Breast cancer is the most frequent cancer in women and consists of heterogeneous types of tumours that are classified into different histological and molecular subtypes. PIK3CA and P53 (also known as TP53) are the two most frequently mutated genes and are associated with different types of human breast cancers. The cellular origin and the mechanisms leading to PIK3CA-induced tumour heterogeneity remain unknown. Here we used a genetic approach in mice to define the cellular origin of Pik3ca-derived tumours and the impact of mutations in this gene on tumour heterogeneity. Surprisingly, oncogenic Pik3ca H1047R mutant expression at physiological levels in basal cells using keratin (K)5-CreER2T2 mice induced the formation of luminal oestrogen receptor (ER)-positive/progesterone receptor (PR)-positive tumours, while its expression in luminal cells using K8-CreER2T2 mice gave rise to luminal ER+ PR+ tumours or basal-like ER- PR- tumours. Concomitant deletion of p53 and expression of Pik3ca H1047R accelerated tumour development and induced more aggressive mammary tumours. Interestingly, expression of Pik3ca H1047R in unipotent basal cells gave rise to luminal-like cells, while its expression in unipotent luminal cells gave rise to basal-like cells before progressing into invasive tumours. Transcriptional profiling of cells that underwent cell fate transition upon Pik3ca H1047R expression in unipotent progenitors demonstrated a profound oncogene-induced reprogramming of these newly formed cells and identified gene signatures characteristic of the different cell fate switches that occur upon Pik3ca H1047R expression in basal and luminal cells, which correlated with the cell of origin, tumour type and different clinical outcomes. Altogether our study identifies the cellular origin of Pik3ca-induced tumours and reveals that oncogenic Pik3ca H1047R activates a multipotent genetic program in normally lineage-restricted populations at the early stage of tumour initiation, setting the stage for future intratumoural heterogeneity. These results have important implications for our understanding of the mechanisms controlling tumour heterogeneity and the development of new strategies to block PIK3CA breast cancer initiation.

Breast cancers can be classified into different histological and molecular subtypes including luminal (ER+ and/or PR+), HER2+ and basal-like/triple-negative (ER- PR- HER2-), cancers which are usually associated with different gene expression and mutation profiles, prognosis and response to therapies. PIK3CA mutations are found in about 30% of breast cancers, more frequently in luminal tumours, although they are also found in basal-like/triple-negative breast cancers. Expression of oncogenic Pik3ca H1047R in all mammary gland lineages using MMTV-Cre mice is preferentially in luminal progenitors using WAP-Cre mice induces heterogeneous mammary tumours. The reason for this tumour heterogeneity upon expression of the Pik3ca H1047R mutant in the mammary gland is currently unknown.

To determine whether breast tumour heterogeneity is determined by the cancer cell of origin, we developed a genetic strategy allowing the expression of the oncogenic Pik3ca mutant at physiological levels using Cre-inducible Pik3ca H1047R knock-in mice, specifically in basal cells (BCs) using K5-CreER2T2 or in luminal cells (LCs) using K8-CreER2T2 mice and followed their fate and tumorigenic potential over time. Tamoxifen (TAM) was administered at a dose that does not impair long-term mammary gland development and homeostasis, and resulted in the specific labelling of about 20% of BCs (Extended Data Fig. 1) in 4–5-week-old K5-CreER2T2/Pik3ca H1047R mice (Fig. 1a). While it has been suggested that the mammary gland contains bipotent basal stem cells, our data using K5-CreER2T2 knock-in or K14-rtTA/TetO-Cre mice, despite the labelling of 20–50% of BCs, showed no contribution of BCs to the luminal lineage (Extended Data Fig. 1). Further lineage-tracing studies that label all BCs or all LCs will be required to determine whether the discrepancy between the different

---

1 Université Libre de Bruxelles, IRIBM, Brussels B-1070, Belgium. 2Institut Jules Bordet, Université Libre de Bruxelles, Brussels B-1000, Belgium. 3Department of Pathology, Erasme Hospital, Université Libre de Bruxelles, Brussels B-1070, Belgium. 4DIAPATH — Center for Microscopy and Molecular Imaging (CMMI), Gosselies B-6041, Belgium. 5Surgical Oncology Research Laboratory, Peter MacCallum Cancer Centre, Melbourne 3002, Australia. 6Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville 3002, Australia. 7WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium.

*These authors contributed equally to this work.
studies arises from the unspecific and simultaneous labelling of BCs and LCs. BC-derived mammary tumours arose with a latency of about 12 ± 4 months (mean ± standard deviation (s.d.)) and were all luminal-like tumour cells that were ER⁺ PR⁺, surrounded by BCs (Fig. 1c, e and Extended Data Fig. 2a–c), classified by pathologists as adenomyoepithelioma in mouse and human (Extended Data Fig. 3a–d). Principal component analysis (PCA) and gene clustering analysis of gene expression profile from fluorescence-activated cell sorting (FACS)-isolated tumour cells using the PAM50 gene set showed that these BC-derived-tumours clustered together with the luminal B breast cancer subtype (Extended Data Figs 3 and 4).

The same dose of TAM was administered to 4–5-week-old K8-CreERT2/Pik3caH1047R mice, resulting in the specific labelling of about 20–30% of LCs (Fig. 1b and Extended Data Fig. 1). Mammary tumours arose with a similar latency (15 ± 4 months) (Fig. 1d). Histological and immunofluorescence analysis revealed that these tumours were more heterogeneous, more aggressive and more proliferative than BC-derived tumours. These tumours comprised adenomyoepithelioma, mixed adenomyoepithelioma with myoepithelial carcinoma, invasive carcinoma of no special type (NST), as well as tumours that show features of metaplastic basal-like breast cancers similar to human breast cancers (Fig. 1f and Extended Data Figs 2, 3). Principal component and gene expression clustering analyses from cells isolated from seven different luminal-derived tumours showed that ER⁺ tumours clustered together with luminal human breast cancers, NST tumours clustered in between luminal B and HER2⁺ tumours, and metaplastic carcinoma clustered with basal-like or HER2⁺ cancers depending on the clustering algorithm (Extended Data Figs 3j, l), consistent with the phenotypic heterogeneity of the tumours. These results revealed that Pik3caH1047R expression in LCs gives rise to distinct types of tumours that are generally more aggressive compared with BC-derived tumours. The greater tumour heterogeneity found in the LC-derived tumours may arise from the greater plasticity of LCs and/or the heterogeneity of the luminal progenitor populations initially targeted in the K8-CreERT2 mice.

We then assessed whether concomitant p53 deletion affects the phenotype of mammary tumours depending on their cellular origin. K5-CreERT2/Pik3caH1047R/p53f/fl mice treated with TAM rapidly developed skin and other cancers that required terminating the experiment before they developed mammary tumours (data not shown). To circumvent this problem, we used mice heterozygous for p53 (K5-CreERT2/Pik3caH1047R/p53f/fl) and another basal Cre driver (K14-tTA/TetO-Cre/Pik3caH1047R/p53f/fl) that alleviated the increased early mortality seen with the K5-CreERT2/Pik3caH1047R/p53f/fl mice. BC-derived p53 heterozygous tumours arose with a latency of 9 ± 2 months and consisted mostly of adenomyoepithelioma luminal-like tumours (42–75%), as well as myoepithelial carcinoma (0–16%), NST tumours (0–12%) and metaplastic carcinoma (12–42%) (Fig. 2 and Extended Data Fig. 5). As previously shown using MMTV-Cre mice, Pik3caH1047R expression together with p53 deletion in LCs dramatically accelerates tumour formation, with a latency of 5 ± 1 months for p53 homozygous and 9 ± 3 months for p53 heterozygous mice (Fig. 2). In contrast to BCs, LC-derived p53-deficient tumours always consisted of aggressive carcinomas consisting mostly of metaplastic carcinoma and high-grade myoepithelial carcinoma with characteristics of epithelial-to-mesenchymal transition (Fig. 2 and Extended Data Fig. 5), as previously reported following Pik3caH1047R expression in all mammary gland cells6,12 and found in human basal-like breast cancers with activation of the PI3K pathway by somatic PIK3CA mutations and gene copy number amplification6,7. Gene expression clustering of these tumours using the PAM50 genes showed that they clustered together with human basal-like or HER2⁺ subtypes depending on the clustering algorithm (Extended Data Fig. 3k, l). These data demonstrate that concomitant Pik3caH1047R expression and p53 deletion accelerates tumour development in basal and luminal lineages and that very aggressive metaplastic tumours arise more frequently from oncogenic targeting of LCs than from BCs.

To define further the cellular basis of intratumoural heterogeneity found in Pik3caH1047R-derived tumours, we combined Rosana26-YFP lineage tracing and Pik3caH1047R expression specifically in LCs or BCs and assessed cell fate change over time. Interestingly, as early as 5 weeks after Pik3caH1047R expression in LCs, yellow fluorescent protein (YFP) was also detected in basal-like cells clustered around LCs (Fig. 3a–e and Extended Data Fig. 6), while, as previously described14, K8-CreERT2-targeted cells consist of a self-sustained unipotent population of LCs (Extended Data Fig. 6a–e). Clonal analysis of LCs expressing oncogenic Pik3caK1047R revealed the presence of bipotent clones containing adjacent LCs and BCs, which were never observed in YFP control LCs (Fig. 3f and Extended Data Fig. 6n–p). The relatively small proportion of K8⁺/K5⁺ BCs compared with K8⁺/K5⁻ BCs suggests that in the initial stage of LC-to-BC transition, these cells expressed markers of both lineages before maturing into basal-like cells and losing expression of LC markers (Fig. 3g and Extended Data Fig. 6q–t), which is consistent with the sequential gene expression shown by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) analysis of FACS-isolated BCs and LCs after Pik3caH1047R expression (Extended Data Fig. 6u, v). The proportion of YFP-expressing LCs increased over time, as well as the proportion of YFP⁺ BCs (Fig. 3e), suggesting that Pik3caH1047R confers a competitive advantage on luminal targeted cells. To determine functionally whether LCs acquired multipotency upon PIK3CA expression, we tested the ability of Pik3caH1047R-expressing LCs and their BC progeny to reconstitute the mammary gland upon transplantation into mammary fat pads. FACS-isolated LCs expressing Pik3caH1047R were able to form outgrowths of mammary epithelium containing both BCs and LCs (observed in 6 out of 28 transplants), while, as previously...
Figure 3 | Oncogenic Pik3ca expression induces multipotency in unipotent luminal and basal progenitors. a, b, Immunofluorescence of K8/YFP at 1 week (a) and K5/YFP at 5 weeks (b) after TAM administration to K8-CreERT2/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice. c, d, FACS analysis of CD24 and CD29 expression in Lin- YFP<sup>+</sup> cells 1 week (c) or 8 weeks (d) after TAM induction. e, Percentage of YFP<sup>+</sup> cells within LCs (CD29<sup>+</sup>/CD24<sup>+</sup>) and BCs (CD29<sup>+</sup>/CD24<sup>-</sup>) at different time points after TAM administration (n = 3, 3, 3, 4, 4 mice for 1 week, 5 weeks, 8 weeks, 4 months and 7 months, respectively). f, Immunofluorescence of K8/K5/YFP 8 weeks after clonal induction of K8-CreERT2/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice. Arrow points to K5/K8/Pik3ca<sup>H1047R</sup>/YFP<sup>+</sup> BC newly generated from a LC. g, Percentage of YFP<sup>+</sup> cells expressing K5 and/or K8 at different time points after Pik3ca<sup>H1047R</sup> expression in LCs (n = 3 mice per condition). See Methods for more details. h, i, Immunofluorescence of K5/K8 of a mammary outgrowth derived from LCs (h) or BCs (i) from K8-CreERT2/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice. j, k, Immunofluorescence of K5/YFP 1 week (j) or of K8/YFP 7 months (k) after TAM administration to K5-CreERT2/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice. l, m, FACS analysis of CD24 and CD29 expression in Lin- YFP<sup>+</sup> cells 1 week (l) or 7 months (m) after TAM administration. n, Percentage of YFP<sup>+</sup> cells within LCs and BCs at different time points after TAM administration (n = 3, 3, 5, 4, 3 mice for 1 week, 5 weeks, 8 weeks, 12 months and 12 months, respectively). o, Immunofluorescence for K8/K5/YFP 7 months after clonal Pik3ca<sup>H1047R</sup>/YFP expression in BCs. Arrow points to newly formed K8<sup>+</sup> YFP<sup>+</sup> LC arising from a BC. p, Mean number of colonies per 1,000 sorted luminal cells in an in vitro colony-forming assay of YFP<sup>+</sup> LCs derived from K5-CreERT2/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice induced for 12 months or wild-type LCs (n = 3 biologically independent experiments per condition). Circles, individual data points. Error bars, standard error of the mean (s.e.m.). Scale bars, 10 μm.

Figure 4 | Genes induced during oncogenic Pik3ca expression. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z. Scale bars, 10 μm.

Figure 5 | Multipotent progenitors. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z. Scale bars, 10 μm.

Figure 6 | Multipotent progenitors. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z. Scale bars, 10 μm.

Figure 7 | Multipotent progenitors. a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z, A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z. Scale bars, 10 μm.
In contrast, the luminal-to-basal multipotent signature was characteristic of a wounding and proliferation response marked by the upregulation of Il33 (also known as alarmin; a cytokine that has been shown to be overexpressed in breast cancers and attenuates NK response against tumour cells), Il24a, Krt16, Itga6b, Itga2, Itga5, Tnc, Cd109, Plau, Wnt10a, Timp3, Inhba, Ngf, Ereg, Ccdn1 and Ccdn2 (Fig. 4e). As found during BC-to-LC transition, most of the luminal-to-basal multipotent signature genes were specific for the newly formed BCs (for example, Ereg, Ccdn1, Wnt10a, Il33); a significant fraction of these genes (for example, Krt16, Il24a, Ccdn2, Inhba, Tnc) were already upregulated in LCs targeted by oncogenic Pik3ca, suggesting that they represent the signature of the LC of origin (Supplementary Tables 2, 3 and Extended Data Fig. 8).

To define the relevance of their multipotency gene signatures to tumour progression, we assessed the expression of these genes in Pik3ca\(^{H1047R}\)-derived tumours. Some luminal-to-basal multipotent signature genes such as Il24a, Krt16 and Plau were only upregulated during the initial stage of reprogramming and downregulated thereafter, while other genes such as Col11a1, the epidermal growth factor receptor (EGFR) ligand Erbg, Inhba, Wnt10a and Tnc continued to be expressed, or even further increased, in basal-like breast cancers arising from LCs (Fig. 4e, f). Similarly, Ntrk2 and Ntrk3 were expressed or even further upregulated in K5-creER\(^{1/2}\)/Pik3ca\(^{H1047R}\)-derived luminal tumours (Fig. 4f). These data indicate that some of the genes associated with cell fate transition during the early steps of tumour initiation increase with tumour progression.

To define the relevance of the Pik3ca\(^{H1047R}\)-induced multipotent gene signatures in human breast cancers, we assessed whether the different multipotent signatures correlated with a particular molecular breast cancer subtype or disease-free survival in a cohort of systematically untreated breast cancer patients. Interestingly, the luminal-to-basal transition gene signature was strongly associated with basal-like breast cancers (Fig. 4g). Higher expression levels of this gene signature or individual genes such as NGF, INHBA, ITGB6 and WNT10A were associated with poor clinical outcome (Fig. 4h and Extended Data Fig. 9), consistent with the more aggressive tumour types induced by Pik3ca\(^{H1047R}\) expression in LCs. In contrast, the BC-to-LC fate signature was associated with luminal A and normal-like human breast cancers (Fig. 4i). High gene expression levels of this gene signature were significantly associated with better prognosis (Fig. 4i), consistent with the less aggressive tumours arising from BCs. These data indicate that the genetic program associated with Pik3ca\(^{H1047R}\)-induced multipotency correlated with distinct molecular subtypes of human breast cancers and their levels of expression correlated with distinct clinical outcome.

Our study shows that the cell of origin controls tumour heterogeneity in Pik3ca\(^{H1047R}\)-induced mammary tumours. Pik3ca\(^{H1047R}\) expression in LCs gives rise to aggressive basal-like tumours while expression in BCs gives rise to less aggressive luminal-like tumours. We demonstrate that Pik3ca\(^{H1047R}\)-induced multipotency in unipotent progenitors. The promotion of multipotency induced by Pik3ca\(^{H1047R}\) is regulated by common and cell-lineage-specific molecular mechanisms that are influenced by the cellular origin in which the oncogene is initially expressed, setting the stage for future tumour heterogeneity and influencing clinical outcome in patients with breast cancers.

**Figure 4** | Molecular characterization of oncogenic Pik3ca-induced multipotency. a, Hierarchical gene expression clustering of BCs and LCs with or without Pik3ca\(^{H1047R}\) expression. Green and red correspond to high and low expressed genes, respectively. The two major branches of the tree are supported by bootstrap values of 100, a, induced for 10–12 months; y, induced for 8 weeks. b, Venn diagram of upregulated genes (>1.5 fold) after Pik3ca\(^{H1047R}\) expression in BCs and LCs. c, e–g, qRT–PCR analysis of genes belonging to the common (c), basal-to-luminal (d), or luminal-to-basal multipotency signature (e) in B-KSPIK and L-KSPIK cell population, 8 weeks and 10–12 months after Pik3ca\(^{H1047R}\) expression, respectively, compared with their age-matched controls. Gene expression was normalized to Gapdh housekeeping gene (n = 4 biologically independent samples). f, qRT–PCR analysis of the multipotency signature genes in control cells, in BC-derived adenomyoepithelium and in LC-derived metastatic tumours. Data were normalized to gene expression in age-matched control LCs (L-K8YFPo) (n = 4 biologically independent samples). g, i, Expression levels of the luminal-to-basal (g) or basal-to-luminal (i) multipotency signature in a large set of breast cancer patients according to their PAM50 subtype. Lum, luminal. h, j, Disease-free survival in untreated patients according to the level of expression of the genes of the luminal-to-basal (h) or basal-to-luminal (j) multipotency signature. k, n, Summary of the role of the cancer cell of origin in regulating Pik3ca\(^{H1047R}\)-induced tumour heterogeneity. k, Expression of Pik3ca\(^{H1047R}\) in BCs gives rise to luminal-like tumours, while in LCs Pik3ca\(^{H1047R}\) gives rise to more heterogeneous and aggressive tumours. Types of carcinoma are noted at the bottom of the panel. Adenomyo, adenomyoepithelium. l, Additional p53 deletion promotes Pik3ca\(^{H1047R}\)-induced tumour heterogeneity in BCs and leads to more aggressive metastatic carcinoma in LCs. m, n, Model of Pik3ca\(^{H1047R}\)-induced multipotency in LCs and BCs. Genes shown are upregulated during cell fate change. Genes highlighted in blue belong to the common multipotency signature. Error bars, s.e.m.

Supplementary Information is available in the online version of the paper.

Acknowledgements C.B. is an investigator of WELBIO. A.V.K. is Chercheur Qualifié of the FNRS. C.S. is Maître de Recherche of the FNRS. M.Y.L. is supported by the Agency for Science, Technology and Research (A*STAR, Singapore) fellowship. M.O. and A.W. are supported by FNRS fellowships. R.R.G. is supported by a TELEVIE fellowship and S.B. is supported by the foundation “Amis de l’institut Jules Bordet”. The Center for Microscopy and Molecular Imaging is supported by the European Regional Development Fund and Wallonia. W.A.P. is supported by project grants from the National Health and Medical Research Council of Australia. This work was supported by the FNRS, TELEVIE, a research grant from the Fondation Contre le Cancer, the ULB fondation, the Fond Yvonne Böll, the Fond Gaston Ithier, the foundation Bettencourt Schueller, the foundation Baillet Latour, and the European Research Council.

Author Contributions C.B. and A.V.K. designed the experiments and performed data analysis. A.V.K., M.Y.L. and M.O. performed all the experiments. S.B. and C.S. performed technical support for cell sorting. A.W. and R.R.G. helped with some experiments. W.A.P. performed the histological classification of mouse mammary tumours with regard to their cancer expression and gene amplification on human samples. S.R. and I.S. helped to analyse the data. M.Y.L., M.O. and S.R. helped to design the experiments. D.O. helped to identify the sequence available at accession number GSE69290. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper.
METHODS

Mice. Rosa26-YFP mice were obtained from the Jackson laboratory. K5-CreER<sup>2</sup>, and K8-CreER<sup>2</sup> mice were described previously<sup>4</sup>. K14-rtTA/TetO-Cre/Rosa26-YFP, K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP, K8-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP, K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP, K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP, K8-CreER<sup>2</sup>/Rosa26-YFP, K5-CreER<sup>2</sup>/Rosa26-YFP and K8-CreER<sup>2</sup>/Rosa26-YFP mice were induced with 15 mg of tamoxifen (TAM) (Sigma; diluted in sunflower seed oil, Sigma) by intraperitoneal injection (3 injections of 5 mg every 3 days). TAM administration induced a transient delay in mammary gland development during puberty but there was no long-term effect on mammary gland development and homeostasis<sup>35</sup>. Five-week-old K14-rtTA/TetO-Cre/Rosa26-YFP mice were induced by oral administration of doxycycline food diet (1 g kg<sup>-1</sup>; BIO-SERV) for 5 days. For clonal analyses, 4- to 5-week-old K8-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP and K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice were respectively induced with 0.05 mg or 2 mg TAM by intraperitoneal injection. For induction in adult mice, 8-week-old K8-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice were induced with 15 mg of TAM by intraperitoneal injection (3 injections of 5 mg every 3 days).

Histology and immunostaining on sections. For immunofluorescence, dissected inguinal mammary glands or tumour samples were pre-fixed for 2 h in 4% paraformaldehyde at room temperature. Tissues were washed three times with PBS for 5 min and incubated overnight in 30% sucrose in PBS at 4 °C. Tissues were embedded in OCT and kept at −80 °C. Sections of 5 μm were cut using a HM550 Microm cryostat (Microm Instruments). For immunofluorescence, tissue sections were incubated in blocking buffer (5% horse serum, 1% BSA, 0.2% Triton-X in PBS) for 1 h at room temperature. The different primary antibodies combinations were incubated overnight at 4 °C. Sections were then rinsed three times for 5 min in PBS and incubated with corresponding secondary antibodies diluted at 1:400 in blocking buffer for 1 h at room temperature. The following primary antibodies were used: anti-GFP (rabbit, 1:1,000, Abcam), anti-K8 (rat, 1:1,000, Abcam), anti-CD24 (1:50, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1:100, clone 500-259, BD Biosciences), anti-CD45 (1:750) for 16 min. Finally, slides were incubated with the OmniMap HRP-conjugated anti-rabbit antibody (Ventana) for 12 min. Standard ABC kit, and ImmPACT DAB (Vector Laboratories) were used for the detection of HRP activity. Nuclear staining was done with Mayer’s Hematoxylin (Labonord), followed by dehydration and mounting with SafeMount (Labonord).

Whole-mount mammary gland immunofluorescence. For clonal analyses, dissected inguinal mammary glands were incubated in 2 ml HBSS plus 30 μM i.mel1 collagenase plus 300 μg ml<sup>-1</sup> hyaluronidase (Sigma) for 30 min at 37 °C under agitation. After three washes of 5 min with HBSS, mammary glands were fixed in 4% paraformaldehyde for 2 h at room temperature, washed three times for 10 min in PBS under agitation and incubated in blocking buffer (5% horse serum, 1% BSA, 0.8% Triton-X in PBS) for 3 h at room temperature. The primary antibody combination, diluted in the blocking buffer, was incubated overnight at room temperature under agitation. Samples were washed three times for 10 min in PBS/0.2% Tween-20 and incubated in secondary antibodies diluted in the blocking buffer for 5 h under agitation. Cell nuclei were stained with Hoechst for 30 min (1:1,000 in PBS/0.2% Tween-20). Samples were mounted on slides in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Staining on human breast cancer sections. Tissue samples were obtained retrospectively from archival formalin-fixed and paraffin-embedded samples in the Department of Pathology of the Erasme Hospital. Histopathological diagnoses were reviewed and assessed according to the 2012 World Health Organization Classification. Sections of 5 μm were subjected to standard immunohistochemistry (IHC) as previously described<sup>4</sup> using respectively monoclonal anti-CK8/18 (1:200; Clone SD3; BioGenex); anti-CK14 (1:100; clone LL02; Leica) and anti-p63 (1:200; clone 7UL; Leica) antibodies. Staining was visualized with streptavidin-biotin-peroxydase complex kit reagents (BioGenex) using diaminobenzidine/H<sub>2</sub>O<sub>2</sub> as the chromogenic substrate. Counterstaining with haematoxylin confluenced the processing. Nuclear staining was done with Mayer’s Haematoxylin (Labonord), followed by dehydration and mounting with SafeMount (Labonord).

Microscope image acquisition. Pictures were acquired on an Axio Observer Z1 Microscope using ×10 and ×40 Zeiss EC Plan-NEOFUAR objectives, with an AxioCam MRc camera and using the Axiosvision software (Carl Zeiss). Confocal images in Fig. 3f, h, i and Extended Data Fig. 6–h, l–o and Extended Data Fig. 7, m were acquired at room temperature using a Zeiss LSM780 multiphoton confocal microscope fitted on an Axiovert M200 inverted microscope equipped with C-Apochromat (×40 = 1.2 numerical aperture) water immersion objectives (Carl Zeiss). Optical sections of 1,024 × 1,024 pixels, were collected sequentially for each fluorochrome. The data sets generated were merged and displayed with the ZEN software.

Mammary gland and tumour cell dissociation. Mammary glands were dissected and lymph nodes removed. Tissues were briefly washed in HBSS, and chopped with a McIlwain tissue chopper. Chopped tissues were placed in HBSS plus 300 U ml<sup>-1</sup> collagenase (Sigma) plus 300 μg ml<sup>-1</sup> hyaluronidase (Sigma) and digested for 2 h at 37 °C under agitation. Physical dissociation using a P1000 pipette was done every 15 min throughout the enzymatic digestion duration. EDTA at a final concentration of 5 mM was added for 10 min to the resultant suspension before filtration through a 70 μm mesh, two successive washes in 2% PBS/PBS and antibody labelling.

Cell labelling, flow cytometry and sorting. Two- to five-million cells per condition were incubated in 250 μl 2% PBS/PBS with fluorochrome-conjugated primary antibodies for 30 min, vortexing every 10 min. Cells were washed with 2% PBS/ PBS and were resuspended in 2.5 μg ml<sup>-1</sup> 4′,6-diamidino-2-phenylindole (DAPI; Invitrogen) before analysis. Primary antibodies used were: PE-Cy7-conjugated anti-CD24 (1:50, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1:50, clone eBioHMMb1-1, eBiosciences), PE-conjugated anti-CD45 (1:50, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1:50, clone MEC 13.3, BD Biosciences), PE-Cy7-conjugated anti-CD105 (1:50, clone APAS. eBiosciences). Data analysis and cell sorting were performed on a FACS Aria sorter using the FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; CD45-, CD31- and CD140a-positive cells were excluded (Lin<sup>+</sup>) before analysis of the FFP<sup>+</sup> cells. For profile analysis, a minimum of 1,000 FFP<sup>+</sup> cells were analysed per sample.

Tumour harvesting and classification. Tumours were detected by mammary gland palpation. Mice were killed when one tumour reached a maximum of 1 cm diameter. The K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice presented 1 tumour in 58%, 2 tumours in 25%, and 3 or more tumours in 17% of the cases at the time of analysis (11 mice were analysed). K8-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/Rosa26-YFP mice presented 1 tumour in 64%, 2 tumours in 18%, and 3 or more tumours in 18% of the cases at the time of analysis (a total of 17 tumours from 11 mice were analysed). The K5-CreER<sup>2</sup>/Pik3ca<sup>H1047R</sup>/p53<sup>+/−</sup>/ Rosa26-YFP mice presented 1 tumour in 83%, and 3 tumours in 17% of the cases.
at the time of analysis (a total of 8 tumours from 6 mice were analysed). The K14
rtTA/TetO-Cre/Pik3ca\H1047R\p53\p53+/\Rosa26-YFP mice presented 1 tumour in 57%, 2 tumours in 36%, and 3 or more tumours in 7% of the cases at the time of anastyis (a total of 19 tumours from 14 mice were analysed). The K8-CreERT2/Pik3ca\H1047R\p53\p53+/\Rosa26-YFP mice presented 1 tumour in 20%, 2 tumours in 45%, and 3 or more tumours in 35% of the cases at the time of analysis (a total of 40 tumours from 20 mice were analysed). The K8-CreERT2/Pik3ca\H1047R\p53\p53+/\Rosa26-YFP mice presented 1 tumour in 71%, and 2 tumours in 29% of the cases at the time of analysis (a total of 22 tumours from 17 mice were analysed). For each harvested tumour, the tumour was cut in three pieces, one for paraffin embedding, one for OCT embedding, and one for cell sorting and RNA extraction. Tumour classification was done based on histological features.

Mammary colony-forming assay. Luminal YFP+ cells from K5-CreERT2/Pik3ca\H1047R\Rosa26-YFP mice induced for 12 months were flow-sorted as a single-cell suspension based on their Lin−CD29hiCD24+24+2 profile. Control YFP+ luminal cells from K5-CreERT2/Rosa26-YFP induced for 12 months were sorted based on their CD29hiCD24+24+2 profile. Luminal cells were cultured with irradiated NIH 3T3 feeder cells in Mouse-Epicult B media (Stem Cell Technologies) supplemented with 10 ng ml−1 epidermal growth factor (Sigma-Aldrich), 10 ng ml−1 basic fibroblast growth factor (R&D Systems), 4 μg ml−1 heparin (Sigma-Aldrich), 1 mg ml−1 bovine serum albumin (BSA; Sigma-Aldrich), 5% FBS (Life Technologies), 50 units ml−1 penicillin and 50 μg ml−1 streptomycin (Life Technologies), as previously described29. After 1 week, colonies were fixed with methanol, stained with Giemsa stain (Sigma-Aldrich) and counted manually.

Mammary fat pad transplantation and analysis. Eight thousand LCs from K8-CreERT2/Pik3ca\H1047R\Rosa26-YFP or control K8-CreERT2/Rosa26-YFP or 1,350 BCs from K8-CreERT2/Pik3ca\H1047R\Rosa26-YFP or control K14-rtTA/TetO-Cre/Rosa26-YFP induced for 4 months were sorted based on their Lin−YFP+CD29hiCD24+24+2 or Lin−YFP+CD29hiCD24+24+2 profiles. LCs were resuspended in 10 μl DMEM plus 50% bovine serum. BCs were sorted in the presence of 10 μM of Rock inhibitor (Y27632, Sigma) and resuspended in 75% DMEM/25% matrigel. Cell suspension was injected into the fourth mammary gland of 3- to 4week-old NOD-SCID mice that had been cleared of endogenous epithelium as previously described30,31. Recipient mice were mated 4 weeks after the transplantation, and killed 2-to-3 weeks later, when fully pregnant. Recipient glands were dissected and stained for GFP, K5 and K8 as whole mounts. An outgrowth was defined as an epithelial structure comprising ducts and lobules and/or terminal end buds.

Quantification of keratin+ cells within YFP+ cells. A total of 1,907, 1,704 and 2,391 YFP+ cells from three different mice per condition were analysed respectively in K8-CreERT2/Rosa26-YFP induced 4 weeks, K8-CreERT2/Pik3ca\H1047R\Rosa26-YFP induced 1 week and 8 weeks on 5 μm cryosections stained for K5, K8 and BCL2. Coexpression of these markers was analysed with a confocal microscope. Cells were scored as K5+ (K8), K5+ K8+ (K5K8) or K5+ K8+ (K5K8) and are shown in Fig. 3g.

Quantification of clone composition. Mammary glands were processed as whole mount and stained for K8, K5 and GFP. Clones were analysed by confocal microscopy. A total of 822, 936, 714 and 360 clones from three independent mice per condition were analysed in K8-CreERT2/Pik3ca\H1047R\Rosa26-YFP induced 1 week, induced for 10 weeks, and in K5-CreERT2/Pik3ca\H1047R\Rosa26-YFP induced for 1 week and induced for 7 months respectively at dose of TAM that labelled very few and isolated clones. The clones were scored in three classes according to their keratin expression: luminal clones, composed only of K8+ cells, basal clones, composed only of K5+ cells, and mixed clones, composed of K5+ and K8+ cells. These data are provided in Extended Data Fig. 6p and Extended Data Fig. 7j.

Quantification of percentage YFP-labelled cells. The percentage of YFP labelled cells within the luminal and basal populations was quantified by FACS. The luminal population was defined as the CD29hiCD24+ population and the basal population was defined as the CD29hiCD24+ population.

Whole-mount carmine staining. Whole-mount mammary fourth mammary glands were fixed in methanol Carnoy (60% methanol, 30% acetic acid, 10% chloroform) for at least 2 h and rehydrated in 70% ethanol, followed by water. Staining in carmine alun (Sigma) was done overnight and excess dye was rinsed with water. This is followed by incubation in 70%, 95%, 100% ethanol (1 h each) and fast-clearing in toluene overnight. All steps were carried out at room temperature.

Epithelial outgrowth measurement. Carmine-stained mammary glands were photographed with a Leica M80 stereomicroscope equipped with a Leica IC80 HD digital camera. The distance from the lymph node of the mammary epitelium was scored by measuring the distance between the distal edge of the lymph node and the most distal tip of the epithelium.

RNA extraction and quantitative real-time PCR. The protocol used for RNA extraction on FACS-isolated cells has been previously described32. Briefly, RNA extraction was performed using the RNeasy micro kit (Qiagen) according to the manufacturer’s recommendations and DNase treatment. After nanodrop RNA quantification and analysis of RNA integrity, purified RNA was used to synthesize the first-strand cDNA in a 50 μl final volume, using Superscript II (Invitrogen) and random hexamers (Roche). Genomic contamination was detected by performing the same procedure without reverse transcriptase. Quantitative PCR analyses were performed with 1 ng of cDNA as template, using FastStart Essential DNA green master (Roche) and a Light Cycler 96 (Roche) for real-time PCR system.

Relative quantitative RNA was normalized using the housekeeping gene Gapdh. Primers were designed using PrimerBank database (http://pga.mgh.harvard.edu/primerbank/) and are listed in Supplementary Table 4. Analysis of the results was performed using Light Cycler 96 software (Roche) and relative quantification was performed using the ΔΔCt method using Gapdh as reference. The entire procedure was repeated in four biologically independent samples. For Extended Data Figs 6 and 7, data are shown as fold change over luminal cells or basal cells derived from 3-month-old wild-type mice (L-WT and B-WT).

Microarray analysis. Total RNA was analysed using mouse whole-genome MG-430 PM array from Affymetrix at the IRB Functional Genomics Core. All the results were normalized with RNA normalization using R-bioconductor package affy with standard parameters33. Two biologically independent samples were analysed for each condition, except for tumours derived from K5-CreERT2/Pik3ca\H1047R\Rosa26-YFP or K8-CreERT2/Pik3ca\H1047R\YFP, for which three and seven samples were analysed, respectively. Sorted BCs from K5-CreERT2/Rosa26-YFP mice induced for 8 weeks or 10–12 months, LCs from K8-CreERT2/Rosa26-YFP mice induced for 8 weeks or 10–12 months, BCS and LCs from K5-CreERT2/Pik3ca\H1047R\Rosa26-YFP mice induced for 10–12 months, BCS and LCs from K8-CreERT2/Pik3ca\H1047R\Rosa26-YFP mice induced for 8 weeks, Lin− cells from K5-CreERT2/Pik3ca\H1047R\Rosa26-YFP or BCs arising from K5-CreERT2/Rosa26-YFP mice. Only genes upregulated or downregulated by at least 1.5 fold were considered in the analysis.

Microarray data clustering. Clustering and bootstrap analyses were performed using the pvclust and gplots packages of the R statistical suite. Clustering was performed with the default parameters of the R hclust function (Euclidean distance and complete linkage) considering only the top 500 most variant genes among all experiments.

Gene expression comparison. Venn diagrams were computed with the R statistical tool. The reported hypergeometric P values for every comparison between two signatures correspond to the probability of an intersection of at least a given size by chance only, knowing the number of genes tested on a microarray chip.

Murine and human breast tumours gene expression profile comparison. To compare the murine tumour gene expression data to human tumour data, we used the METABRIC data set composed of 1,992 patients. METABRIC expression data were downloaded from the EBI website (data sets EGAD00010000210 and EGAD00010000211). When multiple probes mapped to the same Entrez gene identifier, we kept the one with the highest variance in the data set using the genefu package. The PAM50 subtypes were computed using the Bioconductor gene set package dedicated function34, (1,448 basal, 1,027 HER2+, 2,260 LumB, 22 LumA and 323 normal).

Boxplots and Kruskal–Wallis test P values were computed using R. P values reflect the probability that at least one of the cancer subtypes express the tested signature at a significantly different level.

Uni-directional Student’s t-test P values reflect the probability that one signature is significantly more expressed (or repressed) in one subtype compared to all the others. For the t-tests, as they are more robust to the extreme values, median and interquartile ranges were chosen as estimators of the central tendency and of the dispersion (instead of the mean and the standard deviation).

We then merged the murine data set with the METABRIC data set by keeping those genes described as orthologous in the Ensembl database downloaded via Biomart and having exactly the same identifier. Batch effect between murine and human data was corrected using the Combat function of the Bioconductor sva package. The PCA and clustering analyses were performed using the R statistical software considering an expression matrix containing only the expression values of the 46 PAM50 orthologous genes between mouse and human. For clustering, we used the Euclidean distance combined to the complete hierarchical clustering.
method (default parameters). PAM50 subtypes were computed using the R/Bioconductor genefu package.

**Survival analyses in humans.** Mouse-derived signatures were converted to human signatures by considering the orthologous genes in humans. Signatures score were then computed and re-scaled using the dedicated function of the R/Bioconductor genefu package. The scores were computed for each patients of the METABRIC together with those present in 33 other breast tumours reference data sets (7,220 patients). Survival curves were computed using the dedicated function of the genefu package only on untreated patients (1,859 cases) with available survival data. Expression level categories correspond to the tertiles of the expression values in the untreated patients. *P* values correspond to the log-rank *P* value, which reflects the probability that at least one of the class of signature expression presents a significantly differing outcome from the other classes.

Extended Data Figure 1 | Tamoxifen administration has no long-term effect on the mammary gland. a, b, Effect of TAM on mammary epithelial postnatal growth. a, Representative whole-mount preparations of carmine alum-stained mammary epithelium from the fourth mammary gland, showing that TAM induces a delay in mammary epithelium growth at early time points, but no difference is observed 8 weeks after TAM induction (a) and mean distance from lymph node distal edge to the distal epithelial edge 1 week, 5 weeks and 8 weeks after TAM injection or oil injection (b) (n = 6, 6, 4, 3, 5, 4 mice respectively for 1 week control (ctr), 1 week TAM, 5 weeks control, 5 weeks TAM, 8 weeks control, 8 weeks TAM). P value derived from two-sided Student’s t-test is 0.161, 0.035, 0.748 when comparing control and TAM conditions at 1 week, 5 weeks and 8 weeks, respectively. c, Percentage of YFP + cells in LCs (CD29Lo/CD24 Hi) and in BCs (CD29Hi/CD24 Lo) analysed by FACS 48 h after TAM administration in K5-CreERT2/Pik3caH1047R/Rosa26-YFP and K8-CreERT2/Pik3caH1047R/Rosa26-YFP, or 1 week after doxycycline administration to K14-rtTA/TetO-Cre/Pik3caH1047R/p53flox/flox/Rosa26-YFP mice (n = 5, 6, 3 mice respectively for K5-CreERT2, K8-CreERT2 and K14-rtTA/TetO-Cre). Circles, individual data points. Scale bars, 100 μm. Error bars, s.e.m.
Extended Data Figure 2 | Characterization of tumours derived from basal or luminal cells upon oncogenic Pik3ca expression. a–e, Characterization of adenomyoepithelioma (adenomyo) tumours derived from K5-CreERT2/Pik3ca^H1047R/Rosa26-YFP mice. f–y, Characterization of tumours derived from K8-CreERT2/Pik3ca^H1047R/Rosa26-YFP mice. f–j, Characterization of adenomyoepithelioma. k–o, Characterization of myoepithelial carcinoma (C). p–t, Characterization of invasive carcinoma of no special type (NST C). u–y, Characterization of metaplastic carcinoma. a, f, k, p, u, Haematoxylin and eosin staining. b, g, l, q, v, p63 immunohistochemistry. c, h, m, r, w, Immunofluorescence of ER/K8. d, i, n, s, x, Immunofluorescence of K8/K14. e, j, o, t, y, Mean percentage of Ki67^+ cells within tumours (n = 6, 3, 3, 1, 3 tumours and total number of cells counted = 10,408, 10,758, 11,174, 4,622, 5,732 in e, j, o, t, y, respectively). Error bars, s.e.m. Scale bars, 10 μm.
Histological classification of tumours shown:

- **K5PIK TU1** adenomyoepithelioma
- **K5PIK TU2** adenomyoepithelioma
- **K5PIK TU3** adenomyoepithelioma
- **K5PIK TU1** myoepithelial carcinoma
- **K5PIK TU2** invasive carcinoma NST
- **K5PIK TU3** metaplastic carcinoma with mesenchymal differentiation
- **K5PIK TU4** adenomyoepithelioma
- **K5PIK TU5** adenomyoepithelioma
- **K5PIK TU6** adenomyoepithelioma + myoepithelial carcinoma
- **K5PIK TU7** adenomyoepithelioma
- **K5PIKp53 TU1** metaplastic carcinoma with mesenchymal differentiation
- **K5PIKp53 TU2** metaplastic carcinoma with mesenchymal differentiation

---

![Image of histological classifications](image-url)
Extended Data Figure 3 | Similarities between mouse Pik3ca-derived mammary tumours and human breast cancers. a–d, Human breast tumour histologically classified as adenomyoepithelioma resembling K5-CreERT2/Pik3caH1047R/Rosa26-YFP-derived tumours (K5PIK TU). a, Haematoxylin and eosin staining. b–d, p63 (b), K8/K18 (c) and K14 (d) immunohistochemistry in the human adenomyoepithelioma. e–h, Human breast tumour histologically classified as metaplastic carcinoma resembling K8-CreERT2/Pik3caH1047R/Rosa26-YFP derived tumours (K8PIK TU). e, Haematoxylin and eosin staining. f–h, p63 (f), K8/K18 (g) and K14 (h) immunohistochemistry in the human metaplastic carcinoma. i–k, Principal component analysis (PCA) of the METABRIC patients together with murine tumours according to the expression values of the PAM50 genes common to mice and humans. i, PCA of three K5-CreERT2/Pik3caH1047R tumours (black dots) showing that these tumours cluster with human luminal B cancer subtype. j, PCA of seven K8-CreERT2/Pik3caH1047R/Rosa26-YFP-derived tumours (numbered black dots). Histological classification of each numbered tumour is described below the figure. k, PCA of two K8-CreERT2/Pik3caH1047R/p53fl/fl/Rosa26-YFP-derived tumours (K8PIKp53 TU) (black dots) showing that these tumours cluster together with human HER2+ subtype. l, Clustering of the murine tumours among human tumours of the METABRIC data set. Clustering was performed by grouping tumours presenting similar expression patterns of PAM50 genes. Colours on top of the heatmap represent the PAM50 subtypes attributed to the human tumours. The discrepancy between PCA and clustering analysis are due to the influence of HER2 low expression in these tumours, for which around 60% of PC2 relies on ERBB2 expression. Scale bars, 10 μm.
Extended Data Figure 4 | Gating strategy to analyse and isolate tumour cells, LCs, and BCs according to their YFP, CD29 and CD24 profile.

a-e, Dot plot FACS analysis of unicellular suspension of mammary tumour cells (in this example from K8-CreERT2/Pik3ca-H1047R/Rosa26-YFP tumour) stained for Lin (CD31, CD45, CD140a). Debris were eliminated from all events in P1 (a), doublets were discarded in P2 (b), the living cells were gated in P3 by DAPI dye exclusion (c), the non-epithelial Lin− cells were discarded in P4 (d), and the YFP+ cells were gated in P5 (e). f, Gating strategy used for FACS analysis and cell sorting, showing the proportion of parent and total cells for each gate. Tumour cells were isolated based on their Lin− profile for YFP+ tumours (P4 gate), or were isolated based on their YFP profile (P5 gate) for the YFP+ tumours, as described in Methods.

g-m, Dot plot FACS analysis of unicellular suspension of mammary cells (in this example from K5-CreERT2/Pik3caH1047R/Rosa26-YFP mice 12 months after TAM induction) stained for CD24, CD29 and Lin (CD31, CD45, CD140a). Debris were eliminated from all events in P1 (g), doublets were discarded in P2 (h), the living cells were gated in P3 by DAPI dye exclusion (i), the non-epithelial Lin− cells were discarded in P4 (j), and the YFP+ cells were gated in P5 (k). l, m, CD29 and CD24 expression were used to gate the CD29loCD24hi population, corresponding to LCs, and to gate the CD29hiCD24+ population, corresponding to BCs, either in YFP+ cells (l) or in Lin+ cells (m). n, Gating tree showing the gating strategy used for FACS analysis and sorting, showing the proportion of parent and total cells for each gate.
Extended Data Figure 5 | Characterization of tumours derived from BCs or LCs upon concomitant expression of oncogenic Pik3ca and deletion of p53. a–o, Characterization of tumours derived from K14-rtTA/TetO-Cre/Pik3caH1047R/p53fl/fl/Rosa26-YFP mice. a–e, Characterization of adenomyoepithelioma (adenomyo). f–j, Characterization of myoepithelial carcinoma. k–o, Characterization of metaplastic carcinoma. p–a’, Characterization of tumours derived from K8-CreERT2/Pik3caH1047R/p53fl/fl/Rosa26-YFP mice. p–t, Characterization of myoepithelial carcinoma (C). u–a’, Characterization of metaplastic carcinoma. a, f, k, p, u, Haematoxylin and eosin staining. b, g, l, q, v, p63 immunohistochemistry. c, h, m, r, w, Immunofluorescence of K8/ER. d, i, n, s, x, Immunofluorescence of K8/K14. e, j, o, t, y, Mean percentage of Ki67+ cells within tumours (n = 4, 3, 3, 4, 6 tumours and total cells counted = 11,903, 10,670, 6,992, 14,743, 8,172 in e, j, o, t, y, respectively). z, Immunofluorescence of K8/HER2. a’, Immunofluorescence of E-cadherin/vimentin. Error bars, s.e.m. Scale bars, 10 µm.
Extended Data Figure 6 | Oncogenic Pik3ca expression induces multipotency in unipotent luminal progenitors. a–d, Immunofluorescence showing the expression of K8/YFP (a, c) or K5/YFP (b, d) 1 week (a, b) and 7 months (c, d) after TAM injection in control K8-CreERT2/Rosa26-YFP mammary gland. e, Percentage of YFP+ cells in LCs (CD29Lo/CD241) and in BCs (CD29Hi/CD241) at different time points after TAM administration to K8-CreERT2/Rosa26-YFP mice (n = 3 mice per time point) showing that no YFP+ cells expressing CD29Hi/CD241 were detected in control K8-CreERT2/Rosa26-YFP mammary glands at any time point. f–h, Immunofluorescence of K14/YFP (f), p63/YFP (g), SMA/YFP (h) 8 weeks (f, g) or 10 weeks (h) after TAM administration to K8-CreERT2/Pik3caH1047R/Rosa26-YFP mice, shows that the BCs arising from LCs upon oncogenic Pik3ca targeting expressed these classical markers of BCs. i–m, Induction of Pik3caH1047R expression in LCs in adult mice. i–l, Immunofluorescence showing the expression of K8/YFP (i, k) or K5/YFP (j, l) 1 week (i, j) and 8 weeks (k, l) after TAM injection in K8-CreERT2/Pik3caH1047R/Rosa26-YFP mice induced in adulthood. m, Percentage of YFP+ cells in LCs (CD29Lo/CD241) and in BCs (CD29Hi/CD241) at different time points after TAM administration to K8-CreERT2/Pik3caH1047R/Rosa26-YFP mice induced in adulthood (n = 4 mice per time point). n, o, Immunofluorescence of K5/YFP showing the clonal YFP expression in a single isolated LC 1 week after TAM injection (n), and 8 weeks after TAM injection showing a clone that gave rise to an LC and a BC (o) in K8-CreERT2/Pik3caH1047R/Rosa26-YFP mammary gland. Arrow in n points to the isolated LC, while arrow in o points to the newly arisen BC. p, Distribution of clones 1 week or 10 weeks after TAM injection in K8-CreERT2/Pik3caH1047R/Rosa26-YFP at clonal dose. Clones were scored as composed of only luminal cells (luminal clones), composed of only basal cells (basal clones) or composed of luminal and basal cells (mixed clones) (n = 4 mice per time point). See Methods for more details. q–t, Immunofluorescence of K5/K8 (q), K5 (r), K8 (s) and K5/K8/YFP (t) shows that in wild-type mammary gland, K5 and K8 are not co-expressed (q), while K5/K8 double-positive cells are observed in K8-CreERT2/Pik3caH1047R/Rosa26-YFP mammary gland 8 weeks after oncogenic Pik3ca expression in LCs (r–t). Arrows in r–t point to K5+ K8+ YFP+ cells. u, v, RT–PCR analysis of luminal (u) or basal (v) genes in YFP+ LCs and BCs sorted from K8-CreERT2/Pik3caH1047R/Rosa26-YFP mice induced for 1 week, 4 weeks or 8 weeks, in YFP+ LCs derived from K8-CreERT2/Rosa26-YFP and in YFP+ BCs derived from K5-CreERT2/Rosa26-YFP mice induced for 8 weeks. Data for luminal genes are compared to adult wild-type LCs (u) while data for basal genes are compared to adult wild-type BCs (v) (n = 4 biologically independent samples per condition). Circles, individual data points. Scale bars, 10 μm. Error bars, s.e.m.
Oncogenic Pik3ca expression induces multipotency in unipotent basal progenitors. a–c, Immunofluorescence showing the expression of K5/YFP (a, b) or K8/YFP (c) at 1 week (a) and 7 months (b, c) in control K5-CreER\textsuperscript{12}/Rosa26-YFP mammary gland. d, Percentage of YFP\textsuperscript{+} cells in LCs (CD29\textsuperscript{lo}/CD24\textsuperscript{hi}) and in BCs (CD29\textsuperscript{hi}/CD24\textsuperscript{lo}) at different time points after TAM administration to K5-CreER\textsuperscript{12}/Rosa26-YFP (n = 5, 4, 4, 3 mice for 1 week, 8 weeks, 7 months and 12 months, respectively) showing that no YFP\textsuperscript{+} cells expressing CD29\textsuperscript{lo}/CD24\textsuperscript{hi} were detected in control K5-CreER\textsuperscript{12}/Rosa26-YFP mammary glands at any time point. e–h, Immunofluorescence of K19/YFP (e), ER/YFP (f), PR/YFP (g), claudin 3/YFP (h), 8 months after TAM administration to K5-CreER\textsuperscript{12}/Pik3ca\textsuperscript{H1047R}/Rosa26-YFP mice, shows that LCs arising from BCs upon oncogenic Pik3ca targeting expressed these classical markers of LCs. i, Immunofluorescence of K5/YFP showing the YFP expression in a single isolated BC 1 week after TAM injection at a clonal dose. Arrow points to the isolated BC. j, Distribution of clones 1 week or 7 months after TAM injection in K5-CreER\textsuperscript{12}/Pik3ca\textsuperscript{H1047R}/Rosa26-YFP at a clonal dose. Clones were scored as composed of only luminal cells (luminal clones), composed of only basal cells (basal clones) or composed of luminal and basal cells (mixed clones) (n = 3, 4 mice for 1 week and 7 months, respectively). See Methods for more details. k, l, RT–PCR analysis of luminal (k) or basal (l) genes in YFP\textsuperscript{+} LCs and BCs sorted from K5-CreER\textsuperscript{12}/Pik3ca\textsuperscript{H1047R}/Rosa26-YFP mice induced for 10–12 months, in YFP\textsuperscript{+} LCs derived from K8-CreER\textsuperscript{12}/Rosa26-YFP mice and in YFP\textsuperscript{+} BCs derived from K5-CreER\textsuperscript{12}/Rosa26-YFP mice induced for 10–12 months. Data for luminal genes are compared to adult wild-type LCs (k) while data for basal genes are compared to adult wild-type BCs (l) (n = 4 biologically independent samples per condition). m, Confocal microscopy analysis of immunofluorescence of YFP, TrkB and K8 of mammary glands 7 months after Pik3ca expression in BCs, showing that the newly formed LCs after Pik3ca expression in BCs co-expressed Nrtk2 and K8. Arrow points to formed K8\textsuperscript{+}/Nrtk2\textsuperscript{+}/YFP\textsuperscript{+} cell. Circles, individual data points. Scale bars, 10 μm. Error bars, s.e.m.
Extended Data Figure 8 | Molecular characterization of oncogenic *Pik3ca* induced multipotency. a, b, Venn diagram representing the common and distinct upregulated (a) and downregulated (b) genes in BCs and LCs after *Pik3ca* expression in BCs and LCs compared to age-matched control BCs and LCs, respectively, with the name of the list of genes and number of genes in each section. The list of genes in each Venn section is provided in Supplementary Tables 2 and 3. c–l, Venn diagrams representing the common genes upregulated (c, e, g, i, k) or downregulated (d, f, h, j, l) in the newly generated LCs or BCs after *Pik3ca*<sup>H1047R</sup> expression in unipotent progenitors (c, d); in LCs and in BCs after *Pik3ca*<sup>H1047R</sup> expression in LCs (genes regulated following the initial targeting of *Pik3ca*<sup>H1047R</sup> in LCs, and thus reflecting the LC of origin) (e, f); in LCs and in BCs after *Pik3ca*<sup>H1047R</sup> expression in BCs (genes regulated by *Pik3ca*<sup>H1047R</sup> in BCs, and thus reflecting the BC of origin) (g, h); in BCs after *Pik3ca*<sup>H1047R</sup> expression in LCs and in BCs (genes regulated by *Pik3ca*<sup>H1047R</sup> expression in BCs, irrespective of cell of origin) (i, j); in LCs after *Pik3ca*<sup>H1047R</sup> expression in BCs and in LCs (genes regulated by *Pik3ca*<sup>H1047R</sup> expression in BCs, irrespective of cell of origin) (k, l). Diameter of the diagram is proportional to the number of genes it contains. The reported hypergeometric *P* values correspond to the probability of observing an intersection of this size by chance only, knowing the number of genes tested on a microarray chip.
Extended Data Figure 9 | Genes of luminal-to-basal multipotency signature correlate with patient outcome in untreated breast cancer patients. a–d. Disease-free survival in untreated patients according to the level of expression (low = blue, intermediate = green or high = red) of the genes in the luminal-to-basal multipotency signature, namely NGF (a), INHBA (b), ITGB6 (c) and WNT10A (d), showing that genes of luminal-to-basal multipotency signature predict disease-free survival in untreated breast cancer patients. Patients expressing high levels of this signature are more prone to tumour relapse while those expressing lower levels of this signature show lower rates of relapse. The log-rank $P$ values account for the significance of this difference.
PIK3CA mutations are associated with distinct types of human breast cancers but the cellular origin and mechanisms responsible for this heterogeneity were unclear; here, using a genetic approach in mice, PIK3CA mutations are shown to activate a genetic program directing multiple cell fates in normally lineage-restricted cell types.