lu C., Xie J., Fang Q., Wang Y., Yue R., Li T., Huang H., Orkin S.H., Yuan G.C., Chen M. and Guo G. (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172, 1091-107.e17.

Hwang B., Lee J.H. and Bang D. (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Experimental and Molecular Medicine 50, 96.

Inman J.L., Robertson C., Mott J.D. and Bissell M.J. (2015) Mammary gland development: cell fate specification, stem cells and the microenvironment. Development 142, 1028-42.

Maningat P.D., Sen P., Rijnkels M., Sunehag A.L., Hadsell D.L., Bray M. and Haymond M.W. (2009) Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome. Physiol Genomics 37, 12-22.

Medrano J.F.C., A. Islas-Trejo, A. (2010) Comparative analysis of bovine milk and mammary gland transcriptome using RNA-Seq. In: 9th World Congress on Genetics applied to Livestock Production, Leipzig, Germany.

Neville M.C., McFadden T.B. and Forsyth I. (2002) Hormonal regulation of mammary differentiation and milk secretion. J Mammary Gland Biol Neoplasia 7, 49-66.

Nguyen Q.H., Pervolarakis N., Blake K., Ma D., Davis R.T., James N., Phung A.T., Willey E., Kumar R., Jabart E., Driver I., Rock J., Goga A., Khan S.A., Lawson D.A., Werb Z. and Kessenbrock K. (2018) Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. Nature Communications 9, 2028.

Schaum N., Karkanias J., Neff N.F., May A.P., Quake S.R., Wyss-Coray T., Darmanis S., Batson J., Botvinnik O., Chen M.B., Chen S., Green F., Jones R.C., Maynard A., Penland L., Pisco A.O., Sit R.V., Stanley G.M., Webber J.T., Zanini F., Baghel A.S., Bakerman I., Bansal I., Berdnik D., Bilen B., Brownfield D., Cain C., Cho M., Cirolia G., Conley S.D., Demers A., Demir K., de Morree A., Divita T., du Bois H., Dulgeroff L.B.T., Ebadi H., Espinoza F.H., Fish M., Gan Q., George B.M., Gillich A., Genetiano G., Gu X., Gulati G.S., Hang Y., Hosseinzadeh S., Huang A., Iram T., Isobe T., Ives F., Kao K.S., Karnam G., Kershner A.M., Kiss B.M., Kong W., Kumar M.E., Lam J.Y., Lee D.P., Lee S.E., Li G., Li Q., Liu L., Lo A., Lu W.-J., Manjunath A., May K.L., May O.L., McKay M., Metzger R.J., Mignardi M., Min D., Nabhan A.N., Ng K.M., Noh J., Patkar R., Peng W.C., Puccinelli R., Rulifson E.J., Sikandar S.S., Sinha R., Szade K., Tan W., Tato C., Tellez K., Travaglini K.J., Tropini C., Waldburger L., van Weele L.J., Wosczyna M.N., Xiang J., Xue S., Youngyunpipatkul J., Zardeneta M.E., Zhang F., Zhou L., Castro P., Croote D., DeRisi J.L., Kuo C.S., Lehallier B., Nguyen P.K., Tan S.Y., Wang B.M., Yousef H., Beachy P.A., Chan C.K.F., Huang K.C., Weinberg K., Wu S.M., Barres B.A., Clarke M.F., Kim S.K., Krasnow M.A., Nusse R., Rando T.A., Sonnenburg J., Weissman I.L., The Tabula Muris C., Overall c., Logistical c., Organ collection and p., Library preparation and s., Computational data a., Cell type a., Writing g.,

Supplemental text writing g. and Principal i. (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 562, 367-72.

Shackleton M., Vaillant F., Simpson K.J., Stingl J., Smyth G.K., Asselin-Labat M.L., Wu L., Lindeman G.J. and Visvader J.E. (2006) Generation of a functional mammary gland from a single stem cell. Nature 439, 84-8.

Stuart T., Butler A., Hoffman P., Hafemeister C., Papalexi E., Mauck W.M., Stoeckius M., Smibert P. and Satija R. (2018) Comprehensive integration of single cell data. bioRxiv, 460147.

Visvader J.E. and Stingl J. (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev 28, 1143-58.

Wagner A., Regev A. and Yosef N. (2016) Revealing the vectors of cellular identity with single-cell genomics. Nature biotechnology 34, 1145-60.

Wall S.K., Wellnitz O., Bruckmaier R.M. and Schwarz D. (2018) Differential somatic cell count in milk before, during, and after lipopolysaccharide- and lipoteichoic-acid-induced mastitis in dairy cows. J Dairy Sci 101, 5362-73.

Wickramasinghe S., Hua S., Rincon G., Islas-Trejo A., German J.B., Lebrilla C.B. and Medrano J.F. (2011) Transcriptome Profiling of Bovine Milk Oligosaccharide Metabolism Genes Using RNA-Sequencing. PLOS ONE 6, e18895.

Wickramasinghe S., Rincon G., Islas-Trejo A. and Medrano J.F. (2012) Transcriptional profiling of bovine milk using RNA sequencing. BMC genomics 13, 45-.

Zorc M., Dolinar M. and Dovč P. (2024) A Single-Cell Transcriptome of Bovine Milk Somatic Cells. Genes.15,394-