



OPEN

DATA DESCRIPTOR

Single-cell RNA sequencing of freshly isolated bovine milk cells and cultured primary mammary epithelial cells

Doreen Becker¹✉, Rosemarie Weikard¹, Frieder Hadlich¹ & Christa Kühn^{1,2}

Bovine mammary function at molecular level is often studied using mammary tissue or primary bovine mammary epithelial cells (pbMECs). However, bulk tissue and primary cells are heterogeneous with respect to cell populations, adding further transcriptional variation in addition to genetic background. Thus, understanding of the variation in gene expression profiles of cell populations and their effect on function are limited. To investigate the mononuclear cell composition in bovine milk, we analyzed a single-cell suspension from a milk sample. Additionally, we harvested cultured pbMECs to characterize gene expression in a homogeneous cell population. Using the Drop-seq technology, we generated single-cell RNA datasets of somatic milk cells and pbMECs. The final datasets after quality control filtering contained 7,119 and 10,549 cells, respectively. The pbMECs formed 14 indefinite clusters displaying intrapopulation heterogeneity, whereas the milk cells formed 14 more distinct clusters. Our datasets constitute a molecular cell atlas that provides a basis for future studies of milk cell composition and gene expression, and could serve as reference datasets for milk cell analysis.

Background & Summary

Bovine mammary structure, function, development and immune response and the underlying transcriptional regulatory processes in the mammary tissue have been studied extensively for almost half a century. Due to technical advances and new research techniques, molecular mechanisms of lactation, development and immune response can be analysed more accurately^{1,2}. However, bovine mammary gland tissue can only be collected by biopsy or after slaughter, and comprises a number of heterogeneous cell types (e.g., mammary epithelial cells (MECs), myoepithelial cells, adipocytes, fibroblasts) creating potential bias in the outcome of global transcriptome analyses depending on cell composition. Another specific feature of mammary gland tissue is the low complexity of its transcriptome, because milk and whey protein genes are highly expressed, in consequence masking the expression of lowly expressed transcripts^{3,4}.

To overcome these challenges, alternative approaches have been developed. Previous studies have shown that experiments on primary bovine mammary epithelial cells (pbMECs) enable the study of mammary gland functions⁵⁻⁷, but pbMECs are difficult to collect and often require killing the animal to obtain appropriate amounts of tissue. Furthermore, primary cell lines are heterogeneous, and transcriptional variation is substantially impacted by the genetic background. Moreover, after culture in *in-vitro* systems for extended time periods cells lose their original cell-type specific properties⁸, reducing the timeframe for experiments.

Milk-purified MECs can be a valuable, non-invasive source of mammary transcripts^{9,10}. Furthermore, isolation of MECs from milk allows for repetitive sampling of the same animal across several time points, e.g., across the entire lactation. This enables an analysis of transcriptional regulation dynamics. In addition, cells in milk might represent cells in a distinct physiological status, whereas cells of different differentiation and functional stages are present in mammary tissue⁹.

Bovine milk contains about 10^4 to 10^7 cells/ml, although the variation is large between individuals, environmental conditions and time points^{11,12}. Milk cell number is usually reported as number of somatic cells, which

¹Leibniz Institute of Farm Animal Biology (FBN), Institute of Genome Biology, Wilhelm-Stahl-Allee 2, 18196, Dummerstorf, Germany. ²University of Rostock, Faculty of Agricultural and Environmental Sciences, Justus-von-Liebig-Weg 6, 18059, Rostock, Germany. ✉e-mail: becker.doreen@fbn-dummerstorf.de

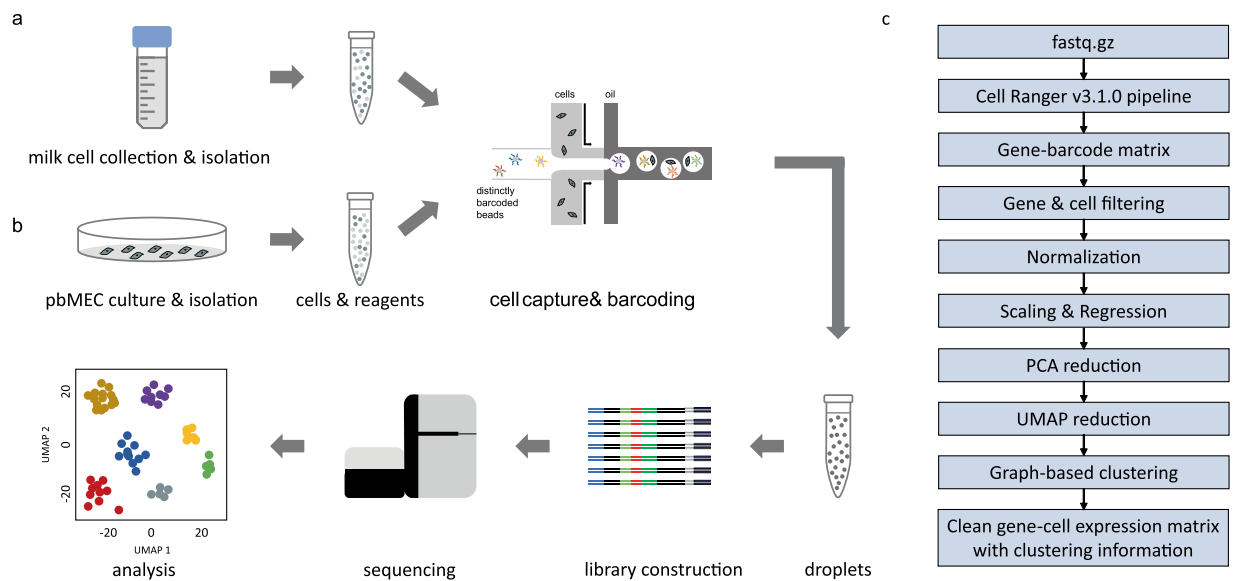


Fig. 1 Outline of laboratory and bioinformatic workflow. The workflow comprises isolation of (a) mononuclear cells from milk or (b) pbMECs from bulk mammary gland tissue, preparation of single-cell suspensions, library construction, sequencing and (c) the final bioinformatic scRNA-seq workflow. Adapted from Liu *et al.*³¹.

comprises the sum across a mixture of different non-epithelial cells (leukocytes)¹³. In contrast, MECs that are exfoliated from the epithelium during lactation are rare in milk and are therefore often not counted separately from somatic cells. While on average around 2% of the total milk cells constitute of epithelial cells, the proportion of MECs within the total milk cell population varies from one sample to another¹⁴. Since the concentration of MECs in milk is low, it is essential to collect sufficiently large volumes of milk to obtain a sufficient RNA quantity for mammary transcript analyses¹⁵. This limits the number of samples processed at a single collection time point, and the number of analyses that can be performed on the same sample. Nevertheless, milk cell samples can be used to monitor the mammary response to an invading pathogen. However, it is difficult to distinguish, if a change in bulk milk cell transcriptome is based on a change in cell composition or on a change in the expression of particular genes in a specific cell type. The same is true for the analysis of bulk mammary gland tissue.

Recent technical advances allow transcripts from thousands of cells to be pooled, sequenced, and subsequently identified in a single experiment at single-cell resolution¹⁶. This approach enables to assess the functional heterogeneity in a cell sample by identifying and characterising subpopulations of cells in a complex cell population^{17–19}.

In spite of this progress in technology, the current understanding of the spectrum of molecular heterogeneity in milk cells as well as in pbMECs is still limited. In this study, we assessed the gene expression across 7,119 single-cells that were isolated from a bovine milk sample and 10,549 pbMECs using single-cell RNA sequencing (scRNA-seq). The cells were processed using the 10x Chromium Single-cell 3' workflow. Single-cell libraries were sequenced on an Illumina HiSeq 2500 platform. Subsequent data analysis was performed with Seurat, a scRNA-seq analysis tool²⁰.

The datasets presented here provide the first single-cell profiles of pbMECs and cells isolated from bovine milk. They provide a suitable reference and basis for future single-cell gene expression studies, e.g., to investigate the response of specific cells to environmental disruptors including pathogen challenges.

Methods

Isolation of mononuclear cells from milk. Residual milk was collected in four 50 ml tubes from one udder quarter (total amount = 200 ml) of a clinically healthy cow immediately subsequent to a regular machine milking. The individual was in its 56th week of the first lactation and had an average milk somatic cell count (SCC) of roughly 182,000 cells/ml across the previous three lactation weeks (Fig. 1a). The collection tubes were centrifuged at $3000 \times g$ for 10 min. Subsequently, the fat layer and the supernatant were removed and 20 ml 1X PBS (Biochrom, Germany) was added to the pellet. The pellet was resuspended using wide-bore tips. After a further centrifugation for 10 min at $3000 \times g$, the supernatant was discarded and the pellet was resuspended in 10 ml 1X PBS. After a third centrifugation of the tube at $3000 \times g$ for 10 min, the supernatant was discarded again, and the pellet was resuspended in 1 ml 1X PBS. The resulting cell suspension was transferred to a 2 ml low-bind DNA tube using a Flowmi cell strainer (Sigma Aldrich, USA) to remove cell debris. Then, all cells were carefully layered over 8 ml of cold Biocoll (Biochrom, Germany) with a density of 1.077 g/ml in a 15 ml tube followed by a centrifugation step at $800 \times g$ for 30 min with the centrifuge brake set to off-mode. After centrifugation, the middle layer of the gradient was carefully collected and transferred to a 2 ml low-bind DNA tube before single-cell libraries were prepared.

Isolation of pbMECs. Primary bovine mammary epithelial cells (pbMECs) were isolated from a healthy mammary gland quarter of a cow in the 5th week of its first lactation according to the methods described by Cifrian *et al.*²¹ and Hensen *et al.*²² and modified as described by Yang *et al.*²³ Initially, the skin of an udder from

a freshly slaughtered cow was rinsed with 70% ethanol and subsequently removed. From the mammary parenchyma, cubes predominantly containing milk ducts and to a lesser extent secretory tissue were collected in HBSS (Hank's Balanced Salt Solution, buffered with HEPES, supplemented with APS solution; Sigma). Minced pieces comprising preferentially epithelial areas were immersed into 30 ml HBSS and shaken continuously for five minutes. Thereafter, large clumps were allowed to sediment until the supernatant was clear of visible clumps. The sediment was homogenized in 20 ml of HBSS supplemented with 200 U/ml collagenase, Type IV. The homogenate was shaken at 37 °C and filtered through a mesh steel grid gradient (500 µm, 300 µm, 150 µm and 90 µm) (Sigma) for retrieving isolated cells. This procedure was repeated four times in 45 min intervals. Cells were collected after centrifugation (1000 × g for 10 min), washed five times in HBSS and finally plated on collagen-coated plates (Greiner Bio-One, Germany). They were cultured in RPMI1640 medium (Biochrom AG), supplemented with prolactin, hydrocortisone, insulin and 10% FCS as described previously²³. After ~four days of unperturbed growth, fibroblasts were removed by selective trypsinization (Trypsin-EDTA (0.25%/0.02%), Biochrom AG)²³. Pure pbMECs were cultured for further six days before harvesting and preparation of a single-cell solution (Fig. 1b).

Concentration and viability of milk cells and pbMECs. A parallel test on inclusion of acridine orange (AO), but resistance to propidium iodide (PI) influx discriminated viable cells from dead cells, cell debris or milk micelles²⁴. Briefly, the cell suspension was mixed with AO/PI (Nexcelom Bioscience Ltd, UK) in a 1:1 ratio and incubated at room temperature for five minutes. Stained objects were examined using the cell counter Auto2000 (Nexcelom, UK). Only viable cells fluoresced green, while nonviable cells fluoresced bright red and milk micelles yielded no signal. The cell counter Auto2000 was also used to determine cell concentration.

Preparation of single-cell libraries, sequencing and alignment. Cells were prepared according to the 10x Genomics Single-cell Protocols Cell Preparation Guide (10x Genomics; CG00053, Rev C). The protocol is optimized to provide droplets containing only a single viable cell without contamination by cell-free nucleic acids or potential inhibitors of subsequent enzymatic reactions (e.g. reverse transcription). Briefly, cells were pelleted by gentle centrifugation and repeatedly washed with 1X PBS containing 0.04% BSA (400 µg/ml) and finally adjusted to a target cell concentration of 1,000 cells/µl for downstream experiments.

Single-cell RNA sequencing (scRNA-seq) libraries from pbMECs and mononuclear milk cells were generated with the Chromium Next GEM Single-cell 3' v3.1 assay (10x Genomics) according to manufacturer's instructions (10x Genomics User Guide Chromium Next GEM Single-cell 3' Reagent Kits v3.1 (CG000204, Rev B)). Briefly, samples were further diluted to a concentration equivalent to a target cell recovery of 10,000 cells after sequencing and subsequently they were loaded onto a 10x Genomics Single-cell 3' Chip together with the reverse transcription enzyme master mix.

On the 10x Genomics Single-cell 3' Chip, cells and gel beads, which were coated with oligonucleotides to enable mRNA capture and barcoding, were partitioned into Gel Beads-in-Emulsions (GEMs). Within GEMs reverse transcription took place. The resulting cDNA for each sample was amplified and used for library preparation using the Single-cell 3' Reagent Kit. The resulting cDNA sequencing libraries were assessed for quality and DNA concentration using a High Sensitivity DNA chip on a BioAnalyzer 2100 (Agilent Technologies) before they were sequenced on the Illumina HiSeq 2500 platform (Illumina) with sequencing parameters as recommended by 10x Genomics using two lanes per sample.

Sequencing reads were analysed using the Cell Ranger v3.1.0 alignment software provided by 10x Genomics (<https://support.10xgenomics.com>). Briefly, FASTQ files were obtained from HiSeq 2500 raw base call files via the Cell Ranger mkfastq pipeline. Subsequently, the FASTQ files were aligned to the *Bos taurus* ARS-UCD1.2 genome²⁵ with the Ensembl 98 annotation release using STAR²⁶ implemented in the Cell Ranger count pipeline, which also conducts the subsequent filtering and counting of cell barcodes and Unique Molecular Identifiers (UMIs). Those reads generated by barcode-associated cells, which passed the pipeline-internal QC, were quantified and used for establishing a gene-barcode matrix.

Bioinformatic analysis. For data analysis, we followed best-practice recommendation for scRNA-seq analysis^{27–29} including quality control, normalization and scaling of the data, regressing out technical and unwanted biological effects, feature selection, dimensionality reduction, clustering and visualization of the data (Fig. 1c) using the scRNA-seq package Seurat v3.1.4²⁰ in the R environment (R version 3.6.3). This workflow with adaptation of filtering thresholds is widely used for scRNA-seq analysis^{30–34} and described in detail below.

QC filtering. For data processing, we imported the gene-barcode matrices established by Cell Ranger into the Seurat environment. We filtered genes that were observed in less than 10 cells to remove genes that might originate from random noise. Regarding cells, thresholds for filtering were <300 genes expressed and <500 UMIs counted. Additionally, we removed cells with mitochondrial reads comprising more than 20% of all reads and with an overall complexity of gene expression under 0.8. After filtering, we retained 7,119 milk cells and 10,549 pbMECs, respectively (Table 1).

Normalization, data scaling & regression. The expression data was normalized using the “NormalizeData” function of the Seurat package to reduce possible technical bias caused by differences in sequencing depths between cells³⁵. The UMI count data for each gene were first divided by the total UMI counts of each cell, subsequently multiplied by the scale factor and finally log-transformed for normalization. A list of cell cycle-specific marker genes³⁶ (Supplementary Table S1) served for inferring each cell's cell cycle phase based on the cell's respective expression levels. The function “CellCycleScoring” in the Seurat package classified the cell regarding cell cycle stage (G1, G2/M or S) and assigned respective scores to each cell. Potential sources of unspecific variation in the data were removed by regressing out the mitochondrial gene proportion, the cell cycle effect and UMI count using

Sample	Estimated number of cells	Number of reads	Reads mapped to genome	Mean reads per cell	Total genes detected	Median genes per cell	Number of cells post filtering
Milk cells	8,218	444,840,327	93.3%	54,130	17,953	829	7,119
pbMECs	15,486	394,325,753	91.2%	25,463	17,971	1,614	10,549

Table 1. Overview of the mapping parameters for the 10x Genomics scRNA-seq datasets established for milk cells and pbMECs. Estimates were generated by Cell Ranger on the raw data. Number of cells after QC filtering in Seurat are given in the last column.

linear models and finally by scaling and centering the residuals as implemented in the function “ScaleData” of the Seurat package.

Dimensional reduction, Uniform Manifold Approximation and Projection (UMAP) & clustering. Based on expression, we identified 2,000 highly variable genes (via the function “FindVariableFeatures”). The scaled and normalized expression data of respective genes served as input for a principal component (PC) analysis, and the first 20 PCs were used to plot the variability between cells in a two-dimensional diagram by means of the UMAP procedure to reduce dimensionality of the input data. Cells were clustered into subpopulations according to the same PCs using the Seurat function “FindClusters” (resolution = 0.8), which is a graph-based clustering approach²⁰.

Identification of cluster marker genes & cell type assignment. The function “FindAllMarkers” of the Seurat package identified genes differentially expressed between a distinct cell cluster and the other clusters in the respective dataset (milk cells, pMECs). The most differentially expressed genes can be considered cluster-specific marker genes. Cell types were assigned manually to clusters based on cell type marker expression specific to each cluster. Cell type markers were retrieved from the literature^{34,37–47} and a single-cell database⁴⁸. Clusters with cells that expressed marker genes for a specific cell type at the highest levels were assigned the corresponding cell type label.

Data Records

The sequencing data have been deposited in the “Functional Annotation of Animal Genomes” (FAANG) data coordination center database (<https://data.faaang.org/home>) with accession PRJEB41576 (<https://data.faaang.org/dataset/PRJEB41576>), which directly channels all information to the European Nucleotide Archive (ENA)⁴⁹. A Seurat object containing the analysed data is available via Figshare for milk cells and pbMECs respectively⁵⁰.

Technical Validation

Mononuclear cells were isolated from milk, and pbMECs were harvested after primary cell isolation and initial cell culture. GEMs were prepared with a Chromium Controller (10x Genomics, Pleasanton, CA). Following initial cDNA synthesis, the output was amplified with 13 or 11 PCR cycles for milk cells and pbMECs, respectively. Quality control as well as quantification of the resulting PCR products was determined with a DNA high sensitivity assay on a BioAnalyzer 2100 (Agilent Technologies). The peak of the fragment size distribution was around 1,300 to 1,500 bp (Fig. 2a) and therefore indicated a good quality of cDNA synthesis. After cDNA fragmentation, adaptor ligation and PCR amplification, the resulting library quality was evaluated via a BioAnalyzer 2100 (Agilent Technologies) (Fig. 2b).

After sequencing of libraries on two lanes per sample on an Illumina HiSeq 2500 and processing of raw sequencing data with Cell Ranger v3.1.0, reads were analysed with Seurat²⁰. Both milk cell and pbMEC libraries achieved a high mapping rate for reads of 93.3% and 91.2%, respectively (Table 1), and the mapping metrics showed a similar pattern between the samples (Supplementary Table S2). For an average of 83.9% and 82.7% bases of milk cell and pbMEC reads, respectively, the phred quality score was 30 and beyond (Q30), indicating that high-quality mapping data were generated for downstream analyses⁵¹. We detected a median number of 829 expressed genes per cell for the milk cells and a median of 1,614 expressed genes per cell for the pbMECs.

Because it is essential that feature count data are obtained from viable cells, we applied strict data thresholds to exclude data that could originate from damaged cells or other technical issues⁵². A high percentage of mitochondrial transcripts indicates cells in apoptosis^{52,53}, presumably due to the loss of cytoplasmic RNA from perforated cells leading to a relative enrichment of mitochondrial transcripts. There is a particular risk for apoptotic cells within the MECs due to a cell turnover of the secretory tissue⁵⁴. Hence, we computed the mitochondrial transcript to gene read ratio in both samples and removed cells with a ratio of >0.2. We also removed cells with <500 distinct reads in a cell and <300 expressed genes detected per cell, as these indicate damaged cells. Additionally, we calculated the library complexity and filtered the data for a complexity >0.8. Supplementary Figure S1 illustrates the number of genes, unique molecular identifiers (UMIs), the percentage of mitochondrial transcript reads in each cell and the library complexity of the milk cell (Supplementary Figure S1a) and the pbMEC (Supplementary Figure S1b) dataset before and after applying the filter criteria. After QC filtering, we retained high quality data for 7,119 cells isolated from milk and 10,549 pbMECs.

Filtered raw read counts were normalized in order to account for differences in sequencing depth per cell. However, normalized data may still contain biological noise such as mitochondrial gene expression or effects of the cell cycle phase causing variability beyond the study design, which should be excluded from the data⁵⁵. Therefore, we assigned each cell a cell cycle score based on the expression of the cell cycle gene markers. To remove all signals related to cell cycle, quantitative scores for the cell cycle phases were used in downstream analyses. According to the cell cycle scores, 3,743 pbMECs and 1,195 milk cells were assigned to the G2M phase, 3,264 pbMECs and 2,146 milk cells to the S phase and 3,542 pbMECs and 3,778 milk cells to the G1 phase. All clusters

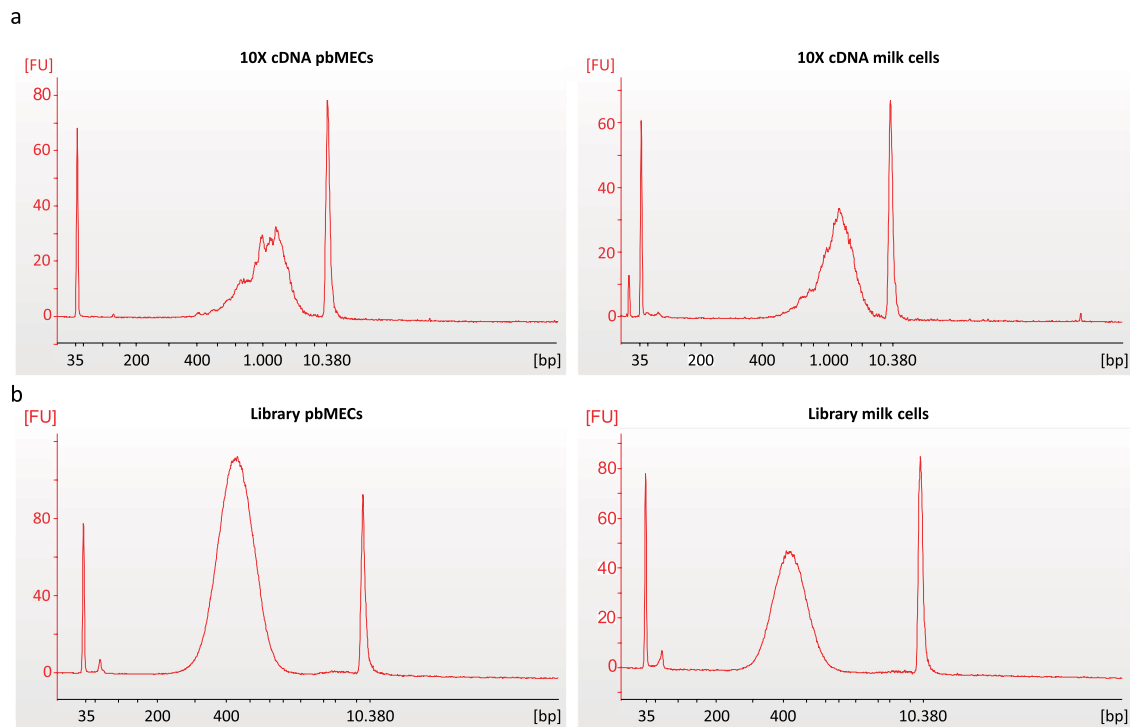


Fig. 2 Quality control of cDNA amplification and library preparation. Concentration and quality were assessed on a BioAnalyzer with a High Sensitivity DNA Chip. Size distribution was evaluated by a standard size ladder. **(a)** Size distribution of amplified cDNA from pbMECs (left) and milk cells (right). **(b)** Fragment size distribution of the sequencing libraries obtained from pbMECs (left) and milk cells (right).

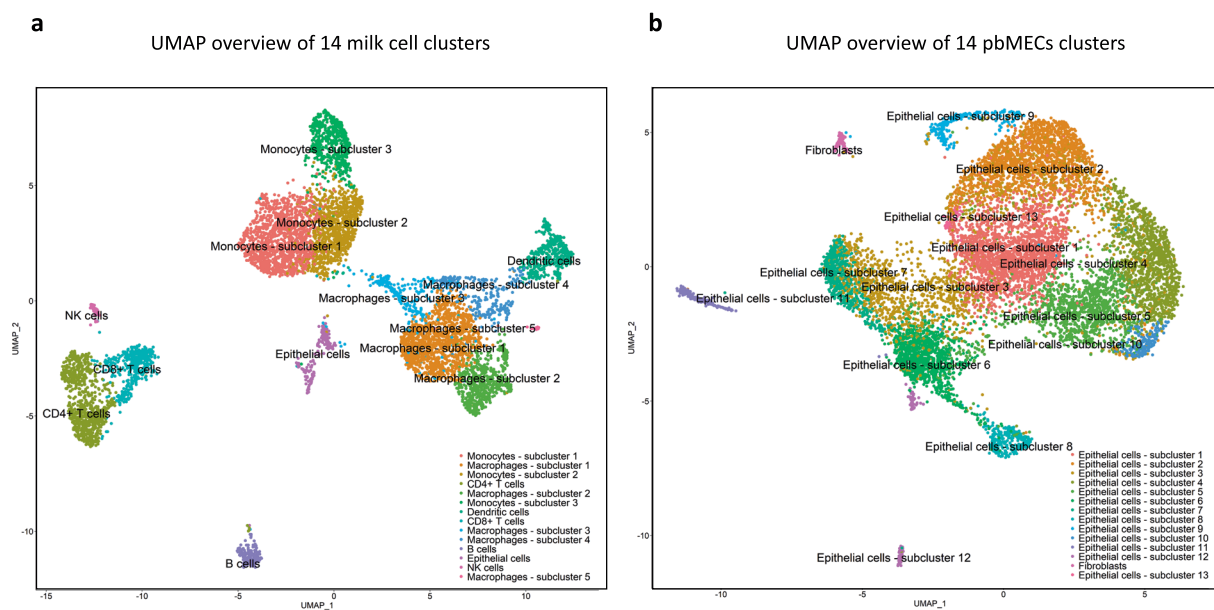


Fig. 3 Clustering of single cell data in milk and pbMECs. **(a)** Two dimensional cell clustering via Uniform Manifold Approximation and Projection (UMAP) from the first 20 principal components for the population of 7,119 milk cells. Cluster allocation of each cell is indicated by color. **(b)** Two dimensional cell clustering via Uniform Manifold Approximation and Projection (UMAP) from the first 20 principal components for the population of 10,549 pbMECs. Cluster allocation of each cell is indicated by color.

of the pbMECs were equally represented in the pbMECs data cell cycle scores, whereas the distribution of the cell cycle scores varied in the milk cell dataset clusters (Supplementary Figure S2). In addition, we scaled the data and regressed against gene count and mitochondrial gene expression.

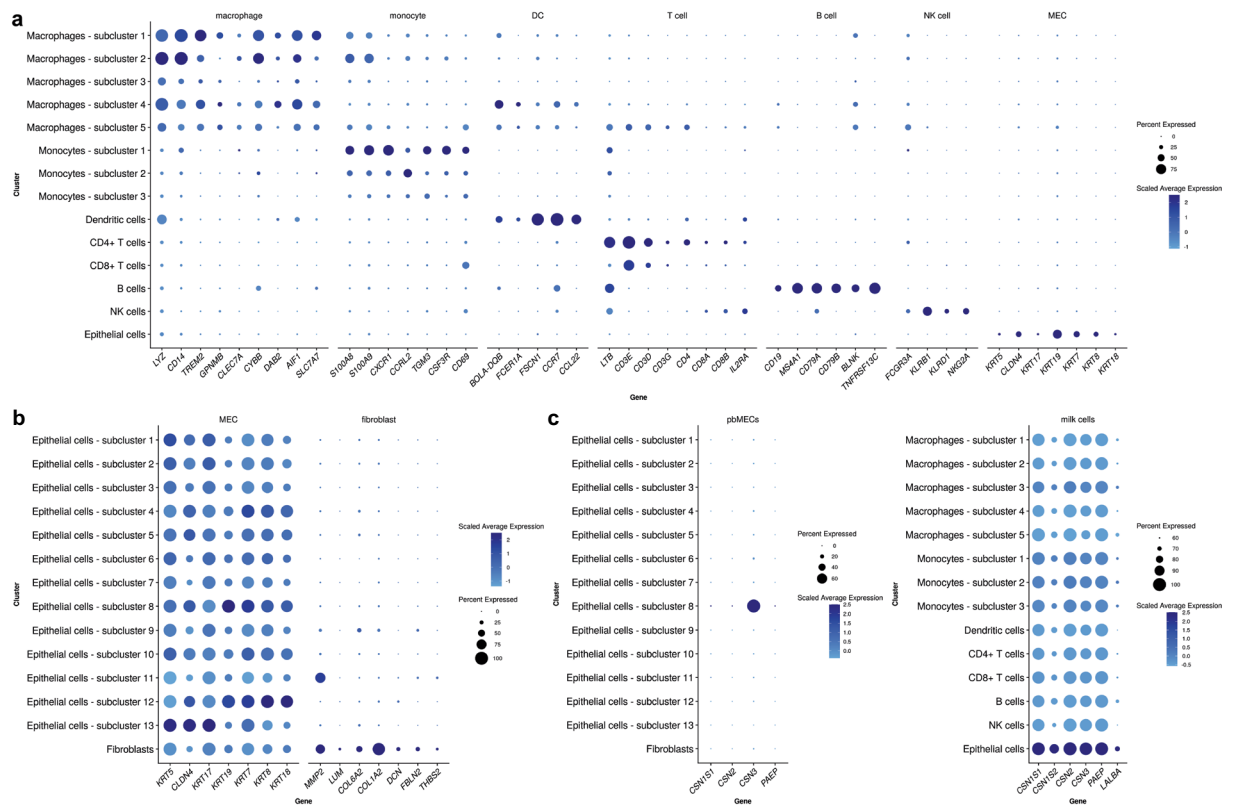


Fig. 4 Analysis of cell type specific marker gene expression. The dot size encodes the proportion of cells that express the gene, while the color encodes the scaled average expression level across those cells (dark blue is high). At the top of each dot plot, the cell types, in which the marker genes are expressed (a,b) or the datasets (c) are indicated (a) Expression of immune cell (macrophage, monocyte, DC, T cell, B cell, NK cell) and epithelial cell marker genes (MEC) in the clusters of the milk cell dataset. (b) Expression of epithelial cell (MEC) and fibroblast (fibroblast) marker genes in the pbMEC clusters. Epithelial cell markers are expressed in all clusters (left), whereas the expression of fibroblast markers is restricted to one cluster (right). (c) Expression of casein (*CSN1S1*, *CSN1S2*, *CSN2*, *CSN3*) and whey protein (*LALBA*, *PAEP*) genes in pbMECs (left) and milk cells (right). *CSN1S2* and *LALBA* were not expressed in pbMECs.

We used the scaled datasets to identify variable expressed genes and performed a dimensional reduction via PC analysis, and visualized the data using UMAP (Fig. 3). The pbMECs formed 14 indistinct clusters, whereas the milk cells formed 14 separated clusters. Differential expressed genes between clusters were identified. The 10 top-scoring genes per cluster can be found in Supplementary Table S3 for milk cell clusters and Supplementary Table S4 for pbMEC clusters, respectively. Cell cycle phase and cluster assignment are reported in Data Table 1 and Data Table 2 for the milk cell and pbMEC dataset, and can be accessed via Figshare⁵⁰.

From the literature^{12,14,56}, we expected to find several immune cell types including monocytes, macrophages and lymphocytes (T cells and B cells) as well as a low number of MECs in the milk cell dataset. Accordingly, we looked at the expression of immune cell^{34,37–42,48} and MEC markers^{43,44} throughout the clusters of this dataset for validation that the established dataset is indeed representative of a milk cell population (Fig. 4a). We detected that the cluster, which comprises 2.5% cells ($n = 178$) of the dataset was the only cluster, in which cells clearly expressed B cell markers, whereas the cluster, which contains 2.47% of all cells ($n = 176$), expressed epithelial cell markers (*KRT5*, *KRT7*, *KRT8*, *KRT17*, *KRT18*, *KRT19* and *CLDN4*, Fig. 4a). Two cell clusters contain cells dominantly expressing T cell markers, whereas cells of another cluster expressed NK cell markers. The expression patterns of three clusters were very similar. These clusters mainly expressed monocyte markers. Dendritic cell markers were expressed by a cluster containing 446 cells. Lymphocytes and macrophages are the dominant cell types in milk of healthy mammary glands^{14,56} and form numerous different sub-populations with distinct functions. Distribution of the milk cells across clusters is reported in Supplementary Table S3.

For the pbMEC dataset, we expected the cells to express characteristic epithelial cell markers, i.e., cytokeratins and claudins^{43,44}. All pbMECs clusters abundantly expressed *KRT5*, *KRT7*, *KRT8*, *KRT17*, *KRT18*, *KRT19* and *CLDN4* (Fig. 4b). Furthermore, eight of the 14 clusters displayed cytokerin genes within the top 10 cluster-specific marker genes (Supplementary Table S4).

Isolation of the pbMECs was performed from bulk mammary tissue with a trypsinization step to exclude fibroblasts. Thus, it was likely that the investigated pbMEC population still contained some fibroblasts. Hence, we checked the expression of fibroblast markers, i.e., *MMP2*, *LUM*, *COL6A2*, *COL1A2*, *DCN*, *FBLN2* and *THBS2*^{45–47}. Indeed, all these genes were mainly expressed in one cluster (Fig. 4b). Therefore, we could conclude that this

cluster of the pbMEC dataset is composed of fibroblasts (0.69% of the cells in the dataset ($n = 73$)). However, taken together, the pbMEC clusters are fairly homogeneous. Distribution of the pbMECs across clusters is reported in Supplementary Table S4. From the results of cluster expression profiles, we can conclude that the cell populations studied display characteristic expression profiles for pbMECs and bovine milk cells, respectively.

Additionally, we checked the expression of casein and whey protein genes⁵⁷ throughout both datasets (Fig. 4c). The mammary tissue and milk cells of lactating dairy is known for a low-complexity transcriptome with milk protein genes (caseins, whey proteins) as the most abundant transcripts accounting for up to 70% of all transcripts expressed^{4,58}. Indeed, milk protein genes were expressed in all clusters of the milk cell dataset, with the highest expression level in the cluster, which contains cells expressing epithelial cell markers and most likely represent MECs. On the other hand, only a few cells in the pbMEC dataset expressed casein and whey protein genes, with epithelial cells in subcluster 8 showing the highest expression level. In the lactating mammary gland, two different types of epithelial cells can be discriminated: myoepithelial and luminal epithelial cells⁵⁹. Luminal epithelial cells comprise ductal epithelial cells lining the ducts of the mammary gland, and alveolar epithelial cells, which are the actual secretory cells in the mammary gland⁶⁰. The expression pattern of pbMECs in our study clearly reflects the sampling protocol of the cells, which have been mainly isolated from ductal regions of udder tissue. Our data on cell type distribution in the pbMEC dataset are in line with this sampling preference, because the proportion of ductal epithelial cells, which do not secrete milk proteins and therefore do not express the respective genes, seems to be considerably higher than the proportion of cells with milk protein expression representing presumably alveolar cells.

Taken together, our datasets represent a valuable resource for dissecting cell-to-cell gene expression variation and cell population heterogeneity of bovine mononuclear milk cells and, to a lesser extent, pbMECs. Additionally, these data could be a reference dataset for bovine milk cell analyses, as the dataset displays the different milk cell populations varying in ontogenetic and original background.

Code availability

Auxiliary File 1 (milk cells) and Auxiliary File 2 (pbMECs) contain the R code (R version 3.6.3) for scRNA-seq data processing and subsequent visualization of the output. They are publicly accessible via Figshare⁵⁰.

Received: 22 January 2021; Accepted: 7 June 2021;

Published: 15 July 2021

References

- Loor, J. J. & Cohick, W. S. ASAS centennial paper: Lactation biology for the twenty-first century. *J. Anim. Sci.* **87**, 813–824 (2009).
- Whelehan, C. J., Meade, K. G., Eckersall, P. D., Young, F. J. & O'Farrelly, C. Experimental *Staphylococcus aureus* infection of the mammary gland induces region-specific changes in innate immune gene expression. *Vet. Immunol. Immunopathol.* **140**, 181–189 (2011).
- Brodhagen, J. *et al.* Development and evaluation of a milk protein transcript depletion method for differential transcriptome analysis in mammary gland tissue. *BMC Genomics* **20**, 400 (2019).
- Canovas, A. *et al.* Comparison of five different RNA sources to examine the lactating bovine mammary gland transcriptome using RNA-Sequencing. *Sci. Rep.* **4**, 5297 (2014).
- Jedrzejczak, M. & Szatkowska, I. Bovine mammary epithelial cell cultures for the study of mammary gland functions. *In Vitro Cell. Dev. Biol. Anim.* **50**, 389–398 (2014).
- Günther, J., Petzl, W., Zerbe, H., Schuberth, H. J. & Seyfert, H. M. TLR ligands, but not modulators of histone modifiers, can induce the complex immune response pattern of endotoxin tolerance in mammary epithelial cells. *Innate Immun.* **23**, 155–164 (2017).
- Huang, Y. *et al.* Metabolomic Profiles of Bovine Mammary Epithelial Cells Stimulated by Lipopolysaccharide. *Sci. Rep.* **9**, 19131 (2019).
- Buehring, G. C., Eby, E. A. & Eby, M. J. Cell line cross-contamination: how aware are Mammalian cell culturists of the problem and how to monitor it? *In Vitro Cell. Dev. Biol. Anim.* **40**, 211–215 (2004).
- Boutinaud, M., Herve, L. & Lollivier, V. Mammary epithelial cells isolated from milk are a valuable, non-invasive source of mammary transcripts. *Front. Genet.* **6**, 323 (2015).
- Martin Carli, J. F. *et al.* 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* (2020).
- Lipkin, E., Shalom, A., Khatib, H., Soller, M. & Friedmann, A. Milk as a Source of Deoxyribonucleic-Acid and as a Substrate for the Polymerase Chain-Reaction. *J. Dairy Sci.* **76**, 2025–2032 (1993).
- Jensen, R. G., Blanc, B. & Patton, S. in *Handbook of Milk Composition* (ed Robert G. Jensen) Ch. 2, pp. 50–62 (Academic Press, 1995).
- Lee, C. S., Wooding, F. B. & Kemp, P. Identification, properties, and differential counts of cell populations using electron microscopy of dry cows secretions, colostrum and milk from normal cows. *J. Dairy Res.* **47**, 39–50 (1980).
- Boutinaud, M. & Jammes, H. Potential uses of milk epithelial cells: a review. *Reprod. Nutr. Dev.* **42**, 133–147 (2002).
- Boutinaud, M., Ben Chedly, M. H., Delamaire, E. & Guinard-Flament, J. Milking and feed restriction regulate transcripts of mammary epithelial cells purified from milk. *J. Dairy Sci.* **91**, 988–998 (2008).
- Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nature Communications* **8**, 14049 (2017).
- Marinov, G. K. *et al.* From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510 (2014).
- Shalek, A. K. *et al.* Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
- Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
- Butler, A., Hoffman, P., Smibert, P., Papalexis, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Cifrian, E., Guidry, A. J., O'Brien, C. N., Keys, J. E. Jr. & Marquardt, W. W. Bovine mammary teat and ductal epithelial cell cultures. *Am. J. Vet. Res.* **55**, 239–246 (1994).
- Hensen, S. M., Pavicic, M. J. A. M. P., Lohuis, J. A. C. M. & Poutrel, B. Use of bovine primary mammary epithelial cells for the comparison of adherence and invasion ability of *Staphylococcus aureus* strains. *J. Dairy Sci.* **83**, 418–429 (2000).

23. Yang, W., Molenaar, A., Kurts-Ebert, B. & Seyfert, H. M. NF-kappaB factors are essential, but not the switch, for pathogen-related induction of the bovine beta-defensin 5-encoding gene in mammary epithelial cells. *Mol. Immunol.* **43**, 210–225 (2006).
24. Bank, H. L. Assessment of islet cell viability using fluorescent dyes. *Diabetologia* **30**, 812–816 (1987).
25. Rosen, B. D. *et al.* in *World Congress of Genetics Applied to Livestock Production*. Auckland, New Zealand, 7–16 February 2018, Paper no. 802, (2018).
26. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
27. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
28. Kulkarni, A., Anderson, A. G., Merullo, D. P. & Konopka, G. Beyond bulk: a review of single cell transcriptomics methodologies and applications. *Curr. Opin. Biotechnol.* **58**, 129–136 (2019).
29. Chen, G., Ning, B. & Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. *Front. Genet.* **10**, 317 (2019).
30. He, L. *et al.* Single-cell RNA sequencing of mouse brain and lung vascular and vessel-associated cell types. *Sci. Data* **5**, 180160 (2018).
31. Liu, X. *et al.* Single-cell RNA-seq of cultured human adipose-derived mesenchymal stem cells. *Sci. Data* **6**, 190031 (2019).
32. Liao, J. *et al.* Single-cell RNA sequencing of human kidney. *Sci. Data* **7**, 4 (2020).
33. Osorio, D., Yu, X., Yu, P., Serpedin, E. & Cai, J. J. Single-cell RNA sequencing of a European and an African lymphoblastoid cell line. *Sci. Data* **6**, 112 (2019).
34. Chen, J. *et al.* PBMC fixation and processing for Chromium single-cell RNA sequencing. *J. Transl. Med.* **16**, 198 (2018).
35. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
36. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
37. Mahnke, Y. D., Brodie, T. M., Sallusto, F., Roederer, M. & Lugli, E. The who's who of T-cell differentiation: human memory T-cell subsets. *Eur. J. Immunol.* **43**, 2797–2809 (2013).
38. Ziegler-Heitbrock, L. *et al.* Nomenclature of monocytes and dendritic cells in blood. *Blood* **116**, e74–80 (2010).
39. Schweighoffer, E. & Tybulewicz, V. L. Signalling for B cell survival. *Curr. Opin. Cell Biol.* **51**, 8–14 (2018).
40. Zhang, L. *et al.* Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
41. Gerrick, K. Y. *et al.* Transcriptional profiling identifies novel regulators of macrophage polarization. *PLoS One* **13**, e0208602 (2018).
42. Cochain, C. *et al.* Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. *Circ. Res.* **122**, 1661–1674 (2018).
43. Moll, R., Franke, W. W., Schiller, D. L., Geiger, B. & Krepler, R. The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* **31**, 11–24 (1982).
44. Troyanovsky, S. M., Guelstein, V. I., Tchipyshva, T. A., Krutovskikh, V. A. & Bannikov, G. A. Patterns of expression of keratin 17 in human epithelia: dependency on cell position. *J. Cell Sci.* **93**(Pt 3), 419–426 (1989).
45. Simian, M. *et al.* The interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithelial cells. *Development* **128**, 3117–3131 (2001).
46. Driskell, R. R. & Watt, F. M. Understanding fibroblast heterogeneity in the skin. *Trends Cell Biol.* **25**, 92–99 (2015).
47. Kalluri, R. & Zeisberg, M. Fibroblasts in cancer. *Nat. Rev. Cancer* **6**, 392–401 (2006).
48. Franzén, O., Gan, L. M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* **2019**, baz046 (2019).
49. *European Nucleotide Archive (ENA)*. <https://identifiers.org/ena.embl:PRJEB41576> (2019).
50. Becker, D. Single-cell RNA sequencing of freshly isolated bovine milk cells and cultured primary mammary epithelial cells. *figshare* <https://doi.org/10.6084/m9.figshare.14611485.v1> (2021).
51. Ewing, B. & Green, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**, 186–194 (1998).
52. Ilicic, T. *et al.* Classification of low quality cells from single-cell RNA-seq data. *Genome Biol.* **17** (2016).
53. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
54. Herve, L., Quesnel, H., Lollivier, V. & Boutinaud, M. Regulation of cell number in the mammary gland by controlling the exfoliation process in milk in ruminants. *J. Dairy Sci.* **99**, 854–863 (2016).
55. Barron, M. & Li, J. Identifying and removing the cell-cycle effect from single-cell RNA-Sequencing data. *Sci. Rep.* **6**, 33892 (2016).
56. Dulin, A. M., Paape, M. J., Berkow, S., Hamosh, M. & Hamosh, P. Comparison of Total Somatic-Cells and Differential Cellular Composition in Milk from Cows, Sheep, Goats, and Humans. *Fed. Proc.* **42**, 1331–1331 (1983).
57. Mercier, J.-C. & Vilotte, J.-L. Structure and Function of Milk Protein Genes. *J. Dairy Sci.* **76**, 3079–3098 (1993).
58. Wickramasinghe, S., Rincon, G., Islas-Trejo, A. & Medrano, J. F. Transcriptional profiling of bovine milk using RNA sequencing. *BMC Genomics* **13**, 45 (2012).
59. Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. *In situ* identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322–327 (2014).
60. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668.

Author contributions

C.K. and D.B. conceived and designed the project; R.W. contributed to establishing the lab methods; D.B. carried out cell isolation experiments; F.H. performed bioinformatic analysis (Cell Ranger pipeline); D.B. analyzed the data. D.B. and C.K. wrote the manuscript. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-00972-1>.

Correspondence and requests for materials should be addressed to D.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021, corrected publication 2021