RESEARCH ARTICLE

Loss of Epigenetic Regulation Disrupts Lineage Integrity, Induces Aberrant Alveogenesis, and Promotes Breast Cancer

Ellen Langille[1,2], Khalid N. Al-Zahrani[1], Zhibo Ma[3], Minggao Liang[4], Liis Uuskula-Reimand[4], Roderic Espin[5], Katie Teng[1,2], Ahmad Malik[1,2], Helga Bergholtz[6], Samah El Ghamrasni[7], Somaieh Afiuni-Zadeh[1], Ricky Tsai[1], Sana Alvi[4], Andrew Elia[7], YiQing Lü[1,2], Robin H. Oh[1,2], Katelyn J. Kozma[2,4], Daniel Trcka[1], Masahiro Narimatsu[1], Jeff C. Liu[2], Thomas Nguyen[1,2], Seda Barutcu[1], Sampath K. Loganathan[1], Rod Bremner[1], Gary D. Bader[2], Sean E. Egan[2,4], David W. Cescon[7], Therese Sørlie[6,8], Jeffrey L. Wrana[1,2], Hartland W. Jackson[1,2], Michael D. Wilson[2,4], Agnieszka K. Witkiewicz[9], Erik S. Knudsen[9], Miguel Angel Pujana[5], Geoffrey M. Wahl[3], and Daniel Schramek[1,2]

ABSTRACT

Systematically investigating the scores of genes mutated in cancer and discerning disease drivers from inconsequential bystanders is a prerequisite for precision medicine but remains challenging. Here, we developed a somatic CRISPR/Cas9 mutagenesis screen to study 215 recurrent "long-tail" breast cancer genes, which revealed epigenetic regulation as a major tumor-suppressive mechanism. We report that components of the BAP1 and COMPASS-like complexes, including KMT2C/D, KDM6A, BAP1, and ASXL1/2 ("EpiDrivers"), cooperate with *PIK3CA*^{H1047R} to transform mouse and human breast epithelial cells. Mechanistically, we find that activation of *PIK3CA*^{H1047R} and concomitant EpiDriver loss triggered an alveolar-like lineage conversion of basal mammary epithelial cells and accelerated formation of luminal-like tumors, suggesting a basal origin for luminal tumors. EpiDriver mutations are found in ~39% of human breast cancers, and ~50% of ductal carcinoma *in situ* express casein, suggesting that lineage infidelity and alveogenic mimicry may significantly contribute to early steps of breast cancer etiology.

SIGNIFICANCE: Infrequently mutated genes comprise most of the mutational burden in breast tumors but are poorly understood. *In vivo* CRISPR screening identified functional tumor suppressors that converged on epigenetic regulation. Loss of epigenetic regulators accelerated tumorigenesis and revealed lineage infidelity and aberrant expression of alveogenesis genes as potential early events in tumorigenesis.

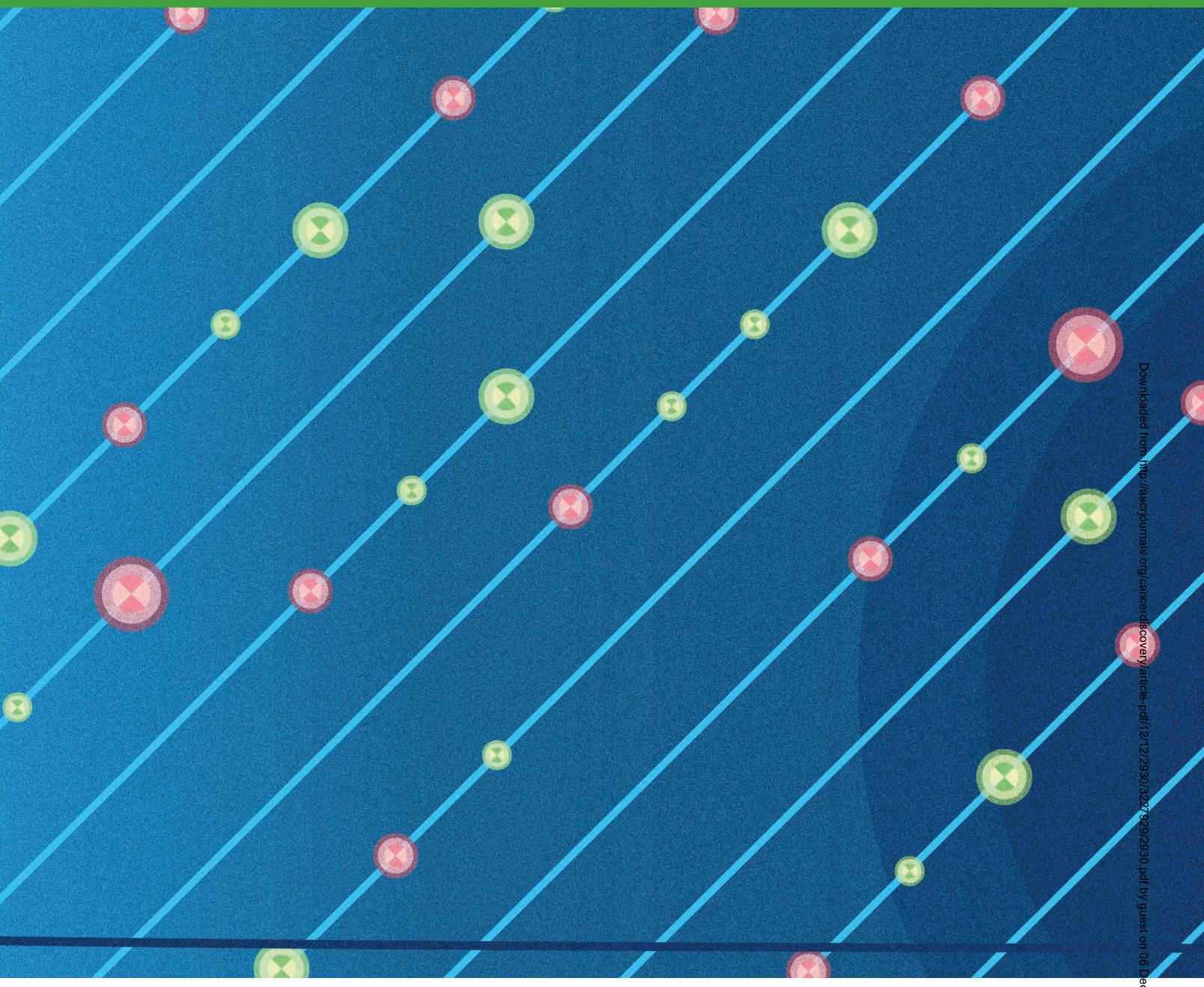
[1]Centre for Molecular and Systems Biology, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [3]Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, California. [4]Hospital for Sick Children, Toronto, Ontario, Canada. [5]Program Against Cancer Therapeutic Resistance (ProCURE), Catalan Institute of Oncology (ICO), Oncobell, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet del Llobregat, Barcelona, Spain. [6]Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. [7]Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. [8]Institute of Clinical Medicine, University of Oslo, Oslo, Norway. [9]Center for Personalized Medicine, Roswell Park Cancer Institute, Buffalo, New York.

Note: K.N. Al-Zahrani and Z. Ma contributed equally to this article.

Corresponding Author: Daniel Schramek, Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. Phone: 416-586-4800; Fax: 416-586-8869; E-mail: schramek@lunenfeld.ca

Cancer Discov 2022;12:2930–53

doi: 10.1158/2159-8290.CD-21-0865



INTRODUCTION

New genomic technologies hold the promise of revolutionizing cancer therapy by allowing treatment decisions guided by a tumor's genetic makeup. However, converting genetic discoveries into tangible clinical benefits requires a deeper understanding of the molecular and cellular mechanisms that underlie disease progression (1). In breast cancer, only a few genes, such as *TP53* and *PIK3CA*, are mutated at high frequencies (~30%–50%), whereas the vast majority are mutated at low frequencies comprising a so-called long-tail gene distribution (2–4). Whether these long-tail genes functionally contribute to breast cancer progression constitutes a significant knowledge gap. Although mutations in these genes seem to be under positive selection, they are only found in relatively small subsets of patients. It has been proposed that these infrequently mutated genes individually confer a small fitness advantage to cancer cells, but when combined synergize to increase fitness (additive-effects model; refs. 5–7). Alternatively, long-tail genes may work in different ways to produce the same phenotype (phenotypic convergence) and/or affect the same pathway or molecular mechanism (pathway convergence; ref. 8). Recently, we reported the latter mechanism in head and neck cancer, in which long-tail genes converge to inactivate NOTCH signaling (9). The biological relevance of long-tail genes in other cancer types remains largely unknown.

Here, we report an *in vivo* CRISPR/Cas9 screening strategy to identify which long-tail breast cancer genes and associated

molecular pathways cooperate with the oncogenic *Pik3ca*^{H1047R} mutation to accelerate breast cancer progression.

We tested 215 long-tail genes and identified several functionally relevant breast cancer genes, many of which converge on regulating histone modifications and enhancer activity (from here onward referred to as "EpiDrivers"). Single-cell multiomics profiling of EpiDriver-mutant mammary glands reveals increased cell-state plasticity and alveogenic mimicry associated with an aberrant alveolar differentiation program during the early specification of luminal breast cancer. Interestingly, EpiDriver loss in basal cells triggers basal-to-alveolar lineage conversion and accelerated tumor formation. Importantly, EpiDriver mutations are found in ~39% of primary breast tumors, supporting the hypothesis that different genes converge to produce the same cell plasticity that facilitates cancer development.

RESULTS

Direct *In Vivo* CRISPR Gene Editing in the Mouse Mammary Gland

First, we developed a multiplexed CRISPR/Cas9 knockout approach in the mammary gland of tumor-prone mice. As *PIK3CA* is the most commonly mutated oncogene in breast cancer, we crossed conditional Lox-Stop-Lox-(LSL)-*Pik3ca*^{H1047R} mice to LSL-*Cas9-GFP* transgenic mice to generate *Pik3ca*^{HR};*Cas9* mice. Intraductal microinjections of a lentivirus that expresses a single-guide RNA (sgRNA) and Cre recombinase (LV-sgRNA-Cre) led to the excision of LSL cassettes and expression of *Cas9*, *GFP*, and oncogenic *Pik3ca*^{HR} in the mammary epithelium (Fig. 1A). We tested the efficacy of CRISPR/Cas9-mediated mutagenesis by injecting sgRNAs targeting *GFP* or the heme biosynthesis gene *Urod*. Knockout of *GFP* was detected as a 86% ± 6% reduction in green fluorescence in transduced cells, whereas knockout of *Urod* was detected as an accumulation of unprocessed fluorescent porphyrins in 30% ± 8% of cells (Supplementary Fig. S1A–S1D; ref. 10). Moreover, *Pik3ca*^{HR};*Cas9* mice transduced with an sgRNA targeting *Trp53* developed tumors significantly faster than littermate mice transduced with a control sgRNA targeting the permissive *Tigre* locus (median tumor-free survival of 83 vs. 152 days; Supplementary Fig. S1E). Together, these data demonstrate that this approach recapitulates cooperation between oncogenic *Pik3ca* and *Trp53* loss of function (11, 12) and can be used to test for genetic interaction between breast cancer genes.

CRISPR Screen Identifies Histone Modifiers as Breast Cancer Driver Genes

In breast cancer, 215 long-tail genes show somatic mutations in 2% to 20% of patients (11, 13). To assess disease relevance of these genes *in vivo*, we established an LV-sgRNA-Cre library targeting the corresponding mouse orthologs (four sgRNAs/gene; 860 sgRNAs) as well as a library of 420 nontargeting control sgRNAs (sgNT; Supplementary Table S1). We optimized the parameters for an *in vivo* CRISPR screen by using a mixture of lentiviruses expressing GFP or RFP to determine the viral titer that transduces the mammary epithelium at clonal density (multiplicity of infection <1). Higher viral titers were associated with double infections, whereas a 15% overall transduction level minimized double infections while generating sufficient clones to screen (Supplementary Fig. S1F–S1I). Flow cytometry revealed that the third and fourth mammary glands each contain >3.5 × 10^5 epithelial cells, and that EPCAM^{hi}/CD49f^{mid} luminal cells showed a higher infectivity (~30%) compared with EPCAM^{mid}/CD49f^{hi} basal cells (~5%; Fig. 1B; Supplementary Fig. S1H and S1I). Thus, at a transduction level of 15% and a pool of 860 sgRNAs, each sgRNA was predicted to be introduced into an average of 60 individual cells within a single gland.

To uncover long-tail genes that cooperate with oncogenic PI3K signaling, we introduced the viral libraries into the third and fourth pairs of mammary glands of 19 *Pik3ca*^{HR};*Cas9* mice, resulting in an overall coverage of >4,000 clones per sgRNA. Next-generation sequencing confirmed efficient lentiviral transduction of all sgRNAs (Supplementary Fig. S2A). Importantly, *Pik3ca*^{HR};*Cas9* mice transduced with the long-tail breast cancer sgRNA library developed mammary tumors significantly faster than littermates transduced with the control sgRNA library (74 vs. 154 days; $P < 0.0001$; Fig. 1C). This result was similar to the accelerated tumorigenesis caused by loss of *Trp53* (Supplementary Fig. S1E), indicating the existence of strong tumor suppressors within the long-tail of breast cancer–associated genes.

We examined the sgRNA representation in 146 tumors to determine the targets responsible for accelerating mammary tumorigenesis. Most tumors showed strong enrichment for a single or occasionally two sgRNAs (Supplementary Fig. S2B). We prioritized genes that were targeted by ≥2 sgRNAs and knocked out in multiple tumors, resulting in 29 candidate tumor suppressor genes (Supplementary Table S2). These candidates included well-known tumor suppressors, such as *Apc* or *Nf1*, as well as genes with poorly understood function, such as *Arhgap35* (14). Intriguingly, several genes encoding histone and DNA modifying enzymes were identified, such as *Arid5b*, *Asxl2*, *Kdm6a* (*Utx*), *Kmt2a* (*Mll1*), *Kmt2c* (*Mll3*), and *Kmt2d* (*Mll4*), indicating a convergence on epigenetic regulation (Fig. 1D; Supplementary Fig. S2C).

KDM6A, KMT2C, ASXL2, BAP1, SETD2, and APC Suppress Breast Cancer in Mice

KMT2C and *KMT2D* encode partly redundant histone methyltransferases within the "complex of proteins associated with SET1" (COMPASS)–like complex, which also contains the histone demethylase KDM6A. The KMT2C/D–COMPASS-like complex catalyzes the monomethylation of lysine 4 as well as demethylation of lysine 27 in histone H3 (H3K4me1/H3K27) at distal enhancers, facilitating recruitment of the CBP/p300 H3K27 histone acetylase (HAT), which ultimately primes enhancers for gene activation (15, 16). The KMT2C/D–COMPASS-like complex is recruited to enhancers by the BAP1–ASXL1/2 complex, which facilitates enhancer priming (17, 18). In addition, the methyltransferase SETD2 deposits H3K36me3 marks at active enhancers and transcribed gene bodies (19, 20). Thus, our top hits converge on regulating enhancer function (Fig. 1E).

We validated each hit by injecting *Pik3ca*^{HR};*Cas9* mice individually with one sgRNA from the library and one newly designed sgRNA targeting *Asxl2*, *Kdm6a*, *Kmt2c*, and *Setd2* (termed EpiDrivers) or *Trp53* and *Apc*. We also transduced

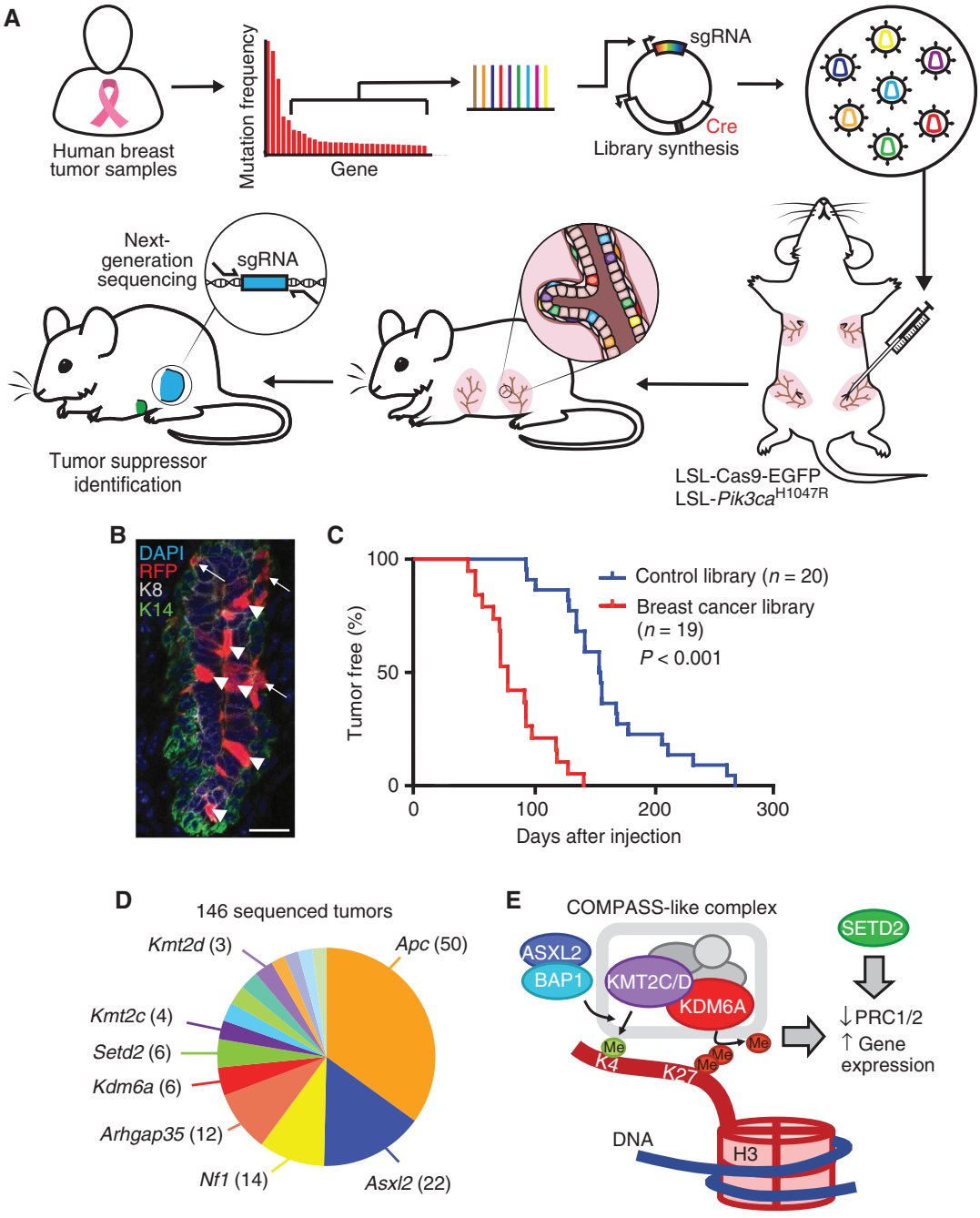
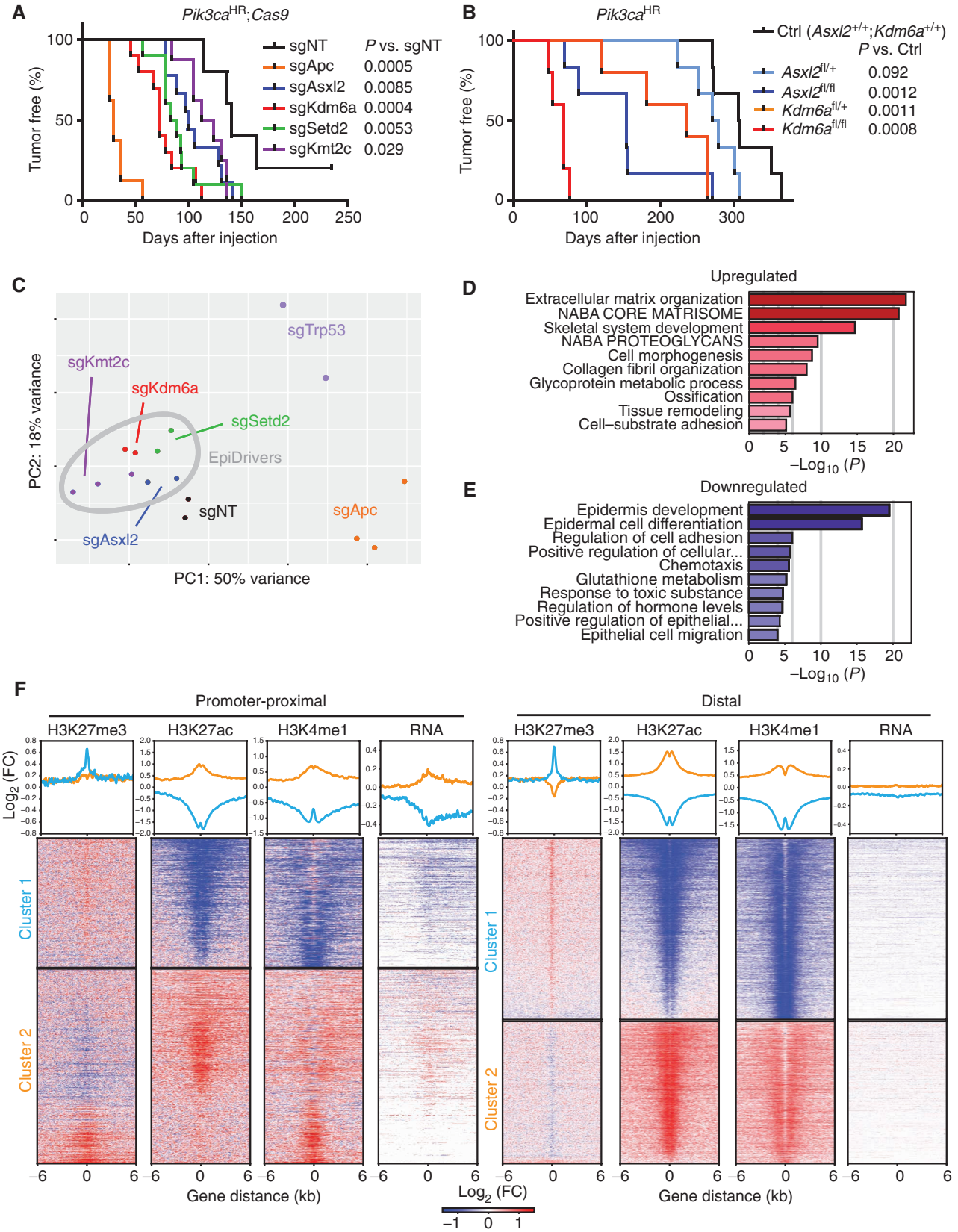


Figure 1. *In vivo* CRISPR screen reveals novel epigenetic breast cancer tumor suppressors, "EpiDrivers." **A,** Experimental design for *in vivo* CRISPR screen showing gene selection from long-tail mutations, intraductal injection of lentiviral libraries, and tumor sequencing. **B,** Mammary epithelium transduced with lentiviral RFP. Arrows denote basal cells, and arrowheads denote luminal cells. Scale bar = 25 μm. **C,** Tumor-free survival of *Pik3ca*^{HR};*Cas9* mice transduced with an sgRNA library targeting putative breast cancer genes or a control sgRNA library. **D,** Pie chart showing putative tumor suppressor genes with enriched sgRNAs in tumor DNA (the number of tumors is denoted in parentheses). **E,** Schematic of COMPASS-like and ASXL/BAP1 complexes in epigenetic control of gene expression.

mice with sgRNAs targeting *Asxl1* and *Bap1*, which were not in the original library. All transduced mice developed multiple highly proliferative breast tumors with much shorter latencies than mice transduced with nontargeting control sgRNAs (sgNT; Fig. 2A; Supplementary Fig. S2D and S2E). All tested tumors harbored biallelic frameshift mutations in the target genes, and Western blot analysis confirmed loss of APC, ASXL2, KDM6A, and p53 expression (Supplementary Fig. S2F–S2K).

Histologically, control tumors and *Asxl2*-, *Kmt2c*-, and *Kdm6a*-mutant tumors presented mostly as invasive ductal carcinoma (IDC) usually with glandular and some papillary differentiation. *Trp53*- and *Apc*-mutant tumors presented mostly as squamous or basal-like tumors. Detailed analysis by



mouse tumor pathologists revealed further glandular, squamous, mixed squamous/glandular (adenomyoepithelioma), or spindle cell differentiation patterns consistent with published reports of *Pik3ca*^{HR}-induced mouse mammary tumors (refs. 12, 21; Supplementary Fig. S3A–S3C and Supplementary Table S2). All tumors were estrogen receptor–positive and recapitulated gland morphology with cells marked by basal keratin 14 (K14) or luminal keratin 8 (K8). The *Trp53*-mutant tumors showed an increased proportion of K14/K8 double-positive cells, which were also seen in invasive microclusters of EpiDriver-mutant tumors (Supplementary Fig. S3D–S3H).

Next, we transduced the mammary epithelium of *Kdm6a*$^{fl/fl}$;*Pik3ca*$^{HR/+}$ and *Asxl2*$^{fl/fl}$;*Pik3ca*$^{HR/+}$ mice with lentiviral Cre and observed significantly accelerated tumor formation (68 and 154 days vs. 308 days for *Pik3ca*$^{HR/+}$, $P < 0.002$), which not only confirmed our CRISPR/Cas9 results, but also revealed that females with *Kdm6a*$^{fl/+}$ tumors presented with significantly shorter tumor-free survival (235 days, $P = 0.001$; Fig. 2B) than males. *KDM6A* is located on the X-chromosome but escapes X-inactivation, and its expression reflects gene copy number (22, 23). Heterozygous *Kdm6a*$^{fl/+}$ tumor cells still expressed KDM6A (Supplementary Fig. S4A and S4B), ruling out loss of heterozygosity and indicating that *Kdm6a* functions as a haploinsufficient tumor suppressor.

To test whether our hits also function as tumor suppressors in a mouse model of basal-like breast cancer, we transduced the mammary epithelium of *Trp53*$^{fl/fl}$;*Rb1*$^{fl/fl}$;*Cas9* mice with LV-sgRNA-Cre targeting *Kmt2c* or *Kdm6a* or with sgNT control. Loss of *Kmt2c* significantly reduced tumor latency (323 vs. 436 days; $P = 0.038$), and ablation of *Kdm6a* resulted in a trend toward reduced tumor latency (348 vs. 436 days; $P = 0.17$; Supplementary Fig. S4C), indicating that these EpiDrivers might function as tumor suppressors in several breast cancer subtypes and genetic backgrounds.

EpiDrivers Regulate Genes Involved in Epithelial-to-Mesenchymal Transition, Inflammatory Pathways, and Differentiation

Next, we set out to molecularly characterize the EpiDriver knockout tumors. Transcriptional profiling of fluorescence activated cell sorting (FACS)–isolated *Asxl2-*, *Kdm6a,- Kmt2c-*, *Setd2-*, *Trp53-*, and *Apc*-mutated *Pik3ca*^{HR} tumor cells revealed a wide range of differentially expressed genes compared with control sgNT transduced tumor cells (450–1,800 genes; FDR <0.05, fold change >2; Supplementary Table S3). Principal component (PC) and Pearson correlation analyses revealed high concordance between tumors transduced with sgRNAs targeting the same gene (Fig. 2C; Supplementary Fig. S4D). Variances along PC1 and PC2 were driven by *Apc* and *Trp53* loss, respectively. Consistent with the squamous histology of *Apc*-mutant tumors, gene set enrichment analysis (GSEA) revealed increased expression of genes linked to keratinization in *Apc*-mutant tumors, whereas *Trp53*-mutant tumors showed downregulation of p53-related pathways (Supplementary Fig. S5A and S5B). In addition, intra- and cross-species comparisons revealed that the transcriptome of several *Trp53*-mutant mammary tumors clustered with basal-like human and mouse breast cancer, whereas the control and EpiDriver-mutant *Pik3ca*^{HR} tumors clustered with human HER2 and/or luminal breast cancers (Supplementary Fig. S5C), further underscoring the distinct biology of *Apc*- and *Trp53*-mutant tumors.

Compared with *Apc*- and *Trp53*-mutant tumors, EpiDriver tumors clustered closely together and closer to control sgNT tumors, indicating that they are transcriptionally less divergent (Fig. 2C). Focusing specifically on EpiDriver-mutant versus control sgNT *Pik3ca*^{HR} tumors revealed that EpiDriver inactivation leads to upregulation of "epithelial-to-mesenchymal transition (EMT)" and "proinflammatory interferon-α/γ responses" and downregulation of cellular metabolism ("oxidative phosphorylation" and "fatty acid metabolism") and "estrogen responses" (Supplementary Fig. S5A). Pairwise comparison revealed differences between EpiDriver-mutant transcriptomes, but that overall EpiDriver tumors were more similar to each other than to the control sgNT tumors (3–40 differential pathways in pairwise EpiDriver-mutant comparisons vs. 46–111 differential pathways between EpiDriver-mutant and sgNT control tumors; Supplementary Fig. S6A–S6G), which is expected for proteins within the same molecular complex. To further elucidate a shared molecular profile, we focused on genes that were commonly dysregulated in all EpiDriver-mutant tumors relative to controls (Supplementary Table S3). Pathway analysis of these 498 commonly dysregulated genes revealed enrichment of "extracellular matrix organization" and EMT and downregulation of "epidermis development" and "epithelial cell differentiation" in EpiDriver-mutant tumors relative to control tumors (Fig. 2D and E; Supplementary Fig. S7A and S7B).

To identify downstream target genes involved in tumor suppression, we screened 283 genes downregulated in EpiDriver-mutant tumors for their ability to suppress mammary tumor formation in *Pik3ca*^{HR};*Cas9* mice (Supplementary Fig. S7C). In this secondary screen, the histone lysine demethylase and nuclear receptor corepressor hairless (*Hr*), the interleukin 4 receptor (*Il4ra*), and the transcription repressor *Bcl6* scored as hits, indicating that these shared downregulated genes function themselves as tumor suppressors (Supplementary Fig. S7D). Of note, *Bcl6* also scored in the primary screen and has a known function in mammary gland biology and lactation (24, 25).

Together, these data show that EpiDriver loss leads to significantly accelerated tumor initiation associated with EMT and altered differentiation but does not affect histologic and molecular subtypes. By contrast, loss of *Apc* or *Trp53* not only accelerated tumor development but also caused dramatic transcriptional and histologic changes.

Figure 2. Validation and transcriptomic profiling of EpiDriver tumors. **A,** Tumor-free survival of *Pik3ca*^{H1047R};*Cas9* mice injected with CRISPR lentivirus targeting the indicated gene or control sgNT. Two independent sgRNAs/genes were used, and data were combined (see Supplementary Fig. S2D for single sgRNA data). **B,** Tumor-free survival of *Pik3ca*^{H1047R} mice with conditional knockout of *Asxl2* or *Kdm6a*. **C,** PC plot of all profiled tumor transcriptomes. **D** and **E,** METASCAPE analysis showing enriched (**D**) and depleted (**E**) pathways in common deregulated genes in EpiDriver knockout tumors compared with control tumors. **F,** K-means clustering of differentially enriched (DE) ChIP peak regions based on the differential signal for H3K27ac, H3K27me3, and H3K4me1 between wild-type and sgKdm6a cells. Peaks were stratified as promoter-proximal or distal based on a minimal distance of ≥2.5 kb to an annotated transcription start site (see Methods). FC, fold change.

Pretumorigenic Cells Display Lineage Plasticity and Aberrant Alveogenesis

To elucidate how EpiDriver loss accelerates tumor initiation, we first assessed the sphere-forming capacity of *Pik3ca*^{HR}-mutant mammary epithelial cells 4 weeks after EpiDriver mutation. Interestingly, *Asxl2-*, *Kdm6a-*, or *Kmt2c*-mutant cells formed significantly more mammospheres that grew to larger diameters compared with LV-sgNT-Cre transduced control mammary epithelial cells (Supplementary Fig. S7E–S7G), indicating a growth advantage early in tumor formation (26).

Next, we assessed how loss of the COMPASS-like complex affects the histone modification landscape of mammary epithelial tumor cells. We focused on *Kdm6a*, a core member of the COMPASS-like complex (15, 16), and performed chromatin immunoprecipitation sequencing (ChIP-seq) for H3K27me3, H3K27ac, and H3K4me1 and transcriptional profiling on cultured primary *Pik3ca*^{HR};*Cas9* mammary tumor cells derived from tumors transduced with either sgKdm6a or control sgNT (Supplementary Fig. S8A). We identified differential peaks and clustered them based on the differential ChIP signal for all three histone marks at promoter-proximal [transcription start site (TSS) ± 2.5 kb] or previously identified distal enhancer regions. For each of the distal and proximal regions, we identified two distinct clusters: cluster 1 displaying increased H3K27me3 and decreased H3K27ac and H3K4me1, indicating repressed regions in *Kdm6a*-mutant cells, and cluster 2 with opposite histone profile, indicating activated regions (Fig. 2F). Indeed, we observed the expected up-/downregulation of transcription at promoter-proximal regions consistent with the histone profiles (Fig. 2F; Supplementary Fig. S8B and S8C). Gene set–based analysis of differentially expressed genes by RNA sequencing (RNA-seq) again revealed EMT and differentiation as the most significant sets upregulated in cultured *Kdm6a*-mutant mammary tumor cells (Supplementary Fig. S8C–S8E), consistent with our findings from the EpiDriver-mutant tumors.

Probing deeper into the mechanism of how inactivation of *Kdm6a* affects transcription and chromatin accessibility at the onset of transformation, we performed parallel single-cell RNA-seq (scRNA-seq) and single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq). First, we analyzed scRNA-seq data from FACS-isolated GFP$^+$ LSL-*Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$; LSL-*Cas9-EGFP* (*Pik3ca*^{HR};*Kdm6a*^{KO}) and LSL-*Pik3ca*^{H1047R}; LSL-*Cas9-EGFP* (*Pik3ca*^{HR}) and LSL-*Cas9-EGFP* control mammary epithelial cells 2 weeks after intraductal Ad-Cre injection. Removing low-quality cells with low read depth (<2,500), high mitochondrial reads (>10%), and/or less than 1,000 detected genes resulted in 14,070 high-quality cells composed of 6,160 control, 2,855 *Pik3ca*^{HR}, and 5,055 *Pik3ca*^{HR};*Kdm6a*^{KO} cells (Supplementary Fig. S9A). Based on canonical markers (27), uniform manifold approximation and projection (UMAP) clustering revealed the three major epithelial populations corresponding to luminal progenitors (LP; *Kit*$^+$, *Elf5*$^+$), hormone-sensing mature luminal cells (HS-ML; *Prlr*$^+$, *Pr*$^+$, *Esr1*$^+$), and basal cells (*Krt5/14*$^+$) with distinct subclusters composed of the three genotypes (Fig. 3A and B).

We performed functional enrichment analysis to reveal the molecular pathways dysregulated upon activation of *Pik3ca*^{HR} and inactivation of *Kdm6a* within each epithelial lineage. Surprisingly, this analysis revealed "lactation" as the most differentially regulated pathway in *Pik3ca*^{HR};*Kdm6a*^{KO} versus control cells. "Lactation" was also upregulated but to a lesser degree in *Pik3ca*^{HR};*Kdm6a*^{KO} versus *Pik3ca*^{HR} cells (Fig. 3C). This signature was driven by genes that are typically expressed only upon differentiation of LPs into secretory alveolar cells in a hormone-dependent manner during gestation/lactation and included caseins (*Csn1s1*, *Csn1s2a, Csn2*, and *Csn3*), milk mucins (*Muc1/15*), lactose synthase (*Lalba*), apolipoprotein D (*Apod*), and milk proteins (*Glycam1*, *Spp1*, and *Wap*; Fig. 3B). Interestingly, we observed upregulation of these genes in the absence of gestation/parity-induced hormones and not only in LP cells but also in some basal and HS-ML *Pik3ca*^{HR};*Kdm6a*^{KO} cells (Fig. 3C and D; Supplementary Fig. S9B). Interestingly, this upregulation of alveogenesis/lactation was associated with a downregulation of genes associated with previously described nonlactation LP cells (Supplementary Fig. S9C; ref. 28). IHC confirmed the increased casein levels in *Pik3ca*^{HR};*Kdm6a*^{KO} versus *Pik3ca*^{HR} mammary tissue cells (Fig. 3E). Importantly, genetic ablation of *Kmt2c* or *Asxl2* in *Pik3ca*^{H1047R}-mutant glands also triggered casein expression (Supplementary Fig. S10A and S10B), indicating a shared phenotype.

Other changes were also evident in *Pik3ca*^{HR};*Kdm6a*^{KO} cells. For example, they exhibited upregulation of genes associated with EMT, hypoxia, and involution (Supplementary Figs. S10C and S11A). *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} cells also exhibited higher expression of characteristic HS-ML genes such as *Cited1* and prolactin receptor (*Prlr*) not only in HS-ML cells but also in a subset of LP and/or basal cells (Fig. 3B; Supplementary Fig. S11B). Conversely, basal markers such as *Krt14*, *Lgr5*, and *Nrtk2* showed aberrant expression in *Pik3ca*^{HR} and/or *Pik3ca*^{HR};*Kdm6a*^{KO} LP cells (Supplementary Fig. S11C). Overall, our data reveal reprogramming of transcriptional landscapes, loss of lineage integrity, and induction of alveogenesis in all mammary epithelial lineages upon oncogenic PI3K signaling, and these cancer hallmarks are exacerbated by loss of EpiDrivers.

Chromatin Profiling Confirms Epigenetic Reprogramming and Mimicry of Alveogenesis

In line with the scRNA-seq results and our previous data (29), unsupervised UMAP clustering of the snATAC-seq data showed that chromatin accessibility clearly separated the three major mammary epithelial lineages (Fig. 4A). Although control, *Pik3ca*^{HR}, and *Pik3ca*^{HR};*Kdm6a*^{KO} cells were intermingled in the HS-ML cluster, indicating that they are indistinguishable with regard to accessible chromatin, they formed distinct subclusters in the LP and to a lesser degree in the basal cluster (Fig. 4A). Within the LP clusters, there was a modest difference between control and *Pik3ca*^{HR} LP cells, and large differences were observed between control and *Pik3ca*^{HR};*Kdm6a*^{KO} and between *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} LP cells (Fig. 4B), showing that loss of *Kdm6a* has a profound effect on chromatin accessibility. In line with KDM6A's H3K27 demethylase function in COMPASS-like enhancer activation, we found substantially more genomic accessibility in *Kdm6a*-mutant cells (Fig. 4B).

We next examined the representation of transcription factor motifs in the differentially accessible genomic regions. The regions with increased accessibility in the *Pik3ca*^{HR};*Kdm6a*^{KO}

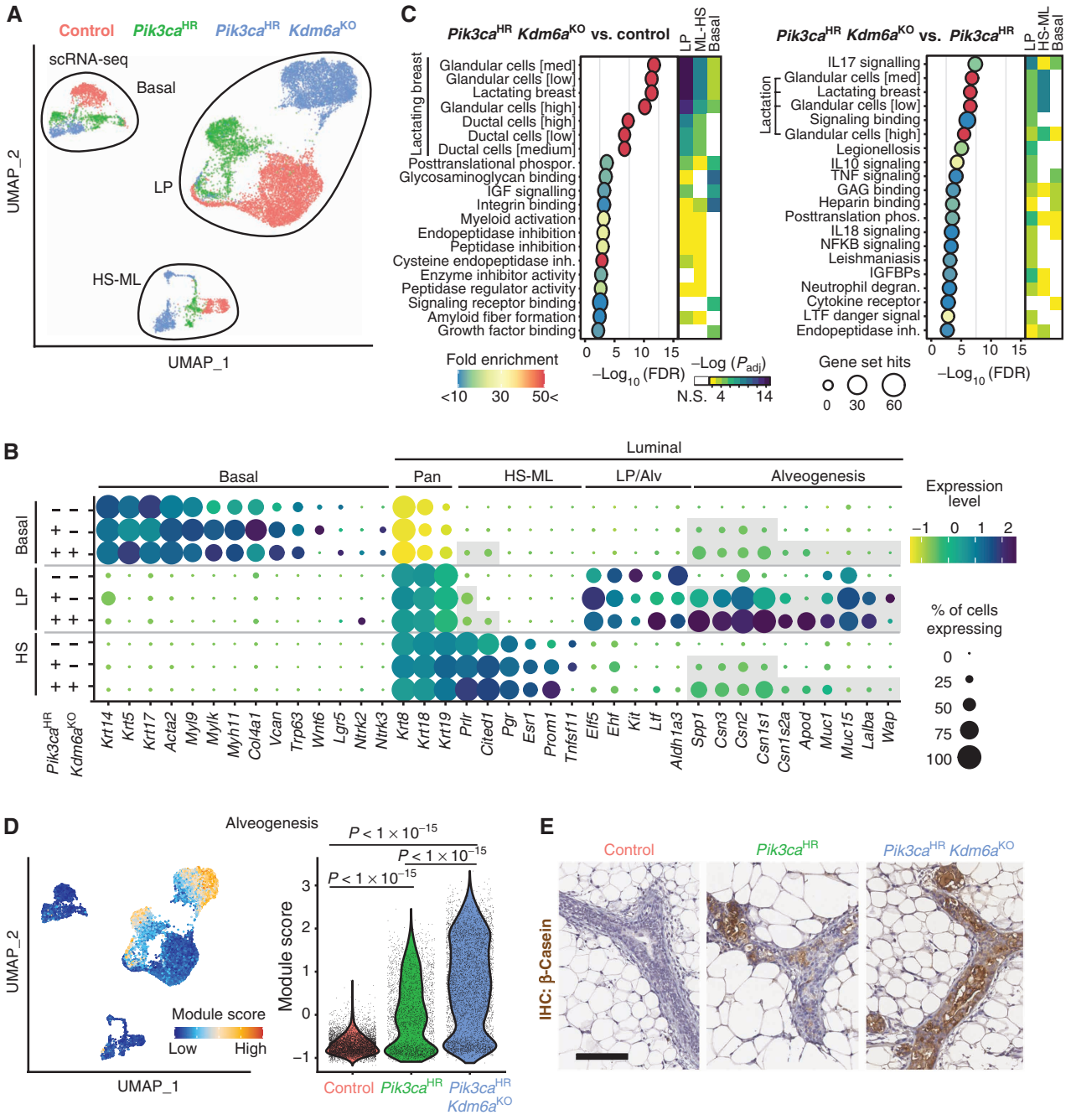
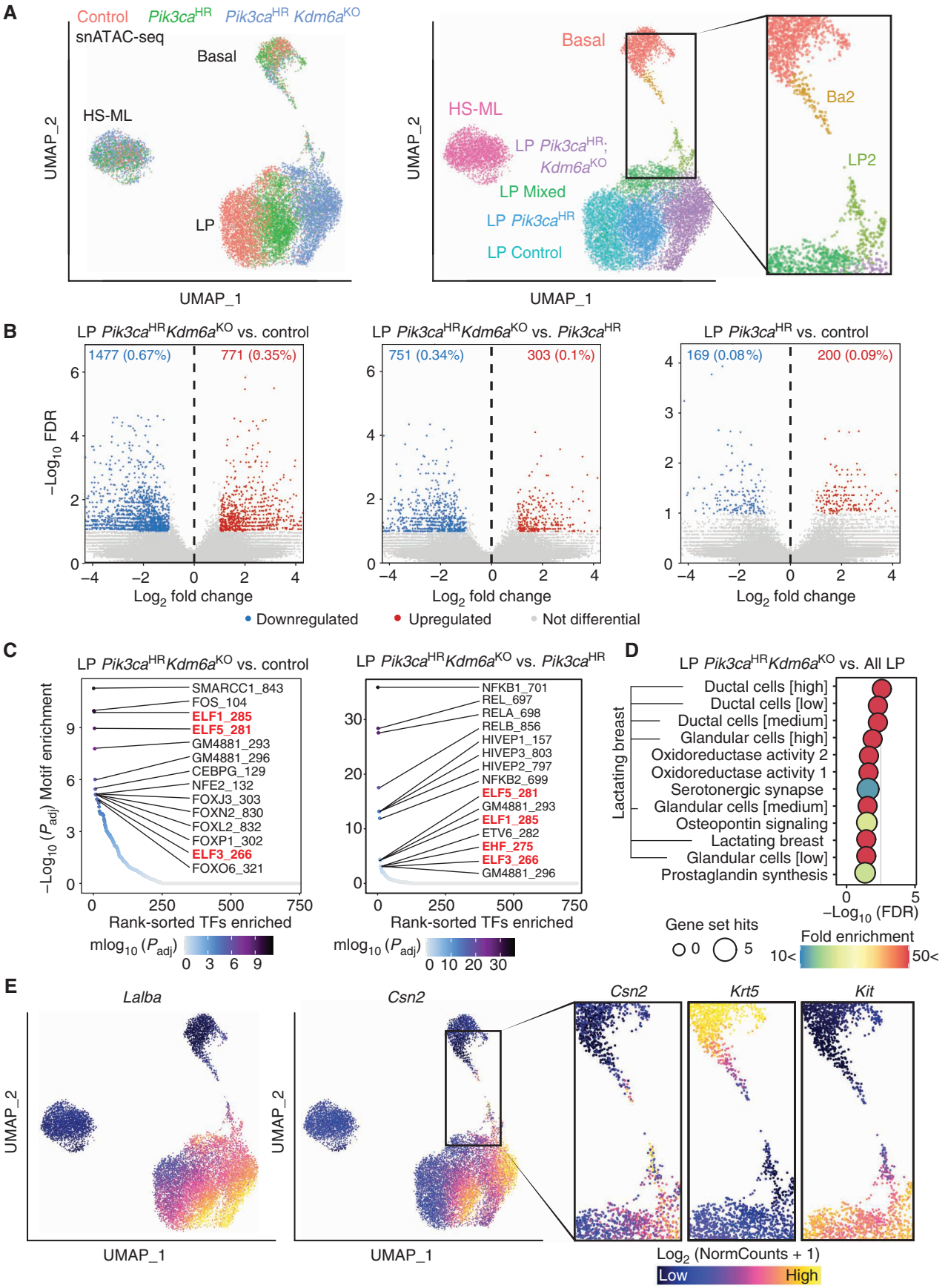


Figure 3. Single-cell transcriptional profiling reveals alveogenic mimicry. **A,** UMAP plot showing mammary epithelial cells from control, *Pik3ca*^{H1047R}-, and *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$-mutant mice 2 weeks after Ad-Cre injection. **B,** Dot plot showing differentially expressed marker genes within the different epithelial lineages stratified by genotypes. **C,** Pathways differentially enriched in *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ versus control and *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ versus *Pik3ca*^{H1047R} mammary epithelial LP cells identified using g:Profiler ($P < 0.05$ with Benjamini–Hochberg FDR correction, >10-fold enrichment). The top 20 enriched pathways are shown. Heat map depicts how these pathways are altered in the three major epithelial lineages. degran., degranulation; inh., inhibition; phospho./phosphor., phosphorylation. **D,** UMAP and violin blots showing alveogenesis signature. **E,** IHC of mammary glands 2 weeks after injection stained with anti-β-casein. Scale bar = 100 μm.

relative to wild-type LP cells were significantly enriched for binding sites of FOS and SMARCC1, followed by the ETS factors ELF1/3/5. Motifs enriched in the *Pik3ca*^{H1047R};*Kdm6a*^{KO} relative to the *Pik3ca*^{HR} LP cells corresponded to NF-κB factors NF-κB1/2 and RELA/B followed again by core LP regulators ELF1/3/5 and EHF (Fig. 4C). Similar enrichment profiles were seen from activity inference using chromVAR (Supplementary Fig. S12A and S12B; ref. 30). Gene set–based analysis of accessible loci revealed "lactation" as the most significant set upregulated in *Pik3ca*^{HR};*Kdm6a*^{KO} LP cells, consistent with the known function of ELF5 and EHF in driving alveolar differentiation (27, 31), and in line with the scRNA-seq data; this association included increased accessibility to multiple alveolar/milk biogenesis-related genes, such as *Apod*,



Csn2/1s1/1s2a, Lalba, Lif, Lipa, and *Spp1* (Fig. 4D and E; Supplementary Fig. S13A).

Further examination of snATAC-seq results identified a basal-like "Ba2" subcluster and a luminal-like "LP2" subcluster enriched in *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} cells that appear to bridge the basal and LP populations (Fig. 4A). Gene set–based analysis of accessible loci in these subclusters revealed sets associated with "chromatin silencing" (Supplementary Fig. S13A and S13B). In addition, the biological Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway "breast cancer" was upregulated in the Ba2 versus the basal cluster, with the identification of prominent WNT (*Wnt10a, Wnt6, Fzd2, Dvl2, Prickle4, Csnk1g2*, and *Dlg4*) and NOTCH (*Dll1* and *Jag2*) signaling genes (Supplementary Fig. S13A and S13C). In line with this notion, chromVAR analysis showed enrichment of binding sites for transcription factors associated with WNT (LEF1, TCF7, TCF7L1, and TCF7L2) and NOTCH (HES1, HEY1/2, and HEYL) signaling in Ba2 cells (Supplementary Fig. S12A). Consistently, we observed upregulation of WNT and NOTCH signaling signatures in *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} basal cells in the scRNA-seq dataset (Supplementary Fig. S13D and S13E). Of note, *Apc* was a major hit in the *in vivo* CRISPR screen (Fig. 1D), suggesting that elevated WNT signaling is oncogenic in the *Pik3ca*^{HR} model. In addition, WNT and NOTCH signaling not only are known drivers of breast cancer but also play critical roles in mammary lineage determination (32–34).

Overall, we found that Ba2 cells have reduced chromatin accessibility at basal markers, such as *Acta2*, *Krt5/14*, *Trp63*, and *Vim*, and increased accessibility of the alveolar genes, such as *Csn2*, whereas LP2 cells have reduced chromatin accessibility at LP markers, such as *Elf5, Ehf*, and *Kit* (Fig. 4E; Supplementary Figs. S14A–S14C and S15A–S15C). These data are consistent with the loss of lineage identity observed in the scRNA-seq data. Together, our scRNA-seq and snATAC-seq data suggest that *Pik3ca*^{HR};*Kdm6a*^{KO} mammary epithelial cells gain lineage plasticity and prior to tumorigenesis reprogram toward the alveolar fate reminiscent of epithelial expansion and differentiation preceding lactation.

To functionally test whether inducing an alveogenic program can indeed accelerate tumorigenesis, we overexpressed ELF5, the key regulator of alveogenesis, in *Pik3ca*^{HR} mammary epithelial cells. Transduction of lentiviruses overexpressing *Elf5* (LV-Elf5-Cre) induced faster tumor formation compared with control LV-Ruby-Cre ($P < 0.05$). This is consistent with previous findings of ELF5 overexpression in a PyMT breast cancer mouse model (refs. 35, 36; Supplementary Fig. S16A and S16B). In addition, overexpression of ELF5 in *Pik3ca*^{HR} mammary epithelial cells triggered casein expression (Supplementary Fig. S16C), reminiscent of the consequences of EpiDriver mutations. Together, these results support a role of alveogenic mimicry in mammary gland tumorigenesis.

The COMPASS-like Complex Inhibits a Tumorigenic Basal-to-Luminal Cell Lineage Conversion

We next determined whether both luminal and basal cells are susceptible to lineage plasticity and contribute to tumor formation using lineage tracing with a basal-specific adenoviral Ad-K5-Cre and luminal-specific Ad-K8-Cre viruses (Supplementary Fig. S17A–S17E; ref. 37). As previously shown (38, 39), expression of oncogenic *Pik3ca*^{HR} can lead to lineage plasticity and convert basal and luminal unipotent progenitors into multipotent cells. In line with these reports, induction of *Pik3ca*^{HR} in basal cells resulted in a gradual lineage conversion to luminal-like cells, which was dramatically accelerated by *Kdm6a* or *Asxl2* mutation (Fig. 5A–C). In line with a haploinsufficiency tumorigenic effect, heterozygous loss of *Kdm6a* also significantly accelerated basal-to-luminal lineage conversion (Supplementary Fig. S17F). In contrast, genetic ablation of *Kdm6a* or *Asxl2* did not accelerate lineage conversion from luminal-to-basal cells (Supplementary Fig. S17G).

To further characterize this basal-to-luminal lineage conversion, we used a K5-Cre^{ERT2} transgenic strain crossed to *Pik3ca*^{HR};*Kdm6a*$^{fl/fl}$;LSL-*Cas9-GFP* mice. We used low-dose tamoxifen treatment to genetically ablate *Kdm6a* and concomitantly activate *Pik3ca*^{HR} at clonal density in the basal mammary compartment. This approach corroborated our findings and allowed us to quantify converting clones along the epithelial tree. At 4 weeks after tamoxifen treatment, we observed that 50% of GFP$^+$ lineage-traced basal clones had generated K8$^+$ luminal-like cells (Supplementary Fig. S17H), demonstrating that this lineage conversion is a frequent event in *Pik3ca*^{HR};*Kdm6a*^{KO} mammary tissue.

Next, we determined if the cell of origin affects the latency and phenotype of tumors arising in *Pik3ca*^{HR};*Kdm6a*$^{fl/fl}$ mice. Loss of *Kdm6a* in the basal compartment significantly accelerated tumor formation, whereas luminal cell–derived *Pik3ca*^{HR};*Kdm6a*^{KO} tumors arose with similar latency as *Pik3ca*^{HR} tumors (Fig. 5D and E). Transcriptome analysis revealed that basal cell–derived tumors clustered with other mouse and human luminal-like tumors (Supplementary Fig. S5C), were indistinguishable from tumors derived upon sgRNA-mediated mutation of *Kdm6a*, and exhibited K5$^+$, K8$^+$, and K5/K8 double-positive cells and casein$^+$ cells (Supplementary Fig. S17I–S17K). Together, these results indicate that loss of the COMPASS-like complex in *Pik3ca*^{HR} basal cells accelerates their reprogramming into tumor-initiating cells that drive luminal-like breast cancer.

To further characterize the basal-to-luminal-like cell transition, we performed scRNA-seq on control, *Pik3ca*^{HR}, or

Figure 4. scATAC-seq reveals alveogenic mimicry and bridge-like clusters. **A,** Unsupervised UMAP plot of snATAC-seq profile colored by genotype (left) and identified clusters (middle). Inset (right) shows Ba2 and LP2 clusters. **B,** Volcano plots showing differentially accessible chromatin peaks between *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ and wild-type control, between *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ and *Pik3ca*^{H1047R}, or between *Pik3ca*^{H1047R} and wild-type control LP cells. **C,** Enrichment of transcription factor (TF) binding sites in differentially accessible chromatin. **D,** Pathways differentially enriched in *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ versus all mammary epithelial LP cells inferred from gene accessibility ArchR Gene Scores. The top 12 enriched pathways are shown as identified using g:Profiler ($P < 0.05$ with Benjamini–Hochberg FDR correction, >10-fold enrichment). **E,** UMAP plots showing open chromatin associated with the alveolar/lactation-associated genes *Lalba* and *Csn2*. Inset (right) shows open chromatin associated with the alveolar/lactation gene *Csn2*, the basal marker gene *Krt5*, and the LP marker gene *Kit* in Ba2 and LP2 clusters.

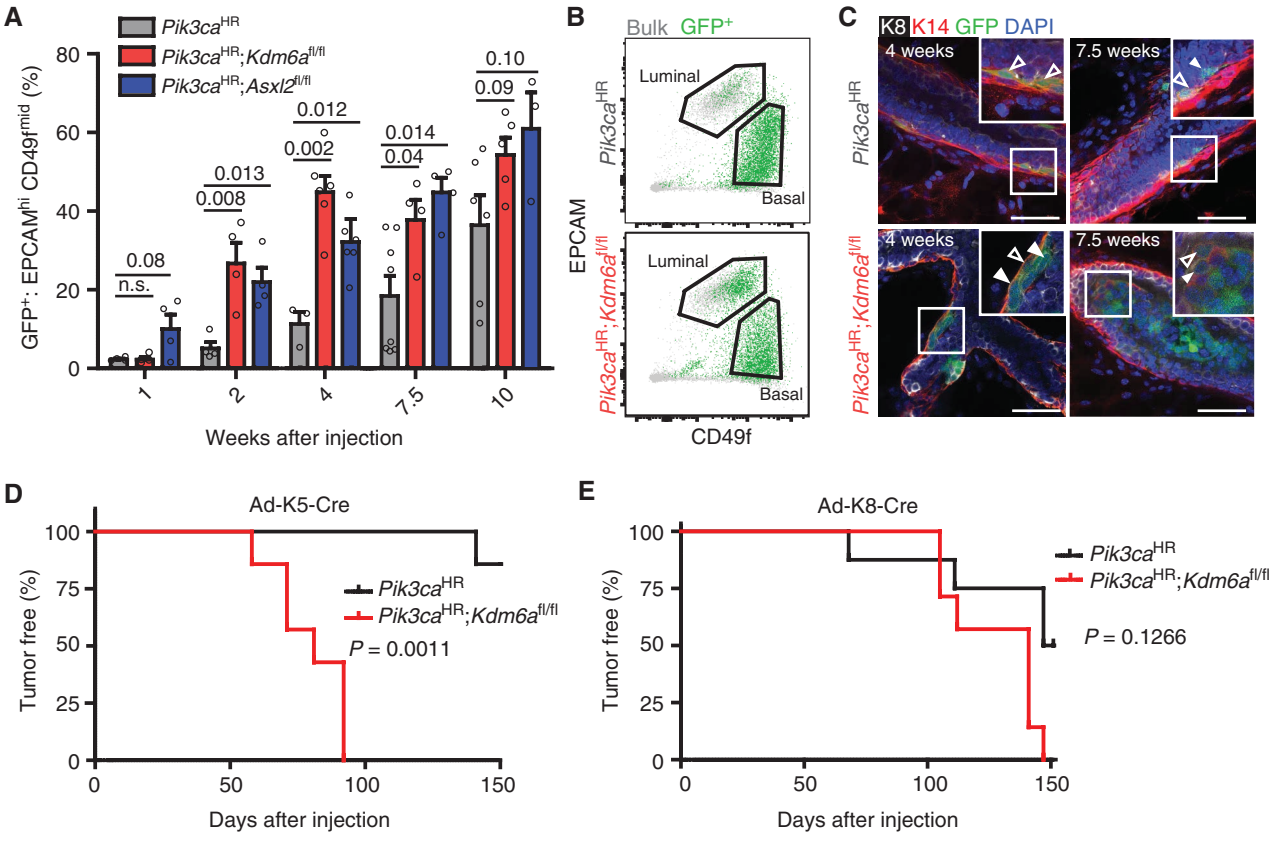


Figure 5. Loss of EpiDrivers induces multipotency. **A,** Percentage of GFP+ cells in the EPCAM^{hi} CD49f^{mid} luminal gate at different time points after Ad-K5-Cre injection into the mammary epithelium of mice with the indicated genotype. n.s., not significant. **B,** Representative FACS plot at 4 weeks after injection with Ad-K5-Cre. **C,** Whole-mount image of mammary glands 4 weeks and 7.5 weeks after Ad-K5-Cre injection showing K14+/K8+ (empty arrowheads) as well as K14/K8 double-positive and K14−/K8+ GFP+ lineage-traced cells (filled arrowheads). Scale bar = 50 μm. **D** and **E,** Tumor-free survival of *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ versus *Pik3ca*^{H1047R} after intraductal injection of Ad-K5-Cre (**D**) and Ad-K8-Cre (**E**).

Pik3ca^{HR};*Kdm6a*^{KO} mammary epithelial cells after 2 weeks of Ad-K5-Cre lineage tracing (Fig. 6A–D; Supplementary Fig. S18A). Consistent with the results above, LP-like cells that lost basal markers and gained LP (e.g., *Cd14, Elf5*, and *Kit*) and alveolar markers (e.g., *Apod, Cns3*, and *Wfdc18*) emerged from *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} basal cells. We even observed rare *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} cells expressing milk genes, such as *Olah* and *Wap*, and HS-ML markers, such as *Prlr* (Fig. 6E–G; Supplementary Figs. S18B and S18C and S19A–S19C).

In addition, *Pik3ca*^{HR};*Kdm6a*^{KO} basal cells were more heterogeneous than wild-type or *Pik3ca*^{HR} cells and comprised three unique subclusters: *Kdm6a*^{KO}-L, adjacent to the LP-like population, a central cluster (*Kdm6a*^{KO}-C), and a cluster enriched in basal/myoepithelial markers (*Kdm6a*^{KO}-B; *Acta2, Igfbp2, Myh11*, and *Myl9*; Fig. 6B), further underscoring the notion of increased phenotypic plasticity upon loss of *Kdm6a*. Importantly, *Kdm6a*^{KO}-L showed a gradual downregulation of basal markers with concomitant upregulation of alveolar/lactation markers such as *Apod*, *Csn2/3, Muc1/15*, or *Wfdc18* (Fig. 6E–G; Supplementary Figs. S18B, S19A–S19C , and S20A–S20C). *Kdm6a*^{KO}-L was also marked by expression of the EMT master regulators *Zeb1* and *Zeb2*, the latent TGFβ binding gene product *Ltbp1*, as well as *Ntrk2* and *Socs2* (Supplementary Fig. S20D). Of note, *Ntrk2* was previously identified as a basal-to-luminal multipotency breast cancer gene (38) and, together with *Ptn*, is a known driver of breast cancer (40). Interestingly, this *Kdm6a*^{KO}-L cluster did not generally express classic luminal progenitor markers (*Aldh1a3, Cd14, Elf5, Kit*, and *Lif*; Fig. 6F; Supplementary Fig. S18C). This observation combined with trajectory analysis suggests that *Pik3ca*^{HR};*Kdm6a*^{KO} basal cells start to gradually activate an aberrant alveolar-like program before acquiring LP characteristics (Fig. 6C–G).

Integrating the Ad-Cre and the Ad-K5-Cre scRNA-seq datasets revealed that luminal-like K5-traced *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} cells clustered with LP cells, further supporting the notion of a basal-to-luminal reprogramming. In addition, luminal-like K5-traced *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} cells with high lactation and involution signatures clustered with *Pik3ca*^{HR} and *Pik3ca*^{HR};*Kdm6a*^{KO} LP cells, whereas those without a lactation/involution signature clustered with wild-type LP cells, suggesting functional heterogeneity (Supplementary Fig. S21A–S21C).

Cells in the proliferating cluster consisted mainly of *Pik3ca*^{HR};*Kdm6a*^{KO} with either basal or luminal characteristics (Fig. 6A, B, E, and F). This cluster also showed marked elevation of RB1/E2F target genes (Supplementary Fig. S22), reminiscent of RB1 inactivation and E2F activation during pregnancy-induced hyperproliferation in the mammary gland (41). These data further support the role of these proliferating cells and the aberrant alveolar program during tumor initiation.

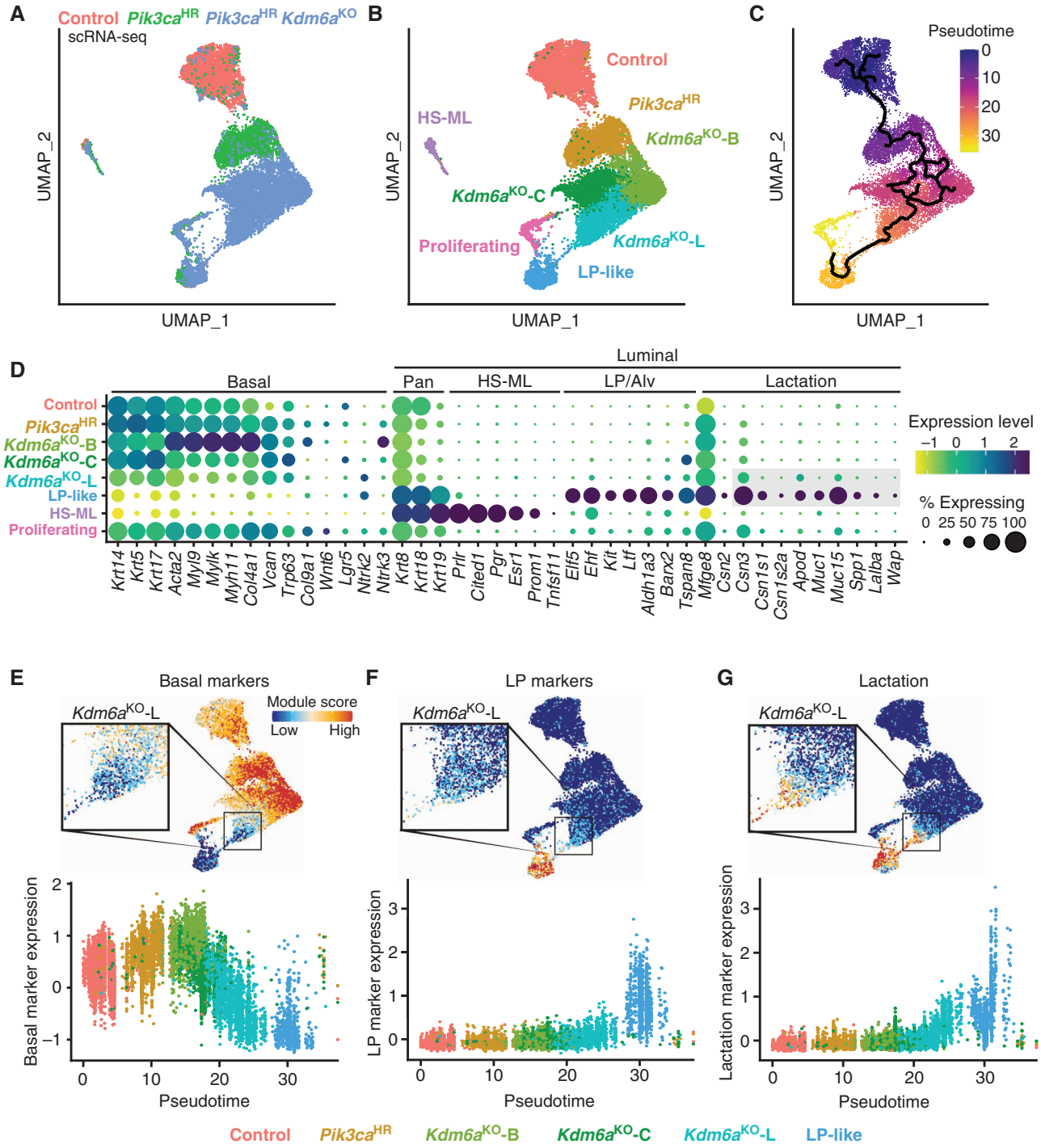


Figure 6. scRNA-seq reveals basal-to-alveolar transdifferentiation at the onset of breast cancer initiation. **A–C,** UMAP plots showing Ad-K5-Cre lineage-traced basal mammary epithelial cells from control, *Pik3ca*^{H1047R}-, and *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$-mutant mice 2 weeks after injection colored by genotype (**A**), clusters (**B**), and trajectories inferred by Monocle 3 (**C**). **D,** Dot plot showing differentially expressed marker genes within the different epithelial clusters. **E–G,** UMAP and pseudotime trajectory plots showing basal (**E**), LP (**F**), and alveolar/lactation (**G**) marker signatures.

Human Breast Cancer Shows Frequent EpiDriver Alterations and Signs of Aberrant Alveogenesis

To extend our findings from mice to humans, we assessed the function of the EpiDrivers in human MCF10A mammary epithelial cells that harbor a *PIK3CA*^{H1047R} knockin mutation (42, 43). Using CRISPR/Cas9, we generated *ASXL2*-, *KDM6A*-, *KMT2C*-, *SETD2*-, *PTEN*-, and *TP53*-mutant cell lines, as well as control sgNT cells (Supplementary Fig. S23A–S23D). Like the parental cells, MCF10A *PIK3CA*^{H1047R} cells formed polarized and hollow, albeit modestly larger, acini in Matrigel

culture (43). In contrast, *ASXL2*-, *KDM6A*-, *KMT2C*-, or *PTEN*-mutant spheres showed a transformed phenotype with large branching protrusions (Supplementary Fig. S23E and S23F). When grafted orthotopically into the fat pads of immunodeficient [NOD scid gamma (NSG)] mice, the *KDM6A-, SETD2-, TP53-*, and *PTEN*-mutant MCF10A *PIK3CA*^{H1047R} cells formed tumors, whereas control sgNT cells did not (Supplementary Fig. S23E). Although the *ASXL2*- and *KMT2C*-mutant cells exhibited a transformed phenotype in 3D cultures, they did not efficiently give rise to xenograft tumors in mice. Together, these data indicate that the EpiDrivers *ASXL2*, *KMT2C, KDM6A*, and *SETD2* suppress transformation of human MCF10A mammary epithelial cells.

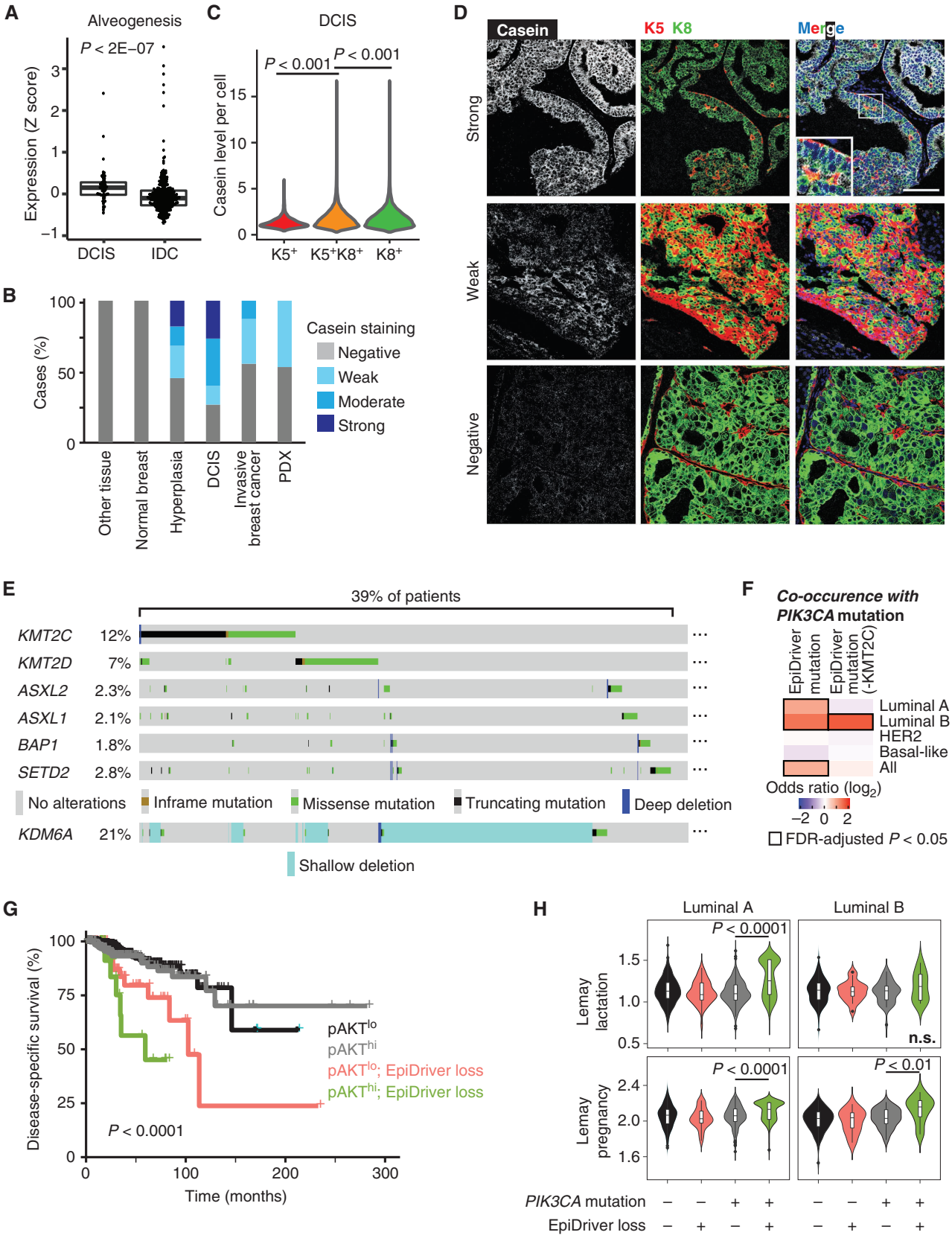
Next, we compared our results from mouse *Kdm6a*-mutant mammary tumor cells to the data obtained from transcriptome and epigenetic profiling of human *KDM6A*-mutant MCF10A *PIK3CA*^{H1047R} cells. We used two independent sg*KDM6A* knockout and two sgNT control clones (Supplementary Fig. S23G) and performed RNA-seq and ChIP-seq for H3K27me2, H3K27ac, and H3K4me1. As expected, the clones clustered together by genotype for both transcriptional and H3K27me3, H3K27ac and H3K4me1 profiles (Supplementary Fig S23H and S23I). Clustering of differential promoter-proximal and distal peaks based on their histone marks again revealed two clusters: cluster 1 displaying increased H3K27me3 and decreased H3K27ac and H3K4me1, indicating repressed regions in KDM6A-mutant cells, and cluster 2 with an opposite histone profile, indicating activated regions. Consistent with these histone profiles, we observed the expected upregulation/downregulation of transcription (Supplementary Figs. S23J–S23L and S24A).

Like mouse *Kdm6a*-mutant mammary tumor cells, *KDM6A*-mutant MCF10A *PIK3CA*^{H1047R} cells showed upregulation of gene sets linked to EMT and mammary stem cells and downregulation of adhesion (Supplementary Fig S24B and S24C). Specifically, we observed upregulation in key mesenchymal markers, such as *CDH2*, *VIM*, and *ZEB1*, and downregulation of *CDH1* and of a repressor of EMT, *GRHL2. KDM6A*-mutant cells also showed some signs of aberrant differentiation, including upregulating *KRT14*, downregulating *KRT18*, and gaining expression of lactation-related genes, including the prolactin receptor (Supplementary Fig S24D–S24F). *KDM6A*-mutant cells also showed upregulation of oncogenes (*MAFB, ETV1, ROS1*, and *EPAS1*) but downregulation of tumor suppressors (*SIRPA, TP63*, and *PTPRB*; Supplementary Fig. S24D and S24E). Overall, these data indicate that knockout of *KDM6A* results in coordinated transcriptional and epigenetic alterations that induce EMT and alter differentiation, which is concordant with our findings in mouse *Kdm6a* knockout cells.

To test whether the alveogenesis program can also be found in human premalignant breast lesions, we analyzed the transcriptional profiles of 57 ductal carcinoma *in situ* (DCIS) and 313 invasive breast cancers (44). Remarkably, we found that curated human gene sets corresponding to mammary gland alveogenesis and lactation exhibited significantly higher expression in DCIS compared with IDC (Fig. 7A; Supplementary Fig. S25A) and correlated with the signatures of EpiDriver loss derived from the mouse tumor studies (Supplementary Fig. S25B and S25C). To corroborate these findings, we optimized and performed IHC for the milk protein casein CSN1S1 on tissue microarrays (TMA). Interestingly, 55% of breast atypical hyperplasia, 73% of DCIS, 44% of invasive breast cancer and 47% of breast cancer patient-derived xenografts exhibited casein staining, whereas no normal breast or any other cancerous or noncancerous tissue exhibited casein staining (Fig. 7B; Supplementary Fig. S26A and S26B). Additional staining of DCIS tumor cores revealed that although casein staining was generally low in KRT5 single-positive cells, stronger casein staining was observed in both KRT5/KRT8 double-positive cells as well as KRT8 single-positive cells, suggesting that alveogenic mimicry can be observed during basal-to-luminal-like conversion or in intermediate lineage cells (Fig 7C and D; Supplementary Fig. S27A). Analysis of an independent panel of 118 clinically annotated DCIS revealed that 50% of hormone receptor–positive (HR$^+$), 56% of HER2$^+$ HR$^+$, 33% of HER2$^+$ HR$^-$, and 20% of HER2$^-$ HR$^-$ DCIS express casein and that HR$^+$ cases showed a higher percentage of casein-positive cells (Supplementary Fig. S27B and S27C). We also found that casein-positive DCIS exhibited more progesterone receptor–positive (PR$^+$) cells, which is in line with progesterone's role during lobuloalveogenesis (Supplementary Fig. S27D). Cases with casein staining did not show statistically significant differences with regard to ipsilateral breast cancer recurrence, although trends toward poorer outcome were observed especially in PR$^+$, as well as HER2$^+$ HR$^+$, cases (Supplementary Fig. S27E).

In human invasive breast cancer, *ASXL2, BAP1, KDM6A, KMT2C, KMT2D*, and *SETD2* are each mutated in 1% to 12% of breast tumors as expected for long-tail genes (Fig. 7E; Supplementary Fig. S28A; refs. 11, 13). The haploinsufficiency of *Kdm6a* in mouse mammary tumorigenesis prompted us to also analyze copy-number alterations. Interestingly, an additional 19% of patients exhibited shallow deletion indicative of heterozygous *KDM6A* loss (Fig. 7E; Supplementary Fig. S28A), which coincided with significantly reduced *KDM6A* expression (Supplementary Fig. S28B). In addition, EpiDriver alterations showed a trend toward mutual exclusivity, and we observed a significant co-occurrence with *PIK3CA* mutations (Fig. 7F;

Figure 7. EpiDrivers function as tumor suppressors in humans. **A,** Average expression of the alveogenesis gene signature from 57 DCIS and 313 IDC. **B,** Casein staining level by IHC in each tissue or tumor type. PDX, patient-derived xenograft. **C,** Casein staining intensity in individual cells in DCIS tumor cores separated by keratin staining. **D,** Representative imaging mass cytometry images of DCIS cores stained for casein, KRT5, KRT8, and nuclear stain. Scale bar = 100 μm. **E,** Prevalence of alterations in EpiDrivers in human breast tumors. Shallow deletion only displayed for *KDM6A*. **F,** Co-occurrence analysis of *PIK3CA* and EpiDriver mutations in the combined breast cancer dataset of The Cancer Genome Atlas (TCGA) and METABRIC. The results are shown for the complete set of identified EpiDrivers (left), or by excluding *KMT2C* (right), considering truncating and deleterious missense mutations. The heat map shows the co-occurrence odds ratios ($\log_2$) across breast cancer subtypes and all tumors considered, and significant (FDR-adjusted $P < 0.05$) associations are highlighted by black rectangles. **G,** Disease-specific survival of patients with breast cancer in the TCGA cohort stratified by phospho-Ser473 AKT (pAKT) and EpiDriver mutations. The long-rank *P* value is shown. **H,** Violin plots showing the expression of the Lemay lactation and pregnancy signatures in TCGA tumors with concurrent *PIK3CA* and EpiDriver mutations relative to other groups in luminal A and B breast cancer. The Mann–Whitney test *P* value is shown (n.s., not significant).



Supplementary Fig. S28A, C, and D; Supplementary Tables S4 and S5). Cases with concurrent *PIK3CA* and EpiDriver mutations did not show statistically significant differences with regard to overall survival when compared with cases with only *PIK3CA* mutation, although we did observe trends toward poorer outcomes in luminal A cases (Supplementary Fig. S29). Given that high PI3K signaling can be a consequence of several genetic alterations in cancer, we performed a survival analysis of The Cancer Genome Atlas (TCGA) breast tumors stratified by PI3K signaling defined by means of phospho-Ser473 AKT (45) or a PI3K transcriptional signature (46). Interestingly, concomitant EpiDiver mutations and high PI3K signaling stratified patients with poor survival across subtypes (Fig. 7G), as well as within luminal A and B breast cancer (Supplementary Fig. S30A–S30C). Concurrent *PIK3CA* and EpiDriver mutations also stratified patients with worse outcome in the independent METABRIC dataset across subtypes and within $HER2^+$ cases (Supplementary Fig. S31A and S31B).

Luminal A and/or B tumors with concurrent *PIK3CA* and EpiDriver mutations were found to be associated with higher expression of gene sets linked to mammary gland alveogenesis and lactation and homologous genes upregulated in EpiDriver-mutant mouse breast cancers (Fig. 7H; Supplementary Fig. S32A and S32B). GSEA identified hallmarks of EMT and immune system function (interferon-α/γ responses, inflammatory responses, TNFα and TGFβ signaling) and downregulation of cellular metabolism (oxidative phosphorylation and fatty acid metabolism) associated with concurrent *PIK3CA* and EpiDriver mutations, especially in luminal B tumors akin to our mouse model (Supplementary Fig. S32C and S32D). Together, these data highlight the relevance of the tumor-suppressive EpiDriver network and alveogenic mimicry during breast cancer initiation.

DISCUSSION

Large international efforts such as the TCGA and the International Cancer Genome Consortium have set out to profile the mutational landscape of many cancers with the goal of cataloging the genes responsible for tumor initiation and progression. The idea was to identify those genes that are mutated more frequently than expected by random chance, and the expectation was that increasing sample size would boost the power to mathematically infer driver mutations (i.e., sensitivity) while weeding out background of random somatic mutations (i.e., specificity). These efforts have considerably expanded the catalog of cancer genes; however, as these studies advance, it is more evident that the individual contribution of most cancer genes to a given cancer burden is very modest. This observation raises important concerns on how confidently we can identify cancer genes based on their mutation profiles and, most importantly, highlights the fundamental question of which common and/or specific mechanisms endorse carcinogenesis.

Here, we devised and deployed an *in vivo* CRISPR/Cas9 screening methodology, which allowed us to identify bona fide cancer drivers in the long tail of breast cancer genes. Our screen identified several tumor suppressor genes, with the top hits converging on epigenetic regulation and mammary epithelial differentiation. Individually, epigenetic regulators are not mutated frequently, but as a group, they are among the most frequently mutated targets in cancer (47–51), indicating that a "dysregulated epigenome" can accelerate tumor development. In particular, we identified several components and auxiliary factors of the COMPASS-like histone methyltransferase complex as potent tumor suppressors and showed that KDM6A might function in a haploinsufficient manner. Our results show that loss of those EpiDrivers accelerates tumor initiation and that the transcriptional profiles of EpiDriver knockout tumors closely cluster together. However, the results do not rule out the possibility that the individual genes also have distinct functions, perhaps depending on cellular or microenvironmental context. It is noteworthy, however, that loss of each of the EpiDrivers analyzed triggers a similar alveogenesis program associated with casein expression. This indicates that their loss, at least in part, reflects involvement in shared biological processes that are distinct from, for example, p53 tumor suppressor loss. Importantly, up to 39% of patients with breast cancer harbor mutations in the COMPASS-like pathway, highlighting the importance of elucidating the mechanisms by which COMPASS inactivation contributes to breast cancer. In human tumors, EpiDriver genes are deleted or harbor nonsense or missense mutations. Most of the missense mutations are variants with uncertain significance and whereas many are predicted to be deleterious (Supplementary Table S4), their exact function and effect on cancer etiology remain to be determined. Further studies will also be needed to elucidate the potential private functions of these tumor suppressors alone or in combination with a sensitizing oncogene such as $PIK3CA^{H1047R}$.

Components of the COMPASS-like complex were recently implicated as tumor suppressors in leukemia (52), medulloblastoma (53), pancreatic cancer (23), and non–small cell lung cancer (54), and their loss was associated with substantial enhancer reprogramming and aberrant transcription. We were surprised to find that EpiDriver inactivation did not substantially affect histology or transcriptional profiles of breast tumors. However, it did significantly accelerate tumor initiation, which was coupled with rapid acquisition of phenotypic plasticity. Plasticity plays a central role in development and during tissue regeneration and wound healing (29, 55, 56). More recently, phenotypic plasticity has also been recognized as a driving force behind tumor initiation and progression (57–59). For example, elegant lineage tracing and single-cell profiling experiments have shown that oncogenic signaling can reactivate multipotency within the two epithelial lineages of the mammary gland (38, 39, 57). Cells that acquire plasticity are thought to gain stem cell features through a process of dedifferentiation (56, 60). However, in the system studied here, we did not observe the acquisition of fetal mammary stem cell–like transcriptomes as observed in basal-like tumor studies (29, 57). Rather, we observed an aberrant differentiation program associated with alveogenesis induced upon PI3K activation and exacerbated by EpiDriver loss. This was most noteworthy in basal cells, which are known to be functionally plastic (61–63). A similar aberrant alveolar differentiation program was recently described in breast cancer models driven by the luminal loss of BRCA1 and p53 (27) and upon luminal overexpression of ELF5 and PyMT (35, 36). Importantly, we show that overexpression of ELF5 in a $Pik3ca^{H1047R}$–mutant

background accelerates mammary tumorigenesis. Although this indicates that alveogenesis is sufficient to increase tumorigenesis, it still remains to be determined whether alveogenesis in the context of EpiDriver mutations is required for the observed accelerated tumor phenotype.

Together, our data indicate that there are different avenues toward transformation and that the innate but poised program coordinating the proliferative burst during gestation and onset of lactation can be highjacked for rapid expansion at the onset of oncogenic transformation—a phenomenon we term "alveogenic mimicry." This phenomenon is exacerbated by the loss of epigenetic control governed by COMPASS-like and associated BAP1–ASXL1/2 complexes and happens not only in luminal cells but—given the right combinations of mutations—also in basal cells. It will be interesting to assess whether other cancers also coerce inherent regenerative or tissue remodeling processes during early transformation.

Another interesting aspect of our study is the potential cell of origin underlying different subtypes of breast cancer. Gene expression studies indicated that mature luminal cells give rise to luminal A/B and HER2 subtypes, whereas LP transform to basal-like cancers and basal cells give rise to the claudin-low subtype (64–66). Mouse lineage tracing studies have supported these observations and have shown that certain mutations in specific lineages can indeed give rise to mouse mammary tumors with features similar to different human breast cancer subtypes (38, 39, 67). Our data now show that, given the right combination of oncogene and cooperating epigenetic alteration, basal cells can also be the cell of origin of luminal tumors. Interestingly, cross-species comparison indicated that *Pik3ca*/EpiDriver-mutant mouse tumors share several dysregulated pathways with human luminal B tumors. This supports the idea that the ultimate epigenomic, transcriptomic, and histopathologic characteristics of a tumor depend on the target cell for the initial mutation, the type of mutations, and the collaborating alterations. Clearly, loss of epigenetic regulation needs to be considered as a significant contributor to the loss of lineage integrity that underlies tumor heterogeneity.

METHODS

Animals

Animal husbandry, ethical handling of mice, and all animal work were carried out according to guidelines approved by the Canadian Council on Animal Care and under protocols approved by the Centre for Phenogenomics Animal Care Committee (18-0272H). All mice used in experiments were female. The animals used in this study were R26-LSL-*Pik3ca*$^{H1047R/+}$ [ref. 11; Gt(ROSA)26Sor$^{tm1(Pik3ca*H1047R)Egan}$ in a clean FVBN background kindly provided by Sean E. Egan, The Hospital for Sick Children], R26-LSL-*Cas9-GFP* [Gt(ROSA)26Sor$^{tm1(CAG-xstpx-cas9,-EGFP)Fezh}$/J, #026175, in C57/Bl6 background from The Jackson Laboratory], LSL-TdTomato [B6;129S6-Gt(ROSA)26Sor$^{tm14(CAG-tdTomato)Hze}$/J, #007908 from The Jackson Laboratory], *Asxl2*$^{fl/fl}$ [C57BL/6N-Asxl2$^{tm1c(EUCOMM)Hmgu}$/Tcp generated by The Canadian Mouse Respiratory], and *Kdm6a*$^{fl/fl}$ [Kdm6a$^{tm1.1Kaig}$] mice kindly provided by Jacob Hanna, Weizmann Institute of Science. *Rb*$^{fl/fl}$;*Trp53*$^{fl/fl}$;LSL-*Cas9-EGFP* mice were generated by crossing B6.129;*Rb1*^{tm1Brn} (#026563 from The Jackson Laboratory), *Trp53*^{tm1Brn} (#008462 from The Jackson Laboratory), and Gt(ROSA)26Sor$^{tm1(CAG-xstpx-cas9,-EGFP)Fezh}$/J mice. CRISPR screens and experiments in the *Pik3ca*$^{H1047R/+}$;*Cas9* cohort were performed in an F1 FVBN/C57Bl6 background. Experiments with *Kdm6a*$^{fl/fl}$ and *Asxl2*$^{fl/fl}$ were conducted by crossing each strain to LSL-*Cas9-EGFP* mice, resulting in *Kdm6a*$^{fl/fl}$;LSL-*Cas9-EGFP* and *Asxl2*$^{fl/fl}$;LSL-*Cas9-EGFP* in a C57Bl6 background. *Kdm6a*$^{fl/fl}$ and *Asxl2*$^{fl/fl}$ were also crossed to R26-LSL-*Pik3ca*^{H1047R} mice, to obtain *Kdm6a*$^{fl/fl}$;R26-LSL-*Pik3ca*^{H1047R} and *Asxl2*$^{fl/fl}$;R26-LSL-*Pik3ca*^{H1047R} mice, which were in a mixed FVBN;C57Bl6 background. These mice were then crossed to produce *Asxl2*$^{fl/fl}$;R26-LSL-*Pik3ca*$^{H1047R/+}$;LSL-*Cas9-EGFP* and *Kdm6a*$^{fl/fl}$;R26-LSL-*Pik3ca*$^{H1047R/+}$;LSL-*Cas9-EGFP* mice, which were of mixed FVBN;C57Bl6 background. *Kdm6a*$^{fl/fl}$;LSL-*Cas9-EGFP* mice were crossed to *Krt5*-CreERT2 mice [Krt5$^{tm1.1(cre/ERT2)Blh}$, #029155 from The Jackson Laboratory] and then to *Kdm6a*$^{fl/fl}$;R26-LSL-*Pik3ca*^{H1047R} mice to generate *Kdm6a*$^{fl/fl}$;R26-LSL-*Pik3ca*^{H1047R};LSL-*Cas9-EGFP;Krt5*-CreERT2 mice. NSG mice used for xenograft experiments were NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ mice (The Jackson Laboratory; #005557). Genotyping was performed by PCR using genomic DNA prepared from mouse ear punches. Tamoxifen was diluted in corn oil at 20 mg/mL and administered as a single dose to each mouse at 75 mg/kg by intraperitoneal injection. For tumor experiments, mice were palpated for tumors weekly by experimenters blinded to the experimental group. When total tumor mass per animal exceeded 1,000 mm^3, mice were monitored biweekly and scored in accordance with standard operating procedure "#AH009 Cancer Endpoints and Tumour Burden Scoring Guidelines."

Lentiviral Constructs and Library Construction

sgRNAs targeting breast cancer long-tail genes were obtained from Hart and colleagues (ref. 68; four sgRNAs/gene), and sgNT were obtained from Sanjana and colleagues (69), ordered as a pooled oligo chip (CustomArray Inc.), and cloned into pLKO sgRNA-Cre plasmid (9) using BsmBI restriction sites. We excluded frequent and known breast cancer tumor suppressor genes such as *TP53* or *CDH1* from the breast long-tail genes library. The sgNT were designed not to target the mouse genome and served as a negative control. Individual sgRNAs used in this study as well as Tracking of Indels by DEcomposition (TIDE) primers for evaluating cutting efficiency are listed in Supplementary Table S6. pLKO-mRFP and pLKO-GFP were kindly provided by Elaine Fuchs, The Rockefeller University (RRID:Addgene_26001 and RRID:Addgene_25999). pLEX-306-iCre was cloned from pLEX-306 (RRID:Addgene_41391) by substituting the Puromycin resistance cassette with Cre. Open reading frames (ORF) for Ruby fluorescent protein or mouse *Elf5* were inserted between the gateway sites. pLKO-mRFP-P2A-Cre was recently described (9) and used for lentiviral injections in *Pik3ca*^{H1047R};*Kdm6a*$^{fl/fl}$ and *Asxl2*$^{fl/fl}$ mice.

Virus Production and Transduction

Large-scale production and concentration of lentivirus were performed as previously described (70–74). Briefly, 293T cells (Invitrogen R700-07, RRID:CVCL_6911) were seeded on poly-L-lysine–coated 15-cm plates and transfected using PEI (polyethyleneimine) method in a nonserum media with a lentiviral construct of interest along with lentiviral packaging plasmids psPAX2 (RRID:Addgene_12260) and pPMD2.G (RRID:Addgene_12259). Eight hours after transfection, media were added to the plates supplemented with 10% fetal bovine serum and 1% pencillin–streptomycin antibiotic solution (w/v). Forty-eight hours later, the viral supernatant was collected and filtered through a Stericup-HV PVDF 0.45-μm filter, and then concentrated ~2,000-fold by ultracentrifugation in an MLS-50 rotor (Beckman Coulter). Viral titers were determined by infecting R26-LSL-tdTomato mouse embryonic fibroblasts (MEF) and FACS-based quantification. *In vivo* viral transduction efficiency was determined by injecting decreasing amounts of a single viral aliquot of known titer, diluted to a constant volume of 8 μL per mammary gland, and analyzed by FACS 7 days after infection. Ad5-K5-Cre (VVC-U of Iowa-1174), Ad5-K8-Cre (VVC-Li-535), or Ad-Cre (VVC-U of Iowa-5) were purchased from the Vector Core at the University of Iowa.

Intraductal Injection and Viral Transduction

Intraductal lentiviral injection has been described. Briefly, to deliver the lentiviral sgRNA library or single sgRNAs targeting gene of interest, a noninvasive injection method was used that selectively transduces mammary epithelium of female mice. Female mice were injected at >8 and <20 weeks of age, with age at injection matched between groups in all experiments. Eight microliters of virus diluted in PBS and visualized with Fast-Green dye were injected into the third and/or fourth mammary glands using pulled glass micropipettes. As previously described (70, 72, 74), we calculated coverage based on the following parameters: mammary epithelium consists of ~3.5×10^5 cells; transduction of ~15% results in a minimal double infection rate (~1/10 infected cells); and at 15% infectivity, every gland has 50,000 infected cells, resulting in 200,000 cells in four glands of a single mouse. To ensure that at least 4,000 individual cells were transduced with a given sgRNA, a pool of 860 sgRNAs requires 3.5×10^6 cells or ~17 animals. To verify the sgRNA abundance and representation in the control and breast long-tail genes libraries, MEFs were transduced with library virus and collected 48 hours after transfection. For single sgRNA or ORF injection, lentivirus was injected at 1×10^7 pfu/mL. Ad5-K5-Cre virus was injected at 8×10^8 pfu/mL, and Ad-K8-Cre virus was injected at $3.5 \times 1{,}010$ pfu/mL, which infected ~2% to 20% of basal or luminal cells.

Deep Sequencing: Sample Preparation, Preamplification, and Sequence Processing

Genomic DNA from epithelial and tumor cells was isolated with the DNeasy Blood and Tissue Kit (Qiagen). Genomic DNA (5 μg) of each tumor was used as a template in a preamplification reaction using a unique barcoded primer combination for each tumor with 20 cycles and Q5 High-Fidelity DNA Polymerase (NEB). The following primers were used:

FW: 5′AATGATACGGCGACCACCGAGATCTACAC**TATAGCCT**ACACTCTTTCCCTACACGACGCTCTTCCGATCTtgtggaaaggacgaaaCACCG-3′

RV: 5′CAAGCAGAAGACGGCATACGAGAT**CGAGTAAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3′

The underlined bases indicate the Illumina (D501-510 and D701-712) barcode locations that were used for multiplexing. PCR products were run on a 2% agarose gel, and a clean ~200-bp band was isolated using the Zymo Gel DNA Recovery Kit as per the manufacturer's instructions (Zymoresearch Inc.). Final samples were quantitated and then sent for Illumina NextSeq sequencing (1 million reads per tumor) to the sequencing facility at the Lunenfeld-Tanenbaum Research Institute (LTRI). Sequenced reads were aligned to the sgRNA library using Bowtie version 1.2.2 with options –v 2 and –m 1. sgRNA counts were obtained using the MAGeCK count command (75).

Analysis of Genome Editing Efficiency

Tumor cells were live sorted for GFP expression, and genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). For cultured cells, genomic DNA extraction was performed on cells harvested during routine passaging. PCR was performed flanking the regions of sgRNA on genomic DNA from both wild-type cells and putative knockout cells and was sent for Sanger sequencing. Sequencing files along with chromatograms were uploaded to http://shinyapps.datacurators.nl/tide/ (76), and genome editing efficiency was estimated. TIDE primers are listed in Supplementary Table S6.

Antibodies

The following primary antibodies were used in this study: rabbit anti-APC (1:200, Santa Cruz sc-896, RRID:AB_2057493), rabbit anti-Kdm6a (1:1,000, CST D3Q1I, RRID:AB_2721244), rabbit anti-Asxl2 (1:500, EMD Millipore, ABE1320, RRID:AB_2923141), mouse anti-TP53 (1:1,000, CST 1C12, RRID:AB_331743), mouse anti-Pten (1:1,000 CST 26H9, RRID:AB_331153), goat anti-Setd2 (1:500 MilliporeSigma, SAB2501940), rabbit anti-Mll3 (1:500 CST D1S1V, RRID:AB_2799442), mouse anti-GAPDH (1:2,500 Santa Cruz sc-32233, RRID:AB_627679), rabbit anti–histone H3 (1:1,000 CST 4499, RRID:AB_10544537), rabbit anti-Keratin14 (PRB-155P, 1:200 for whole mount, 1:700 for sections, RRID:AB_292096), rat anti-Keratin8 (1:50, TROMA-1, RRID:AB_2891089), mouse anti-ERalpha (R&D Systems, RRID:AB_10890942), APC-conjugated anti-CD45 (1:500 rat monoclonal clone 30 F11, RRID:AB_10376146), APC-conjugated anti-CD31 (1:250 rat monoclonal clone MEC133, BioLegend, RRID:AB_312917), APC-conjugated anti-Ter119 (1:250 BioLegend, RRID:AB_313712), PECy7 anti-human/mouse CD49f (1:50 clone GoH3, BioLegend, RRID:AB_2561705), and APCVio770 mouse anti-CD326 EpCAM (1:50, Miltenyi, RRID:AB_2657525). Casein staining of mouse tissue used HRP-conjugated anti–β-casein (1:20 sc-166530HRP H-4). Staining of human tissues used anti-casein (polyclonal NBP2-55090, Novusbio, 1:5,000 dilution, Opal 520, RRID:AB_2923142) and anti–pan-cytokeratin (AE1AE3, Agilent DAKO, Opal 620, RRID:AB_2132885). The antibodies used for IMC were Pr14-conjugated anti–Keratin8-18 (clone C51, CST-4546BF, RRID:AB_2134843), Nd144-conjugated anti-Keratin5 (Abcam ab214586, RRID:AB_869890), and Eu151-conjugated anti-casein (Novus Biologicals NBP2-55090, RRID:AB_2923142). The antibodies used for ChIP-seq were anti-H3K27ac (Active Motif #39133, RRID:AB_2561016), anti-H3K4me1 (EpiCypher #13-0040, RRID:AB_2923143), and anti-H3K27me3 (Millipore #07–449, RRID:AB_310624).

Mammary Gland Isolation and Flow Cytometry for Lineage Tracing and Mammosphere Assay

Mice were injected with the indicated virus in the third or fourth mammary glands with no greater than two replicates of a single condition per mouse. Individual mammary glands were harvested and digested according to STEMCELL Technologies' gentle collagenase/hyaluronidase protocol. Briefly, glands were digested overnight while shaking at 37°C in 250 μL Gentle Collagenase (STEMCELL Technologies, #07919) and 2.25 mL of complete Basal Epicult media formulated according to the manufacturer's instructions (Epicult Basal Medium, STEMCELL Technologies, #05610, 10% proliferation supplement, 5% FBS, 1% penicillin–streptomycin, 10 ng/mL EGF, 10 ng/mL bFGF, 0.0004% heparin). Glands were then treated with ammonium chloride and triturated for 2 minutes in prewarmed trypsin followed by dispase. Cells were stained with CD45, CD31, Ter119, CD49f, and EPCAM for luminal and basal cell identification.

Cell Culture

Primary mouse tumor cells were isolated directly from tumors, which were minced and treated with collagenase for 45 minutes and trypsin for 10 minutes. Single-cell suspensions from tumors were sorted to isolate GFP^+ cells using FACS and were then plated. Primary mouse tumor cells were cultured in DMEM/F12 (1:1) supplemented with MEGS supplement, FBS, and penicillin–streptomycin. MCF10A-$PIK3CA^{H1047R}$ cells were purchased from Horizon (cat. #HD 101-011, RRID:CVCL_LD55, acquired in May 2018) and were cultured as previously described (77) in DMEM/F12 + 5% horse serum, 1% penicillin–streptomycin, 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, and 10 μg/mL insulin. For sgRNA transfection, cells were cultured in a monolayer for growth and transfected with a lentiviral CRISPR/Cas9 construct containing puro resistance and sgRNA targeting genes of interest. Cells were tested for cutting efficiency after selection with TIDE analysis and by Western blot. All cells were negative for *Mycoplasma* via monthly PCR testing. All cell culture experiments were conducted in less than 25 passages after either derivation from tumors (for primary mammary tumor cells) or thaw of the original

vial (for MCF10A *PIK3CA*^{H1047R} cells). Cell line authentication was not performed after receiving MCF10A *PIK3CA*^{H1047R} cells.

Xenograft Assay

MCF10A *PIK3CA*^{H1047R} cells were infected with the lentiviruses carrying Cas9 and the indicated sgRNAs as well as a puro selection marker. After puro selection and TIDE to determine the more effective guide, cells were used for sphere formation assay or xenograft. For xenograft, 500,000 cells were resuspended in 50 μL PBS, mixed 1:1 with chilled Corning Matrigel (Fisher Scientific, cat. #CB-40234), and injected into the fourth mammary fat pads of NSG mice. Mice were monitored for tumor formation by mammary gland palpation for 6 months. Each fat pad was counted individually.

Sphere Formation

For sphere experiments, MCF10A cells were plated on growth factor–reduced Matrigel (Corning, Fisher Scientific, cat. #CB-40230C) as described previously (77) and imaged by bright field after 10 days of sphere growth. Primary mammospheres were isolated from mouse mammary glands and were plated on Corning Costar Ultra-Low Attachment 24-Well Plates (CLS3473-24EA) in serum-free Epicult Basal sphere media (Epicult Basal Medium, STEMCELL Technologies, #05610, 10% proliferation supplement, 1% penicillin–streptomycin, 10 ng/mL EGF, 10 ng/mL bFGF, 0.0004% heparin, +2% W21 growth supplement). Mammospheres were counted and imaged 10 days after plating.

Immunofluorescence

Cryosections were fixed with 4% paraformaldehyde for 10 minutes. Following fixation, slides were rinsed 3 times with PBS for 5 minutes. Samples were blocked at room temperature with blocking serum (recipe: 1% BSA, 1% gelatin, 0.25% goat serum 0.25% donkey serum, and 0.3% Triton-X 100 in PBS) for 1 hour. For paraffin sections, samples were embedded in paraffin, sectioned, and rehydrated, and antigen retrieval was performed with sodium citrate buffer. Samples were incubated with primary antibody diluted in blocking serum overnight at 4°C followed by 3 washes for 5 minutes in PBS. The secondary antibody was diluted in blocking serum with DAPI and incubated for 1 hour at room temperature in the dark. Following incubation, samples were washed 3 times for 5 minutes in PBS. Coverslips were added on slides using MOWIOL/DABCO-based mounting medium and imaged under a microscope the next day. For quantification, laser power and gain for each channel and antibody combination were set using secondary-only control and confirmation with primary positive control and applied to all images.

Casein Staining of Breast Cancer Specimens, Tissue Imaging, and Analysis

Formalin-fixed, paraffin-embedded (FFPE) TMA slides were dried at 60°C for 4 hours. After drying, the slides were placed on the BOND RX^{m} Research Stainer (Leica Biosystems) and deparaffinized with BOND Dewax solution (AR9222, Lecia Biosystems). The multispectral immunofluorescent (mIF) staining process involved serial repetitions of the following for each biomarker: epitope retrieval/stripping with ER1 (citrate buffer pH 6, AR996, Leica Biosystems) or ER2 (Tris-EDTA buffer pH 9, AR9640, Leica Biosystems), blocking buffer (AKOYA Biosciences), primary antibody, Opal Polymer HRP secondary antibody (AKOYA Biosciences), and Opal Fluorophore (AKOYA Biosciences). All AKOYA reagents used for mIF staining come as a kit (NEL821001KT). Spectral DAPI (AKOYA Biosciences) was applied once slides were removed from the BOND. They were coverslipped using an aqueous method and Diamond antifade mounting medium (Invitrogen Thermo Fisher). The duplex mIF panel consisted of the following antibodies: casein (polyclonal NBP2-55090, Novusbio, 1:5,000 dilution, Opal 520) and pan-cytokeratin (AE1AE3, Agilent DAKO, Opal 620).

Slides were imaged on the Vectra Polaris Automated Quantitative Pathology Imaging System (AKOYA Biosciences). Further analysis of the slides was performed using inForm Software v2.4.11 (AKOYA Biosciences). Whole TMA spectral unmixing was achieved using the synthetic spectral library supplied within inForm. The operator then created a batch TMA map, which encircles each TMA core as its own individual region of interest. Next, a unique algorithm was created using a machine learning technique, in which the operator selects positive and negative cell examples for each marker. These algorithms were then batch applied across the entire TMA. The operator then conducted a visual review of the phenotyping across all cores to ensure accuracy. Finally, the individual files resulting from the batch analysis were consolidated in RStudio using phenoptr reports to determine the percentage of total casein per TMA core, and this information was aligned with known clinical data.

Mammary Gland Whole Mount

Mammary gland whole mounts were prepared as previously described for visualization of endogenous proteins and fluorescent labeling (78). Briefly, 2-mm^3 pieces of the mammary gland were fixed for 45 minutes in 4% PFA, followed by a 30-minute wash in WB buffer, 2 hours in WB1, and an overnight incubation in anti-Keratin8 and anti-Keratin14 antibodies diluted in WB2 buffer. The following day, the pieces underwent 3 × 1 hour washes in WB2 buffer prior to overnight incubation in secondary antibody (at 1:200 dilution) with DAPI added at 4°C. Finally, pieces were washed 3 times for 1 hour each and then cleared using FUnGI solution for 2+ hours at room temperature until glands appeared sufficiently cleared, and then were mounted and imaged using confocal microscopy.

RNA-seq and GSEA

Tumors were minced and treated with collagenase for 45 minutes and trypsin for 15 minutes. Single-cell suspensions from tumors were sorted to isolate GFP^{+} cells using FACS. RNA was extracted from FACS-isolated cells using Quick-RNA Plus Mini Kit (Zymoresearch Inc., #R1057) as per the manufacturer's instructions. RNA quality was assessed using an Agilent 2100 Bioanalyzer, with all samples passing the quality threshold of RNA integrity number score of >7.5. The library was prepared using an Illumina TrueSeq mRNA sample preparation kit at the LTRI sequencing facility, and complementary DNA was sequenced on an Illumina NextSeq platform. For *in vivo* mouse tumor samples, sequencing reads were aligned to mouse genome (mm10) using Hisat2 version 2.1.0. For cultured cells, human and mouse RNA-seq datasets were aligned using STAR v2.5.1b (79) to hg38 + GENCODE v27 and mm10 + GENCODE vM4, respectively. Counts were obtained using featureCounts (Subread package version 2.0.0) with the settings -s2 and -t gene (80). Differential expression was performed using DESeq2 (81) release 3.8. GSEA was performed using GSEA version 4.0, utilizing gene sets obtained from the Molecular Signatures Database (MSigDB; ref. 82). GSEA lists were weighted by −log(*P*)*sign(FC) for mouse tumors, mouse cells, and MCF10A-PIK3CA^{H1047R} cells. For integration with human and existing mouse tumor models, clustering was conducted after normalization and filtering only for intrinsic genes as described previously (83, 84). Metascape analysis was performed using default settings (85). g:Profiler (86) was run using the following parameters: version e104_eg51_p15_3922dba; ordered: true; sources: GO:MF, KEGG, REAC, HPA, HP; with all other parameters at default settings. Gene sets are available in Supplementary Table S7.

ChIP-seq Sample Preparation and Sequencing

For ChIP-seq, two biological replicates (separately cultured cell populations) of wild-type and *Kdm6a*-mutant mouse mammary

tumor cells and separate clones of wild-type and *KDM6A*-mutant MCF10A-HR cells were cross-linked with 1% formaldehyde in Solution A (50 mmol/L Hepes–KOH, 100 mmol/L NaCl, 1 mmol/L EDTA, and 0.5 mmol/L EGTA) for 10 minutes at room temperature. Fixation was stopped by the addition of glycine at a final concentration of 125 mmol/L. Fixed cells were washed with PBS and lysed using low SDS Chromatin EasyShear Kit (Diagenode, #C01020013) following the manufacturer's instructions. Briefly, cells were resuspended in Lysis Buffer iL1b, incubated for 20 minutes at 4°C on a rotator, and pelleted by centrifugation at 500 × *g* for 5 minutes at 4°C. Cells were resuspended in Lysis Buffer iL2 and incubated for 10 minutes at 4°C while rotating. After centrifugation of 5 minutes at 500 × *g* at 4°C, cell pellets were resuspended in iS1b Shearing Buffer (Diagenode, #C01020013) supplemented with Protease Inhibitor Cocktail (Roche). Chromatin was shared into 200- to 500-bp fragments with 8 cycles of 30 seconds sonication and 30 seconds of pause at 4°C using the Bioruptor Pico sonicator (Diagenode). Chromatin was clarified by centrifugation at 21,000 × *g* at 4°C for 10 minutes. An aliquot of 50 μL of shared chromatin from each sample was removed for input DNA extraction. For each ChIP, chromatin lysates from ~6 million cells were combined with 10 μg of anti-H3K27ac (Active Motif, #39133), anti-H3K4me1 (EpiCypher, #13-0040), or anti-H3K27me3 (Millipore, #07-449) antibodies and incubated overnight rotating at 4°C. Chromatin-antibody lysates were then incubated for 4 hours with 100 μL of Dynabeads protein G beads (Thermo Fisher, #10004D) preblocked with 0.5 mg/mL BSA while rotating at 4°C. Beads were collected with a magnetic separator (Invitrogen DynaMag-2), washed 6 times with RIPA buffer (50 mmol/L Hepes–KOH, pH 7.5; 500 Mm LiCl; 1 mmol/L EDTA; 1% NP-40 or Igepal CA-630; and 0.7% Na–Deoxycholate) and once with TBS (20 mmol/L Tris–HCl, pH 7.6; 150 mmol/L NaCl), and resuspended in ChIP Elution buffer (50 mmol/L Tris–HCl, pH 8; 10 mmol/L EDTA; and 1% SDS). Cross-linking was reversed by incubating the beads at 65°C for 16 hours. Cellular proteins and RNA were digested with Proteinase K (Invitrogen, #25530049) and RNaseA (Ambion, #2271). ChIP and input DNA were purified with phenol:chloroform:isoamyl alcohol (25:24:1) extraction and ethanol precipitation, and used for ChIP-seq library preparation with NEBNext Ultra II DNA Library Prep Kit (NEB, #E7645S). In brief, ChIP and input DNA samples were blunt-end repaired and ligated to Illumina sequencing adapters containing uracil hairpin loop structure and 3′ T overhangs (NEB, #E7337A). Looped adapter sequences were opened by removal of uracil from hairpin structures by adding 3 units of USER enzyme (Uracil-Specific Excision Reagent; NEB, M5505S) and incubation at 37°C for 15 minutes. This made DNA accessible for PCR amplification with barcoded primers for Illumina sequencing (NEB, #E7335 and #E7500). Agencourt AMPure XP beads (Beckman Coulter) were used to clean up adapter-ligated DNA without size selection. PCR amplification was carried out at 98°C for 30 seconds, followed by 9 cycles of 10 seconds at 98°C and 75 seconds at 65°C, and a final 5 minutes extension at 72°C. PCR reactions were cleaned and size selected (200–500 bp) with Agencourt AMPure XP beads (Beckman Coulter). Library concentration and size distribution were assessed by Bioanalyzer High Sensitivity DNA chip (Agilent) followed by sequencing on the Illumina NovaSeq 6000 (150-bp, paired-end reads).

ChIP-seq Alignment and Peak Calling

Human and mouse in fastq format were aligned to their respective genomes (hg38 and mm10) using BWA mem v0.7.8 (87) with default settings and filtered to retain properly paired and uniquely mapping reads with the following command: Samtools view -Shb -q 5 -f 0 × 2 -F 0 × 100 -F 0 × 800. Resultant bam files were processed with Picard MarkDuplicates v2.5.0 to remove PCR and optical duplicates. Peak calling was performed with merged replicates and paired input files using MACS v2.1.2 (88) with a q-value cutoff <0.005 and a fold-enrichment cutoff >4 for punctate histone modifications (H3K27ac and H3K4me1). A fold-enrichment cutoff = 2 and –broad was used for H3K27me3 datasets. A consensus peak set was generated per histone modification by merging peak sets from wild-type and knockout conditions. Normalized signal tracks (bedgraph/bigwig) were generated during peak calling using the flags –B –SPMR. Fold change over input tracks was generated using the macs2 bdgcmp utility.

Differential Analysis of ChIP-seq Regions

Peak level read counts were obtained using bedtools multiBamCov v2.29.2. Differential ChIP enrichment was assessed using DESeq2 v1.34.0 (81). DE peaks were designated as regions passing an FDR-adjusted *P* value cutoff of <0.05 (Wald test).

Designation and Clustering of Promoter-Proximal and Distal ChIP Peaks

To properly align and cluster ChIP peaks, we overlapped all peaks with previously published accessible chromatin regions in matched human and mouse cell types (ATAC-seq and snATAC-seq; GSE89013 and ref. 29, respectively). Accessible regions were designated as distal or proximal based on a threshold of ≤2.5 kb from the nearest annotated TSS (GENCODE v27 for human, GENCODE vM4 for mouse). Accessible regions overlapping >1 differential ChIP peak were then clustered based on differential ChIP signal using deepTools v3.6.7 as follows: Differential ChIP signal was calculated genome-wide using the fold change over input tracks (described above) with the bigwigCompare utility with pseudocount values of 0.1, 0.01, and 0.05 for H3K27ac, H3K27me3, and H3K4me1 datasets, respectively. Differential signal was extracted at peak regions using computeMatrix reference point with the settings -b 6000 -a 6000 -bs 30 –missingDataAsZero –referencePoint center. Clustering was performed using the plotHeatmap utility with –kmeans 2 (number of clusters selected by visual inspection for $k = 2$–4).

Single-Cell Processing and Library Preparation

R26-LSL-*Pik3ca*$^{H1047R/+}$;LSL-*Cas9-EGFP* (*Pik3ca*^{HR}) and *Kdm6a*$^{fl/fl}$; R26-LSL-*Pik3ca*$^{H1047R/+}$;LSL-*Cas9-EGFP* (*Pik3ca*^{HR}*Kdm6a*^{KO}) were cohoused for at least 14 days prior to injection to synchronize estrus cycles, with control LSL-*Cas9-EGFP* mice housed separately due to limitations in mouse numbers per cage. Each mouse was injected with 5×10^8 pfu/mL Ad-Cre or 8×10^8 pfu/mL Ad5-Cre in the left and right fourth mammary glands. Two mice per group were harvested except for the *Pik3ca*^{HR}*Kdm6a*^{KO} sample in the K5-Cre experiment, which was performed on one mouse. Mammary gland digestion was carried out as described in the "Mammary Gland Isolation and Flow Cytometry for Lineage Tracing and Mammosphere Assay" section except two glands were pooled per mouse, and glands were digested in 2× gentle collagenase/hyaluronidase for 2 hours with trituration by P1000 pipette halfway through digestion instead of overnight. Cells were then sorted for GFP$^+$ infected cells and immediately processed for snATAC-seq or scRNA-seq according to the 10X Genomics protocol (scRNA-seq 3′ kit v.3.1 and snATAC-seq kit v1.1). Approximately 5,000 cells per sample were sequenced with targeted 50,000 reads/cell.

10X scRNA-seq Data Processing

The raw sequencing data from each channel were first aligned in Cell Ranger 4.0.0 using a customized reference based on refdata-gex-mm10-2020-A-R26 to allow quantification of EGFP expression. The EGFP reporter transgene was added to the refdata-gex-mm10-2020-A-R26 reference and rebuilt by running cellranger mkref with default parameters (10X Genomics). To minimize the batch effects from sequencing depth variation, we further used the cellranger aggr function to match the depth of mapped reads. The filtered gene-by-cell count matrices from 10X cellranger aggr underwent further quality control (QC) and were analyzed in R package Seurat (v3.2.3;

ref. 89). The merged library was first processed in Seurat with the NormalizeData (normalization.method = "LogNormalize") function. The normalized data were further linear transformed by ScaleData() function prior to dimension reduction. PC analysis was performed on the scaled data by using only the most variable 2,000 genes (identified using the default "vst" method). Cells were examined in each sample across all clusters to determine the low-quality cell QC threshold that accommodates the variation between cell types. Low-quality cells were removed with the same filtering parameters on the merged object (percent.mt ≤10 & nCount_RNA ≥2,500 & nCount_RNA <50,000 & nFeature_RNA ≥1,000). Stromal cell contamination from FACS and doublet clusters was removed to keep only mammary epithelium cells. The QCed merged dataset was further integrated using the RunHarmony()function in SeuratWrappers R package to minimize the batch effect between the Ad-Cre batch and K5-Cre batch. Top 30 harmony PCs were used for subsequent UMAP embedding and neighborhood graph construction of the integrated dataset. To investigate Ad-Cre and K5-Cre separately, the QCed dataset was split into Ad-Cre and K5-Cre subsets and then reprocessed as described above and clusters were labeled with cell types based on marker gene expression and sample/library identity. The first 30 PCs in the K5-Cre subset and the first 40 PCs in the Ad-Cre subset were selected as significant PCs for downstream UMAP embedding and neighborhood graph construction in Seurat. Pseudotime analysis was performed using Monocle 3 on K5-Cre basal cells. A central point within the wild-type control cluster was set as the root node, and pseudotime was calculated with automatic partitioning. The proliferating cluster was included for pseudotime calculation but excluded from pseudotime visualization. The HS-ML cluster was portioned separately from the remaining cells and was excluded from visualization. Diffusion mapping was performed on epithelial cells excluding the HS-ML cluster using the destiny package (v3.4.0). The first three eigenvectors were used for visualization using the plot3d package.

Cerebro Shinyapp of scRNA-seq Data

Final processed Seurat objects from the harmony integrated dataset, Ad-Cre subset, and K5-Cre subset were further processed using the cerebroApp functions in the cerebroApp R package (v1.3.0; ref. 90). Cerebro processed data were hosted on shinyapps.io server, accessible through https://wahl-lab-salk.shinyapps.io/Kdm6aKO_scRNAseq/.

10X snATAC-seq Data Processing

The raw sequencing data from each mouse were first processed separately in 10X cellranger-atac 1.2.0 pipeline using refdata-cellranger-atac-mm10-1.2.0 reference. To minimize the batch effects from sequencing depth variation, we further used the aggr function in cellranger-atac pipeline to match the depth of mapped reads across samples. The postnormalization fragments output from the 10X cellranger-atac aggr pipeline were imported into ArchR (91) and further QCed and analyzed. Arrow files were created with the initial filtering: minTSS = 4 and minFrags =1,000. Each library was inspected separately to determine the QC filtering thresholds. All samples were further QC filtered with TSS enrichment >6 and $\log_{10}$(nFrags) ≥3.4 with the exception of the wild-type sample, which used a higher threshold $\log_{10}$(nFrags) ≥3.55). The merged samples were first embedded in UMAP by running latent semantic indexing with 1 iteration with the iterative latent semantic indexing (LSI) function. Clusters identified were inferred based on the gene score of marker genes. Clusters of doublets were marked by shared marker gene expression from two different lineages and a higher number of reads per cell on average as previously described (29). Clusters of stromal cell contamination and doublets were removed from subsequent analysis based on gene expression and average read-depth distribution as previously described (29). The cleaned mammary epithelial cell dataset was reprocessed through a 1-iteration interative LSI with default parameters. All top 20 PCs were used to embed cells in 2D UMAP. Clusters were called by using the addClusters (method = 'Seurat',resolution = 1.1,dimsToUse = 1:20) function and subsequently labeled with cell types using gene scores of marker genes and sample identity. Pseudobulk profile with replicates was generated, and reproducible peaks were identified by calling peaks specifically in each cluster or cell type across replicates using the macs2 method. Differentially accessible peaks were identified using the getMarkerFeatures and getMarkers (cutOff = "FDR ≤0.1 & Log_2FC ≥1"). Transcription factor motif activity was inferred by using the chromVAR transcription factor enrichment deviation z-scores in ArchR (30, 91). Heat maps were generated using the ComplexHeatmap R package using scaled and centered values across cell type groups (92).

TCGA and METABRIC Data

Clinical and pathologic data, somatic genetic mutations, and genomic copy numbers were obtained from the cBioPortal (93). Gene expression (RNA-seq fragments per kilobase of transcript per million mapped reads upper quartile normalized) data were obtained from the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov). In survival analyses, EpiDriver mutations were defined as somatic gene mutations and/or homozygous genomic deletions of *ASXL2*, *BAP1*, *KMT2C*, *KMT2D*, *KDM6A*, and/or *SETD2*. The TCGA breast cancers were previously scored for PI3K/AKT/mTOR signaling using a transcription-based CMAP signature, in which high values were associated with poor outcome (94). The measures of phospho-Ser473 AKT were downloaded from The Cancer Proteome Atlas (45) and corresponded to level 4 normalized values from assays using reverse-phase protein arrays. The high/low threshold (value = 0) of CMAP and pAKT was confirmed by examining the value distributions in all primary tumors. The Kaplan–Meier curve and log-rank test analyses were performed in R software using the *survival* and *survminer* packages. The signature expression scores were derived from the combined expression analysis of the corresponding gene constituents using the single-sample GSEA algorithm (95), calculated with the gene set variation analysis application (96). Pregnancy, lactation, involution, and alveogenesis gene signatures corresponded to Gene Ontology Biological Processes terms and to gene sets defined in the study of mouse mammary development (36, 97, 98). The genes in signature can be found in Supplementary Table S7, the MsigDB, and the corresponding referenced papers (36, 97, 98).

Transcriptomic Analyses of Ductal Carcinoma vs. Invasive Breast Cancer

Gene expression data from 57 DCIS and 313 IDC were obtained and processed as previously described (44). For each gene, standardized gene expression values were calculated by subtracting the mean (across all samples) from the sample's gene expression value and then dividing by the standard variation. Signature Z-scores were calculated as the mean of standardized gene expression for all genes included in the signature and present in the dataset. The genes in each signature can be found in Supplementary Table S7 or can be found by name in the MSigDB.

IMC Staining

Immunofluorescence was used to validate antibodies, and metal conjugations were carried out using Maxpar Conjugation Kits (Fluidigm). FFPE slides were baked for 1 hour at 60°C, deparaffinized using xylene washes, and rehydrated in an ethanol gradient (100%, 95%, 80%, and 75%). Heat-induced antigen retrieval was performed using antigen retrieval buffer (Tris-EDTA pH 9.2) at 95°C for 30 minutes. Slides were blocked at room temperature for 1 hour using a blocking solution (3% BSA, 5% horse serum, 0.1% Tween in

TBS) followed by overnight incubation at 4°C with a panel of metal-conjugated antibodies. The following day, slides were washed using TBS and DNA staining was performed using iridium in TBS for 5 minutes at room temperature. Slides were washed 3 times in TBS and dipped in milliQ before being air-dried. The Hyperion Imaging System (Fluidigm) was calibrated using a tuning slide, and IMC images were acquired at 1-μm resolution at 200 Hz.

IMC Data Analysis Pipeline

Data were preprocessed, segmented, and analyzed using an in-house integrated flexible data analysis pipeline, ImcPQ, available upon request. The analysis pipeline was implemented in Python.

Briefly, data were converted to TIFF format and segmented into single cells using the pipeline to classify pixels based on a combination of antibody stains to identify membranes/cytoplasm and nuclei. The stacks were then segmented into single-cell object masks. Single cells were clustered into cell categories based on prespecified markers and cell phenotypes.

IMC raw data were converted to TIFF format without normalization. The ImcPQ pipeline was used for segmentation and to process images to single-cell data. Then, based on membranes/cytoplasm and nuclei markers, the analysis stacks were generated. First, image layers, or channels, were split into nuclear or cytoplasm/membrane channels and added together to sum all markers that represent nuclei or cytoplasm/membrane. Then Mesmer model (99) was used for segmentation as a deep learning method. The resulting single-cell mask was used to quantify the expression of each marker of interest and spatial features of each cell. Single-cell marker expressions were summarized by mean pixel values for each channel. The single-cell data were normalized and scaled per marker channel. Then data were censored at the 99th percentile to remove outliers.

Clusters of interest KRT5$^+$, KRT8–18$^+$, and double-positive population were gated based on the phenotypes. For quantification, the normalized density of marker in gated cell populations is reported.

Statistics and Reproducibility

All quantitative data are expressed as the mean±SE. Significance of the difference between groups was calculated by a two-tailed Student *t* test (with Welch correction when variances were significantly different), Wilcoxon rank-sum test (when data were not normally distributed), or log-rank test for survival data using Prism 7 (GraphPad Software) unless otherwise specified in the figure legends. Where adjustment is indicated and the method is not otherwise specified, the *P* value was adjusted using Bonferroni correction.

Data Availability

All RNA-seq, scRNA-seq, snATAC-seq, and ChIP-seq data are available at NCBI Gene Expression Omnibus (GEO) under GEO accession GSE178424. Cerebro processed data are hosted on the shinyapps.io server and are accessible through https://wahl-lab-salk.shinyapps.io/Kdm6aKO_scRNAseq/.

Authors' Disclosures

E. Langille reports grants from the Government of Ontario and the Canadian Institutes of Health Research during the conduct of the study. T. Nguyen reports other support from the Lunenfeld-Tanenbaum Research Institute–Mount Sinai during the conduct of the study. G.D. Bader reports personal fees from Adela Bio and grants and personal fees from Deep Genomics outside the submitted work. D.W. Cescon reports grants from the Terry Fox Research Institute and the Canadian Institutes of Health Research during the conduct of the study; grants, personal fees, and other support from AstraZeneca and Pfizer, personal fees from Exact Sciences, Eisai, and Novartis, personal fees and other support from Gilead, GSK, Merck, and Roche, and other support from Inivata outside the submitted work; and a patent (US62/675,228) for methods of treating cancers characterized by a high expression level of spindle and kinetochore associated complex subunit 3 (ska3) gene issued. J.L. Wrana reports grants from the Canadian Institutes of Health Research, the Canadian Cancer Society Research Institute, and the Terry Fox Research Institute during the conduct of the study. H.W. Jackson reports other support from Standard BioTools and Abcam during the conduct of the study, as well as personal fees and nonfinancial support from Standard BioTools outside the submitted work. D. Schramek reports grants from the Susan G. Komen Foundation, the Terry Fox Research Institute, and the Nicol Family Foundation during the conduct of the study, as well as personal fees from Tango Therapeutics outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

E. Langille: Validation, investigation. **K.N. Al-Zahrani:** Investigation. **Z. Ma:** Formal analysis. **M. Liang:** Investigation. **L. Uuskula-Reimand:** Investigation. **R. Espin:** Formal analysis. **K. Teng:** Formal analysis. **A. Malik:** Formal analysis. **H. Bergholtz:** Data curation, formal analysis. **S. El Ghamrasni:** Data curation. **S. Afiuni-Zadeh:** Data curation, formal analysis. **R. Tsai:** Investigation. **S. Alvi:** Investigation. **A. Elia:** Investigation. **Y. Lu:** Data curation. **R.H. Oh:** Investigation. **K.J. Kozma:** Formal analysis. **D. Trcka:** Investigation. **M. Narimatsu:** Data curation, investigation. **J.C. Liu:** Formal analysis. **T. Nguyen:** Formal analysis, investigation. **S. Barutcu:** Formal analysis, investigation. **S.K. Loganathan:** Validation. **R. Bremner:** Conceptualization, investigation. **G.D. Bader:** Conceptualization. **S.E. Egan:** Conceptualization, formal analysis. **D.W. Cescon:** Data curation, investigation. **T. Sorlie:** Data curation. **J.L. Wrana:** Conceptualization, formal analysis. **H.W. Jackson:** Conceptualization. **M.D. Wilson:** Data curation, supervision. **A.K. Witkiewicz:** Data curation, investigation. **E.S. Knudsen:** Conceptualization. **M.A. Pujana:** Conceptualization, data curation, investigation. **G.M. Wahl:** Conceptualization, data curation, formal analysis, supervision, investigation, writing–original draft. **D. Schramek:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, writing–original draft, project administration.

Acknowledgments

We thank all members of our laboratories for helpful comments, with additional thanks to K. Schleicher, and G. Mbamalu for their insight and assistance. We thank H. Melo and D. Durocher for assistance with the visualization of g:Profiler data. We also thank the Centre for Phenogenomics, Network Biology Collaborative Centre, and Flow Cytometry Facility at the Lunenfeld-Tanenbaum Research Institute as well as the Flow Cytometry Facility at the University of Toronto. This work was supported by a Career Catalyst Research Grant to D. Schramek from the Susan G. Komen Foundation (CCR16377321), a Terry Fox Research Institute Program Projects Grant to J.L. Wrana and D. Schramek and colleagues (TFRI Project #1107), and by the Nicol Family Foundation. E. Langille is a recipient of the Ontario Graduate Scholarship and the Frank Fletcher Memorial Fund, K.N. Al-Zahrani is a recipient of the Medicine By Design fellowship and supported by a Sinai Health System Foundation donation (Mr. Ah Shai), S.K. Loganathan is a Canadian Cancer Society Fellowship recipient (BC-F-16#31919), and L. Uuskula-Reimand is a recipient of the Next Generation of Scientists Scholarship from the Cancer Research Society (PIN25558). G.M. Wahl and Z. Ma are supported by a Cancer Center Core Grant (5 P30CA014195), the NIH/NCI (R35 CA197687), and the Breast Cancer Research Foundation. S.E. Egan is supported by the Canadian Institutes of Health Research. R. Espin and M.A. Pujana were supported by grants from the Carlos III Institute of Health (PI18/01029 and PI21/01306; cofunded by European Regional Development Fund, a way to build Europe), Generalitat de Catalunya (SGR 2017–449), and the CERCA Program to IDIBELL.

The publication costs of this article were defrayed in part by the payment of publication fees. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

Note

Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

Received June 30, 2021; revised July 15, 2022; accepted September 13, 2022; published first September 15, 2022.

REFERENCES

1. Mateo J, Steuten L, Aftimos P, Andre F, Davies M, Garralda E, et al. Delivering precision oncology to patients with cancer. Nat Med 2022;28:658–65.
2. Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;153:17–37.
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. Science 2013;339:1546–58.
4. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in cancer and somatic tissues. Cell 2017;171:1029–41.
5. Castro-Giner F, Ratcliffe P, Tomlinson I. The mini-driver model of polygenic cancer evolution. Nat Rev Cancer 2015;15:680–5.
6. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci U S A 2010;107:18545–50.
7. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, et al. Passenger mutations in more than 2,500 cancer genomes: overall molecular functional impact and consequences. Cell 2020;180:915–27.
8. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in The Cancer Genome Atlas. Cell 2018;173:321–37 e10.
9. Loganathan SK, Schleicher K, Malik A, Quevedo R, Langille E, Teng K, et al. Rare driver mutations in head and neck squamous cell carcinomas converge on NOTCH signaling. Science 2020;367:1264–9.
10. Ablain J, Durand EM, Yang S, Zhou Y, Zon LI. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. Dev Cell 2015;32:756–64.
11. Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive molecular portraits of invasive lobular breast cancer. Cell 2015;163:506–19.
12. Adams JR, Xu K, Liu JC, Agamez NM, Loch AJ, Wong RG, et al. Cooperation between Pik3ca and p53 mutations in mouse mammary tumor formation. Cancer Res 2011;71:2706–17.
13. Pereira B, Chin SF, Rueda OM, Vollan HK, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. Nat Commun 2016;7:11479.
14. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495–501.
15. Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. Science 2016;352:aad9780.
16. Steffen PA, Ringrose L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. Nat Rev Mol Cell Biol 2014;15:340–56.
17. Wang L, Zhao Z, Ozark PA, Fantini D, Marshall SA, Rendleman EJ, et al. Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. Nat Med 2018;24:758–69.
18. Campagne A, Lee MK, Zielinski D, Michaud A, Le Corre S, Dingli F, et al. BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. Nat Commun 2019;10:348.
19. Zentner GE, Tesar PJ, Scacheri PC. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res 2011;21:1273–83.
20. Huang C, Zhu B. Roles of H3K36-specific histone methyltransferases in transcription: antagonizing silencing and safeguarding transcription fidelity. Biophys Rep 2018;4:170–7.
21. Meyer DS, Koren S, Leroy C, Brinkhaus H, Muller U, Klebba I, et al. Expression of PIK3CA mutant E545K in the mammary gland induces heterogeneous tumors but is less potent than mutant H1047R. Oncogenesis 2013;2:e74.
22. Gazova I, Lengeling A, Summers KM. Lysine demethylases KDM6A and UTY: the X and Y of histone demethylation. Mol Genet Metab 2019;127:31–44.
23. Andricovich J, Perkail S, Kai Y, Casasanta N, Peng W, Tzatsos A. Loss of KDM6A activates super-enhancers to induce gender-specific squamous-like pancreatic cancer and confers sensitivity to BET inhibitors. Cancer Cell 2018;33:512–26.
24. Tran TH, Utama FE, Lin J, Yang N, Sjolund AB, Ryder A, et al. Prolactin inhibits BCL6 expression in breast cancer through a Stat5a-dependent mechanism. Cancer Res 2010;70:1711–21.
25. Logarajah S, Hunter P, Kraman M, Steele D, Lakhani S, Bobrow L, et al. BCL-6 is expressed in breast cancer and prevents mammary epithelial differentiation. Oncogene 2003;22:5572–8.
26. Dontu G, Abdallah WM, Foley JM, Jackson KW, Clarke MF, Kawamura MJ, et al. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev 2003;17:1253–70.
27. Bach K, Pensa S, Zarocsinceva M, Kania K, Stockis J, Pinaud S, et al. Time-resolved single-cell analysis of Brca1 associated mammary tumourigenesis reveals aberrant differentiation of luminal progenitors. Nat Commun 2021;12:1502.
28. Pervolarakis N, Nguyen QH, Williams J, Gong Y, Gutierrez G, Sun P, et al. Integrated Single-cell transcriptomics and chromatin accessibility analysis reveals regulators of mammary epithelial cell identity. Cell Rep 2020;33:108273.
29. Chung CY, Ma Z, Dravis C, Preissl S, Poirion O, Luna G, et al. Single-cell chromatin analysis of mammary gland development reveals cell-state transcriptional regulators and lineage relationships. Cell Rep 2019;29:495–510.
30. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods 2017;14:975–8.
31. Oakes SR, Naylor MJ, Asselin-Labat ML, Blazek KD, Gardiner-Garden M, Hilton HN, et al. The Ets transcription factor Elf5 specifies mammary alveolar cell fate. Genes Dev 2008;22:581–6.
32. Farnie G, Clarke RB. Mammary stem cells and breast cancer–role of Notch signalling. Stem Cell Rev 2007;3:169–75.
33. Bouras T, Pal B, Vaillant F, Harburg G, Asselin-Labat ML, Oakes SR, et al. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell Stem Cell 2008;3:429–41.
34. Gu B, Watanabe K, Sun P, Fallahi M, Dai X. Chromatin effector Pygo2 mediates Wnt-notch crosstalk to suppress luminal/alveolar potential of mammary stem and basal cells. Cell Stem Cell 2013;13:48–61.
35. Gallego-Ortega D, Ledger A, Roden DL, Law AM, Magenau A, Kikhtyak Z, et al. ELF5 drives lung metastasis in luminal breast cancer through recruitment of Gr1+ CD11b+ myeloid-derived suppressor cells. PLoS Biol 2015;13:e1002330.
36. Valdes-Mora F, Salomon R, Gloss BS, Law AMK, Venhuizen J, Castillo L, et al. Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype. Cell Rep 2021;35:108945.
37. Tao L, van Bragt MP, Laudadio E, Li Z. Lineage tracing of mammary epithelial cells using cell-type-specific Cre-expressing adenoviruses. Stem Cell Reports 2014;2:770–9.
38. Van Keymeulen A, Lee MY, Ousset M, Brohee S, Rorive S, Giraddi RR, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature 2015;525:119–23.
39. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature 2015;525:114–8.
40. Chang Y, Zuka M, Perez-Pinera P, Astudillo A, Mortimer J, Berenson JR, et al. Secretion of pleiotrophin stimulates breast cancer

progression through remodeling of the tumor microenvironment. Proc Natl Acad Sci U S A 2007;104:10888–93.
41. Andrechek ER, Mori S, Rempel RE, Chang JT, Nevins JR. Patterns of cell signaling pathway activation that characterize mammary development. Development 2008;135:2403–13.
42. Gustin JP, Karakas B, Weiss MB, Abukhdeir AM, Lauring J, Garay JP, et al. Knockin of mutant PIK3CA activates multiple oncogenic pathways. Proc Natl Acad Sci U S A 2009;106:2835–40.
43. Croessmann S, Wong HY, Zabransky DJ, Chu D, Rosen DM, Cidado J, et al. PIK3CA mutations and TP53 alterations cooperate to increase cancerous phenotypes and tumor heterogeneity. Breast Cancer Res Treat 2017;162:451–64.
44. Bergholtz H, Lien TG, Swanson DM, Frigessi A, Oslo Breast Cancer Research C, Daidone MG, et al. Contrasting DCIS and invasive breast cancer by subtype suggests basal-like DCIS as distinct lesions. NPJ Breast Cancer 2020;6:26.
45. Li J, Lu Y, Akbani R, Ju Z, Roebuck PL, Liu W, et al. TCPA: a resource for cancer functional proteomics data. Nat Methods 2013;10:1046–7.
46. Creighton CJ, Fu X, Hennessy BT, Casa AJ, Zhang Y, Gonzalez-Angulo AM, et al. Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. Breast Cancer Res 2010;12:R40.
47. Valencia AM, Kadoch C. Chromatin regulatory mechanisms and therapeutic opportunities in cancer. Nat Cell Biol 2019;21:152–61.
48. Morgan MA, Shilatifard A. Chromatin signatures of cancer. Genes Dev 2015;29:238–49.
49. Timp W, Feinberg AP. Cancer as a dysregulated epigenome allowing cellular growth advantage at the expense of the host. Nat Rev Cancer 2013;13:497–510.
50. Plass C, Pfister SM, Lindroth AM, Bogatyrova O, Claus R, Lichter P. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. Nat Rev Genet 2013;14:765–80.
51. Zack TI, Schumacher SE, Carter SL, Cherniack AD, Saksena G, Tabak B, et al. Pan-cancer patterns of somatic copy number alteration. Nat Genet 2013;45:1134–40.
52. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. Cancer Cell 2014;25:652–65.
53. Dhar SS, Zhao D, Lin T, Gu B, Pal K, Wu SJ, et al. MLL4 is required to maintain broad H3K4me3 peaks and super-enhancers at tumor suppressor genes. Mol Cell 2018;70:825–41.
54. Alam H, Tang M, Maitituoheti M, Dhar SS, Kumar M, Han CY, et al. KMT2D deficiency impairs super-enhancers to confer a glycolytic vulnerability in lung cancer. Cancer Cell 2020;37:599–617.
55. Giraddi RR, Chung CY, Heinz RE, Balcioglu O, Novotny M, Trejo CL, et al. Single-cell transcriptomes distinguish stem cell state changes and lineage specification programs in early mammary gland development. Cell Rep 2018;24:1653–66.
56. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell Stem Cell 2019;24:65–78.
57. Dravis C, Chung CY, Lytle NK, Herrera-Valdez J, Luna G, Trejo CL, et al. Epigenetic and transcriptomic profiling of mammary gland development and tumor models disclose regulators of cell state plasticity. Cancer Cell 2018;34:466–82.
58. LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. Cancer Cell 2020;38:212–28.
59. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a high-plasticity cell state during lung cancer evolution. Cancer Cell 2020;38:229–46.
60. Carvalho J. Cell reversal from a differentiated to a stem-like state at cancer initiation. Front Oncol 2020;10:541.
61. Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, et al. Purification and unique properties of mammary epithelial stem cells. Nature 2006;439:993–7.
62. Shackleton M, Vaillant F, Simpson KJ, Stingl J, Smyth GK, Asselin-Labat ML, et al. Generation of a functional mammary gland from a single stem cell. Nature 2006;439:84–8.
63. Centonze A, Lin S, Tika E, Sifrim A, Fioramonti M, Malfait M, et al. Heterotypic cell-cell communication regulates glandular stem cell multipotency. Nature 2020;584:608–13.
64. Tharmapalan P, Mahendralingam M, Berman HK, Khokha R. Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. EMBO J 2019;38:e100852.
65. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med 2009;15:907–13.
66. Saeki K, Chang G, Kanaya N, Wu X, Wang J, Bernal L, et al. Mammary cell gene expression atlas links epithelial cell remodeling events to breast carcinogenesis. Commun Biol 2021;4:660.
67. Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, et al. BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell 2010;7:403–17.
68. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell 2015;163:1515–26.
69. Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods 2014;11:783–4.
70. Beronja S, Livshits G, Williams S, Fuchs E. Rapid functional dissection of genetic networks via tissue-specific transduction and RNAi in mouse embryos. Nat Med 2010;16:821–7.
71. Endo M, Zoltick PW, Peranteau WH, Radu A, Muvarak N, Ito M, et al. Efficient in vivo targeting of epidermal stem cells by early gestational intraamniotic injection of lentiviral vector driven by the keratin 5 promoter. Mol Ther 2008;16:131–7.
72. Beronja S, Fuchs E. RNAi-mediated gene function analysis in skin. Methods Mol Biol 2013;961:351–61.
73. Beronja S, Janki P, Heller E, Lien WH, Keyes BE, Oshimori N, et al. RNAi screens in mice identify physiological regulators of oncogenic growth. Nature 2013;501:185–90.
74. Schramek D, Sendoel A, Segal JP, Beronja S, Heller E, Oristian D, et al. Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. Science 2014;343:309–13.
75. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 2014;15:554.
76. Brinkman EK, Chen T, Amendola M, van Steensel B. Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res 2014;42:e168.
77. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods 2003;30:256–68.
78. Rios AC, Capaldo BD, Vaillant F, Pal B, van Ineveld R, Dawson CA, et al. Intraclonal plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. Cancer Cell 2019;35:618–32.
79. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15–21.
80. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 2014;30:923–30.
81. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
82. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
83. Liu JC, Voisin V, Wang S, Wang DY, Jones RA, Datti A, et al. Combined deletion of Pten and p53 in mammary epithelium accelerates triple-negative breast cancer with dependency on eEF2K. EMBO Mol Med 2014;6:1542–60.
84. Jones RA, Robinson TJ, Liu JC, Shrestha M, Voisin V, Ju Y, et al. RB1 deficiency in triple-negative breast cancer induces mitochondrial protein translation. J Clin Invest 2016;126:3739–57.

85. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
86. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 2019;47:W191–W8.
87. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60.
88. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-seq (MACS). Genome Biol 2008;9:R137.
89. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive integration of single-cell data. Cell 2019;177:1888–902.
90. Hillje R, Pelicci PG, Luzi L. Cerebro: interactive visualization of scRNA-seq data. Bioinformatics 2020;36:2311–3.
91. Granja JM, Corces MR, Pierce SE, Bagdatli ST, Choudhry H, Chang HY, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat Genet 2021;53:403–11.
92. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 2016;32:2847–9.
93. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401–4.
94. Zhang Y, Kwok-Shing Ng P, Kucherlapati M, Chen F, Liu Y, Tsang YH, et al. A pan-cancer proteogenomic atlas of PI3K/AKT/mTOR pathway alterations. Cancer Cell 2017;31:820–32.
95. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108–12.
96. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf 2013;14:7.
97. Lemay DG, Lynn DJ, Martin WF, Neville MC, Casey TM, Rincon G, et al. The bovine lactation genome: insights into the evolution of mammalian milk. Genome Biol 2009;10:R43.
98. Lemay DG, Neville MC, Rudolph MC, Pollard KS, German JB. Gene regulatory networks in lactation: identification of global principles using bioinformatics. BMC Syst Biol 2007;1:56.
99. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nat Biotechnol 2022;40:555–65.