Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia

Heng Sun‡, Zhengqiang Miao*, Xin Zhang†, Un In Chan‡, Sek Man Su‡, Sen Guo‡, Chris Koon Ho Wong*, Xiaoling Xu†, and Chu-Xia Deng‡

From the ‡Cancer Center, §Zhuhai Research Institute, ‡Genomics and Bioinformatics Core, and ‡Transgenic and Knockout Core, Faculty of Health Sciences, University of Macau, Macau SAR, China

Edited by Xiao-Fan Wang

The mammary gland is very intricately and well organized into distinct tissues, including epithelia, endothelia, adipocytes, and stromal and immune cells. Many mammary gland diseases, such as breast cancer, arise from abnormalities in the mammary epithelium, which is mainly composed of two distinct lineages, the basal and luminal cells. Because of the limitation of traditional transcriptome analysis of bulk mammary cells, the hierarchy and heterogeneity of mammary cells within these two lineages remain unclear. To this end, using single-cell RNA-Seq coupled with FACS analysis and principal component analysis, we determined gene expression profiles of mammary epithelial cells of virgin and pregnant mice. These analyses revealed a much higher heterogeneity among the mammary cells than has been previously reported and enabled cell classification into distinct subgroups according to signature gene markers present in each group. We also identified and verified a rare CDH5+ cell subpopulation within a basal cell lineage as quiescent mammary stem cells (MaSCs). Moreover, using pseudo-temporal analysis, we reconstructed the developmental trajectory of mammary epithelia and uncovered distinct changes in gene expression and in biological functions of mammary cells along the developmental process. In conclusion, our work greatly refines the resolution of the cellular hierarchy in developing mammary tissues. The discovery of CDH5+ cells as MaSCs in these tissues may have implications for our understanding of the initiation, development, and pathogenesis of mammary tumors.

The mammary gland is one of the most typical characteristics of mammals. This milk-producing organ is very intricate and exactly organized, composed of variable cell types, such as epithelia, endothelia, adipocytes, and stromal and immune cells. During puberty, the mammary gland quickly develops and becomes functionally mature in response to sexual hormones; it undergoes particularly dramatic structural and functional changes in pregnancy. So, the mammary gland serves for decades as a perfect example to study the self-renewal and differentiation of adult stem cells, as well as interplaying among distinct types of cells with different functions and differentiation potentials, and also the regulations of cells from other factors of the microenvironment, such as sexual hormones (1).

The human and mouse mammary epithelia share the similar tree-like ductal and lobular structures and consist of two main cellular lineages, termed the basal (or myoepithelial) cells and luminal cells. Previous studies indicate that the mammary stem cells (MaSCs) are enriched in the basal population and show a similar gene expression profile as basal cells and can further differentiate into basal/luminal cells via the progenitor (unipotent stem cells) stage (2–4). Furthermore, only the basal cells instead of the luminal cells possess the potency to reconstruct the whole mammary gland when a limited number of cells are implanted into the cleaned fat pad (5). In the mouse, the combination of a few genes, such as Cd24a, Itgb1, Itga6, Procr, etc., but no single gene, was reported as exclusive markers for MaSCs or progenitors (3, 4, 6). Besides, the details of differentiation progress from MaSCs to functional mammary cells are largely unknown.

From a human health perspective, many mammary gland diseases, in particular breast cancer, are closely related to the abnormality of mammary epithelia, especially the MaSCs and/or progenitors (7, 8). Moreover, the characteristics, aggressiveness, prognosis, and treatment strategies of breast cancer are highly relevant to the heterogeneity and hierarchy of mammary epithelia, whose complexity has not been well revealed.

Single cell RNA-seq is a recently devised and continuously developing technology, which achieves the refinement of gene expression analysis from the bulk of cells deep into the single cell level, and so it greatly facilitates the studies in the stem cell,
embryonic and adult development, cancer research, and many other fields (9, 10). The valuable information harbored in gene-expression profiles contributes greatly to elucidating the previously imperceptible but quite significant differences of seemingly identical single cells. By utilizing this powerful tool, the cellular hierarchy and developmental trajectory of multiple tissues or organs have been discovered, such as lung, kidney, intestine, pancreas, heart, and the immune, hematopoietic, and neural systems (11–23). In this study, we applied single-cell RNA-seq to uncover the heterogeneity and differentiation trajectory of mouse mammary epithelia as well as the novel markers for the MaSCs and other subgroups and also the molecular changes, especially pathway regulation during the MaSCs differentiation in virgin and pregnant states. We identified a rare basal population with exclusive expression of Cdh5 and other genes as quiescent MaSCs. As parts of these candidate markers for MaSCs, including Cdh5, which have been previously reported to be specifically expressed in the endothelial precursors and progeny, play important roles in cell adhesion and other cellular functions, we hypothesize that these lineages of adult stem cells might share similar gene expression features, biological functions, and differentiation processes at given developmental stages.

Results
Single-cell transcriptome profiling of mouse mammary epithelium

To obtain gene expression profiles of single mouse mammary epithelial cells, we devised a workflow, including isolation and sorting of single cells, automated capture and reverse transcription on a Fluidigm C1 platform (24), barcoding, and construction of libraries followed by high-throughput sequencing and bioinformatics analysis (Fig. 1A). As pregnancy greatly stimulates mammary gland development, we aimed to assess the effect of pregnancy on the heterogeneity of the mammary...
cells, and so we isolated the mammary cells from both mature virgin (3–4 months old) and age-matched pregnant day 12 (P12) female mice. To facilitate the data analysis, we pre-classified the mouse mammary epithelia via fluorescence-activated cell sorting (FACS) into basal and luminal lineages separately based on the distinct expression pattern of cell-surface markers CD24 and CD29 (encoded by \( \text{Cd}24\text{a} \) and \( \text{Itgb}1 \), respectively). The CD24\(^{\text{mid}} \)CD29\(^{\text{hi}} \) signature marks for basal cells and the CD24\(^{\text{hi}} \)CD29\(^{\text{lo}} \) characters for luminal cells, regardless of the virgin or pregnant status (Fig. 1B). In line with a previous report (25), pregnancy increases the population of mouse mammary epithelium compared with the virgin stage (Fig. 1C). In total, we obtained 239 high-quality samples, including 34 basal and 54 luminal cells from virgin female mice as well as 102 basal and 49 luminal cells from P12 mice (termed as VB, VL, PB, and PL cells in this study), which were selected from a total of 404 captured single cells via rigorous quality control (Fig. S1 and “Materials and methods”). By comparing the single cell RNA-seq samples, we noticed that the overall gene expression level (number of genes with FPKM >1 or >10) of the basal cells from virgin mice was much less than those from pregnant mice; although the luminal cells showed little difference with or without pregnancy (Fig. 1D). Notably, we totally captured 103 single basal cells from virgin mice, but 69 cells failed to pass the quality control mainly due to low gene-expression levels (Fig. S1). In addition, the 34 VB samples had fewer expressed genes than the VL samples (Fig. 1D), indicating an inactive status of gene expression in basal cells of virgin mice.

### Hierarchy and heterogeneity analysis of mouse mammary epithelium

To evaluate the heterogeneity among the single mammary cells, after collection of the high-resolution RNA-seq data (Table S1), we first performed the principal component analysis (PCA) (Fig. 2A and Fig. S2A). The result showed that the basal and luminal cells were distinctly separated mostly by the PC1 dimension, regardless of the pregnancy status. Moreover, the PC2 dimension to some extent reflected the effect of pregnancy on the basal mammary cells; and in the PC3 dimension, the luminal cells from virgin and P12 mice showed distinct patterns.
**Single-cell RNA-seq of mouse mammary epithelia**

(Fig. S2A). To figure out which genes and related pathways mainly contribute to classification of the cells, we checked the top 1000 genes of PC1, PC2, and PC3. Further analysis indicated these gene sets are enriched for distinct biological functions, such as PC1 for transport and oxidation-reduction process; PC2 for translation and cell adhesion; and PC3 for lipid metabolism process and innate immune response, etc. (Table S2 and Fig. S2B), implying such functions might be closely related to and highly affected by cell lineages and pregnancy.

Next, we examined the co-relationship of the single cells within the four groups based on overall gene expression patterns and found the basal cells showed much higher heterogeneity compared with the luminal cells, and pregnancy reduced the heterogeneity of both basal and luminal cells (Fig. 2B). Then, to reconstruct the hierarchy of the mammary epithelia, we unbiasedly clustered the single cells into six subgroups (C1–C6) based on overall gene-expression profiles (hierarchy clustering analysis) and checked the expression patterns of some housekeeping genes, basal, luminal, and MaSC markers in distinct subgroups at single cell levels (Fig. 2C). In detail, the single basal cells and only a few luminal cells were clustered into subgroups C1–C3, and subgroups C4–C6 consisted of the rest of the luminal cells. Based on hierarchy-clustering analysis, we further divided C2 into two subclusters C2A and C2B, where C2A only included PB cells except for two PL cells, and most members of C2B were from the VB group (Fig. 2C). Studying expression levels of several selected genes, we found the expression of housekeeping genes Actb, Actg1, and B2m was roughly even among all the single cells, confirming there is no obvious technical bias among the RNA-seq libraries (Fig. 2C). In accordance with our FACS setting, the basal cells had high expression of lineage markers Itgb1, Itga6, Krt5, Krt14, and krt17, whereas the luminal cells had high expression of Cd24a, Krt8, and Krt18. Notably, the Esr1, Pgr, and Erbb2, which play significant roles in mammary development and breast cancer formation, were exclusively expressed in subgroups C5 and C6 of luminal cells mainly from virgin mice instead of the P12 mice (Fig. 2C). In line with previous reports (26–28), the damped expression of the receptors should be due to a negative feedback control under the sustained hormone stimuli during pregnancy.

In addition, previously described MaSC markers showed distinct expression patterns in our results, for example, about 70% of all the cells (168/239) expressed Lgr5, and eight cells expressed the Proc (six basal and two luminal cells). Although expression of Axin2 and Inpp5d (also termed as s-ship) could only be detected in one VL cell and one VB cell, respectively, six of the 136 basal cells showed the expression of Bcl11b, a recently identified quiescent MaSC marker (29). Meanwhile, no luminal cell expressed this gene. Another new marker for quiescent MaSCs, Tspan8, was found randomly expressed in 7/136 basal and 12/103 luminal cells (Fig. 2C) (30). Together, the inconsistent expression patterns of previously identified MaSC markers in our single-cell analysis raise the possibility that there might be multiple subpopulations within the mouse mammary gland that behave as stem cells to drive mammary development.

**Rare subpopulation of basal cells bearing bi-lineage gene signature**

To reveal the relationship of the single mammary cells from each subgroup, we performed the t-Distributed Stochastic Neighbor Embedding (tSNE) analysis and found similarity with the PCA results, i.e., the basal and luminal cells were separately clustered (Fig. 3A). Here, we noticed the single cells from the same subgroups mainly grouped together with only a few exceptions. The relative location of these subgroups, to some extent, reflected their similarity of gene expression profiling and biological functions. Notably, subgroup C3 of basal cells showed a close relationship with the luminal cells, indicating their potential for bi-lineage differentiation (Fig. 3A). Additionally, the transcription activity was quite low in subgroup C3 as fewer genes were expressed in this subgroup (Fig. 3B), which is a typical characteristic of some adult stem cells (31, 32). Besides, several C3 basal cells highly expressed the supposed MaSC markers Procr and Inpp5d and only showed low-level expression of basal or luminal lineages markers like Krt14 or Krt18, strongly suggesting their multipotency (Fig. 2C and Fig. S3). Taken together, the above features indicate that C3 cells might serve as bipotent MaSCs.

Next, we focused on these cells and checked the exclusively expressed genes and enriched pathways and functions of the C3 cells, which can be further divided into two subgroups, C3A and C3B (Fig. S4). Although the overall gene expression level is quite low for C3B cells (data not shown), about 90 genes were highly expressed in C3A cells, including some previously reported candidate MaSC markers, such as Cd93 (33), Cldn5 and Flt1 (34), and Igfbp7 (35), as well as some other genes (Fig. 3C and Fig. S4). We then analyzed the protein–protein interactions (Fig. 3D) and enriched biological functions (Fig. 3E) of these signature genes. These results demonstrate a cluster of proteins formed a close-knit interactions network and suggest they might be integrally involved in some signaling pathways or biological functions. Furthermore, the following Gene Ontology (GO) analysis pinpointed that C3 cell signature genes might contribute to vasculature development, angiogenesis, cell adhesion, as well as other cellular processes, some of which are closely relevant to features of adult stem cells. Moreover, we noticed that G1 (like Ppp3ca and Itgb1) but not S or M phase-related genes (like Rad21, Cdk2 and Rad51, and Mki67, respectively) were expressed in C3 cells (Fig. S5), indicating a quiescent status of C3 cells.

**CDH5+ subpopulation cells behave as quiescent MaSCs**

As the C3 cells showed the greatest possibility to behave as MaSCs among all the subgroups via transcriptome analysis, we carefully checked the genes exclusively expressed in C3 and mainly focused on some receptor-encoding genes to see whether any could serve as MaSC markers. Among several candidates, Cdhh5 (Fig. 4A), encoding the VE-cadherin, a previously identified endothelial-specific adhesion molecule located at junctions between endothelial cells (36), drew our attention. Because cellular adhesion proteins, such as several cadherin and integrin family members, play very significant roles in stem cell maintenance and differentiation, such as E-cadherin.
(CDH1), N-cadherin (CDH2), CD24, CD29, and CD49f (37, 38), we hypothesized that CDH5 may not only serve as a MaSC marker but may also contribute to the biological function of the population of MaSCs.

To confirm the Cdh5⁺ cells mainly reside in the basal lineage of mouse mammary epithelia and behave as MaSCs, we next detected the mRNA expression pattern of Cdh5 in mouse mammary glands and found it was indeed mainly expressed in the basal cells instead of the luminal cells (Fig. 4B). Besides, flow cytometry analysis demonstrated that about one-third of basal cells were CDH5⁺ in the basal lineage of 3-month-old virgin female mice, whereas this ratio was reduced to one-fifth in P12 mice (Fig. 4C and Fig. S6). Notably, the total CDH5⁺ basal cell percentage of all mammary epithelia stays roughly equivalent in the two developmental stages (Fig. 4C), ~6% (6.3 ± 0.5% in virgin and 5.4 ± 0.8% in P12 mice). However, only about 4.6 and 0.1% luminal cells expressed CDH5 in virgin and P12 stages, respectively (Fig. 4C).

Then, to assess the regenerative capacity of CDH5⁺ basal cells, we performed the mammosphere assays, and the results
demonstrated a higher mammosphere-forming capacity of CDH5^+ basal cells compared with the CDH5^- basal cells (Fig. 5A). To further confirm the CDH5^+ basal cells are bona fide MaSCs, we performed the in vivo mammary gland reconstruction experiment by implanting different numbers of CDH5^+ cells, or overall basal cells into the cleaned fat pads of nude mice to assess their capacities for rebuilding the mammary epithelia, and the result verified a higher multipotency in mouse mammary gland.

Next, to examine the transcriptional profile of the single Cdhs5^+ mammary cells (with FPKM of Cdhs5 > 1 in single-cell RNA-seq data), we separated the basal population of single cell RNA-seq samples into Cdhs5^+ VB (7), Cdhs5^- VB (27), Cdhs5^+ PB (97), and Cdhs5^- PB (9) subgroups and compared their differentially expressed genes. Based on these samples, ~21% of basal cells in virgin female mice expressed Cdh5 and only ~5% of Cdhs5^- cells were contained in the basal population of P12 mice (Fig. S8A). In addition, we only found 1 of 49 PL cells, and none of the VL cells showed detectable Cdh5 expression (Fig. S8A). Then, we checked the differentially expressed genes and relevant biological functions among the four basal groups. By comparing Cdhs5^+ VB cells with Cdhs5^- PB cells, we found that cellular functions like cell division, adhesion, regulation of transcription, and removal of superoxide radicals were enriched in Cdhs5^- VB cells, although gene ontologies such as RNA splicing, RNA transcription, protein folding, tricarboxylic acid cycle, and 2-oxoglutarate metabolic process, etc. were enriched in Cdhs5^+ PB cells, which reflected the Cdhs5^- basal cells in virgin mice might have a higher proliferation rate but less biosynthesis activity than those in the P12 mice (Fig. 5C).

Meanwhile, some expressed genes and enriched functions were consistent in the Cdhs5^- VB and PB cells, when compared with Cdhs5^- PB and VB cells, such as Cd36, Igf1bp7, and Eng enriched for cell adhesion and Nrp1, Nrp2, Pecam1, and S1pr1 enriched for angiogenesis, etc. Similarly, in both of the Cdhs5^- VB and PB cells, highly expressed Lamb1, Lamb3, Itgav, and Mmp2 enriched for endodermal cell differentiation; Ezr, Nipbl, and Csf1 enriched for positive regulation of multicellular organism growth; Cks1b, Cdh1, Mdk, Myc, and Lgr4 enriched for positive regulation of transcription, DNA-templated, etc. (Fig. 5D and Table S3). Collectively, these results demonstrate that Cdh5^+ is mainly expressed in the basal lineage, and pregnancy (P12) dramatically reduces the ratio of Cdh5^+ basal cells, although parts of the expressed genes and the biological functions are kept conservative regardless of the pregnancy status.

Esr1 and Pgr expression profile in mammary luminal lineage

As sexual hormones play critical roles in development of mammary epithelia, we decided to further check the expression profiles of sexual hormone receptors Esr1 and Pgr in our single cell RNA-seq data. Analysis results of Fig. 2C revealed the absence of Esr1 and Pgr expression in P12 mammary luminal cells, which is consistent with previous reports (26–28). Actually, only 2/49 and 1/49 of PL cells expressed Esr1 and Pgr, respectively. The ratios were even less than those in PB cells (9 and 6%, Fig. S8B). However, about 65 and 39% of VL cells showed detectable expression of Esr1 and Pgr, and 35% of cells expressed both. Meanwhile, only two VB cells expressed Esr1, and none expressed Pgr (Fig. S8B). To understand the behaviors of the VL cells with distinct Esr1 and Pgr expression patterns, we analyzed the differentially expressed genes among Esr1^+ Pgr^- VL cells, Esr1^- Pgr^- VL cells, and Esr1^- Pgr^+ VL cells, which count for 35, 30, and 31% of VL cells, respectively (Table S4). By comparing Esr1^- Pgr^- VL cells with Esr1^- Pgr^+ VL cells, we found that the former enriched cellular function/pathways like nucleosome assembly, chromatin remodeling, and histone H3-K4 trimethylation, all of which are related to
Figure 5. CDH5$^+$ basal cells serve as mouse MaSCs. A, mammosphere assay of CDH5$^+$ and CDH5$^-$ basal cells sorted from 3-month-old virgin female mice. B, mammary epithelial reconstruction assay. C, top GO terms from GO analysis results of differentially expressed genes (Cdh5$^+$ VB versus Cdh5$^-$ PB) enriched in Cdh5$^+$ PB or PB single cells. D, average expression levels of some differentially expressed genes (Cdh5$^+$ versus Cdh5$^-$ VB or PB cells) and their enriched functions (yellow and green colors stand for enriched GO terms in Cdh5$^+$ and Cdh5$^-$ cells, respectively).
epigenetic modification; negative regulation of cell proliferation, indicating a cell cycle-arrested status; and Wnt- and Notch-signaling pathways. Notably, the $\text{Esr1}^+\text{Pgr}^+$ VL cells highly expressed many signaling factors, for example, Bmpr1, Wnt4, Wnt5a, Tnf5f11 (also known as Rankl), and Vegfa. Coincidently, several kinase cascades were activated in the $\text{Esr1}^+\text{Pgr}^+$ VL cells, such as Wnt, TNF, Tgf-$\beta$, PKB, and ERK pathways. In contrast to $\text{Esr1}^+\text{Pgr}^+$ LV cells, the proliferation of the $\text{Esr1}^+\text{Pgr}^-$ VL cells appeared to be activated, which might have resulted from the stimulation of paracrine signals from $\text{Esr1}^+\text{Pgr}^+$ VL cells (Fig. S9A). When comparing $\text{Esr1}^+\text{Pgr}^+$ VL cells with $\text{Esr1}^+\text{Pgr}^-$ VL cells, few significant gene ontology terms were enriched, indicating only a slight difference existed between the two groups. Consistently, Fig. S9B shows cellular function analysis results of the comparison of $\text{Esr1}^+\text{Pgr}^-$ VL cells and $\text{Esr1}^+\text{Pgr}^-$ VL cells, which is largely similar to that from the comparison of $\text{Esr1}^+\text{Pgr}^+$ VL cells and $\text{Esr1}^+\text{Pgr}^-$ VL cells.

We also compared the $\text{Esr1}^+\text{Pgr}^-$ VL cells with $\text{Esr1}^+\text{Pgr}^-$ PL cells and observed many variances on cellular behaviors, some of which were summarized in Fig. S9C. We noticed the biosynthesis process was obviously activated in $\text{Esr1}^+\text{Pgr}^-$ luminal cells at P12 stage, including fatty acid/lipid metabolism, translation, tRNA processing as well as translation, etc. In contrast, $\text{Esr1}^+\text{Pgr}^-$ VL cells were kept in quiescent status; for example, the proliferation was negatively regulated and the MAPK activity was also inhibited, although the ERK signaling and cell migration were activated. Interestingly, apoptotic process was enriched in both groups of cells, although the involved genes were different. So, apoptosis might be consistently activated in the hormone-irrespective luminal cells, indicating a fast turnover of such cells.

**Reconstruction of the mammary epithelia developmental trajectory**

To reveal the continuous developmental process of mammary cells, via hierarchy clustering, we separate the mammary cells based on gene expression profiles into 5 and 6 subgroups with distinct selected markers for each subgroups, respectively (Fig. 6A and Table S5). For the single mammary cells from virgin mice, VB cells are mainly divided into VC1 (as basal progenitor groups with marker genes Trp63, Sna12, C6d3, etc.) and VC2 (as MaSC groups exclusively expressing the MaSC markers identified in this study) subgroups, whereas VL cells are separated into VC3 (lipid biosynthetic luminal cells (LBCDs), which highly express the lipid biosynthesis-associated genes Acls1, Aess2, etc.), VC4 (keratinized luminal cells (KLCs), which highly express many genes related to keratinization, such as Krt6a, Krt6b, Krt76, etc.), and VC5 (stimulus-responsive luminal cells (SRLCs), which express many receptors and signaling transducers such as Erbb2, Smad2, Smo, etc.) subgroups (Fig. 6A and Table S5). In the P12 mice, PC1–PC4 subgroups are mainly composed of PB cells and mixed with a few PL cells, whereas PC5 and PC6 subgroups include the rest of the PL cells. Among these subgroups, eight PC1 cells are proliferative basal cells (PBCs), which highly express Mki67, Cnb1, and Cdk1. The majority of PB cells belong to PC2, which might serve as the mature basal cells. We did not identify specific markers of this subgroup. Cells of PC3 serve as the MaSCs according to our identification of MaSC markers. The PC4 basal cells highly express Wnt5a, Fzd6, and some other genes, which are termed as Wnt signaling–responsive basal cells (WRBCs) here. For the PL cells, the PCs groups are the mature luminal cells, which show activation of biogenesis, whereas the PC6 cells are proliferative luminal cells (PLCs), which highly express mitotic markers Birc5 (Fig. 6A and Table S5).

Having confirmed Cdh5$^+$ C3 subcluster cells (here separated into VC2 and PC3 subclusters) as MaSCs in our single cells pool, we attempted to reconstruct the MaSCs differentiation trajectory via pseudo-temporal Monocle analysis (Fig. 6B) (39). Of note, we found that the MaSCs differentiation showed distinct trajectories in virgin and P12 mice. The MaSCs appeared as a unidirectional differentiation process in the virgin mice, going through a basal stage to the luminal lineage (cells differentiation order: VC2, VC1, VC3, VC5, and VC4); however, in P12 mice, the MaSCs differentiated along two distinct routes to the basal or luminal lineage, respectively (cells differentiation order: PC3, PC2, PC4, and PC1 for basal lineage and PC3, PC5, and PC6 for luminal lineage) (Fig. 6B). This result indicates the VB cells might have the plasticity to be transdifferentiated into a luminal lineage, whereas such potency is blocked in pregnancy under the control of a highly–sexual hormone level.

Having determined the MaSCs differentiation trajectory, we wanted to analyze the corresponding molecular changes during such a process. Therefore, we extracted the differentially expressed genes and inspected the related biological functions (Fig. 7 and Table S6). For the mammary cells from virgin mice, the MaSCs highly express genes involved in vasculature development and regulation of cell proliferation and other gene ontology terms. Also, such pathways were immediately inactivated once the MaSCs differentiated into the next stage (Fig. 7A). Meanwhile, biological functions such as tissue morphogenesis, cytoskeleton organization, RNA processing, and translation began to be activated, and great transcriptome variation showed up in the basal-luminal lineage turning point (Fig. 7A). Furthermore, expression of genes involved in lipid biosynthesis, oxidation reduction, protein transport, TCA cycle, and cell cycle was dramatically up-regulated, sequentially followed by the activation of genes relevant to chromosome organization, telomere maintenance, DNA metabolic process, small GTPase, and MAPK signaling as well as gland morphogenesis (Fig. 7A). In brief, during MaSC differentiation in the virgin stage, cellular functions involved in tissue morphogenesis, biosynthesis, cell activity, proliferation, and differentiation were gradually activated, which appeared quite similar to the P12 mice (Fig. 7, B and C). Notably, for the basal lineage differentiation from MaSCs in P12 mice, genes involved in epithelial–mesenchymal transition (EMT), such as Sox9 and Rbpj, showed a rising then falling pattern, indicating EMT and mesenchymal-epithelial transition might play phased roles in MaSC differentiation (Fig. 7B and Fig. S10A). Furthermore, in both the basal and luminal lineages, there appeared a distinct subcluster (stage) of enriched cells within the mitotic phase of the cell cycle, which implies that the cell cycle might be closely related to the MaSCs differentiation (Fig. S10B).
Discussion

We have studied the heterogeneity and the hierarchy of mouse mammary epithelia via single-cell transcriptome profiling. Previously, the hierarchy of mouse mammary epithelia was only vaguely elucidated via distinguishing the mammary cells by just a handful of representative markers, like CD24, CD29, CD49f, CD61, KIT, ESR1, PGR, and ERBB2, etc. (40). Based on the distinct expression patterns of these well-accepted markers, the mouse mammary epithelial cells could be classified into several groups, for example the MaSCs, bipotent progenitors,
Figure 7. Differentially expressed genes and related functions during MaSCs differentiation. A–C, differentially expressed genes and related functions during MaSCs differentiation in virgin (A) and P12 mice (B for basal lineage and C for luminal lineage). Left, the expression trends of each cluster of genes; middle, the heatmap of differentially expressed genes along the pseudo-differentiation trajectory of MaSCs; right, enriched GO terms and representative genes of each gene cluster.
Identification and characterization of MaSCs are primarily based on the definition and features of adult stem cells, in simple terms, the capacity of cell self-renewal and diverse differentiation. Heretofore, the in vivo genetic-lineage tracing, cleaned fat pad implantation assays, and in vitro colony- and mammosphere-forming assays have been widely used to verify the authenticity of MaSCs (47). Nevertheless, one shortage of the lineage-tracing technology is only the proliferating cells can be traced, whereas the quiescent MaSCs with almost no progeny cannot be identified unless they are aroused into cell cycling by some stimulation like hormones. Notably, several recent studies indicated the existence of quiescent MaSCs in mouse mammary glands and identified several markers for this subpopulation (29, 30). For example, Cai et al. (29) verified that a Bcl11bHigh subpopulation of basal cells behaved as quiescent mammary stem cells and demonstrated a Cdkn2a-dependent mechanism for Bcl11b maintaining the quiescence and homeostasis of MaSCs. Another report from Visvader and co-workers (30) identified tetraspanin8 as a novel marker for dormant MaSCs and found these tetraspanin8+ MaSCs could be activated in response to steroid hormones. In our single-cell RNA-seq analysis, we confirmed the exclusive expression of Bcl11b in basal cells, but tetraspanin8 seems to be sporadically expressed in both basal and luminal lineages (Fig. 2C). Moreover, neither of the two novel populations overlaps with the Cdh5+ MaSCs discovered in this study (Table S7), indicating there might be multiple dormant subpopulations with distinct gene signatures and even entirely different origins within the mammary epithelia, all of which could undertake the multilineage development. Hence, more investigations are necessary to verify these seemingly inconsistent results, especially in consideration of the effect of the mammary niche for MaSC homestasis and differentiation. Nevertheless, all of these studies including ours confirm the existence of quiescent MaSCs that are poised for activation in response to suitable environmental signals. Furthermore, the MaSC differentiation trajectory revealed by our single cell transcriptome profile adds more details to the activation process of quiescent MaSCs. For instance, we found the RNA processing, protein translation and transport, lipid biosynthesis, oxidation reduction, TCA cycle, as well as many other pathways involved in tissue morphogenesis, biosynthesis, cell activity, proliferation, and differentiation are indeed activated when the Cdh5+ dormant cells begin to differentiate. Of course, more studies are required to illustrate what mechanisms control the trigger of the dormant MaSCs proliferation and differentiation.

The mammary tumor cells, especially the mammary/breast cancer stem cells, share many features with the mammary stem cells or progenitors. For example, the genomic mutation and/or epigenetic variations accumulate in the stem cells, or the progenitors more easily drive the tumorigenesis compared with the differentiated cells. In addition, the tumor initiation and developmental processes always couple with the dedifferentiation, reprogramming the normal cells countercurrent to a stem cell-like status. In our study, we identified a rare basal population that serves as quiescent MaSCs with a distinct gene expression profile. By checking the expression levels of MaSCs markers in TCGA human breast cancer samples, we discovered a signifi-
Single-cell RNA-seq of mouse mammary epithelia

cant negative correlation between the expression of MaSCs markers and the outcomes of the patients (Fig. S11). Hence, the determination of MaSC characters could greatly facilitate the further understanding of initiation, development, and pathogenesis of mammary tumors as well as contribute to the diagnosis and treatment of the disease.

Materials and methods
Suspension preparation of single mammary cells

Three-to-four month-old female virgin or pregnant at day 12.5 FVB mice were sacrificed to isolate the fourth pair of mammary glands for single-cell preparation as described previously (48). In brief, the mammary glands were minced, washed in PBS, and digested in DMEM/F-12 containing 300 units/ml collagenase III (S4M7602S, Worthington), 100 units/ml hyaluronidase (H3506; Sigma), 5% FBS (Gibco), 5 μg/ml insulin (350–020; BIOSOURCE), 10 ng/ml EGF (13247–051, Invitrogen), and 500 ng/ml hydrocortisone (H8088; Sigma) for about 1 h at 37 °C, 5% CO2. The resultant organoid suspension was sequentially resuspended in DMEM/F-12 supplemented with 5 mg/ml dispase II (825-001; Roche Diagnostics) and 0.1 mg/ml DNase I (58C10349; Worthington) for 5 min at 37 °C, and then digested with 0.25% trypsin/EDTA for 1–2 min, and treated with RBC lysis buffer (00433–57; eBioscience, San Diego) for 3 min to remove the red blood cells before filtration through a 40-μm cell strainer (352340, Falcon) to obtain single-cell suspension. Enrichment of Lin− epithelial cells was achieved through selective depletion of hematopoietic and endothelial cells by using the EasySeq mouse epithelial cell enrichment kit according to the manufacturer (19758; Stemcell). Our studies involving mice were approved by the University of Macau Animal Ethics Committee under protocol UMAEC-050-2015.

FACS

The enriched Lin− mouse mammary epithelial cells were stained at a concentration of 1 × 10⁶ cells per 100 μl of FACS buffer (Hanks’ balanced salt solution with 0.5% BSA and 2 mM EDTA). Antibodies against mouse CD24 (anti-CD24-PE-Cy™7, BD Biosciences), CD29 (anti-CD29-APC, Biolegend), and CDH5 (also known as CD144, anti-CD144-PE, BD Biosciences) were used. DAPI (1 μg/ml) was also used as an indicator for cell viability. Antibody incubations were performed at 4 °C for 30 min. Then the cells were washed with PBS twice before sorting by using FACSaria III Cell Sorter (BD Biosciences). DAPI−CD24−/CD29− cells and DAPI−CD24+/CD29− cells were gated and sorted out as basal and luminal mammary epithelial cells, respectively (3).

Single-cell capture and RNA-seq

The sorted mammary basal or luminal cells resuspended in FACS buffer (300–500 cells/μl) were mixed (3:2 ratio) with C1 Cell Suspension Reagent (Fluidigm) before loading onto a 5–10– or 10–17–μm diameter C1 integrated fluidic circuit (Fluidigm). Each capture site was carefully examined and recorded under a Zeiss microscope in bright-field channel for cell doublets. Cell lysis, reverse transcription, and cDNA amplification were performed on the C1 single-cell auto-preparation integrated fluidic circuit, as specified by the manufacturer (protocol 100-7168 E1). The SMARTer ultra low RNA kit (Clontech) was used for cDNA synthesis from the single cells. Illumina NGS library was constructed with Nextera XT DNA sample preparation kit (Illumina), according to the manufacturer’s recommendations (protocol 100-716 E1). Sequencing was performed on Illumina HiSeq2500 (Illumina) high output run mode by multiplexed paired end run with 60, 100, or 125 cycles.

Data processing and analysis

The single-cell RNA-seq reads were aligned to the reference mm10 mouse genome with TopHat 2.1.1 by using default parameters (49). Gene expression was quantified as FPKM using Cufflinks 2.2.1 (50). We set rigorous quality control to remove the libraries with poor quality. The libraries with unique mapped reads of <1 million were removed. We detected the coverage of mapped reads on housekeeping genes and then clustered the libraries based on the coverage distribution and removed the libraries with low coverage. We also detected the coverage of mapped reads on all the genes and then clustered the libraries based on the coverage distribution and removed the libraries with low coverage. We further examined the libraries distribution based on read counts and gene expression features (numbers of expressed genes) and removed the libraries with low read counts and gene expression features. Most removed libraries could not meet two or more criteria. Moreover, we only retained the genes that expressed >5% samples in any of the four groups of single cells (VB, VL, PB, and PL). This resulted in a matrix of 239 single cells × 12,688 genes. For the following analyses, the expression values (log2(FPKM + 1)) of all expressed genes were utilized with R. To cluster the single cells into subgroups, PCA, hierarchical clustering, and tSNE were performed using R with default parameters. To assess the heterogeneity of single cells among different subgroups, the Spearman correlation across all cells within a given subgroup was calculated; while the higher heterogeneity exists, the smaller correlation should be expected and vice versa. Differentially expressed genes among subclusters were identified by using DESeq2 (51). Also, the following cellular function analysis of differentially expressed genes was performed using DAVID (52). Protein–protein network analysis of interested genes was performed using STRING (53). To reconstruct the mammary developmental trajectory based on single-cell dynamic transcriptionomes, pseudo-temporal analysis was performed using Monocle as reported previously (39). To investigate the relationship of breast cancer patient outcomes and MaSC markers expression, the TCGA database for breast cancer cohorts was utilized for Kaplan-Meier survival analysis.

Real-time PCR

Total RNA was extracted from about 5 × 10⁵ freshly fluorescence-activated cell-sorted mouse mammary cells after washing with PBS by using TRIzol reagent according to the manufacturer’s instructions (Life Technologies, Inc., 15596026). Reverse transcription to cDNA was initiated with 1 μg of each RNA sample using a QuantiTect reverse transcription kit (QIagen Inc., 205313) following the manufacturer’s instructions.
Quantitative real-time PCR was performed using the Applied Biosystems™ QuantStudio™ 7 Flex real-time PCR system (Thermo Fisher Scientific). The reaction mixture contained 6 μl of 2× SYBR Green (Roche Diagnostics, 04707516001), 4 μl of PCR-qualified H₂O, 0.5 μl of forward primer (10 μM), 0.5 μl of reverse primer (10 μM), and 1 μl of cDNA from the 100 μl of diluted (1:5) stock solution prepared above. Each sample was analyzed in a 384-well PCR plate. The data were analyzed initially using the SPSS 13.0 software (SPSS Co.) included with the PCR machine. The results were analyzed statistically and graphed using Microsoft Excel 2010. The primers used for real-time RT-PCR are shown in Table S8. Actb was used as an internal control.

**Mammospheres assay**

For generation of mammospheres, 1 × 10⁵ single cells were plated on 24-well ultra-low attachment plates (Corning) in DMEM/F-12 medium containing 20 ng/ml basic fibroblast growth factor (13256-029; Invitrogen), 20 ng/ml epidermal growth factor (13247-051; Invitrogen), B-27 supplement (1:50 dilution, 17590-044; Invitrogen), and antibiotic/antimycotic (1:100 dilution, 15240-062; Invitrogen) and cultured at 37 °C, 5% CO₂. Another 0.5 ml of new medium was added to each well every other 5 days. After 12 days, the mammospheres with 100 μm and over were examined and counted under microscope.

**In vivo mammary gland formation**

Three-week-old virgin female nude mice underwent surgery to remove the fourth inguinal mammary gland outgrowths between the nipple and lymph node. For mammary fat pad implantation, the primary mouse mammary cells isolated from 3-month-old FVB female mice after sorting were washed two times with PBS and then resuspended at given concentrations in DMEM/F-12 with 30% FBS and 0.04% trypan blue. 10 μl of the cell suspension from each group was injected into a pre-cleared mammary fat pad. Eight weeks after injection, the outgrowths were harvested and processed for whole-mount imaging. All animals were handled and housed in accordance with the guidelines of the Institutional Animal Care and Use Committee of the University of Macau. This study was approved by the University of Macau Animal Ethics Committee with protocol UMAEC-050-2015.

**Whole-mount staining and analysis**

Mammary gland was fixed by Carnoy’s fix buffer (60% ethanol, 30% CHCl₃, and 10% acetic acid) overnight. Fixed tissue was washed with a gradient of ethanol (100, 95, 70, 50, and 30%) and PBS and then stained with carmine-alum staining solution (0.2% Carmine, Sigma C1022 with 0.5% aluminum potassium sulfate, Sigma A7210) overnight. The next day, tissue was washed with PBS and a gradient of ethanol (70, 95, and 100%). Fat tissue was removed by xylene treatment, and tissue was mounted using Permount (Thermo Fisher Scientific) for long-term preservation. The pictures of stained whole mounts were taken under a Leica-M165FC stereo microscope, and the sizes of reconstructed epithelia and whole mounts were measured by using ImageJ. Then, the mammary repopulating unit frequency and confidence interval were determined by ELDA (54).
ral gene expression during heart development. *Dev. Cell* **39**, 480–490

13. Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255

14. La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L. E., Stott, S. R. W., Toledo, E. M., Villaescusa, J. C., Lönnberg, P., Ryge, J., Barker, R. A., Arenas, E., and Linnarsson, S. (2016) Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**, 566–580.e19

15. Li, G., Xu, A., Sim, S., Priest, J. R., Tian, X., Khan, T., Quertermous, T., Zhou, B., Tsao, P. S., Quake, S. R., and Wu, S. M. (2016) Transcriptionic profiling maps anatomically patterned subpopulations among single embryonic cardiac cells. *Dev. Cell* **39**, 491–507

16. Li, J., Klughammer, J., Farlik, M., Penz, T., Spittler, A., Barbieux, C., Beresewich, E., Bock, C., and Kubicek, S. (2016) Single-cell transcriptionomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* **17**, 178–187

17. Llorens-Bobadilla, E., Zhao, S., Baser, A., Saiz-Castro, G., Zwaldo, K., and Martin-Villalba, A. (2015) Single-cell transcriptomics reveals a population of dormant neural stem cells that become activated upon brain injury. *Cell Stem Cell* **17**, 329–340

18. Shahle, A. K., Satija, R., Shaju, J., Trombeta, J. J., Gennert, D., Lu, D., Chen, P., Gerttner, R. S., Gaudabile, J. T. J., Yosef, N., Schwartz, S., Fowler, B., Weaver, S., Wang, J., Wang, X., et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369

19. Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., Enikolopov, G., Nauen, D. W., Christian, K. M., Ming, G. L., and Song, H. (2015) Single-cell RNA-Seq with waterfall reveals molecular cascades underlying islet cell types. *Science** 347**, 381–386

20. Treutlein, B., Brownfield, D. G., Wu, A. R., Neff, N. F., Mantalas, G. L., Espinoza, F. H., Desai, T. J., Krasnow, M. A., and Quake, S. R. (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375

21. Xin, Y., Kim, J., Ni, M., Wei, Y., Okamoto, H., Lee, J., Adler, C., Cavino, K., Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Pollen, A. A., Nowakowski, T. J., Shuga, J., Wang, X., Leyrat, A. A., Liu, J. H., Li, N., Szpankowski, L., Fowler, B., Chen, P., Ramalingam, N., Sun, G., Yu, J., Wang, F., Chen, E. Y., Scheeren, F. A., Kuo, A. H., Sikandar, S. S., Hisamori, S., et al. (2017) A quiescent Bcl11b high stem cell population is required for maintenance of the mammary gland. *Cell Stem Cell* **20**, 247–260.e5

22. Fu, N. Y., Rios, A. C., Pal, B., Law, C. W., Jamieson, P., Liu, R., Vaillant, F., Jackling, F., Liu, K. H., Smyth, G. K., Lindeman, G. J., Ritchie, M. E., and Visvader, E. J. (2017) Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. *Nat. Cell Biol.* **19**, 164–176

23. Arai, F., Hirao, A., Ohmura, M., Sato, H., Matsuoka, S., Takubo, K., Ito, K., Koh, G. Y., and Suda, T. (2004) Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149–161

24. Cheshier, S. H., Morrison, S. J., Liao, X., and Weissman, I. L. (1999) In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3120–3125

25. dos Santos, C. O., Rebbeck, C., Rozhkov, E., Valentine, A., Samuels, A., Kadiri, L. R., Osten, P., Harris, E. Y., Uren, P. J., Smith, A. D., and Hannon, G. J. (2013) Molecular hierarchy of mammary differentiation yields refined markers of mammary stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7123–7130

26. Soody, K. J., Kendrick, H., Gao, Q., Tutt, A., Zvelebil, M., Ordonez, L. D., Quist, J., Tan, D. W., Isacce, C. M., Grigoridias, A., and Small, M. J. (2015) Mouse mammary stem cells express prognostic markers for triple-negative breast cancer. *Breast Cancer Res.* **17**, 31

27. Proia, T. A., Keller, P. J., Gupta, P. B., Klebba, I., Jones, A. D., Sedic, M., Gilmore, H., Tung, N., Naber, S. P., Schnitt, S., Lander, E. S., and Kuperwasser, C. (2011) Genetic predisposition directs breast cancer phenotype by dictating progenitor cell fate. *Cell Stem Cell* **8**, 149–163

28. Lampugnani, M. G., Resnati, M., Rafteri, M., Pigott, R., Piscane, A., Houen, G., Roçu, L. P., and Dejana, E. (1992) A novel endothelial-specific membrane protein is a marker of cell-cell contacts. *J. Cell Biol.* **118**, 1511–1522

29. Browse, A. B., Chong, F., Gray, P. P., and Munro, T. P. (2011) Stem cell integrins: implications for *ex vivo* culture and cellular therapies. *Stem Cell Res.* **6**, 1–12

30. Raymond, K., Deugnier, M. A., Faraldo, M. M., and Glukhova, M. A. (2009) Adhesion within the stem cell niches. *Curr. Opin. Cell Biol.* **21**, 623–629

31. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. (2015) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386

32. Visvader, I. E. (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Ges. Dev.* **23**, 2563–2577

33. Fina, L., Molgaard, H. V., Robertson, D., Bradley, N. J., Monaghan, P., Delia, D., Sutherland, D. R., Baker, M. A., and Greaves, M. F. (1990) Expression of the CD34 gene in vascular endothelial cells. *Blood* **75**, 2417–2426

34. Morita, K., Sasaki, H., Furuse, M., and Tsukita, S. (1999) Endothelial claudin-5/TMVCF constitutes tight junction strands in endothelial cells. *J. Cell Biol.* **147**, 185–194

35. Soker, S., Takashima, S., Miao, H. Q., Neufeld, G., and Klagsbrun, M. (1998) Neurelin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell* **92**, 735–745

36. Newman, P. J., Berndt, M. C., Gorski, J., White, G. C., 2nd., Lyman, S., Paddock, C., and Muller, W. A. (1990) PECAM-1 (CD31) cloning and relation to adhesion molecules of the immunoglobulin gene superfamily. *Science* **247**, 1219–1222
45. Yu, Q. C., Song, W., Wang, D., and Zeng, Y. A. (2016) Identification of blood vascular endothelial stem cells by the expression of protein C receptor. Cell Res. 26, 1079–1098 CrossRef Medline

46. Jiang, S., Lee, B. C., Fu, Y., Avraham, S., Lim, B., and Avraham, H. K. (2010) Reconstitution of mammary epithelial morphogenesis by murine embryonic stem cells undergoing hematopoietic stem cell differentiation. PLoS ONE 5, e9707 CrossRef Medline

47. Visvader, J. E., and Stingl, J. (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. 28, 1143–1158 CrossRef Medline

48. Vassilopoulos, A., Wang, R. H., Petrovas, C., Ambrozak, D., Koup, R., and Deng, C. X. (2008) Identification and characterization of cancer initiating cells from BRCA1 related mammary tumors using markers for normal mammary stem cells. Int. J. Biol. Sci. 4, 133–142 Medline

49. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36 CrossRef Medline

50. Trapnell, C., Hendrickson, D. G., Sauvageau, M., Goff, L., Rinn, J. L., and Pachter, L. (2013) Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53 CrossRef Medline

51. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 CrossRef Medline

52. Huang da, W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 CrossRef Medline

53. Snel, B., Lehmann, G., Bork, P., and Huynen, M. A. (2000) STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res. 28, 3442–3444 CrossRef Medline

54. Hu, Y., and Smyth, G. K. (2009) ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cells and other assays. J. Immunol. Methods 347, 70–78 CrossRef Medline

4. Single-cell RNA-seq of mouse mammary epithelia

J. Biol. Chem. (2018) 293(22) 8315–8329 8329