Seminars in Cell and Developmental Biology

Cell and developmental biology of the mammary gland

Manuscript number: YSCDB-D-20-00002

Mammary gland development from a single cell 'omics view

Alecia-Jane Twigger[1,2*], Walid Khaled[1,2*]

Affiliation:

[1]Department of Pharmacology, University of Cambridge, Cambridge, UK.

[2]Wellcome -MRC Cambridge Stem Cell Institute, Cambridge, UK

***Corresponding authors: ajt215@cam.ac.uk, wtk22@cam.ac.uk**

Abstract

Understanding the complexity and heterogeneity of mammary cell subpopulations is vital to delineate the mechanisms behind breast cancer development, progression and prevention. Increasingly sophisticated tools for investigating these cell subtypes has led to the development of a greater understanding of these cell subtypes, complex interplay of certain subtypes and their developmental potential. Of note, increasing accessibility and affordability of single cell technologies has led to a plethora of studies being published containing data from mammary cell subtypes and their differentiation potential in both mice and human data sets. Here, we review the different types of single cell technologies and how they have been used to improve our understanding of mammary gland development.

Introduction

Whilst the key motivating factor to study the development and function of the mammary gland is to better understand the tissue of origin for the deadliest cancer in females worldwide [1], under normal homeostatic conditions it is the only secretory organ to mature in adulthood. Fundamentally, the organ is a bilayer branched ductal tree tipped by alveoli consisting of inner luminal cells and outer contractile myoepithelial basal cells which interface directly with surrounding supportive stromal cells. The main stages of mammary gland development occur during embryogenesis, puberty and in the adult in the case of pregnancy, lactation and subsequent involution [2]. Dramatic restructuring of the mammary gland occurs at each of these developmental stages, where only during lactation is the mammary gland considered functional and able to fulfill its purpose of milk production. Another source of dramatic reconstruction occurs in the case of breast cancer.

Historically, breast cancer patients were typically stratified for treatment based on the clinical parameters of age, node status, tumour stage and histological grade together with the presence or absence of key hormone receptors for estrogen (ER), progesterone (PR) and human epidermal growth factor receptor 2 (HER2) [3]. However, since then molecular portraits of human breast cancers have been identified [4] and have been developed to separate tumours into six major subtypes: Luminal subtype A, Luminal subtype B, Luminal subtype C, ERBB2+, basal-like and normal-breast like [5, 6]. Due to the clear differences between these cell types, questions subsequently arose as to whether these diverse breast cancer phenotypes arose from the same “cell of origin” or from different subpopulations.

The molecular profile of these tumour types were compared to normal human mammary subpopulations which can be enriched for using fluorescence activated cell sorting (FACS) and the markers integrin alpha-6 (CD49f) and epithelial cell adhesion marker (EpCAM) [7].

Following stromal depletion of $CD45^+$ hematopoietic and $CD31^+$ endothelial cells, enrichment and transcriptomic profiling of basal/mammary stem cells ($EpCAM^-/CD49f^+$), mature luminal cells ($EpCAM^+/CD49f^-$) and luminal progenitor cells ($EpCAM^+/CD49f^+$) revealed that each normal cell type resembled different breast cancer subtypes [7]. Despite these findings and the fact that the molecular subtypes can be used to help predict patient outcomes, we still cannot make accurate predictions as to patients' response to treatment and indeed still do not know conclusively from which mammary cell subpopulation breast cancer is derived [8]. The major hindrance to answering these questions is that studying the human mammary gland is a highly static *ex vivo* process, where only opportunistic samples from donated tissue can be studied. This means that in any given study, it is only possible to determine cell types and possible cellular hierarchies from the single developmental time point at which the genetically unique individual donates their tissue.

Examining the murine mammary gland offers the possibility to not only examine different stages of development at selected time points in genetically identical animals but also provides the opportunity to conduct extensive *in vivo* studies. Previous mammary repopulation studies, which involve the transplantation of one or more cells into a cleared murine mammary fat pad, identified a bipotent stem cell within the basal compartment of the mammary gland which was able to generate a fully functional mammary gland [9-11]. However, due to the nature of this experiment that involves radical changes in the microenvironment of the cell, it has been suggested that this bipotent mammary cell plasticity may be an induced effect that would not necessarily exist in normal biology [12]. Lineage tracing on the other hand allows for the tracking of individual cells and their progeny *in vivo*, by inducing cells expressing a specific marker to concurrently express a fluorescent and/or colorimetric protein that can be easily identified using a fluorescence microscope [13]. Whilst this technique would seemingly shed

light on whether bipotent or unipotent lineage-restricted progenitors drive glandular tissue expansion, this technique is not without its flaws. Many of the lineage tracing studies conducted in the mammary gland have produced contradictory findings where some suggest the presence of multipotent progenitors [14-17], whilst others find only unipotent progenitors contributing to mammary gland development postpartum [18-25]. Discrepancies in these studies may have arisen from a number of factors that could be attributed to the lineage tracing technique or non-specific lineage marker selection. The result of these studies is that the evidence suggesting the presence of mammary multipotent stem cells postpartum remains inconclusive and better tools to identify different cell populations in both the mouse and the human are required to answer this question. With the rise of single cell technologies, we are able to discern cell subpopulations and state from different tissues on an unprecedented level.

Single cell technologies

Technology to molecularly characterise single cells on a protein, transcriptomic, genomic or epigenomic level has rapidly developed over the past decade and allows for a better understanding of cell subpopulations and differentiation potential. With the development of single cell RNA-sequencing (scRNA-seq) and single cell ATAC-seq (scATAC-seq), cells dissociated from healthy or diseased tissue can be grouped by similar transcriptomes or epigenomes to determine cell subpopulations and maturation states in an unbiased fashion. On the other hand, single cell technologies such as single cell DNA-sequencing (scDNA-seq) can be used to determine copy number variations which provide important insights into cancer cell heterogeneity and tracking of tumour cell evolution. [26].

Single cell DNA-sequencing

Clonal diversity is an important feature of human tumours and can be explored through the use of scDNA-sequencing. After initial isolation of single nuclei from single cells, whole genome amplification is conducted before next generation sequencing is used to determine genome wide copy number profiles of single cells. Initial methodologies allowed for ~10% physical coverage of a single cell genome allowing for identification of copy number aberrations alone [27]. Methodologies such as BGI [28], nuc-seq [29], and SNES [30] that use a Phi29 enzyme to perform multiple-displacement amplifications, generated a greater than 90% coverage of the single cell genome allowing for mutations at a base pair level to be detected [31]. Another early amplification-based scDNA-seq technique, multiple annealing and looping based amplification cycles (MALBAC), also allows for the identification of single-nucleotide variations (SNVs) due to the high genome coverage (93%) offered by this method [32]. More recent transposition-based methodologies direct library preparation (DLP) [33] and DLP+ [34], do not require preamplification of DNA and therefore overcome disadvantages of coverage and polymerase bias. Careful extraction of single cells using laser-capture microdissection and catapulting, combined with adaption of highly multiplexed single nucleus sequencing (HM-SNS) [35] and tissue section image analysis, allows for topographic single cell sequencing (TSCS). TSCS allows for the capture of the genomic profile of single cells whilst preserving their cellular position and morphology within the tissue of interest [36]. This technique, along with the others described, are integral to better understanding tumor cell copy number mutations and breast cancer development, progression and evolution. However, scDNA-seq analysis alone cannot provide information on cell state or type and hence other single techniques are required to comprehensively understand mammary gland development and pathogenesis.

Flow and mass cytometry

Flow cytometry is one of the most widely used techniques of single cell analysis, using fluorescently labelled antibodies to conduct multiparameter profiling of individual cells. This technique uses polychromatic lasers and filters to detect cell protein expression profiles and is generally limited to 10 simultaneous measurements that can be made at once [37]. Building on similar concepts of flow cytometry, mass cytometry was developed to measure cells that have been labelled with isotope conjugated antibodies which can then be examined with mass spectrometry [38]. Each labelled cell is sprayed in a single droplet into inductively coupled argon plasma which vaporises the cell. During this process the cell's atomic constituents are ionised and the resulting elemental ions can be sampled using time-of-flight (TOF) mass spectrometry and quantified. The result is that the cell is permanently destroyed, however this technique allows for characterisation of a large number of proteins where up to 100 parameters per cell can be measured [37-39]. Whilst this technique is well suited to cost-effective high throughput use to investigate rare cell subpopulations, it relies on a priori knowledge of cell markers limiting its use for novel cell type discoveries.

Single cell RNA-sequencing

Another single cell tool that has been widely adopted is scRNA-sequencing which has been predominantly developed to discover and unbiasedly characterize cell subpopulations from a large number of tissue dissociated cells. Since the development of the first scRNA-seq protocol in 2009 [40], many different protocols have been published, where major differences between them include: how the cells are separated and whether mRNA transcripts are amplified to generate full length cDNA or partial coverage. Many of these techniques, along with bioinformatic tools required to analyse them, have been previously expertly reviewed [41-44]. Hence, for the purposes of this review, we will focus on the major techniques that have been

adopted in the field of mammary gland biology. These techniques include SMART-Seq 2 and the Fluidigm C1 platform as well as Drop-seq utilised by the Chromium platform from 10x Genomics.

Switching mechanism at the 5' end of the RNA template (SMART-Seq) was first developed in 2012 [45] and subsequent chemical refinements of the system led to the development of SMART-Seq2 [46], which has an increased yield and length of cDNA libraries generated for each individual cell. Cells tagged with multi fluorophore antibodies are sorted using a FACS machine into individual wells of a 96- or 384-well plate, allowing for optional index sorting of the cells. Confirmation of single cells per individual well can be subsequently conducted with a microscope before proceeding with cell lysis and downstream library preparation. cDNA is then generated from the full length of the mRNA using specially designed primers which also add sequences including a cell barcode and Illumina sequence, which allow for subsequent amplifications and analysis including multiplexing of up to 96 samples [47]. Library preparation is then conducted before pooling samples. Generation of full-length cDNA provides good read coverage across the entire transcript and allows for detection of gene isoforms or allele-specific expression by determining single nucleotide polymorphisms (SNPs). SMART-Seq 2 provides a cost-effective method to analyse hundreds of cells with high transcript coverage, however it is a laborious process involving many pipetting steps which limits the number of cells that can be easily sequenced. SMART-Seq technology has been incorporated into an automated microfluidic system, Fluidigm C1, which allows for the capture, lysis and reverse transcription of up to 96 cells on an integrated fluidic chip. As with the manual method of SMART-Seq, one can verify the capture of a single cell on the chip using a microscope. However, in this case the user is restricted to only preparing 96 cells per cartridge on the machine, limiting the number of cells that can be analysed. This problem has been

somewhat overcome by the introduction of C1 mRNA Seq HT allowing for up to 800 cells to be captured in a single run. This method, however, compromises on read coverage across the entire script due to utilising 3' end counting mRNA sequencing [41].

The technique of droplet barcoding (or Drop-seq) revolutionised the field of scRNA-seq and allowed this technology to become a more accessible tool for molecular biology. The first protocols captured single cells in nanolitre droplets together with DNA-barcoded beads using a microfluidic system [48, 49]. Within these droplets, cells could be lysed, barcoded and cDNA generated. Subsequently, the cDNA can then be pooled for PCR amplification and sheared to allow for short-read sequencing permitting partial coverage of cDNA sequences that can be aligned to infer gene expression profiles. After development of this technique, a commercially available version was provided by 10x Genomics called the Chromium controller platform. This platform allows for the capture and analysis of thousands of cells on a single cell level from 8 different samples [50] and is a cost-effective technique when calculated per cell compared to many other scRNA-seq techniques. Evidently, it is not possible to check the morphology of the cells prior to sequencing them, nor verify single cell rather than doublet capture, although this has been optimised previously [50]. With the development of cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq), detection of multiplexed protein markers can be done in conjunction with unbiased transcriptome profiling of the cells using this platform among others [51].

Single-cell ATAC-Seq

Whilst scRNA-sequencing provides in depth information of cell subpopulation gene expression profiles at a snapshot in time, only through examining gene promoter and enhancer chromatin accessibility profiles can we understand a cell's potential for different cell states.

Many of the techniques of single cell capture including array-based technologies, droplet microfluidics and combinatorial indexing through split pooling have allowed for the development of many single-cell Assay for Transposase Accessible Chromatin using sequencing (scATAC-seq) protocols [52]. The first method of scATAC-seq involved microfluidic capture of single cells using the Fluidigm system, and subsequent transposition using Tn5 transposases to tag regulatory regions by inserting sequencing adapters into accessible regions of the genome [53]. Combinatorial cellular indexing through split pooling is a technique involving two rounds of nuclei barcoding, which allows for scATAC-seq analysis to be conducted on pooled single cells, without the need to first physically separate them [54]. An alternative method of scATAC-seq, which allows for the profiling of an even greater number of cells has also been developed by 10x Genomics [55]. Nuclei of cells are isolated from a bulk single cell suspension and transposed before being loaded onto the microfluidic chip. Within the chip, gel beads in emulsion (GEMs) are generated and contain a single nucleus and a bead. The bead contains: the reagents to lyse the nuclei, oligonucleotide sequencing adapter and a priming sequence which during the linear amplification reaction incorporates a unique barcode into the transposed DNA [55]. Barcoded DNA can then be amplified before sequencing. A major disadvantage of scATAC-seq is inherent sparsity of the generated data, which arises due to the fact that diploid cells have low copy numbers. Despite the powerful computational tools developed to analyse such sparse data, it is thought that only 1-10% of the total accessible peaks are detected [52], where much of the scATAC-seq analysis heavily relies on single cell transcriptomic data to infer cell types.

Limitations of single cell tools

Evidently for each of the above single cell analysis techniques, it is possible to gain invaluable insights into cellular behaviour or differentiation potential, however no one technique is

omnipotent. Whilst scDNA-seq provides information about mutational states of cells through copy number variations or SNVs, it may be limited in its coverage and does not provide any information on cell type, function or state. On the other hand, techniques that aim to characterise immediate cell subpopulations either by examining the direct transcriptional profiles (scRNA-seq) or gene expression products (flow/mass cytometry), are not always in agreement due to post-transcriptional modifications. Mass cytometry allows for identification of lowly expressed proteins which otherwise might be undetected in scRNA-seq data (particularly droplet-based platforms) due to dropout of low-level transcripts. Unlike scRNA-seq which unbiasedly classifies cells based on their whole measured transcriptional profile, mass cytometry relies on appropriate marker selection and antibody availability, without which cells might be misidentified. As mentioned above, only a fraction of the total chromatin accessible sites can be extracted using scATAC-seq, where identification of different cell subpopulations relies on mapping this limited data onto available transcriptomic information. To overcome the limitations of these different single cell techniques, multi-omic approaches aiming to combine different single cell technologies (such as CITE-seq) are emerging and have recently been expertly reviewed [56]. Along with emerging techniques to collect data from different single cell profiling methods, it is important that methods to analyse the increasing amounts of data involving machine learning and novel algorithms continue to evolve.

Single cell analysis tools

Development of machine learning bioinformatic tools to analyse single cell data is as important as improving the techniques themselves to advance our understanding of tissue subpopulations and differentiation trajectories. As demonstrated above each single cell technology is unique and often requires its own analysis pipelines and tools. Many of these have been previously expertly reviewed for analysis of: scDNA-seq [57], mass cytometry [58, 59], scRNA-seq [60-

62] and scATAC-seq [52] data. Visualisation of single cell datasets has been improved as dimension reduction tools have evolved from: principle component analysis to t-distributed stochastic neighbour embedding (tSNE)[63] to uniform manifold approximation and projection for dimension reduction (UMAP) [64] to diffusion map graphical representation [65]. Tools such as diffusion pseudotime analysis use input from different single cell techniques and construct a temporal order of differentiating cells by measuring transitions between cells using diffusion-like random walks [66] and allow for cellular hierarchies to be inferred. Other interesting tools for scRNA-seq data include: scGen which uses machine learning to model perturbation and infection response of cells across cell types, studies and species [67] and Markov affinity-based graph imputation of cells (MAGIC) which determines gene interactions [68]. On the other hand, ChromVar uses scATAC-seq data to predict transcription factor associated accessibility and allow clustering of epigenomically profiled cells [69], whereas more recent tools such as Model-based AnalysEs of Transcriptome and RegulOme (MAESTRO) directly integrate data from both scRNA-seq and scATAC-seq experiments [70]. These are a few of the hundreds of tools available that can be used to aid in the biological interpretation of the big data generated by single cell analysis.

Using single cell technologies to understand mammary gland biology

Since the publication of the first single cell manuscripts characterising the postnatal mammary gland [71, 72], an increasing number of studies (**Table 1**) have focused on comparing mammary cells across different stages of normal development (**Figure 1**). The focus of these studies is to comprehensively characterise mammary cell subpopulations and differentiation trajectories using mass cytometry, scRNA-seq and scATAC-seq, to answer the key question as to whether a rare population of bipotent progenitors exist past embryonic development.

Embryonic mammary gland composition

Recent studies have compared embryonic and post birth mammary cell subpopulations (**Figure 1**) using scRNA-seq and ATAC-seq (**Table 1**) in an attempt to answer whether bipotent progenitors giving rise to both luminal and basal cells exist postnatally. Using multicolour lineage tracing and scRNA-seq, Wuidart et al. identified that at embryonic day 14 (E14) a single population of embryonic multipotent progenitors (EMPs) exist in the mouse mammary gland. These cells exhibit a hybrid transcriptional signature comprising of marker genes from both the luminal and basal lineages [23]. Within this study it was found that p63 is a master regulator of basal cells, where overexpression in luminal cells is enough to convert them into a basal phenotype. One of the limitations of this study is that only a small number of cells were analysed by single cell analysis and that comparisons between different developmental time points were limited (69 EMPs compared to 51 adult basal cells and 73 adult luminal cells). However, within the same year, another study was published by Giraddi et al. which compared the single cell profile of cells during embryogenesis (E16 and E18), postnatally at day 4 (P4) and from the adult mouse mammary gland. As was observed in the previous study, mammary cells from E16 and E18 generated single separate clusters of foetal cells, whereas postnatally both luminal and basal cells could be identified which was further separated in the adult gland into mature and alveolar luminal clusters and a separate basal cluster [73]. Original bulk ATAC-seq analysis by Dravis et al., referenced in Girraddi et al., showed that E18 fetal mammary stem cells (fMaSCs) presented open features at distal enhancer and proximal promoter regions of both luminal and basal genes [74]. Using single nucleus ATAC-seq (snATAC-seq), follow up studies confirmed that during late embryogenesis (E18), individual cells displayed either a basal-like (*Krt5, Acta2*) or pan luminal/luminal progenitor-like (*Krt8, Krt18, Kit*) chromatin accessibility profile together with fetal-enriched genes such as Sox10 and Sox21 [75]. snATAC-seq and scRNA-seq data from Giraddi et al. was integrated to

examine the developmental trajectory which found that both E16 and E18 fetal cells remain tightly clustered. Whilst clustered cells at E18 were found to be indistinguishable on a transcriptomic level, lineage priming was found using snATAC-seq, where cells displayed accessibility either to luminal or basal cell fate [75]. Subsequently, only postnatal lineage restricted luminal or basal cells were present, which upon onset of puberty the luminal cells further split into luminal alveolar and hormone responsive luminal cells (see **Table 2** for a summary of the commonly used nomenclature for these populations).

Adult mammary gland epithelial cellular composition

Even through the dramatic changes occurring during pregnancy, lactation and involution in the adult mammary gland, scRNA-seq findings from Bach et al. support the findings that only lineage restricted progenitors exist postnatally. Epithelial cells taken from virgin, pregnant, lactating and involuting mice were characterised to find distinct subpopulations of cells from either the luminal or basal lineages [71]. These cell clusters were ordered along a pseudotime trajectory and it was found that both luminal and basal cells did not have a common progenitor [71]. The luminal cells however were found to have a continuum of differentiation stemming from *Aldh1a3*$^{+}$ luminal progenitors and extend to the two major lineages of secretory alveolar and hormone responsive. Clearly replicating such experiments in humans would prove difficult, hence studies conducting scRNA-seq of human milk cells have provided novel findings on the maturation of functional human mammary cells during lactation [76, 77]. Recent work from our lab examining milk cells from 4 donors and 4 non-lactating breast tissue samples found two populations of secretory luminal milk derived cells which are transcriptionally similar to luminal progenitor cells from non-lactating breast [76]. This work suggests that luminal progenitors may give rise to secretory alveolar cells in humans, as well as in mice. Furthermore, another recently published study has found that aberrant

differentiation of luminal progenitor cells, resulting in expression of key milk proteins *Lalba*, *Csn2* and *Wap*, may provide an indication of early tumorigenesis in Brca/p53 mouse models [78]. This highlights the importance of resolving the role luminal progenitors play in development to better understand breast cancer development. Findings of a common luminal progenitor in the adult mammary gland is in agreement with findings from Giraddi et al. but are in contrast to the recent lineage tracing studies which suggest that these two luminal cell lineages are unipotent in adult mice [22, 79-82]. It is possible that the bipotent luminal cells identified in these scRNA-seq studies are active during early embryonic and prepubertal stages and that later in development, the two luminal lineages are maintained by their respective progenitors. Interesting, it was found that parity induced *Aldh1a3*$^{+}$ luminal progenitor cells exist post involution, that were seemingly primed towards the alveolar lineage and expressed many lactation associated genes such as *Lipa*, *Xdh* and casein genes such as *Csn2* and *Csn3* [71]. Findings from a recent study that combined scATAC-Seq and scRNA-seq to study 10-week-old adult mice corroborated findings from the Bach study, identifying a single myoepithelial population as well as a *Foxa1*$^{+}$ hormone responsive luminal population and *Elf5*$^{+}$*/Kit*$^{+}$ luminal cells (termed L-sec for luminal secretory, see **Table 2**) [83]. The L-sec cells were further divided into *Rspo1*$^{+}$*/Aldh1a3*$^{+}$ luminal progenitor and *Lalba*$^{+}$*/Csn2*$^{+}$*/Lipa*$^{+}$ lactation progenitor cells which according to pseudo temporal analysis suggests that the former gives rise to the latter cell subtype [83]. A limitation of this study was that analysis was only conducted at one time point, where comparisons between the gene expression profiles of the discovered lactation progenitor and previously described secretory alveolar cells [71] would have developmentally contextualised these findings.

Similar cell populations were reported in the adult human mammary gland at a single time point [84, 85]. Mass cytometry of normal primary mammary tissue identified epithelial subsets

of basal, luminal (corresponding to mature luminal, see **Table 2**) and luminal progenitor cells [85]. Strikingly, within the latter population, a small subset of cells displayed an elevated content of active caspase-3 whilst maintaining clonogenicity, suggesting potential greater genomic instability and increased risk of oncogenic transformation in this set of luminal progenitor cells [85]. Nguyen et al. also identified 3 major epithelial cell types using scRNA-seq which they termed myoepithelial, luminal 1 (L1) and luminal 2 (L2) cells (see **Table 2**). The *SLPI*$^{+}$ L1 population was split into *ELF5*$^{+}$*/KIT*$^{+}$ L1.2 (resembling luminal progenitor cells, **Table 2**) and LTF^{+} L1.1 cells (resembling secretory alveolar cells, **Table 2**). L1.2 luminal progenitor cells were found to sit above the other luminal cell types on the differentiation hierarchy. L2 cells expressed the marker *ANKRD30A* together with hormone receptors *ESR1*, *PGR* and *AR* indicative of a hormone responsive subpopulation (**Table 2**). Interestingly, upon closer examination of the basal cluster identified using 10x Genomics scRNA-seq, a subcluster of cells highly expressed contractile genes (*ACTA2, TGLN, KRT14*) which the authors subsequently identified as specific myoepithelial cells [84]. However, data from a recent pre-print suggests that differential expression of adhesion markers may be an artifact of long digestion duration rather than due to distinct subclusters in human mammary cell subtypes [86]. Within the Nguyen paper, the authors manually set the start of the pseudotime to be within the basal cell type and found that the resulting trajectory differentiated from the basal cells to a myoepithelial branch and a bifurcated luminal branch, topped by L1.2 luminal progenitors [84]. Whilst the studies described thus far appear to represent a concordant view of mammary cell subpopulations in the embryonic and postnatal mammary gland of both mice and humans, contradictory findings have been published [72, 87-89] and reviewed [90, 91] suggesting bipotent progenitors existing postnatally.

Multipotent lineage progenitors in the adult mammary gland

Similar to Bach et al., another study published at the same time [72] investigated postnatal mammary cell subpopulations in mice at slightly different developmental time points (pre-puberty, puberty, adult and pregnancy as well as different stages of the oestrus cycle) however came to different conclusions. Findings from this study suggest that prior to puberty a majority of epithelial cells consist of a basal phenotype, where a rare basal-like population expressing some luminal features is positive for the marker *Cd55* [72]. During puberty this population seemingly expands, where after puberty the *Cd55*$^{+}$ cells predominantly reside in the luminal cell compartment. Further luminal subpopulations were identified including intermediate (sharing markers of both luminal progenitor and hormone responsive cells) and mixed (expressing both *Kit/Elf5* luminal markers and *Acta/Krt14* basal markers) cell subpopulations [72]. Detailed discussion on this study, together with other early mammary scRNA-seq studies that classified new cell types based on expression of markers such as c-Kit (*Kit*), can be found in [92]. A more recent study in mice also found a hybrid luminal cell subpopulation, which expressed genes characteristic of both hormone sensing (HS) and alveolar progenitors (AV) termed HS-AV cells, which decreased with age [93]. Both the described "intermediate" [72] and "HS-AV" [93] luminal cells expressed higher levels of *Prlr* and *Cited* than luminal progenitor/alveolar cells, and higher levels of *Csn3* and *Trf* than hormone responsive cells. Li et al. went further and confirmed using immunofluorescence these double luminal lineage cells by illustrating co-staining of progesterone receptor with milk fat globule-EGF factor 8 (PR^{+}/$MFGE8^{+}$) and estrogen receptor with lactoferrin (ER^{+}/LTF^{+}) [93]. Despite the identification of new luminal cell types, Pal et al. did not observe a separate rare population of mammary stem cells positive for *Epcam* and *Procr* or *Lgr5* and *Tspan8* that was expected according to data from previous studies [72]. Similarly, Sun et al. examined epithelial cells taken from the adult mouse in the virgin and pregnant state and also noted an absence of a

unique population of cells that expressed *Lgr5* and *Tspan8* [87]. Findings from this study observed a *Cdh5*$^{+}$/*Procr*$^{+}$ population of cells that resided in the basal compartment, which proliferated *in vitro* and had *in vivo* mammary repopulation potential [87]. Analysis of developmental trajectories suggested that this cell type was at the top of a mammary epithelial cell hierarchy during pregnancy, however this was not the case in the virgin gland.

Follow up studies utilised novel bioinformatic tools to compare across different data sets in an attempt to reconcile the differences found between the single cell studies to date. One such study developed a model called Landscape of Single Cell Entropy (LandSCENT) which provides a potency score for single cells independent of using lineage markers which they then used to define root states for lineage trajectory algorithms [88]. Using this method, this study reanalysed an adult human data set [84], where they identified high potency cells mapping to the periphery of the basal and immature alveolar clusters. Cells that sit in this bipotent state overexpressed transcription factors *YBX1* and *ENO1* which have been implicated in basal breast cancer risk and were suggested as potential markers identifying multipotent progenitors [88]. However, throughout this paper it was noted that data from the Nguyen et al. study was not of a sufficient quality to identify stem cell populations without this technique and that that future studies should investigate more cells to a greater depth. A more recent study also conducting retrospective comparisons between many data sets and, in addition, produced a novel dataset characterising normal human mammoplasty cells from 3 donors together with cultured cells [89]. Whilst this study's main focus was on the epithelial compartment of mammary cells, they captured all noted epithelial populations together with stromal and immune cell subtypes [89]. Primary mammary cells were cultured in 2D on top of irradiated fibroblasts to conditionally program the cells [94]. After which, the authors identified different "hybrid" epithelial cell populations emerging including basal/luminal (*KRT14*$^{+}$/*KRT18*$^{+}$),

epithelial/mesenchymal (*EPCAM*$^{+}$*/VIM*$^{+}$) and some that expressed all 4 markers (quadruple positive cells, *KRT14*$^{+}$*/KRT18*$^{+}$*/EPCAM*$^{+}$*/VIM*$^{+}$). In particular, they found by comparing their data with previous studies [71, 73, 84] that normal human mammary tissue mapped to the adult mammary tissue of mice and conditionally reprogrammed cells map closer to the stem-like population of embryonic mouse cells. After comparison with all these time points, they noted that the "hybrid" cells existed at different time points predominantly during the *in utero* period, gestation and lactation. This paper provides a first look at all the mammary cell subpopulations (including stroma) that exist in the adult human mammary gland but does not take into account differences in marker expression profiles between mice and human cells, namely that KRT14 expression has been routinely observed by pathologists in both luminal and basal cells in the adult human [95]. Accurate use of markers to define different cell types is important when describing whether cell subpopulations contain markers from both luminal and basal cells. In addition, it is important to remember that cells captured in these studies are done so at a single snapshot in time and that their temporal trajectories are inferred only from these single time points. Therefore conclusive statements of biopotency must be further verified using *in vivo* models, as discussed by Watson and Khaled [96], and may take advantage of emerging technologies such as scGASTALT/LINNAEUS [97, 98] that combine unbiased lineage tracing with scRNA-seq-detectable genetic scaring.

Challenges remaining in the face of generating a unified mammary cell atlas

Before we attempt to understand the cells driving breast cancer formation, it is becoming increasingly clear that we need to unify findings from mammary development studies to be able to describe which mammary cell subpopulations universally exist, which species-specific markers they express and define commonly used nomenclature for each of these cell subtypes. To this end, we have attempted to summarize the major mammary single cell studies (**Table**

1) and common nomenclature used for the four major adult epithelial subtypes (**Table 2**). Not all studies identify the same mammary subtypes, however most studies agree that cell subpopulations such as adipocytes or secretory alveolar cells during lactation (see Bach et al.) have been incompletely profiled due to difficulties in isolating intact cells during dissociation. Future efforts to build a mammary cell atlas must explore different methods to dissociate tissue/isolate cells, such as modifying digestion duration [86] or temperature of enzymes [99], to overcome certain cell types being disproportionately enriched or depleted during sample preparation.

As described above, some subclusters classified as epithelial cells were inimitably observed, including $Cd55^+$ [72] and $Cdh5^+$ [87] subclusters or *ZEB1*$^+$*/TCF4*$^+$ [84], *VIM*$^+$*/EPCAM*$^+$ and *KRT18*$^+$*/KRT14*$^+$ cells [89]. $Cd55^+$ and $Cdh5^+$ clusters highly express genes such as *Axl* and *Sparc* or *Cd36* and *Pecam1*, that have been associated to fibroblasts and endothelial stromal subpopulations [100]. This suggests the utilised CD24/CD29 sorting strategy may not have been sufficient to exclude contaminating stromal cells. Similarly, a cluster of $Procr^+$ cells classified as basal cells also expressed some but not all markers of pericytes and hence the authors conceded that these cells may be contaminating non-epithelial cells [71]. Further rare epithelial cell subtypes such as *ZEB1*$^+$*/TCF4*$^+$ cells [84], *VIM*$^+$*/EPCAM*$^+$ cells and *KRT18*$^+$*/KRT14*$^+$ cells [89] were identified by marker expression alone, where cells did not form separate clusters arising from clustering or dimension reduction analysis. As such, without orthogonal validation (such as tissue staining) it is difficult to determine the biological relevance of these results due to potential technical error such as doublet artefacts or detection of spurious transcripts. Clearly, it is remarkably difficult to characterise mammary cell subpopulations based on a limited number of markers. Hence, moving forward, cells profiled using single cell technologies should be classified into subpopulations using the cell's global

transcriptome, preferentially utilising cell signatures made up of an abundance of different validated genes.

Indeed, the issue of unifying diverse cell types across different organs is a universal problem of cell biology and various consortiums have begun to catalogue different cell types using single cell technologies. A recent study by Han et al. characterized a range of different tissues including: the mammary gland, lung, kidney, testis and placenta using scRNA-seq as a resource in developing the mouse cell atlas [101]. Specifically, in regard to the mammary gland, tissue was examined from virgin, pregnant, lactating and involution tissue. Whilst many of the briefly described findings fit what has been described in studies above [101], due to the lymph nodes not being removed prior to tissue digestion it is hard to distinguish whether the reported immune cells are lymph node-resident or mammary tissue-resident. Another similar study conducted by the Tabula Muris consortium attempted to discern how many different organs aged differently [102]. Cells of different organs were extracted and sequenced from mice between one month to 2.5 years. Interestingly, it was noted that for the mammary gland, there was a significant decline in T cells (which could be confounded by the lymph nodes potentially not being removed prior to gland dissociation), and overall there was an upregulation of API transcription factor family (*Junb, Jund, Fos*) with age [102]. Findings from these studies highlight that it is important to consider factors outside of normal development such as age that might influence mammary cell subpopulations and gene expression profiles.

The impact aging, menopause, parity and obesity have on the adult mammary gland

The effect ageing, menopause, parity and obesity have on mammary cell subpopulation proportions has been investigated across a number of studies [93, 100, 103, 104].

Findings from an article examining primary mammary cells of healthy women using mass cytometry observed age-related increase in certain luminal cell subpopulations with a corresponding decrease in myoepithelial cell proportions [104]. Similarly, a recent article examining the mammary glands of young compared to aged mice has found that the gland contains altered proportions and gene expression profiles of both epithelial and stromal cells [93]. Interestingly, they find an increase in the proportion of epithelial cells as well as identifying an increase specifically in the numbers of alveolar cells. It was also noted that in myoepithelial cells, there was a decrease in the expression levels of genes associated with basement membrane protein synthesis as well as an increase in inflammatory cytokine production by the myoepithelial, vascular endothelial and macrophage cells [93]. These findings suggest dramatic changes occurring in the stroma in aging cells that may provide a pro-tumourigenic microenvironment.

As a result of aging, menopause is a risk factor for breast cancer development and similarly various changes in the mammary cell subpopulations may provide a mechanism behind this phenomenon. In the context of surgical menopause (through an ovariectomy), the impact estrogen (17beta-estadiol, E2) supplementation with or without exposure to harmful endocrine disrupting environmental contaminant PBDE (polybrominated diphenyl ethers) has on mammary cell subpopulation proportions was investigated [100]. This study finds in the menopausal vehicle mouse there is an abundance of fibroblasts and stomal cells with a lower proportion of epithelial cells. However, the use of E2 induces the regrowth of terminal end bud like structures which is enhanced in the presence of PBDE. Administering ER and PBDE together induced expression of progesterone receptor (PR) independent of estrogen receptor (ERα22, $Esr1^{-}$) cells in two luminal cell populations as well as an increase in M2 macrophage

populations which likely contribute to the mammary tissue remodeling and production of a pro-tumour microenvironment.

Another set of factors thought to change the risk of developing breast cancer include parity and obesity. Murrow et al. in a recently updated preprint explored differences in the human mammary gland composition as a result of either obesity or parity [103]. It was found that parity increased the number of myoepithelial cells and transcriptional response of hormone responsive cells in the mammary gland, as well as altering the size of alveoli. Interestingly, they found that luminal progenitor proportions do not seem to correlate with parity and that hormone responsive cells decrease in number in obese women. Coming from a slightly different perspective, another study examined the impact high fat diet (HFD)-induced obesity had on the stomal compartment of the mammary glands of Brca1-/-, p53+/- breast cancer model mice [105]. This study found that HFD induced an increase in the expression of extracellular matrix genes (*Col3a1, Col6a3, Eln* and *Sparc*) and pro-tumourigenic M2 macrophage markers in monocytes suggesting an alteration in stromal cell function and microenvironment in obese breast-cancer prone mice. From these studies, it is clear that certain risk factors such as age, menopause and obesity alter both the cellular states and proportions of epithelial and stromal subtypes and suggest that careful characterization of stomal cells is essential to understand breast cancer development.

Understanding mammary stroma diversity is integral to understanding breast cancer development

Theories surrounding the mammary stroma inducing a wound healing phenotype elicited in breasts with a high mammographic density, during obesity and in post-lactational remodelling have been put forward as a mechanism for breast cancer development [90]. As such, it is

552 important to carefully characterize normal stromal cell compositions and how these might
553 change as a mammary gland acquires a tumourigenic phenotype. Indeed, an early scRNA-seq
554 study that profiled cells from breast tumours found that many of the cells with a normal
555 phenotype (as identified by estimates of copy number variations) were immune cells [106].
556 Increasing interest in the macrophage proportion of mammary immune cells led one group to
557 investigate the $CD45^+$ population of 10-week-old murine mammary glands. Using scRNA-seq
558 they identified a *Lyve-1*$^+$ subpopulation of tissue resident macrophages which expressed high
559 levels of extracellular matrix remodeling associated genes [107]. In human tissue, arising from
560 either breast tumour or matched normal breast tissue, scRNA-seq analysis of the $CD45^+$
561 compartment found that immune cell subpopulations identified in normal tissue were only a
562 small subset of those identified in tumours [108]. Interestingly they often found M1 and M2
563 macrophage associated genes in the same cell which were positively correlated suggesting that
564 each state might not be as mutually exclusive as previously thought and cell state might exist
565 instead along a continuum.
566
567 Another key stomal cell type of interest are mammary fibroblasts and in particular their
568 similarity to cancer associated fibroblasts (CAFs). Focusing on sorted fibroblasts extracted
569 from a precancer MMTV-PyMT mouse mammary cells, three major populations of CAFs have
570 been identified including: vascular (vCAF, with a subset of cycling CAFs, cCAFs), matrix
571 (mCAF) and developmental (dCAF) associated CAFs [109] . These cells were found to be
572 transcriptionally and spatially distinct, where vCAFs were found to come from a perivascular
573 location, resident fibroblasts gave rise to mCAFs and dCAFs are malignant cells that have
574 undergone epithelial to mesenchymal transition [109]. On the other hand, a subsequent study
575 investigating the CAFs and normal fibroblasts in triple-negative breast cancer (TNBC)
576 modelling BALB/c-derived 4T1 mammary tumours and identified 6 CAF subpopulations

[110]. These subpopulations included: Ly6c1^{high}, α-SMA^{high}, Cd53^{high}, Crabp1^{high}, Cd74^{high} and cycling CAFs which were identified in the above study. This study further compared their findings to those of Bartoschek et al. and found that mCAFs were enriched for *Crabp1*, and other markers expressed by Crabp1^{high} cells, suggesting that these cells may represent similar subpopulations. A recent study by Wu et al. examined the epithelial and stromal compartment of primary TNBC tumours from 6 patients and found aside from the expected normal mammary subpopulations a basal epithelial cancer cluster, two CAF subpopulations and two perivascular like (PVL) subpopulations [111]. The noted subpopulations included: myofibroblast-like CAFs (myCAFs), inflammatory CAFs (iCAFs) and mature and immature PVL cells. Here the authors suggest that the cells described as vCAFs in the Bartoschek et al. study, may be PVL cells identified by this study [111]. Not all CAF subsets were identified across studies, highlighting the need to synchronize findings as much as possible to determine translatable cell subpopulations.

Profiling breast cancer tumours

Many single cell techniques have been utilized to examine human mammary tumours and may be considered the beginning of a breast cancer cell atlas. A recent study by Wagner et al. used mass cytometry to examine the expression of 73 proteins across 144 human breast tumours and 50 non-tumour tissue samples to determine different cell subpopulations [112]. It was found that despite tumour subtype, there was a significant amount of cellular variation within individual tumours. Common to all breast cancer subtypes however, were PD-1^{+} T-cells and PD-L1^{+} tumour associated macrophages (TAMs). Another recent mass cytometry study also generated a cell atlas but instead examined 62 breast cancer cell lines and five lines from healthy tissue [113].

Comparisons between breast cancer cell lines and primary breast cancer tumours were made by Gao et al. using a novel technique of single-nucleus RNA-sequencing [114]. Similar to Wagner et al., different subpopulations of cells could be found within a single tumour. However, through examining all populations together, a rare subpopulation of highly proliferative cells was identified that upregulated breast cancer associated genes [114]. Focusing more closely on six triple negative breast cancer tumours, Karaayvaz et al. found that although most clusters arose from separate tumours from different patients, one cluster was contributed by all tumour samples [115]. This subcluster had a gene expression profile associated with multiple signatures of treatment resistance and metastasis and was characterized by activated glycosphingolipid metabolism and associated innate immunity pathways [115].

Using spatial transcriptomics, Salmén et al. instead examined another breast cancer subtype, HER-2 positive tumours and found that by integrating topographical information from the tumour section images and spatial spots, immune cells infiltrated into the invasive regions of the cancers [116]. In contrast, Andersson et al. who also used spatial transcriptomics but instead integrated image analysis with scRNA-seq data, found that epithelial cells remained segregated from the cross-talking stromal cells [117]. In many tumours, plasma cells were spatially segregated from B-cells, which were found to rather colocalize near T-cells in several patients [117]. Spatial information has also been integrated with scDNA-sequencing, where analysis of breast tissue from patients with ductal carcinoma *in situ* (DCIS) with invasive ductal carcinoma (IDC), found that genome evolution occurs in the ducts before tumour cells escape the basement membrane and begin generation of the IDC tissue [36].

Analysis of tumour cell genome and epigenome is essential to identifying how cancer cells acquire mutations and progress to become metastatic. Use of HM-SNS (see single cell DNA-sequencing section), found that copy number aberrations (CNA) were acquired in the earliest stages of triple negative breast cancer evolution in short, punctuated bursts rather than gradual evolution, as was previously thought [35]. Interestingly, Wang et al. found that triple-negative tumour cells had an increased mutation rate compared to ER+ cells [29]. Further to findings of Gao et al. [35], Wang found that aneuploid arrangements arose early in tumour development and remained stable across clonal expansion [29]. A more recent study using single-cell copy number analysis found that pseudo-diploid single cells existed across different PAM50 subtypes and contained cancer specific alterations (such as 1q gain and 16q loss), suggesting intrinsic genomic stability and providing a potential identifying feature that could be used in early disease stage risk assessment [118]. Analysis of chromatin states in patient-derived treatment-resistant xenograft models found that a subset of cells across different tumours shared a common chromatin signature consisting of a loss of H3K27me3 chromatin mark [119]. Another study examining the epigenome of breast cancer cells focused particularly on the DNA methylation patterns of circulating tumour cells and found that binding sites for stemness and proliferation-associated transcripts are specifically hypomethylated [120]. Utilization of many different single cell tools provides new understandings about breast tumour cell diversity and growth potential; however, in many cases normal mammary cells may not have been profiled in the same way and thus cannot be compared. Where similar technologies have compared tumour and normal mammary cells it is often the case that they have vastly different profiles, making it difficult to identify intermediate cells and discern the cell of origin for cancer. Increasing numbers of studies are now focused on examining normal mammary tissue from individuals with high breast cancer risk that may assist in identifying intermediate pre-clinical cells that can be targeted for early detection of disease.

Conclusions

As is evident from the increasing number of studies utilising these tools, single cell technologies is playing an increasingly important role in understanding the cellular composition of the mammary gland. Here we presented an overview of many of the findings from the emerging studies, which together are beginning to bring us a clearer picture of the different mammary subpopulations that exist in both the human and murine gland and how they change in normal development and cancer. Increasingly, there is a need to better unify findings from different studies through the development of consortiums such as the human cell atlas. Importantly to remember, is that disparate isolation techniques may influence findings of different cell subpopulations, and computational tools may be required to correct for batch or indeed cellular isolation effects that may have provided discordant views of cell subpopulations in previous literature. Whilst the power of computational tools to overcome limitations of different cell isolation techniques may be required, it is important to remember that *in silico* modelling also has its limitations. All single cell technologies to date only allow for the capture and examination of single cell profiles at a snap-shot in time, where any differentiation trajectories in the cell populations have only been inferred *in silico*. Hence, it is important that other *in vitro* and *in vivo* techniques such as organoid modelling or lineage tracing experiments are used to validate findings inferred from single cell analysis. This is, however, an exciting era for molecular biology where future studies should take full advantage of current and emerging techniques. Topographical/spatial transcriptomics such as Slide-seq [121] permits scRNA-seq of cells from known positions in *ex vivo* tissue sections. Techniques such as these allow for verification of cellular identity determined by transcriptomic profile by examining the cell's position within a tissue, as well as interaction partners such as surrounding cells or extracellular matrix. Techniques that combine two single cell technologies such as CITE-seq allows for verification of cell subpopulation identity by combining cell surface marker analysis

with scRNA-seq generated transcriptomic profiles. Similarly, SHARE-seq allows for the prediction of future cell state by integrating scRNA-seq and scATAC-seq in the same cells. This technique provides information on cell identity and function through transcriptomic profiling and cell states through chromatin accessibility. Together, techniques such as these will allow us to answer some of the most fundamental and interesting questions of mammary gland biology research, such as how the mammary gland is arranged, which definitive mammary subpopulations exist and the differentiation potential of different cell types.

Acknowledgment

Our thanks are extended to our reviewers whose insightful comments allowed us to develop a more comprehensive review. Apologies are extended for any studies we did not include due to space limitations. AJT is funded by a BBSRC project grant (BB/S006745/1) to WTK. WTK is funded by CRUK grant (C56829/A22053).

References

[1] N. Azamjah, Y. Soltan-Zadeh, F. Zayeri, Global Trend of Breast Cancer Mortality Rate: A 25-Year Study, Asian Pac J Cancer Prev 20(7) (2019) 2015-2020.

[2] C.J. Watson, W.T. Khaled, Mammary development in the embryo and adult: a journey of morphogenesis and commitment, Development 135(6) (2008) 995-1003.

[3] A. Prat, C.M. Perou, Deconstructing the molecular portraits of breast cancer, Mol Oncol 5(1) (2011) 5-23.

[4] C.M. Perou, T. Sorlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, H. Johnsen, L.A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S.X. Zhu, P.E. Lonning, A.L. Borresen-Dale, P.O. Brown, D. Botstein, Molecular portraits of human breast tumours, Nature 406(6797) (2000) 747-52.

[5] A. Prat, C.M. Perou, Mammary development meets cancer genomics, Nat Med 15(8) (2009) 842-4.

[6] T. Sorlie, C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, T. Thorsen, H. Quist, J.C. Matese, P.O. Brown, D. Botstein, P.E. Lonning, A.L. Borresen-Dale, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications, Proc Natl Acad Sci U S A 98(19) (2001) 10869-74.

[7] E. Lim, F. Vaillant, D. Wu, N.C. Forrest, B. Pal, A.H. Hart, M.L. Asselin-Labat, D.E. Gyorki, T. Ward, A. Partanen, F. Feleppa, L.I. Huschtscha, H.J. Thorne, kConFab, S.B. Fox, M. Yan, J.D. French, M.A. Brown, G.K. Smyth, J.E. Visvader, G.J. Lindeman, Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers, Nat Med 15(8) (2009) 907-13.

[8] J.E. Visvader, Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis, Genes Dev 23(22) (2009) 2563-77.

[9] M. Shackleton, F. Vaillant, K.J. Simpson, J. Stingl, G.K. Smyth, M.L. Asselin-Labat, L. Wu, G.J. Lindeman, J.E. Visvader, Generation of a functional mammary gland from a single stem cell, Nature 439(7072) (2006) 84-8.

[10] J. Stingl, P. Eirew, I. Ricketson, M. Shackleton, F. Vaillant, D. Choi, H.I. Li, C.J. Eaves, Purification and unique properties of mammary epithelial stem cells, Nature 439(7079) (2006) 993-7.

[11] K.E. Sleeman, H. Kendrick, A. Ashworth, C.M. Isacke, M.J. Smalley, CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells, Breast Cancer Res 8(1) (2006) R7.

[12] J.E. Visvader, J. Stingl, Mammary stem cells and the differentiation hierarchy: current status and perspectives, Genes Dev 28(11) (2014) 1143-58.

[13] S.S. Wu, J.H. Lee, B.K. Koo, Lineage Tracing: Computational Reconstruction Goes Beyond the Limit of Imaging, Mol Cells 42(2) (2019) 104-112.

[14] D. Wang, C. Cai, X. Dong, Q.C. Yu, X.O. Zhang, L. Yang, Y.A. Zeng, Identification of multipotent mammary stem cells by protein C receptor expression, Nature 517(7532) (2015) 81-4.

[15] A.C. Rios, N.Y. Fu, G.J. Lindeman, J.E. Visvader, In situ identification of bipotent stem cells in the mammary gland, Nature 506(7488) (2014) 322-7.

[16] R. van Amerongen, A.N. Bowman, R. Nusse, Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland, Cell Stem Cell 11(3) (2012) 387-400.

[17] W. Song, R. Wang, W. Jiang, Q. Yin, G. Peng, R. Yang, Q.C. Yu, J. Chen, J. Li, T.H. Cheung, N. Jing, Y.A. Zeng, Hormones induce the formation of luminal-derived basal cells in the mammary gland, Cell Res 29(3) (2019) 206-220.

[18] F.M. Davis, B. Lloyd-Lewis, O.B. Harris, S. Kozar, D.J. Winton, L. Muresan, C.J. Watson, Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny, Nat Commun 7 (2016) 13053.
[19] A. Van Keymeulen, A.S. Rocha, M. Ousset, B. Beck, G. Bouvencourt, J. Rock, N. Sharma, S. Dekoninck, C. Blanpain, Distinct stem cells contribute to mammary gland development and maintenance, Nature 479(7372) (2011) 189-93.
[20] A. Wuidart, M. Ousset, S. Rulands, B.D. Simons, A. Van Keymeulen, C. Blanpain, Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells, Genes Dev 30(11) (2016) 1261-77.
[21] S. Elias, M.A. Morgan, E.K. Bikoff, E.J. Robertson, Long-lived unipotent Blimp1-positive luminal stem cells drive mammary gland organogenesis throughout adult life, Nat Commun 8(1) (2017) 1714.
[22] A.M. Lilja, V. Rodilla, M. Huyghe, E. Hannezo, C. Landragin, O. Renaud, O. Leroy, S. Rulands, B.D. Simons, S. Fre, Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland, Nat Cell Biol 20(6) (2018) 677-687.
[23] A. Wuidart, A. Sifrim, M. Fioramonti, S. Matsumura, A. Brisebarre, D. Brown, A. Centonze, A. Dannau, C. Dubois, A. Van Keymeulen, T. Voet, C. Blanpain, Early lineage segregation of multipotent embryonic mammary gland progenitors, Nat Cell Biol 20(6) (2018) 666-676.
[24] M.D. Prater, V. Petit, I. Alasdair Russell, R.R. Giraddi, M. Shehata, S. Menon, R. Schulte, I. Kalajzic, N. Rath, M.F. Olson, D. Metzger, M.M. Faraldo, M.A. Deugnier, M.A. Glukhova, J. Stingl, Mammary stem cells have myoepithelial cell properties, Nat Cell Biol 16(10) (2014) 942-50, 1-7.
[25] C.L. Scheele, E. Hannezo, M.J. Muraro, A. Zomer, N.S. Langedijk, A. van Oudenaarden, B.D. Simons, J. van Rheenen, Identity and dynamics of mammary stem cells during branching morphogenesis, Nature 542(7641) (2017) 313-317.
[26] A. Shlien, D. Malkin, Copy number variations and cancer, Genome Med 1(6) (2009) 62.
[27] N. Navin, J. Kendall, J. Troge, P. Andrews, L. Rodgers, J. McIndoo, K. Cook, A. Stepansky, D. Levy, D. Esposito, L. Muthuswamy, A. Krasnitz, W.R. McCombie, J. Hicks, M. Wigler, Tumour evolution inferred by single-cell sequencing, Nature 472(7341) (2011) 90-4.
[28] X. Xu, Y. Hou, X. Yin, L. Bao, A. Tang, L. Song, F. Li, S. Tsang, K. Wu, H. Wu, W. He, L. Zeng, M. Xing, R. Wu, H. Jiang, X. Liu, D. Cao, G. Guo, X. Hu, Y. Gui, Z. Li, W. Xie, X. Sun, M. Shi, Z. Cai, B. Wang, M. Zhong, J. Li, Z. Lu, N. Gu, X. Zhang, L. Goodman, L. Bolund, J. Wang, H. Yang, K. Kristiansen, M. Dean, Y. Li, J. Wang, Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor, Cell 148(5) (2012) 886-95.
[29] Y. Wang, J. Waters, M.L. Leung, A. Unruh, W. Roh, X. Shi, K. Chen, P. Scheet, S. Vattathil, H. Liang, A. Multani, H. Zhang, R. Zhao, F. Michor, F. Meric-Bernstam, N.E. Navin, Clonal evolution in breast cancer revealed by single nucleus genome sequencing, Nature 512(7513) (2014) 155-60.
[30] M. Leung, Y. Wang, J. Waters, N.E. Navin, SNES: single nucleus exome sequencing, Genome Biol 16 (2015) 55.
[31] N. Navin, The first five years of single-cell cancer genomics and beyond, Genome Res 25(10) (2015) 1499-507.
[32] C. Zong, S. Lu, A.R. Chapman, X.S. Xie, Genome-wide detection of single-nucleotide and copy-number variations of a single human cell, Science 338(6114) (2012) 1622-6.

[33] H. Zahn, A. Steif, E. Laks, P. Eirew, M. VanInsberghe, S.P. Shah, S. Aparicio, C.L. Hansen, Scalable whole-genome single-cell library preparation without preamplification, Nat Methods 14(2) (2017) 167-173.
[34] E. Laks, A. McPherson, H. Zahn, D. Lai, A. Steif, J. Brimhall, J. Biele, B. Wang, T. Masud, J. Ting, D. Grewal, C. Nielsen, S. Leung, V. Bojilova, M. Smith, O. Golovko, S. Poon, P. Eirew, F. Kabeer, T. Ruiz de Algara, S.R. Lee, M.J. Taghiyar, C. Huebner, J. Ngo, T. Chan, S. Vatrt-Watts, P. Walters, N. Abrar, S. Chan, M. Wiens, L. Martin, R.W. Scott, T.M. Underhill, E. Chavez, C. Steidl, D. Da Costa, Y. Ma, R.J.N. Coope, R. Corbett, S. Pleasance, R. Moore, A.J. Mungall, C. Mar, F. Cafferty, K. Gelmon, S. Chia, C.I.G.C. Team, M.A. Marra, C. Hansen, S.P. Shah, S. Aparicio, Clonal Decomposition and DNA Replication States Defined by Scaled Single-Cell Genome Sequencing, Cell 179(5) (2019) 1207-1221 e22.
[35] R. Gao, A. Davis, T.O. McDonald, E. Sei, X. Shi, Y. Wang, P.C. Tsai, A. Casasent, J. Waters, H. Zhang, F. Meric-Bernstam, F. Michor, N.E. Navin, Punctuated copy number evolution and clonal stasis in triple-negative breast cancer, Nat Genet 48(10) (2016) 1119-30.
[36] A.K. Casasent, A. Schalck, R. Gao, E. Sei, A. Long, W. Pangburn, T. Casasent, F. Meric-Bernstam, M.E. Edgerton, N.E. Navin, Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing, Cell 172(1-2) (2018) 205-217 e12.
[37] O. Ornatsky, D. Bandura, V. Baranov, M. Nitz, M.A. Winnik, S. Tanner, Highly multiparametric analysis by mass cytometry, J Immunol Methods 361(1-2) (2010) 1-20.
[38] D.R. Bandura, V.I. Baranov, O.I. Ornatsky, A. Antonov, R. Kinach, X. Lou, S. Pavlov, S. Vorobiev, J.E. Dick, S.D. Tanner, Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry, Anal Chem 81(16) (2009) 6813-22.
[39] S.C. Bendall, E.F. Simonds, P. Qiu, A.D. Amir el, P.O. Krutzik, R. Finck, R.V. Bruggner, R. Melamed, A. Trejo, O.I. Ornatsky, R.S. Balderas, S.K. Plevritis, K. Sachs, D. Pe'er, S.D. Tanner, G.P. Nolan, Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum, Science 332(6030) (2011) 687-96.
[40] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B.B. Tuch, A. Siddiqui, K. Lao, M.A. Surani, mRNA-Seq whole-transcriptome analysis of a single cell, Nat Methods 6(5) (2009) 377-82.
[41] P. See, J. Lum, J. Chen, F. Ginhoux, A Single-Cell Sequencing Guide for Immunologists, Front Immunol 9 (2018) 2425.
[42] L. Valihrach, P. Androvic, M. Kubista, Platforms for Single-Cell Collection and Analysis, Int J Mol Sci 19(3) (2018).
[43] B. Hwang, J.H. Lee, D. Bang, Single-cell RNA sequencing technologies and bioinformatics pipelines, Exp Mol Med 50(8) (2018) 96.
[44] G. Chen, B. Ning, T. Shi, Single-Cell RNA-Seq Technologies and Related Computational Data Analysis, Front Genet 10 (2019) 317.
[45] D. Ramskold, S. Luo, Y.C. Wang, R. Li, Q. Deng, O.R. Faridani, G.A. Daniels, I. Khrebtukova, J.F. Loring, L.C. Laurent, G.P. Schroth, R. Sandberg, Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells, Nat Biotechnol 30(8) (2012) 777-82.
[46] S. Picelli, A.K. Bjorklund, O.R. Faridani, S. Sagasser, G. Winberg, R. Sandberg, Smart-seq2 for sensitive full-length transcriptome profiling in single cells, Nat Methods 10(11) (2013) 1096-8.
[47] S. Picelli, O.R. Faridani, A.K. Bjorklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2, Nat Protoc 9(1) (2014) 171-81.

[48] A.M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D.A. Weitz, M.W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells, Cell 161(5) (2015) 1187-201.
[49] E.Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martersteck, J.J. Trombetta, D.A. Weitz, J.R. Sanes, A.K. Shalek, A. Regev, S.A. McCarroll, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets, Cell 161(5) (2015) 1202-1214.
[50] G.X. Zheng, J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, S.B. Ziraldo, T.D. Wheeler, G.P. McDermott, J. Zhu, M.T. Gregory, J. Shuga, L. Montesclaros, J.G. Underwood, D.A. Masquelier, S.Y. Nishimura, M. Schnall-Levin, P.W. Wyatt, C.M. Hindson, R. Bharadwaj, A. Wong, K.D. Ness, L.W. Beppu, H.J. Deeg, C. McFarland, K.R. Loeb, W.J. Valente, N.G. Ericson, E.A. Stevens, J.P. Radich, T.S. Mikkelsen, B.J. Hindson, J.H. Bielas, Massively parallel digital transcriptional profiling of single cells, Nat Commun 8 (2017) 14049.
[51] M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P.K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells, Nat Methods 14(9) (2017) 865-868.
[52] H. Chen, C. Lareau, T. Andreani, M.E. Vinyard, S.P. Garcia, K. Clement, M.A. Andrade-Navarro, J.D. Buenrostro, L. Pinello, Assessment of computational methods for the analysis of single-cell ATAC-seq data, Genome Biol 20(1) (2019) 241.
[53] J.D. Buenrostro, B. Wu, U.M. Litzenburger, D. Ruff, M.L. Gonzales, M.P. Snyder, H.Y. Chang, W.J. Greenleaf, Single-cell chromatin accessibility reveals principles of regulatory variation, Nature 523(7561) (2015) 486-90.
[54] D.A. Cusanovich, R. Daza, A. Adey, H.A. Pliner, L. Christiansen, K.L. Gunderson, F.J. Steemers, C. Trapnell, J. Shendure, Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing, Science 348(6237) (2015) 910-4.
[55] A.T. Satpathy, J.M. Granja, K.E. Yost, Y. Qi, F. Meschi, G.P. McDermott, B.N. Olsen, M.R. Mumbach, S.E. Pierce, M.R. Corces, P. Shah, J.C. Bell, D. Jhutty, C.M. Nemec, J. Wang, L. Wang, Y. Yin, P.G. Giresi, A.L.S. Chang, G.X.Y. Zheng, W.J. Greenleaf, H.Y. Chang, Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion, Nat Biotechnol 37(8) (2019) 925-936.
[56] J. Lee, D.Y. Hyeon, D. Hwang, Single-cell multiomics: technologies and data analysis methods, Exp Mol Med 52(9) (2020) 1428-1442.
[57] X.F. Mallory, M. Edrisi, N. Navin, L. Nakhleh, Methods for copy number aberration detection from single-cell DNA-sequencing data, Genome Biol 21(1) (2020) 208.
[58] A.K. Kimball, L.M. Oko, B.L. Bullock, R.A. Nemenoff, L.F. van Dyk, E.T. Clambey, A Beginner's Guide to Analyzing and Visualizing Mass Cytometry Data, J Immunol 200(1) (2018) 3-22.
[59] X. Liu, W. Song, B.Y. Wong, T. Zhang, S. Yu, G.N. Lin, X. Ding, A comparison framework and guideline of clustering methods for mass cytometry data, Genome Biol 20(1) (2019) 297.
[60] W. Huber, V.J. Carey, R. Gentleman, S. Anders, M. Carlson, B.S. Carvalho, H.C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K.D. Hansen, R.A. Irizarry, M. Lawrence, M.I. Love, J. MacDonald, V. Obenchain, A.K. Oles, H. Pages, A. Reyes, P. Shannon, G.K. Smyth, D. Tenenbaum, L. Waldron, M. Morgan, Orchestrating high-throughput genomic analysis with Bioconductor, Nat Methods 12(2) (2015) 115-21.
[61] M.D. Luecken, F.J. Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol Syst Biol 15(6) (2019) e8746.
[62] A.T. Lun, D.J. McCarthy, J.C. Marioni, A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor, F1000Res 5 (2016) 2122.

[63] L.V.D. Maaten, G. Hinton, Visualizing data using t-SNE, Journal of machine learning research 9 (2008) 2579-2605.
[64] L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction, arxiv (2018).
[65] L. Haghverdi, F. Buettner, F.J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data, Bioinformatics 31(18) (2015) 2989-98.
[66] L. Haghverdi, M. Büttner, A.f. Wolf, F. Buettner, F.J. Theis, Diffusion pseudotime robustly reconstructs lineage branching, Nature Methods 13(10) (2016) 845-848.
[67] M. Lotfollahi, F.A. Wolf, F.J. Theis, scGen predicts single-cell perturbation responses, Nat Methods 16(8) (2019) 715-721.
[68] D. van Dijk, R. Sharma, J. Nainys, K. Yim, P. Kathail, A.J. Carr, C. Burdziak, K.R. Moon, C.L. Chaffer, D. Pattabiraman, B. Bierie, L. Mazutis, G. Wolf, S. Krishnaswamy, D. Pe'er, Recovering Gene Interactions from Single-Cell Data Using Data Diffusion, Cell 174(3) (2018) 716-729 e27.
[69] A.N. Schep, B. Wu, J.D. Buenrostro, W.J. Greenleaf, chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data, Nat Methods 14(10) (2017) 975-978.
[70] C. Wang, D. Sun, X. Huang, C. Wan, Z. Li, Y. Han, Q. Qin, J. Fan, X. Qiu, Y. Xie, C.A. Meyer, M. Brown, M. Tang, H. Long, T. Liu, X.S. Liu, Integrative analyses of single-cell transcriptome and regulome using MAESTRO, Genome Biol 21(1) (2020) 198.
[71] K. Bach, S. Pensa, M. Grzelak, J. Hadfield, D.J. Adams, J.C. Marioni, W.T. Khaled, Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing, Nat Commun 8(1) (2017) 2128.
[72] B. Pal, Y. Chen, F. Vaillant, P. Jamieson, L. Gordon, A.C. Rios, S. Wilcox, N. Fu, K.H. Liu, F.C. Jackling, M.J. Davis, G.J. Lindeman, G.K. Smyth, J.E. Visvader, Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling, Nat Commun 8(1) (2017) 1627.
[73] R.R. Giraddi, C.Y. Chung, R.E. Heinz, O. Balcioglu, M. Novotny, C.L. Trejo, C. Dravis, B.M. Hagos, E.M. Mehrabad, L.W. Rodewald, J.Y. Hwang, C. Fan, R. Lasken, K.E. Varley, C.M. Perou, G.M. Wahl, B.T. Spike, Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development, Cell Rep 24(6) (2018) 1653-1666 e7.
[74] C. Dravis, C.Y. Chung, N.K. Lytle, J. Herrera-Valdez, G. Luna, C.L. Trejo, T. Reya, G.M. Wahl, Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity, Cancer Cell 34(3) (2018) 466-482 e6.
[75] C.Y. Chung, Z. Ma, C. Dravis, S. Preissl, O. Poirion, G. Luna, X. Hou, R.R. Giraddi, B. Ren, G.M. Wahl, Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships, Cell Rep 29(2) (2019) 495-510 e6.
[76] A.-J. Twigger, L.K. Engelbrecht, K. Bach, I. Schultz-Pernice, S. Petricca, C.H. Scheel, W. Khaled, Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk, bioRxiv (2020).
[77] 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 (2020).

[78] K. Bach, S. Pensa, M. Zarocsinceva, K. Kania, J. Stockis, S. Pinaud, K.A. Lazarus, M. Shehata, B.M. Simoes, A.R. Greenhalgh, S.J. Howell, R.B. Clarke, C. Caldas, T.Y.F. Halim, J.C. Marioni, W.T. Khaled, Time-resolved single-cell analysis of Brca1 associated mammary tumorigenesis reveals aberrant differentiation of luminal progenitors, Nature Comms (2021).
[79] C. Wang, J.R. Christin, M.H. Oktay, W. Guo, Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages, Cell Rep 18(12) (2017) 2825-2835.
[80] A. Van Keymeulen, M. Fioramonti, A. Centonze, G. Bouvencourt, Y. Achouri, C. Blanpain, Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage, Cell Rep 20(7) (2017) 1525-1532.
[81] V. Rodilla, A. Dasti, M. Huyghe, D. Lafkas, C. Laurent, F. Reyal, S. Fre, Luminal progenitors restrict their lineage potential during mammary gland development, PLoS Biol 13(2) (2015) e1002069.
[82] T.H. Chang, K. Kunasegaran, G.A. Tarulli, D. De Silva, P.M. Voorhoeve, A.M. Pietersen, New insights into lineage restriction of mammary gland epithelium using parity-identified mammary epithelial cells, Breast Cancer Res 16(1) (2014) R1.
[83] N. Pervolarakis, Q.H. Nguyen, J. Williams, Y. Gong, G. Gutierrez, P. Sun, D. Jhutty, G.X.Y. Zheng, C.M. Nemec, X. Dai, K. Watanabe, K. Kessenbrock, Integrated Single-Cell Transcriptomics and Chromatin Accessibility Analysis Reveals Regulators of Mammary Epithelial Cell Identity, Cell Rep 33(3) (2020) 108273.
[84] Q.H. Nguyen, N. Pervolarakis, K. Blake, D. Ma, R.T. Davis, N. James, A.T. Phung, E. Willey, R. Kumar, E. Jabart, I. Driver, J. Rock, A. Goga, S.A. Khan, D.A. Lawson, Z. Werb, K. Kessenbrock, Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity, Nat Commun 9(1) (2018) 2028.
[85] D. Knapp, N. Kannan, D. Pellacani, C.J. Eaves, Mass Cytometric Analysis Reveals Viable Activated Caspase-3(+) Luminal Progenitors in the Normal Adult Human Mammary Gland, Cell Rep 21(4) (2017) 1116-1126.
[86] L.K. Engelbrecht, A.-J. Twigger, H.M. Ganz, C.J. Gabka, A.R. Bausch, H. Lickert, M. Sterr, I. Kunze, W.T. Khaled, C.H. Scheel, A strategy to address dissociation-induced compositional and transcriptional bias for single-cell analysis of the human mammary gland, bioRxiv (2021).
[87] H. Sun, Z. Miao, X. Zhang, U.I. Chan, S.M. Su, S. Guo, C.K.H. Wong, X. Xu, C.X. Deng, Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia, J Biol Chem 293(22) (2018) 8315-8329.
[88] W. Chen, S.J. Morabito, K. Kessenbrock, T. Enver, K.B. Meyer, A.E. Teschendorff, Single-cell landscape in mammary epithelium reveals bipotent-like cells associated with breast cancer risk and outcome, Commun Biol 2 (2019) 306.
[89] T. Thong, Y. Wang, M.D. Brooks, C.T. Lee, C. Scott, L. Balzano, M.S. Wicha, J.A. Colacino, Hybrid Stem Cell States: Insights Into the Relationship Between Mammary Development and Breast Cancer Using Single-Cell Transcriptomics, Front Cell Dev Biol 8 (2020) 288.
[90] L.J. Anstine, R. Keri, A new view of the mammary epithelial hierarchy and its implications for breast cancer initiation and metastasis, J Cancer Metastasis Treat 5 (2019).
[91] S. Cristea, K. Polyak, Dissecting the mammary gland one cell at a time, Nat Commun 9(1) (2018) 2473.
[92] J.L. Regan, M.J. Smalley, Integrating single-cell RNA-sequencing and functional assays to decipher mammary cell states and lineage hierarchies, npj Breast Cancer 6(1) (2020).
[93] C.M. Li, H. Shapiro, C. Tsiobikas, L.M. Selfors, H. Chen, J. Rosenbluth, K. Moore, K.P. Gupta, G.K. Gray, Y. Oren, M.J. Steinbaugh, J.L. Guerriero, L. Pinello, A. Regev, J.S.

Brugge, Aging-Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA Sequencing, Cell Rep 33(13) (2020) 108566.
[94] X. Liu, V. Ory, S. Chapman, H. Yuan, C. Albanese, B. Kallakury, O.A. Timofeeva, C. Nealon, A. Dakic, V. Simic, B.R. Haddad, J.S. Rhim, A. Dritschilo, A. Riegel, A. McBride, R. Schlegel, ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells, Am J Pathol 180(2) (2012) 599-607.
[95] G. Dontu, T.A. Ince, Of mice and women: a comparative tissue biology perspective of breast stem cells and differentiation, J Mammary Gland Biol Neoplasia 20(1-2) (2015) 51-62.
[96] C.J. Watson, W.T. Khaled, Mammary development in the embryo and adult: new insights into the journey of morphogenesis and commitment, Development 147(22) (2020).
[97] B. Spanjaard, B. Hu, N. Mitic, P. Olivares-Chauvet, S. Janjuha, N. Ninov, J.P. Junker, Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars, Nat Biotechnol 36(5) (2018) 469-473.
[98] B. Raj, J.A. Gagnon, A.F. Schier, Large-scale reconstruction of cell lineages using single-cell readout of transcriptomes and CRISPR-Cas9 barcodes by scGESTALT, Nat Protoc 13(11) (2018) 2685-2713.
[99] C.H. O'Flanagan, K.R. Campbell, A.W. Zhang, F. Kabeer, J.L.P. Lim, J. Biele, P. Eirew, D. Lai, A. McPherson, E. Kong, C. Bates, K. Borkowski, M. Wiens, B. Hewitson, J. Hopkins, J. Pham, N. Ceglia, R. Moore, A.J. Mungall, J.N. McAlpine, C.I.G.C. Team, S.P. Shah, S. Aparicio, Dissociation of solid tumor tissues with cold active protease for single-cell RNA-seq minimizes conserved collagenase-associated stress responses, Genome Biol 20(1) (2019) 210.
[100] N. Kanaya, G. Chang, X. Wu, K. Saeki, L. Bernal, H.J. Shim, J. Wang, C. Warden, T. Yamamoto, J. Li, J.S. Park, T. Synold, S. Vonderfecht, M. Rakoff, S.L. Neuhausen, S. Chen, Single-cell RNA-sequencing analysis of estrogen- and endocrine-disrupting chemical-induced reorganization of mouse mammary gland, Commun Biol 2 (2019) 406.
[101] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S.H. Orkin, G.C. Yuan, M. Chen, G. Guo, Mapping the Mouse Cell Atlas by Microwell-Seq, Cell 172(5) (2018) 1091-1107 e17.
[102] C. Tabula Muris, A single-cell transcriptomic atlas characterizes ageing tissues in the mouse, Nature 583(7817) (2020) 590-595.
[103] L.M. Murrow, R.J. Weber, J. Caruso, C.S. McGinnis, K. Phong, P. Gascard, A.D. Borowsky, T.A. Desai, M. Thomson, T. Tlsty, Z.J. Gartner, Changes in epithelial proportions and transcriptional state underlie major premenopausal breast cancer risks, bioRxiv (2020).
[104] F.A. Pelissier Vatter, D. Schapiro, H. Chang, A.D. Borowsky, J.K. Lee, B. Parvin, M.R. Stampfer, M.A. LaBarge, B. Bodenmiller, J.B. Lorens, High-Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia, Cell Rep 23(4) (2018) 1205-1219.
[105] P.K. Lo, Y. Yao, Q. Zhou, Single-Cell RNA-seq Reveals Obesity-Induced Alterations in the Brca1-Mutated Mammary Gland Microenvironment, Cancers (Basel) 12(8) (2020).
[106] W. Chung, H.H. Eum, H.O. Lee, K.M. Lee, H.B. Lee, K.T. Kim, H.S. Ryu, S. Kim, J.E. Lee, Y.H. Park, Z. Kan, W. Han, W.Y. Park, Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer, Nat Commun 8 (2017) 15081.
[107] Y. Wang, T.S. Chaffee, R.S. LaRue, D.N. Huggins, P.M. Witschen, A.M. Ibrahim, A.C. Nelson, H.L. Machado, K.L. Schwertfeger, Tissue-resident macrophages promote extracellular matrix homeostasis in the mammary gland stroma of nulliparous mice, Elife 9 (2020).

[108] E. Azizi, A.J. Carr, G. Plitas, A.E. Cornish, C. Konopacki, S. Prabhakaran, J. Nainys, K. Wu, V. Kiseliovas, M. Setty, K. Choi, R.M. Fromme, P. Dao, P.T. McKenney, R.C. Wasti, K. Kadaveru, L. Mazutis, A.Y. Rudensky, D. Pe'er, Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment, Cell 174(5) (2018) 1293-1308 e36.
[109] M. Bartoschek, N. Oskolkov, M. Bocci, J. Lovrot, C. Larsson, M. Sommarin, C.D. Madsen, D. Lindgren, G. Pekar, G. Karlsson, M. Ringner, J. Bergh, A. Bjorklund, K. Pietras, Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing, Nat Commun 9(1) (2018) 5150.
[110] A. Sebastian, N.R. Hum, K.A. Martin, S.F. Gilmore, I. Peran, S.W. Byers, E.K. Wheeler, M.A. Coleman, G.G. Loots, Single-Cell Transcriptomic Analysis of Tumor-Derived Fibroblasts and Normal Tissue-Resident Fibroblasts Reveals Fibroblast Heterogeneity in Breast Cancer, Cancers (Basel) 12(5) (2020).
[111] S.Z. Wu, D.L. Roden, C. Wang, H. Holliday, K. Harvey, A.S. Cazet, K.J. Murphy, B. Pereira, G. Al-Eryani, N. Bartonicek, R. Hou, J.R. Torpy, S. Junankar, C.L. Chan, C.E. Lam, M.N. Hui, L. Gluch, J. Beith, A. Parker, E. Robbins, D. Segara, C. Mak, C. Cooper, S. Warrier, A. Forrest, J. Powell, S. O'Toole, T.R. Cox, P. Timpson, E. Lim, X.S. Liu, A. Swarbrick, Stromal cell diversity associated with immune evasion in human triple-negative breast cancer, EMBO J (2020) e104063.
[112] J. Wagner, M.A. Rapsomaniki, S. Chevrier, T. Anzeneder, C. Langwieder, A. Dykgers, M. Rees, A. Ramaswamy, S. Muenst, S.D. Soysal, A. Jacobs, J. Windhager, K. Silina, M. van den Broek, K.J. Dedes, M. Rodriguez Martinez, W.P. Weber, B. Bodenmiller, A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer, Cell 177(5) (2019) 1330-1345 e18.
[113] M. Tognetti, A. Gabor, M. Yang, V. Cappelletti, J. Windhager, K. Charmpi, N. de Souza, A. Beyer, P. Picotti, J. Saez-Rodriguez, B. Bodenmiller, Deciphering the signaling network landscape of breast cancer improves drug sensitvity prediction, bioRxiv (2020).
[114] R. Gao, C. Kim, E. Sei, T. Foukakis, N. Crosetto, L.K. Chan, M. Srinivasan, H. Zhang, F. Meric-Bernstam, N. Navin, Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer, Nat Commun 8(1) (2017) 228.
[115] M. Karaayvaz, S. Cristea, S.M. Gillespie, A.P. Patel, R. Mylvaganam, C.C. Luo, M.C. Specht, B.E. Bernstein, F. Michor, L.W. Ellisen, Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq, Nat Commun 9(1) (2018) 3588.
[116] F. Salmén, S. Vickovic, L. Larsson, L. Stenbeck, J. Vallon-Christersson, A. Ehinger, J. Häkkinen, Å. Borg, J. Frisén, P.L. Ståhl, J. Lundeberg, Multidimensional transcriptomics provides detailed information about immune cell distribution and identity in HER2+ breast tumors, bioRxiv (2018).
[117] A. Andersson, L. Larsson, L. Stenbeck, F. Salmén, A. Ehinger, S. Wu, G. Al-Eryani, D. Roden, A. Swarbrick, Å. Borg, J. Frisén, C. Engblom, J. Lundeberg, Spatial deconvolution of HER2-positive breast tumors reveals novel intercellular relationships, bioRxiv (2020).
[118] T. Baslan, J. Kendall, K. Volyanskyy, K. McNamara, H. Cox, S. D'Italia, F. Ambrosio, M. Riggs, L. Rodgers, A. Leotta, J. Song, Y. Mao, J. Wu, R. Shah, R. Gularte-Merida, K. Chadalavada, G. Nanjangud, V. Varadan, A. Gordon, C. Curtis, A. Krasnitz, N. Dimitrova, L. Harris, M. Wigler, J. Hicks, Novel insights into breast cancer copy number genetic heterogeneity revealed by single-cell genome sequencing, Elife 9 (2020).
[119] K. Grosselin, A. Durand, J. Marsolier, A. Poitou, E. Marangoni, F. Nemati, A. Dahmani, S. Lameiras, F. Reyal, O. Frenoy, Y. Pousse, M. Reichen, A. Woolfe, C. Brenan, A.D. Griffiths, C. Vallot, A. Gerard, High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer, Nat Genet 51(6) (2019) 1060-1066.

[120] S. Gkountela, F. Castro-Giner, B.M. Szczerba, M. Vetter, J. Landin, R. Scherrer, I. Krol, M.C. Scheidmann, C. Beisel, C.U. Stirnimann, C. Kurzeder, V. Heinzelmann-Schwarz, C. Rochlitz, W.P. Weber, N. Aceto, Circulating Tumor Cell Clustering Shapes DNA Methylation to Enable Metastasis Seeding, Cell 176(1-2) (2019) 98-112 e14.
[121] S.G. Rodriques, R.R. Stickels, A. Goeva, C.A. Martin, E. Murray, C.R. Vanderburg, J. Welch, L.M. Chen, F. Chen, E.Z. Macosko, Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution, Science 363(6434) (2019) 1463-1467.

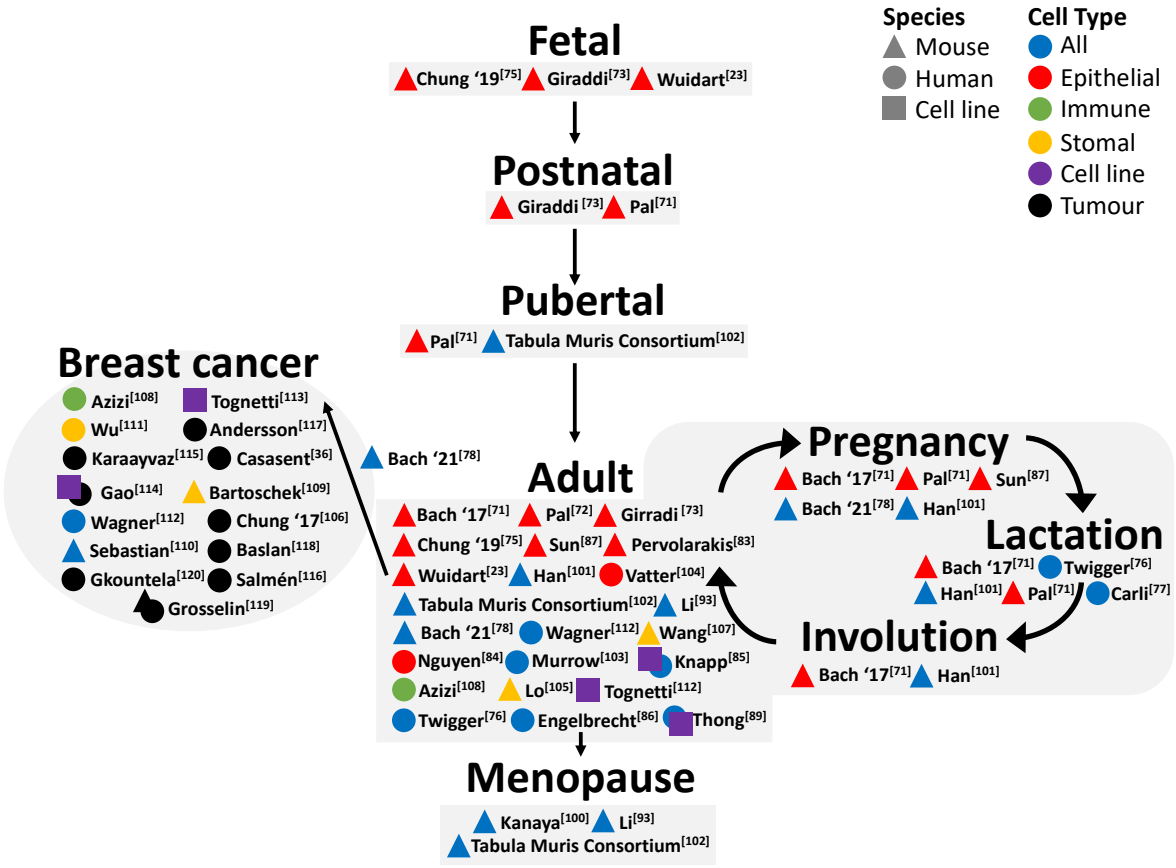

Figure 1: Overview of mouse and human mammary development and cancer single cell studies.

Table 1: Single cell datasets of normal mammary and tumour cells

Reference	**Lab**	**Species**	**Tissue type (normal/ cancer/ both)**	**Sample time points**	**Mammary fraction (epithelial/ Stroma-immune/ both)**	**Preparation of cells prior to single cell sequencing**	**Technique + technology**	**Number of cells/nuclei profiled**
Andersson et al. *bioRxiv* 2020 (preprint)	Lundeberg	Human	Cancer	8x HER2+ breast tumors	Both	Cells spatially selected from fixed tissue sections	***Spatial transcriptomics*** developed by Ståhl et al. *Science* 2016	1007 spots (where spot refers to a small neighborhood populated by multiple cells)
Azizi et al. *Cell* 2018	Pe'er and Rudensky	Human	Both	8x primary breast carcinomas tumour and matched normal breast tissue, peripheral blood and lymph nodes	Stroma-imune	$CD45^+$ FACS sorted cells	***scRNA-seq*** using the inDrop platform	47,016 cells
Bach et al. *Nature Comm* 2017	Khaled	Mouse	Normal	Mammary cells taken during pregnancy (day 14.5), lactation (day 6), involution (day 11) and from virgin 8 week old C57BL/6N mice	Epithelial	Live lineage negative $EpCAM^+$ cells sorted using FACS	***scRNA-seq*** using Chromium 10x Drop-Seq platform	23,184 cells
Bach and Pensa et al. *Nature Comm* 2021	Khaled and Marioni	Mouse	Both	Tumourogenesis data set: 13 x Blg-Cre; Brca1f/f;p53+/- mice (aged 30-48 weeks, nulliparous) and 2x C57BL/6N mice (aged 36-40 weeks old) Pregnancy data set: 9x C57BL/6N mice at 4.5/9.5/14.5 days gestation as well as 3x C57BL/6N mice (aged 12 weeks old)	Both	Viable cells were isolated using MACS Dead Cell Removal Kit	***scRNA-seq*** using Chromium 10x Drop-Seq platform	102,829 cells
Bartoschek et al. *Nature Comm* 2018	Pietras	Mouse	Cancer	Tumours from 14 week old MMTV-PyMT mice	Stroma-imune	$EpCAM^-/CD45^-/CD31^-/NG2^-$ FACS sorted mesenchymal cells	***scRNA-seq*** using Smart-Seq2	768 cells
Baslan et al. *eLife* 2020	Hicks	Human	Cancer	Fresh pre-treatment core biopsies were taken from 16 patients' enrolled in phase II clinical trials	Both	Nuclei was isolated and FACS sorted for ploidy	***Single-nuclei copy number analysis*** developed by	Mean of 116 single-nuclei per tumour

				conducted by the Brown University Oncology Group (BrUOG)			Baslan et al. *Nature Protocols* 2012	
Carli et al. *J Mammary Gland Biol Neoplasia* 2020	Rudolph	Human	Normal	Human milk cells from two participants	Both	Viable cells were isolated using FACS	***scRNA-seq*** using Chromium 10x Drop-Seq platform	3740 cells
Casasent et al. *Cell* 2018	Edgerton and Navin	Human	Cancer	10x ductal carcinoma *in situ* (DCIS) and invasive ductal carcinoma (IDC) tumour samples	Both	Cells selected using single cell laser dissection	Topographic single cell sequencing (***TSCS***, scDNA sequencing)	On average 129 cells per patient
Chung et al. *Nature Comm* 2017	Park	Human	Cancer	11x tumor cells from different breast cancer subtypes	Both	Dead cells were removed using Ficoll-Paque PLUS	***scATAC-seq*** using the Fluidigm C1 platform	515 cells
Chung et al. *Cell Reports* 2019	Wahl	Mouse	Normal	Mammary cells from E18 fetal and 8-week old adult CD1 mice	Epithelial	$EpCAM^{+}/Lin^{-}$ FACS sorted cells	***snATAC-seq***	7,846 high quality single nuclei
Engelbrecht and Twigger et al. *bioRxiv* 2020 (preprint)	Scheel and Khaled	Human	Normal	Mammary tissue from one participant was digested 3 ways (3hr vs. 16 hr. digest and 10rpm vs. 100 rpm. shaking speed).	Both	Viable cells were isolated using MACS Dead Cell Removal Kit	***scRNA-seq*** using Chromium 10x Drop-Seq	11,191 cells
Gao et al. *Nature Comm* 2017	Navin	Human	Cancer	Breast cancer cell lines for technique validation and a single triple negative breast tumour sample	Both	Nuclei isolated from frozen tumour and cell lines	Nanogrid ***snRNA-seq***	796 primary cell nuclei
Giraddi et al. *Cell Rep* 2019	Wahl and Spike	Mouse	Normal	Mammary cells isolated from day 16/18 embryonic (E16, E18), postnatal day 10 (P10) and 10-16-week-old adult C57BL/6 mice	Epithelial	$EpCAM^{+}$ FACS sorted cells	***scRNA-seq*** using Chromium 10x Drop-Seq and Fluidigm C1 platforms	6,060 cells using 10x and 262 cells using Fluidigm C1
Gkountela et al. *Cell* 2019	Aceto	Human	Cancer	Blood samples containing circulating tumour cells (CTCs) from 43 patients with progressive breast cancer	Both	Live CTSs were stained for EpCAM, HER2 and EGFR and sorted for CTCs	***Single Cell Whole-genome Bisulfite Sequencing*** developed by Farlik et al. *Cell Rep.* 2015	89 single CTCs and 71 CTC clusters from patients and xenographs
Grosselin et al. *Nature Genetics* 2019	Griffiths, Vallot and Gérard	Mouse and Human	Cancer	Patient derived xenogrft mouse models of luminal and treatment resistant breast cancers	Both	Viable cells were isolated using MACS Dead Cell Removal Kit	***scChIP-seq and scRNA-seq***	2,728 cells profiled by scRNA-seq
Han et al. *Cell* 2018	Guo	Mouse	Normal	Mammary cells taken during pregnancy, lactation, involution and from virgin C57BL/6 mice	Both*	Cells from all major mice organs including mammary gland. *However, no	***scRNA-seq*** using microwell-seq	61,196 mammary gland cells

						mention was made of removal of the mammary lymph nodes prior to mammary gland dissociation.		
Kanaya et al. *Commun Biol* 2019	Chen	Mouse	Normal	Mammary cells from 9-week-old BALB/cj mice which underwent ovariectomy (surgical menopause) and were treated with vehicle, E2 and/or PBDE	Both	Dead cells were removed using microbeads	***scRNA-seq*** using Chromium 10x Drop-Seq platform	14,856 cells
Karaayvaz et al. *Nature Comm* 2018	Ellisen	Human	Cancer	6x primary triple negative breast tumours	Both	Viable tumour cells were isolated by FACS sorting	***scRNA-seq*** using Smart-seq2	1,189 cells
Knapp et al. *Cell Reports* 2017	Eaves	Human	Both	7x breast cancer cell lines and 8x normal primary breast tissue samples	Both	N/A	***Mass cytometry*** using 35 markers (16 extracellular and 19 intracellular)	Not disclosed
Li et al. *Cell Reports* 2020	Brugge	Mouse	Normal	Mammary cells from young (3-4 month, n=3) and aged (13-14 month, n=4) old virgin C57BL/6J mice	Both	None	***scRNA-seq*** using Chromium 10x Drop-Seq platform	13,684 cells
Lo et al. *Cancers* 2020	Zhou	Mouse	Normal	Mammary cells were isolated from normal and high fat diet (HFD) fed two month old C57BL/6 *Brca1*$^{-/-}$;*p53*$^{+/-}$ mice	Stroma-imune	EpCAM^{-} FACs sorted viable stromal and immune cells	***scRNA-seq*** using Chromium 10x Drop-Seq platform	3,892 cells
Murrow et al. *bioRxiv* 2020 (preprint v4)	Gartner	Human	Normal	Mammary cells isolated from 28x premenopausal women of varying BMI and age	Both	Live (DAPI^{-}) cells were sorted using CD31^{-}/CD45^{-}/EpCAM$^{+/-}$/ CD49f$^{+/-}$ and CD45^{+}	***scRNA-seq*** using Chromium 10x Drop-Seq platform	87,793 cells
Nguyen et al. *Nature Comms* 2018	Kessenbrock	Human	Normal	Mammary cells isolated from 7x individuals whom underwent reduction mammoplasties	Epithelial	CD31^{-}/CD45^{-} epithelial cells enriched by FACs sorting using CD49f and EpCAM	***scRNA-seq*** using Chromium 10x Drop-Seq and Fluidigm C1 platforms	24,646 cells using 10x and 868 cells using Fluidigm C1
Pal et al. *Nature Comms* 2017	Visvader	Mouse	Normal	Mammary cells were isolated from early postnatal (2 week old), mid-pubertal (5 weeks old) and mature adult (10 weeks, either from virgin or pregnant) FVB/NJ mice	Epithelial	CD45^{-}/CD31^{-}/CD24^{+} epithelial cells were FACS enriched	***scRNA-seq*** using Chromium 10x Drop-Seq and Fluidigm C1 platforms	3,308 cells using 10x and 460 cells using Fluidigm C1

Pervolarakis et al. *Cell Reports* 2020	Watanabe and Kessenbrock	Mouse	Normal	Mammary cells were isolated from 10-week-old FVB/NJ mice	Both	FACS sorting using markers CD31, CD45, EpCAM and CD49f separated live cells for scRNA-seq and basal/luminal cells for scATAC-seq	***scRNA-seq*** and ***scATAC-seq*** was performed using 10x Genomics Chromium platforms	26,859 cells were profiled using scRNA-seq and 23,338 cells were profiled using scATAC-seq
Salmén et al. *bioRxiv* 2018 (preprint)	Lundeberg	Human	Cancer	Tumour tissue sections from 10 patients diagnosed with HER2+ breast cancer	Both	N/A	***Spatial Transcriptomics*** developed by Ståhl et al. *Science* 2016 and others	1007 spatial spots
Sebastian et al. *Cancers* 2020	Loots	Mouse	Cancer	Mammary cells were extracted from 10-week-old BALB/c mice with 4T1 derived mammary tumours (a synergistic model for triple negative breast cancer)	Stroma	Using magnetic cell separation, CD45 depleted or CD45/CD90.1 depleted cells were sequenced, as well as bulk cells and $CD140a^{+}/EpCAM^{-}/CD45^{-}/7AAD^{-}$ FACS sorted fibroblasts.	***scRNA-seq*** using Chromium 10x Drop-Seq platform	6,420 cells
Sun et al. *J. Biol. Chem.* 2018	Deng	Mouse	Normal	Mammary cells were isolated from 3-4 month-old virgin or day 12.5 pregnant FVB mice	Epithelial	Lineage positive (Lin^{+}, endothelial or immune) cells were excluded using the EasySeq mouse epithelial cell enrichment kit. Luminal and basal cells were FACS sorted for using CD24 and CD29	***scRNA-seq*** using Fluidigm C1 platform	239 cells
Thong et al. *Front. Cell Dev. Biol.* 2020	Colacino	Human	Normal	Mammary cells were isolated from 3x normal mammoplasty tissue samples. Additionally, mammary cells from the 3 patients were conditionally reprogrammed *in vitro* and also sequenced.	Both	None	***scRNA-seq*** using drop-seq protocols developed by Macosko et al. 2015.	Not disclosed
The Tabula Muris Consortium *Nature* 2020	-	Mouse	Normal	Mammary cells taken from at 3, 18 and 21 months from C57BL/6JN mice	Both*	Cells from all major mice organs including mammary gland. *However, no mention was made of removal of the mammary lymph nodes prior to	***scRNA-seq*** using Chromium 10x Drop-Seq platform	15,577 mammary cells

						mammary gland dissociation.		
Tognetti et al. *bioRxiv* 2020 (preprint)	Bodenmiller	Human cell lines	Both	62x breast cancer cell lines and 5x normal cell lines	N/A	N/A	***Mass cytometry*** using 34 markers	>80 million cells
Twigger et al. *bioRxiv* 2020 (preprint)	Khaled and Scheel	Human	Normal	4x samples from participants donating mammoplasty tissue and 4x human milk samples from lactating women	Both	None	***scRNA-seq*** using Chromium 10x Drop-Seq platform	24,666 non-lactating breast cells and 27,023 human milk cells
Vatter et al., Cell Reports 2018	Bodenmiller, LaBarge and Lorens	Human	Normal	57 samples were profiled consisting of cultured cells from 44 women and 13 uncultured breast epithelia samples.	Epithelia	N/A	***Mass cytometry*** using 29 markers	880,000 cells
Wagner at al. *Cell* 2019	Bodenmiller	Human	Both	144x human breast tumours and 50x normal breast tissue samples	Both	N/A	***Mass cytometry*** using 73 markers	26 million cells
Wang et al. *eLife* 2020	Schwertfeger	Mouse	Normal	Mammary cells isolated from 10-week-old diestrus FVB/NJ mice	Stroma-imune	$CD45^{+}$ cells were enriched for using magnetic cell separation	***scRNA-seq*** using Chromium 10x Drop-Seq platform	13,000 cells were targeted
Wu et al. *EMBO J* 2020	Swarbrick	Human	Cancer	5x triple negative breast tumours	Both	Viable cells were enriched for using EasySep Dead Cell Removal Kit	***scRNA-seq*** using Chromium 10x Drop-Seq platform	24,271 cells
Wuidart et al. *Nat Cell Biol* 2018	Blanpain	Mouse	Normal	Mammary cells were isolated from embryonic day 14 (E14) and >8-week-old adult mice	Epithelial	Embryonic $CD49f^{Hi}/Lgr5^{Hi}$ cells, adult $CD24^{+}/CD29^{Hi}$ basal and adult $CD24^{+}/CD29^{Lo}$ luminal cells were FACS enriched	***scRNA-seq*** using Smart-seq2	193 cells

Table 2: Summary of nomenclature of adult mammary epithelial cell subtypes

			Basal		**Luminal**			
Reference	**Organism**	**Technique**	***Basal***		***Luminal Progenitor***	***Secretory alveolar***		***Hormone responsive***
Bach et al. 2017	Mouse	scRNA-seq	Basal		Luminal progenitor	Differentiated secretory alveolar		Hormone sensing
Chung et al. 2019	Mouse	snATAC-seq	Basal		Luminal progenitor	N.A.*		Mature luminal
Engelbrecht et al. 2020	Human	scRNA-seq	Basal/ myoepithelial (BA)		Luminal hormone receptor negative progenitors (LHR^-)	N.A.*		Luminal hormone-receptor positive mature cells (LHR^+)
Giraddi et al. 2018	Mouse	scRNA-seq	Basal		Alveolar precursor	N.A.*		Mature Luminal
Han et al. 2018	Mouse	scRNA-seq	Myoepithelial cells		Luminal progenitor	Ductal luminal	Secretory alveoli	N.A.
Knapp et al. 2017	Human	Mass Cytometry	Basal cells (BCs)		Luminal Progenitors (LPs)		Luminal cells (LCs)	
Li et al. 2020	Mouse	scRNA-seq	Myoepithelial		HS-AV	Alveolar (AV)	N.A.*	Hormone sensing (HS)
Murrow et al. 2020	Human	scRNA-seq	Basal/Myoepithelial		Secretory luminal	N.A.*		Hormone responsive (HR+)
Nguyen et al. 2017	Human	scRNA-seq	Basal	Myoepithelial	Secretory L1-type	N.A.*		Hormone responsive L2-type
Pal et al. 2017	Mouse	scRNA-seq	Basal/ Myoepithelial	Mammary stem cells (MaSCs)	Luminal progenitor	Luminal intermediate (Lum Int)	Alveolar	Mature luminal/Ductal
Pervolarakis et al. 2020	Mouse	scRNA-seq/	Myoepithelial		Luminal secretory (L-sec)			Hormone responsive
					Luminal progenitor	Lactation progenitor/precursor	Secretory alveolar	

		scATAC-seq					
Regan and Smalley 2020	N.A.	N.A.	Myoepithelial cells		Estrogen receptor negative (ER^-) ductal cells	Secretory alveolar cells	Hormone-responsive estrogen receptor-positive (ER^+) cells
Sun et al. 2018	Mouse	scRNA-seq	Myoepithelial/Basal: Proliferative basal cells (PBs); Wnt signaling - responsive basal cells (WRBCs); Mammary stem cells (MaSCs)		Proliferative luminal cells (PLCs); Keratinized luminal cells (KLCs)	Mature luminal cells (Mature LCs); Lipid biosynthetic luminal cells (LBLCs)	Stimulus-responsive luminal cells (SRLCs)
Thong et al. 2020	Human	scRNA-seq	Myoepithelial		Luminal 1 (L1)	N.A.*	Luminal 2 (L2)
Twigger et al. 2020	Human	scRNA-seq	Basal cells (BA)		Luminal progenitor (LP)	Secretory alveolar: Luminal 1 (LC1); Luminal 2 (LC2)	Hormone responsive (HR)
Vatter et al. 2018	Human	Mass Cytometry	Basal myoepithelium (MEP)		Luminal Epithelium (LEP)		
Wuidart et al. 2018	Mouse	scRNA-seq	Basal cells (BCs)		Luminal cells (LCs)		

**N.A. pregnancy/lactation time point wasn't included in this study, hence secretory alveolar cells not described*