 doi:10.20944/preprints201906.0128.v1

Integrating Single Cell RNA-Sequencing Analyses and Functional Assays to Decipher Cell Differentiation Hierarchies in the Mammary Epithelium

Joseph L. Regan[1*] & Matthew J. Smalley[2]

[1]Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

[2]European Cancer Stem Cell Research Institute, School of Biosciences, Hadyn Ellis Building, Cardiff University, CF24 4HQ, UK

*Correspondence: joseph.regan@charite.de

Abstract

The identification and molecular characterization of cellular hierarchies in complex tissues is key to understanding both normal cellular homoeostasis and tumorigenesis. The mammary epithelium is a heterogeneous tissue consisting of two main cellular compartments, an outer basal layer containing myoepithelial cells and an inner luminal layer consisting of estrogen receptor negative (ER^-) ductal cells and secretory alveolar cells (in the fully functional differentiated tissue) and hormone responsive estrogen receptor positive (ER^+) cells. Recent publications in Nature Communications used single cell RNA-sequencing (scRNA-seq) analysis to decipher epithelial cell differentiation hierarchies in human (Nguyen et al., 2018) and murine (Pal et al., 2017) mammary glands and report the identification of new cell types based on the expression of the luminal progenitor cell marker KIT (c-Kit). However, there are several inaccuracies and unfortunate omissions in the citation of previous research in each of these studies. As a result, the overall conclusions on the significance of these reports, in particular the claimed identification of new cell types is not accurate. Here we discuss these studies in the context of our previous research (Regan

 doi:10.20944/preprints201906.0128.v1

et al., 2012), in which we identified c-Kit as a luminal progenitor cell marker and functionally characterized cellular subpopulations analogous to those reported in the recent scRNA-seq studies.

Keywords: stem cells; luminal progenitors; mammary; breast; cell hierarchy; differentiation; single cell RNA-sequencing; lineage tracing; c-Kit

To investigate cellular heterogeneity and lineage relationships in the human breast, Nguyen *et al.* performed scRNA-seq analysis on fluorescence-activated cell sorted (FACS) breast epithelial cells and reported the identification of previously unidentified cell types within the three main mammary epithelial cell populations, previously identified as basal (B: $CD49f^{High}$ $EPCAM^{+}$, $K14^{+}$), luminal progenitors (L1: $CD49f^{+}$ $EPCAM^{+}$, ER^{-}, $K8/18^{+}$), and mature luminal (L2: $CD49f^{-}$ $EPCAM^{+}$, ER^{+}, $K8/18^{+}$) cells (Nguyen et al., 2018; Shehata et al., 2012). Significantly, the authors detected replicating cells in all three main populations (Basal, L1, and L2), suggesting that each cluster may be maintained by its own progenitor cell population and proposed a continuous lineage hierarchy connecting the basal lineage to the two luminal branches via a bipotent MaSC. Furthermore, the authors identify what they describe as previously unidentified adult luminal cells that co-express both luminal (KRT8/18) and basal (KRT14) markers *in situ*.

When discussing previous reports demonstrating heterogeneity within the mammary epithelium Nguyen *et al.* refer to a review article and state, “while several distinct subpopulations of murine basal and luminal cells have been reported anecdotally (Visvader and Stingl, 2014), comprehensive knowledge about expression signatures and cellular identities of these subpopulations remains sparse”. Anecdotal evidence refers to casual observations or hearsay rather than conclusions supported by rigorous scientific study. Previous studies involving the functional and molecular characterization of mammary epithelial subpopulations were not based on anecdotal evidence. The cellular identities of mammary epithelial

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

subpopulations have been rigorously characterized using flow cytometric and functional transplantation assays (DeOme et al., 1959; Regan and Smalley, 2007) as well as, more recently, lineage tracing studies (Davis et al., 2016; Koren et al., 2015; Lloyd-Lewis et al., 2018; Rios et al., 2014; Scheele et al., 2017; van Amerongen, 2015; van de Moosdijk et al., 2017; Van Keymeulen et al., 2011; Van Keymeulen et al., 2017; Wang et al., 2014). Transplantation experiments have generally supported a model in which MaCSs, cells capable of regenerating the epithelium when injected into a cleared mammary fat pad (one free of endogenous epithelium)(DeOme et al., 1959; Regan and Smalley, 2007), are localised to the basal cell layer (Asselin-Labat et al., 2007; Regan et al., 2012; Shackleton et al., 2006; Sleeman et al., 2006; Stingl et al., 2006). Progenitor cells, which are functionally defined by high colony forming and proliferative potential *in vitro* and limited repopulating ability when transplanted into cleared fat pads, are localised to the luminal layer (Asselin-Labat et al., 2007; Regan et al., 2012; Shehata et al., 2012; Sleeman et al., 2007; Taddei et al., 2008). Differentiated cells do not transplant or generate colonies *in vitro*. The molecular profiling of mammary epithelial subpopulations functionally defined by their transplantation potential has also been extensive (Grigoriadis et al., 2006; Jones et al., 2004; Kendrick et al., 2008; Lim et al., 2009; Lim et al., 2010; Pal et al., 2015; Pece et al., 2010; Raouf et al., 2008; Shipitsin et al., 2007; Spike et al., 2012; Stingl et al., 2006; Wansbury et al., 2011).

However, lineage tracing experiments have yielded results which in some cases conflict with this model. *In situ* evidence from early mammary development, puberty and alveolargenesis during pregnancy has shown that basal stem cells can contribute to the luminal layer (Regan et al., 2013; van Amerongen et al., 2012; Van Keymeulen et al., 2011; Wuidart et al., 2018). We previously proposed, based on *in situ* analysis, that basal MaSCs located in the cap cell layer of terminal end buds (TEBS), the outermost cell layer of the specialized growth structure that drives ductal growth during puberty, are bipotent and produce daughter cells that contribute to both the basal and luminal cell lineages (Regan et al., 2013). Lineage tracing experiments from Rios *et al.* (2014) and Wang et al. (2014) were in agreement with

transplantation data and our *in situ* analysis suggesting that MaSCs in the postnatal gland are bipotent (Regan et al., 2013; Rios et al., 2014; Wang et al., 2014). In contrast, work from Van Keymeulan *et al*. (2011 and 2017) using cell type-specific promoters in adult tissue suggested that in the resting post-pubertal gland, MaSCs are unipotent and that basal and luminal lineages are maintained by separate stem/progenitor populations (Van Keymeulen et al., 2011; Van Keymeulen et al., 2017). Studies in mouse tumor models suggested that during oncogenic transformation basal and luminal cell populations lose this restricted lineage potential and acquire multipotency (Koren et al., 2015; Van Keymeulen et al., 2017).

Throughout their paper, Nguyen *et al.* discuss KIT (c-Kit) as a marker of mammary progenitor cells. When discussing KIT as a progenitor cell marker, the authors consistently refer to two papers, namely Stingl *et al.* (2001)(Stingl et al., 2001) and Shehata *et al.* (2012)(Shehata et al., 2012) (references 7 and 8 in Nguyen *et al.* 2018). The Stingl *et al.* paper from 2001, which characterized bipotent mammary epithelial progenitor cells in normal adult human breast tissue, does not include any mention KIT as a progenitor cell marker (or otherwise). Shehata *et al.* (2012) performed phenotypic and functional characterization of the luminal cell hierarchy of the mammary gland and discussed our previous work (Regan et al., 2012) when describing their reasons for further investigating c-Kit as a marker of luminal progenitor cells. The Shehata *et al*. study is, therefore, not the primary research paper demonstrating c-Kit as a luminal progenitor cell marker.

Using c-Kit as a marker in combination with CD24, Sca-1 and CD49f, we isolated and characterized, using functional limiting dilution (down to a single cell) cleared mammary gland fat pad transplants, 2D and 3D *in vitro* colony forming assays, gene expression and immunofluorescence analysis, the most highly enriched subpopulation of MaSCs so far reported. In addition, we described distinct c-Kit-expressing cells within each of the mammary epithelial basal (CD24$^{+/Low}$ Sca-1^{-} CD49f$^{+/High}$ c-Kit^{-} and c-Kit^{+}), myoepithelial (CD24$^{+/Low}$ Sca-1^{-} CD49f$^{+/Low}$ c-Kit^{-} and c-kit^{+}), luminal ER^{-} (CD24$^{+/High}$ Sca-1^{-} c-Kit$^{+/Low}$ and c-Kit$^{+/High}$)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

and luminal ER^{+} (CD24$^{+/Low}$ Sca-1^{-} c-Kit^{-}, CD24$^{+/Low}$ Sca-1^{+} c-kit^{-} and c-kit^{+}) cellular compartments (Regan et al., 2012). These studies resulted in the identification of nine distinct cell types within the mammary epithelium and demonstrated that luminal ER^{-} progenitors, the cells of origin of BRCA1 mutation-associated breast cancer (Lim et al., 2009; Molyneux et al., 2010) (which are identified as L1.1 cells by Nguyen *et al*.), are c-Kit^{+}.

Similar to our 2012 study, Nguyen *et al*. also identify *KIT* expressing cells in each of the cellular subtypes described in their paper. The expression of *KIT* (as well as the luminal markers *KRT8/18* and *ESR1* and basal marker *KRT14*) in each of Nguyen *et al*.'s human breast populations of B, Myo, L1.1, L1.2 and L2 are consistent with the expression levels reported by us in the equivalent murine basal, myoepithelial, luminal ER^{-} c-Kit$^{+/High}$, luminal ER^{-} c-Kit$^{+/Low}$, and luminal ER^{+} cells, respectively (Figure 1). The progenitor cells identified by Nguyen *et al*. in the three main mammary epithelial compartments (Basal, L1, and L2) are therefore equivalent to the c-Kit-expressing basal and luminal progenitor cells previously reported by us.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

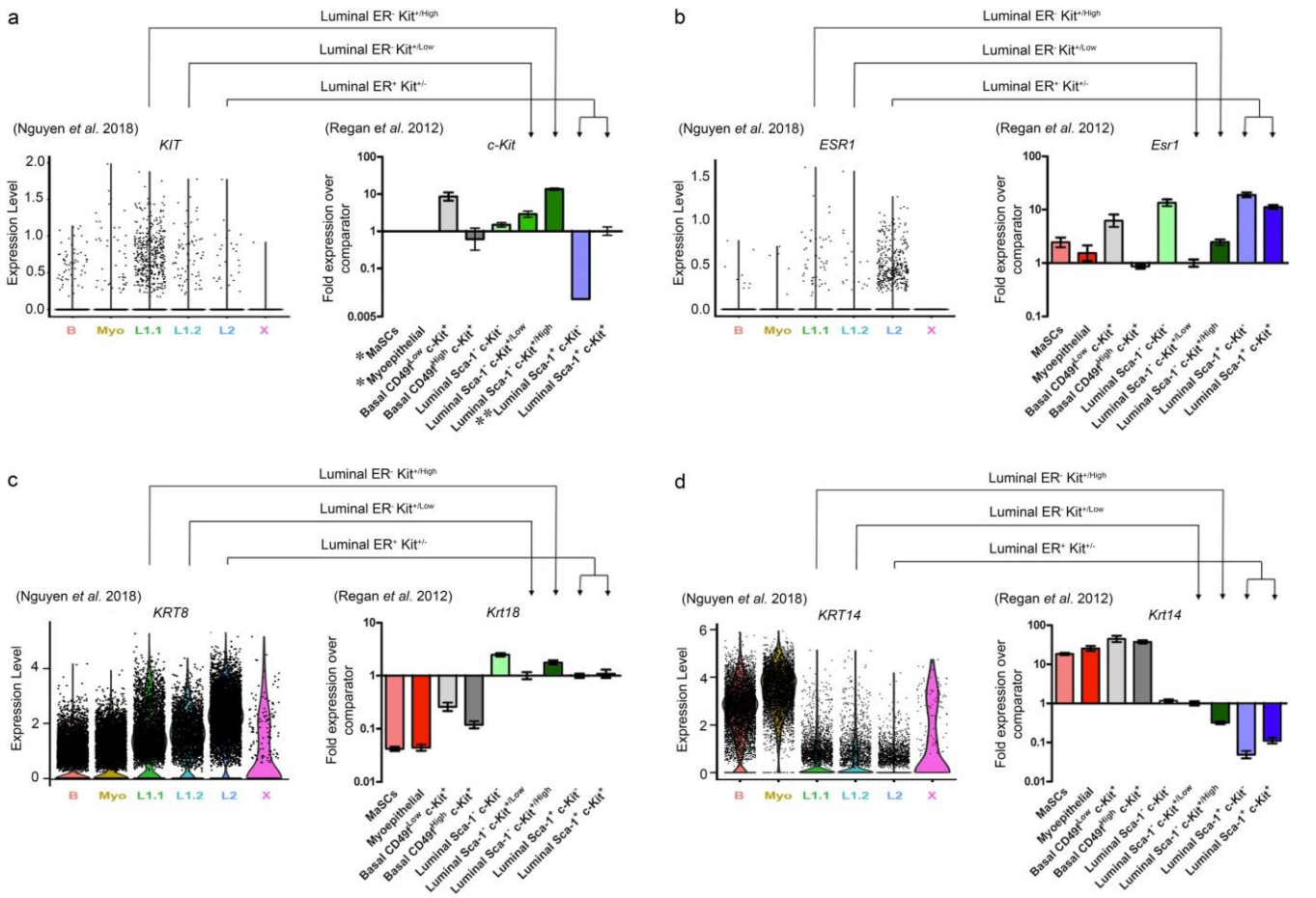

Figure 1: Comparison of gene expression in cell populations identified by Nguyen *et al.* (2018) and Regan *et al*. (2012). Nguyen *et al.* (2018) violin plots showing the expression pattern of progenitor gene *KIT* (a; LHS), luminal genes *ESR1* and *KRT8* (b – c; LHS) and basal gene *KRT14* (d; LHS) grouped by final cluster determination in human mammary epithelium. B = Basal (containing MaSCs), Myo = Myoepithelial. Regan *et al*. (2012) gene expression in the different cellular subpopulations as determined by qPCR relative to comparator (luminal Sca-1$^+$ c-Kit$^+$ cells) for progenitor gene *c-Kit* (a; RHS), luminal genes *Esr1* and K*rt18* (b – c; RHS) and basal gene *Krt14* (d; RHS) in murine mammary epithelium. Data are presented as fold expression levels ±95% confidence intervals (n=three independently harvested isolates of each cell population). *Gene expression was undetectable in these populations in all three independent isolates. **Gene expression was only detected (at very low levels) in two of three isolates of the luminal Sca-1$^+$ c-Kit$^-$ population. Therefore, no error bars are shown for this sample.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

Nguyen *et al.* observed fractions of cells that co-express both luminal KRT8 and basal KRT14 markers and report that such KRT8$^+$/KRT14$^+$ cells had previously been observed in mouse fetal MaSCs(Spike et al., 2012) but not in adult human tissue in homeostasis. This statement is incorrect. Firstly, in addition to the Spike *et al.* (2012) paper cited by Nguyen *et al.*, fetal MaSCs that co-express KRT8$^+$ and KRT14$^+$ have previously also been reported by Sun *et al.*, (2010)(Sun et al., 2010). Secondly, numerous other studies have identified mammary cells that co-express basal and luminal keratins in adult tissue homeostasis. In ductal cells(Su et al., 1996) and terminal duct lobular units (TDLUs) of the human breast (equivalent to the secretory alveoli in the mouse), luminal KRT8/18$^+$ cells have previously been shown to express basal KRT5/6 and KRT14 (Gusterson et al., 2005; Malzahn et al., 1998; Nagle et al., 1986; Otterbach et al., 2001; Werner and Horst, 2003; Wetzels et al., 1991). Furthermore, human luminal progenitors have been shown to contain a substantial population of cells co-expressing luminal KRT18 and basal KRT5/6 which are also KIT$^+$ (Lim et al., 2009). Our analysis of separated cell populations confirmed that luminal c-Kit$^+$ progenitor cells in the adult mouse mammary epithelium also co-express KRT18 and KRT14 (Regan et al., 2012), although the latter is at low levels and we have not been able to observe this in sections of the normal mammary gland by routine IHC staining. Co-expression of proteins associated with different lineages in the normal mammary gland and in breast cancer has been suggested to indicate plasticity (Petersen and Polyak, 2010; Spike et al., 2012). Overall, these data suggest that proliferative capacity and phenotypic plasticity in the mammary epithelium are tightly linked. Plasticity in the expression of keratins and other genes within c-Kit$^+$ luminal progenitors may relate to their potential to contribute to multiple cell lineages during epithelial remodeling, e.g. at involution of the mammary gland after weaning (Wagner et al., 2002). In addition, the phenotypic plasticity and multilineage differentiation potential of these luminal progenitors is consistent with their ability to give rise to tumors with basal features (Lim et al., 2009; Molyneux et al., 2010) as well as lineage switching in response to oncogene activation (Koren et al., 2015; Van Keymeulen et al., 2017).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

There are currently two opposing concepts as to whether homeostasis and development in the mammary gland is maintained by bipotent MaSCs (Regan et al., 2013; Rios et al., 2014) or lineage-restricted basal and luminal cells (Van Keymeulen et al., 2011; Van Keymeulen et al., 2017). Considering these two models, Nguyen *et al*. performed pseudotemporal reconstruction based lineage hierarchy analysis. This analysis identified a continuous lineage connecting the basal lineage, via a bipotent MaSC, to the two luminal branches. These results agree with previous models of mammary differentiation wherein a bipotent basal MaSC generates daughter cells that differentiate into myoepithelial and luminal cell lineages (Regan et al., 2012; Regan et al., 2013; Rios et al., 2014; Wang et al., 2014). However, Nguyen *et al*. propose that their results differ from previous studies in that L1.2 cells (luminal ER^{-} c-kit$^{+/Low}$ cells) are progenitors to L1.1 cells (luminal ER^{-} c-Kit$^{+/High}$ cells) and that c-Kit$^{+/High}$ L1.1 cells are another type of mature differentiated luminal cell rather than a luminal progenitor upstream of luminal ER^{+} L2 cells. Based on this pseudotemporal analysis the authors suggest that c-Kit is not a marker of luminal progenitor cells. This is a surprising conclusion, considering that Nguyen *et al.'s* L1.2 progenitor cells actually express c-Kit (Figure 1). c-Kit has been demonstrated to be a defining marker of the gene expression signatures of mouse and human luminal progenitor cells (Kendrick et al., 2008; Lim et al., 2009; Lim et al., 2010). Furthermore, and in contrast to Nguyen *et al*., we functionally tested c-Kit^{-}, c-Kit$^{+/Low}$ and c-Kit$^{+/High}$ luminal cells for progenitor properties. Luminal c-Kit$^{+/Low}$ (L1.2) and c-Kit$^{+/High}$ (L1.1) cells had high colony forming and proliferative capacity *in vitro* and even repopulated cleared mammary fat pads *in vivo*, thus demonstrating the defining functional features of stem/progenitor cells. c-Kit^{-} luminal cells, on the other hand, were found to be differentiated. These studies clearly demonstrated that c-Kit is a marker of luminal stem/progenitor cells and that luminal cells do indeed contain a population with 'stem cell potential' as defined by the cleared fat pad transplantation assay.

In addition to identifying two ER^{-} (L1) progenitor subtypes (c-Kit$^{+/High}$ and c-Kit$^{+/Low}$), we also identified three luminal ER^{+} (L2) cell types (CD24$^{+/Low}$ Sca-1^{+} c-kit^{-} and c-kit^{+} and CD24$^{+/Low}$ Sca-1^{-} c-Kit^{-}), which

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

were not reported by Nguyen *et al*. Our data demonstrated that luminal ER$^+$ Sca-1$^+$ c-Kit$^+$ cells are hormone responsive ductal progenitor cells and that ER$^+$ Sca-1$^+$ c-Kit$^-$ cells are terminally differentiated. Luminal ER$^+$ Sca-1$^-$ c-Kit$^-$ cells clustered with the luminal ER$^-$ c-Kit$^+$ cells but also expressed genes in common with the ER$^+$ Sca-1$^+$ cells, such as *Prom1* and *Prlr*. We therefore proposed that they may represent an alternative intermediate route for generating ER$^+$ cells from luminal ER$^-$ c-Kit$^+$ progenitors, rather than from the luminal ER$^+$ c-Kit$^+$ progenitor population. A potential mammary epithelial cell hierarchy based on these data and integrating the scRNA-seq analysis from Nguyen *et al*. as well as the lineage-primed cell types reported by Pal *et al*. (Pal et al., 2017) (discussed further below) is shown in Figure 2.

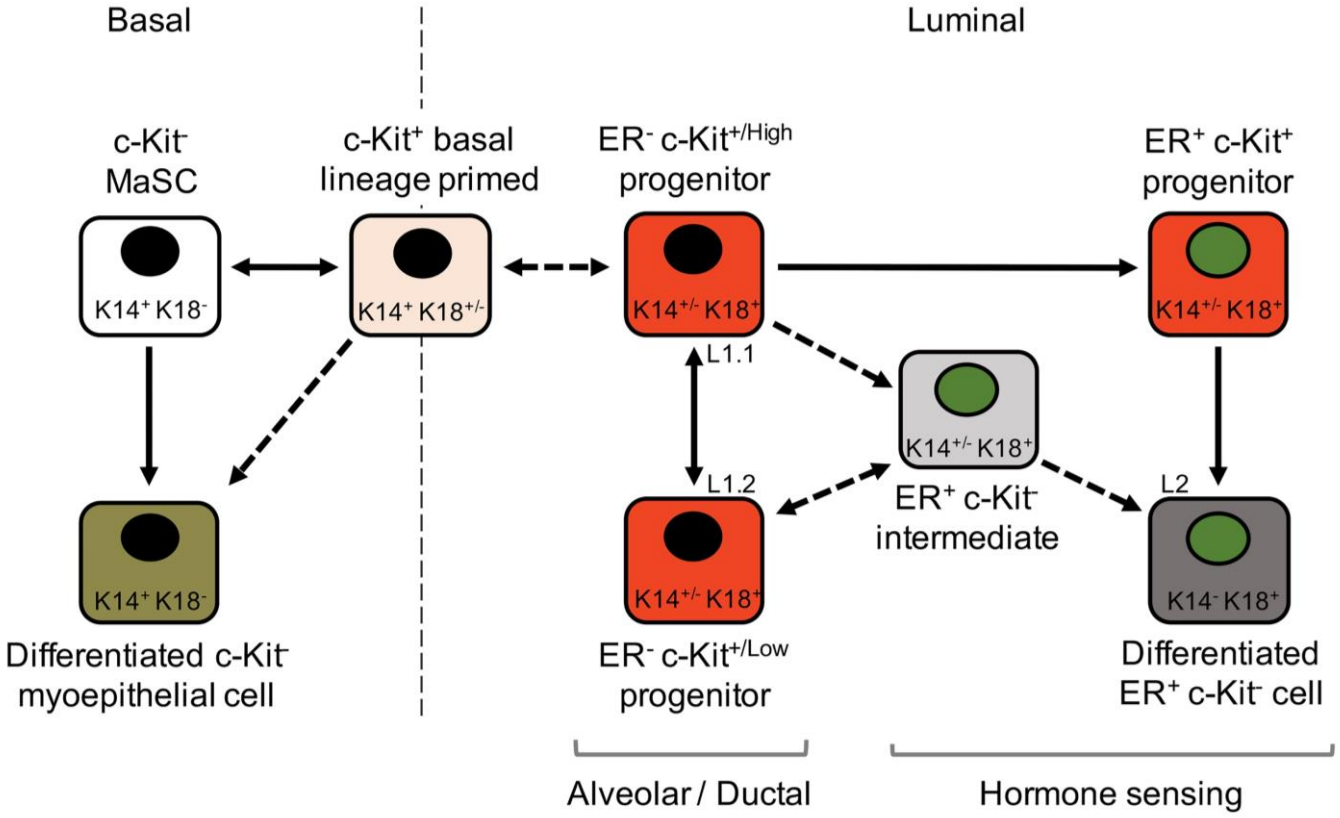

Figure 2: Proposed model of the mammary epithelial lineage cell hierarchy in the post-natal gland. Bipotent basal MaSCs generate both luminal and basal/myoepithelial cell lineages. During homeostasis, each lineage is maintained by its own c-Kit-expressing progenitor cell subpopulation. Lineage-primed c-Kit$^+$ basal cells that express intermediate levels of luminal genes may represent a transient or intermediate population that precedes commitment to the luminal lineage (Regan et al., 2012; Pal et al., 2017). Luminal

ER$^-$ c-Kit$^+$ progenitor cells give rise to secretory alveolar cells and ER$^+$ hormone sensing ductal cells. Gene expression analysis suggests that an alternative route for generating ER$^+$ cells from intermediate luminal cells may also exist.

Similar to Nguyen *et al*., Pal *et al*. (2017) used scRNA-seq to identify lineage relationships in the mouse mammary gland and also demonstrated that bipotent basal MaSCs give rise to basal and luminal lineages (Pal et al., 2017). Supporting our previous assessment of intermediate cells in the luminal lineage (Regan et al., 2012), the authors also described the identification of intermediate luminal cells. Significantly, Pal *et al*. report the identification of rare mixed-lineage or "lineage-primed" c-Kit-expressing basal cells in the adult mammary gland and state, "It is presumed that these cells represent a transient population that is poised for commitment to the luminal lineage, reminiscent of 'lineage-primed' stem and progenitor cells initially reported in the hematopoietic system." These lineage-primed c-Kit$^+$ basal cells comprised approximately 5% of the basal compartment and expressed luminal genes such as *Esr1, Prlr, Csn2* and *Areg* in addition to basal genes. Pal *et al*. state, "these data suggest that the basal state may precede commitment to a luminal cell fate in the post-natal mammary gland."

In our 2012 study we also identified cells that we described as lineage-primed basal cells (CD24$^{+/Low}$ Sca-1$^-$ CD49f$^{+/High}$ c-Kit$^+$) in the adult mammary gland that expressed luminal genes, including those described by Pal *et al*. (*Esr1, Prlr, Csn2, Areg*), but that clustered with the basal MaSCs (Regan et al., 2012). Furthermore, we functionally tested these cells by single cell cleared mammary fat pad transplantation and demonstrated that they can reconstitute an entire ductal tree, although at a lower frequency than c-Kit$^-$ MaSCs. Based on these data we came to the same conclusion as Pal *et al*. (2017) and described these c-Kit$^+$ basal cells as intermediate MaSCs that were undergoing "lineage priming", in which stem cells express genes associated with their differentiated daughter populations (Huang et al., 2007; Månsson et

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2019 doi:10.20944/preprints201906.0128.v1

al., 2007). This was the first time that lineage-primed basal cells in the adult mammary gland had been reported and functionally tested.

In contrast, in recent work from Bach *et al*. (2017)(Bach et al., 2017) in which scRNA-seq was performed on mouse mammary epithelial cells at four developmental stages of nulliparous, mid gestation, lactation and post involution, no mixed-lineage population with both basal and luminal markers was detected. Furthermore, Bach *et al.* (2017) found that, rather than clearly defined clusters maintained by their own stem / progenitor population, a continuous spectrum of differentiation exists in which a common luminal progenitor cell, which notably expressed c-Kit at high levels, gives rise to intermediate, restricted alveolar, and hormone-sensitive progenitors. However, this highly relevant study from Bach *et al*. was also not discussed or referenced in the subsequent Nguyen *et al*. paper, despite appearing in the same journal five months previously.

Concluding Remarks

Taken together, therefore, the weight of evidence supports c-Kit as a stem/progenitor marker in the mammary epithelium and, more importantly, one which is functionally characterized and can be used to isolate stem/progenitor cells. Indeed, we have already begun to understand the signaling pathways downstream of c-Kit in mammary stem/progenitor cells (Tornillo et al., 2018). scRNA-seq studies, which allow for comprehensive and unbiased analysis of the different cell types that constitute a heterogeneous tissue (Cristea and Polyak, 2018), have been extremely valuable in contributing to this evidence. However, in order to fully understand the significance of these studies it is essential to link them to functional data, in particular where such data already exists, and future studies should aim to do so. This will avoid confusion as to the state of knowledge in the field and prevent the misreporting of previously identified cell types as new.

References

 doi:10.20944/preprints201906.0128.v1

Asselin-Labat, M. L., Sutherland, K. D., Barker, H., Thomas, R., Shackleton, M., Forrest, N. C., Hartley, L., Robb, L., Grosveld, F. G., van der Wees, J., et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9**, 201–209.

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

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

Davis, F. M., Lloyd-Lewis, B., Harris, O. B., Kozar, S., Winton, D. J., Muresan, L. and Watson, C. J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nat. Commun.* **7**, 13053.

DeOme, K. B., Faulkin Jr., L. J., Bern, H. A. and Blair, P. B. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res.* **19**, 515–520.

Grigoriadis, A., Mackay, A., Reis-Filho, J. S., Steele, D., Iseli, C., Stevenson, B. J., Jongeneel, C. V, Valgeirsson, H., Fenwick, K., Iravani, M., et al. (2006). Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data. *Breast Cancer Res* **8**, R56.

Gusterson, B. A., Ross, D. T., Heath, V. J. and Stein, T. (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res* **7**, 143–148.

Huang, S., Guo, Y. P., May, G. and Enver, T. (2007). Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev Biol* **305**, 695–713.

Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J. S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., et al. (2004). Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res* **64**, 3037–3045.

Kendrick, H., Regan, J. L., Magnay, F.-A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M. and Smalley, M. J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**, 591.

Koren, S., Reavie, L., Couto, J. P., De Silva, D., Stadler, M. B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., et al. (2015). PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114.

Lim, E., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., Hart, A. H., Asselin-Labat, M. L., Gyorki, D. E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907–913.

Lim, E., Wu, D., Pal, B., Bouras, T., Asselin-Labat, M.-L., Vaillant, F., Yagita, H., Lindeman, G. J., Smyth, G. K. and Visvader, J. E. (2010). Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21–R21.

Lloyd-Lewis, B., Davis, F. M., Harris, O. B., Hitchcock, J. R. and Watson, C. J. (2018). Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. *Development* **145**, dev164079.

Malzahn, K., Mitze, M., Thoenes, M. and Moll, R. (1998). Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. *Virchows Arch.* **433**, 119–129.

Månsson, R., Hultquist, A., Luc, S., Yang, L., Anderson, K., Kharazi, S., Al-Hashmi, S., Liuba, K., Thorén, L., Adolfsson, J., et al. (2007). Molecular Evidence for Hierarchical

Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors. *Immunity* **26**, 407–419.

Molyneux, G., Geyer, F. C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., MacKay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. *Cell Stem Cell* **7**, 403–417.

Nagle, R. B., Böcker, W., Davis, J. R., Heid, H. W., Kaufmann, M., Lucas, D. O. and Jarasch, E. D. (1986). Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J. Histochem. Cytochem.* **34**, 869–881.

Nguyen, Q. H., Pervolarakis, N., Blake, K., Ma, D., Davis, R. T., James, N., Phung, A. T., Willey, E., Kumar, R., Jabart, E., et al. (2018). Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat. Commun.* **9**, 2028.

Otterbach, F., Bànkfalvi, À., Bergner, S., Decker, T., Krech, R. and Boecker, W. (2001). Cytokeratin 5/6 immunohistochemistry assists the differential diagnosis of atypical proliferations of the breast. *Histopathology* **37**, 232–240.

Pal, B., Chen, Y., Bert, A., Hu, Y., Sheridan, J. M., Beck, T., Shi, W., Satterley, K., Jamieson, P., Goodall, G. J., et al. (2015). Integration of microRNA signatures of distinct mammary epithelial cell types with their gene expression and epigenetic portraits. *Breast Cancer Res.* **17**, 85.

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A. C., Wilcox, S., Fu, N., Liu, K. H., Jackling, F. C., et al. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat. Commun.* **8**, 1627.

Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P. G. and Di Fiore, P. P. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62–73.

Petersen, O. W. and Polyak, K. (2010). Stem Cells in the Human Breast. *Cold Spring Harb. Perspect. Biol.* **2**, a003160.

Raouf, A., Zhao, Y., To, K., Stingl, J., Delaney, A., Barbara, M., Iscove, N., Jones, S., McKinney, S., Emerman, J., et al. (2008). Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* **3**, 109–118.

Regan, J. and Smalley, M. (2007). Prospective isolation and functional analysis of stem and differentiated cells from the mouse mammary gland. *Stem Cell Rev.* **3**, 124–136.

Regan, J. L., Kendrick, H., Magnay, F.-A., Vafaizadeh, V., Groner, B. and Smalley, M. J. (2012). c-Kit is required for growth and survival of the cells of origin of Brca1-mutation-associated breast cancer. *Oncogene* **31**, 869.

Regan, J. L., Sourisseau, T., Soady, K., Kendrick, H., McCarthy, A., Tang, C., Brennan, K., Linardopoulos, S., White, D. E. and Smalley, M. J. (2013). Aurora A Kinase Regulates Mammary Epithelial Cell Fate by Determining Mitotic Spindle Orientation in a Notch-Dependent Manner. *Cell Rep.* **4**, 110–123.

Rios, A. C., Fu, N. Y., Lindeman, G. J. and Visvader, J. E. (2014). In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322.

Scheele, C. L. G. J., Hannezo, E., Muraro, M. J., Zomer, A., Langedijk, N. S. M., van Oudenaarden, A., Simons, B. D. and van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature* **542**, 313–317.

Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M. L., Wu, L., Lindeman, G. J. and Visvader, J. E. (2006). Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88.

Shehata, M., Teschendorff, A., Sharp, G., Novcic, N., Russell, I. A., Avril, S., Prater, M., Eirew, P., Caldas, C., Watson, C. J., et al. (2012). Phenotypic and functional characterisation

of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **14**, R134.

Shipitsin, M., Campbell, L. L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., Nikolskaya, T., Serebryiskaya, T., Beroukhim, R., Hu, M., et al. (2007). Molecular definition of breast tumor heterogeneity. *Cancer Cell* **11**, 259–273.

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

Sleeman, K. E., Kendrick, H., Robertson, D., Isacke, C. M., Ashworth, A. and Smalley, M. J. (2007). Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *J Cell Biol* **176**, 19–26.

Spike, B. T., Engle, D. D., Lin, J. C., Cheung, S. K., La, J. and Wahl, G. M. (2012). A Mammary Stem Cell Population Identified and Characterized in Late Embryogenesis Reveals Similarities to Human Breast Cancer. *Cell Stem Cell* **10**, 183–197.

Stingl, J., Eaves, C. J., Zandieh, I. and Emerman, J. T. (2001). Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res. Treat.* **67**, 93–109.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H. I. and Eaves, C. J. (2006). Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997.

Su, L., Morgan, P. R. and Lane, E. B. (1996). Expression of cytokeratin messenger RNA versus protein in the normal mammary gland and in breast cancer. *Hum. Pathol.* **27**, 800–806.

Sun, P., Yuan, Y., Li, A., Li, B. and Dai, X. (2010). Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochem. Cell Biol.* **133**, 213–221.

Taddei, I., Deugnier, M. A., Faraldo, M. M., Petit, V., Bouvard, D., Medina, D., Fassler, R., Thiery, J. P. and Glukhova, M. A. (2008). Beta1 integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nat Cell Biol* **10**, 716–722.

Tornillo, G., Knowlson, C., Kendrick, H., Cooke, J., Mirza, H., Aurrekoetxea-Rodríguez, I., Vivanco, M. d. M., Buckley, N. E., Grigoriadis, A. and Smalley, M. J. (2018). Dual Mechanisms of LYN Kinase Dysregulation Drive Aggressive Behavior in Breast Cancer Cells. *Cell Rep.* **25**, 3674-3692.e10.

van Amerongen, R. (2015). Lineage Tracing in the Mammary Gland Using Cre/lox Technology and Fluorescent Reporter Alleles. In: Vivanco M. (eds) Mammary Stem Cells. Methods in Molecular Biology. In (ed. Vivanco, M. del M.), pp. 187–211. New York, NY: Springer New York.

van Amerongen, R., Bowman, A. N. and Nusse, R. (2012). Developmental Stage and Time Dictate the Fate of Wnt/β-Catenin-Responsive Stem Cells in the Mammary Gland. *Cell Stem Cell* **11**, 387–400.

van de Moosdijk, A. A. A., Fu, N. Y., Rios, A. C., Visvader, J. E. and van Amerongen, R. (2017). Lineage Tracing of Mammary Stem and Progenitor Cells. In: Martin F., Stein T., Howlin J. (eds) Mammary Gland Development. Methods in Molecular Biology, vol 1501. Humana Press, New York, NY. In (ed. Martin, F.), Stein, T.), and Howlin, J.), pp. 291–308. New York, NY: Springer New York.

Van Keymeulen, A., Rocha, A. S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S. and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189.

Van Keymeulen, A., Fioramonti, M., Centonze, A., Bouvencourt, G., Achouri, Y. and Blanpain, C. (2017). Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. *Cell Rep.* **20**, 1525–1532.

 doi:10.20944/preprints201906.0128.v1

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

Wagner, K. U., Boulanger, C. A., Henry, M. D., Sgagias, M., Hennighausen, L. and Smith, G. H. (2002). An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* **129**, 1377–1386.

Wang, D., Cai, C., Dong, X., Yu, Q. C., Zhang, X.-O., Yang, L. and Zeng, Y. A. (2014). Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81.

Wansbury, O., Mackay, A., Kogata, N., Mitsopoulos, C., Kendrick, H., Davidson, K., Ruhrberg, C., Reis-Filho, J. S., Smalley, M. J., Zvelebil, M., et al. (2011). Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. *Breast Cancer Res.* **13**, R79–R79.

Werner, B. and Horst, B. (2003). Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept. *Cell Prolif.* **36**, 73–84.

Wetzels, R. H., Kuijpers, H. J., Lane, E. B., Leigh, I. M., Troyanovsky, S. M., Holland, R., van Haelst, U. J. and Ramaekers, F. C. (1991). Basal cell-specific and hyperproliferation-related keratins in human breast cancer. *Am. J. Pathol.* **138**, 751–763.

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., Van Keymeulen, A., et al. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat. Cell Biol.* **20**, 666–676.

Author Information

Affiliations

Charité Universitätsmedizin Berlin, Charitéplatz 1, Berlin, Germany
Joseph L. Regan

European Cancer Stem Cell Research Institute, School of Biosciences, Cardiff University, Wales, UK
Matthew J. Smalley

Contributions

J.L.R. wrote the manuscript. M.J.S. revised and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Corresponding author

Correspondence to Joseph L. Regan (joseph.regan@charite.de).
ORCID iD: https://orcid.org/0000-0001-9987-7942