

Multi-modal transcriptomic analysis unravels enrichment of hybrid epithelial/mesenchymal state and enhanced phenotypic heterogeneity in basal breast cancer

Running title: Hybrid E/M phenotype and phenotypic heterogeneity in basal breast cancer

Sarthak Sahoo¹, Soundharya Ramu¹, Madhumathy G Nair², Maalavika Pillai^{1,3}, Beatriz P San Juan⁴, Heloisa Zaccaron Milioli⁴, Susmita Mandal¹, Chandrakala M Naidu², Apoorva D Mavatkar², Harini Subramaniam¹, Arpita G Neogi¹, Christine L Chaffer^{4,5}, Jyothi S Prabhu², Jason A Somarelli⁶, Mohit Kumar Jolly^{1,*}

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 Current affiliation: Feinberg School of Medicine, Northwestern University, Chicago, 60611, USA

⁴ Garvan Institute of Medical Research, Darlinghurst, NSW, 2010, Australia

⁵ University of New South Wales, UNSW Medicine, UNSW Sydney, NSW, 2052, Australia

⁶ Department of Medicine, Duke University, Durham, NC 27708, USA

* Author to whom correspondence should be addressed: mkjolly@iisc.ac.in (M.K.J)

Abstract

Intra-tumoral phenotypic heterogeneity promotes tumor relapse and therapeutic resistance and remains an unsolved clinical challenge. It manifests along multiple phenotypic axes and decoding the interconnections among these different axes is crucial to understand its molecular origins and to develop novel therapeutic strategies to control it. Here, we use multi-modal transcriptomic data analysis – 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. These patterns were inherent in methylation profiles, suggesting an epigenetic crosstalk between EMT and lineage plasticity in breast cancer. Mathematical modelling of core underlying gene regulatory networks representative of the crosstalk between the luminal-basal and epithelial-mesenchymal axes recapitulate and thus 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 to identify possible interventions to restrict it.

Keywords: lineage plasticity, epithelial-mesenchymal heterogeneity, regulatory networks, hybrid epithelial/mesenchymal, luminal-basal transition

Introduction

Intra-tumoral heterogeneity in breast cancer remains a key obstacle in the effective management of the disease (Marusyk et al., 2020). A major determinant of molecular heterogeneity in breast cancer is attributed to molecular subtype characteristics, which can be broadly classified as luminal or basal (Skibinski and Kuperwasser, 2015). 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 an increased overall phenotypic heterogeneity (Brown et al., 2022; Hong et al., 2018; Jain et al., 2022; Wahl and Spike, 2017). 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 (Jia et al., 2021; Sahoo et al., 2021a). Specifically, in breast cancer, two mainstays of molecular heterogeneity that are often used interchangeably are the luminal-basal and epithelial-hybrid-mesenchymal states (Lesniak et al., 2013). This assumed equivalence, based at least partly on gene set enrichment analysis, largely considers EMT as a binary process (Prat et al., 2010). 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 (Bierie et al., 2017; Grosse-Wilde et al., 2015; Jolly et al., 2017; Yu et al., 2013). 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 (Mohamed et al., 2023). Similarly, the extent to which associations between partial/full EMT and luminal-basal plasticity in breast cancer are generalizable also remain to be elucidated.

Current therapeutic approaches often target specific molecular subtypes of breast cancer (Ignatiadis and Sotiriou, 2013; Yin et al., 2020). However, therapy-driven adaptive plasticity and consequent phenotypic heterogeneity pose challenges in achieving durable responses (Qin et al., 2020). In addition, the clinical implications of understanding coupling between EMT and luminal-basal plasticity, and its impact on the estrogen receptor (ER) signalling, 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 signalling (Lindström et al., 2018; Mohamed et al., 2023; Sahoo et al., 2021a; Yu et al., 2021). Heterogeneity can also drive collective behaviour by facilitating division of labour among different cancer cell subpopulations, and consequently varying interactions with the surrounding microenvironment (Bhattacharya et al., 2021). 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 (Burkhardt et al., 2022; Pillai et al., 2023).

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

A luminal gene expression signature is closely associated with an epithelial 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 (Łukasiewicz et al., 2021). 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 (Turner et al., 2021). 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 sought to compare how these four gene sets - luminal, basal, epithelial and mesenchymal - correlated with one another independently. To do this, we first calculated single sample gene set enrichment analysis (ssGSEA) scores for corresponding gene expression signatures to quantify the activity of these four gene sets. The epithelial and mesenchymal gene signatures were adapted from a previous pan-cancer analysis (Tan et al., 2014). The luminal and basal nature of breast cancer samples was assessed using a previously curated set of 15 luminal and basal specific genes each (Nair et al., 2022). There was minimal overlap between the genes included in these signatures (**Table S1**). As expected, in both CCLE cohort of cell lines and in TCGA breast cancer samples, which cover all major breast cancer subtypes (luminal A, luminal B, HER2+ and basal), the epithelial signature score correlated negatively with the mesenchymal score (CCLE: $r = -0.73$, $p < 0.01$, TCGA: $r = -0.414$, $p < 0.01$), and the luminal signature score correlated negatively with the basal score (CCLE: $r = -0.331$, $p < 0.05$, TCGA: $r = -0.369$, $p < 0.05$) (**Fig S1A-B**).

In CCLE, we observed a positive correlation of epithelial signature with the luminal signature ($r = 0.657$, $p < 0.01$) (**Fig 1A, i**). Conversely, the luminal signature correlated negatively with the mesenchymal score ($r = -0.775$, $p < 0.01$) (**Fig S1C, i**), indicating that the luminal nature was strongly associated with an elevated epithelial and reduced mesenchymal state. The basal signature, on the other hand, showed relatively weaker trends. It correlated positively with the mesenchymal nature ($r = 0.409$, $p < 0.01$) (**Fig 1A, ii**), but showed no significant correlation with the epithelial nature of cell lines (**Fig S1C, ii**). These observations corroborate previous reports of basal cells expressing some mesenchymal markers (VIM, SLUG) (Gudjonsson et al., 2005; Phillips and Kuperwasser, 2014), while also considered as a part of mammary epithelium in normal tissues and in breast cancer (Gusterson et al., 2005; Nguyen et al., 2018).

Next, we examined if the above trends were observed in primary tumor samples. Thus, we analysed the TCGA breast cancer cohort, considering all subtypes – Luminal A, Luminal B, HER2+ and Basal-like. Similar to cell line data, the luminal signature positively correlated with the epithelial signature ($r = 0.657$, $p < 0.01$) (**Fig 1B, i**). However, in contrast to the cell line data, the basal

nature was not correlated with the mesenchymal phenotype ($r = 0.047$, $p > 0.05$) (**Fig 1B, ii**), but showed negative correlation trends with the epithelial nature ($r = -0.259$, $p < 0.01$) (**Fig S1D, i**). Further, the luminal nature was not significantly correlated with the mesenchymal signature (**Fig S1D, ii**). Upon quantifying the different correlations especially in the Luminal A and the Basal-like subtypes, we found that the luminal-epithelial association was consistently positive and significant in both cases (**Table S2**). Further, in the Luminal A subtype patients, the following was observed: a) mesenchymal score does not correlate with the basal score ($r = 0.16$, $p < 0.01$) or with the luminal score ($r = -0.19$, $p < 0.01$), and b) the basal score correlates positively with epithelial score ($r = 0.35$, $p < 0.01$), suggesting that within luminal A subtype, tumors enriched for a basal character were likely to be both high in epithelial and mesenchymal nature. Similar trends, albeit to a weaker extent, were noted in the basal-like subtype (**Table S2**).

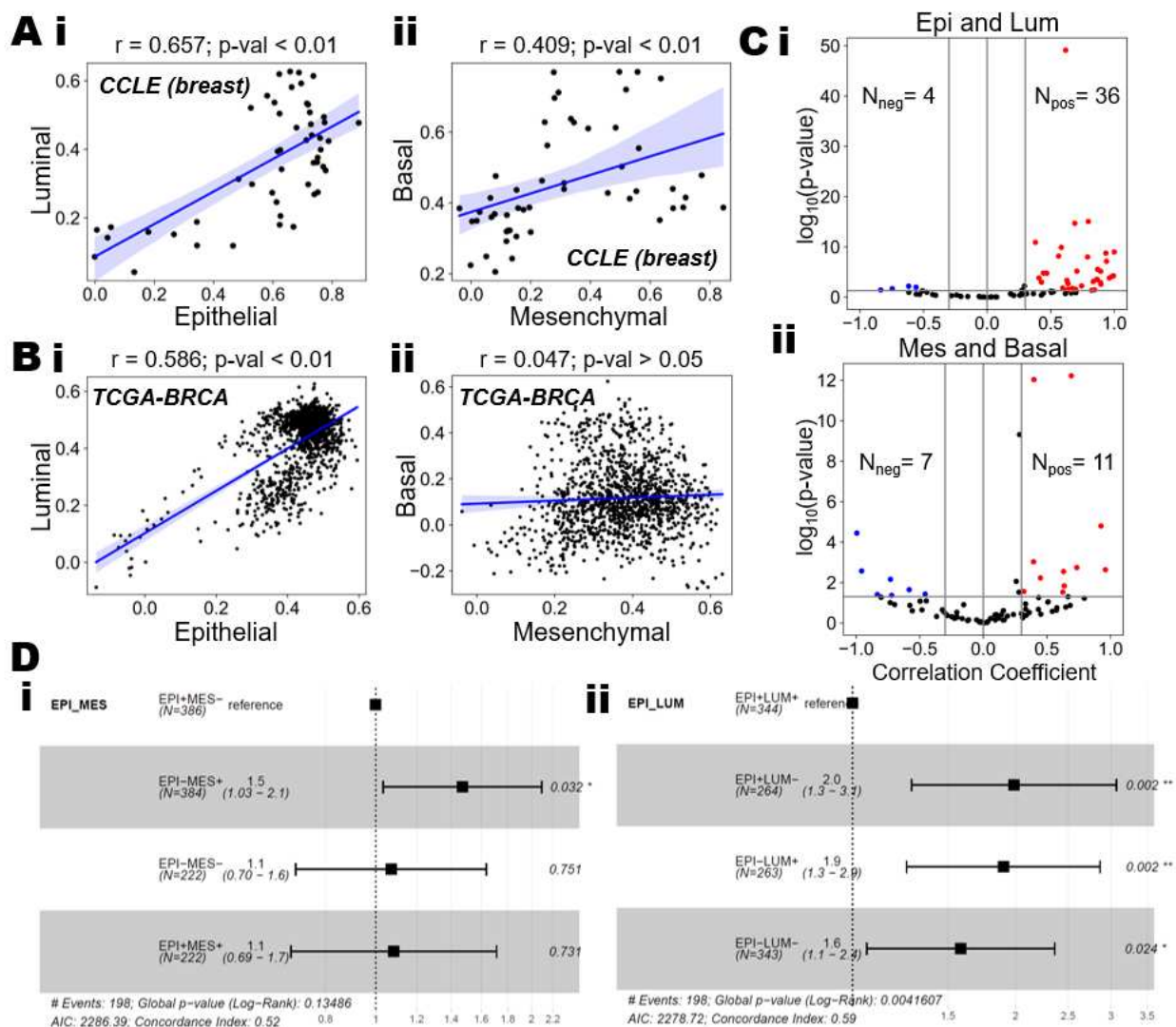


Figure 1: Associations between luminal-epithelial and basal-mesenchymal programs in breast cancer. **A)** Scatterplots showing the correlations between ssGSEA scores of **i)** Epithelial-Luminal and **ii)** Basal-Mesenchymal programs in CCLE breast cancer cell lines. Pearson's correlation coefficients are reported with the corresponding p-values. **B)** Same as A) but for TCGA BRCA patients. **C)** Meta analysis of 80 breast cancer specific datasets showing volcano plots of correlation coefficient values for **i)** Epithelial-Luminal and **ii)** Basal-Mesenchymal programs. Pearson's correlation coefficients and the corresponding p-values are shown. **D)** Forest plots showing survival analysis in TCGA breast cancer patients stratified based on their ssGSEA scores on the **(i)** epithelial-mesenchymal spectrum and **(ii)**

*epithelial-luminal spectrum. The reference cell population is set to high epithelial, low mesenchymal (EPI+ MES-) and high epithelial, high luminal (EPI+, LUM+) states, respectively. * represents a $0.01 < p\text{-value} < 0.05$ while ** represents a $p\text{-value} < 0.01$.*

Having shown that the luminal-epithelial association is stronger than a basal-mesenchymal association in a representative cell line dataset (CCLE) and patient samples (TCGA), we examined whether these trends were more broadly conserved in other breast cancer datasets. We performed a meta-analysis of the associations between these pairs of pathways in 80 transcriptomic datasets (**Table S3**). 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 4 datasets showed a significant negative trend ($r < -0.3$, $p < 0.05$) (**Fig 1C, i**). Such a skew towards the positive side was not present for the correlations between the basal and the mesenchymal signatures (11 vs 4 respectively) (**Fig 1C, ii**). Similarly, such large skews were not found for the luminal-mesenchymal pair or the epithelial-basal pair (**Fig S1E**). These results indicate that the 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 the patients with high mesenchymal (EPI-MES+) tumors showed a poorer prognosis (Hazard ratio = 1.5, $p < 0.05$) (**Fig 1D, 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 (**Fig 1D, 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 effectively to stratify patients into prognostic survival groups.

Basal-like breast cancer samples are associated with a partial EMT signature in bulk transcriptomic datasets.

Given the unclear nature of association of the 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 earlier (Puram et al., 2017) to estimate the pEMT nature of the bulk samples by calculating the ssGSEA scores.

In the CCLE breast cancer cell line set, we observed that the pEMT signature was positively correlated with basal signature ($r = 0.603$, $p < 0.01$) (**Fig 2A**) and, to a lesser extent, with the basal-mesenchymal correlation previously observed ($r = 0.409$, $p < 0.01$) (**Fig 1A, ii**). Similarly, in TCGA samples, the correlation of pEMT and basal signatures was higher than that of a mesenchymal signature with basal ($r = 0.153$, $p < 0.01$ vs. $r = 0.047$, $p > 0.05$ respectively) (**Fig 2B** and **Fig 1B, ii**). We next investigated these trends in a subtype-specific manner and found a numerically higher association of the basal signature with a pEMT state rather than a mesenchymal state (compare **Fig 2B, ii** with **Fig 2B, i**; compare **Fig 2C, ii** with **Fig 2C, i**). In other words, within

the luminal A and B subtypes, as the basal nature increases, a consistent upregulation of a pEMT state, but not necessarily that of mesenchymal state, is observed.

Next, we interrogated whether the association of the pEMT and basal signatures was observed more broadly in breast cancer samples. 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 in only 3 of them ($r < -0.3$, $p < 0.05$) (**Fig 2D**). This skew towards a positive correlation is greater than the one observed for the basal-mesenchymal pair (**Fig 1C, 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 of basal-pEMT set with those of basal-mesenchymal set in a paired manner and found a significant increase in basal-pEMT correlation ($T = 2.9$, $p < 0.01$) (**Fig 2E**). These results collectively suggest that at a bulk transcriptomic level, breast cancer samples show a stronger association of basal-like 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 (Landragin et al., 2022). Moreover, mammary basal epithelial cells have been shown to exhibit a pEMT state, i.e. co-expressing typical epithelial and mesenchymal markers – *ZEB1* and *OVOL2* respectively (Han et al., 2022). 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 invasion (Grasset et al., 2022). Together, these observations underscore the association of the basal breast cancer cell state with a pEMT phenotype.

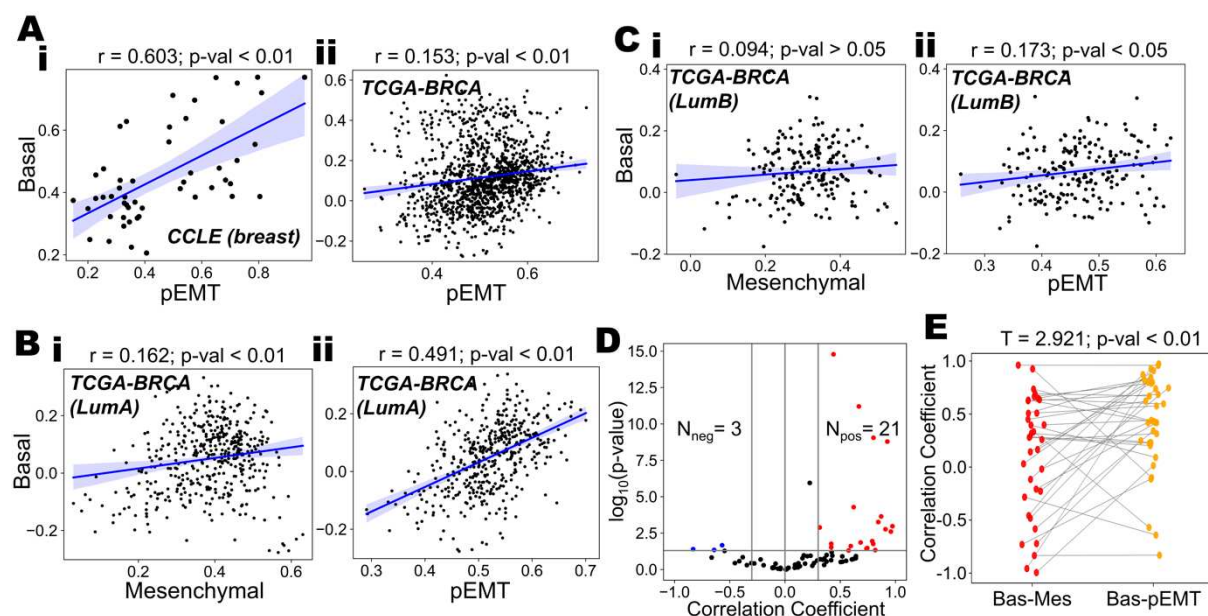


Figure 2: Associations between pEMT and basal programs in breast cancer. **A**) Scatterplots showing the correlations between the ssGSEA scores of pEMT and Basal in **(i)** CCLE breast cancer cell lines and **(ii)** TCGA breast cancer patients. Pearson's correlation coefficients have been reported with the corresponding p-values. **B**) Scatterplots showing the correlations between the ssGSEA scores of **(i)** Mesenchymal-Basal and **(ii)** pEMT-basal programs in Luminal A subtype of TCGA BRCA patients. Pearson's correlation coefficients have been reported with the corresponding p-values. **C**) Scatterplots

showing the correlations between the ssGSEA scores of (i) Mesenchymal-Basal and (ii) pEMT-basal programs in Luminal B subtype of TCGA BRCA patients. Pearson's correlation coefficients have been reported with the corresponding p-values. D) Meta analysis of 80 breast cancer specific datasets showing volcano plots of correlations for pEMT and basal programs. Pearson's correlation coefficients and the corresponding p-values have been used. E) Pair plot showing the comparison of the Pearson correlation coefficients between the Basal-Mesenchymal and the Basal-pEMT programs. Paired T test has been performed and the T statistic and p-value have been reported.

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 analysed 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 (Cope et al., 2014) – were distinctly positioned on the two-dimensional EMT plane (**Fig 3A, i**). The luminal subtype of cell lines clustered towards the (high epithelial, low mesenchymal) part of the plane (**Fig 3A, i**), suggesting a strong association with the epithelial characteristics. Conversely, the Basal B cell lines 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 (**Fig 3A, ii**) while being positively associated with the pEMT score (**Fig S2A**).

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 (**Fig 3A, iii**). The converse was true for the Basal B subtype of breast cancer cells. The Basal A subtype of breast cancer cells which were intermediate in terms of their EMT status exhibited lower levels of methylation in both the epithelial and mesenchymal genes (**Fig 3A, iii**). This indicates that while 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)) (**Fig S2B, i**), a key gene contributing to luminal behaviour (**Fig S2B, 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 (**Fig S2B, iii**) and the ones belonging to basal signature (**Fig S2B, iv**).

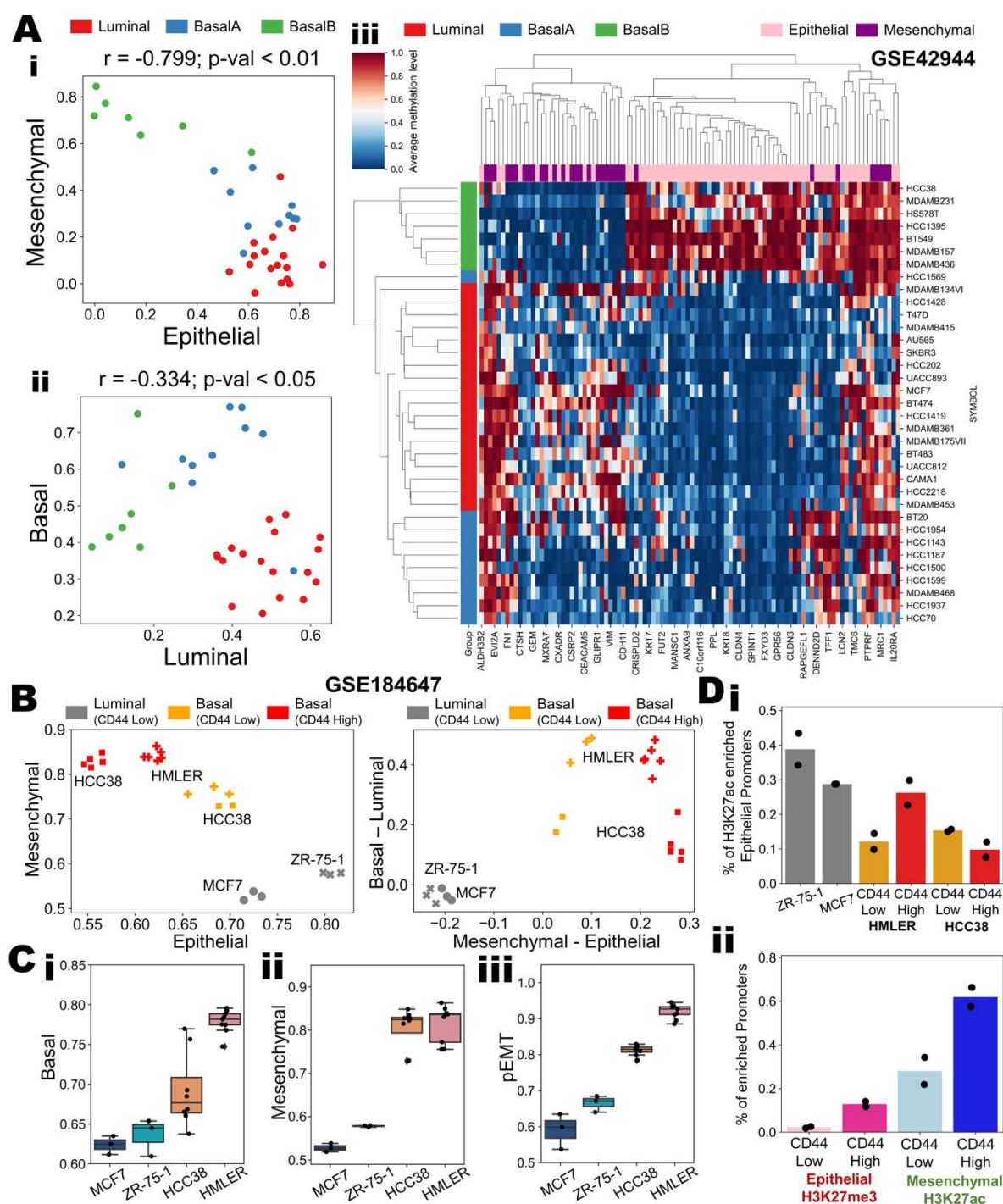


Figure 3: Epigenetic underpinnings of epithelial-mesenchymal phenotypes in luminal and basal breast cancer cell lines. **A)** Scatterplots for ssGSEA scores of **i)** Epithelial-Mesenchymal and **ii)** Luminal-Basal programs in breast cancer cell lines. Pearson's correlation coefficient values and corresponding 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. **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.

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 (Bhatia et al., 2019; Deshmukh et al., 2021; Mani et al., 2008). Thus, we focused on heterogeneity within a cell line in terms of their epithelial-mesenchymal nature and their luminal-basal characteristics. We analysed the RNA-seq data of four cell lines (2 luminal – MCF7 and ZR-75, 2 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 harboured distinct CD44-low and CD44-high subpopulations (San Juan et al., 2022). We observed that luminal cell lines were clustered on the high epithelial, low mesenchymal section of the two-dimensional EMT spectrum, while the CD44-high subpopulation of basal cell lines were clustered towards the low epithelial, high mesenchymal end of it (**Fig 3B, left**). Intriguingly, the CD44 low basal subpopulations were clustered in the medium epithelial, medium mesenchymal region, indicative of a pEMT state (**Fig 3B, 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, while 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 (**Fig 3B, 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 (**Fig 3B, 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 (**Fig 3C**). 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 (**Fig 3C**). 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 (**Fig 3D, i**). Furthermore, in HMLER, the CD44-low subpopulation had consistently lower levels of activation marks (H3K27Ac) on mesenchymal genes as well as lower levels of suppressive marks (H3K27me3) on the epithelial genes (**Fig 3D, 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 (Dravis et al., 2018). Given the established role of NP63 in driving a pEMT program (Dang et al., 2015; Jolly et al., 2017; Westcott et al., 2020) and that of ELF5 in inhibiting EMT (Chakrabarti et al., 2012; Wu et al., 2015; Yao et al., 2015), 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 analysed publicly available spatial transcriptomics datasets of breast tissue sections (Wu et al., 2021) 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 amongst the $n=6$ patients for whom spatial transcriptomics data was available, the ER+ breast cancer patients had distinctively higher levels of luminal nature but downregulated basal scores (**Fig 4A**). On the contrary, the basal breast cancer patients, while showing reduced levels of luminal nature, were quite heterogenous for basal signature expression (**Fig 4A**). 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 towards the high epithelial, low mesenchymal portion of the plane, TNBC patients were highly variable and scattered across the spectrum (**Fig S2C**).

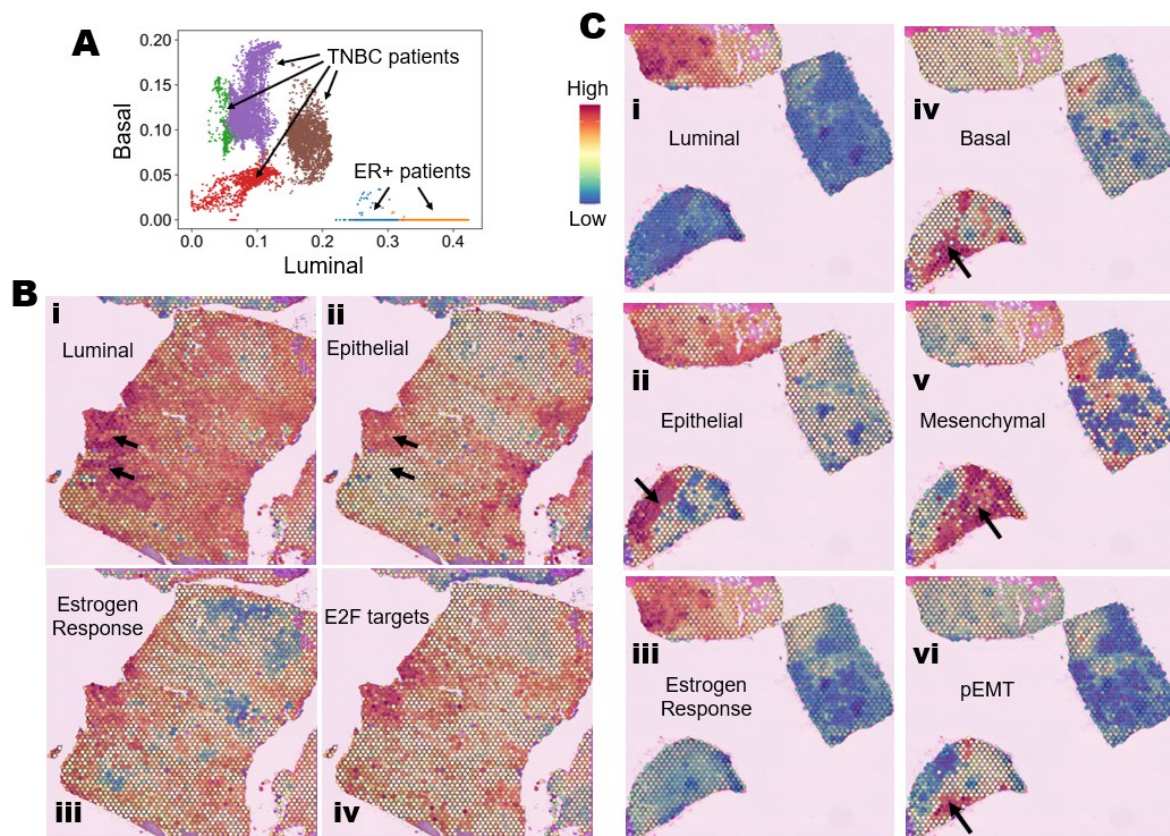


Figure 4: Spatial transcriptomic analysis of estrogen receptor positive and triple negative breast cancer patients. **A)** Scatterplot showing the position of the estrogen receptor positive (ER+) and triple negative breast cancer (TNBC) patients on a two dimensional luminal-basal plane. Each point the 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 coloured 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 coloured 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.

We observed spatial heterogeneity in epithelial and luminal scores in ER+ tumors (**Fig 4B, i-ii**) but the signatures relating to basal, mesenchymal, and pEMT phenotypes were largely absent (**Fig S2D**). 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 **Fig 4B, i-ii**). This observation is particularly important in the context of our survival analysis highlighting that EPI+LUM- or EPI-LUM+ phenotype show significantly worse survival compared to EPI+ LUM+ (**Fig 1D, ii**). We also investigated the spatial activity patterns of ER-driven genes that are generally found to be concordant with luminal and epithelial cell-states (Bouris et al., 2015; Lin et al., 2004). 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) (**Fig 4B, iii-iv**). Such patterns of heterogeneity can dramatically impact sensitivity to anti-ER drugs (Chang, 2012). Thus, the degree of underlying heterogeneity of these molecular programs may limit the efficacy and evolution of resistance in ER+ breast cancer.

Next, we analysed 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 analysed case of luminal disease. We observed that the areas of normal breast tissue were high for luminal, epithelial and ER response gene set scores (**Fig 4C, i-iii**, top left tissue section). However, the tissue slices of the tumor were significantly more enriched in basal scores (**Fig 4C, iv**), with marked variability in terms of their epithelial and mesenchymal natures (**Fig 4C, ii, v**). 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 (**Fig 4C, compare iv with vi**) was higher compared to that between the basal and mesenchymal signature (**Fig 4C, 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. This 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: a) the basal subtype is primarily composed of cells that are pEMT, or b) the basal subtype is comprised 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 analysed single-cell RNA-seq of 32 breast cancer cell lines spanning all the clinical subtypes (GSE173634) (Gambardella et al., 2022). A previous study (Grasset et al., 2022) 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 (**Fig 5A, i**). The extreme peaks were labelled as epithelial and mesenchymal while the intermediate ones were labelled 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 had varied heterogeneity patterns along EMT axis. The TNBC type A cells were found to be either epithelial

or hybrid E/M in nature, while the TNBC type B ones belonged to epithelial, hybrid E/M and mesenchymal states (**Fig 5A, ii-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 (**Fig 5A, ii-iii**). These results indicate that while 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 harbour a more pEMT cell state.

We analysed 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 (**Fig 5B**). 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 (**Fig 5B**).

Next, we probed whether these trends observed in RNA-seq data analysis could be recapitulated experimentally as well. Thus, we took 5 representative cell lines – 2 belonging to luminal subtype (MCF7 and ZR-75-1) and 3 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 through immunofluorescence experiments. It was observed that the luminal cell lines of MCF7 and ZR-75-1 were exclusively high for E-cadherin while have very little to no expression of Vimentin (**Fig 5C, i-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 (co expression of E-cadherin and vimentin in the same cells) (**Fig 5C, iii-iv**). Finally, MDA-MB-231 showed largely mesenchymal cells (low E-cadherin, high Vimentin) with a few cells co-expressing E-cadherin and Vimentin (**Fig 5C, 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 (**Fig 5D, i**). Similar patterns were observed from projection of single-cell RNA-seq data from tumor cells isolated from ER+ and TNBC patients (GSE 176078) (Wu et al., 2021) into this two-dimensional space. (**Fig 5D, ii**). These results collectively indicate that luminal cell lines (and ER+ tumors) were more restricted or homogenous 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 (**Fig 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 (Jolly et al., 2022; Puram et al., 2017). 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.

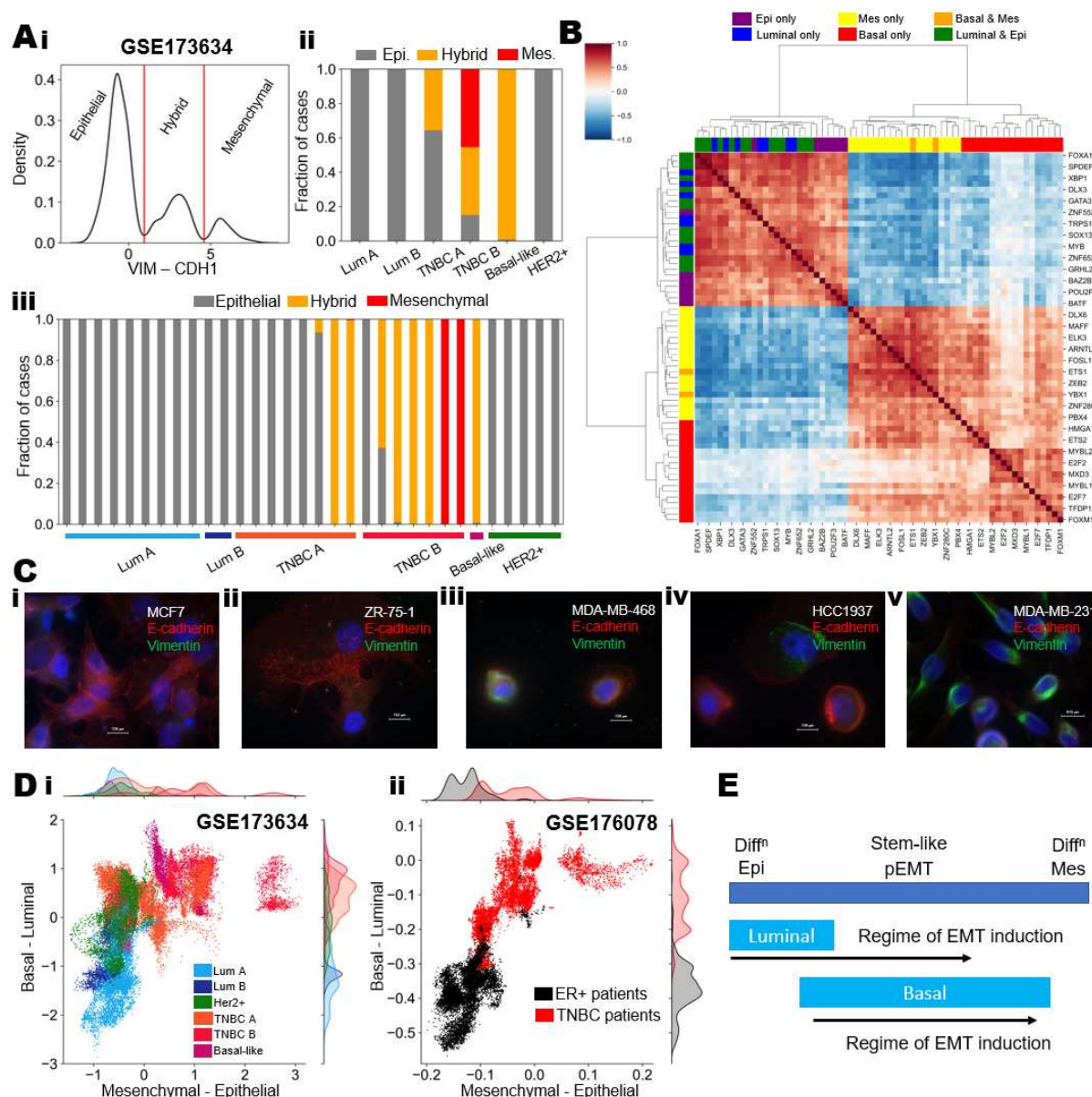


Figure 5: Heterogeneity patterns in single-cell analysis of breast cancer cell lines & patient derived cells. **A) i)** Kernel density estimate of VIM – CDH1 levels for cells belonging to all breast cancer cell lines (GSE173634). Red lines split the distribution into epithelial, mesenchymal and intermediate/hybrid 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) and 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.

Mathematical modelling 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.

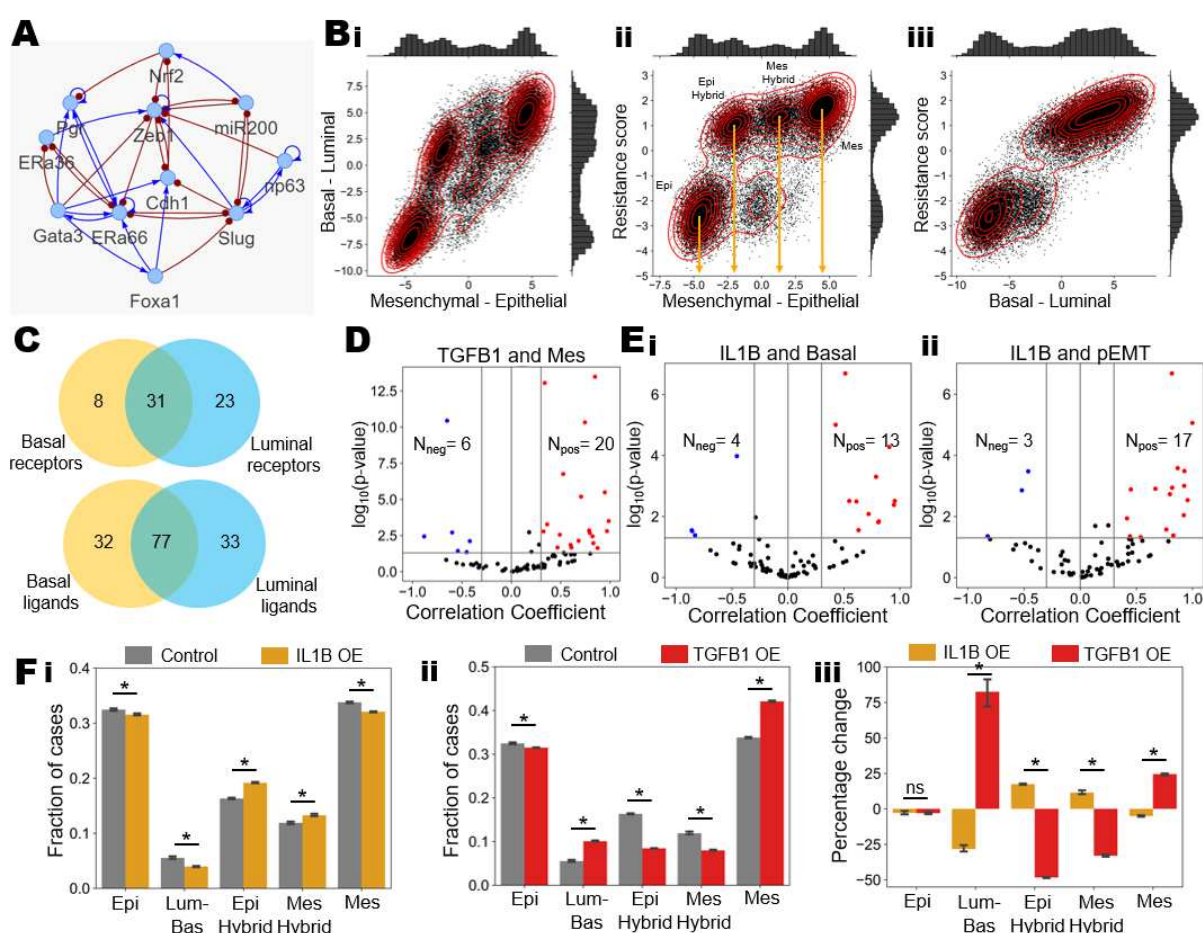


Figure 6: Gene regulatory network analysis for coupled luminal-basal and epithelial-mesenchymal plasticity. **A)** Gene regulatory network showing the representative genes from luminal, basal, epithelial and mesenchymal programs and the associated regulatory links between them. Blue links represent activation, while 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 TGFβ1 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 over-expression (OE) and ii) TGFβ1 OE scenarios. **iii)** Simulation results showing

percentage change in each observed phenotype compared to the control scenario, in TGFB1 and IL1B OE scenarios.

This GRN consists of representative TFs and genes associated with luminal (ER α 66, PGR, GATA3 and FOXA1) (Bernardo et al., 2012, 2010; Eeckhoutte et al., 2007; Kouros-Mehr et al., 2006), basal (SLUG and Δ NP63) (Nekulova et al., 2016; Storci et al., 2008), epithelial (CDH1 and miR-200) and mesenchymal (ZEB1, SLUG) phenotypes (Subbalakshmi et al., 2022), and regulatory interactions among them. We also incorporated two additional players: a) ER α 36 as a marker for anti-ER therapy resistance (Thiebaut et al., 2020) and b) NRF2, as a reported stabilizer of the hybrid E/M state and also associated with drug resistance though its impact on cellular metabolism (Bocci et al., 2019; Kim et al., 2008; Wang et al., 2008) (**Fig 6A**). To understand the emergent dynamics of this GRN, we used RACIPE (Huang et al., 2017), 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: a) luminal score as the sum of normalised steady state values of ER α 66, GATA3, PGR, and FOXA1, b) basal score as the sum of normalised steady state values of Δ NP63 and SLUG, c) epithelial score as the sum of normalised steady state values of CDH1 and miR-200, d) mesenchymal score as the sum of normalised steady state values of ZEB1 and SLUG, and e) resistance score as the difference between normalised steady state levels of ER α 36 and ER α 66. 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 (**Fig S2E**). Thus, our GRN dynamics can recapitulate the multiple hybrid E/M phenotypes that have been well-reported experimentally (Brown et al., 2022; Jolly et al., 2022), as well as a lumino-basal phenotype identified recently in breast cancer (Fan et al., 2020; Mohamed et al., 2023).

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 with the luminal-basal axis (**Fig 6B, i**) as observed for breast cancer cell lines (**Fig 5D, i**) and patient-derived tumor cell data (**Fig 5D, 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 (**Fig 6B, 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.

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 (**Fig 6B, ii**). These model predictions are reminiscent of prior experimental observations that partial and/or full EMT can drive resistance to tamoxifen and *vice versa* (Kastrati et al., 2020; Sahoo et al., 2021a). 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 (**Fig 6B, iii**). We found that the two hybrid E/M clusters had higher levels of NRF2 (**Fig S2F**), consistent with the reported literature (Bocci et al., 2019; Kim et al.,

2018; Pasani et al., 2020). 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 analysed a single-cell patient atlas of 7 ER+ breast cancer patients and 5 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 (Dimitrov et al., 2022) 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 (Balko et al., 2012; Semba et al., 2023). We observed substantial overlap between the ligand and receptors from the basal and luminal cell types (**Fig 6C**), suggesting potential overlaps in signalling 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- β 1 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. We found that TGF β 1 was amongst the top genes that correlated positively with a more mesenchymal phenotype (**Fig 6D**). On the other hand, we found IL-1 β to be one of the top genes to be correlated with a basal phenotype (**Fig 6E, i**) as well as pEMT phenotype (**Fig 6E, ii**). Further, we interrogated whether our GRN could reproduce the plasticity patterns that can be driven by TGF- β 1 or IL-1 β . To do this we extended our GRN (**Table S4**) to include IL-1 β as well as TGF β and performed simulations to over express these nodes *in silico*. Our model predicted that IL-1 β 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 (**Fig 6F, i, iii**). On the other hand, TGF- β 1 over expression caused a marked increase in the frequency of mesenchymal phenotype, and a simultaneous reduction in epithelial, epithelial-hybrid and mesenchymal-hybrid states (**Fig 6F, ii-iii**). There was also a significant increase in the proportion of the luminal basal hybrid phenotype (**Fig 6F, ii-iii**). Although both IL-1 β and TGF- β 1 cause a net decrease in the epithelial phenotype and increase the mesenchymal nature, their impact is quite distinct: IL-1 β 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-1 β *in vitro* and *in vivo*: a) IL-1 β treatment can induce two stabilizers of hybrid E/M phenotype NRF2 and SLUG (González-Carnicero et al., 2023; Li et al., 2020), b) IL-1 β treatment of MCF7 luminal breast cancer cells can induce Δ NP63 and mediate subsequent therapy resistance (Mendoza-Rodríguez et al., 2019), and c) IL-1 β treatment can prevent differentiation of metastatic-initiating cells to highly proliferative epithelial cells, inhibiting overt metastatic growth (Castaño et al., 2018). Together, our analysis suggests IL-1 β to be a potent target to prevent luminal-to-basal lineage plasticity.

Overall, our simulation results show that the core GRN modelled 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 signalling 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 (van Boxtel et al., 2017). Not surprisingly, cancer cells are heterogenous 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 (Brown et al., 2022; Chauhan et al., 2021; Pastushenko et al., 2018; Rambow et al., 2019; Yu et al., 2013). Single-cell approaches have been instrumental in characterizing such heterogeneity, but usually along one axis (Deshmukh et al., 2021; Jia et al., 2019; Karacosta et al., 2019). Analysing 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 (Beltran et al., 2019; Kern et al., 2022; Mohamed et al., 2023; Padua et al., 2018). Often, these axes have been assumed to be synonymous or largely overlapping, but many questions remained unanswered: a) how do luminal and basal features map to epithelial, hybrid E/M and mesenchymal phenotypes? ; b) which breast cancer subtype (luminal/basal) have higher heterogeneity in terms of EMT? ; and c) what feedback loops connect these two plasticity axes?

These analyses of breast cancer cell lines and primary tumors have unravelled that while luminal breast cancer samples are predominantly epithelial and quite homogenous, 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 (Brown et al., 2022; Jolly et al., 2022; Vipparthi et al., 2022). Thus, their co-existence can possibly explain the aggressive behaviour and limited therapeutic response of basal-like breast cancers (Milioli et al., 2017). Further, given the relatively higher plasticity of hybrid E/M phenotypes as compared to ‘fully epithelial’ or ‘fully mesenchymal’ phenotypes (Hari et al., 2022; Kröger et al., 2019), 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 (Ali et al., 2015) and tamoxifen resistance (Fan et al., 2020) in basal-like tumors, given that both these traits have been previously linked to hybrid E/M cells (Sahoo et al., 2021a, 2021b). Even among luminal tumors, those expressing low levels of ER can display higher basal-like phenotypes (Mohamed et al., 2023), endorsing previous results that *ESR1* expression closely associates with luminal breast cancer cells and that silencing ER can drive EMT (Al Saleh et al., 2011). Because EMT and tamoxifen resistance can both drive each other (Hiscox et al., 2006; Sahoo et al., 2021a), 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 (Jindal et al., 2023; Mu et al., 2017) and small cell lung cancer (Groves et al., 2023). These correspondences indicate possible generalisability 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 (Kern et al., 2022). Similarly, basal-like tumorigenesis involves luminal-to-basal reprogramming with gain in stemness (Christin et al., 2020; Landragin et al., 2022). Further, luminal-basal hybrid cells can express NP63 (basal marker) while maintaining functional levels of ER-alpha (associated with luminal phenotype) (Padua et al., 2018). 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.

METHODS

ssGSEA scores for bulk transcriptomics

We used previously published signatures for luminal and basal breast cancer (Nair et al., 2022), for epithelial and mesenchymal state (Tan et al., 2014) and for pEMT state (Puram et al., 2017). Hallmark signatures (estrogen response, E2F target genes) were taken from MSigDB database (Liberzon et al., 2011). ssGSEA scores were calculated for bulk transcriptomic samples using the gseapy python package (Fang et al., 2023) 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 was considered. The beta values for CpG islands vary between 0 (unmethylated) and 1 (fully methylated). The heatmap included only previously identified epithelial and mesenchymal genes (Tan et al., 2014) and cell lines included in CCLE breast cancer cohort to facilitate direct comparison between the RNA-Seq and methylation data. 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) (Wu et al., 2021) were reanalysed for assessing the activity of various biological pathways. Count matrices were first imputed by MAGIC algorithm (van Dijk et al., 2018) and activity scores were calculated on imputed values using AUCell (Aibar et al., 2017). Pre-processing of spatial data and images was done as per the Seurat pipeline (Hao et al., 2021).

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.

Dual immunofluorescence

Cells (1×10^4) were seeded on poly-L-lysine coated coverslips. Immunofluorescence was performed as reported previously (Nair et al., 2016) 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 (van Dijk et al., 2018). Activity scores were calculated on imputed values using AUCell (Aibar et al., 2017). Lists of top cell-cell communication receptor ligand pairs were estimated using the LIANA package (Dimitrov et al., 2022). 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) (Huang et al., 2017). 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 ER α 36 and ER α 66. 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 (no 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.

Conflict of Interest

The authors declare no conflicts of interest.

Author contributions

Conceptualized and designed research: MKJ

Supervised research: CLC, JSP, JAS, MKJ

Performed research: SS, SR, MGN, MP, BPSN, SM, CMN, ADM, HS, AGN

Interpreted data: SS, SR, MGN, MP, BPSN, SM, CMN, ADM, CLC, JSP, JAS, MKJ

Funding acquisition: JSP, JAS, MKJ

Manuscript writing/editing: SS (prepared first draft), JAS, MKJ (edited with inputs from all authors)

Funding

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 MKJ (SB/S2/RJN-049/2018). SS is supported by PMRF (Prime Ministers Research Fellowship) awarded by DST, Government of India. JASs is supported by NCI 1R01CA233585-04.

References

- Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083–6. <https://doi.org/10.1038/nmeth.4463>.
- Ali HR, Glont SE, Blows FM, Provenzano E, Dawson SJ, Liu B, et al. PD-L1 protein expression in breast cancer is rare, enriched in basal-like tumours and associated with infiltrating lymphocytes. *Ann Oncol* 2015;26:1488–93. <https://doi.org/10.1093/ANNONC/MDV192>.
- Balko JM, Miller TW, Morrison MM, Hutchinson K, Young C, Rinehart C, et al. The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. *Proc Natl Acad Sci U S A* 2012;109:221–6. https://doi.org/10.1073/PNAS.1115802109/SUPPL_FILE/PNAS.201115802SI.PDF.
- Beltran H, Hruszkewycz A, Scher HI, Hildesheim J, Isaacs J, Yu EY, et al. The Role of Lineage Plasticity in Prostate Cancer Therapy Resistance. *Clin Cancer Res* 2019;25:6916–24. <https://doi.org/10.1158/1078-0432.CCR-19-1423>.
- Bernardo GM, Bebek G, Ginther CL, Sizemore ST, Lozada KL, Miedler JD, et al. FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene* 2013 32:5 2012;32:554–63. <https://doi.org/10.1038/onc.2012.62>.
- Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, et al. FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis. *Development* 2010;137:2045–54. <https://doi.org/10.1242/DEV.043299>.
- Bhatia S, Monkman J, Blick T, Pinto C, Waltham A, Nagaraj SH, et al. Interrogation of Phenotypic Plasticity between Epithelial and Mesenchymal States in Breast Cancer. *J Clin Med* 2019;8:893. <https://doi.org/10.3390/jcm8060893>.
- Bhattacharya S, Mohanty A, Achuthan S, Kotnala S, Jolly MK, Kulkarni P, et al. Group Behavior and Emergence of Cancer Drug Resistance. *Trends Cancer* 2021;7:323–34. <https://doi.org/10.1016/J.TRECAN.2021.01.009>.
- Bierie B, Pierce SE, Kroeger C, Stover DG, Pattabiraman DR, Thiru P, et al. Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proceedings of the National Academy of Sciences* 2017;114:E2337-2346. <https://doi.org/10.1073/pnas.1618298114>.
- Bocci F, Tripathi SC, Vilchez MSA, George JT, Casabar J, Wong P, et al. NRF2 activates a partial Epithelial-Mesenchymal Transition and is maximally present in a hybrid Epithelial/Mesenchymal phenotype. *Integrative Biology* 2019;11:251–63. <https://doi.org/10.1101/390237>.
- Bouris P, Skandalis SS, Piperigkou Z, Afratis N, Karamanou K, Aletras AJ, et al. Estrogen receptor alpha mediates epithelial to mesenchymal transition, expression of specific matrix effectors and functional properties of breast cancer cells. *Matrix Biol* 2015;43:42–60. <https://doi.org/10.1016/J.MATBIO.2015.02.008>.
- van Boxtel C, van Heerden JH, Nordholt N, Schmidt P, Bruggeman FJ. Taking chances and making mistakes : non-genetic phenotypic heterogeneity and its consequences for surviving in dynamic environments. *Journal of Royal Society Interface* 2017;14:20170141. <https://doi.org/10.1098/rsif.2017.0141>.

Brown MS, Abdollahi B, Wilkins OM, Lu H, Chakraborty P, Ognjenovic NB, et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Sci Adv* 2022;8:eabj8002. <https://doi.org/10.1101/2021.03.17.434993>.

Burkhardt DB, San Juan BP, Lock JG, Krishnaswamy S, Chaffer CL. Mapping Phenotypic Plasticity upon the Cancer Cell State Landscape Using Manifold Learning. *Cancer Discov* 2022;12:1847–59. <https://doi.org/10.1158/2159-8290.CD-21-0282/705079/P/MAPPING-PHENOTYPIC-PLASTICITY-UPON-THE-CANCER-CELL>.

Castaño Z, San Juan BP, Spiegel A, Pant A, DeCristo MJ, Laszewski T, et al. IL-1 β inflammatory response driven by primary breast cancer prevents metastasis-initiating cell colonization. *Nat Cell Biol* 2018;20:1084. <https://doi.org/10.1038/S41556-018-0173-5>.

Chakrabarti R, Hwang J, Andres Blanco M, Wei Y, Lukačičin M, Romano R-A, et al. Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. *Nat Cell Biol* 2012;14:1212–22. <https://doi.org/10.1038/ncb2607>.

Chang M. Tamoxifen resistance in breast cancer. *Biomol Ther (Seoul)* 2012;20:256–67. <https://doi.org/10.4062/biomolther.2012.20.3.256>.

Chauhan L, Ram U, Hari K, Jolly MK. Topological signatures in regulatory network enable phenotypic heterogeneity in small cell lung cancer. *Elife* 2021;10:e64522. <https://doi.org/https://doi.org/10.7554/eLife.64522>.

Christin JR, Wang C, Chung CY, Liu Y, Dravis C, Tang W, et al. Stem Cell Determinant SOX9 Promotes Lineage Plasticity and Progression in Basal-like Breast Cancer. *Cell Rep* 2020. <https://doi.org/10.1016/j.celrep.2020.107742>.

Cope LM, Fackler MJ, Lopez-Bujanda Z, Wolff AC, Visvanathan K, Gray JW, et al. Do Breast Cancer Cell Lines Provide a Relevant Model of the Patient Tumor Methylome? *PLoS One* 2014;9:e105545. <https://doi.org/10.1371/journal.pone.0105545>.

Dang TT, Esparza MA, Maine EA, Westcott JM, Pearson GW. Δ Np63 α promotes breast cancer cell motility through the selective activation of components of the epithelial-to-mesenchymal transition program. *Cancer Res* 2015;75:3925–35. <https://doi.org/10.1158/0008-5472.CAN-14-3363>.

Deshmukh AP, Vasaikar S V., Tomczak K, Tripathi S, Den Hollander P, Arslan E, et al. Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. *Proc Natl Acad Sci U S A* 2021;118:e2102050118. <https://doi.org/10.73/pnas.2102050118>.

van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* 2018;174:716–729.e27. <https://doi.org/10.1016/j.cell.2018.05.061>.

Dimitrov D, Türei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. *Nature Communications* 2022 13:1 2022;13:1–13. <https://doi.org/10.1038/s41467-022-30755-0>.

Dravis C, Chung CY, Lytle NK, Herrera-Valdez J, Luna G, Trejo CL, et al. Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell* 2018;34:466–482.e6. <https://doi.org/10.1016/J.CCELL.2018.08.001>.

Eeckhoutte J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res* 2007;67:6477–83. <https://doi.org/10.1158/0008-5472.CAN-07-0746>.

Fan M, Chen J, Gao J, Xue W, Wang Y, Li W, et al. Triggering a switch from basal- to luminal-like breast cancer subtype by the small-molecule diptoindonesin G via induction of GABARAPL1. *Cell Death & Disease* 2020 11:8 2020;11:1–13. <https://doi.org/10.1038/s41419-020-02878-z>.

Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 2023;39. <https://doi.org/10.1093/bioinformatics/btac757>.

Gambardella G, Viscido G, Tumaini B, Isacchi A, Bosotti R, di Bernardo D. A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nat Commun* 2022;13. <https://doi.org/10.1038/S41467-022-29358-6>.

González-Carnicero Z, Hernanz R, Martínez-Casales M, Barrús MT, Martín Á, Alonso MJ. Regulation by Nrf2 of IL-1 β -induced inflammatory and oxidative response in VSMC and its relationship with TLR4. *Front Pharmacol* 2023;14:1058488. <https://doi.org/10.3389/FPHAR.2023.1058488/BIBTEX>.

Grasset EM, Dunworth M, Sharma G, Loth M, Tandurella J, Cimino-Mathews A, et al. Triple-negative breast cancer metastasis involves complex epithelial-mesenchymal transition dynamics and requires vimentin. *Sci Transl Med* 2022;14:7571. https://doi.org/10.1126/SCITRANSLMED.ABN7571/SUPPL_FILE/SCITRANSLMED.ABN7571_MДАР_REPRODUCIBILITY_CHECKLIST.PDF.

Grosse-Wilde A, Fouquier d' Herouei A, McIntosh E, Ertaylan G, Skupin A, Kuestner RE, et al. Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One* 2015;10:e0126522. <https://doi.org/10.1371/journal.pone.0126522>.

Groves SM, Panchy N, Tyson DR, Harris LA, Quaranta V, Hong T. Involvement of Epithelial-Mesenchymal Transition Genes in Small Cell Lung Cancer Phenotypic Plasticity. *Cancers (Basel)* 2023;15. <https://doi.org/10.3390/CANCERS15051477>.

Gudjonsson T, Adriance MC, Sternlicht MD, Petersen OW, Bissell MJ. Myoepithelial Cells: Their Origin and Function in Breast Morphogenesis and Neoplasia. *J Mammary Gland Biol Neoplasia* 2005;10:261–72. <https://doi.org/10.1007/s10911-005-9586-4>.

Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Research* 2005. <https://doi.org/10.1186/bcr1041>.

Han Y, Villarreal-Ponce A, Gutierrez G, Nguyen Q, Sun P, Wu T, et al. Coordinate control of basal epithelial cell fate and stem cell maintenance by core EMT transcription factor Zeb1. *Cell Rep* 2022;38:110240. <https://doi.org/10.1016/J.CELREP.2021.110240>.

Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–3587.e29. <https://doi.org/10.1016/J.CELL.2021.04.048/ATTACHMENT/1E5EB5C1-59EE-4B2B-8BFA-14B48A54FF8F/MMC3.XLSX>.

Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK. Landscape of epithelial mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. *Elife* 2022;11:e76535. <https://doi.org/10.7554/elife.76535>.

- Hiscox S, Jiang WG, Obermeier K, Taylor K, Morgan L, Burmi R, et al. Tamoxifen resistance in MCF7 cells promotes EMT-like behaviour and involves modulation of β -catenin phosphorylation. *Int J Cancer* 2006;118:290–301. <https://doi.org/10.1002/ijc.21355>.
- Hong D, Fritz AJ, Zaidi SK, van Wijnen AJ, Nickerson JA, Imbalzano AN, et al. Epithelial-to-mesenchymal transition and cancer stem cells contribute to breast cancer heterogeneity. *J Cell Physiol* 2018;233:9136–44. <https://doi.org/10.1002/JCP.26847>.
- Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, Onuchic J. Interrogating the Topological Robustness of Gene Regulatory Circuits. *PLoS Comput Biol* 2017;13:e1005456. <https://doi.org/http://dx.doi.org/10.1101/084962>.
- Ignatiadis M, Sotiriou C. Luminal breast cancer: from biology to treatment. *Nature Reviews Clinical Oncology* 2013 10:9 2013;10:494–506. <https://doi.org/10.1038/nrclinonc.2013.124>.
- Jain P, Bhatia S, Thompson EW, Jolly MK. Population Dynamics of Epithelial-Mesenchymal Heterogeneity in Cancer Cells. *Biomolecules* 2022;12:348. <https://doi.org/10.3390/biom12030348>.
- Jia D, Lu M, Jung KH, Park JH, Yu L, Onuchic JN, et al. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. *Proc Natl Acad Sci U S A* 2019;116:3909–18. <https://doi.org/10.1073/pnas.1816391116>.
- Jia D, Park JH, Kaur H, Jung KH, Yang S, Tripathi S, et al. Towards decoding the coupled decision-making of metabolism and epithelial-mesenchymal transition in cancer. *Br J Cancer* 2021;124:1902–11.
- Jindal R, Nanda A, Pillai M, Ware KE, Singh D, Sehgal M, et al. Emergent dynamics of underlying regulatory network links EMT and androgen receptor-dependent resistance in prostate cancer. *Comput Struct Biotechnol J* 2023;21:1498–509. <https://doi.org/10.1016/J.CSBJ.2023.01.031>.
- Jolly MK, Boareto M, Debeb BG, Aceto N, Farach-Carson MC, Woodward WA, et al. Inflammatory breast cancer: a model for investigating cluster-based dissemination. *Npj Breast Cancer* 2017 3:1 2017;3:1–8. <https://doi.org/10.1038/s41523-017-0023-9>.
- Jolly MK, Murphy RJ, Bhatia S, Whitfield HJ, Redfern A, Davis MJ, et al. Measuring and Modelling the Epithelial- Mesenchymal Hybrid State in Cancer: Clinical Implications. *Cells Tissues Organs* 2022;211:110–33. <https://doi.org/10.1159/000515289>.
- Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping Lung Cancer Epithelial-Mesenchymal Transition States and Trajectories with Single-Cell Resolution. *Nat Commun* 2019;10:5587. <https://doi.org/10.1101/570341>.
- Kastrati I, Joosten SEP, Semina SE, Alejo LH, Brovkovich SD, Stender JD, et al. The NF κ B pathway promotes tamoxifen tolerance and disease recurrence in estrogen receptor positive breast cancers. *Mol Cancer Res* 2020;18:1018. <https://doi.org/10.1158/1541-7786.MCR-19-1082>.
- Kern JG, Tilston-Lunel AM, Federico A, Ning B, Mueller A, Peppler GB, et al. Inactivation of LATS1/2 drives luminal-basal plasticity to initiate basal-like mammary carcinomas. *Nature Communications* 2022 13:1 2022;13:1–13. <https://doi.org/10.1038/s41467-022-34864-8>.
- Kim D, Choi B, Ryoo I, Kwak M-K. High NRF2 level mediates cancer stem cell-like properties of aldehyde dehydrogenase (ALDH)-high ovarian cancer cells: inhibitory role of all-trans retinoic acid in ALDH/NRF2 signaling. *Cell Death Dis* 2018;9:896. <https://doi.org/10.1038/s41419-018-0903-4>.

Kim SK, Yang JW, Kim MR, Roh SH, Kim HG, Lee KY, et al. Increased expression of Nrf2/ARE-dependent anti-oxidant proteins in tamoxifen-resistant breast cancer cells. *Free Radic Biol Med* 2008. <https://doi.org/10.1016/j.freeradbiomed.2008.05.011>.

Kouros-Mehr H, Slorach EM, Sternlicht MD, Werb Z. GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. *Cell* 2006;127:1041–55. <https://doi.org/10.1016/J.CELL.2006.09.048>.

Kröger C, Afeyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, et al. Acquisition of a hybrid E / M state is essential for tumorigenicity of basal breast cancer cells. *Proc Natl Acad Sci U S A* 2019;116:7353–62. <https://doi.org/10.1073/pnas.1812876116>.

Landragin C, Saichi M, Prompsy P, Durand A, Mesple J, Trousset A, et al. Luminal progenitors undergo partial epithelial-to-mesenchymal transition at the onset of basal-like breast tumorigenesis. *BioRxiv* 2022:2022.06.08.494710. <https://doi.org/10.1101/2022.06.08.494710>.

Lesniak D, Sabri S, Xu Y, Graham K, Bhatnagar P, Suresh M, et al. Spontaneous Epithelial-Mesenchymal Transition and Resistance to HER-2-Targeted Therapies in HER-2-Positive Luminal Breast Cancer. *PLoS One* 2013. <https://doi.org/10.1371/journal.pone.0071987>.

Li R, Ong SL, Tran LM, Jing Z, Liu B, Park SJ, et al. Chronic IL-1 β -induced inflammation regulates epithelial-to-mesenchymal transition memory phenotypes via epigenetic modifications in non-small cell lung cancer. *Sci Rep* 2020;10. <https://doi.org/10.1038/S41598-019-57285-Y>.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40. <https://doi.org/10.1093/bioinformatics/btr260>.

Lin CY, Ström A, Vega VB, Kong SL, Yeo AL, Thomsen JS, et al. Discovery of estrogen receptor alpha target genes and response elements in breast tumor cells. *Genome Biol* 2004;5:1–18. <https://doi.org/10.1186/GB-2004-5-9-R66/FIGURES/7>.

Lindström LS, Yau C, Czene K, Thompson CK, Hoadley KA, Van't Veer LJ, et al. Intratumor heterogeneity of the estrogen receptor and the long-term risk of fatal Breast cancer. *J Natl Cancer Inst* 2018. <https://doi.org/10.1093/jnci/djx270>.

Łukasiewicz S, Czelewska M, Forma A, Baj J, Sitarz R, Stanisławek A. Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers (Basel)* 2021;13:4287. <https://doi.org/10.3390/cancers13174287>.

Mani SA, Guo W, Liao M-J, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* 2008;133:704–15. <https://doi.org/10.1016/j.cell.2008.03.027>.

Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* 2020;37:471–84. <https://doi.org/10.1016/j.ccell.2020.03.007>.

Mendoza-Rodríguez MG, Ayala-Sumano JT, García-Morales L, Zamudio-Meza H, Pérez-Yepez EA, Meza I. IL-1 β Inflammatory Cytokine-Induced TP63 Isoform Δ NP63 α Signaling Cascade Contributes to Cisplatin Resistance in Human Breast Cancer Cells. *Int J Mol Sci* 2019;20. <https://doi.org/10.3390/IJMS20020270>.

Milioli HH, Tishchenko I, Riveros C, Berretta R, Moscato P. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC Med Genomics* 2017;10:1–17. <https://doi.org/10.1186/S12920-017-0250-9/TABLES/3>.

Mohamed GA, Mahmood S, Ognjenovic NB, Lee MK, Wilkins OM, Christensen BC, et al. Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits. *Breast Cancer Research* 2023;25:1–21. <https://doi.org/10.1186/S13058-023-01621-8/FIGURES/5>.

Mu P, Zhang Z, Benelli M, Karthaus WR, Hoover E, Chen C-C, et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* 2017;355:84–8. <https://doi.org/10.1126/science.aah4307>.

Nair MG, Apoorva D, Chandrakala M, Vp S, Ce A. 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. *BioRxiv* 2022:505398. <https://doi.org/10.1101/2022.09.05.505398>.

Nair MG, Desai K, Prabhu JS, Hari PS, Remacle J, Sridhar TS. $\beta 3$ integrin promotes chemoresistance to epirubicin in MDA-MB-231 through repression of the pro-apoptotic protein, BAD. *Exp Cell Res* 2016;346:137–45. <https://doi.org/10.1016/j.yexcr.2016.05.015>.

Nekulova M, Holcakova J, Gu X, Hrabal V, Galtsidis S, Orzol P, et al. Δ Np63 α expression induces loss of cell adhesion in triple-negative breast cancer cells. *BMC Cancer* 2016;16:1–12. <https://doi.org/10.1186/S12885-016-2808-X/FIGURES/6>.

Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, et al. Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* 2018;9:2028. <https://doi.org/10.1038/s41467-018-04334-1>.

Padua MB, Bhat-Nakshatri P, Anjanappa M, Prasad MS, Hao Y, Rao X, et al. Dependence receptor UNC5A restricts luminal to basal breast cancer plasticity and metastasis. *Breast Cancer Res* 2018;20. <https://doi.org/10.1186/S13058-018-0963-5>.

Pasani S, Sahoo S, Jolly MK. Hybrid E/M Phenotype(s) and Stemness: A Mechanistic Connection Embedded in Network Topology. *J Clin Med* 2020;10:60. <https://doi.org/10.3390/jcm10010060>.

Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature* 2018;556:463–8. <https://doi.org/10.1038/s41586-018-0040-3>.

Phillips S, Kuperwasser C. SLUG: Critical regulator of epithelial cell identity in breast development and cancer. *Cell Adh Migr* 2014;8:578–87. <https://doi.org/10.4161/19336918.2014.972740>.

Pillai M, Hojel E, Jolly MK, Goyal Y. Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. *Nature Computational Science* 2023 3:4 2023;3:301–13. <https://doi.org/10.1038/s43588-023-00427-0>.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, et al. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research* 2010. <https://doi.org/10.1186/bcr2635>.

Puram S V., Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* 2017;171:1611–24. <https://doi.org/10.1016/j.cell.2017.10.044>.

Qin S, Jiang J, Lu Y, Nice EC, Huang C, Zhang J, et al. Emerging role of tumor cell plasticity in modifying therapeutic response. *Signal Transduction and Targeted Therapy* 2020 5:1 2020;5:1–36. <https://doi.org/10.1038/s41392-020-00313-5>.

Rambow F, Marine JC, Goding CR. Melanoma plasticity and phenotypic diversity: Therapeutic barriers and opportunities. *Genes Dev* 2019;33:1295–318. <https://doi.org/10.1101/gad.329771.119>.

Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. *NAR Cancer* 2021a;3:zcab027. <https://doi.org/10.1101/2021.03.14.435359>.

Sahoo S, Nayak SP, Hari K, Purkait P, Mandal S, Kishore A, et al. Immunosuppressive traits of the hybrid epithelial/mesenchymal phenotype. *Front Immunol* 2021b;12:797261. <https://doi.org/10.3389/fimmu.2021.797261>.

Al Saleh S, Al Mulla F, Luqmani YA. Estrogen Receptor Silencing Induces Epithelial to Mesenchymal Transition in Human Breast Cancer Cells. *PLoS One* 2011;6:e20610. <https://doi.org/10.1371/journal.pone.0020610>.

San Juan BP, Hediye-Zadeh S, Rangel L, Milioli HH, Rodriguez V, Bunkum A, et al. Targeting phenotypic plasticity prevents metastasis and the development of chemotherapy-resistant disease. *MedRxiv* 2022:2022.03.21.22269988. <https://doi.org/10.1101/2022.03.21.22269988>.

Semba R, Horimoto Y, Sakata-Matsuzawa M, Ishizuka Y, Denda-Nagai K, Fujihira H, et al. Possible correlation of apical localization of MUC1 glycoprotein with luminal A-like status of breast cancer. *Scientific Reports* 2023 13:1 2023;13:1–8. <https://doi.org/10.1038/s41598-023-32579-4>.

Skibinski A, Kuperwasser C. The origin of breast tumor heterogeneity. *Oncogene* 2015;34:5309–16. <https://doi.org/10.1038/onc.2014.475>.

Storci G, Sansone P, Trere D, Tavorali S, Taffurelli M, Ceccarelli C, et al. The basal-like breast carcinoma phenotype is regulated by SLUG gene expression. *J Pathol* 2008;214:25–37. <https://doi.org/10.1002/PATH.2254>.

Subbalakshmi AR, Sahoo S, Biswas K, Jolly MK. A computational systems biology approach identifies SLUG as a mediator of partial Epithelial-Mesenchymal Transition (EMT). *Cells Tissues Organs* 2022;211:105–18. <https://doi.org/10.1159/000512520>.

Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RY-J, et al. Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO Mol Med* 2014;6:1279–93. <https://doi.org/10.15252/emmm.201404208>.

Thiebaut C, Konan HP, Guerquin MJ, Chesnel A, Livera G, Le Romancer M, et al. The Role of ERα36 in Development and Tumor Malignancy. *Int J Mol Sci* 2020;21:1–23. <https://doi.org/10.3390/IJMS21114116>.

Turner KM, Yeo SK, Holm TM, Shaughnessy E, Guan J-L. Heterogeneity within molecular subtypes of breast cancer. *Am J Physiol Cell Physiol* 2021;321:C343–54. <https://doi.org/10.1152/ajpcell.00109.2021>.

Vipparthi K, Hari K, Chakraborty P, Ghosh S, Patel AK, Ghosh A, et al. Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. *IScience* 2022;25:104317. <https://doi.org/10.1016/j.isci.2022.104317>.

Wahl GM, Spike BT. Cell state plasticity, stem cells, EMT, and the generation of intra-tumoral heterogeneity. *Npj Breast Cancer* 2017 3:1 2017;3:1–13. <https://doi.org/10.1038/s41523-017-0012-z>.

Wang X-J, Sun Z, Villeneuve NF, Zhang S, Zhao F, Li Y, et al. Nrf2 enhances resistance of cancer cells to chemotherapeutic drugs, the dark side of Nrf2. *Carcinogenesis* 2008;29:1235–43. <https://doi.org/10.1093/carcin/bgn095>.

Westcott JM, Camacho S, Nasir A, Huysman ME, Rahhal R, Dang TT, et al. ΔNp63 regulated epithelial-to-mesenchymal transition state heterogeneity confers a leader-follower relationship that drives collective invasion. *Cancer Res* 2020;in press. <https://doi.org/10.1158/0008-5472.CAN-20-0014>.

Wu B, Cao X, Liang X, Zhang X, Zhang W, Sun G, et al. Epigenetic regulation of Elf5 is associated with epithelial-mesenchymal transition in urothelial cancer. *PLoS One* 2015;10:e0117510. <https://doi.org/10.1371/journal.pone.0117510>.

Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* 2021;53:1334. <https://doi.org/10.1038/S41588-021-00911-1>.

Yao B, Zhao J, Li Y, Li H, Hu Z, Pan P, et al. Elf5 inhibits TGF-β-driven epithelial-mesenchymal transition in prostate cancer by repressing SMAD3 activation. *Prostate* 2015;75:872–82. <https://doi.org/10.1002/pros.22970>.

Yin L, Duan JJ, Bian XW, Yu SC. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research* 2020;22:1–13. <https://doi.org/10.1186/S13058-020-01296-5/TABLES/3>.

Yu K Da, Cai YW, Wu SY, Shui RH, Shao ZM. Estrogen receptor-low breast cancer: Biology chaos and treatment paradox. *Cancer Commun* 2021. <https://doi.org/10.1002/cac2.12191>.

Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* 2013;339:580–4. <https://doi.org/10.1126/science.1228522>.

Supplementary Figures

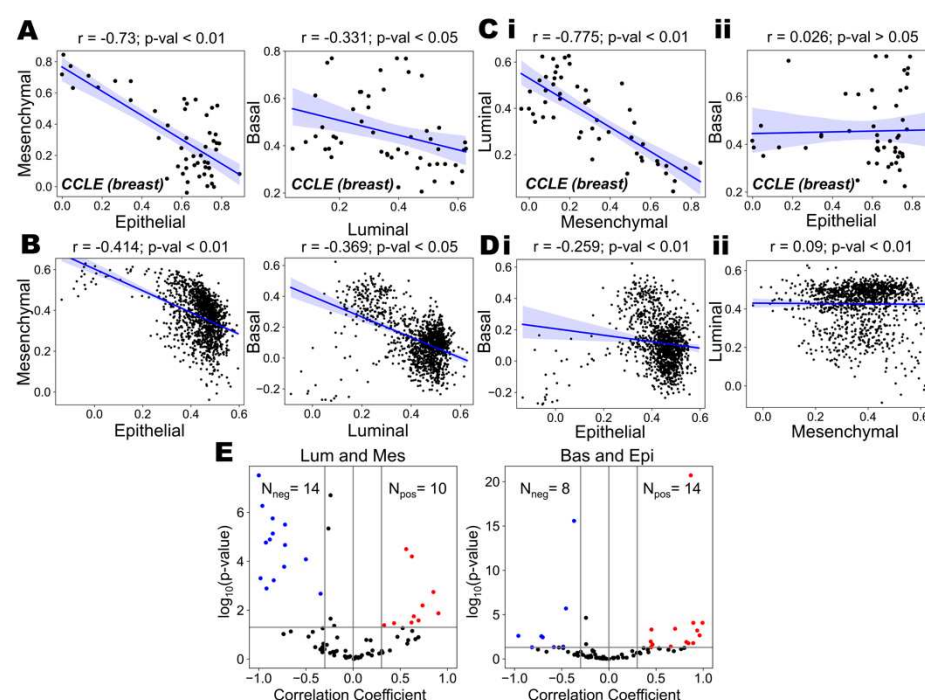


Figure S1: Associations between the luminal-mesenchymal and basal-epithelial programs in breast cancer.

A) Scatterplots showing the correlations between the ssGSEA scores of i) Epithelial-Mesenchymal and ii) Luminal-Basal programs in CCLL breast cancer cell lines. Pearson's correlation coefficients have been reported with corresponding p -values. **B)** Scatterplots showing the correlations between the ssGSEA scores of i) Epithelial-Mesenchymal and ii) Luminal-Basal programs in TCGA breast cancer patients. Pearson's correlation coefficients have been reported with corresponding p -values. **C)** Scatterplots showing the correlations between the ssGSEA scores of i) Mesenchymal-Luminal and ii) Basal-Epithelial programs in CCLL breast cancer cell lines. Pearson's correlation coefficients have been reported with corresponding p -values. **D)** Scatterplots showing the correlations between ssGSEA scores of i) Mesenchymal-Luminal and ii) Basal-Epithelial programs in TCGA breast cancer patients. Pearson's correlation coefficients have been reported with corresponding p -values. **E)** Meta analysis of 80 breast cancer specific datasets showing volcano plots of correlations for i) Mesenchymal-Luminal and ii) Basal-Epithelial programs. Pearson's correlation coefficients and corresponding p -values are given.

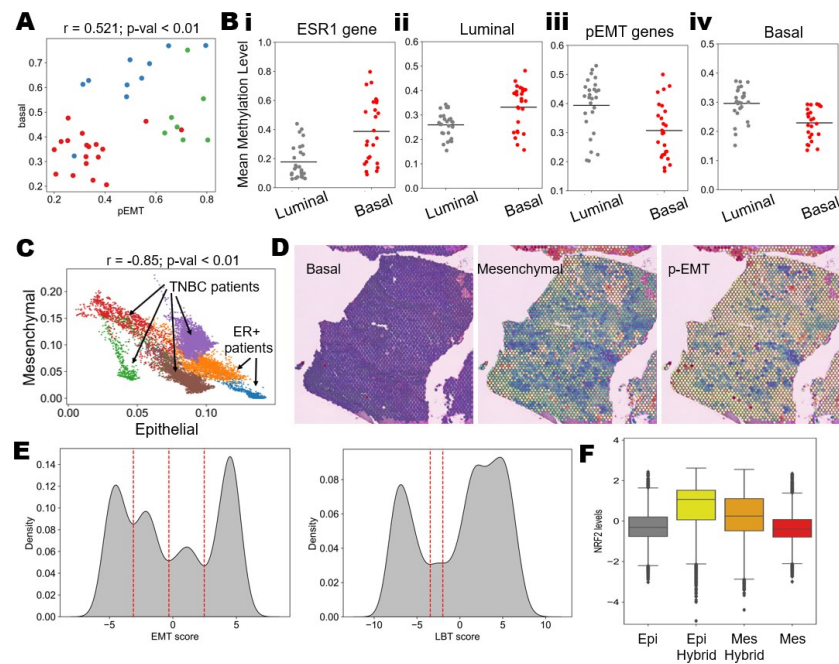


Figure S2: Epigenetic, Spatial and Simulation results of breast cancer heterogeneity. **A)** Scatterplots showing correlations between the ssGSEA scores of pEMT-basal programs in breast cancer cell lines. Pearson's correlation coefficients have been reported with corresponding p-values. **B)** Comparison of mean methylation levels across luminal and basal breast cancer for i) ESR1 gene ii) Luminal genes iii) pEMT genes and iv) Basal genes. **C)** Scatterplot showing position of the ER+ and TNBC patients on a two dimensional epithelial-mesenchymal plane. Each point on the plane is based on gene expression values for a specific spot on the spatial transcriptomic datasets. **D)** Spatial transcriptomic slides from an ER+ breast cancer patient coloured by the activity scores of basal, mesenchymal and pEMT programs. Red represents a higher activity score, and blue represents a lower activity score. **E)** Kernel density estimate of epithelial-mesenchymal score and luminal-basal scores from RACIPE simulations. Red dotted lines show the position of minima used to define phenotypes. **F)** Simulated gene expression values for NRF2 gene in different EM phenotypes.