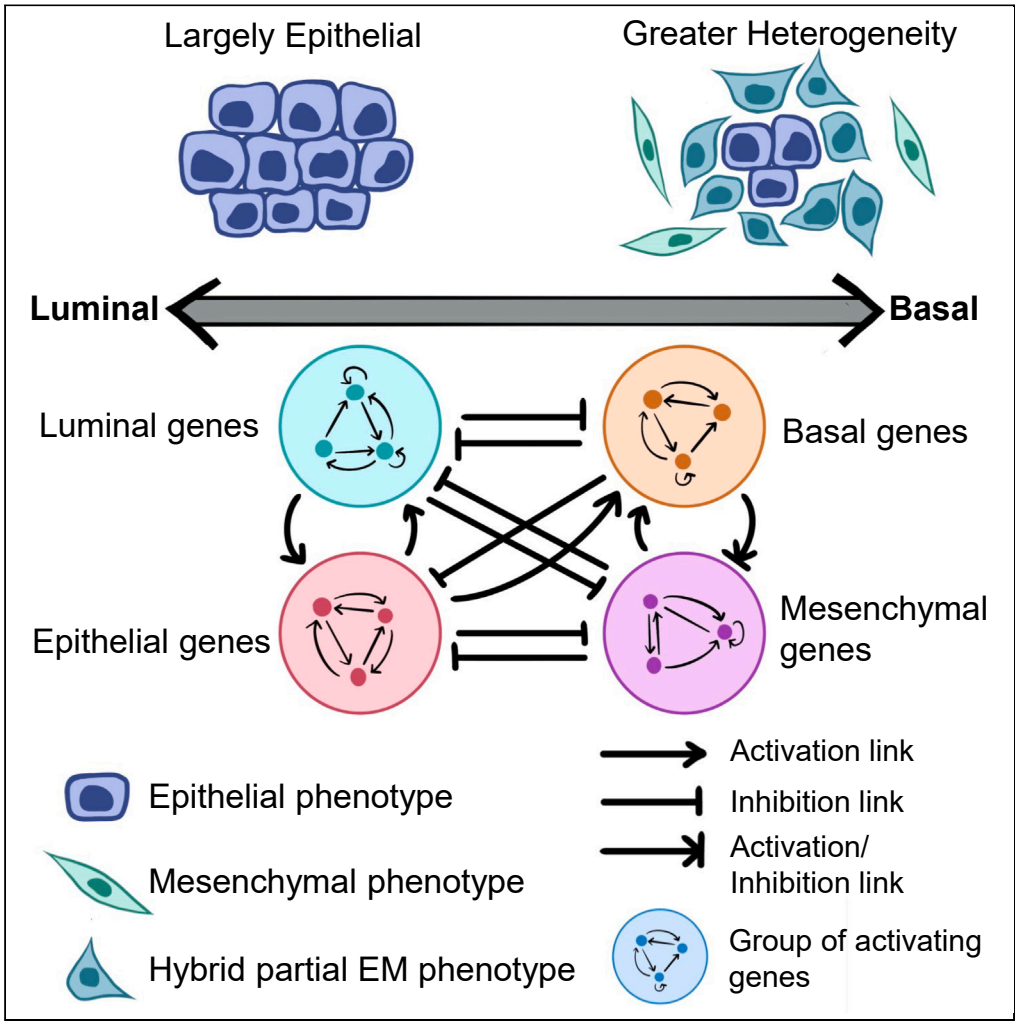
iScience

CellPress
OPEN ACCESS

Article

Increased prevalence of hybrid epithelial/mesenchymal state and enhanced phenotypic heterogeneity in basal breast cancer



Sarthak Sahoo, Soundharya Ramu, Madhumathy G. Nair, ..., Jyothi S. Prabhu, Jason A. Somarelli, Mohit Kumar Jolly

mkjolly@iisc.ac.in

Highlights

Luminal signature is closely associated with epithelial signature in breast cancer

Basal signature correlates well with a hybrid epithelial-mesenchymal signature

Basal breast cancer exhibits higher epithelial-mesenchymal heterogeneity patterns

Mathematical modeling of underlying gene networks explains observed heterogeneity

Sahoo et al., iScience 27, 110116
July 19, 2024
https://doi.org/10.1016/j.isci.2024.110116

iScience

CellPress
OPEN ACCESS

Article

Increased prevalence of hybrid epithelial/mesenchymal state and enhanced phenotypic heterogeneity in basal breast cancer

Sarthak Sahoo,[1] Soundharya Ramu,[1] Madhumathy G. Nair,[2] Maalavika Pillai,[1,6] Beatriz P. San Juan,[3] Heloisa Zaccaron Milioli,[3] Susmita Mandal,[1] Chandrakala M. Naidu,[2] Apoorva D. Mavatkar,[2] Harini Subramaniam,[1] Arpita G. Neogi,[1] Christine L. Chaffer,[3,4] Jyothi S. Prabhu,[2] Jason A. Somarelli,[5] and Mohit Kumar Jolly[1,7,*]

SUMMARY

Intra-tumoral phenotypic heterogeneity promotes tumor relapse and therapeutic resistance and remains an unsolved clinical challenge. Decoding the interconnections among different biological axes of plasticity is crucial to understand the molecular origins of phenotypic heterogeneity. Here, we use multi-modal transcriptomic data—bulk, single-cell, and spatial transcriptomics—from breast cancer cell lines and primary tumor samples, to identify associations between epithelial-mesenchymal transition (EMT) and luminal-basal plasticity—two key processes that enable heterogeneity. We show that luminal breast cancer strongly associates with an epithelial cell state, but basal breast cancer is associated with hybrid epithelial/mesenchymal phenotype(s) and higher phenotypic heterogeneity. Mathematical modeling of core underlying gene regulatory networks representative of the crosstalk between the luminal-basal and epithelial-mesenchymal axes elucidate mechanistic underpinnings of the observed associations from transcriptomic data. Our systems-based approach integrating multi-modal data analysis with mechanism-based modeling offers a predictive framework to characterize intra-tumor heterogeneity and identify interventions to restrict it.

INTRODUCTION

Intra-tumoral heterogeneity in breast cancer remains a key obstacle in the effective management of the disease.[1] A major determinant of molecular heterogeneity in breast cancer is attributed to molecular subtype characteristics, which can be broadly classified as luminal or basal.[2] In addition, cancer cells can exhibit different interconvertible cellular states along varied axes of plasticity such as epithelial-mesenchymal transition (EMT), stemness, metabolic reprogramming, and immune evasion traits to create increased overall phenotypic heterogeneity.[3–6] The extent of the crosstalk among these different axes, which is often mediated via feedback loops, can have major implications in dependence and coordination between plasticity axes on each other as well as overall disease progression.[7,8] Specifically, in breast cancer, two mainstays of molecular heterogeneity that are often used interchangeably are the luminal-basal and epithelial-hybrid-mesenchymal states.[9] This assumed equivalence, based at least partly on gene set enrichment analysis, largely considers EMT as a binary process.[10] However, it has now been extensively reported that EMT in breast cancer exists more as a spectrum of phenotypes residing along the epithelial-mesenchymal axis.[11–14] Therefore, the association of luminal-basal lineage characteristics and associated plasticity with a partial EMT (pEMT) or hybrid epithelial/mesenchymal (E/M) plasticity remains largely unclear.[15] Similarly, the extent to which associations between partial/full EMT and luminal-basal plasticity in breast cancer are generalizable also remains to be elucidated.

Current therapeutic approaches often target specific molecular subtypes of breast cancer.[16,17] However, therapy-driven adaptive plasticity and consequent phenotypic heterogeneity pose challenges in achieving durable responses.[18] In addition, the clinical implications of understanding coupling between EMT and luminal-basal plasticity, and its impact on the estrogen receptor (ER) signaling, in breast cancer are substantial. Intra-tumor heterogeneity and phenotypic plasticity can impact treatment response and patient survival by altering the dependence of breast cancer cells on a luminal program and ER signaling.[8,15,19,20] Heterogeneity can also drive collective behavior by facilitating division of

[1]Department of Bioengineering, Indian Institute of Science, Bangalore 560012, India
[2]Division of Molecular Medicine, St. John's Research Institute, St. John's Medical College, Bangalore 560012, India
[3]Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[4]University of New South Wales, UNSW Medicine, Sydney, NSW 2010, Australia
[5]Department of Medicine, Duke University, Durham, NC 27708, USA
[6]Present address: Feinberg School of Medicine, Northwestern University, Chicago 60611, USA
[7]Lead contact
*Correspondence: mkjolly@iisc.ac.in
https://doi.org/10.1016/j.isci.2024.110116

labor among different cancer cell subpopulations and consequently varying interactions with the surrounding microenvironment.[21] Thus, higher phenotypic heterogeneity can have both cell-autonomous and non-cell-autonomous contributions to population-level fitness. Gaining insights into the underlying dynamics and mechanisms of breast cancer cell heterogeneity is critical to understanding plasticity-associated therapy relapse.[22,23]

Here, we uncover the associations between epithelial-hybrid-mesenchymal cell state and the luminal-basal axis using multi-modal transcriptomics (bulk, single-cell, and spatial transcriptomics) data from breast cancer cell lines and primary tumor samples. Our analysis demonstrates that luminal cell lines and tumors strongly associate with an epithelial phenotype, but basal cell lines and tumors are not fully mesenchymal. Instead, basal breast cancer associates with pEMT phenotype as well as an enriched phenotypic heterogeneity along the EMT spectrum. These patterns were also recapitulated in methylation profiles, indicating an epigenetic crosslinking between EMT and lineage plasticity along the luminal-basal axis. Finally, we propose an underlying gene regulatory network including players mediating EMT and luminal and basal differentiation axes. The emergent dynamics of this network could recapitulate the trends observed in transcriptomic data, offering novel insights into molecular underpinnings of basal breast cancer. Our integrative systems biology approach reveals hallmarks of basal breast cancer heterogeneity (enrichment of hybrid E/M phenotypes and higher diversity in terms of EMT cell states) and proposes a mechanistic computational model of crosstalk between EMT and lineage plasticity as a framework to test possible therapeutic interventions to restrict intra-tumor heterogeneity.

RESULTS

Luminal gene expression signature is closely associated with an epithelial state whereas basal gene expression is associated with a pEMT state

Clinical and molecular classification of breast cancer has yielded key insights into breast cancer disease biology and has provided a much-needed stratification of the disease for its effective management in a clinical setting.[24] However, the activities of the various biological pathways/processes, including the level of activity of the various gene sets corresponding to the breast cancer subtypes, contributing to the overall phenotypes exhibited by cancer cells can still be quite varied even within a well-stratified breast cancer subtype.[25] While such heterogeneities can significantly impact the disease outcome, the origins and molecular underpinnings of said heterogeneities remain largely unexplored. Hence, we sought to uncover associations relating to two key biological axes of plasticity—the extent to which breast cancer samples express the luminal/basal gene expression programs and the extent to which the samples are epithelial/mesenchymal, whether we examine a particular subtype of breast cancer or consider the entire cohort of subtypes as a unified group.

To uncover the associations between the luminal-basal phenotypes and the EMT status of breast cancer cells, we compared how these five gene sets—luminal, basal, epithelial, mesenchymal, and pEMT—correlated with one another independently. We first calculated the single-sample gene set enrichment analysis (ssGSEA) scores for corresponding gene expression signatures to quantify the activity of these five gene sets. The epithelial and mesenchymal gene signatures (cell-line-specific and tumor-specific) were adapted from a previous pan-cancer analysis.[26] We used a pEMT signature reported earlier[27] to estimate the pEMT nature of the bulk samples. Furthermore, we also considered breast-cancer-specific EMT gene sets to contrast differences between the pan-cancer EMT gene signatures and breast-cancer-specific EMT signatures. We used three different breast-cancer-specific gene sets—EMT_up (mesenchymal), EMT_down (Epithelial), and EMT_partial (pEMT) to assess the activity of these pathways in our transcriptomic data analysis.[28] The luminal and basal nature of breast cancer samples was assessed using a previously curated set of 15 luminal- and basal-specific genes each.[29] There was minimal overlap between the genes included in these signatures as summarized in the upset plots (Figure 1A).

We calculate the activity of each of the abovementioned gene expression signatures using ssGSEA in a set of 80 breast-cancer-specific datasets (Table S3) that span different experimental setups and clinical samples. The objective of this analysis is to uncover associations, if any, that hold between the luminal-basal nature and the EM nature of breast cancer samples across diverse datasets. We observed that in 36 datasets the correlations between luminal and epithelial signatures were significant and positive ($r > 0.3$, $p < 0.05$) while only four datasets showed a significant negative trend ($r < -0.3$, $p < 0.05$) (Figure 1B, i). Such a skew toward the positive side was minimal for the correlations between the basal and the mesenchymal signatures (11 vs. 7, respectively) (Figure 1B, ii). We obtained similar results when we compared EMT_down breast-cancer-specific signature with the luminal signature across the 80 datasets (40 vs. 6, respectively) (Figure 1C, i). A minimal skew in the number of datasets was seen when we compared EMT_up and the basal gene expression signatures (10 vs. 5, respectively) (Figure 1C, ii). These results collectively indicate that the positive association between the luminal and epithelial signatures is more common across breast cancer samples both *in vitro* and *in vivo* as compared to those between the basal and mesenchymal signatures.

To determine if the association between the epithelial and luminal programs can provide additional power to stratify breast cancer patients, we compared the prognostic capacity of an epithelial-mesenchymal gene set with the epithelial-luminal gene set in the TCGA breast cancer cohort. Using the epithelial-mesenchymal signatures, we observed that patients with high mesenchymal (EPI−MES+) tumors showed a worse prognosis (hazard ratio [HR] = 1.5, $p < 0.05$) (Figure S1A, i). The groups of patients with tumors belonging to mixed epithelial and mesenchymal characteristics (EPI+MES+ and EPI−MES−) were not well segregated from the reference group. However, when we performed the analysis with our epithelial-luminal classification, we found that with respect to the reference distribution of patients, i.e., patients with high epithelial and high luminal tumors (EPI+LUM+), all the other groups showed significant differences in survival (Figure S1A, ii). Patients with tumors showing low luminal and/or epithelial signatures (EPI+LUM−, EPI−LUM+, EPI−LUM−) had worse prognosis. Collectively, these results indicate a positive correlation between the luminal and epithelial programs that can be used to potentially stratify patients into prognostic survival groups.

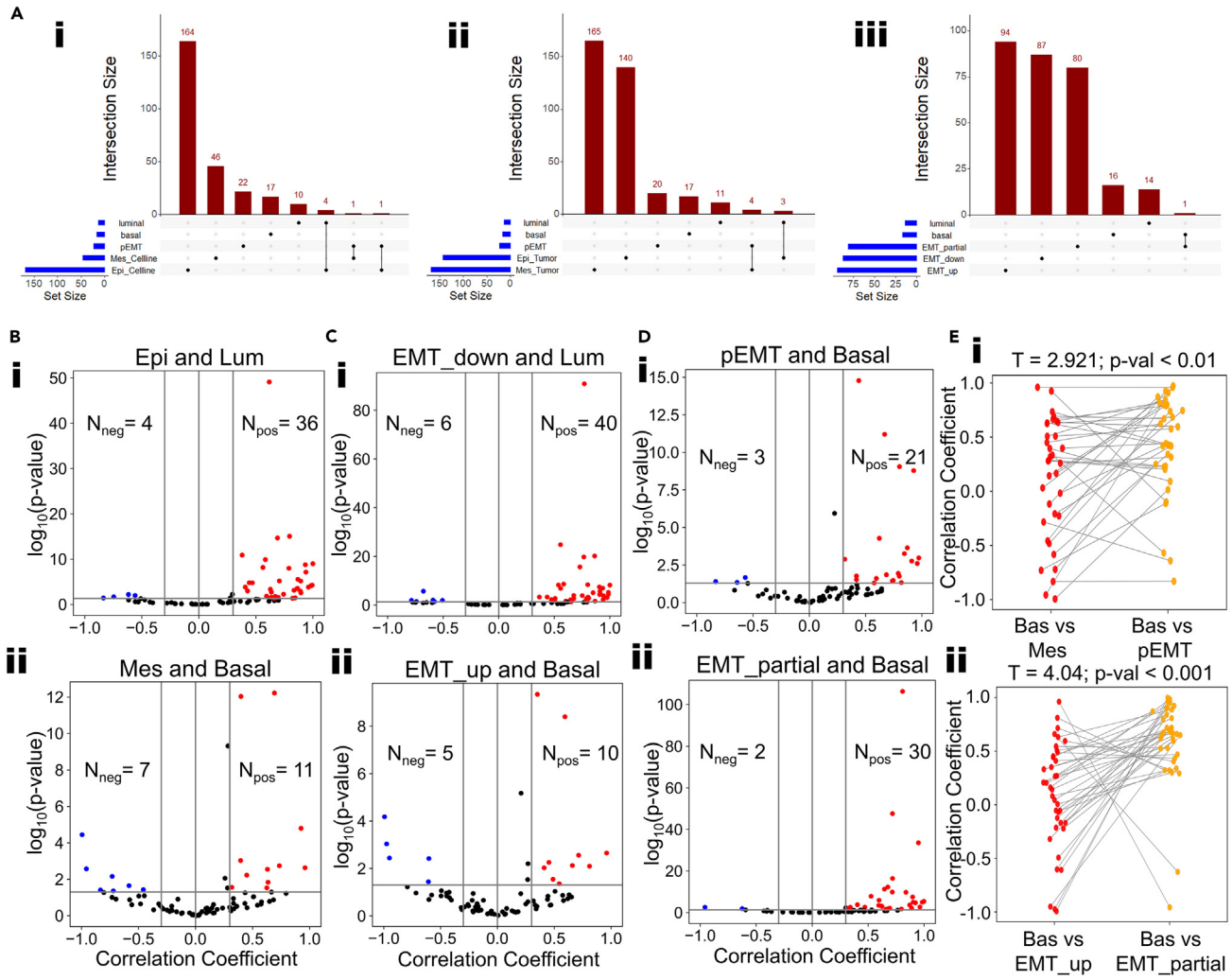


Figure 1. Associations between luminal-epithelial and basal-pEMT programs in breast cancer

(A) Upset plots showing the extent of overlap between the gene sets used. Luminal-basal gene sets from Nair et al. (2022) and pEMT gene set from Puram et al. (2017) are compared against (i) Tan et al. (2013) cell-line-specific epithelial and mesenchymal gene sets, (ii) Tan et al. (2013) tumor-specific epithelial and mesenchymal gene sets, and (iii) Knutsen et al. (2023), EMT_up, EMT_down, and EMT_partial gene sets.

(B) Meta analysis of 80 breast-cancer-specific datasets showing volcano plots of correlation coefficient values for (i) epithelial-luminal and (ii) basal-mesenchymal programs for gene signatures from Tan et al. (2013).

(C) Meta analysis of 80 breast-cancer-specific datasets showing volcano plots of correlation coefficient values for (i) EMT_down-Luminal and (ii) Basal-EMT_up programs for gene signatures from Knutsen et al. (2023).

(D) Meta analysis of 80 breast-cancer-specific datasets showing volcano plots of correlations for (i) pEMT (Puram et al. 2017) and (ii) EMT_partial (Knutsen et al. 2023) and basal programs. Pearson's correlation coefficients and the corresponding *p* values are shown.

(E) Pair plot showing the comparison of the Pearson correlation coefficients between the (i) basal-mesenchymal vs. basal-pEMT and (ii) basal-EMT_partial and basal-EMT_up programs. Paired t test has been performed and the T statistic and *p* value have been reported.

Given the unclear nature of association of basal nature with either the epithelial or mesenchymal markers signatures exclusively, we postulated that the basal signature may be correlated with a pEMT signature. We used the pEMT signature reported previously[27] as well as the breast-cancer-specific EMT_partial gene set to estimate the pEMT nature of the bulk samples by calculating ssGSEA scores. In our meta-analysis of 80 datasets (across many breast cancer subtypes), we observed that the basal signature was positively correlated with the pEMT state in 21 of them ($r > 0.3$, $p < 0.05$) but negatively correlated in only three of them ($r < -0.3$, $p < 0.05$) (Figure 1D, i). This skew toward a positive correlation is greater than the one observed for the basal-mesenchymal pair (Figure 1B, ii). Similar results were obtained when we used a breast-cancer-specific pEMT signature, EMT_partial, to compare with the basal nature of the samples (30 vs. 2, respectively) (Figure 1D, ii). These results indicate that irrespective of the clinical/molecular subtype of breast cancer samples/cell lines, there exists a positive association between the basal nature and the pEMT nature of samples. We also compared the Pearson's correlation coefficients

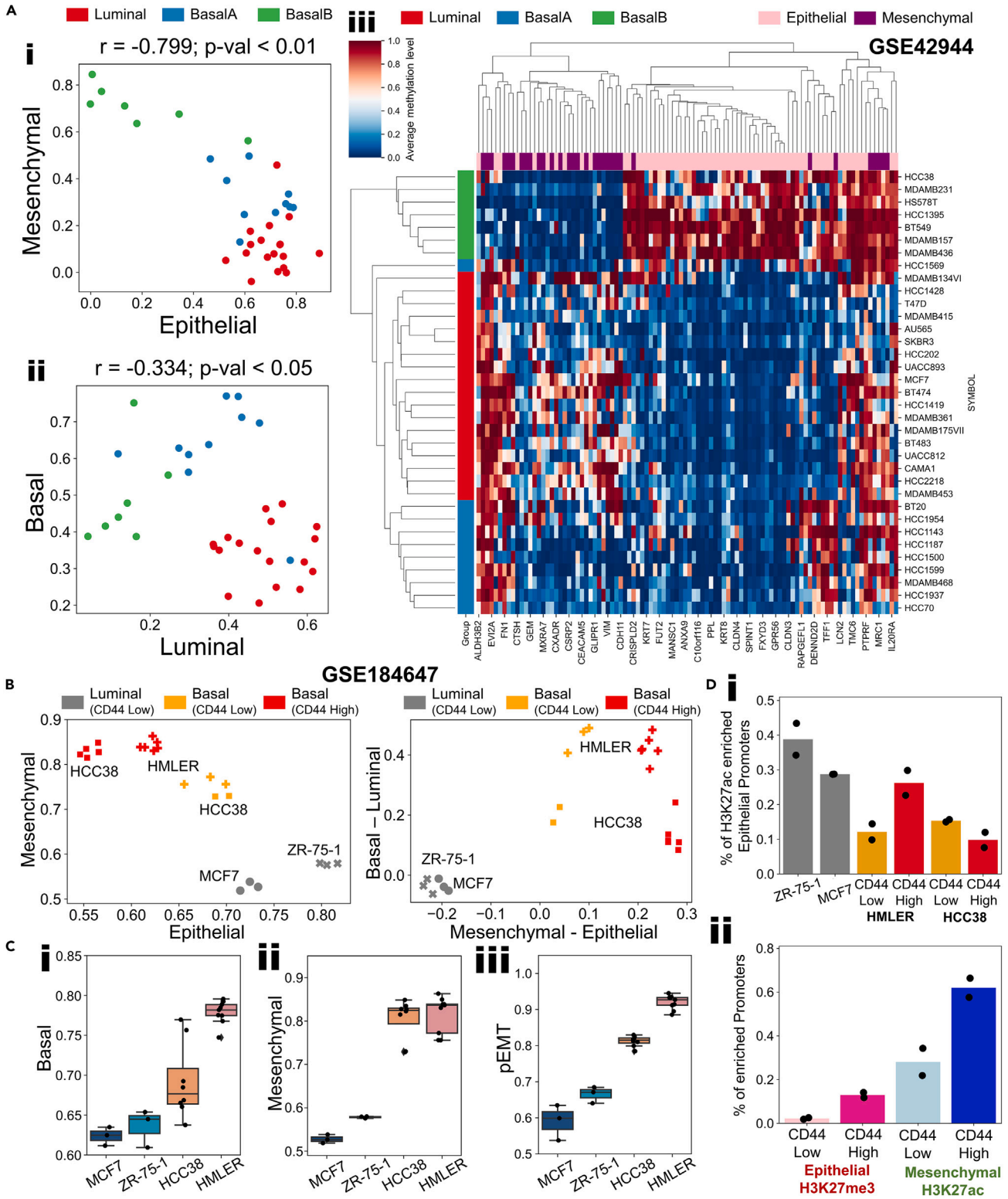


Figure 2. Epigenetic underpinnings of epithelial-mesenchymal phenotypes in luminal and basal breast cancer cell lines

(A) Scatterplots for ssGSEA scores of (i) epithelial-mesenchymal and (ii) luminal-basal programs in breast cancer cell lines. Pearson's correlation coefficient values and corresponding *p* values are mentioned. (iii) Heatmap showing the extent of methylation of CpG islands in the promoters of epithelial (pink) and mesenchymal (purple) genes for different breast cancer cell lines belonging to luminal (red), basal A (blue), and basal B (green) subtypes.

Figure 2. *Continued*

(B) Scatterplots showing the position of luminal, basal (CD44 low), and basal (CD44 high) cell lines on the two-dimensional (left) epithelial-mesenchymal and (right) luminal-basal plane.

(C) Boxplots showing expression of (i) basal, (ii) mesenchymal, and (iii) pEMT gene expression programs across four different cell lines.

(D) (i) Bar plots showing the percentage of H3K27ac activation marks in different breast cancer cell lines. (ii) Bar plots showing the percentage of enriched promoters of epithelial inactivation marks and the mesenchymal activation marks across the CD44 high and low subpopulations in basal breast cancer cell lines.

of basal-pEMT/basal-EMT_partial set with those of basal-mesenchymal/basal-EMT_up set in a paired manner and found a significant increase in basal-pEMT correlation ($T = 2.9$, $p < 0.01$ and $T = 4.0$, $p < 0.001$, respectively) (Figure 1E). The overall trends for association between the luminal-basal pathways and the EM pathways also hold for the specific cases of breast-cancer-specific cell lines (CCLE) as well as TCGA-BRCA patient samples (Figures S1B–S1D and S2A–S2E and Table S2). These results collectively suggest that at a bulk transcriptomic level, breast cancer samples show a stronger association of basal breast cancer with a pEMT signature instead of a mesenchymal one.

These trends are further supported by recent experimental observations that luminal progenitors, the proposed cell of origin of basal-like tumors, undergo a pEMT at onset of tumorigenesis.[30] Moreover, mammary basal epithelial cells have been shown to exhibit a pEMT state, i.e., coexpressing typical epithelial and mesenchymal markers—ZEB1 and OVOL2, respectively.[31] Furthermore, genetically engineered mouse models, patient-derived xenografts, and patient samples of triple-negative breast cancer cells demonstrated large populations of hybrid E/M cells *in vivo* that lead to invasion.[32] Together, these observations underscore the association of basal breast cancer cell state with a pEMT phenotype.

Epigenetic status of epithelial and mesenchymal genes underlies associations between the luminal-epithelial and basal-pEMT programs

Having shown the close association of a luminal program with an epithelial state and that of basal program with a pEMT state at a bulk transcriptomic level, we wished to interrogate whether these associations may have an epigenetic basis. Thus, we analyzed genome-wide methylation data of CCLE breast cancer cell lines (GSE42944) and compared it with the corresponding gene expression data. We observed that the cell lines that were primarily classified into three groups—Luminal, Basal A, and Basal B[33]—were distinctly positioned on the two-dimensional EMT plane (Figure 2A, i). The luminal subtype of cell lines clustered toward the high epithelial, low mesenchymal part of the plane (Figure 2A, i), suggesting a strong association with the epithelial characteristics. Conversely, the Basal B cell lines are positioned diametrically opposite, aligning with the mesenchymal end. The Basal A subtype, however, occupied intermediate regions of the EMT plane, indicating a mixed epithelial-mesenchymal profile. Consistent with our previous observations, the basal score of the cell lines was negatively associated ($r = -0.34$, $p < 0.05$) with their luminal score (Figure 2A, ii) while being positively associated with the pEMT score (Figure S3A).

We next quantified the methylation level for each gene from the epithelial and the mesenchymal gene sets across all cell lines belonging to the three subtypes. We observed that the luminal cell lines were extensively methylated in promoters of mesenchymal genes and had lower methylation levels on epithelial genes (Figure 2A, iii). The converse was true for the Basal B subtype of breast cancer cells. The Basal A subtype of breast cancer cells that was intermediate in terms of their EMT status exhibited lower levels of methylation in both the epithelial and mesenchymal genes (Figure 2A, iii). This indicates that although the Basal B subtype of breast cancer cells had mostly silenced epithelial genes, the Basal A subtype of cells have both epithelial and mesenchymal genes active that may explain the pEMT-like phenotype observed. Furthermore, the basal cell lines (Basal A and Basal B taken together) exhibited higher levels of methylation of the ESR1 gene (encoding for Estrogen receptor [ER]) (Figure S3B, i), a key gene contributing to luminal behavior (Figure S3B, ii) compared to the more luminal cell lines. On the contrary, basal cell lines had lower methylation levels in genes belonging to the pEMT signature (Figure S3B, iii) and the ones belonging to basal signature (Figure S3B, iv).

Previous analysis has demonstrated that an isogenic cell line can have different subpopulations in terms of its EMT nature such as EpCAM-hi and EpCAM-lo subpopulations in PMC-42LA cells or cells with varying CD24 and/or CD44 levels in multiple breast cancer cell lines.[34–36] Thus, we focused on heterogeneity within a cell line in terms of their epithelial-mesenchymal nature and their luminal-basal characteristics. We analyzed RNA sequencing (RNA-seq) data from four representative cell lines belonging to luminal and basal subtypes of breast cancer (two luminal—MCF7 and ZR-75; two basal—HCC38 and HMLER) (GSE184647). The luminal cell lines were CD44-low; thus, no CD44-high subgroup was observed in them, but basal cell lines harbored distinct CD44-low and CD44-high subpopulations.[37] We observed that luminal cell lines were clustered on the high epithelial, low mesenchymal section of the two-dimensional EMT spectrum, whereas the CD44-high subpopulations of basal cell lines were clustered toward the low epithelial, high mesenchymal end of it (Figure 2B, left). Intriguingly, the CD44 low basal subpopulations were clustered in the medium epithelial, medium mesenchymal region, indicative of a pEMT state (Figure 2B, left). We also observed that on the combined EMT and luminal-basal plane, the luminal cell lines were clustered together in the high epithelial, high luminal region, whereas the basal cell lines showed a larger spread, with the CD44-high subpopulations being more mesenchymal, but not necessarily more basal than their CD44-low counterparts as assessed by ssGSEA scores of the corresponding gene signatures (Figure 2B, right). We observed that HMLER cells were more basal compared to HCC38, i.e., basal nature was better explained by the cell line rather than CD44 (Figure 2B, right).

We also noted that luminal cell lines had lower ssGSEA scores for basal, mesenchymal, and pEMT gene signatures as compared to basal cell lines (Figure 2C). Intriguingly, both the basal cell lines had distinct ssGSEA scores of basal gene set activity but comparable ssGSEA scores for mesenchymal signature. The pEMT signature was better able to capture the trends of the basal signature compared to that of the

mesenchymal signature (Figure 2C). This trend strengthens our observations that the pEMT signature is a better predictor of basal nature compared to that of a mesenchymal signature. Finally, we probed the MINT-CHIP data for these samples and found that the promoters of epithelial genes were specifically higher in H3K27ac (activation) marks for luminal cell lines compared to basal cell lines (Figure 2D, i). Furthermore, in HMLER, the CD44-low sub-population had consistently lower levels of activation marks (H3K27Ac) on mesenchymal genes as well as lower levels of suppressive marks (H3K27me3) on the epithelial genes (Figure 2D, ii). This pattern offers a potential explanation for the association of basal subtype with a pEMT state. Recent data from mammary stem cell subpopulations revealed higher accessibility and enrichment of P63 DNA-binding motifs in basal cells and that of ELF5 DNA-binding motif in luminal progenitors.[38] Given the established role of NP63 in driving a pEMT program[13,39,40] and that of ELF5 in inhibiting EMT,[41–43] these observations together support the possibility of an epigenetic control in the association of basal breast cancer with hybrid E/M phenotype(s).

Spatial transcriptomics reveals intra-patient variability in EMT phenotypes in basal subtypes of breast cancer

After demonstrating a higher heterogeneity of basal breast cancer cell lines along the EMT spectrum *in vitro*, we sought to investigate the same in breast cancer patients. Thus, we analyzed publicly available spatial transcriptomics datasets of breast tissue sections[44] to infer the patterns of spatial heterogeneity in associated luminal-epithelial and basal-mesenchymal/pEMT status, using therapeutically relevant markers of gene expression. Firstly, we observed that among the $n = 6$ patients for whom spatial transcriptomics data were available, the ER+ breast cancer patients had distinctively higher levels of luminal nature but downregulated basal scores (Figure 3A). On the contrary, the basal breast cancer patients, while showing reduced levels of luminal nature, were quite heterogeneous for basal signature expression (Figure 3A). The overall correlation for the epithelial-mesenchymal score pair was strongly negative ($r = -0.85$, $p < 0.01$)—while the ER+ breast cancer patients clustered toward the high epithelial, low mesenchymal portion of the plane, TNBC patients were highly variable and scattered across the spectrum (Figure S4A).

We observed spatial heterogeneity in epithelial and luminal scores in ER+ tumors (Figure 3B, i and ii) but the signatures relating to basal, mesenchymal, and pEMT phenotypes were largely absent (Figures S4B and S4C). More specifically, we observed that even though different areas within the tumor displayed a similar extent of luminal signature, they had varied epithelial scores (arrows shown in Figure 4B, i and ii). This observation is particularly important in the context of our survival analysis highlighting that EPI+LUM− or EPI−LUM+ phenotype shows significantly worse survival compared to EPI+LUM+ (Figure S1A, ii). We speculated that patients having a higher proportion of EPI+LUM− or EPI−LUM+ phenotypes would likely have a poorer prognosis in comparison to patients who show homogeneous levels of EPI+LUM+ phenotypes. We also investigated the spatial activity patterns of ER-driven genes that are generally found to be concordant with luminal and epithelial cell states.[45,46] We observed that similar to luminal and epithelial scoring patterns, spatial transcriptomic slides had heterogeneous activity levels of ER response genes and E2F targets (a proxy for cell-cycle activity) (Figures 3B, iii and iv, and S4C). Such patterns of heterogeneity can dramatically impact sensitivity to anti-ER drugs[47] as these pathways are the primary target of currently used targeted therapies. Thus, the degree of underlying heterogeneity of these molecular programs may limit the efficacy and evolution of resistance in ER+ breast cancer.

Next, we analyzed spatial transcriptomics data from breast cancer patients classified to have basal disease. In one such patient sample, we noticed a more heterogeneous expression of luminal and basal scores across the tissue samples as compared to a previously analyzed case of luminal disease. We observed that the areas of normal breast tissue were high for luminal, epithelial, and ER response gene set scores (Figures 3C, i–iii, top left tissue section). However, the tissue slices of the tumor were significantly more enriched in basal scores (Figure 3C, iv), with marked variability in terms of their epithelial and mesenchymal natures (Figures 3C, ii and v, and S4D). This trend supports our results showing higher diversity in tumor epithelial-hybrid-mesenchymal states among basal tumors as compared to a luminal breast cancer case. Further, the concordance between the basal and the pEMT signature (Figure 3C, compare iv with vi) was higher compared to that between the basal and mesenchymal signature (Figure 3C, compare iv with v), with the pEMT signature being a proper subset of the spatial sections enriched for the basal signature of the tissue sections. Finally, we quantified the overall activity of the luminal-basal nature, epithelial-mesenchymal nature, and the pEMT nature of the six patients for which spatial transcriptomics data were available. We observed qualitatively that while TNBC patients had a higher net basal score, the EM score and the pEMT scores were variable across the six patients. On the other hand, the samples from ER+ patients were relatively more luminal as well as more epithelial in the two patients for which data were available (Figure S4E). Collectively, these results may explain the earlier observed bulk expression patterns where pEMT signatures correlated more closely with the basal signature compared to a mesenchymal signature.

Basal breast cancer cell lines and tumor samples have higher phenotypic heterogeneity in terms of EMT

Next, we focused on pinpointing molecular underpinnings that may explain the association of the basal signature with the pEMT program. This association can be explained primarily by two scenarios: (1) the basal subtype is primarily composed of cells that are pEMT or (2) the basal subtype is composed of separate populations of cells that are epithelial and mesenchymal, resulting in a higher pEMT signature.

To gain a more comprehensive understanding of subtype-specific heterogeneity with respect to EMT status, we analyzed single-cell RNA-seq of 32 breast cancer cell lines spanning all the clinical subtypes (GSE173634).[48] A previous study[32] showed that CDH1+VIM+ TNBC cells were enriched during invasion and had higher colony-forming ability. Thus, we classified this single-cell data based on gene expression values of CDH1 and VIM. The distribution of the difference between the imputed expression values of VIM and CDH1 showed multiple peaks (Figure 4A, i). The extreme peaks were labeled as epithelial and mesenchymal, whereas the intermediate ones were labeled as hybrid phenotypes. We observed that all cells belonging to luminal A and luminal B cell lines were epithelial in nature. However, the basal cell lines

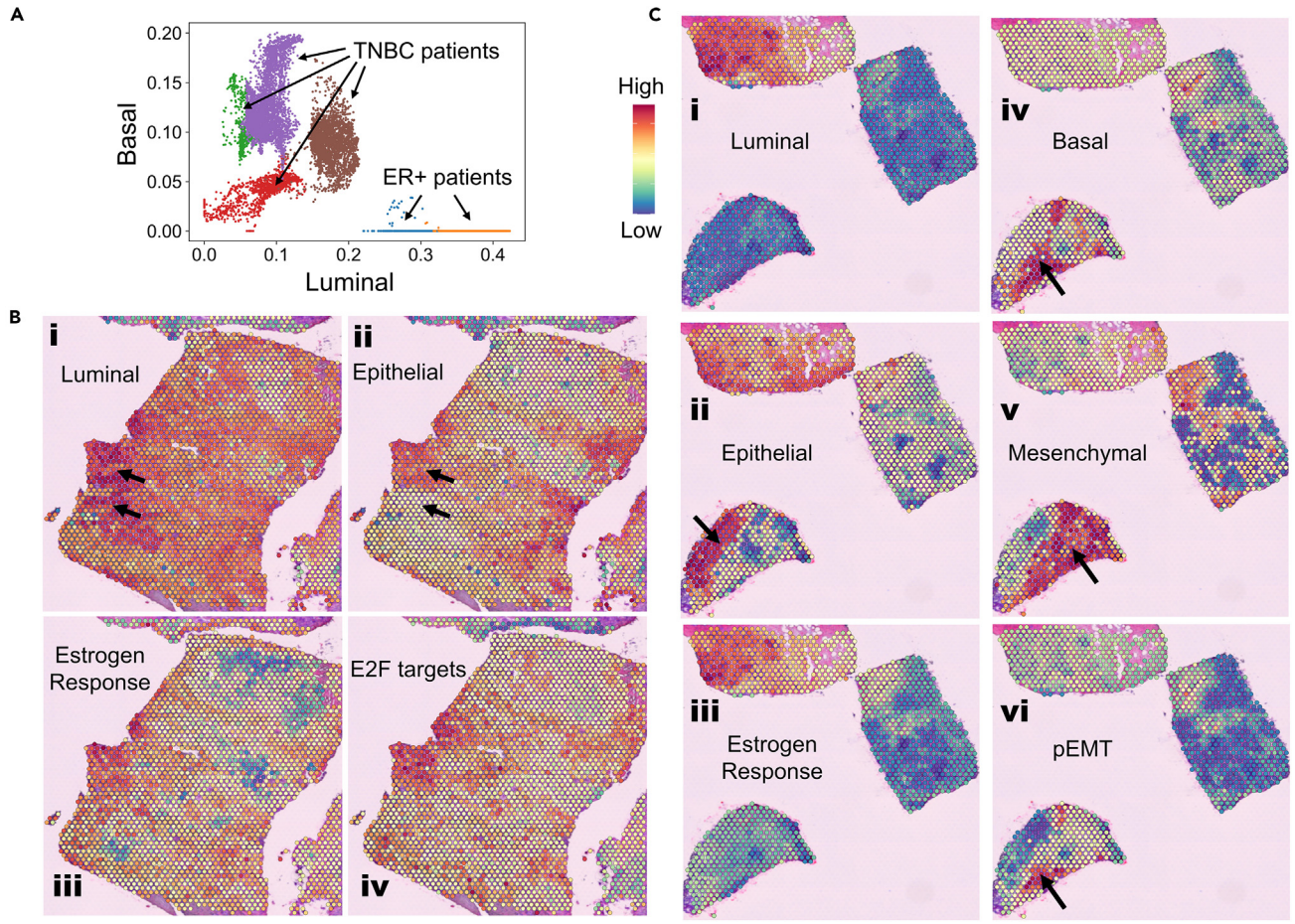


Figure 3. Spatial transcriptomic analysis of estrogen-receptor-positive and triple-negative breast cancer patients

(A) Scatterplot showing the position of the estrogen-receptor-positive (ER+) and triple-negative breast cancer (TNBC) patients on a two-dimensional luminal-basal plane. Each point in the plane is based on the gene expression values for a specific spot on the spatial transcriptomic datasets.

(B) Spatial transcriptomic slides from an ER+ breast cancer patient colored by the activity scores of (i) luminal, (ii) epithelial, (iii) hallmark estrogen response, and (iv) hallmark E2F target genes. Red represents a higher activity score, and blue represents a lower activity score. The arrows point to areas of specific interest based on differences in heterogeneity of the different biological pathways.

(C) Spatial transcriptomic slides from a TNBC breast cancer patient colored by the activity scores of (i) luminal, (ii) epithelial, (iii) hallmark estrogen response, (iv) basal, (v) mesenchymal, and (vi) pEMT gene sets. Red represents a higher activity score, and blue represents a lower activity score. The arrows point to areas of specific interest based on differences in heterogeneity of the different biological pathways.

had varied heterogeneity patterns along EMT axis. The TNBC type A cells were found to be either epithelial or hybrid E/M in nature, whereas the TNBC type B ones belonged to epithelial, hybrid E/M, and mesenchymal states (Figure 4A, ii and iii). Interestingly, the HER2 subtype of breast cancer was largely epithelial in nature but the basal-like cell line MCF12A was predominantly in a hybrid E/M state (Figure 4A, ii and iii). Similar results were obtained when gene-set-based scoring of epithelial and mesenchymal pathways were employed to characterize the EM status of the cells instead of CDH1 and VIM (Figure S5). These results indicate that although the luminal subtype is constituted primarily from an epithelial phenotype, the basal subtypes are more heterogeneous in terms of their E/M status and more likely to harbor a more pEMT cell state.

We analyzed the top 20 transcription factors (TFs) that correlated with each of the four signatures: luminal, basal, epithelial, and mesenchymal (in GSE173634). Among the 20 TFs correlating with luminal and epithelial, 13 TFs were common, but among the 20 TFs correlating with mesenchymal and basal, only two TFs were common. These analyses suggest that the transcriptional programs that regulate luminal and epithelial states overlap while basal and mesenchymal states are regulated by distinct TFs. A hierarchically clustered pairwise correlation map of the top TFs revealed antagonism between the luminal-epithelial and basal-mesenchymal group of TFs (Figure 4B). Furthermore, the basal and the mesenchymal transcription factors clustered separately with each other, indicative of the weak coupling between these two biological axes in contrast to the stronger associations of the luminal and epithelial TFs (Figure 4B).

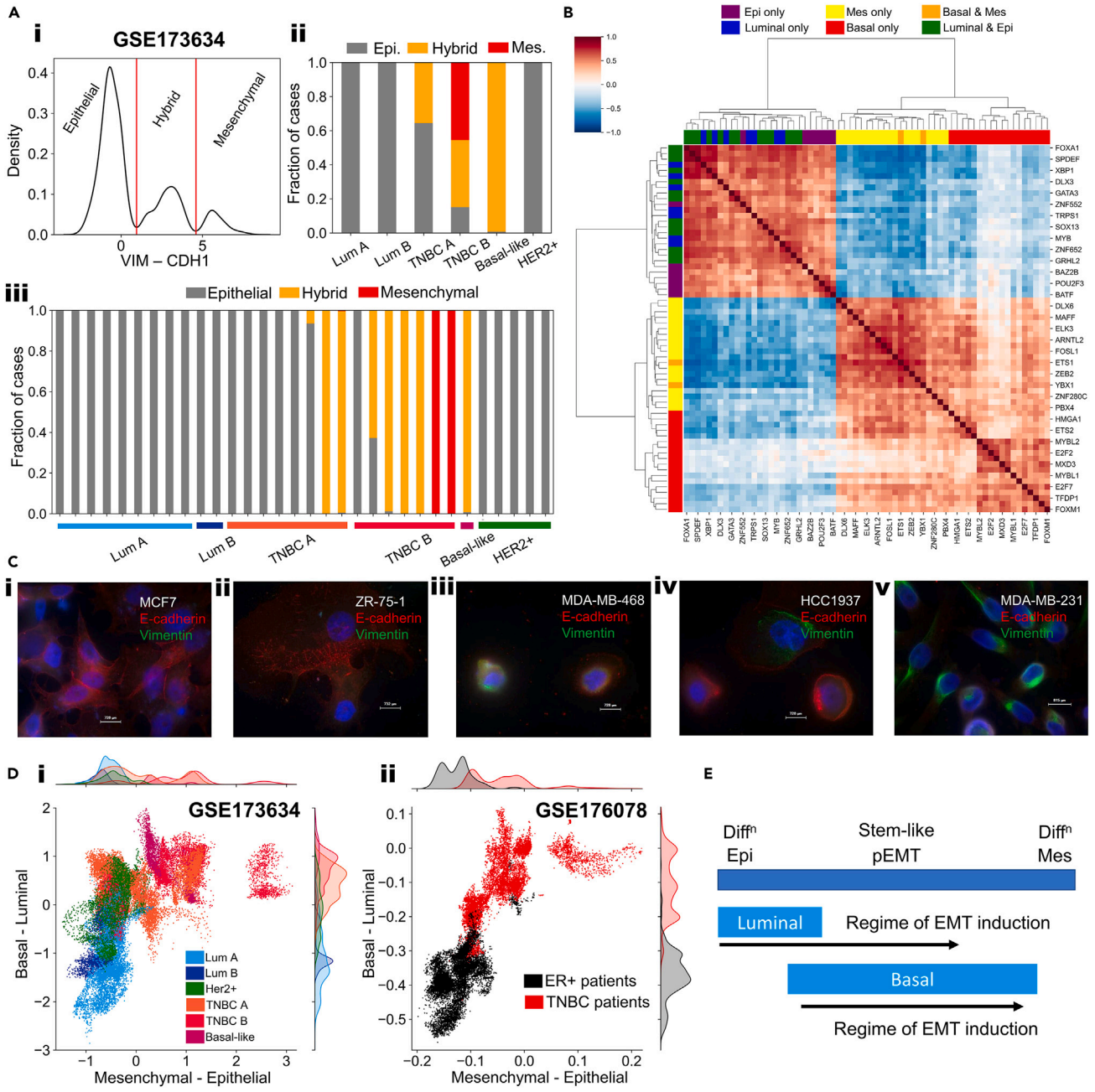


Figure 4. Heterogeneity patterns in single-cell analysis of breast cancer cell lines and patient-derived cells

(A) (i) Kernel density estimate of VIM-CDH1 levels for cells belonging to all breast cancer cell lines (GSE173634). Red lines split the distribution into epithelial, mesenchymal, and intermediate/hybrid states, based on difference in VIM and CDH1 levels. (ii) Composition of different subtypes of breast cancer cell lines in terms of classified epithelial, mesenchymal, and hybrid states. (iii) Composition of different cell lines belonging to subtypes of breast cancer cell lines in terms of classified epithelial, mesenchymal, and hybrid states.

(B) Gene-gene pairwise correlation heatmap showing top 20 transcription factors correlated with the luminal, basal, epithelial, and mesenchymal signatures each in GSE173634.

(C) Immunofluorescence imaging of representative breast cancer cell lines showing E-cadherin and vimentin levels.

(D) (i) Scatterplot showing the distribution of single-cell RNA-seq data of breast cancer cell lines on the two-dimensional epithelial-mesenchymal status and luminal-basal status (GSE173634). x axis represents the epithelial-mesenchymal score (mesenchymal score–epithelial score); y axis represents the luminal-basal score (basal score–luminal score). (ii) Scatterplot showing the distribution of breast cancer cells from ER+ and TNBC patient samples on the two-dimensional epithelial-mesenchymal status and luminal-basal status (GSE176078). x axis represents the epithelial-mesenchymal score (mesenchymal score–epithelial score); y axis represents the luminal-basal score (basal score–luminal score).

(E) Schematic showing the extent of heterogeneity and mapping of epithelial to mesenchymal transition status of luminal and basal breast cancer cases.

Next, we probed whether these trends observed in RNA-seq data analysis could be recapitulated experimentally as well. Thus, we took five representative cell lines—two belonging to luminal subtype (MCF7 and ZR-75-1) and three belonging to basal subtype (HCC1937, MDA-MB-468, and MDA-MB-231). Each of these cell lines were probed for E-cadherin and vimentin levels though immune-fluorescence experiments. It was observed that the luminal cell lines of MCF7 and ZR-75-1 were exclusively high for E-cadherin, whereas they have very little to no expression of vimentin (Figure 4C, i and ii), indicating that the luminal cell lines were largely epithelial in nature. On the other hand, basal cell lines such as MDA-MB-468 and HCC1937 exhibited a mix of epithelial (high specifically for E-cadherin) and hybrid E/M phenotypes (coexpression of E-cadherin and vimentin in the same cells) (Figure 4C, iii and iv). Finally, MDA-MB-231 showed largely mesenchymal cells (low E-cadherin, high vimentin) with a few cells coexpressing E-cadherin and vimentin (Figure 4C, v). The experimental results thus support our observations from bulk, single-cell, and spatial transcriptomic data.

We next projected single-cell RNA-seq from both breast cancer cell lines and primary tumors on a two-dimensional plane, where the x axis indicates EMT status (defined as difference in ssGSEA scores of mesenchymal and epithelial gene sets) and the y axis denotes a difference between the ssGSEA scores of luminal and basal gene sets. For the 32 breast cancer cell lines (GSE173634), the trend between luminal-basal transition and EMT was non-linear; while a majority of the samples with high luminal scores clustered close to the epithelial end of the EMT axis, the basal-high samples spanned the entire EMT (Figure 4D, i). Similar patterns were observed from projection of single-cell RNA-seq data from tumor cells isolated from ER+ and TNBC patients (GSE176078)[44] into this two-dimensional space (Figure 4D, ii). These results collectively indicate that luminal cell lines (and ER + tumors) were more restricted or homogeneous in terms of their EMT state and exhibited predominantly an epithelial state, while the basal breast cancer cell lines (and TNBC tumors) were not only more likely to exhibit a pEMT phenotype but also more heterogeneous in terms of their EMT status (Figure 5E).

The enrichment of hybrid E/M phenotypes and/or higher heterogeneity along the EMT axis has been associated with worse survival in many cancer types.[27,49] Currently, no specific therapy targets either a hybrid E/M state or higher phenotypic heterogeneity, thus the enrichment of these attributes may explain the underlying mechanistic basis for difficulty in targeting basal-like tumors and TNBC.

Mathematical modeling of gene regulatory networks captures phenotypic heterogeneity in breast cancer and pinpoints determinants of luminal-basal plasticity

Having uncovered the complex associations between the lineage characteristics of breast cancer cells and the status of cells along the epithelial-mesenchymal spectrum, we sought to understand the mechanistic underpinnings of these interconnected axes of plasticity. Specifically, we asked how breast cancer cells undergoing EMT may drive lineage plasticity and vice versa. Thus, we first assembled a gene regulatory network (GRN) based on experimental evidence (Table S4) to investigate the associations between the luminal-basal and epithelial-mesenchymal axes of cellular plasticity. This GRN is not inferred via statistical tools. It is expected to capture key factors involved in EMT and luminal-basal plasticity and to be capable of recapitulating underlying phenotypic heterogeneity and observed associations between different biological axes.

This GRN consists of representative TFs and genes associated with luminal (ERα66, PGR, GATA3, and FOXA1),[50–53] basal (SLUG and ΔNP63),[54,55] epithelial (CDH1 and miR-200), and mesenchymal (ZEB1, SLUG) phenotypes[56] and regulatory interactions among them. We also incorporated two additional players: (1) ERα36 as a marker for anti-ER therapy resistance[57] and (2) NRF2 as a reported stabilizer of the hybrid E/M state and also associated with drug resistance through its impact on cellular metabolism[58–60] (Figure 5A). The creation of this gene regulatory network establishes direct and indirect feedback regulations between the luminal-basal genes as well as the epithelial-mesenchymal genes, which can give rise to emergent cellular phenotypes, thus coupling both these axes of plasticity. To understand the emergent dynamics of this GRN, we used RACIPE,[61] a computational framework to identify the possible phenotypic space for a given GRN. RACIPE uses a set of coupled ordinary differential equations to simulate the dynamics of interconnected nodes (via regulatory links/edges) in a GRN and outputs the different possible steady state values of all nodes in the network. The relative expression levels of different genes/nodes constitute different cell states that can correspond to observed phenotypes in breast cancer cell-state heterogeneity.

We defined the following scores to better understand the association between EMT and luminal-basal plasticity: (1) luminal score as the sum of normalized steady state values of ERα66, GATA3, PGR, and FOXA1; (2) basal score as the sum of normalized steady state values of ΔNP63 and SLUG; (3) epithelial score as the sum of normalized steady state values of CDH1 and miR-200; (4) mesenchymal score as the sum of normalized steady state values of ZEB1 and SLUG; and (5) resistance score as the difference between normalized steady state levels of ERα36 and ERα66. We observed that the epithelial-mesenchymal score (= mesenchymal score–epithelial score) was multimodal in nature with two hybrid states (referred to as epithelial-hybrid and mesenchymal-hybrid) in addition to the canonical epithelial and mesenchymal states. The luminal-basal score (= basal score–luminal score) distribution was largely trimodal (Figure S6A). Thus, our GRN dynamics can recapitulate the multiple hybrid E/M phenotypes that have been well reported experimentally,[3,49] as well as a luminal-basal phenotype identified recently in breast cancer.[15,62]

Upon projecting these simulated node values on a two-dimensional plane of luminal-basal and epithelial-mesenchymal axes, we could recapitulate the non-linear relationship between the epithelial-mesenchymal axis and the luminal-basal axis (Figure 5B, i) as observed for breast cancer cell lines (Figure 4D, i) and patient-derived tumor cell data (Figure 4D, ii) at a qualitative level. More specifically, we observed that while the solutions corresponding to a luminal state are almost exclusively epithelial in nature, those that correspond to a basal-like state could be epithelial, mesenchymal, or hybrid E/M (pEMT) on the epithelial-mesenchymal spectrum (Figure 5B, i). This observation suggests that the GRN considered here can explain and reproduce the major associations noted in extensive multi-modal transcriptomic data (bulk, single-cell, spatial) analysis across breast cancer cell lines and tumors.

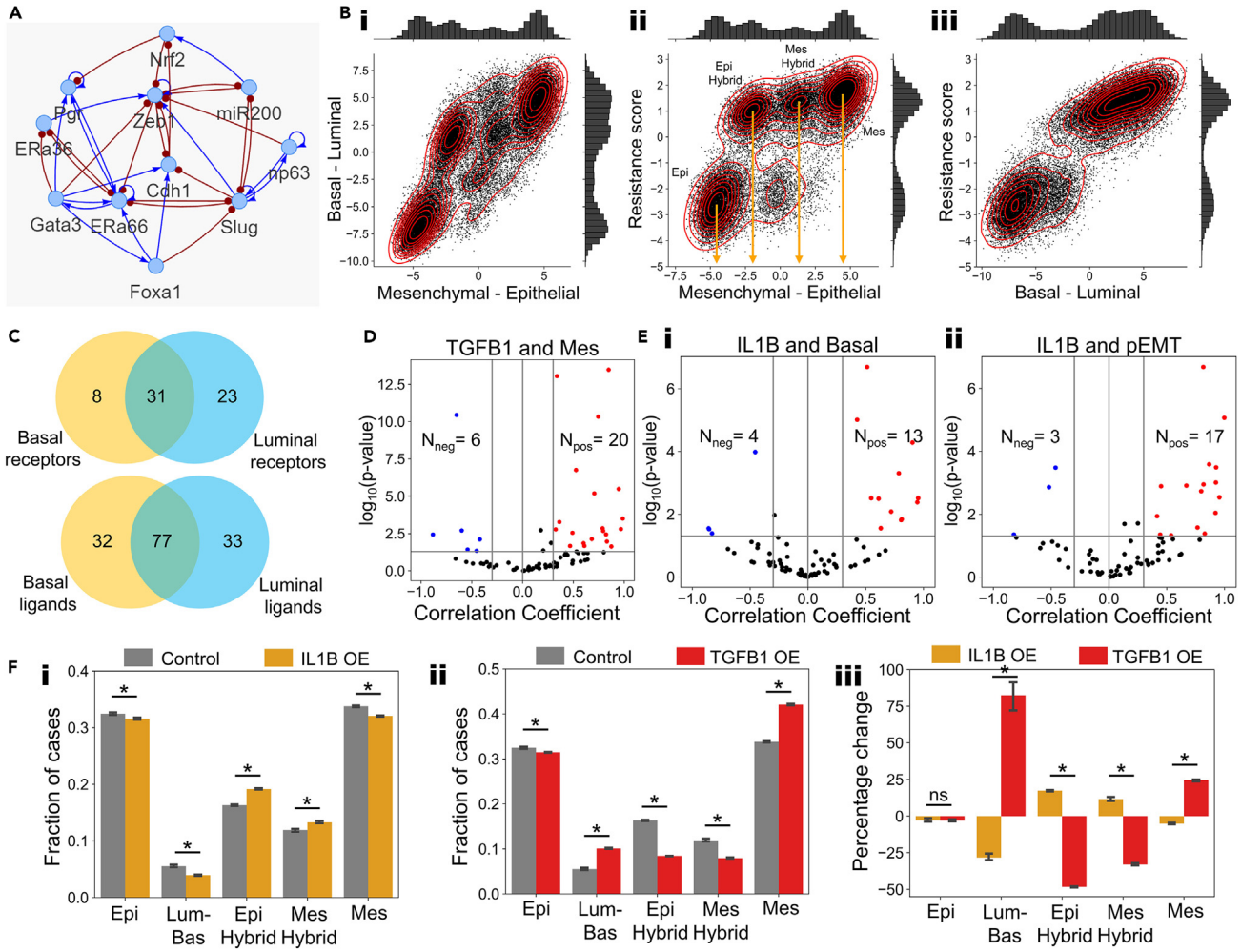


Figure 5. Gene regulatory network analysis for coupled luminal-basal and epithelial-mesenchymal plasticity

(A) Gene regulatory network showing the representative genes from luminal, basal, epithelial, and mesenchymal programs and the associated regulatory links between them. Blue links represent activation, whereas red links represent inhibition.

(B) Simulated steady state solutions projected on a two-dimensional (i) epithelial-mesenchymal axis and luminal-basal axis, (ii) epithelial-mesenchymal axis and resistance score, and (iii) luminal-basal axis and the resistance score.

(C) Number of common and unique receptors that were associated with luminal and basal cell type and respective tumor microenvironments.

(D) Meta-analysis of 80 breast-cancer-specific datasets showing volcano plots of correlations for the TGFB1 gene and the mesenchymal program. Each dot denotes a unique GSE ID (dataset). Pearson's correlation coefficients and the corresponding *p* values are mentioned.

(E) Meta-analysis of 80 breast-cancer-specific datasets showing volcano plots of correlations for IL1B gene expression with (i) basal signature and (ii) the pEMT program. Pearson's correlation coefficients and corresponding *p* values are indicated.

(F) Simulation results showing the fraction of steady state solutions belonging to the different cell states in control and (i) IL1B overexpression (OE) and (ii) TGFB1 OE scenarios. (iii) Simulation results showing percentage change in each observed phenotype compared to the control scenario, in TGFB1 and IL1B OE scenarios. * represents a statistically significant difference as assessed by a two-tailed Student's t test (*p*-value <0.05). Data are represented as mean +/− standard deviation.

Further, our simulation results highlight that the most epithelial cluster is predominantly low in resistance score, i.e., being sensitive to anti-ER drugs such as tamoxifen. However, the other three clusters (epithelial-hybrid, mesenchymal-hybrid, and mesenchymal) had significantly higher resistance scores (Figure 5B, ii). These model predictions are reminiscent of prior experimental observation that partial and/or full EMT can drive resistance to tamoxifen and vice versa.[8,63] We also observed that the luminal-basal status strongly correlated with the resistance score, suggesting that lineage determination is a crucial factor for sensitivity to anti-ER drugs (Figure 5B, iii). We found that the two hybrid E/M clusters had higher levels of NRF2 (Figure S6B), consistent with the reported literature.[58,59,64] These results indicate that the GRN considered here can sufficiently capture the observed associative trends in breast cancer and further correlate luminal-basal and epithelial-mesenchymal trends to their corresponding sensitivity to anti-estrogen drugs such as tamoxifen.

Having simulated a GRN that couples the biological axes of lineage characteristics (luminal/basal), epithelial/mesenchymal status, and the degree of sensitivity to the widely used anti-ER drugs, we wished to identify potential ligands and receptors that could drive a luminal to basal

phenotypic switch, thus limiting the efficacy of the anti-ER drugs especially in the context of ER+ breast cancer. To do this, we analyzed a single-cell patient atlas of seven ER+ breast cancer patients and five TNBC patients (GSE176078). Each of these samples had annotated tumor cells as well as corresponding stromal and immune cell types. Thus, we used LIANA pipeline[65] to score for top ligands and receptors for each patient separately. Specifically, we identified the ligands and receptors that were specific to the target cell type pre-annotated to be luminal A or basal subtype of cancer cells in ER+ and TNBC patients, respectively. We focused on ligands and receptors that were consistently expressed in all patient samples with at least 100 target breast cancer cells. We created separate lists for ligands and receptors for luminal A and for basal breast cancer cells; for instance, the top receptors unique to luminal A tumor cells contained ESR1, MUC1, and ERBB3, each of which has been associated with a luminal subtype.[66,67] We observed substantial overlap between the ligand and receptors from the basal and luminal cell types (Figure 5C), suggesting potential overlaps in signaling between luminal and basal cell types with the tumor microenvironment.

Next, we focused on the list of common ligands, as they can potentially act on luminal cells and cause a transition to a basal-like phenotype. TGF-β1 was one such common ligand, which is a well-known driver of EMT in many carcinomas. To assess how likely a ligand is to affect the luminal-basal phenotype of the cells, we performed meta-analysis of each of the common ligands in 80 bulk transcriptomics datasets from breast cancer (Table S3). We found that TGFβ1 was among the top genes that correlated positively with a more mesenchymal phenotype (Figure 5D). On the other hand, we found IL-1β to be one of the top genes to be correlated with a basal phenotype (Figure 5E, i) as well as pEMT phenotype (Figure 5E, ii). Further, we interrogated whether our GRN could reproduce the plasticity patterns that can be driven by transforming growth factor β1 (TGF-β1) or interleukin-1β (IL-1β). To do this, we extended our GRN (Table S4) to include IL-1β as well as TGF-β and performed simulations to overexpress these nodes in silico. Our model predicted that IL-1β overexpression led to a significant increase in the population of the epithelial-hybrid and the mesenchymal-hybrid populations, with a concomitant decrease in the frequency of epithelial and mesenchymal phenotypes (Figure 5F, i and iii). On the other hand, TGF-β1 overexpression caused a marked increase in the frequency of mesenchymal phenotype and a simultaneous reduction in epithelial, epithelial-hybrid, and mesenchymal-hybrid states (Figure 5F, ii and iii). There was also a significant increase in the proportion of the luminal-basal hybrid phenotype (Figure 5F, ii and iii). Although both IL-1β and TGF-β1 cause a net decrease in the epithelial phenotype and increase the mesenchymal nature, their impact is quite distinct: IL-1β can enrich the hybrid E/M phenotypes that are basal in nature. This prediction is consistent with recent experimental observations about the impact of IL-1β *in vitro* and *in vivo*: (1) IL-1β treatment can induce two stabilizers of hybrid E/M phenotype NRF2 and SLUG,[68,69] (2) IL-1β treatment of MCF7 luminal breast cancer cells can induce ΔNP63 and mediate subsequent therapy resistance,[70] and (3) IL-1β treatment can prevent differentiation of metastatic-initiating cells to highly proliferative epithelial cells, inhibiting overt metastatic growth.[71] Together, our analysis suggests IL-1β to be a potent target to prevent luminal-to-basal lineage plasticity.

Overall, our simulation results show that the core GRN modeled here can capture complex associations between the luminal-basal and epithelial-mesenchymal axes and explain underlying mechanisms of the phenotypic heterogeneity in basal breast cancer. This GRN can also serve as a robust framework to simulate the effect of other signaling molecules to enrich for different phenotypes as a consequence of emergent properties of cross-linked feedback loops among different factors.

DISCUSSION

Phenotypic heterogeneity is a fundamental feature of biological systems implicated in better chances of survival of a population under various dynamically varying environmental stress levels.[72] Not surprisingly, cancer cells are heterogeneous along various functional and molecular axes. This heterogeneity often helps them to evade therapeutic attacks and adapt to their changing environments, eventually driving their metastatic dissemination and colonization.[3,14,73–75] Single-cell approaches have been instrumental in characterizing such heterogeneity, but usually along one axis.[35,76,77] Analyzing interconnections among different axes of heterogeneity is relatively poorly understood.

Here, we used a multi-modal (single-cell, bulk, and spatial) transcriptomic data analysis approach to identify the associations between two key axes of heterogeneity in breast cancer—EMT and luminal-basal (lineage) plasticity.[15,78–80] Often, these axes have been assumed to be synonymous or largely overlapping, but many questions remained unanswered, such as the following: (1) how do luminal and basal features map to epithelial, hybrid E/M, and mesenchymal phenotypes? (2) which breast cancer subtype (luminal/basal) have higher heterogeneity in terms of EMT? and (3) what feedback loops connect these two plasticity axes?

Through our analysis of bulk transcriptomics datasets, we uncover that the luminal-epithelial association and the basal-pEMT association are strongly positively correlated across breast cancer samples/model systems. Further, we show that the strong association of basal with pEMT could be due to both increased prevalence of the pEMT cell state in the basal subtype of the disease as well as the coexistence of both epithelial and mesenchymal cell states in the basal subtype of the disease. Finally, we show using mathematical models of representative gene regulatory networks how these associations could be realized in scenarios in breast cancer, thus establishing a conceptual framework to better characterize EM heterogeneity along the luminal-basal spectrum of the disease.

Our analyses of breast cancer cell lines and primary tumors have unraveled that although luminal breast cancer samples are predominantly epithelial and quite homogeneous, basal breast cancer samples correspond to hybrid E/M state(s) with more phenotypic heterogeneity along the EMT spectrum. Both of these hallmark features of basal breast cancer have independently been associated with worse patient survival in many cancer types.[3,49,81] Thus, their coexistence can possibly explain the aggressive behavior and limited therapeutic response of basal-like breast cancers.[82] Further, given the relatively higher plasticity of hybrid E/M phenotypes as compared to "fully epithelial" or "fully

mesenchymal" phenotypes,[83,84] the presence of hybrid E/M phenotypes can facilitate more phenotypic heterogeneity in a population. Our results are consistent with observations of higher PD-L1 protein levels[85] and tamoxifen resistance[62] in basal-like tumors, given that both these traits have been previously linked to hybrid E/M cells.[8,86] Even among luminal tumors, those expressing low levels of ER can display higher basal-like phenotypes,[15] endorsing previous results that ESR1 expression closely associates with luminal breast cancer cells and that silencing ER can drive EMT.[87] Because EMT and tamoxifen resistance can both drive each other,[8,88] our results suggest that tamoxifen resistance can govern lineage plasticity, i.e., luminal-to-basal switch, as well. Similar interconnections about lineage plasticity, EMT, and anti-androgen resistance have been reported in prostate cancer[89,90] and small cell lung cancer.[91] These correspondences indicate possible generalizability of our results to other cancer types as well.

Lineage plasticity is being increasingly reported in the context of breast cancer but with limited mechanistic understanding. For instance, mature mammary luminal epithelial cells can give rise to Krt14- and Sox9-expressing basal-like carcinomas that can metastasize.[78] Similarly, basal-like tumorigenesis involves luminal-to-basal reprogramming with gain in stemness.[30,92] Further, luminal-basal hybrid cells can express NP63 (basal marker) while maintaining functional levels of ER-alpha (associated with luminal phenotype).[79] Our systems-level analysis integrating multi-modal transcriptomic data with mechanism-based models for underlying regulatory networks explains these *in vitro* and *in vivo* observations, reveals hallmarks of basal breast cancer in terms of EMT, and offers a predictive platform to better characterize and control intra-tumor phenotypic heterogeneity in breast cancer.

Limitations of the study

Our analysis of breast cancer subtypes is focused on the broad categories of being clinically or molecularly labeled as luminal or basal. However, there are other subtypes of breast cancer that are not studied here, for example, HER2 subtype or subclassifications within the basal or luminal subtypes are not considered. We have considered majorly methylation as a mode of epigenetic regulation, which does not consider accessibility of promoters/enhancers or three-dimensional structures of the chromatin that can contribute to gene regulation crosstalk. Our minimalistic gene regulatory network explains how different cell phenotypes are emergent of pathways that are coupled with one another but does not have the granularity to distinguish among more specific biological microstates or phenotypes seen in breast cancer. Including other relevant genes to the constructed network considered here would be key to explaining more nuanced phenotypes that are biologically observed.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines and culture
- METHOD DETAILS
  - ssGSEA scores for bulk transcriptomics
  - Survival analysis
  - Methylation data analysis
  - Spatial transcriptomics data analysis
  - Dual immunofluorescence
  - Single cell RNA sequencing and cell-cell communication data analysis
  - RACIPE simulations
  - MINT-chip data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110116.

ACKNOWLEDGMENTS

This work was supported by Ramanujan Fellowship awarded by SERB (Science and Engineering Research Board), Department of Science and Technology (DST), Government of India, awarded to M.K.J. (SB/S2/RJN-049/2018). S.S. is supported by PMRF (Prime Ministers Research Fellowship) awarded by DST, Government of India. J.A.S. is supported by NCI 1R01CA233585-04. M.K.J. is also supported by Param Hansa Philanthropies.

AUTHOR CONTRIBUTIONS

Conceptualized and designed research: M.K.J.; supervised research: C.L.C., J.S.P., J.A.S., and M.K.J.; performed research: S.S., S.R., M.G.N., M.P., B.P.S.N., S.M., C.M.N., A.D.M., H.S., and A.G.N.; interpreted data: S.S., S.R., M.G.N., M.P., B.P.S.N., H.Z.M., S.M., C.M.N., A.D.M., C.L.C., J.S.P., J.A.S., and M.K.J.; funding acquisition: J.S.P., J.A.S., and M.K.J.; manuscript writing/editing: S.S. (prepared first draft), J.A.S., and M.K.J. (edited with inputs from all authors).

DECLARATION OF INTERESTS

The authors declare no conflicts of interest.

Received: November 6, 2023
Revised: April 4, 2024
Accepted: May 23, 2024
Published: May 27, 2024

REFERENCES

1. Marusyk, A., Janiszewska, M., and Polyak, K. (2020). Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. Cancer Cell *37*, 471–484. https://doi.org/10.1016/j.ccell.2020.03.007.
2. Skibinski, A., and Kuperwasser, C. (2015). The origin of breast tumor heterogeneity. Oncogene *34*, 5309–5316. https://doi.org/10.1038/onc.2014.475.
3. Brown, M.S., Abdollahi, B., Wilkins, O.M., Lu, H., Chakraborty, P., Ognjenovic, N.B., Muller, K.E., Jolly, M.K., Christensen, B.C., Hassanpour, S., and Pattabiraman, D.R. (2022). Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. Sci. Adv. *8*, eabj8002. https://doi.org/10.1126/sciadv.abj8002.
4. Hong, D., Fritz, A.J., Zaidi, S.K., van Wijnen, A.J., Nickerson, J.A., Imbalzano, A.N., Lian, J.B., Stein, J.L., and Stein, G.S. (2018). Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity. J. Cell. Physiol. *233*, 9136–9144. https://doi.org/10.1002/jcp.26847.
5. Wahl, G.M., and Spike, B.T. (2017). Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. NPJ Breast Cancer *3*, 14. https://doi.org/10.1038/s41523-017-0012-z.
6. Jain, P., Bhatia, S., Thompson, E.W., and Jolly, M.K. (2022). Population Dynamics of Epithelial-Mesenchymal Heterogeneity in Cancer Cells. Biomolecules *12*, 348. https://doi.org/10.3390/biom12030348.
7. Jia, D., Park, J.H., Kaur, H., Jung, K.H., Yang, S., Tripathi, S., Galbraith, M., Deng, Y., Jolly, M.K., Kaipparettu, B.A., et al. (2021). Towards decoding the coupled decision-making of metabolism and epithelial-to-mesenchymal transition in cancer. Br. J. Cancer *124*, 1902–1911. https://doi.org/10.1038/s41416-021-01385-y.
8. Sahoo, S., Mishra, A., Kaur, H., Hari, K., Muralidharan, S., Mandal, S., and Jolly, M.K. (2021). A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer *3*, zcab027. https://doi.org/10.1093/narcan/zcab027.
9. Lesniak, D., Sabri, S., Xu, Y., Graham, K., Bhatnagar, P., Suresh, M., and Abdulkarim, B. (2013). Spontaneous Epithelial-Mesenchymal Transition and Resistance to HER-2-Targeted Therapies in HER-2-Positive Luminal Breast Cancer. PLoS One *8*, e71987. https://doi.org/10.1371/journal.pone.0071987.
10. Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res. *12*, R68. https://doi.org/10.1186/bcr2635.
11. Bierie, B., Pierce, S.E., Kroeger, C., Stover, D.G., Pattabiraman, D.R., Thiru, P., Liu Donaher, J., Reinhardt, F., Chaffer, C.L., Keckesova, Z., and Weinberg, R.A. (2017). Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. Proc. Natl. Acad. Sci. USA *114*, E2337–E2346. https://doi.org/10.1073/pnas.1618298114.
12. Grosse-Wilde, A., Fouquier d'Hérouël, A., McIntosh, E., Ertaylan, G., Skupin, A., Kuestner, R.E., del Sol, A., Walters, K.-A., and Huang, S. (2015). Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. PLoS One *10*, e0126522. https://doi.org/10.1371/journal.pone.0126522.
13. Jolly, M.K., Boareto, M., Debeb, B.G., Aceto, N., Farach-Carson, M.C., Woodward, W.A., and Levine, H. (2017). Inflammatory breast cancer: a model for investigating cluster-based dissemination. NPJ Breast Cancer *3*, 21. https://doi.org/10.1038/s41523-017-0023-9.
14. Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., et al. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science *339*, 580–584. https://doi.org/10.1126/science.1228522.
15. Mohamed, G.A., Mahmood, S., Ognjenovic, N.B., Lee, M.K., Wilkins, O.M., Christensen, B.C., Muller, K.E., and Pattabiraman, D.R. (2023). Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits. Breast Cancer Res. *25*, 23. https://doi.org/10.1186/s13058-023-01621-8.
16. Ignatiadis, M., and Sotiriou, C. (2013). Luminal breast cancer: from biology to treatment. Nat. Rev. Clin. Oncol. *10*, 494–506. https://doi.org/10.1038/nrclinonc.2013.124.
17. Yin, L., Duan, J.-J., Bian, X.-W., and Yu, S.C. (2020). Triple-negative breast cancer molecular subtyping and treatment progress. Breast Cancer Res. *22*, 61. https://doi.org/10.1186/s13058-020-01296-5.
18. Qin, S., Jiang, J., Lu, Y., Nice, E.C., Huang, C., Zhang, J., and He, W. (2020). Emerging role of tumor cell plasticity in modifying therapeutic response. Signal Transduct. Target. Ther. *5*, 228. https://doi.org/10.1038/s41392-020-00313-5.
19. Lindström, L.S., Yau, C., Czene, K., Thompson, C.K., Hoadley, K.A., Van't Veer, L.J., Balassanian, R., Bishop, J.W., Carpenter, P.M., Chen, Y.Y., et al. (2018). Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal Breast cancer. J. Natl. Cancer Inst. *110*, 726–733. https://doi.org/10.1093/jnci/djx270.
20. Yu, K.D., Cai, Y.W., Wu, S.Y., Shui, R.H., and Shao, Z.M. (2021). Estrogen receptor-low breast cancer: Biology chaos and treatment paradox. Cancer Commun. *41*, 968–980. https://doi.org/10.1002/cac2.12191.
21. Bhattacharya, S., Mohanty, A., Achuthan, S., Kotnala, S., Jolly, M.K., Kulkarni, P., and Salgia, R. (2021). Group Behavior and Emergence of Cancer Drug Resistance. Trends Cancer *7*, 323–334. https://doi.org/10.1016/j.trecan.2021.01.009.
22. Burkhardt, D.B., San Juan, B.P., Lock, J.G., Krishnaswamy, S., and Chaffer, C.L. (2022). Mapping Phenotypic Plasticity upon the Cancer Cell State Landscape Using Manifold Learning. Cancer Discov. *12*, 1847–1859. https://doi.org/10.1158/2159-8290.CD-21-0282.
23. Pillai, M., Hojel, E., Jolly, M.K., and Goyal, Y. (2023). Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. Nat. Comput. Sci. *3*, 301–313. https://doi.org/10.1038/s43588-023-00427-0.
24. Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., and Stanisławek, A. (2021). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. Cancers *13*, 4287. https://doi.org/10.3390/cancers13174287.
25. Turner, K.M., Yeo, S.K., Holm, T.M., Shaughnessy, E., and Guan, J.-L. (2021). Heterogeneity within molecular subtypes of breast cancer. Am. J. Physiol. Cell Physiol. *321*, C343–C354. https://doi.org/10.1152/ajpcell.00109.2021.
26. Tan, T.Z., Miow, Q.H., Miki, Y., Noda, T., Mori, S., Huang, R.Y.J., and Thiery, J.P. (2014). Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer

patients. EMBO Mol. Med. *6*, 1279–1293. https://doi.org/10.15252/emmm.201404208.
27. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.
28. Knutsen, E., Das Sajib, S., Fiskaa, T., Lorens, J., Gudjonsson, T., Mælandsmo, G.M., Johansen, S.D., Seternes, O.-M., and Perander, M. (2023). Identification of a core EMT signature that separates basal-like breast cancers into partial- and post-EMT subtypes. Front. Oncol. *13*, 1249895. https://doi.org/10.3389/fonc.2023.1249895.
29. Nair, M.G., Apoorva, D., Chandrakala, M., Snijesh, V., Anupama, C., Rajarajan, S., Sahoo, S., Mohan, G., Jayakumar, V.S., Ramesh, R.S., et al. (2022). Acquisition of hybrid E/M phenotype associated with increased migration , drug resistance and stemness is mediated by reduced miR-18a levels in ER-negative breast cancer. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.05.505398.
30. Landragin, C., Saichi, M., Prompsy, P., Durand, A., Mesple, J., Trouchet, A., Faraldo, M., Salmon, H., and Vallot, C. (2022). Luminal progenitors undergo partial epithelial-to-mesenchymal transition at the onset of basal-like breast tumorigenesis. Preprint at bioRxiv. https://doi.org/10.1101/2022.06.08.494710.
31. Han, Y., Villarreal-Ponce, A., Gutierrez, G., Nguyen, Q., Sun, P., Wu, T., Sui, B., Berx, G., Brabletz, T., Kessenbrock, K., et al. (2022). Coordinate control of basal epithelial cell fate and stem cell maintenance by core EMT transcription factor Zeb1. Cell Rep. *38*, 110240. https://doi.org/10.1016/j.celrep.2021.110240.
32. Grasset, E.M., Dunworth, M., Sharma, G., Loth, M., Tandurella, J., Cimino-Mathews, A., Gentz, M., Bracht, S., Haynes, M., Fertig, E.J., and Ewald, A.J. (2022). Triple-negative breast cancer metastasis involves complex epithelial-mesenchymal transition dynamics and requires vimentin. Sci. Transl. Med. *14*, eabn7571. https://doi.org/10.1126/scitranslmed.abn7571.
33. Cope, L.M., Fackler, M.J., Lopez-Bujanda, Z., Wolff, A.C., Visvanathan, K., Gray, J.W., Sukumar, S., and Umbricht, C.B. (2014). Do Breast Cancer Cell Lines Provide a Relevant Model of the Patient Tumor Methylome? PLoS One *9*, e105545. https://doi.org/10.1371/journal.pone.0105545.
34. Bhatia, S., Monkman, J., Blick, T., Pinto, C., Waltham, M., Nagaraj, S.H., and Thompson, E.W. (2019). Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. J. Clin. Med. *8*, 893. https://doi.org/10.3390/jcm8060893.
35. Deshmukh, A.P., Vasaikar, S.V., Tomczak, K., Tripathi, S., Den Hollander, P., Arslan, E., Chakraborty, P., Soundararajan, R., Jolly, M.K., Rai, K., et al. (2021). Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. Proc. Natl. Acad. Sci. USA *118*, e2102050118. https://doi.org/10.1073/pnas.2102050118.
36. Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715. https://doi.org/10.1016/j.cell.2008.03.027.
37. San Juan, B.P., Hediyeh-Zadeh, S., Rangel, L., Milioli, H.H., Rodriguez, V., Bunkum, A., Kohane, F.V., Purcell, C.A., Bhuva, D.D., Kurumlian, A., et al. (2022). Targeting phenotypic plasticity prevents metastasis and the development of chemotherapy-resistant disease. Preprint at medRxiv. https://doi.org/10.1101/2022.03.21.22269988.
38. Dravis, C., Chung, C.Y., Lytle, N.K., Herrera-Valdez, J., Luna, G., Trejo, C.L., Reya, T., and Wahl, G.M. (2018). Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. Cancer Cell *34*, 466–482.e6. https://doi.org/10.1016/j.ccell.2018.08.001.
39. Dang, T.T., Esparza, M.A., Maine, E.A., Westcott, J.M., and Pearson, G.W. (2015). ΔNp63α promotes breast cancer cell motility through the selective activation of components of the epithelial-to-mesenchymal transition program. Cancer Res. *75*, 3925–3935. https://doi.org/10.1158/0008-5472.CAN-14-3363.
40. Westcott, J.M., Camacho, S., Nasir, A., Huysman, M.E., Rahhal, R., Dang, T.T., Riegel, A.T., Brekken, R.A., and Pearson, G.W. (2020). ΔNp63-Regulated Epithelial-to-Mesenchymal Transition State Heterogeneity Confers a Leader-Follower Relationship That Drives Collective Invasion. Cancer Res. *80*, 3933–3944. https://doi.org/10.1158/0008-5472.CAN-20-0014.
41. Chakrabarti, R., Hwang, J., Andres Blanco, M., Wei, Y., Lukačišin, M., Romano, R.A., Smalley, K., Liu, S., Yang, Q., Ibrahim, T., et al. (2012). Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. Nat. Cell Biol. *14*, 1212–1222. https://doi.org/10.1038/ncb2607.
42. Yao, B., Zhao, J., Li, Y., Li, H., Hu, Z., Pan, P., Zhang, Y., Du, E., Liu, R., and Xu, Y. (2015). Elf5 inhibits TGF-β-driven epithelial-mesenchymal transition in prostate cancer by repressing SMAD3 activation. Prostate *75*, 872–882. https://doi.org/10.1002/pros.22970.
43. Wu, B., Cao, X., Liang, X., Zhang, X., Zhang, W., Sun, G., and Wang, D. (2015). Epigenetic regulation of Elf5 is associated with epithelial-mesenchymal transition in urothelial cancer. PLoS One *10*, e0117510. https://doi.org/10.1371/journal.pone.0117510.
44. Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N., et al. (2021). A single-cell and spatially resolved atlas of human breast cancers. Nat. Genet. *53*, 1334–1347. https://doi.org/10.1038/s41588-021-00911-1.
45. Bouris, P., Skandalis, S.S., Piperigkou, Z., Afratis, N., Karamanou, K., Aletras, A.J., Moustakas, A., Theocharis, A.D., and Karamanos, N.K. (2015). Estrogen receptor alpha mediates epithelial to mesenchymal transition, expression of specific matrix effectors and functional properties of breast cancer cells. Matrix Biol. *43*, 42–60. https://doi.org/10.1016/j.matbio.2015.02.008.
46. Lin, C.Y., Ström, A., Vega, V.B., Kong, S.L., Yeo, A.L., Thomsen, J.S., Chan, W.C., Doray, B., Bangarusamy, D.K., Ramasamy, A., et al. (2004). Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. Genome Biol. *5*, R66. https://doi.org/10.1186/gb-2004-5-9-r66.
47. Chang, M. (2012). Tamoxifen resistance in breast cancer. Biomol. Ther. *20*, 256–267. https://doi.org/10.4062/biomolther.2012.20.3.256.
48. Gambardella, G., Viscido, G., Tumaini, B., Isacchi, A., Bosotti, R., and di Bernardo, D. (2022). A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. Nat. Commun. *13*, 1714. https://doi.org/10.1038/s41467-022-29358-6.
49. Jolly, M.K., Murphy, R.J., Bhatia, S., Whitfield, H.J., Redfern, A., Davis, M.J., and Thompson, E.W. (2022). Measuring and Modelling the Epithelial- Mesenchymal Hybrid State in Cancer: Clinical Implications. Cells Tissues Organs *211*, 110–133. https://doi.org/10.1159/000515289.
50. Bernardo, G.M., Bebek, G., Ginther, C.L., Sizemore, S.T., Lozada, K.L., Miedler, J.D., Anderson, L.A., Godwin, A.K., Abdul-Karim, F.W., Slamon, D.J., and Keri, R.A. (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. Oncogene *32*, 554–563. https://doi.org/10.1038/onc.2012.62.
51. Bernardo, G.M., Lozada, K.L., Miedler, J.D., Harburg, G., Hewitt, S.C., Mosley, J.D., Godwin, A.K., Korach, K.S., Visvader, J.E., Kaestner, K.H., et al. (2010). FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. Development *137*, 2045–2054. https://doi.org/10.1242/dev.043299.
52. Eeckhoute, J., Keeton, E.K., Lupien, M., Krum, S.A., Carroll, J.S., and Brown, M. (2007). Positive cross-regulatory loop ties GATA-3 to estrogen receptor α expression in breast cancer. Cancer Res. *67*, 6477–6483. https://doi.org/10.1158/0008-5472.CAN-07-0746.
53. Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. (2006). GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. Cell *127*, 1041–1055. https://doi.org/10.1016/j.cell.2006.09.048.
54. Storci, G., Sansone, P., Trere, D., Tavolari, S., Taffurelli, M., Ceccarelli, C., Guarnieri, T., Paterini, P., Pariali, M., Montanaro, L., et al. (2008). The basal-like breast carcinoma phenotype is regulated by SLUG gene expression. J. Pathol. *214*, 25–37. https://doi.org/10.1002/path.2254.
55. Nekulova, M., Holcakova, J., Gu, X., Hrabal, V., Galtsidis, S., Orzol, P., Liu, Y., Logotheti, S., Zoumpourlis, V., Nylander, K., et al. (2016). ΔNp63α expression induces loss of cell adhesion in triple-negative breast cancer cells. BMC Cancer *16*, 782. https://doi.org/10.1186/s12885-016-2808-x.
56. Subbalakshmi, A.R., Sahoo, S., Biswas, K., and Jolly, M.K. (2022). A Computational Systems Biology Approach Identifies SLUG as a Mediator of Partial Epithelial-Mesenchymal Transition (EMT). Cells Tissues Organs *211*, 689–702. https://doi.org/10.1159/000512520.
57. Thiebaut, C., Chamard-Jovenin, C., Chesnel, A., Morel, C., Djermoune, E.H., Boukhobza, T., and Dumond, H. (2017). Mammary epithelial cell phenotype disruption *in vitro* and *in vivo* through ERalpha36 overexpression. PLoS One *12*, e0173931. https://doi.org/10.1371/journal.pone.0173931.
58. Bocci, F., Tripathi, S.C., Vilchez Mercedes, S.A., George, J.T., Casabar, J.P., Wong, P.K., Hanash, S.M., Levine, H., Onuchic, J.N., and Jolly, M.K. (2019). NRF2 activates a partial Epithelial-Mesenchymal Transition and is maximally present in a hybrid Epithelial/Mesenchymal phenotype. Integr.

Biol. *11*, 251–263. https://doi.org/10.1101/390237.
59. Kim, S.K., Yang, J.W., Kim, M.R., Roh, S.H., Kim, H.G., Lee, K.Y., Jeong, H.G., and Kang, K.W. (2008). Increased expression of Nrf2/ARE-dependent anti-oxidant proteins in tamoxifen-resistant breast cancer cells. Free Radic. Biol. Med. *45*, 537–546. https://doi.org/10.1016/j.freeradbiomed.2008.05.011.
60. Wang, X.-J., Sun, Z., Villeneuve, N.F., Zhang, S., Zhao, F., Li, Y., Chen, W., Yi, X., Zheng, W., Wondrak, G.T., et al. (2008). Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. Carcinogenesis *29*, 1235–1243. https://doi.org/10.1093/carcin/bgn095.
61. Huang, B., Lu, M., Jia, D., Ben-Jacob, E., Levine, H., and Onuchic, J.N. (2017). Interrogating the topological robustness of gene regulatory circuits by randomization. PLoS Comput. Biol. *13*, e1005456. https://doi.org/10.1101/084962.
62. Fan, M., Chen, J., Gao, J., Xue, W., Wang, Y., Li, W., Zhou, L., Li, X., Jiang, C., Sun, Y., et al. (2020). Triggering a switch from basal- to luminal-like breast cancer subtype by the small-molecule diptoindonesin G via induction of GABARAPL1. Cell Death Dis. *11*, 635. https://doi.org/10.1038/s41419-020-02878-z.
63. Kastrati, I., Joosten, S.E.P., Semina, S.E., Alejo, L.H., Brovkovych, S.D., Stender, J.D., Horlings, H.M., Kok, M., Alarid, E.T., Greene, G.L., et al. (2020). The NF-κB Pathway Promotes Tamoxifen Tolerance and Disease Recurrence in Estrogen Receptor-Positive Breast Cancers. Mol. Cancer Res. *18*, 1018–1027. https://doi.org/10.1158/1541-7786.MCR-19-1082.
64. Pasani, S., Sahoo, S., and Jolly, M.K. (2020). Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology. J. Clin. Med. *10*, 60. https://doi.org/10.3390/jcm10010060.
65. Dimitrov, D., Türei, D., Garrido-Rodriguez, M., Burmedi, P.L., Nagai, J.S., Boys, C., Ramirez Flores, R.O., Kim, H., Szalai, B., Costa, I.G., et al. (2022). Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat. Commun. *13*, 3224. https://doi.org/10.1038/s41467-022-30755-0.
66. Semba, R., Horimoto, Y., Sakata-Matsuzawa, M., Ishizuka, Y., Denda-Nagai, K., Fujihira, H., Noji, M., Onagi, H., Ichida, M., Miura, H., et al. (2023). Possible correlation of apical localization of MUC1 glycoprotein with luminal A-like status of breast cancer. Sci. Rep. *13*, 5281. https://doi.org/10.1038/s41598-023-32579-4.
67. Balko, J.M., Miller, T.W., Morrison, M.M., Hutchinson, K., Young, C., Rinehart, C., Sánchez, V., Jee, D., Polyak, K., Prat, A., et al. (2012). The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. Proc. Natl. Acad. Sci. USA *109*, 221–226. https://doi.org/10.1073/pnas.1115802109.
68. Li, R., Ong, S.L., Tran, L.M., Jing, Z., Liu, B., Park, S.J., Huang, Z.L., Walser, T.C., Heinrich, E.L., Lee, G., et al. (2020). Chronic IL-1β-induced inflammation regulates epithelial-to-mesenchymal transition memory phenotypes via epigenetic modifications in non-small cell lung cancer. Sci. Rep. *10*, 377. https://doi.org/10.1038/s41598-019-57285-y.
69. González-Carnicero, Z., Hernanz, R., Martínez-Casales, M., Barrús, M.T., Martín, Á., and Alonso, M.J. (2023). Regulation by Nrf2 of IL-1β-induced inflammatory and oxidative response in VSMC and its relationship with TLR4. Front. Pharmacol. *14*, 1058488. https://doi.org/10.3389/fphar.2023.1058488.
70. Mendoza-Rodríguez, M.G., Ayala-Sumuano, J.T., García-Morales, L., Zamudio-Meza, H., Pérez-Yepez, E.A., and Meza, I. (2019). IL-1β inflammatory cytokine-induced TP63 isoform ΔNP63α signaling cascade contributes to cisplatin resistance in human breast cancer cells. Int. J. Mol. Sci. *20*, 270. https://doi.org/10.3390/ijms20020270.
71. Castaño, Z., San Juan, B.P., Spiegel, A., Pant, A., DeCristo, M.J., Laszewski, T., Ubellacker, J.M., Janssen, S.R., Dongre, A., Reinhardt, F., et al. (2018). IL-1β inflammatory response driven by primary breast cancer prevents metastasis-initiating cell colonization. Nat. Cell Biol. *20*, 1084–1097. https://doi.org/10.1038/s41556-018-0173-5.
72. van Boxtel, C., van Heerden, J.H., Nordholt, N., Schmidt, P., and Bruggeman, F.J. (2017). Taking chances and making mistakes : non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. J. R. Soc. Interface *14*, 20170141. https://doi.org/10.1098/rsif.2017.0141.
73. Chauhan, L., Ram, U., Hari, K., and Jolly, M.K. (2021). Topological signatures in regulatory network enable phenotypic heterogeneity in small cell lung cancer. Elife *10*, e64522.
74. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468. https://doi.org/10.1038/s41586-018-0040-3.
75. Rambow, F., Marine, J.C., and Goding, C.R. (2019). Melanoma plasticity and phenotypic diversity: Therapeutic barriers and opportunities. Genes Dev. *33*, 1295–1318. https://doi.org/10.1101/gad.329771.119.
76. Jia, D., Lu, M., Jung, K.H., Park, J.H., Yu, L., Onuchic, J.N., Kaipparettu, B.A., and Levine, H. (2019). Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proc. Natl. Acad. Sci. USA *116*, 3909–3918. https://doi.org/10.1073/pnas.1816391116.
77. Karacosta, L.G., Anchang, B., Ignatiadis, N., Kimmey, S.C., Benson, J.A., Shrager, J.B., Tibshirani, R., Bendall, S.C., and Plevritis, S.K. (2019). Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. Nat. Commun. *10*, 5587. https://doi.org/10.1101/570341.
78. Kern, J.G., Tilston-Lunel, A.M., Federico, A., Ning, B., Mueller, A., Peppler, G.B., Stampouloglou, E., Cheng, N., Johnson, R.L., Lenburg, M.E., et al. (2022). Inactivation of LATS1/2 drives luminal-basal plasticity to initiate basal-like mammary carcinomas. Nat. Commun. *13*, 7198. https://doi.org/10.1038/s41467-022-34864-8.
79. Padua, M.B., Bhat-Nakshatri, P., Anjanappa, M., Prasad, M.S., Hao, Y., Rao, X., Liu, S., Wan, J., Liu, Y., McElyea, K., et al. (2018). Dependence receptor UNC5A restricts luminal to basal breast cancer plasticity and metastasis. Breast Cancer Res. *20*, 35. https://doi.org/10.1186/s13058-018-0963-5.
80. Beltran, H., Hruszkewycz, A., Scher, H.I., Hildesheim, J., Isaacs, J., Yu, E.Y., Kelly, K., Lin, D., Dicker, A., Arnold, J., et al. (2019). The role of lineage plasticity in prostate cancer therapy resistance. Clin. Cancer Res. *25*, 6916–6924. https://doi.org/10.1158/1078-0432.CCR-19-1423.
81. Vipparthi, K., Hari, K., Chakraborty, P., Ghosh, S., Patel, A.K., Ghosh, A., Biswas, N.K., Sharan, R., Arun, P., Jolly, M.K., and Singh, S. (2022). Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. iScience *25*, 104317. https://doi.org/10.1016/j.isci.2022.104317.
82. Milioli, H.H., Tishchenko, I., Riveros, C., Berretta, R., and Moscato, P. (2017). Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. BMC Med. Genom. *10*, 19. https://doi.org/10.1186/s12920-017-0250-9.
83. Hari, K., Ullanat, V., Balasubramanian, A., Gopalan, A., and Jolly, M.K. (2022). Landscape of epithelial mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. Elife *11*, e76535. https://doi.org/10.7554/elife.76535.
84. Kröger, C., Afeyan, A., Mraz, J., Eaton, E.N., Reinhardt, F., Khodor, Y.L., Thiru, P., Bierie, B., Ye, X., Burge, C.B., and Weinberg, R.A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. Proc. Natl. Acad. Sci. USA *116*, 7353–7362. https://doi.org/10.1073/pnas.1812876116.
85. Ali, H.R., Glont, S.E., Blows, F.M., Provenzano, E., Dawson, S.J., Liu, B., Hiller, L., Dunn, J., Poole, C.J., Bowden, S., et al. (2015). PD-L1 protein expression in breast cancer is rare, enriched in basal-like tumours and associated with infiltrating lymphocytes. Ann. Oncol. *26*, 1488–1493. https://doi.org/10.1093/annonc/mdv192.
86. Sahoo, S., Nayak, S.P., Hari, K., Purkait, P., Mandal, S., Kishore, A., Levine, H., and Jolly, M.K. (2021). Immunosuppressive Traits of the Hybrid Epithelial/Mesenchymal Phenotype. Front. Immunol. *12*, 797261. https://doi.org/10.3389/fimmu.2021.797261.
87. Al Saleh, S., Al Mulla, F., and Luqmani, Y.A. (2011). Estrogen Receptor Silencing Induces Epithelial to Mesenchymal Transition in Human Breast Cancer Cells. PLoS One *6*, e20610. https://doi.org/10.1371/journal.pone.0020610.
88. Hiscox, S., Jiang, W.G., Obermeier, K., Taylor, K., Morgan, L., Burmi, R., Barrow, D., and Nicholson, R.I. (2006). Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of β-catenin phosphorylation. Int. J. Cancer *118*, 290–301. https://doi.org/10.1002/ijc.21355.
89. Mu, P., Zhang, Z., Benelli, M., Karthaus, W.R., Hoover, E., Chen, C.C., Wongvipat, J., Ku, S.Y., Gao, D., Cao, Z., et al. (2017). SOX2 promotes lineage plasticity and antiandrogen resistance in TP53-and RB1-deficient prostate cancer. Science *355*, 84–88. https://doi.org/10.1126/science.aah4307.
90. Jindal, R., Nanda, A., Pillai, M., Ware, K.E., Singh, D., Sehgal, M., Armstrong, A.J., Somarelli, J.A., and Jolly, M.K. (2023). Emergent dynamics of underlying regulatory network links EMT and androgen receptor-dependent resistance in prostate cancer. Comput. Struct. Biotechnol. J. *21*, 1498–1509. https://doi.org/10.1016/j.csbj.2023.01.031.
91. Groves, S.M., Panchy, N., Tyson, D.R., Harris, L.A., Quaranta, V., and Hong, T. (2023). Involvement of Epithelial–Mesenchymal Transition Genes in Small Cell Lung Cancer Phenotypic Plasticity. Cancers *15*, 1477. https://doi.org/10.3390/cancers15051477.

92. Christin, J.R., Wang, C., Chung, C.Y., Liu, Y., Dravis, C., Tang, W., Oktay, M.H., Wahl, G.M., and Guo, W. (2020). Stem Cell Determinant SOX9 Promotes Lineage Plasticity and Progression in Basal-like Breast Cancer. Cell Rep. *31*, 107742. https://doi.org/10.1016/j.celrep.2020.107742.

93. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.

94. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: Single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

95. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112. https://doi.org/10.1038/nature08460.

96. Fang, Z., Liu, X., and Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics *39*, btac757. https://doi.org/10.1093/bioinformatics/btac757.

97. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27. https://doi.org/10.1016/j.cell.2018.05.061.

98. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

99. Nair, M.G., Desai, K., Prabhu, J.S., Hari, P.S., Remacle, J., and Sridhar, T.S. (2016). β3 integrin promotes chemoresistance to epirubicin in MDA-MB-231 through repression of the pro-apoptotic protein, BAD. Exp. Cell Res. *346*, 137–145. https://doi.org/10.1016/j.yexcr.2016.05.015.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Analyzed dataset	Cope et al.[33]	GSE42944
Analyzed dataset	San Juan et al.[37]	GSE184647
Analyzed dataset	Gambardella et al.[48]	GSE173634
Analyzed dataset	Wu et al.[44]	GSE176078
Epithelial & Mesenchymal gene signatures	Tan et al.[26]	N/A
pEMT gene signature	Puram et al.[27]	N/A
Hallmark EMT, E2F and Estrogen Response signatures	Liberzon et al.[93]	N/A
Breast cancer specific Epithelial, Mesenchymal and pEMT gene signatures	Knutsen et al.[28]	N/A
Antibodies		
E-cadherin	Rabbit monoclonal EP700Y- ab40772 Abcam	RRID: AB_731493
Vimentin	Mouse clone V9- AM074-5M Bio-genex	RRID: AB_3101770
Experimental models: Cell lines		
MCF-7	ATCC	N/A
BT-474	Gift from IISC-Dr Annapoorni Rangarajan lab	N/A
MDA-MB-453	Gift from IISC-Dr Annapoorni Rangarajan lab	N/A
HCC-1937	ATCC	N/A
MDA-MB-231	ATCC	N/A
MDA-MB-468	NCCS	N/A
SKBR3	ATCC	N/A
ZR-75-1	NCCS	N/A
Software and algorithms		
RACIPE (Random Circuit Perturbation)	Huang et al.[61]	https://github.com/simonhb1990/RACIPE-1.0
AUCell	Aibar et al.[94]	https://github.com/aertslab/AUCell
Single Sample Gene Set Enrichment Analysis (ssGSEA)	Barbie et al.[95]	https://github.com/zqfang/GSEApy
GSEApy	Fang et al.[96]	https://github.com/zqfang/GSEApy
MAGIC	van Dijk et al.[97]	https://github.com/KrishnaswamyLab/MAGIC/
LIANA	Dimitrov et al.[65]	https://github.com/saezlab/liana
Analysis codes	This paper	https://github.com/sarthak-sahoo-0710/luminal_basal_EMT_crosstalk

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Mohit Kumar Jolly (mkjolly@iisc.ac.in).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The accession numbers for the existing, publicly available datasets are listed in the key resources table.
- All codes used in the manuscript are available at: https://github.com/sarthak-sahoo-0710/luminal_basal_EMT_crosstalk.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and culture

Cell lines MDA-MB-231, MCF7 and HCC1937 were obtained from the American Type Culture Collection (ATCC- Manassas, VA). MDA-MB-468 and ZR-75-1 were obtained from NCCS (Pune, India) where cell authentication was performed using STR profiling. MDA-MB-231 and MDA-MB-468 were maintained in L-15 (Leibovitz) medium (Sigma-Aldrich), MCF7 in DMEM-Hi Glucose medium (Sigma-Aldrich) HCC1937 and ZR-75-1 in RPMI 1640 media (Gibco), HEPES buffered and supplemented with 10% (v/v) heat inactivated Fetal Bovine Serum (Himedia) and 100 U/ml penicillin and streptomycin (Gibco). All cells were maintained in a humidified incubator with 5% CO2 at 37°C except for MDA-MB-231 and MDA-MB-468 that were maintained with 0% CO2. For all experimental assays using cell lines, a passage number below 20 was used and all cell lines were subjected to frequent recharacterization by immunophenotyping and testing of mycoplasma.

METHOD DETAILS

ssGSEA scores for bulk transcriptomics

We used previously published signatures for luminal and basal breast cancer,[29] for epithelial and mesenchymal state[26] and for pEMT state[27] Hallmark signatures (estrogen response, E2F target genes) were taken from MSigDB database.[93] Breast cancer specific epithelial (EMT_down), pEMT (EMT_partial) and mesenchymal (EMT_up) were taken from previously published literature.[28] ssGSEA scores were calculated for bulk transcriptomic samples using the gseapy python package[96] to estimate the activity of biological pathway of interest. A correlation was considered significant if the Pearson's correlation coefficient is greater than 0.3 or lesser than 0.3 with a p-value lesser than 0.05. Meta analysis was performed on a list of 80 breast cancer specific bulk RNA/microarray transcriptomic datasets (Table S3).

Survival analysis

Overall survival data was acquired from TCGA. Based on the median of sample scores, all samples were split into 4 groups : epithelial-high mesenchymal-low (EPI+MES-) (reference group), epithelial -high mesenchymal-high (EPI+MES+), epithelial-low mesenchymal-high (EPI-MES+), and epithelial-low mesenchymal-low (EPI-MES-). Similarly, epithelial-high luminal-high (EPI+LUM+) (reference group), epithelial-high luminal-low (EPI+LUM-), epithelial-low luminal-high (EPI-LUM+), and epithelial-low luminal-low (EPI-LUM-). The R package 'survival' was employed to perform the Kaplan–Meier analysis. Reported p-values were calculated using a log-rank test. Cox regression was used to determine the hazard ratio (HR) and confidence interval (95% CI) for TCGA cohorts, and forest plots were made using 'ggforest' function from 'survminer' package.

Methylation data analysis

Methylation data (beta values) from GSE42944 for breast cancer cell lines were downloaded from Gene Expression Omnibus. The beta values for CpG islands vary between 0 (unmethylated) and 1 (fully methylated). The heatmap included only previously identified epithelial and mesenchymal genes[26] and cell lines included in CCLE breast cancer cohort to facilitate direct comparison between the RNA-Seq and methylation data. The subtype classification for each cell line was taken from GSE42944. We further performed z-normalisation and scaled the values between 0 and 1 to portray only the relative amount of methylation values for each gene across the cell lines.

Spatial transcriptomics data analysis

Spatial transcriptomics datasets in the public domain for 6 patients (2 estrogen receptor positive and 4 triple negative patients)[44] were reanalysed for assessing the activity of various biological pathways. Count matrices were first imputed by MAGIC algorithm[97] and activity scores were calculated on imputed values using AUCell.[94] Pre-processing of spatial data and images was done as per the Seurat pipeline.[98]

Dual immunofluorescence

Cells ($1x10^4$) were seeded on poly-L-lysine coated coverslips. Immunofluorescence was performed as reported previously[99] by incubating cells in primary antibodies- anti-E-cadherin (Abcam-EP700Y) and anti-Vimentin (BioGenex) overnight at 4°C at specific dilutions- 1:500 and 1:25 respectively. This was followed by labelling with specific secondary antibodies - Alexa Fluor® 488 Chicken Anti-Mouse IgG (H+L) for Anti-Vimentin and Alexa Fluor 568 Donkey Anti-Rabbit IgG for anti-E-cadherin for 1 h at room temperature. The slide was then mounted on gold antifade reagent with DAPI and examined under a fluorescent microscope (Olympus BX51).

Single cell RNA sequencing and cell-cell communication data analysis

Count matrices for single-cell RNA sequencing data were imputed by MAGIC algorithm.[97] Activity scores were calculated on imputed values using AUCell.[94] Lists of top cell-cell communication receptor ligand pairs were estimated using the LIANA package.[65] For estimating the top ligand receptor pairs, pre-labelled luminal A and basal cell types from each patient were considered and all other cell types including themselves were considered as potential ligand producing cells while the receptors were assumed to be only expressed on the chosen luminal A or basal subtype of cells. Only those ligand receptor pairs which had a cellphonedb p-value < 0.05 and sca LRscore > 0.8 and were expressed in all ER+ or TNBC were considered as unique/common ligands/receptors. Ligands or receptors that were only specific

to luminal cells in ER+ breast cancer patients were considered to be luminal specific ligands/ receptors. Similar analysis was done for basal cells in TNBC as well. Top genes were subjected to meta-analysis with relevant pathways & their expression in bulk transcriptomics.

RACIPE simulations

Random Circuit Perturbation (RACIPE) was employed to generate an ensemble of kinetic models for a given GRN. The GRN contains nodes and edges (inhibitory or activating) among them. The dynamics of each node was determined using a set of coupled ordinary differential equations (ODEs).[61] Each node/gene had basal production and degradation rates as ODE parameters. Shifted Hill functions were multiplied to the production rate to incorporate the effects of excitatory and inhibitory links incoming to that node/gene. All steady-state values obtained from RACIPE, which were initially in log2 scale, were converted into z-scores, to indicate relative levels. RACIPE simulations were done in triplicates, each replicate with 10,000 parameter sets, and 100 initial conditions for each parameter set. Euler's Method was employed for numerical integration. RACIPE chooses kinetic parameters from a large range of biologically realistic parameter values to identify a majority of states that are allowed by a given GRN. A single RACIPE parameter set and associated random initial conditions has the potential to produce one or more stable steady-state solutions. However, for this analysis, up to six stable steady-state solutions were considered.

The luminal score was calculated as sum of normalised steady state values of ERα66, GATA3, PGR, and FOXA1. Similarly, basal score included ΔNP63 and SLUG; epithelial score incorporated CDH1 and miR-200 and the mesenchymal score consisted of ZEB1 and SLUG. Furthermore, the resistance (to anti-ER drugs) score was calculated as the difference between steady-state values of ERα36 and ERα66. Additionally, RACIPE was used to perform overexpression (OE; 100x) of TGFB1 & IL1B genes separately for a modified GRN consisting of both these genes and the steady state results obtained were then compared with control (not OE) RACIPE simulations.

MINT-chip data analysis

MINT-Chip data for 4 breast cancer cell lines (MCF7, ZR-75-1, HCC38 and HMLER) were processed to obtain the enriched promoters (considered to be 5000 base pairs either side of the transcriptional start site) for different cell lines and CD44 status. First, a list of top correlated epithelial and mesenchymal genes (Spearman's correlation coefficient > 0.5 and p-value < 0.05) were obtained from RNA seq data of the same cell lines (GSE184647). Amongst these genes which were labelled to be epithelial or mesenchymal, the proportion of genes that had an enriched promoter in either activation (H3K27ac) or inhibitory (H3K27me3) marks were quantified and compared across the cell lines or cell lines with specific CD44 status.

QUANTIFICATION AND STATISTICAL ANALYSIS

We computed the Pearson's correlation coefficients and used corresponding p-values to gauge the strength of correlations. For statistical comparison between groups, we used a two-tailed Student's t-test under the assumption of unequal variances and computed significance. Details of statistical analysis, definitions for significance, dataset IDs and the description of n (sample size) and abbreviations can be found in main text and figure legends.