Oncogene (2012) 31, 869–883

www.nature.com/onc

ORIGINAL ARTICLE

c-Kit is required for growth and survival of the cells of origin of *Brca1*-mutation-associated breast cancer

JL Regan[1], H Kendrick[1], F-A Magnay[1], V Vafaizadeh[2], B Groner[2] and MJ Smalley[1]

[1]*Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK and* [2]*Georg-Speyer-Haus, Institute for Biomedical Research, Frankfurt, Germany*

***BRCA1* mutation-associated breast cancer originates in oestrogen receptor-alpha-negative (ER^-) progenitors in the mammary luminal epithelium. These cells also express high levels of the *Kit* gene and a recent study demonstrated a correlation between *Brca1* loss and *Kit* over-expression in the mammary epithelium. However, the functional significance of c-Kit expression in the mammary gland is unknown. To address this, $c\text{-}Kit^-$ and $c\text{-}Kit^+$ mammary epithelial subsets were isolated by flow cytometry, characterised for expression of lineage-specific cell markers and functionally analysed by *in vitro* colony forming and *in vivo* transplantation assays. The results confirm that the majority of luminal ER^- progenitors are $c\text{-}Kit^+$, but also that most stem cells and the differentiated cell populations are $c\text{-}Kit^-$. A subset of $c\text{-}Kit^+$ cells with high proliferative potential was found in the luminal ER^+ population, however, suggesting the existence of a distinct luminal ER^+ progenitor cell type. Analysis of mouse *Brca1* mammary tumours demonstrated that they expressed *Kit* and its downstream effector *Lyn* at levels comparable to the most strongly $c\text{-}Kit^+$ luminal ER^- progenitors. Consistent with c-Kit being a progenitor cell marker, *in vitro* three-dimensional differentiation of $c\text{-}Kit^+$ cells resulted in a loss of c-Kit expression, whereas c-Kit over-expression prevented normal differentiation *in vivo*. Furthermore, c-Kit was a functional marker of proliferative potential, as c-Kit inhibition by short hairpin knockdown prevented normal epithelial growth and caused cells to undergo apoptosis. Therefore, c-Kit defines distinct progenitor populations in the mammary epithelium and is critical for mammary progenitor survival and proliferation. Importantly, c-Kit is only the second mammary epithelial stem/progenitor marker to be shown to have a functional role in the mammary epithelium and the first marker to be shown to be required for progenitor cell function. The c-Kit signalling network has potential as a target for therapy and/or prevention in *BRCA1*-associated breast cancer.**

Oncogene (2012) **31,** 869–883; doi:10.1038/onc.2011.289; published online 18 July 2011

Keywords: *Brca1*; c-Kit; mammary; progenitor

Correspondence: Dr MJ Smalley, Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.
E-mail: matthew.smalley@icr.ac.uk
Received 20 January 2011; revised and accepted 12 June 2011; published online 18 July 2011

Introduction

The identification and molecular characterisation of stem, progenitor and differentiated cell hierarchies in complex tissues is key to understanding not only normal cellular homoeostasis but also the origins of cancer (Molyneux *et al*., 2010). Identifying the key biological processes common to tumours and the cells from which they develop will help define the pathways/networks required for tumour initiation and maintenance. Such pathways are potential targets for therapeutic and preventative intervention.

In the adult breast, the epithelium is organised into two major cell types defined by their spatial location, the luminal and basal epithelial cells. However, mammary epithelial cells can also be defined functionally. Stem cells regenerate the epithelium with high efficiency when injected into a 'cleared' mammary fat pad (one free of endogenous epithelium). Progenitors are cells derived from stem cells that have high *in vitro* colony forming and proliferative potential, but limited ability to repopulate a cleared fat pad. Terminally differentiated cells do not transplant or form colonies *in vitro* (Shackleton *et al*., 2006; Stingl *et al*., 2006; Asselin-Labat *et al*., 2007; Sleeman *et al*., 2007). Using flow cytometry and functional assays, stem cells have been localised to the basal cell layer, whereas progenitors are found in large numbers in the luminal population (Shackleton *et al*., 2006; Stingl *et al*., 2006; Asselin-Labat *et al*., 2007; Sleeman *et al*., 2007; Taddei *et al*., 2008). We have recently demonstrated that oestrogen receptor-alpha-negative (ER^-) luminal progenitors are the origin of *BRCA1*-associated breast cancers and likely of the majority of 'basal-like' breast cancers not linked to a family history (Molyneux *et al*., 2010). Understanding the molecular regulation of these cells in particular is therefore important for identifying novel approaches to breast cancer prevention and therapy.

Gene expression analysis showed that *Kit* was highly expressed by luminal ER^- progenitors in both mouse (Kendrick *et al*., 2008) and human (Lim *et al*., 2009). Furthermore, *Brca1* loss in mouse mammary epithelial cells resulted in defective differentiation and increased c-Kit expression (Smart *et al*., 2011). However, a role for c-Kit in directly regulating mammary epithelial differentiation could not be determined. Therefore, the function of c-Kit in the mammary epithelium remains unclear.

We have analysed c-Kit-expressing subpopulations within the mammary epithelial stem, progenitor and differentiated cells. We demonstrate that the luminal ER$^-$ progenitors, the cells of origin of *Brca1* mutation-associated breast cancer, are strongly c-Kit$^+$. Expression of c-Kit is also seen in a subpopulation of luminal ER$^+$ cells. Using functional assays, we show that c-Kit expression marks progenitors with high *in vitro* proliferative potential, but that the majority of mammary stem cells (MaSCs) are c-Kit$^-$. We demonstrate that in a mouse model of *Brca1* breast cancer that phenocopies the human disease (Molyneux *et al*., 2010), the tumours express *Kit* and its downstream effector *Lyn* at levels comparable to those luminal ER$^-$ progenitors, which are most strongly c-Kit$^+$. Importantly, we demonstrate that c-Kit is required for mammary epithelial cell survival. This makes the c-Kit signalling network not only a key component in the regulation of normal mammary epithelial homoeostasis, but also a potential target for therapy and prevention in *BRCA1* mutation carriers and potentially other at-risk groups.

Results

The majority of c-Kit$^+$ mammary epithelial cells are luminal ER$^-$ progenitors

To identify c-Kit$^+$ populations and enable functional characterisation, freshly harvested primary mammary cells were stained for CD24, Sca-1, CD49f and c-Kit expression and analysed by flow cytometry. As previously, CD24$^{+/Low}$ basal epithelial cells and CD24$^{+/High}$ luminal epithelial cells were identified (Sleeman *et al*., 2006). Within the basal population, a CD24$^{+/Low}$ Sca-1$^-$ CD49f^{Low} population (previously identified as mainly myoepithelial cells) and a CD24$^{+/Low}$ Sca-1$^-$ CD49f^{High} population (previously shown to be highly enriched for mammary stem cells) could be recognised (Stingl *et al*., 2006; Sleeman *et al*., 2007). Within the luminal population, CD24$^{+/High}$ Sca-1$^-$ luminal cells (mainly ER$^-$ progenitors) and CD24$^{+/High}$ Sca-1$^+$ luminal cells (mainly differentiated ER$^+$ cells) could be distinguished (Figure 1a and Supplementary Figure 1) (Sleeman *et al*., 2007; Molyneux *et al*., 2010).

Consistent with the transcriptomic data (Kendrick *et al*., 2008), the majority (up to 90%) of luminal Sca-1$^-$ progenitors were c-Kit$^+$ (equal to 50% of all mammary epithelial cells). A small number of these (designated luminal Sca-1$^-$ c-Kit$^{+/High}$) had very high levels of c-Kit expression. In contrast, only approximately 20% of the luminal Sca-1$^+$ cells were c-Kit$^+$ (3% of the total epithelium). The majority of basal cells were c-Kit$^-$. However, occasional c-Kit$^+$ cells could be detected in both CD24$^{+/Low}$ Sca-1$^-$ CD49f^{Low} and CD24$^{+/Low}$ Sca-1$^-$ CD49f^{High} populations. It was unclear at this stage whether they represented minor contamination from the luminal c-Kit$^+$ populations, a biologically relevant group of cells or a sorting artefact. These rare cells were termed basal CD49f^{Low} c-Kit$^+$ and basal CD49f^{High} c-Kit$^+$ to distinguish them from the majority basal CD49f^{Low} c-Kit$^-$ myoepithelial and basal CD49f^{High} c-Kit$^-$ mammary stem cells, respectively (Figures 1a and b).

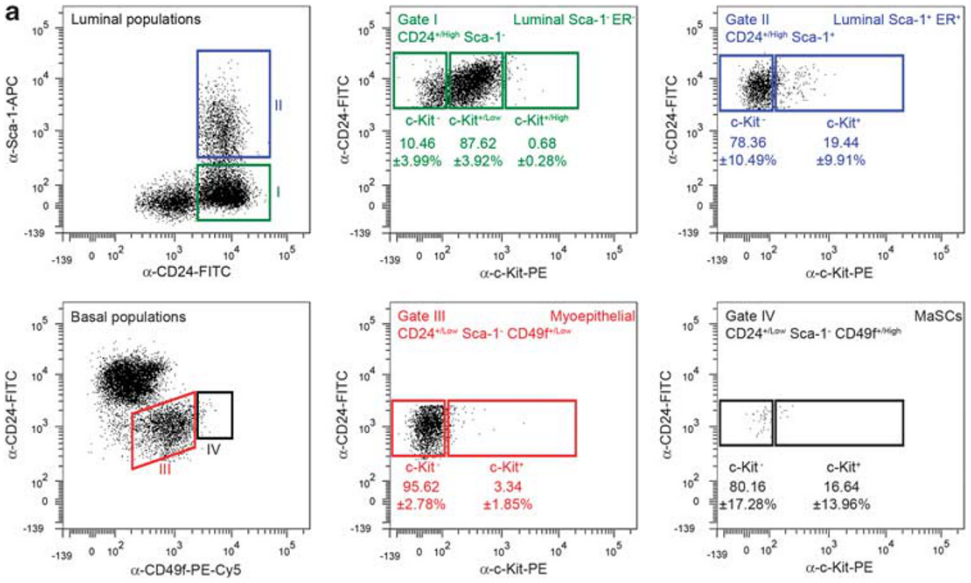
To confirm the flow cytometry data, analysis of *Kit* gene expression by quantitative real-time RT–PCR (qPCR) was carried out on each mammary epithelial subpopulation. *Kit* was undetectable in the basal myoepithelial cells and MaSCs, but its expression could be detected in the rare basal CD49f^{Low} c-Kit$^+$ and basal CD49f^{High} c-Kit$^+$ cells. As expected, *Kit* was detected at only very low levels (and in only one of three independent isolates) in the luminal Sca-1$^+$ c-Kit$^-$ population, but was robustly expressed in the luminal Sca-1$^+$ c-Kit$^+$ cells. Surprisingly, however, while *Kit* was strongly expressed in the luminal Sca-1$^-$ c-Kit$^{+/Low}$ and luminal Sca-1$^-$ c-Kit$^{+/High}$ populations, it was also detected in the luminal Sca-1$^-$ c-Kit$^-$ cells at levels comparable to the luminal Sca-1$^+$ c-Kit$^+$ population, suggesting post-transcriptional regulation.

In addition, levels of the c-Kit receptor ligand, *Kit ligand*/*Stem cell factor* (*Kitl*), were measured (Kent *et al*., 2008). *Kitl* expression could be detected in all cells; however, it was most strongly expressed by the basal populations. Given the mainly luminal pattern of c-Kit expression, this suggests that c-Kit signalling in the mammary epithelium is primarily driven by paracrine signals from the basal to the luminal cells (Figure 1c).

c-Kit$^+$ expression defines cells with high clonogenic potential and which express markers of multiple cell lineages

Previously, luminal CD24$^{+/High}$ Sca-1$^-$ progenitors were found to have the highest *in vitro* colony forming potential of any mammary epithelial subpopulation (Sleeman *et al*., 2007). To determine whether colony forming potential correlated with c-Kit$^+$ expression, the relative *in vitro* colony forming abilities of c-Kit$^+$ and c-Kit$^-$ cells from the mammary epithelial subpopulations were compared in a single cell per well 96-well plate assay. The results showed that in the two luminal populations, the c-Kit$^+$ cells had significantly higher ($P<0.01$) colony forming ability relative to the c-Kit$^-$ cells. Basal CD49f^{Low} c-Kit$^+$ cells also had a slight, but significant ($P<0.05$) higher colony forming ability relative to the other basal populations (Figure 2a). Given the scarcity of these cells, however, it was clear that the majority of *in vitro* colony-forming activity in the mammary epithelium was associated with luminal c-Kit$^+$ cells.

The high *in vitro* clonogenic potential of c-Kit$^+$ cells is consistent with a progenitor identity. Plasticity of luminal progenitor differentiation is thought to underlie the 'basal-like' gene expression profile observed in tumours derived from these cells (Molyneux *et al*., 2010). Therefore, we examined whether mammary epithelial populations showed evidence of differentiative plasticity by analysing expression of the basal marker Keratin 14 (K14), the luminal marker Keratin 18 (K18) and ER in freshly isolated subpopulations sorted as single cells on to slides. The results (Table 1)



b

Luminal Populations

	CD24+/High Sca-1⁻			CD24+/High Sca-1⁺	
	c-Kit⁻	c-Kit+/Low	c-Kit+/High	c-Kit⁻	c-Kit⁺
Mean frequency (% of total epithelium±SD)	5.97±2.18	50.51±5.61	0.40±0.19	13.85±4.83	3.34±1.71

Basal Populations

	CD24+/Low Sca-1⁻ CD49fLow		CD24+/Low Sca-1⁻ CD49fHigh	
	c-Kit⁻ (Myoepithelial)	c-Kit⁺	c-Kit⁻ (MaSCs)	c-Kit⁺
Mean frequency (% of total epithelium±SD)	20.59±6.12	0.70±0.43	0.36±0.14	0.07±0.06

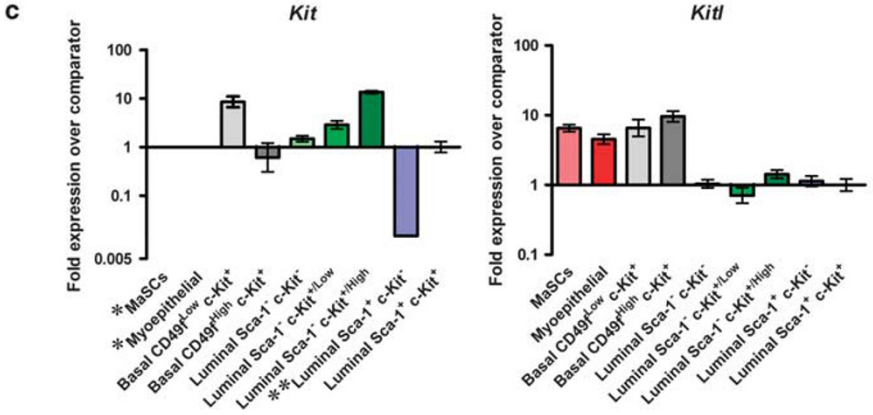


Figure 1 c-Kit is predominantly expressed by luminal Sca-1$^-$ cells in the mammary epithelium. (**a**) Flow cytometry dot plots of primary mouse mammary epithelial cells isolated from 10 week virgin FVB mice and stained with antibodies against CD24, Sca-1, CD49f and c-Kit. Only the epithelial cells are shown (for the full gating strategy used to define the epithelial-only compartment see Supplementary Figure 1). In the upper series of plots, the c-Kit status of the luminal subpopulations was determined by first plotting CD24 against Sca-1 staining (left hand plot), to define the CD24$^{+/High}$ Sca-1$^-$ (Gate I; green) and CD24$^{+/High}$ Sca-1$^+$ (Gate II; blue) luminal cells. Dot plots of c-Kit against CD24 staining were then drawn for each of these gated populations (middle and right hand plots). The numbers indicate the proportion of c-Kit$^-$ or c-Kit$^+$ cells as a percentage of the total cells in that gate (± s.d.; n = five independent sorts). In the lower series of plots, CD49f was plotted against CD24 (data from the same samples) to similarly define the CD24$^{+/Low}$ Sca-1$^-$ CD49f^{Low} myoepithelial cells (Gate III; red) and CD24$^{+/Low}$ Sca-1$^-$ CD49f^{High} mammary stem cell enriched population (MaSCs; Gate IV; black). c-Kit and CD24 dot plots were then drawn as before. (**b**) Table of frequencies of each population as a percentage of the total epithelium (± s.d.) calculated from percentages shown in (**a**). (**c**) Levels of *Kit* and *Kitl* gene expression in the different mammary epithelial subpopulations as determined by qPCR relative to comparator (the luminal Sca-1$^+$ c-Kit$^+$ population). Data are presented as fold expression levels ± 95% confidence intervals (n = three independently harvested isolates of each cell population). **Kit* gene expression was undetectable in these populations in all three independent isolates. ***Kit* gene expression was only detected (at very low levels) in one of three isolates of the luminal Sca-1$^+$ c-Kit$^-$ population. No error bars are therefore shown for this sample.

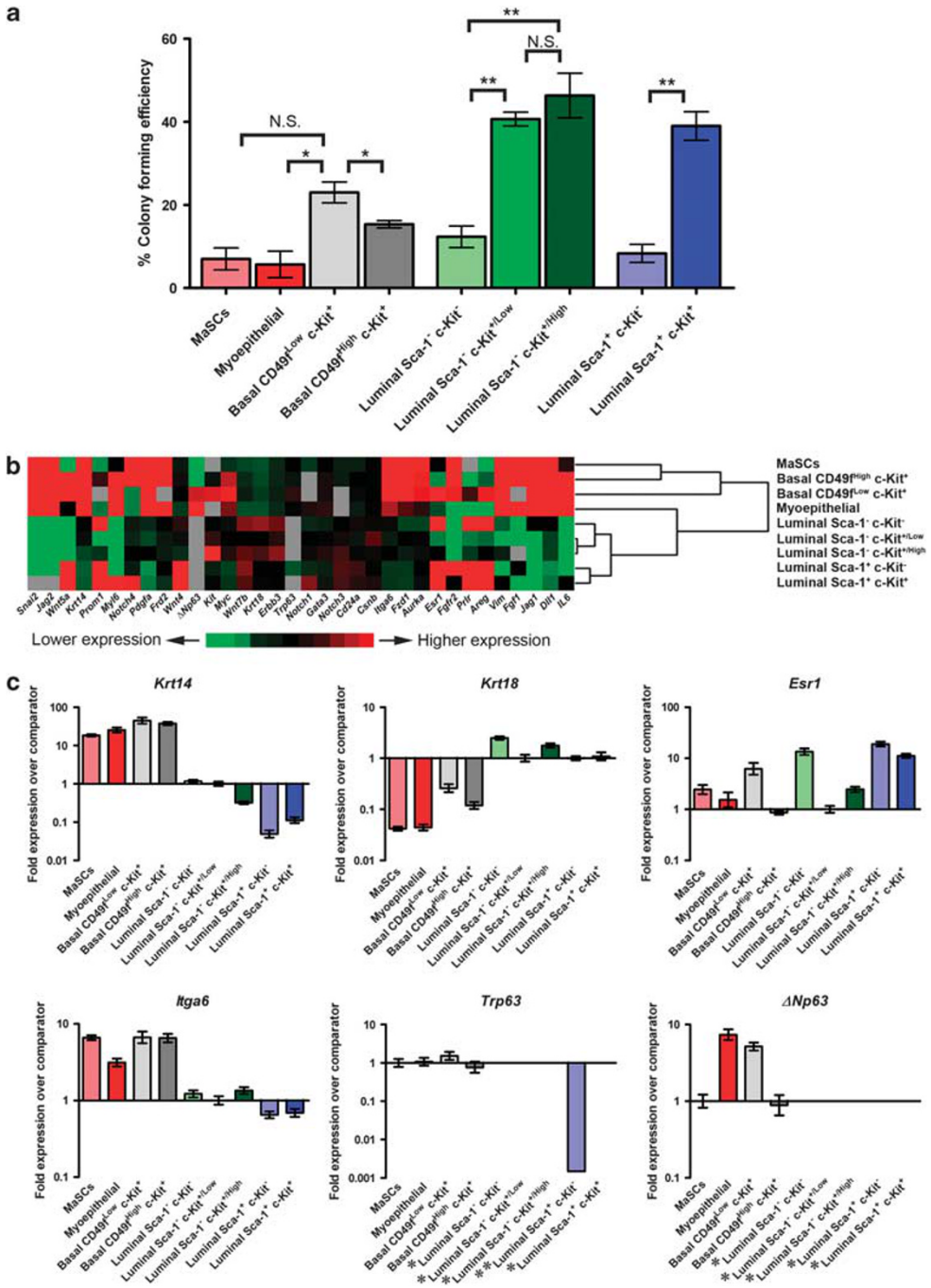


Figure 2 c-Kit expression defines cells with progenitor features. (**a**) Colony forming efficiency in single-cell cloning 96-well plate assays of mammary epithelial subpopulations separated by c-Kit status. Data are presented as mean percentage colony forming ability ± s.d. (n = three independent experiments each of which analysed three 96-well plates of each cell population). *$P<0.05$, **$P<0.01$ (t-tests on Log10-transformed data). (**b**) Heat map of qPCR gene expression data for mammary cell subpopulations. Relative gene expression levels ($n=3$ independent cell preparations) over the comparator population (luminal Sca-1$^-$ c-Kit$^{+/Low}$ cells for all genes except *Trp63* and *ΔNp63*, for which the MaSCs were the comparator) were used to determine relationships between the mammary epithelial subpopulations by unsupervised hierarchical clustering. All relative expression data together with 95% confidence intervals are given in Supplementary Table 1. Grey boxes indicate samples in which the gene was undetectable. (**c**) Fold expression ± 95% confidence intervals in mammary epithelial subpopulations ($n=3$ independent cell preparations) over the comparator population (as above) of genes associated with the basal (*Krt14*, *Itga6* and *Trp63*/*ΔNp63*) and luminal (*Krt18, Esr1*) lineages (Kendrick *et al*., 2008). *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. No error bars are therefore shown for this sample.

demonstrated that the c-Kit$^-$ and c-Kit$^+$ luminal Sca-1$^-$ cells, and the c-Kit$^+$ luminal Sca-1$^+$ cells, were weakly K14-positive (K14$^{+/-}$) as well as K18-positive (K18$^+$), suggesting potential for differentiative plasticity. The luminal Sca-1$^+$ c-Kit$^-$ cells, however, were K18$^+$ only, consistent with terminal differentiation. Luminal markers

Table 1 Staining of mammary epithelial subpopulations for Keratin 14 (K14), Keratin 18 (K18) and ER (data from three independent cell preparations)

Population	*c-Kit status*	*Antigen*	−[a]	−/+	+	*Summarised cell phenotype*
MaSCs	c-Kit$^-$	K14	0	0	613	
		K18	669[b]	0	0	K14$^+$ (ER$^-$ and K18$^-$)
		ER[c]	623		0	
Myoepithelial	c-Kit$^-$	K14	0	0	685	
		K18	697	0	0	K14$^+$ (K18$^-$ and ER$^-$)
		ER	653		0	
Basal CD49f^{Low}	c-Kit$^+$	K14	0	0	235	
		K18	298	0	0	K14$^+$ (K18$^-$ and ER$^-$)
		ER	292		0	
Basal CD49f^{High}	c-Kit$^+$	K14	0	0	297	
		K18	0	222	0	K14$^+$/K18$^{+/-}$ (ER$^-$)
		ER	267		0	
Luminal Sca-1$^-$	c-Kit$^-$	K14	0	286	0	
		K18	0	0	292	K14$^{-/+}$/K18$^+$/ER$^+$
		ER	0		270	
	c-Kit$^{+/Low}$	K14	0	663	0	
		K18	0	0	720	K14$^{-/+}$/K18$^+$ (ER$^-$)
		ER	651		0	
	c-Kit$^{+/High}$	K14	0	306	0	
		K18	0	0	313	K14$^{-/+}$/K18$^+$ (ER$^-$)
		ER	325		0	
Luminal Sca-1$^+$	c-Kit$^-$	K14	698	0	0	
		K18	0	0	677	K18$^+$/ER$^+$ (K14$^-$)
		ER	0		709	
	c-Kit$^+$	K14	0	296	0	
		K18	0	0	316	K14$^{-/+}$/K18$^+$/ER$^+$
		ER	0		280	

Abbreviations: ER, oestrogen receptor; MaSCs, mammary stem cells.
[a]− = no staining, −/+ = weak positive staining, + = positive staining.
[b]Numbers indicate the number of cells counted with the indicated intensity of staining.
[c]ER staining was defined as − = no nuclear staining, + = positive nuclear staining, with no intermediate grade.

could not be detected in the basal CD49f^{Low} c-Kit$^+$ cells but the basal CD49f^{High} c-Kit$^+$ cells were K14$^+$ and also K18$^{+/-}$. Both the myoepithelial cells and MaSCs were K14$^+$ but K18$^-$. Nuclear ER staining was found in both the luminal Sca-1$^+$ c-Kit$^+$ and luminal Sca-1$^+$ c-Kit$^-$ cells, supporting a progenitor—differentiated cell relationship between them. However, nuclear ER staining was also found in the luminal Sca-1$^-$ c-Kit$^-$ cells.

To investigate further the relationships between the populations, a qPCR-based gene expression analysis was carried out on a panel of 33 genes. The panel included genes previously shown to be specifically expressed in basal and luminal cells, as well as genes thought to be expressed uniformly throughout the mammary epithelium (Kendrick *et al*., 2008). Unsupervised hierarchical clustering of the data was carried out (Figure 2b; the full data set including relative expression levels and 95% confidence intervals for all genes in all populations is given in Supplementary Table 1; the expected luminal or basal expression pattern of each gene based on our previous transcriptomic and qPCR analysis is also shown) and patterns of expression of individual genes were analysed (Figure 2c). The clustering analysis divided the populations into three main branches. In one branch were the MaSCs together with the basal CD49f^{Low} c-Kit$^+$ and basal CD49f^{High} c-Kit$^+$ cells. This suggested that the rare basal c-Kit$^+$ cells, despite their phenotypic differences, were in fact closely related to the MaSCs. The luminal populations were clustered on a second branch, with the two luminal Sca-1$^+$ populations on one sub-branch and the three luminal Sca-1$^-$ populations on another. Finally, the myoepithelial cells were on a branch on their own.

Looking at individual genes in more detail, analysis of the expression patterns of *Krt14*, *Krt18* and *Esr1* largely supported the K14, K18 and ER staining on slides, although in populations which showed no staining for these markers, low levels of gene expression could always be detected, consistent with previous observations (Sleeman *et al*., 2007; Kendrick *et al*., 2008). As expected, the basal populations expressed the highest levels of *Krt14*, while the luminal populations expressed very low levels. Consistent with the staining of cells on slides, *Krt14* expression was lowest in the luminal Sca-1$^+$ c-Kit$^-$ cells, in which K14 staining could not be detected, with slightly higher levels in the luminal Sca-1$^+$ c-Kit$^+$ and luminal Sca-1$^-$ c-Kit$^-$, c-Kit$^{+/Low}$ and c-Kit$^{+/High}$ populations, which were weakly K14-positive. *Krt18* expression was highest in the luminal populations and lowest in the basal populations. *Esr1* gene expression was highest in the luminal Sca-1$^+$ c-Kit$^+$, luminal Sca-1$^+$ c-Kit$^-$ and the luminal Sca-1$^-$ c-Kit$^-$ cells, consistent with the ER staining. Remarkably, in the basal CD49f^{Low} c-Kit$^+$ cells *Esr1* was expressed at high levels, despite being undetected by staining.

Consistent with the clustering data, but in contrast with their flow cytometry profile, the basal CD49f^{Low} c-Kit$^+$ cells expressed high levels of *Itga6* (the gene coding for CD49f/α6 integrin) comparable to those of the MaSCs and basal CD49f^{High} c-Kit$^+$ cells. Furthermore, analysis of *Trp63* levels with two different probes, one of which detects both the full length and ΔN splice isoforms, and the other only the ΔN isoform, demonstrated that MaSCs express mainly full length *Trp63*, while the myoepithelial cells likely express mainly ΔNp63 (Figure 2c). The significance of this is unclear at this time. Overall, the gene expression patterns were consistent with the basal or luminal identities of the cell populations (the expression of *Esr1* in the CD49f^{Low} c-Kit$^+$ cells being one of the notable exceptions; Supplementary Table 1) (Kendrick *et al*., 2008).

These data demonstrate that the large majority of c-Kit$^+$ cells in the mammary epithelium are found in the luminal cell layers, and primarily in the luminal Sca-1$^-$ ER$^-$ populations. Furthermore, c-Kit$^+$ mammary epithelial cells have high proliferative capacity and show plasticity of developmental potential consistent with a progenitor phenotype.

c-Kit$^+$ mammary epithelial cells behave as progenitors in three-dimensional culture

In three-dimensional culture on Matrigel, mammary epithelial progenitors form small and simple acini with a single layer of epithelial cells. In contrast, stem cell-derived colonies on Matrigel are large and tend to be solid, complex structures (Stingl *et al*., 2006). To assess the response of c-Kit$^+$ and c-Kit$^-$ mammary epithelial cells to three-dimensional culture, freshly isolated populations were plated at low density, in equal numbers, on growth factor-reduced Matrigel. Within 24–48 h of being placed in culture, c-Kit$^+$ cells began to form colonies. In contrast, colony formation by c-Kit$^-$ cells was observed only after 4–5 days in culture. After 20 days in culture, c-Kit$^-$ cells had generated approximately five-fold fewer colonies than c-Kit$^+$ cells (Figure 3a). However, those colonies that did grow from the c-Kit$^-$ cells were significantly larger than colonies derived from c-Kit$^+$ cells, and tended to be solid multilayered structures. In contrast, the smaller c-Kit$^+$ cell-derived structures tended to have a typical differentiated acinus-like morphology, consisting of a single layer of cells around a lumen (Figures 3b–d). qPCR analysis of the c-Kit$^+$ cell-derived colonies demonstrated that they had significantly lower levels of *Kit* expression compared with their cells of origin. In contrast, c-Kit$^-$ cell-derived colonies had significantly upregulated levels of both *Kit* and *Kitl* gene expression compared with the freshly isolated parental cells (Figure 3e). Therefore, not only do c-Kit$^+$ cells behave as progenitors in three-dimensional culture, but a subset of c-Kit$^-$ cells form colonies resembling stem cell-derived colonies. Furthermore, the data suggest that differentiation of c-Kit$^+$ cells on Matrigel is associated with downregulation of c-Kit expression.

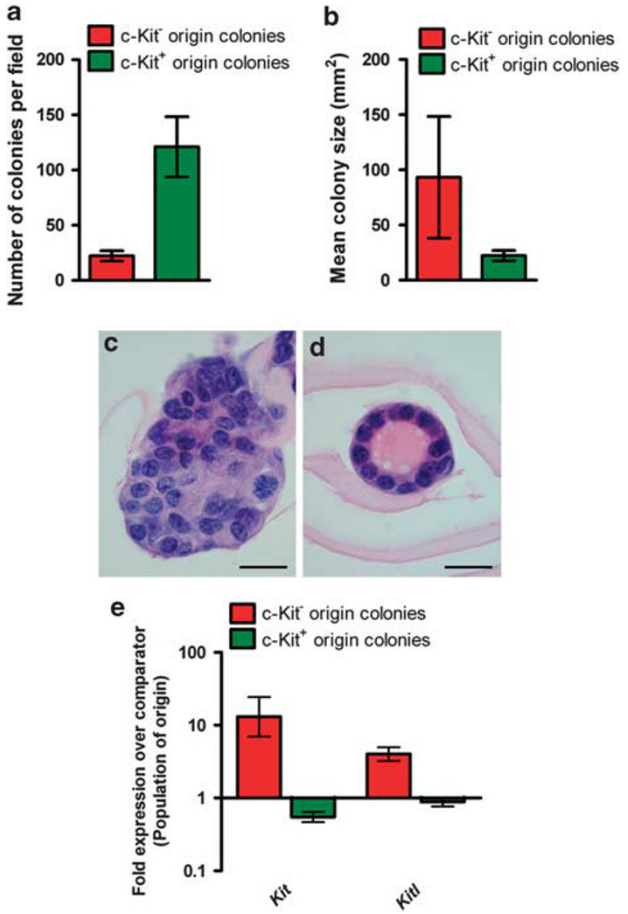


Figure 3 c-Kit$^+$ cells have a distinct colony morphology in three-dimensional culture**.** (**a**) Mean (± s.d.) number of colonies per field of view in Matrigel cultures of c-Kit$^-$ and c-Kit$^+$ mammary epithelial cells. Data from three independent cell preparations after 20 days in culture (each of which had three separate fields of view in each of three separate wells counted). Identical numbers of cells were plated per well in all cultures. (**b**) Mean (± s.d.) area of colonies in Matrigel cultures of c-Kit$^-$ and c-Kit$^+$ mammary epithelial cells. Data from three independent cell preparations after 20 days in culture (each of which had three separate fields of view in each of three separate wells analysed). (**c**, **d**) Haematoxylin and eosin sections of c-Kit$^-$ origin (**c**) and c-Kit$^+$ origin (**d**) colonies. Bar = 20 μm. (**e**) qPCR analysis of *Kit* and *Kitl* gene expression in c-Kit$^-$ and c-Kit$^+$ origin colonies compared with freshly isolated parental c-Kit$^-$ and c-Kit$^+$ cells. Data are mean fold expression levels over the parental populations ± 95% confidence intervals. Data from three independent cell preparations after 20 days in culture.

In vivo *transplantation potential is highest in c-Kit$^-$ basal cells*

Our data suggest a model in which mammary progenitors are c-Kit$^+$, while the c-Kit$^-$ fraction of the mammary epithelium includes both terminally differentiated cells and stem cells. To provide *in vivo* functional evidence for this model, we determined the transplantation potential of mammary epithelial subpopulations prospectively isolated on the basis of CD24 Sca-1 CD49f and c-Kit staining (Figures 4a–f).

Transplants of basal epithelial cells (CD24$^{+/Low}$ Sca-1$^-$) separated by c-Kit expression showed that the total basal c-Kit$^-$ fraction was highly enriched for

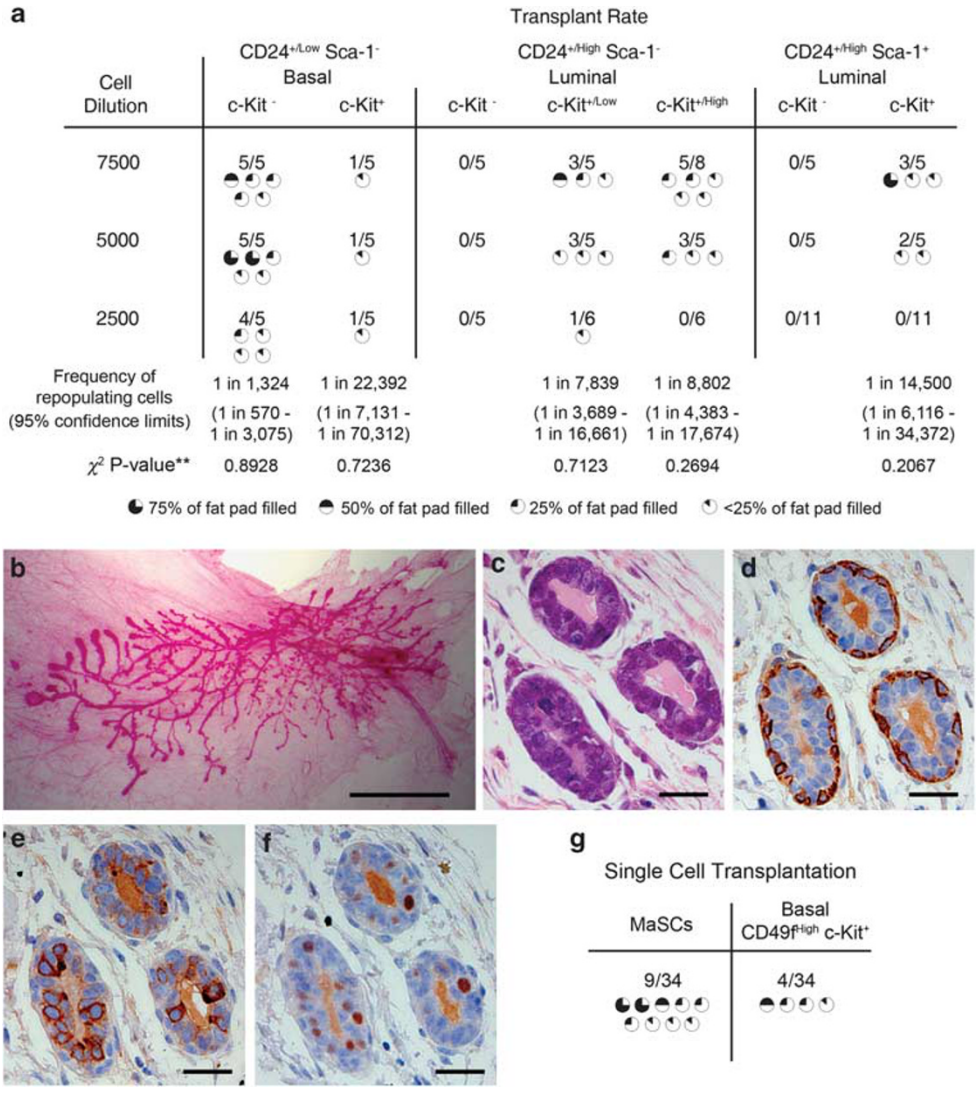


Figure 4 MaSCs are c-Kit$^-$. (**a**) Results of limited dilution transplant series for mammary epithelial cells separated on the basis of CD24, Sca-1 and c-Kit staining. The number of successful outgrowths as a fraction of the number of cleared fat pads transplanted is given. The extent to which each outgrowth filled the fat pad is indicated by the 'pie chart' symbols. The χ^2 statistic is a test for internal consistency in the data. Non-significant values indicate that all data sets are consistent with the Poisson distribution assumed to calculate the stem cell frequency. Data are from eight independent transplant sessions. (**b**) Carmine stained wholemount example of a transplanted mammary epithelial outgrowth. Bar = 3.5 mm. (**c**–**f**) Serial sections through epithelium from a successful transplant showing haematoxylin and eosin staining (**c**), K14 staining (**d**), K18 staining (**e**) and ER staining (**f**). All outgrowths, regardless of origin, contained basal, luminal ER$^-$ and luminal ER$^+$ cells. Bars = 25 μm. (**g**) Results of single cell transplants for MaSCs and basal CD49f^{High} c-Kit$^+$ cells. Data are from five independent transplant sessions.

transplantable cells, whereas the rare basal c-Kit$^+$ cells (which are mainly CD49f^{Low} c-Kit$^+$; Figure 1b) had poor transplantation activity. No transplants were observed from either the luminal Sca-1$^-$ c-Kit$^-$ or the luminal Sca-1$^+$ c-Kit$^-$ populations. In contrast, when c-Kit$^+$ cells from the luminal populations were transplanted, some outgrowths were observed (Figure 4a), especially when larger cell numbers were used. Importantly, histological analysis and immunohistochemical staining of sections through all outgrowths demonstrated that outgrowth formation was always associated with the ability to generate myoepithelial, luminal ER$^-$ and luminal ER$^+$ cells (Figures 4b–f).

Owing to their scarcity, basal CD49f^{High} c-Kit$^+$ cells could not be assessed by standard limiting dilution transplants. Therefore, to define the transplantation potential of these cells, MaSCs and basal CD49f^{High} c-Kit$^+$ cells were transplanted at one cell per fat pad into fat pads of syngeneic FVB mice. Of the 34 fat pads that received a single MaSC, 9 produced outgrowths, the highest enrichment of a mammary epithelial subpopulation for MaSCs reported to date. The basal c-Kit$^+$ transplants produced fewer outgrowths, 4 out of 34, but this difference was non-significant ($P > 0.05$ by χ^2-test) (Figure 4g). This contrasts with the poor transplantation ability of c-Kit$^+$ cells isolated from the total basal epithelial population (CD24$^{+/Low}$ Sca-1$^-$) without CD49f fractionation (Figure 4a).

These results demonstrate that basal MaSCs are c-Kit$^-$. However, rare basal CD49f^{High} c-Kit$^+$ cells that

are related to the MaSCs can be isolated (Figure 2b), which have good transplantation potential and which also express the luminal marker K18 at low levels. The basal CD49f^{Low} c-Kit^{+} cells, which are also apparently related to the MaSCs, have poor transplant potential and do not express K18. By contrast, the limited transplantation capacity which has been previously observed in luminal mammary epithelial cells (Sleeman *et al*., 2007) is a property of the luminal c-Kit^{+} fractions, supporting a progenitor function for these cells.

Brca1 mammary tumours have a progenitor-like phenotype

We have previously demonstrated that *Brca1* mutation-associated breast cancers, and the majority of sporadic 'basal-like' breast cancers, develop from luminal Sca-1^{-} ER^{-} progenitors. We also demonstrated that high expression of the Src-family kinase Lyn is associated with 'basal-like' tumours (Molyneux *et al*., 2010). As Lyn is a downstream effector of c-Kit, we examined its expression pattern in the normal mammary epithelial subpopulations we have defined by CD24 Sca-1 CD49f and c-Kit staining. *Lyn* expression could be detected in all populations, although it was detected most robustly and reproducibly in the luminal populations (Figure 5a). The relative expression levels of *Lyn* in the different luminal populations correlated with *Kit* expression levels (Figure 1c), while there was no such correlation in the basal populations.

Expression of both *Kit* and *Lyn* were next determined by qPCR in *Blg-Cre Brca1$^{flox/flox}$p53$^{+/-}$* mouse mammary tumours, which are derived from the luminal ER^{-} progenitors and phenocopy human *BRCA1* breast cancers (Molyneux *et al*., 2010). Both *Kit* and *Lyn* were expressed at levels comparable to those in the luminal Sca-1^{-} c-Kit$^{+/High}$ progenitors (Figure 5b) and staining of the tumours confirmed the presence of Lyn-positive neoplastic cells (data not shown). These findings were consistent with the progenitor origins of *Blg-Cre Brca1$^{flox/flox}$p53$^{+/-}$* mouse mammary tumours and support the hypothesis that these tumours have a strong mammary progenitor-like phenotype.

c-Kit activity is functionally required for progenitor growth and survival

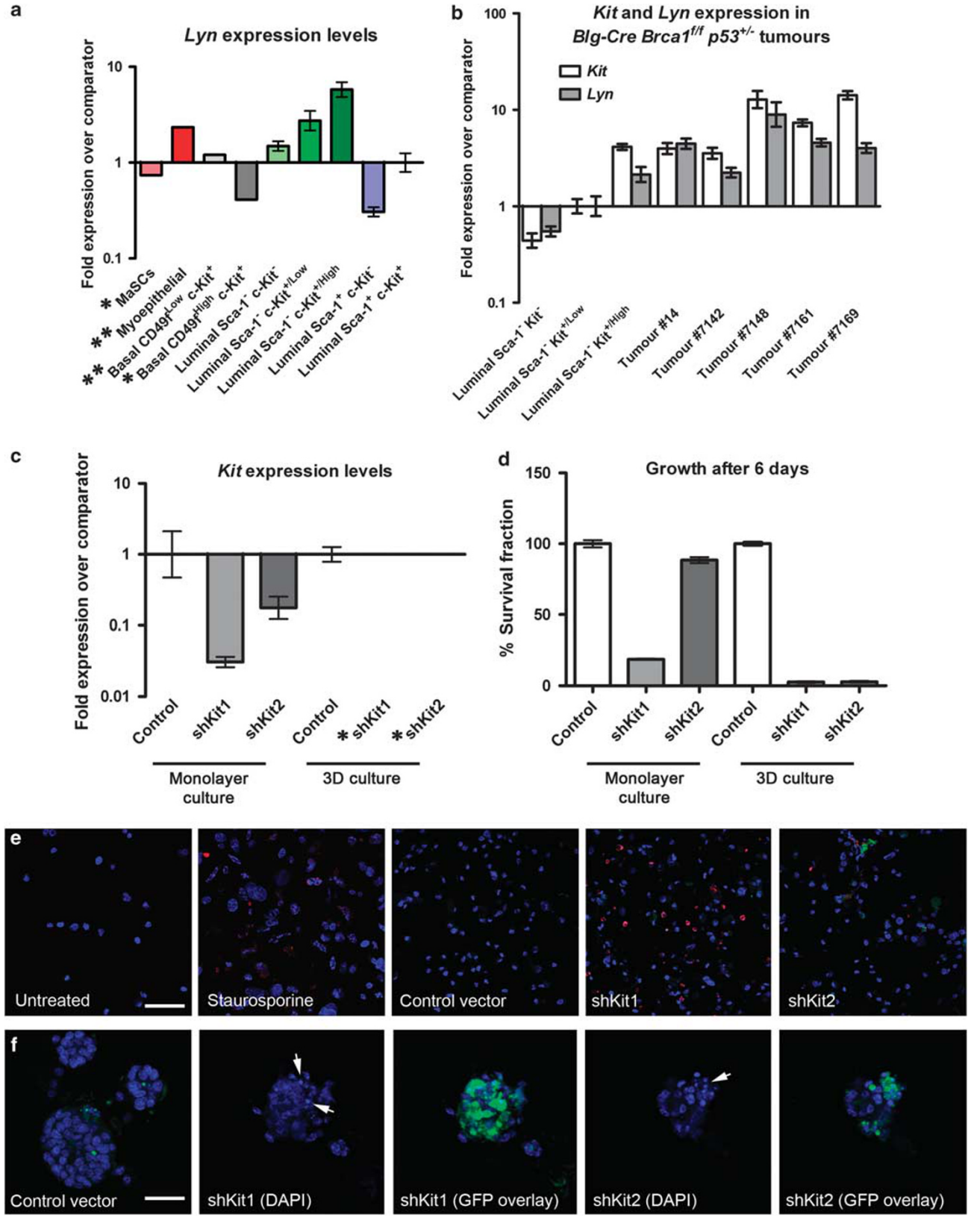
To determine whether c-Kit activity is functionally required for progenitor activity, primary mammary cells grown in both monolayer culture and as three-dimensional acini on Matrigel were transduced with lentiviruses expressing either of two different short hairpin sequences (shKit1 and shKit2) or a control virus. Cells were harvested after 2 days to determine the extent of c-Kit knockdown and cell numbers were determined in parallel cultures after 6 days. In monolayer, cultures transduced with shKit1 expressed approximately thirty-fold less *Kit* mRNA than control cells. Cultures transduced with shKit2 expressed approximately six-fold less *Kit* mRNA than control. In three-dimensional culture, however, *Kit* expression was undetectable in both shKit1 and shKit2 cultures (Figure 5c). After 6 days in monolayer culture, the overall number of cells in the shKit2-transduced cultures was reduced by 10% compared with control. However, in the shKit1-transduced cultures the number of surviving cells was reduced by 80% compared with control, consistent with the extent of knockdown with this virus. On Matrigel, cell numbers were reduced by >95% compared with control in both shKit1- and shKit2-transduced cultures after 6 days (Figure 5d).

To determine whether c-Kit knockdown blocked cell proliferation or compromised cellular viability, the extent of apoptosis in treated cultures was measured. In monolayer culture, shKit1 transduction caused extensive apoptosis at levels comparable to staurosporine-treated positive controls (Figure 5e). Transduction with shKit2 also resulted in apoptosis, although at lower levels than shKit1 (Figure 5e). In three-dimensional culture, c-Kit knockdown also caused extensive apoptosis, as demonstrated by disrupted acinar morphology and presence of numerous pyknotic nuclei (Figure 5f, arrowheads).

To determine the effects of deregulated c-Kit overexpression *in vivo*, primary mammary epithelial cells were transduced with either a lentivirus carrying a c-Kit (GNNK^{+} isoform) cDNA, as well as green fluorescent protein (GFP) in a bicistronic expression cassette, or a control virus coding for GFP only

Figure 5 c-Kit signalling is required for mammary epithelial progenitor growth. (**a**) Levels of *Lyn* gene expression in mammary epithelial subpopulations relative to comparator (the luminal Sca-1^{+} c-Kit^{+} population). Data are presented as fold expression levels ± 95% confidence intervals (n = three independently harvested isolates of each cell population). **Lyn* gene expression was only detected in one of three isolates of these populations. No error bars are therefore shown. ***Lyn* gene expression was only detected in two of three isolates of these populations. No error bars are therefore shown. (**b**) *Kit* and *Lyn* gene expression in normal luminal Sca-1^{-} c-Kit^{-}, c-Kit$^{+/Low}$ and c-Kit$^{+/High}$ populations, as well as in five luminal Sca-1^{-}-origin *Blg-Cre Brca1$^{f/f}$p53$^{+/-}$* mammary tumours (Molyneux *et al*., 2010). (**c**) *Kit* gene expression levels in primary mammary epithelial cells in *in vitro* monolayer or three-dimensional culture 2 days after transduction with control or *Kit*-knockdown lentiviruses. Data expressed as fold expression (mean ± 95% confidence intervals; data from three independent experiments) over comparator population (the control transduced cells). * = Kit expression was below the limits of detection in these samples. (**d**) Mammary epithelial cell growth six days after transduction with control or *Kit*-knockdown lentiviruses. Data expressed as percentage (mean ± s.d.; data from three independent experiments) of cells surviving relative to control transduced monolayer or three-dimensional cultures. (**e**) Apoptosis detected by indirect TUNEL labelling (indicated by red stain) in control untreated, staurosporine treated (1 μm/ml for four hours) or transduced with control, shKit1 or shKit2 viruses (and then cultured for four days) monolayer primary mammary epithelial cells. Nuclei are stained blue with DAPI. The virus transduced cells also express GFP at low levels. Bar = 40 μm. (**f**) Nuclear morphology of control-, shKit1- and shKit2-transduced mammary acini cultured on Matrigel. Note the disrupted acinar morphology in the knockdown cultures and pyknotic nuclei. Bar = 20 μm.

(Kendrick *et al*., 2008) and transplanted into cleared fat pads. Flow cytometric analysis confirmed that cells transduced with the c-Kit virus had higher levels of c-Kit expression than control cells (Supplementary Figure 3a). After 8 weeks, fat pads were isolated and examined for the presence of GFP$^+$ outgrowths. Of the



24 fat pads transplanted with cells overexpressing c-Kit in three separate experiments, only two gave rise to very small GFP$^+$ transplants (8%). In contrast, transplantation of control cells gave rise to 16 extensive GFP$^+$ outgrowths out of 27 fat pads (59%) transplanted ($P<0.01$; χ^2-test) (Supplementary Figure 3b). Therefore, although c-Kit expression is a marker of mammary epithelial progenitors and is required for their survival, forced c-Kit over-expression prevents cells from contributing to outgrowths in cleared fat pads. The proper regulation of c-Kit is therefore required for normal mammary development.

Discussion

The origins of the different subtypes of breast cancer can be explained, in part, by transformation of different stem and/or progenitor cell types (Dontu *et al*., 2003; Molyneux *et al*., 2007, 2010). We have shown that ER$^-$ progenitors in the luminal cell layer of the mammary epithelium are the cells of origin of *BRCA1*-associated breast cancers (Molyneux *et al*., 2010). Here, we demonstrate that c-Kit is a marker of these progenitors, consistent with data from the human (Lim *et al*., 2009). Unlike in the human, however, a c-Kit$^+$ luminal ER$^+$ progenitor population was also identified. Such cells have been suggested as the origin of ER-positive breast cancer (Dontu *et al*., 2003; Molyneux *et al*., 2007; Arendt *et al*., 2011), but until now it has not proven possible to prospectively isolate them. Importantly, c-Kit was required for progenitor function and is therefore only the second mammary stem/progenitor marker, after β1 integrin (Taddei *et al*., 2008), to be shown to have a functional role in the gland.

Luminal Sca-1$^-$ ER$^-$ c-Kit$^+$ cells formed >50% of the epithelial content of the mammary gland, consistent with a report showing reciprocal staining of c-KIT and ER in the human breast (Westbury *et al*., 2009). c-Kit$^+$ luminal cells expressed markers of multiple lineages, as has been reported for a subset of luminal cells in the human (Gusterson *et al*., 2005; Lim *et al*., 2009), and had high *in vitro* proliferative potential. When transplanted, the limited ability to form small outgrowths in cleared fat pads that we previously observed in the luminal epithelial populations (Sleeman *et al*., 2007) was found to be a property of c-Kit$^+$ cells. Together these results suggest that c-Kit expression principally marks luminal progenitors with high proliferative potential and the capacity for multilineage differentiation. This is a huge reservoir of potential proliferative capacity in a tissue in which there is normally little proliferation (Clarke *et al*., 1997; Fernandez-Gonzalez *et al*., 2009). This potential is required, however, during the massive growth stimulus, which occurs during pregnancy, when epithelial cell numbers can increase up to 20-fold (Kordon and Smith, 1998). Importantly, the multilineage differentiation potential of these luminal progenitors is consistent with their ability to give rise to tumours with basal features (Molyneux *et al*., 2010).

The large majority of basal cells, both CD49f^{Low} myoepithelial cells and CD49f^{High} MaSCs, were c-Kit$^-$. However, rare c-Kit$^+$ cells were observed in both the CD49f^{Low} and CD49f^{High} populations. These rare cells were too scarce to be confidently classed as distinct populations, however, analysis of staining for K14, K18 and ER, as well as gene expression analysis suggested that while they did have similarities to other mammary epithelial populations, they also had their own unique features. Gene expression analysis confirmed that both the CD49f^{Low} and CD49f^{High} c-Kit$^+$ cells expressed *Kit*, whereas *Kit* was undetectable in the majority of basal cells. Their strong K14 staining argued against these rare cells being contaminants from the luminal c-Kit$^+$ populations. The basal CD49f^{Low} c-Kit$^+$ cells had similarities to myoepithelial cells, including being CD49f^{Low}, K14$^+$ and K18$^-$. However, although they were CD49f^{Low}, they expressed *Itga6* at levels comparable to CD49f^{High} cells. They also expressed luminal associated genes, such as *Esr1*, although they did not stain for ER. The basal CD49f^{High} c-Kit$^+$ cells had similarities to MaSCs, being CD49f^{High} and K14$^+$, as well as to luminal cells, being weakly K18$^+$. Crucially, however, gene expression analysis and clustering demonstrated that both the basal CD49f^{Low} c-Kit$^+$ and basal CD49f^{High} c-Kit$^+$ cells were in fact most closely related to the MaSCs. This suggests that the occasional c-Kit$^+$ basal cells may be variant MaSCs undergoing 'lineage priming', in which stem/progenitor cells express genes associated with their differentiated daughter populations (Huang *et al*., 2007; Mansson *et al*., 2007). Such cells are unlikely to comprise a stable population, but will revert to a canonical stem cell state if they do not receive signals that promote their differentiation.

The identity of the luminal Sca-1$^-$ c-Kit$^-$ cells is more difficult to interpret. These cells expressed levels of the Kit gene comparable to other Kit$^+$ populations (Figure 1c), but were not significantly contaminated by luminal Sca-1$^-$ c-Kit$^+$ ER$^-$ cells as they were ER$^+$ (Table 1). Furthermore, they clustered with the luminal Sca-1$^-$ c-Kit$^+$ ER$^-$ cells (Figure 2b), but also expressed genes in common with the luminal Sca-1$^+$ ER$^+$ cells, such as *Prom1* and *Prlr* (Supplementary Table 1). We suggest that they may represent an alternative route for generating ER$^+$ cells from luminal Sca-1$^-$ progenitors, rather than from the luminal Sca-1$^+$ ER$^+$ progenitor population. This is supported by lineage tracing experiments, which suggest that a small minority of luminal Sca-1$^+$ cells are indeed derived from the luminal Sca-1$^-$ population and not from basal stem cells (Molyneux *et al*., 2010). A potential mammary epithelial cell hierarchy based on these data is shown in Supplementary Figure 3a.

Not only was c-Kit a marker of progenitors, but also its inhibition *in vitro* by shRNA knockdown resulted in their apoptosis. The effects of c-Kit loss of function *in vivo* are more difficult to determine. Homozygotes for naturally occurring mutations in *Kit* and *Kitl* (the *dominant white-spotting* and *steel* mutants of mice, respectively) are often lethal and those that do survive are sterile and anaemic (Baxter *et al*., 2004). In *W-sash*

mice, which carry a hypomorphic *Kit* allele, mammary ductal outgrowth but not differentiation were impaired (Lilla and Werb, 2010). These mice had decreased mammary epithelial proliferation, however, it was not possible to distinguish cell autonomous from non-cell autonomous (such as mast cell depletion) effects of c-Kit loss. Our findings confirm a cell-autonomous role for c-Kit in the mammary epithelium although it may be that *in vivo*, interactions with the microenvironment alter the response to c-Kit inhibition in proliferating mammary cells from apoptosis to proliferative arrest.

Forced over-expression of c-Kit was also deleterious, as it blocked the ability of primary cells to form outgrowths in cleared mammary fat pads. It may be that forced c-Kit expression prevented differentiation and this stopped normal outgrowths from developing (for example, by preventing the appearance of functionally mature luminal Sca-1$^+$ ER$^+$ cells required for the paracrine interactions necessary for ductal extension) (Mallepell *et al*., 2006). Consistent with this, downregulation of c-Kit expression was observed when c-Kit$^+$ cells were cultured under conditions that generate an *in vivo*-like morphology. These data therefore suggest that c-Kit activity must be downregulated to allow normal differentiation.

Signalling pathways downstream of c-Kit are complex (Roskoski, 2005a, b). This complexity is increased by the existence of two different splice variants of c-Kit, GNNK$^-$ and GNNK$^+$. Both are thought to be present in all cells in which c-Kit is expressed (Hayashi *et al*., 1991; Reith *et al*., 1991). The two isoforms have different signalling activities, although these appear to be cell-type specific and possibly context dependent. In NIH3T3 cells (which have low levels of GAB2), both isoforms activate PI3K/Akt signalling but the GNNK$^-$ variant also strongly activates MAPK signalling in a Src-dependent manner (Caruana *et al*., 1999; Voytyuk *et al*., 2003; Young *et al*., 2007). However, in haematopoietic cells, PI3K activity is at least partly dependent on Src-family kinase mediated phosphorylation of GAB2 and this can lead to isoform-specific c-Kit activation of PI3K signalling (Yu *et al*., 2006; Sun *et al*., 2008). In NIH3T3 cells both isoforms promote resistance to anoikis, but GNNK$^-$ also strongly promotes loss of contact inhibition and increases tumourigenicity (Caruana *et al*., 1999). It is unclear at this time, which isoform predominates in mammary epithelial cells or whether the balance of isoforms is different between the subpopulations. We used the GNNK$^+$ variant, which is less strongly transforming, for our over-expression studies.

Kit ligand/SCF is also found in two isoforms which differ by the presence or absence of exon 6. It localises to the cell membrane, where the isoform that includes exon 6 can undergo proteolytic cleavage of the extracellular domain to release soluble SCF. The isoform that lacks exon 6 cannot be cleaved and remains membrane bound (Miyazawa *et al*., 1995; Gommerman *et al*., 2000; Driessen *et al*., 2003). The soluble form of SCF rapidly activates c-Kit but the signalling activity quickly decays owing to receptor internalisation. The membrane bound form gives a more persistent signalling activity, possibly owing to increased stability of the receptor-ligand complex and reduced internalisation (Miyazawa *et al*., 1995; Gommerman *et al*., 1997). *In vivo*, only the membrane bound form supports survival of long-term haematopoietic progenitors (Gommerman *et al*., 2000; Driessen *et al*., 2003). The Steel-Dickie mutation in mice generates an SCF protein lacking both transmembrane and cytoplasmic domains and results in production of only soluble SCF (Brannan *et al*., 1991), supporting the importance of the membrane bound form in a number of biological processes. It is not yet clear which form predominates in the mammary epithelium.

Anti-c-KIT therapy has had limited success in human breast cancer treatment (Barton *et al*., 2006; Chew *et al*., 2006; Haley *et al*., 2007; Waterhouse *et al*., 2008), suggesting that in transformed mammary cells the c-KIT signalling network may be activated downstream of the receptor. Consistent with this is the strong association of basal-like breast cancers with high expression of the Src-family kinase member Lyn (Molyneux *et al*., 2010), a transducer of c-Kit signalling (Linnekin *et al*., 1997). Previous transcriptomic analysis suggested that *Lyn* is strongly expressed in the normal luminal ER$^-$ progenitors (Kendrick *et al*., 2008). Here, analysis of *Lyn* expression in the mammary subpopulations by qPCR demonstrated that in the luminal cells, but not the basal cells, the relative levels of *Kit* and *Lyn* expression had a very similar pattern. Although correlative, these data support a role for Lyn as a transducer of c-Kit signalling in the mammary luminal epithelium. Further work will be needed to delineate the complexities of c-Kit signalling in the different luminal progenitor populations.

Kit and *Lyn* were expressed in mouse *Brca1* tumours derived from the luminal Sca-1$^-$ ER$^-$ progenitor population (Molyneux *et al*., 2010) at levels comparable to those seen in the c-Kit$^{+/High}$ fraction of that population. This suggests that *Brca1* tumours are blocked in a progenitor-like state. The expression of basal markers at low levels in the normal c-Kit$^+$ luminal progenitors (Table 1) and the presence in the tumours of neoplastic cells with both luminal and basal features, would be consistent with such a block. Lyn is a potential regulator of Brca1, acting via Akt (Linnekin *et al*., 1997; Altiok *et al*., 1999; Shivakrupa and Linnekin, 2005). Brca1 is itself a potential regulator of Lyn, acting via CDC2 (Pathan *et al*., 1996; Wang *et al*., 1997). Furthermore, the downstream transducer of c-Kit signalling, Stat5a, is negatively regulated by Brca1 (Brizzi *et al*., 1999; Vidarsson *et al*., 2002), while Stat1 and Jak2 are positively regulated by both Brca1 and c-Kit (Deberry *et al*., 1997; Chin *et al*., 1998; Ouchi *et al*., 2000; Gao *et al*., 2001). Taken together, these studies suggest a model in which Brca1, c-Kit and Lyn function in a feedback loop that modulates Jak/Stat survival signalling (Supplementary Figure 3b). Although this remains to be directly tested in the mammary epithelium, in a recent study *Brca1* loss in mouse mammary epithelial cells was found to result in defective differentiation and increased *Kit* expression (Smart *et al*., 2011), consistent

with our data on *Kit* and *Lyn* expression in mouse *Brca1* tumours.

Our data show that the c-Kit signalling pathway is required for the normal function of mammary epithelial progenitors. Regulation of this pathway by Brca1 activity may explain why these cells are sensitive to transformation by *Brca1* loss, while over-expression of Lyn in basal-like breast cancers may explain why anti-c-Kit therapy has had poor results in breast cancer. Understanding how c-Kit and Brca1 regulate each other has important implications for therapeutic and preventative approaches in basal-like breast cancer.

Materials and methods

Preparation of mammary epithelial cells

All animal work was carried out under UK Home Office project and personal licences following local ethical approval and in accordance with local and national guidelines. Single cells were prepared from fourth mammary fat pads of 10–week-old virgin female FVB mice as described (Smalley, 2010). Cell suspensions at 10^6 cells/ml were stained with combinations of anti-CD24-FITC (1.0 μg/ml; BD Biosciences, Oxford, UK), anti-Sca-1-APC (1.0 μg/ml; eBioscience, Hatfield, UK), anti-CD45-PE-Cy7 (1.0 μg/ml; BD Biosciences), anti-CD49f-PE-Cy5 (5.0 μl/ml; BD Biosciences) and anti-c-Kit-PE (1.0 μg/ml; BD Biosciences). Cells were sorted on a FACSAria (Becton Dickenson, Oxford, UK) and dead cells, $CD45^+$ leucocytes and non-single cells were excluded as shown in Supplementary Figure 1 (Sleeman *et al*., 2006). Cells incubated in non-specific IgG were used to set the limits that defined negative and positive staining for each antibody. Within the $CD24^+$ population, two separate cell populations were annotated as $CD24^{+/High}$ and $CD24^{+/Low}$ as previously described (Sleeman *et al*., 2006). Within the basal $CD49f^+$ population $CD49f^{High}$ mammary stem cell enriched cells were defined by previous understanding and with reference to the outermost contours of a 5% interval linear density contour plot as described (Supplementary Figure 1) (Britt *et al*., 2009). For consistency, c-$Kit^{+/High}$ cells within the luminal Sca-1^- c-Kit^+ population were defined with reference to the outermost contours of a 5% interval linear density contour plot.

Gene expression analysis by quantitative real-time rtPCR

Freshly sorted normal cells were resuspended in RLT buffer (Qiagen, Crawley, West Sussex, UK) and stored at −80 °C until required for RNA extraction. RNA was isolated from cultured cells with Trizol (Invitrogen, Paisley, UK). qPCR reactions were performed largely as previously described (Kendrick *et al*., 2008) using TAQMAN Assays-on-Demand probes (Supplementary Table 3). *Actb* (β-actin) was used as an endogenous control and results calculated using the Δ-ΔC_t method. Data were expressed as the mean fold gene expression difference in three independently isolated cell preparations over a comparator sample with 95% confidence intervals.

The *Blg-Cre* $Brca1^{flox/flox}p53^{+/-}$ mouse model and the tumours it generates have been fully described previously (Molyneux *et al*., 2010). For analysis of *Kit* and *Lyn* gene expression by qPCR, archived RNA samples were used. Previous analysis of these samples with a probe targeting *Brca1* exon 22 (contained within the *floxed* region) had demonstrated a very high contribution of neoplastic cells and minimal normal tissue contamination (Molyneux *et al*., 2010).

For unsupervised hierarchical clustering of qPCR gene expression, data on epithelial subpopulations and median-centred data for mean fold expression levels (n = three independent samples) over the comparator population were uploaded into CLADIST (http://www.brcabase.icr.ac.uk/software/cladist.jsp). Average linkage clustering with a Euclidean distance similarity metric was used.

Lentivirus production

Virus over-expressing the $GNNK^+$ isoform of c-Kit was generated using the Tronolab (http://www.tronolab.epfl.ch/) pWPI system (Kendrick *et al*., 2008). Oligonucleotide pairs for shKit1 (CACCGCCTTCCCTTGTTTGGCAAAGCGAA CTTTGCCAAACAAGGGAAGGC and AAAAGCCTTCC CTTGTTTGGCAAAGTTCGCTTTGCCAAACAAGGGAA GGC) and shKit2 (CACCGCAGAAGCGGCACTTTATAA GCGAACTTATAAAGTGCCGCTTCTGC and AAAAGC AGAAGCGGCACTTTATAAGTTCGCTTATAAAGTGCC GCTTCTGC) were ligated into pENTR/U6 Gateway system entry vector (Invitrogen) according to manufacturer's instructions. Hairpin sequences were verified and then transferred, together with the U6 promoter, into a Gateway-modified pSEW lentiviral vector backbone (Vafaizadeh *et al*., 2010) by LR reaction (Invitrogen).

Viral supernatants were generated as described (Kendrick *et al*., 2008). shKit1 and shKit2 viruses were concentrated with Lenti-X concentrator before use (Clontech Europe, Saint-Germain-en-Laye, France). Supernatants were stored at −80 °C until use. Primary mammary cells were transduced with virus using the suspension method as described (Kendrick *et al*., 2008). Relative titres of viral supernatants were determined by transducing NIH 3T3 cells and analysing the number of GFP^+ cells after 6 days. The relative titre of the pWPI-c-Kit over-expression virus was found to be 60% of that of the control. The relative titres of the shKit1 and shKit2 viruses were found to be 181 and 127% that of the control (which was not concentrated before use), respectively.

Immunofluorescence staining

Immunophenotyping of cell populations sorted on to slides and of mouse mammary epithelial cell clones grown on glass coverslips was carried out as described (Sleeman *et al*., 2007) with the exception that cells sorted on to slides were incubated with primary antibody either for 40 min at room temperature or overnight at 4 °C. Antibodies used were LLOO2 for K14 (0.26 μg/ml; Abcam, Cambridge, UK), Ks18.04 for K8/18 (1:2 dilution; Progen Biotechnik, Heidelberg, Germany) and ID5 for ERα (9.9 μg/ml; DAKO, Ely, UK). Negative controls with isotype-matched clones gave no staining.

Staining intensities of K14 and K18 were assessed by comparison with unstained controls using identical settings on a Leica TCS SP2 confocal microscope (Leica Microsystems, Milton Keynes, UK). Images captured with Leica confocal software were used for scoring staining. The following criteria were used to score cells as negative, weakly positive or positively stained. In 'negative' populations, no staining could be distinguished above background using either protocol. In populations consisting of cells with 'weak' staining, such staining could only be distinguished above the background in cells stained overnight at 4 °C. In samples stained at room temperature for 40 min, staining could not be distinguished above background. In populations scored as 'positive', the staining was still distinct even with the shorter incubation time.

For Lyn staining in tumour samples, fresh sections were cut from archived formalin-fixed and paraffin-embedded tissue. Dewaxed and re-hydrated sections underwent antigen retrieval

in citrate buffer (0.01 M, pH 6.0) for 18 min in a microwave (900W) before blocking in 0.05% hydrogen peroxide solution then protein block (Dako UK Ltd, Cambridgeshire, UK) for 30 min. The Lyn primary antibody (ab53141, rabbit polyclonal, Abcam; diluted 1:200) was incubated overnight at 4 °C. Detection was carried out using the ABC method (ABC, Vector Laboratories, Peterborough, UK). Negative controls were performed using the same protocols with substitution of the primary antibody with IgG-matched controls.

In vitro *analysis of mammary progenitors*
Analysis of *in vitro* colony forming potential of single cells was carried out in a 96-well plate format as described (Sleeman *et al*., 2007). Single cellularity of sorted populations was ensured using time-of-flight gating on both forward and side scatter. Statistical comparisons were carried out using t-tests of Log10 transformed data.

For three-dimensional culture of mammary epithelial cells, 24-well plates were prepared with 200 µl growth factor reduced Matrigel (BD Biosciences) per well. Wells were seeded at 10^4 mammary epithelial cells per well in 1 ml/well of 1:1 DMEM:-Ham's F12 medium (Invitrogen) with 10% FCS (PAA, Yeovil, UK), 5 ug/ml insulin (Sigma, Poole, UK), 10 ng/ml cholera toxin (Sigma) and 10 ng/ml epidermal growth factor (Sigma). Cultures were maintained at 37 °C in a 5%O_2 atmosphere in a Heraeus BB6060 low oxygen incubator (Thermo Fisher Scientific, Leicestershire, UK) for up to 20 days. Numbers and sizes of colonies were determined from three fields of view of photomicrographs of cultures taken at identical magnifications. Colony areas were determined by image analysis using ImageJ image analysis software (http://www.rsbweb.nih.gov/ij/) or Adobe Photoshop CS4 Extended (Adobe Systems, Uxbridge, UK). Owing to the optical properties of Matrigel cultures and the necessity to change focal planes, when examining the centre as opposed to the edges of wells, it was not possible to take a single photograph of an entire well for automated analysis.

For haematoxylin and eosin stained sections of three-dimensional colonies, gels were fixed overnight in 4% paraformaldehyde in PBS at room temperature. The fixed Matrigel containing the colonies was then removed from the wells and subsequently processed by standard histological techniques.

Mammary epithelial cell transplantation
Transplantation of freshly isolated primary mouse mammary epithelial cells at limiting dilutions into cleared fats pads was carried out as described (Sleeman *et al*., 2007; Britt *et al*., 2009). For single cell transplants, freshly isolated cells were directly sorted into the wells of low-volume (10 µl) Terasaki plates using the 'single cell' purity protocol. The presence of a cell was visually confirmed and then the full volume of each well injected. Fat pads were wholemounted, carmine stained and analysed 8 weeks after transplant as described (Sleeman *et al*., 2007; Britt *et al*., 2009). The χ^2-test comparing the expected versus the actual number of outgrowths in transplanted fat pads was calculated using Excel.

Survival assays
Cell viability in monolayer cultures was quantified using the CellTitre Glo cell viability reagent (Promega, Southampton, Hampshire, UK), according to the manufacturer's instructions. For three-dimensional culture, the reagent was added directly to 24-well plates containing cells in Matrigel following media removal. Plates were left to equilibrate at room temperature for 20 min before lysates being transferred to 96-well plates. Assay plates were read in a luminescence plate reader (PerkinElmer, Cambridge, Cambridgeshire, UK).

Quantification of apoptotic cells
Quantification of apoptosis in monolayer cultures was carried out using the ApopTag Red indirect TUNEL labelling kit (Millipore, Watford, UK). Following treatment, both floating and adherent cell fractions were combined as cytospins and then analysed for an overall assessment of apoptotic fraction. Staurosporine-treated cells (1 µl/ml for four hours) were used as a positive control. Technical difficulties in recovering cells from three-dimensional Matrigel cultures meant the presence of apoptosis in these samples had to be determined by direct visualisation of nuclear morphology. Eight-well chamber slides (LabTek, Nunc, Roskilde, Denmark) were prepared with 40 µl of Matrigel (BD Biosciences) per well and seeded with 8000 cells per well. Cells were fixed *in situ* in 4% paraformaldehyde after 2 days, stained with DAPI and analysed on a Leica TCS SP2 confocal microscope using the immersion lens. Apoptosis was determined by the assessment of nuclear morphology.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Eric So for the c-Kit cDNA, Fredrik Wallberg for help with flow cytometry analysis and Jorge Reis-Filho and Arno Gauthier for assistance with assessment of Lyn staining on mouse tumours. This work was funded by Breakthrough Breast Cancer. We acknowledge NHS funding to the NIHR Biomedical Research Centre. This work was funded by Breakthrough Breast Cancer. We acknowledge NHS funding to the NIHR Biomedical Research Centre.

References

Altiok S, Batt D, Altiok N, Papautsky A, Downward J, Roberts TM *et al*. (1999). Heregulin induces phosphorylation of BRCA1 through phosphatidylinositol 3-Kinase/AKT in breast cancer cells. *J Biol Chem* **274**: 32274–32278.

Arendt LM, Rugowski DE, Grafwallner-Huseth TA, Garcia-Barchino MJ, Rui H, Schuler LA. (2011). Prolactin-induced mouse mammary carcinomas model estrogen resistant luminal breast cancer. *Breast Cancer Res* **13**: R11.

Asselin-Labat ML, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC *et al*. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9**: 201–209.

Barton J, Liggett W, Mainwaring M, Hainsworth JD, Simons L, Spigel DR *et al*. (2006). Phase II pilot trial of imatinib mesylate with weekly docetaxel in metastatic breast cancer. *J Clin Oncol* **24**: 10716.

Baxter LL, Hou L, Loftus SK, Pavan WJ. (2004). Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders. *Pigment Cell Res* **17**: 215–224.

Brannan CI, Lyman SD, Williams DE, Eisenman J, Anderson DM, Cosman D *et al*. (1991). Steel-Dickie mutation encodes a c-kit ligand lacking transmembrane and cytoplasmic domains. *Proc Natl Acad Sci USA* **88**: 4671–4674.

Britt KL, Kendrick H, Regan JL, Molyneux G, Magnay FA, Ashworth A *et al*. (2009). Pregnancy in the mature adult mouse does not alter the proportion of mammary epithelial stem/progenitor cells. *Breast Cancer Res* **11**: R20.

Brizzi MF, Dentelli P, Rosso A, Yarden Y, Pegoraro L. (1999). STAT protein recruitment and activation in c-Kit deletion mutants. *J Biol Chem* **274**: 16965–16972.

Caruana G, Cambareri AC, Ashman LK. (1999). Isoforms of c-KIT differ in activation of signalling pathways and transformation of NIH3T3 fibroblasts. *Oncogene* **18**: 5573–5581.

Chew HK, Barlow W, Albain K, Lew D, Budd T, Allen G *et al*. (2006). SWOG 0338: A phase II trial of imatinib mesylate in combination with capecitabine in metastatic breast cancer. *J Clin Oncol* **24**: 10529.

Chin H, Arai A, Wakao H, Kamiyama R, Miyasaka N, Miura O. (1998). Lyn physically associates with the erythropoietin receptor and may play a role in activation of the Stat5 pathway. *Blood* **91**: 3734–3745.

Clarke RB, Howell A, Potten CS, Anderson E. (1997). Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res* **57**: 4987–4991.

Deberry C, Mou S, Linnekin D. (1997). Stat1 associates with c-kit and is activated in response to stem cell factor. *Biochem J* **327**(Part 1): 73–80.

Dontu G, Al-Hajj M, Abdallah WM, Clarke MF, Wicha MS. (2003). Stem cells in normal breast development and breast cancer. *Cell Prolif* **36**(Suppl 1): 59–72.

Driessen RL, Johnston HM, Nilsson SK. (2003). Membrane-bound stem cell factor is a key regulator in the initial lodgment of stem cells within the endosteal marrow region. *Exp Hematol* **31**: 1284–1291.

Fernandez-Gonzalez R, Illa-Bochaca I, Welm B, Fleish MC, Werb Z, Ortiz-deSolorzano C *et al*. (2009). Mapping mammary gland architecture using multi-scale *in situ* analysis. *Integr Biol* **1**: 80–89.

Gao B, Shen X, Kunos G, Meng Q, Goldberg ID, Rosen EM *et al*. (2001). Constitutive activation of JAK-STAT3 signaling by BRCA1 in human prostate cancer cells. *FEBS Lett* **488**: 179–184.

Gommerman JL, Rottapel R, Berger SA. (1997). Phosphatidylinositol 3-kinase and Ca2+ influx dependence for ligand-stimulated internalization of the c-Kit receptor. *J Biol Chem* **272**: 30519–30525.

Gommerman JL, Sittaro D, Klebasz NZ, Williams DA, Berger SA. (2000). Differential stimulation of c-Kit mutants by membrane-bound and soluble Steel Factor correlates with leukemic potential. *Blood* **96**: 3734–3742.

Gusterson BA, Ross DT, Heath VJ, Stein T. (2005). Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res* **7**: 143–148.

Haley BB, Ashfaq R, DeHaas M, Ramaswami A, Sikder K, Tripathy D. (2007). A phase I/II study of imatinib and docetaxel as neoadjuvant therapy in locally advanced breast cancer. *J Clin Oncol* **25**: 11039.

Hayashi S, Kunisada T, Ogawa M, Yamaguchi K, Nishikawa S. (1991). Exon skipping by mutation of an authentic splice site of c-kit gene in W/W mouse. *Nucleic Acids Res* **19**: 1267–1271.

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

Kendrick H, Regan JL, Magnay FA, Grigoriadis A, Mitsopoulos C, Zvelebil M *et al*. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics* **9**: 591.

Kent D, Copley M, Benz C, Dykstra B, Bowie M, Eaves C. (2008). Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway. *Clin Cancer Res* **14**: 1926–1930.

Kordon EC, Smith GH. (1998). An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**: 1921–1930.

Lilla JN, Werb Z. (2010). Mast cells contribute to the stromal microenvironment in mammary gland branching morphogenesis. *Dev Biol* **337**: 124–133.

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH *et al*. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**: 907–913.

Linnekin D, DeBerry CS, Mou S. (1997). Lyn associates with the juxtamembrane region of c-Kit and is activated by stem cell factor in hematopoietic cell lines and normal progenitor cells. *J Biol Chem* **272**: 27450–27455.

Mallepell S, Krust A, Chambon P, Brisken C. (2006). Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci USA* **103**: 2196–2201.

Mansson R, Hultquist A, Luc S, Yang L, Anderson K, Kharazi S *et al*. (2007). Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**: 407–419.

Miyazawa K, Williams DA, Gotoh A, Nishimaki J, Broxmeyer HE, Toyama K. (1995). Membrane-bound Steel factor induces more persistent tyrosine kinase activation and longer life span of c-kit gene-encoded protein than its soluble form. *Blood* **85**: 641–649.

Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R *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.

Molyneux G, Regan J, Smalley MJ. (2007). Mammary stem cells and breast cancer. *Cell Mol Life Sci* **64**: 3248–3260.

Ouchi T, Lee SW, Ouchi M, Aaronson SA, Horvath CM. (2000). Collaboration of signal transducer and activator of transcription 1 (STAT1) and BRCA1 in differential regulation of IFN-gamma target genes. *Proc Natl Acad Sci USA* **97**: 5208–5213.

Pathan NI, Geahlen RL, Harrison ML. (1996). The protein-tyrosine kinase Lck associates with and is phosphorylated by Cdc2. *J Biol Chem* **271**: 27517–27523.

Reith AD, Ellis C, Lyman SD, Anderson DM, Williams DE, Bernstein A *et al*. (1991). Signal transduction by normal isoforms and W mutant variants of the Kit receptor tyrosine kinase. *EMBO J* **10**: 2451–2459.

Roskoski Jr R. (2005a). Structure and regulation of Kit protein-tyrosine kinase—the stem cell factor receptor. *Biochem Biophys Res Commun* **338**: 1307–1315.

Roskoski Jr R. (2005b). Signaling by Kit protein-tyrosine kinase—the stem cell factor receptor. *Biochem Biophys Res Commun* **337**: 1–13.

Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML *et al*. (2006). Generation of a functional mammary gland from a single stem cell. *Nature* **439**: 84–88.

Shivakrupa R, Linnekin D. (2005). Lyn contributes to regulation of multiple Kit-dependent signaling pathways in murine bone marrow mast cells. *Cell Signal* **17**: 103–109.

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

Sleeman KE, Kendrick H, Ashworth A, Isacke CM, Smalley MJ. (2006). CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res* **8**: R7.

Smalley MJ. (2010). Isolation, culture and analysis of mouse mammary epithelial cells. In: Ward A, Tosh D (eds). *Methods in Molecular Biology*. Springer: Berlin, pp 139–170.

Smart CE, Wronski A, French JD, Edwards SL, Asselin-Labat ML, Waddell N *et al*. (2011). Analysis of Brca1-deficient mouse mammary glands reveals reciprocal regulation of Brca1 and c-kit. *Oncogene* **30**: 1597–1607.

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

Sun J, Pedersen M, Ronnstrand L. (2008). Gab2 is involved in differential phosphoinositide 3-kinase signaling by two splice forms of c-Kit. *J Biol Chem* **283**: 27444–27451.

Taddei I, Deugnier MA, Faraldo MM, Petit V, Bouvard D, Medina D *et al*. (2008). Beta1 integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nat Cell Biol* **10**: 716–722.

Vafaizadeh V, Klemmt P, Brendel C, Weber K, Doebele C, Britt K *et al*. (2010). Mammary epithelial reconstitution with gene-modified stem cells assigns roles to Stat5 in luminal alveolar cell fate decisions, differentiation, involution, and mammary tumor formation. *Stem Cells* **28**: 928–938.

Vidarsson H, Mikaelsdottir EK, Rafnar T, Bertwistle D, Ashworth A, Eyfjord JE *et al*. (2002). BRCA1 and BRCA2 bind Stat5a and suppress its transcriptional activity. *FEBS Lett* **532**: 247–252.

Voytyuk O, Lennartsson J, Mogi A, Caruana G, Courtneidge S, Ashman LK *et al*. (2003). Src family kinases are involved in the differential signaling from two splice forms of c-Kit. *J Biol Chem* **278**: 9159–9166.

Wang H, Shao N, Ding QM, Cui J, Reddy ES, Rao VN. (1997). BRCA1 proteins are transported to the nucleus in the absence of serum and splice variants BRCA1a, BRCA1b are tyrosine phosphoproteins that associate with E2F, cyclins and cyclin dependent kinases. *Oncogene* **15**: 143–157.

Waterhouse DM, Mainwaring M, Barton J, Webb C, Markus TM, Spigel DR *et al*. (2008). Phase II pilot results of imatinib mesylate with weekly docetaxel in metastatic breast cancer. *J Clin Oncol* **26**: May 20 Suppl; abstr 1090.

Westbury CB, Reis-Filho JS, Dexter T, Mahler-Araujo B, Fenwick K, Iravani M *et al*. (2009). Genome-wide transcriptomic profiling of microdissected human breast tissue reveals differential expression of KIT (c-Kit, CD117) and oestrogen receptor-alpha (ERalpha) in response to therapeutic radiation. *J Pathol* **219**: 131–140.

Young SM, Cambareri AC, Odell A, Geary SM, Ashman LK. (2007). Early myeloid cells expressing c-KIT isoforms differ in signal transduction, survival and chemotactic responses to Stem Cell Factor. *Cell Signal* **19**: 2572–2581.

Yu M, Luo J, Yang W, Wang Y, Mizuki M, Kanakura Y *et al*. (2006). The scaffolding adapter Gab2, via Shp-2, regulates kit-evoked mast cell proliferation by activating the Rac/JNK pathway. *J Biol Chem* **281**: 28615–28626.

Supplementary Information accompanies the paper on the Oncogene website (http://www.nature.com/onc)