Estrogen receptor signaling is reprogrammed during breast tumorigenesis

David Chi[a,b,1], Hari Singhal[a,1], Lewyn Li[c], Tengfei Xiao[a], Weihan Liu[c], Matthew Pun[c], Rinath Jeselsohn[c], Housheng He[d], Elgene Lim[e], Raga Vadhi[c], Prakash Rao[c], Henry Long[c], Judy Garber[a], and Myles Brown[a,c,2]

[a]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215; [b]Department of Medicine, Harvard Medical School, Boston, MA 02115; [c]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215; [d]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada; and [e]Cancer Research Division, Garvan Institute of Medical Research, Darlinghurst NSW 2010, Australia

Contributed by Myles Brown, April 17, 2019 (sent for review November 8, 2018; reviewed by Cathrin Brisken and Bert W. O'Malley)

Limited knowledge of the changes in estrogen receptor (ER) signaling during the transformation of the normal mammary gland to breast cancer hinders the development of effective prevention and treatment strategies. Differences in estrogen signaling between normal human primary breast epithelial cells and primary breast tumors obtained immediately following surgical excision were explored. Transcriptional profiling of normal ER$^+$ mature luminal mammary epithelial cells and ER$^+$ breast tumors revealed significant difference in the response to estrogen stimulation. Consistent with these differences in gene expression, the normal and tumor ER cistromes were distinct and sufficient to segregate normal breast tissues from breast tumors. The selective enrichment of the DNA binding motif GRHL2 in the breast cancer-specific ER cistrome suggests that it may play a role in the differential function of ER in breast cancer. Depletion of GRHL2 resulted in altered ER binding and differential transcriptional responses to estrogen stimulation. Furthermore, GRHL2 was demonstrated to be essential for estrogen-stimulated proliferation of ER$^+$ breast cancer cells. DLC1 was also identified as an estrogen-induced tumor suppressor in the normal mammary gland with decreased expression in breast cancer. In clinical cohorts, loss of DLC1 and gain of GRHL2 expression are associated with ER$^+$ breast cancer and are independently predictive for worse survival. This study suggests that normal ER signaling is lost and tumor-specific ER signaling is gained during breast tumorigenesis. Unraveling these changes in ER signaling during breast cancer progression should aid the development of more effective prevention strategies and targeted therapeutics.

breast cancer | estrogen receptor | mammary gland | cancer genomics | tumorigenesis

Breast cancer is the most common noncutaneous malignancy affecting women worldwide. The luminal subtype of breast tumors accounts for ~70% of all breast tumors and is primarily driven by the estrogen receptor (ER) (1). In addition, ER is the dominant target of highly effective therapies for both the prevention and treatment of ER$^+$ breast cancers (2, 3). ER also functions as a key regulator of normal mammary development and differentiation, though knowledge remains limited regarding how ER signaling is altered to become a key oncogenic driver of luminal breast cancers. To allow for insights into ER's role in tumorigenesis, characterization of ER signaling in the normal mammary epithelium and primary ER$^+$ breast tumors is needed.

ER is a ligand-dependent transcription factor that upon the binding of estrogen will bind specific targets in the genome and regulate the expression of downstream genes (4). Differences between normal and tumor ER signaling may illuminate the early steps played by ER during mammary tumorigenesis. Ultimately, understanding the normal-specific and cancer-specific ER signaling patterns of genomic regulation could facilitate the specific targeting of tumorigenic ER functions while preserving normal ER function in the treatment of breast cancer patients.

The process of breast tumorigenesis has been historically difficult to investigate because most cell culture models of normal human breast epithelial cells have not maintained expression of ER (5, 6). Reduction mammoplasty specimens thus provide a potential source of normal breast tissue. The human breast is made up not only of ER-expressing mammary epithelial cells but also contains large numbers of adipocytes and stromal cells lacking the expression of ER. The ER-expressing mammary epithelial cells make up only a small fraction of the total cells of the normal breast, thereby making it challenging to isolate sufficient numbers of ER$^+$ cells for genomic ER experiments (7). Appropriate models that allow the study of ER biology in primary tumors have also been somewhat lacking as the majority of available ER$^+$ breast cancer cell lines have been derived from late-stage cancers. To circumvent these problems, fluorescence-activated cell sorting (FACS) following the generation of a single cell suspension from reduction mammoplasty specimens was used to study the role of ER in the normal mammary gland. These studies facilitated the comparison of estrogen signaling in the normal breast with data obtained from fresh frozen ER$^+$ breast tumors.

Results

Differential Transcriptional Signatures Distinguish Normal Breast and Breast Tumor. Normal mammary epithelial cells were isolated from reduction mammoplasty specimens by FACS using a combination of EpCAM and CD49f as previously described (Fig. 1*A*) (8). The mature luminal (ML) mammary epithelial cells

Significance

Abnormal estrogen receptor (ER) signaling drives the majority of breast cancers and is targeted by endocrine therapies. However, in normal breast tissue, ER signaling has been demonstrated to promote benign functions such as development and differentiation. Using genomic techniques to characterize ER function in normal breast and breast tumors, this study reveals differential patterns of ER signaling, suggesting that normal ER signaling is lost and tumorigenic ER signaling gained during breast tumor formation. Better understanding of this process can aid the development of improved breast cancer prevention strategies and therapies.

Author contributions: D.C., H.S., H.H., E.L., and M.B. designed research; D.C., H.S., T.X., W.L., M.P., H.H., R.V., and P.R. performed research; L.L., R.J., and J.G. contributed new reagents/analytic tools; D.C., H.S., L.L., T.X., H.H., H.L., and M.B. analyzed data; and D.C., H.S., and M.B. wrote the paper.

Reviewers: C.B., ISREC School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne; and B.W.O., Baylor College of Medicine.

Conflicts of interest statement: M.B. receives sponsored research support from Novartis and serves on the scientific advisory board of Kronos Bio, Inc.

Data deposition: The sequencing data and processed files reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, https://www.ncbi.nlm.nih.gov/geo (accession no. GSE99680).

[1]D.C. and H.S. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: myles_brown@dfci.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819155116/-/DCSupplemental.

Published online May 20, 2019.

characterized by high expression of EpCAM and low expression of CD49f express the highest levels of ER, and gene expression profiling by RNA-seq from five sorted normal ML samples treated with estradiol (E2) or vehicle after FACS was performed. In addition, RNA-seq was performed on 13 primary ER$^+$ breast tumors (*SI Appendix* and Dataset S1). Unsupervised hierarchical clustering of these gene expression patterns based on the top 500 differential genes segregated the samples into tumor- and normal-specific groups (Fig. 1*B*). Furthermore, significant sample-to-sample differences were observed regardless of the presence or absence of estrogen stimulation in the normal cells, suggesting that the transcriptomes of normal breast samples consistently differ from breast tumors across multiple patient samples regardless of the estrogenic state (*SI Appendix*, Fig. S1 *A*–*C*). Collectively, RNA-seq transcriptional profiling of the normal breast, modeled by ER-expressing ML cells and ER$^+$ breast tumors, revealed consistent and significant differences between the normal and the tumor state. This analysis identifies a small cluster of genes with expression consistently higher in tumor compared with normal and two clusters of genes that are higher in normal compared with tumor (Fig. 1*C*). These gene expression differences in ER$^+$ normal breast epithelial cells and primary breast tumors are likely to arise during the process of breast tumorigenesis, prompting further investigation into whether ER signaling may play a role in mediating these differences.

Estrogen Stimulation Induces Differential Transcriptional Responses in Normal and Tumor. Given the major role of estrogen in driving luminal breast cancer and the observed transcriptional differences in normal and tumor state (Fig. 1 *B* and *C*), the transcriptional response to estrogen (E2) stimulation was explored through comparing the gene expression by RNA-seq of eight different primary breast tumor samples previously profiled by Singhal et al. (9) and five normal ML samples under E2-stimulated and vehicle control conditions (Dataset S2). E2-responsive genes in the normal and tumor samples were identified by significant expression fold change ($P < 0.05$) to generate a list of genes whose expression is altered in response to estrogen stimulation (*SI Appendix*, Fig. S2 *A* and *B*). Many of these E2-responsive genes induced in the normal breast were down-regulated in breast tumors and vice versa, suggesting differences in transcriptional response to estrogen stimulation between normal and tumor (Fig. 2*A*). Independent gene set enrichment analysis (GSEA) further corroborated these observed differences between the normal and tumor estrogen response by finding publicly available oncogenic estrogen-induced gene signatures to be enriched among gene sets induced in tumor but depleted in the normal breast following estrogen stimulation (Fig. 2 *B* and *C*). Furthermore, quantitative PCR of canonical E2-induced genes for normal and breast tumor samples in short-term cell culture supported the hypothesis of a differential estrogen response in normal and tumor states (*SI Appendix*, Fig. S2 *C* and *D*). Of this panel, PR, TFF1, and GREB1 were significantly up-regulated in the tumor compared with the normal ($P < 0.05$).

RNA-seq analysis of the estrogen response in normal breast found DLC1 to be one of the top genes induced by E2 at nearly fourfold, whereas in primary tumors, there was an equivocal response by E2 stimulation (*SI Appendix*, Fig. S2 *E* and *F*). Because DLC1 has been previously reported by others as a tumor suppressor in multiple solid tumors, it was hypothesized that DLC1 may have an E2-induced tumor suppressor function in these normal breast epithelial cells (10–14). When DLC1 was silenced by siRNA in the luminal ER$^+$ breast cancer cell line MCF-7, a proproliferative phenotype was observed with almost a doubling in the rate of E2-induced cellular proliferation compared with the control siRNA (Fig. 2 *D* and *E*). However, in the absence of E2, loss of DLC1 is not sufficient to stimulate E2-independent growth, suggesting that DLC1 can function as an inhibitor of E2-induced proliferation in ER$^+$ breast cancer. The hypothesis of DLC1's tumor suppressive role was further supported by the observations that higher expression of DLC1 in ER$^+$ breast cancer patients is associated with a significant recurrence-free survival benefit (HR = 0.67, P = 1.7E-4) (Fig. 2*F* and *SI Appendix*, Fig. S2 *G* and *H*) (15). Additional analysis of 10 normal and luminal breast cancer clinical datasets found that DLC1 is either underexpressed or lost in breast cancer compared with the normal breast (Fig. 2*G*) (16).

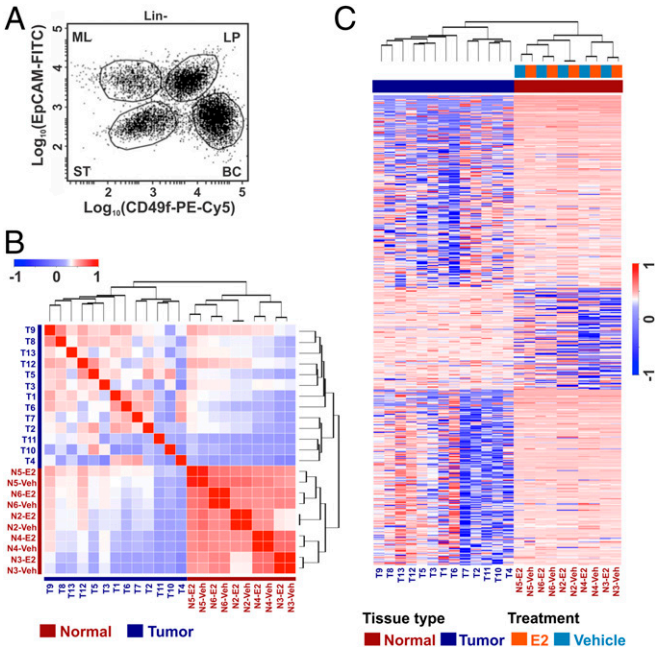

Fig. 1. ER-expressing cells from normal breast tissue and breast tumors differ significantly in gene expression. (*A*) A representative FACS contour plot displays four mammary cell subpopulations using EpCAM and CD49f antibodies. EpCAM$^-$/CD49f$^-$ represent the stromal cell (ST) subpopulation, EpCAM$^-$/CD49f$^+$ the basal cell (BC) subpopulation, EpCAM$^+$/CD49f$^+$ the luminal progenitor (LP) subpopulation, and EpCAM$^+$/CD49f$^-$ the mature luminal (ML) cell subpopulation. (*B*) Heatmap displays the sample–sample correlation following RNA-seq transcriptional profiling and unsupervised hierarchical clustering of the normal breast samples (N2–6) treated with estradiol (E2) and breast tumors (T1–13). (*C*) Heatmap displays the normalized signal intensity of read counts following RNA-seq transcriptional profiling and *k*-means (k = 3) clustering of differentially expressed genes between the normal breast (N2–6) treated with E2 and breast tumors (T1–13).

Differential ER Binding Separates Normal Breast from Breast Tumors. As ER is a ligand-activated transcriptional factor, differences in E2-responsive gene expression between the normal breast and breast tumor were hypothesized to be facilitated by differential ER chromatin binding between the two tissue types. To test the hypothesis, chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) of ER was performed on five normal breast samples and 10 ER$^+$ breast tumors to define the normal breast epithelial cell and ER$^+$ breast cancer cistromes (*SI Appendix*).

The union of all normal and tumor ER binding sites from all samples resulted in a combined set of 237,928 binding sites with the vast majority of these originating from the breast tumor ER cistromes with a relatively small contribution from the normal breast (2,145 union binding sites). This relatively lower number of ER binding sites in the normal breast is in concordance with prior reports that observed an increasing number of ER binding sites during cancer progression from primary tumors to metastases (17). As tumors have much more frequent copy number variations than normal tissue, the ChIP-seq output was normalized according to the control (input) DNA samples to account for potential confounders arising from differences in copy number for any genomic region (*SI Appendix*, Fig. S3*A*). Given the overall lower number of ER binding sites in the normal breast compared with tumor, the relevance of these normal-derived ER

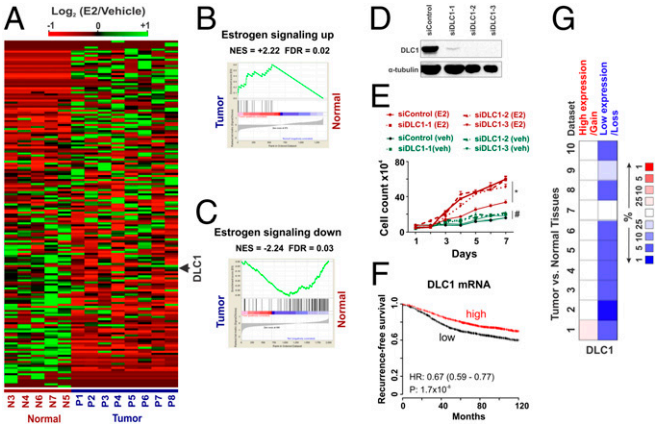

Fig. 2. Normal breast cells and breast tumors differ significantly in their transcriptional response to estrogen stimulation. (*A*) Heatmap displays the fold induction of significantly ($P < 0.05$) E2-responsive genes in RNA-seq experiments performed in normal breast cells (N3–7) and explanted breast tumors (P1–8) ranked from highest induction in tumor to lowest (9). (*B* and *C*) GSEA analyses on gene sets differentially expressed in normal and malignant breast tissues in response to estrogen stimulation. The gene sets from *A* are used for GSEA analysis. (*D*) Western blot of DLC1 following RNAi-mediated depletion of DLC1 in MCF-7 breast cancer cells. (*E*) Cell proliferation curves of MCF-7 cells in response to E2 stimulation following RNAi-mediated depletion of DLC1. (*F*) Kaplan–Meier recurrence-free survival curves of 2766 ER^+ breast cancer patients according to high and low expression levels of DLC1 (15). (*G*) Relative expression of DLC1 in breast cancer and normal breast in 10 patient tissue datasets from Oncomine (16).

binding sites was explored. When unsupervised hierarchical clustering analysis was restricted to the normal ER cistrome, it was still sufficient to distinguish the normal breast from breast tumor samples (*SI Appendix*, Fig. S3*B*). Adjustments for copy number alterations and screening for the top differential binding sites resulted in the identification of 2,750 high confidence ER binding sites (Dataset S3). Of these high-confidence ER binding sites, 1,195 sites were specific to the normal breast, 859 sites specific to the breast tumors, and 696 sites shared between both normal breast and breast tumors (Fig. 3*A*). Representative examples of these tumor-specific, normal-specific, and common ER binding sites are shown (*SI Appendix*, Fig. S3*C*). Consistent with the normal-specific expression of DLC1, ER binding to the DLC1 gene was also more frequently observed in the normal versus tumor samples (*SI Appendix*, Fig. S3*C*).

To further support the robustness of these data, similar analyses were performed utilizing the normal breast cistromes in conjunction with a previously published independent cohort of 15 ER^+ breast tumor cistromes (17). Unsupervised hierarchical clustering with these combined datasets was similarly able to distinguish between normal breast and breast tumor solely on the basis of ER binding (*SI Appendix*, Fig. S4*A*). Characterization of the ER binding sites found 1,445 sites specific to the normal breast, 591 sites specific to the breast tumors, and 312 sites shared between both normal breast and breast tumors in high concordance with the prior analysis (*SI Appendix*, Fig. S4*B*) (17). These independent analyses suggest that the differential ER binding in normal breast and breast tumors observed in this study are generalizable beyond the individual cohorts. Functional annotation of these ER binding sites to their putative gene targets was then performed by the MARGE algorithm (18). Patterns of genes regulated by differential ER binding between normal and malignant breast tissue were sufficient to segregate these two tissue types, suggesting that differential ER chromatin binding could be contributing to the differences in E2-mediated transcriptional responses reported in Fig. 1 (*SI Appendix*, Fig. S4*C*).

To identify the key ER binding sites that distinguish normal from tumor, the binding peak intensities were assessed at every binding site using a *t* test with cutoff of $P < 0.05$. These analyses revealed 291 ER binding sites with higher binding intensity in ER^+ breast tumors and 270 ER binding sites with higher binding intensity in the normal breast (Fig. 3*B*). Unsupervised hierarchical clustering of these ER binding sites was sufficient to segregate the samples into normal breast and breast tumors, demonstrating that there are consistent and significant differences in ER chromatin binding that occur during breast tumorigenesis (Fig. 3*C*).

Because it remains to be seen whether these differential ER binding sites in the tumors are associated with the functionally active genomic regions, further investigation of the enhancer-associated H3K27Ac chromatin modification was performed by H3K27Ac ChIP-seq on the breast tumor samples (*SI Appendix*). When these breast tumor H3K27Ac histone modifications were overlapped with the tumor-specific and normal-specific ER binding sites, 69% of the tumor-specific ER binding sites were marked by the H3K27Ac activating histone modification, whereas only 29% of the normal-specific ER binding sites were marked (Fig. 3*D*). The differential enrichment of tumor-specific activating epigenetic marks at the tumor ER binding sites but not normal-specific ER binding sites indicate that tumor-specific ER binding sites are transcriptionally active and functionally meaningful in the context of breast tumors.

GRHL2 Is a Tumorigenic ER-Cooperating Transcription Factor. Having identified distinct ER binding sites, motif analysis of the normal-specific and tumor-specific cistromes was performed to identify ER-cooperating transcription factors as ER binding can lead to recruitment of other transcription factors (Dataset S4) (19, 20). Interestingly, motifs for different potential cooperating transcription factors were enriched in the tumor- and normal-specific ER cistromes, indicating that differential factors may participate with ER in the two tissue types. For example, RUNX and SPDEF binding motifs were the top hits in the normal-specific ER cistrome while the ER, FOXA, GRHL2, and AR binding motifs were the top hits in the tumor-specific ER cistrome (Fig. 4*A* and Dataset S4) (17). To determine which of the transcription factors implicated by the motif analysis might be playing a

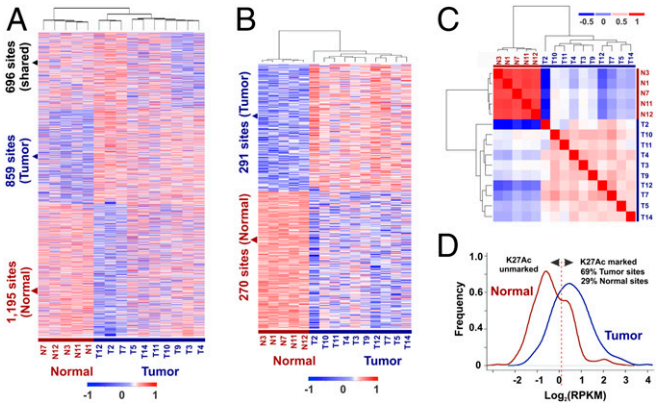

Fig. 3. ER cistromes in normal breast cells and breast tumors differentiate normal from tumor. (*A*) Heatmap displays the signal intensity of read counts following ER ChIP-seq and *k*-means clustering (k = 3) of five normal breast samples and 10 primary breast tumors for the 2,750 high-confidence ER binding sites defining the normal-specific (1,195), tumor-specific (859), and shared common (696) binding sites. (*B*) Heatmap displays the signal intensity of high-confidence differential ER binding sites following a *t* test cutoff of $P < 0.05$ to identify the high-confidence ER binding sites that are different between normal breast and breast tumor tissues. (*C*) Heatmap displays the sample–sample correlation based on the 561 high-confidence differential ER binding sites following ER ChIP-seq of five normal breast samples and 10 primary breast tumors that distinguish normal from tumor. (*D*) Aggregated H3K27Ac ChIP-seq signal from 13 breast tumors centered on the normal-specific and tumor-specific ER cistromes demonstrates tumor-specific H3K27 acetylation at the tumor-specific ER cistrome.

functional role, the list of transcription factors was compared with genome-wide CRISPR screens to identify regulated genes essential in ER$^+$ breast cancer cells (21).

In addition to the known ER$^+$ breast cancer-essential transcription factors FOXA1, GATA3, and ER, GRHL2 was significantly negatively selected in multiple ER$^+$ breast cancer cell lines (Fig. 4*B*) (21–23). Further analysis of the ER and H3K27Ac ChIP-seq datasets found a substantial overlap of the H3K27Ac histone modification and ER binding at the GRHL2 locus in each of the breast tumors, whereas there was no ER binding to the GRHL2 gene in the normal breast, indicating that GRHL2 may be a functionally active direct target of ER in tumor but not normal tissue (Fig. 3*C* and *SI Appendix*, Fig. S3*C*). Because the GRHL2 binding motif was a top hit in the motif analysis of the tumor-specific ER cistrome, a tumor-specific target of ER binding, and an essential gene in ER$^+$ breast cancer cell lines, GRHL2 likely plays an important role in driving the oncogenic functions of ER in breast cancer (Fig. 4 *A* and *B* and *SI Appendix*, Fig. S3*C*).

To investigate GRHL2's potential role in ER-mediated breast tumorigenesis, targeted CRISPR-mediated knockouts of GRHL2 in the MCF-7 cell line were performed (Dataset S5). Comparing the transcriptional output by RNA-seq of the control and GRHL2-depleted cells also revealed marked differences in their E2-responsive genes ($P = 0.002$) as well as an overall attenuation in the number of genes responsive to E2 stimulation (Fig. 4*C* and *SI Appendix*, Fig. S5 *A* and B). ER ChIP-seq of GRHL2-depleted MCF-7 demonstrates that the ER cistrome is significantly altered in the absence of GRHL2 with the observed loss of 6,640 binding sites from the control and gain of 2,975 novel binding sites (Fig. 4*D*). Furthermore, separate ChIP-seq experiments of GRHL2 and ER in the control MCF-7 cells demonstrated a significant overlap of 2,720 binding sites, suggesting that GRHL2 could potentially modulate estrogen-regulated expression of select target genes (Fig. 4*E*). To predict GRHL2-dependent direct target genes of ER binding, GRHL2-dependent ER binding sites were annotated to genes possessing transcription start sites (TSSs) within 50 kb. These genes were found to be differentially regulated in the absence of GRHL2 following estrogen stimulation, suggesting that they are transcriptionally regulated by GRHL2-dependent ER binding (Fig. 4*F*). Pathway analysis of the GRHL2-dependent direct ER targets found significant positive enrichment for cell cycling, G1/S/G2/M phase progression, DNA synthesis, and telomere maintenance in the MCF-7 control cells. However, when GRHL2 was depleted, positive enrichment of these pathways was not only absent, but pathways for cell cycle and mitotic phases were significantly negatively enriched (Dataset S5).

To understand the functional consequences of GRHL2 depletion, changes in cell proliferation of GRHL2-depleted MCF-7 cells in response to E2 stimulation was measured. While the control MCF-7 line displayed the expected increase in cell proliferation following E2 treatment, the GRHL2-depleted cell line demonstrated a minimal response in cell proliferation (Fig. 4*G*). This finding suggests that GRHL2 may be one of the mediators of E2-induced cell proliferation seen in luminal breast cancer cell lines. Interrogation of 10 clinical datasets of ER$^+$ breast cancers and normal breast found that GRHL2 is either overexpressed or copy number gained in breast tumors compared with the normal breast (Fig. 4*H*) (16). In addition, GRHL2 also appears to possess prognostic value as higher expression of GRHL2 in ER$^+$ breast tumors is significantly associated with decreased recurrence-free survival (HR = 1.27, P = 4.6E-4) (Fig. 4*I*) (15). These findings suggest a tumorigenic role of GRHL2 in ER$^+$ breast cancers.

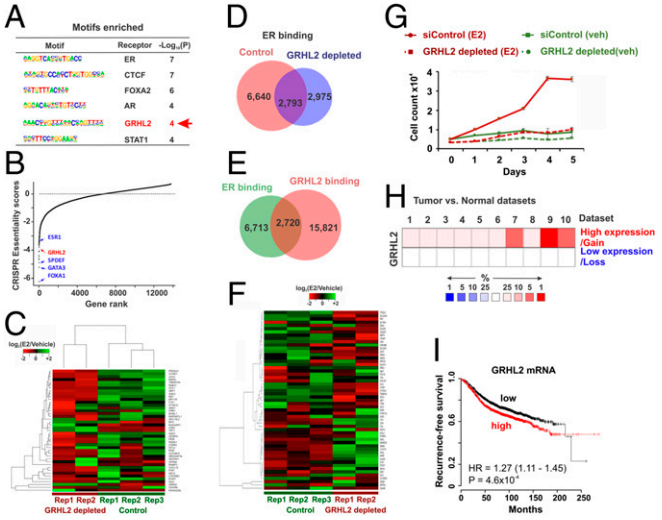

Fig. 4. GRHL2 is an ER-cooperating transcription factor necessary for estrogen-induced cell proliferation in breast cancer cells. (*A*) DNA binding motif analysis of the tumor-specific ER cistrome identifies the GRHL2 motif to be significantly enriched. (*B*) Essentiality scores of genome-wide CRISPR screening of luminal breast cancer cell lines MCF-7 and T47D identifies GRHL2 as a gene necessary for growth in the presence of estrogen (21). (*C*) Heatmap displays the expression fold change ($P < 0.05$, fold-change >1.5) of estrogen-responsive genes by RNA-seq of MCF-7 cells following targeted CRISPR knockout of GRHL2. (*D*) Venn diagram demonstrates the differences in ER binding sites in ER ChIP-seq experiments performed on MCF-7 cells following GRHL2 depletion. (*E*) Venn diagram demonstrates the overlap in binding sites between ER ChIP-seq and GRHL2 ChIP-seq experiments performed on MCF-7 cells. (*F*) Heatmap displays the expression fold change following estrogen stimulation of predicted GRHL2-dependent direct ER targets in control and GRHL2-depleted MCF-7 cells. These predicted direct ER targets are those genes with TSSs within 50 kb of a GRHL2-dependent ER binding site derived from *E*. (*G*) Cell proliferation following E2 stimulation in MCF-7 cells following GRHL2 knockout. (*H*) Relative expression of GRHL2 in breast cancer and normal breast in 10 patient tissue datasets from Oncomine (16). (*I*) Kaplan–Meier recurrence-free survival curves of 2,766 ER$^+$ breast cancer patients according to high and low expression level of GRHL2 (15).

Discussion

Estrogen signaling drives the majority of luminal breast cancers, and ER is the key target of both effective breast cancer prevention and therapy. An understanding of how estrogen signaling is altered during breast tumorigenesis is thus critical to improving the effectiveness of current ER-targeted therapies. Here, the estrogen-responsive transcriptome and ER cistrome has been characterized in the normal mammary epithelium and compared with the ER signaling network in ER$^+$ luminal breast cancers. Significant gene expression differences were found between the ER-expressing normal mammary epithelium and ER$^+$ breast cancer cells that were driven by differences in the active ER cistrome.

Many of the key transcription factors implicated in ER$^+$ luminal breast cancer, including ER itself, FOXA1, GATA3, and PDEF, are also necessary for normal mammary development (19, 24, 25). While gain-of-function ER mutations are frequently found in advanced endocrine therapy-resistant metastatic breast cancer, these mutations are rarely observed in primary breast cancers, suggesting that ER mutations do not play a major role in the progression from the normal ER$^+$ mammary epithelium to primary ER$^+$ breast cancer (26, 27). Similarly, mutations in ER coregulators and collaborating transcription factors are also rarely found in primary breast cancers (28–30). These findings suggest other mechanisms, including potentially cis-regulatory mutations or changes in epigenetic programming, may play important roles in altering ER signaling during the progression from the normal ER$^+$ mammary epithelium to ER$^+$ breast cancer.

The possibility of whether the loss of estrogen-induced tumor suppressors, gain of estrogen-regulated oncogenes, or both influenced the progression from the normal ER$^+$ mammary epithelium to ER$^+$ breast cancer was explored. These studies revealed that one of the top estrogen-induced genes in the normal breast but unresponsive in ER$^+$ tumor is DLC1, which has been previously identified in several studies as a tumor suppressor

frequently lost in solid tumors (10, 31). As a Rho-GAP protein, DLC1 catalyzes GTP hydrolysis to GDP, inhibiting Rho-GTPases such as RhoA and CDC42, which have known oncogenic functions in breast cancer (32, 33). This mechanism suggests that when DLC1 is silenced or deleted, these oncogenic Rho-GTPases would be free to promote estrogen-responsive proliferation. Here, experimental silencing of DLC1 in ER^+ breast cancer cell lines led to increased estrogen-induced proliferation, supporting the finding that DLC1 may be an estrogen-induced tumor suppressor in the normal breast. Prior studies that demonstrate DLC1 overexpression in triple-negative breast cancer cell lines also inhibits proliferation, suggesting that E2 regulation of DLC1 is likely upstream of its tumor suppressive function and is subtype dependent (10). In contrast to breast tumor cells, significant ER binding at the DLC1 locus was observed in the normal ER^+ breast epithelial cells, lending further support for DLC1's role as an estrogen-induced tumor suppressor.

As loss of tumor suppressors and/or gain of oncogenes are both involved in breast tumorigenesis, transcription factor motif enrichment analyses were performed on the normal and tumor-specific ER cistromes to identify evidence for binding of ER-associated oncogenic transcription factors. The motif analysis of the tumor-specific ER cistrome revealed the GRHL2 DNA binding motif as one of the top-most enriched motifs. GRHL2 has previously been linked to the progression of ER^+ luminal breast cancers as a potential oncogene, and the lack of enrichment of the GRHL2 motif in the normal ER cistrome further implicates GRHL2 as an ER-cooperating transcription factor driving ER^+ breast tumorigenesis (21, 34–36). In support of this hypothesis, CRISPR-mediated depletion of GRHL2 in MCF-7 cells was found to alter ER binding and the regulation of estrogen-responsive genes. The GRHL2 cistrome also significantly overlapped with ER binding sites, and these shared genomic regions correlated strongly with differentially regulated estrogen-responsive genes. Interestingly, GRHL2 appears to possess a breast cancer subtype-specific role, since GRHL2 depletion inhibited estrogen-induced proliferation in this study, consistent with previously reported findings that GRHL2 overexpression promotes cellular proliferation only in luminal breast cancer lines and not basal cancer cell lines (35, 36). When GRHL2 was overexpressed in nonluminal breast cancer lines and injected into mouse tumor xenograft models, tumor initiation was inhibited, demonstrating that GRHL2's oncogenic function could be specific to luminal breast cancers (35, 36). GRHL2's subtype specificity for the luminal phenotype may be explained by its simultaneous repression of ZEB1-mediated epithelial–mesenchymal transition of luminal breast tumors and promotion of mesenchymal–epithelial transition (34, 36–38). Additionally, the finding of higher expression of GRHL2 in luminal breast tumors relative to normal controls and poorer recurrence-free survival of patients whose tumors express higher levels of GRHL2 is in agreement with previous reports that GRHL2 locus is amplified in 29% of ER^+ breast cancers and the higher expression of GRHL2 correlates with poorer survival outcomes (15, 29, 39, 40).

One limitation of this study is the difference in the protocols used to obtain the gene expression and ER cistrome data from the normal mammary epithelial cells and the ER^+ tumors. The normal luminal mammary epithelial cells were purified by FACS as they represent only a fraction of the cells in the normal mammary gland (Fig. 1*A*). In contrast, entire sections of fresh-frozen ER^+ breast tumors were utilized without sorting as these were highly enriched for the luminal cancer cells (*SI Appendix*). The COMBAT batch correction algorithm was then employed to remove potential confounders from the differences in the protocols used (41). Following batch correction, GSEA analysis revealed robust up-regulation of oncogenic estrogen signaling in the tumors compared with the normal, supporting the biological relevance of the findings (Fig. 2 *A*–*C*).

The characterization of these contrasting functions of ER in the normal breast and breast tumors has important implications for the targeting of ER in breast cancer prevention and first-line therapy for ER^+ breast cancers. These findings suggest that improved ER-targeted therapies that retain the ability to induce the expression of tumor suppressors such as DLCI in the normal mammary epithelium while inhibiting the ability of ER to collaborate with oncogenic transcription factors such as GRHL2 in primary ER^+ breast cancers might be developed for breast cancer prevention and therapy. Optimal ER-targeted therapies would also lack the known adverse effects of ER modulation, including vasomotor symptoms, hypercoagulation, osteoporosis, and endometrial hyperplasia and cancer. A deeper understanding of ER's role in normal physiology and breast tumorigenesis should lead to the development of safer and more effective breast cancer prevention and treatment strategies.

Materials and Methods

Primary Human Breast Tissue. Histologically confirmed normal breast samples were obtained from reduction mammoplasty specimens through the pathology departments of the Brigham and Women's Hospital and Beth Israel Deaconess Medical Center. These breast specimens are in excess of normal diagnostic requirements, hence commonly discarded and completely deidentified except for patient age. Primary breast tumors were obtained from cancer surgery specimens through the Pathology Department of the Dana-Farber/Harvard Cancer Center (DF/HCC). Any leftover remaining tumor tissue following clinical requirements was made available. These specimens are deidentified except for patient age, tumor type, grade/stage, and tumor receptor status (ER/PR/HER2). All tissue was collected under the DF/HCC institutional review board-approved protocol 10-458.

FACS was used to enrich the luminal epithelial cells expressing ER (*SI Appendix*). A different protocol was used to process the fresh-frozen ER^+ tumors as single-cell dissociation, and FACS on the frozen ER^+ tumors was not performed. Given these differing experimental protocols, computational correction for batch effect was performed, and the experimentally observed findings remained biologically relevant as they capture significantly validated true biological differences via GSEA analysis (Fig. 2 *A*–*C*).

Preparation of a Human Mammary Cell Suspension for FACS Analysis. All mammary tissue was obtained from consenting patients in the absence of formalin fixative and typically processed on the same day as surgery. Mechanical dissociation into ~8-mm^3 pieces was performed and placed into charcoal-stripped tissue digest media (*SI Appendix*) to enzymatically dissociate for 5–14 h in an orbital shaking incubator set at 37 °C (8). Normal breast cells were hormone-deprived using charcoal-stripped digest media for 24 h to facilitate synchronization of ER cycling while the ER^+ primary breast tumors were frozen immediately after surgical resection (42). Following further enzymatic digestion (*SI Appendix*), a single cell suspension was obtained through filtration through a 40-μm cell strainer (BD Biosciences) and counted for FACS analysis. Charcoal-stripped digest media was utilized through the entire process to the end of FACS to ensure continuous hormone deprivation.

Immunolabeling of cells was performed at 4 °C using PBS as a buffer and wash solution at a concentration of 10^6 cells per 40 μL volume. Phycoerythrin-linked antibodies directed against human lineage markers CD31 (endothelial cells), CD45 (leukocytes), and CD235a (red blood cells) were used to prepare a nonepithelial lineage-negative cell population that was selectively depleted and filtered from the epithelial cells of interest (Dataset S6). Primary antibodies against EpCAM and CD49f were added at optimized dilutions and incubated for 25 min at 4 °C (Dataset S6). Propidium iodide (Sigma) was used for selection of only viable cells. Cell analysis and sorting was performed on either a FACS DIVA or FACS Aria (BD Biosciences), and reanalysis of the sorted populations routinely revealed a purity of more than 95%.

In Vitro Cell Culture and Assays. MCF-7 cells were cultured as previously described (43). Primary mammary epithelial and EpCAM-enriched primary tumor cells were isolated as described above and maintained in primary MEC media (*SI Appendix*) for the qPCR experiments described in *SI Appendix*, Fig. S2 *C* and *D* (Dataset S6).

Cell Proliferation Assays. Cells were plated in 24-well plates in the appropriate medium. After the cells had settled, the medium was changed to phenol red-free medium containing 10% charcoal dextran-treated FBS (CDT) for hormone deprivation, then treated with either 10 nM E2 or vehicle (EtOH). The number of viable cells was determined by Trypan blue exclusion and counted using a hemocytometer. Data represent mean ± SD from three independent replicates. *P* values were calculated using an unpaired Student's

t test. ATP-based measurements of cellular proliferation were performed by CellTiter 96 non-radioactive cell proliferation assay (Promega) and biological replicated three times.

ChIP-Seq. Immediately after FACS, the ML cells were resuspended in primary MEC media with 10 nM E2 for 45 min. Chromatin was then extracted using the truChIP chromatin shearing low-cell protocol (Covaris). Using this protocol and undergoing parameter optimization, the cells were fixed at room temperature for 2 min with 1% formaldehyde (Fisher) and quenched with 0.125 M glycine (Sigma) for 15 min. Once extracted and placed in the microtube (Covaris), the chromatin was sheared to 200–700 bp in size using the Covaris E210 ultrasonicator. The sheared chromatin was then incubated overnight with 0.3 μg of each ER antibody per million cells, HC-20 (Santa Cruz), and Ab-10 (Abcam). DNA sequencing libraries were prepared using the ThruPLEX-FD Prep kit (Rubicon Genomics) using on average three fewer amplification cycles than suggested by the protocol. Libraries were sequenced using 75-bp single-end reads on the NextSeq500 (Illumina) platform at the Dana-Farber Cancer Institute Molecular Biology Core Facility.

The fresh frozen ER^+ primary breast tumors were fresh frozen in OCT and cut for ∼10 scrolls at 20-μm thickness each. They were then fixed for 20 min at 25 °C in 1% formaldehyde (Fisher) and extracted, sonicated, and incubated with the same ER antibodies as described above. The antibody for H3K27Ac was C15410196 (Diagenode). Libraries were prepared and sequenced as previously described above.

MCF-7 cells were prepared for ChIP-seq experiments as previously described (9). The antibody used for ER was HC-20 (Santa Cruz) and for GRHL2 it was HPA004820 (Sigma). Libraries were prepared and sequenced as previously described above.

ChIP-Seq Analyses. ChIP-seq analyses were performed using ChiLin, a ChIP-seq pipeline developed at Center for Functional Cancer Epigenomics, Dana-Farber Cancer Institute (44). Short reads were aligned to the HG19 human genome using Bowtie2 and subsequently peaks were called using model-based analysis of ChIP-seq (MACS2) peak caller (45). Subsequently statistically significant peaks were selected based on the false discovery rate and the peak height of the reported peaks.

To perform clustering analyses on a group of samples, a union of all of the peaks within that group was generated. Normalization by length of peak and sequence depth was performed by normalizing read counts for each peak to per kilobase of reads per million mapped reads (RPKM), creating a read count matrix for the group of interest. Subsequently quantile normalization was applied to this read count matrix to control for outliers. Copy number variation differences between normal and tumor tissue was controlled for by generating a 15-kb window surrounding each binding site (7.5 kb on each side) and normalizing the read counts in this window from the input files. The sample–sample correlation heatmaps represent the correlation observed between any two samples. The sample-feature heatmaps represent the signal intensity of a feature for any given sample. CoverageView, GGPlot2, heatmap.2, and Pheatmap packages in R were used to build ChIP-seq heatmaps.

Peaks from all study samples were merged to create a union set of ER binding sites. Read densities were calculated for each peak for each sample in RPKM, which were used for comparison of cistromes across samples. Sample similarity was determined by hierarchical clustering using the Spearman correlation between samples. The high-confidence ER binding sites were generated by examining the top differential binding sites that met the sufficient RPKM threshold. Tissue-specific sites were then identified from these high-confidence ER binding sites by *t* test using limma with adjusted $P \leq 0.05$. DNA binding motif analysis by the motif search algorithm HOMER (v3.0.0) was then performed on these tissue-specific ER binding sites (46).

RNA-Seq. Total RNA was isolated from the experimental cell populations and primary breast tumors using TRIzol (Invitrogen). Poly-A selected, stranded RNA-seq libraries were constructed using the TruSeq Stranded mRNA kit (Illumina).

RNA-Seq Analyses. The RNA-seq analyses were performed using the VIPER analysis pipeline (47). Following alignment to hg19 with STAR, cufflinks packages were used to perform transcript assemblies. Differential gene expression calling was performed using DESEq. (48). To perform clustering analyses on a group of samples, a union of all of the genes and their expression RPKM values within that group was generated to build a read count matrix for the group of interest. The sample–sample correlation heatmaps represent the correlation observed between any two samples. The sample-feature heatmaps represent the signal intensity of a feature for any given sample. GGPlot2, heatmap.2, and Pheatmap packages in R were used to build various RNA-seq heatmaps. Functional analysis was performed using GSEA.

CRISPR Experiments. Genome-wide CRISPR screens were performed and analyzed as previously described by Xiao et al. (21). Lentiviral gRNAs targeting GRHL2 were generated by ligation of targeted oligos into LentiCRISPR-v2 vector (Addgene) linearized with BsmBI using quick ligase (NEB). Lentivirus-enriched supernatant targeting GRHL2 was then generated from 293FT cells as previously described, and MCF-7 cells were infected at a low MOI (0.3–0.5) (21). Puromycin selection was then performed after 48 h to select for cells that have a targeted depletion of GRHL2.

ACKNOWLEDGMENTS. We acknowledge the collaborating pathology departments at the Brigham and Women's Hospital and Beth Israel Deaconess Medical Center through the Dana-Farber/Harvard Cancer Center for providing the primary breast samples and clinical correlates essential to this project. Also, we thank the Flow Cytometry and the Molecular Biology Core Facilities at the Dana-Farber Cancer Institute for performing the FACS experiments and next generation sequencing. The Center for Functional Cancer Epigenetics at Dana-Farber Cancer Institute provided the computing infrastructure in our analysis of genomic data. This work was supported by funding from the NIH CA192477 to D.C. and CA080111 to M.B. and from the V Foundation to M.B.

1. M. H. Forouzanfar *et al*., Breast and cervical cancer in 187 countries between 1980 and 2010: A systematic analysis. *Lancet* **378**, 1461–1484 (2011).
2. E. V. Jensen, V. C. Jordan, The estrogen receptor: A model for molecular medicine. *Clin. Cancer Res.* **9**, 1980–1989 (2003).
3. A. Howell, The endocrine prevention of breast cancer. *Best Pract. Res. Clin. Endocrinol. Metab.* **22**, 615–623 (2008).
4. J. Liang, Y. Shang, Estrogen and cancer. *Annu. Rev. Physiol.* **75**, 225–240 (2013).
5. K. M. Vincent, S. D. Findlay, L. M. Postovit, Assessing breast cancer cell lines as tumour models by comparison of mRNA expression profiles. *Breast Cancer Res.* **17**, 114 (2015).
6. Y. Qu *et al*., Evaluation of MCF10A as a reliable model for normal human mammary epithelial cells. *PLoS One* **10**, e0131285 (2015).
7. X. Sun *et al*., Benign breast tissue composition in breast cancer patients: Association with risk factors, clinical variables, and gene expression. *Cancer Epidemiol. Biomarkers Prev.* **23**, 2810–2818 (2014).
8. E. Lim *et al*.; kConFab, Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
9. H. Singhal *et al*., Genomic agonism and phenotypic antagonism between estrogen and progesterone receptors in breast cancer. *Sci. Adv.* **2**, e1501924 (2016).
10. B.-Z. Yuan *et al*., DLC-1 gene inhibits human breast cancer cell growth and in vivo tumorigenicity. *Oncogene* **22**, 445–450 (2003).
11. S. Goodison *et al*., The RhoGAP protein DLC-1 functions as a metastasis suppressor in breast cancer cells. *Cancer Res.* **65**, 6042–6053 (2005).
12. P. Basak, R. Dillon, H. Leslie, A. Raouf, M. R. A. Mowat, The Deleted in Liver Cancer 1 (Dlc1) tumor suppressor is haploinsufficient for mammary gland development and epithelial cell polarity. *BMC Cancer* **15**, 630 (2015).
13. D. Wang, X. Qian, M. Rajaram, M. E. Durkin, D. R. Lowy, DLC1 is the principal biologically-relevant down-regulated DLC family member in several cancers. *Oncotarget* **7**, 45144–45157 (2016).
14. B. K. Tripathi, D. R. Lowy, DLC1: A tumor suppressor that regulates Rho signaling. *Oncotarget* **8**, 27674–27675 (2017).
15. B. Győrffy *et al*., An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).
16. D. R. Rhodes *et al*., ONCOMINE: A cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
17. C. S. Ross-Innes *et al*., Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
18. S. Wang *et al*., Modeling cis-regulation with a compendium of genome-wide histone H3K27ac profiles. *Genome Res.* **26**, 1417–1429 (2016).
19. A. Hurtado, K. A. Holmes, C. S. Ross-Innes, D. Schmidt, J. S. Carroll, FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
20. J. S. Carroll *et al*., Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
21. T. Xiao *et al*., Estrogen-regulated feedback loop limits the efficacy of estrogen receptor-targeted breast cancer therapy. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7869–7878 (2018).
22. M. Lupien *et al*., FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
23. R. Mehra *et al*., Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res.* **65**, 11259–11264 (2005).
24. H. Kouros-Mehr *et al*., GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* **13**, 141–152 (2008).

25. G. Buchwalter *et al*., PDEF promotes luminal differentiation and acts as a survival factor for ER-positive breast cancer cells. *Cancer Cell* **23**, 753–767 (2013).
26. R. Jeselsohn *et al*., Allele-specific chromatin recruitment and therapeutic vulnerabilities of ESR1 activating mutations. *Cancer Cell* **33**, 173–186.e5 (2018).
27. R. Jeselsohn *et al*., Emergence of constitutively active estrogen receptor-α mutations in pretreated advanced estrogen receptor-positive breast cancer. *Clin. Cancer Res.* **20**, 1757–1767 (2014).
28. S. Nik-Zainal *et al*., Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
29. D. C. Koboldt *et al*.; Cancer Genome Atlas Network Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
30. C. Curtis *et al*.; METABRIC Group, The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
31. S. L.-K. Au, C. C.-L. Wong, J. M.-F. Lee, C.-M. Wong, I. O.-L. Ng, EZH2-Mediated H3K27me3 is involved in epigenetic repression of deleted in liver cancer 1 in human cancers. *PLoS One* **8**, e68226 (2013).
32. K. O'Connor, M. Chen, Dynamic functions of RhoA in tumor cell migration and invasion. *Small GTPases* **4**, 141–147 (2013).
33. J. Ma *et al*., Role of activated Rac1/Cdc42 in mediating endothelial cell proliferation and tumor angiogenesis in breast cancer. *PLoS One* **8**, e66275 (2013).
34. X. Xiang *et al*., Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One* **7**, e50781 (2012).
35. S. Werner *et al*., Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. *J. Biol. Chem.* **288**, 22993–23008 (2013).
36. B. Cieply, J. Farris, J. Denvir, H. L. Ford, S. M. Frisch, Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. *Cancer Res.* **73**, 6299–6309 (2013).
37. B. Cieply *et al*., Suppression of the epithelial-mesenchymal transition by Grainyhead-like-2. *Cancer Res.* **72**, 2440–2453 (2012).
38. S. M. Mooney *et al*., The GRHL2/ZEB feedback loop-A key axis in the regulation of EMT in breast cancer. *J. Cell. Biochem.* **118**, 2559–2570 (2017).
39. Y. Li *et al*., Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat. Med.* **16**, 214–218 (2010).
40. N. Dompe *et al*., A whole-genome RNAi screen identifies an 8q22 gene cluster that inhibits death receptor-mediated apoptosis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E943–E951 (2011).
41. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
42. Y. Shang, X. Hu, J. DiRenzo, M. A. Lazar, M. Brown, Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852 (2000).
43. R. M. Neve *et al*., A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527 (2006).
44. Q. Qin *et al*., ChiLin: A comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. *BMC Bioinformatics* **17**, 404 (2016).
45. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
46. S. Heinz *et al*., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
47. M. Cornwell *et al*., VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. *BMC Bioinformatics* **19**, 135 (2018).
48. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).