Rearrangement of 3D genome organization in breast cancer epithelial - mesenchymal transition and metastasis organotropism

Priyojit Das[1,†,#], Rebeca San Martin[2,‡,#], Tian Hong[1,2,+], and Rachel Patton McCord[1,2,*]

[1] UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN, 37996, USA

[2] Biochemistry & Cellular and Molecular Biology, University of Tennessee, Knoxville, TN, 37996, USA

* To whom correspondence should be addressed.

† Current address: Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA; Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

‡ Current address: Department of Cell Biology, Albert Einstein College of Medicine Bronx, NY 10461, USA. Department of Oncology, Albert Einstein College of Medicine Bronx, NY 10461, USA. Ruth L. and David S. Gottesman Institute for Stem Cell and Regenerative Medicine Research, Albert Einstein College of Medicine, Bronx, NY, USA. Cancer Dormancy Institute, Montefiore Einstein Comprehensive Cancer Center Bronx, NY 10461, USA.

+ Current address: Department of Biological Sciences, The University of Texas at Dallas, Richardson, TX 75080, USA.

# Contributed equally.

Running Title: 3D genome rearrangement in metastasis organotropism

Corresponding Author: Rachel Patton McCord, 309 Ken and Blaire Mossman Building, 1311 Cumberland Ave, Knoxville, TN 37996; Tel: 865-974-3149; Email: rmccord@utk.edu

Conflict of Interest Statement: The authors declare no potential conflicts of interest

Word Count: 5,114

4 Figures, Supplementary Information: 5 Figures, 3 Tables

ABSTRACT

Breast cancer cells exhibit organotropism during metastasis, showing preferential homing to certain organs such as bone, lung, liver, and brain. One potential explanation for this organotropic behavior is that cancer cells gain properties that enable thriving in certain microenvironments. Such specific metastatic traits may arise from gene regulation at the primary tumor site. Spatial genome organization plays a crucial role in oncogenic transformation and progression, but the extent to which chromosome architecture contributes to organ-specific metastatic traits is unclear. This work characterizes chromosome architecture changes associated with organotropic metastatic traits. By comparing a collection of genomic data from different subtypes of localized and lung metastatic breast cancer cells with both normal and cancerous lung cells, we find important trends of genomic reorganization. The most striking differences in 3D genome compartments segregate cell types according to their epithelial vs. mesenchymal status. This EMT compartment signature occurs at genomic regions distinct from transcription-defined EMT signatures, suggesting a separate layer of regulation. Specifically querying organotropism, we find 3D genome changes consistent with adaptations needed to survive in a new microenvironment, with lung metastatic breast cells exhibiting compartment switch signatures that shift the genome architecture to a lung cell-like conformation and brain metastatic prostate cancer cells showing compartment shifts toward a brain-like state. TCGA

patient data reveals gene expression changes concordant with these organ-permissive compartment changes. These results suggest that genome architecture provides an additional level of cell fate specification informing organotropism and enabling survival at the metastatic site.

SIGNIFICANCE

Computational analysis of a cohort of cancer cell lines reveals some 3D genome spatial compartment changes are associated with transitions in cancer cell state that relate to metastasis (EMT) and others reflect the characteristics of the metastatic organ context.

INTRODUCTION

Breast cancer is the leading malignant disease in women worldwide. It is highly heterogeneous; based on molecular characteristics it can be divided into four broad subgroups: luminal A (estrogen receptor ER+ progesterone receptor PR+/- human epidermal growth factor receptor 2 HER2-), luminal B (ER+ PR+/- HER2+), HER2-enriched (ER- PR- HER2+), and triple-negative/basal (ER- PR- HER2-) (1). About twenty to thirty percent of early-stage breast cancer patients have a chance to ultimately develop distant metastasis, which accounts for ninety percent of all breast cancer-related deaths (2). Overall, metastasis implies the dissemination of cancer cells from the primary tumor site into the circulation and later extravasation to a new site which leads to the formation of a secondary tumor. However, the organs affected by breast cancer metastasis vary remarkably, depending on the disease subtype, molecular features, primary tumor location, and secondary site microenvironment. This phenomenon is known as 'metastasis organotropism' (3). For example, breast cancer cells can metastasize to the bone, lung, liver, and brain. However, bone metastasis accounts for a major percentage of metastatic cases compared to other organs (4). Moreover, luminal A and basal subtypes of breast cancer show a higher incidence of bone and lung metastasis, respectively, whereas luminal B preferentially metastasizes to both organs (5). Organotropic behavior is congruent with the 'seed and soil' hypothesis, according to which a cancer cell (seed) having specific intrinsic abilities will only be able to survive in a host local microenvironment (soil) that is favorable to it (6).

Secondary organ colonization by circulating tumor cells is a complex process, ruled by dynamic and evolving interactions between cancer cells and the host microenvironment. In most situations, the microenvironment is unfavorable, and the disseminated cells fail to survive in that 'soil'. Despite the low survival rate of cancer cells during the metastatic colonization process, cells with the intrinsic capabilities to interact with the host microenvironment can survive and proliferate in the secondary site (7). Alternatively, cells might arrive at the secondary site and remain dormant under the influence of the local microenvironment (8-11) slowly acquiring characteristics that would later allow their proliferation. Importantly, the microenvironment surrounding primary breast tumors seems to condition the cells to specific metastatic profiles. For example, if the primary tumor stroma is enriched in transforming growth factor beta (TGFB), lung metastasis prevails through the induction of angiopoietin-like 4 (12). In contrast, a tumor stroma enriched with mesenchymal cytokines like C-X-C motif chemokine 12 (CXCL12) and insulin-like growth factor 1 (IGF1) predisposes primary cancer cell population to home into the bone marrow (13). How these different reactive microenvironments condition the origin of specific metastatic traits still poses an outstanding question.

The current view in the cancer biology field is that transcriptional gene signatures associated with organ-specific metastatic traits are regulated in part by cues from the reactive microenvironment that accompanies primary tumorigenesis (12-15). However, whether regulating those genes involves changes in 3D genome organization is still an open question. Recent advancements in imaging and molecular biology techniques have shown that the genome organizes in a non-random multi-layered fashion inside the nucleus, and each layer of organization contributes both independently and synergistically to gene expression regulation (16). At a large scale, the reorganization of genome compartments, in which euchromatic and heterochromatic genomic regions are spatially segregated, is also found to be associated with cancer progression (17-19). It has been shown that as breast cancer becomes endocrine-resistant, the change in compartment

organization is accompanied by gene expression changes and differential estrogen receptor binding (17). Other studies have also found that an unintended consequence of various drug treatments and therapies is the rearrangement of chromatin architecture at different levels (17,20,21). Though these types of studies are essential for elucidating the association between the reorganization of the genome in three dimensions and oncogenic transformation, they do not consider the genome organization changes experienced by the metastatic cell that occur as a result of the interaction with the microenvironment of the secondary site. Genome spatial compartmentalization is strongly associated with cell type definition: genomic regions segregated into the A (euchromatic) or B (heterochromatic) compartments differ among cell types and tissue origin (22). These compartmentalization profiles help encode stable cell fate gene expression (23-25). Therefore, we hypothesize that similar to gene expression, metastatic cancer cells experience 3D genome structure rearrangements that mimic the compartment signatures of their new host microenvironment. To test this hypothesis, here we focused on the breast cancer-lung metastasis system. We compared genome organization data from different types of breast epithelial cell lines (normal, localized cancer, and metastatic) with normal and cancerous lung cells. Our analysis revealed that metastatic breast cancer cells have a higher proportion of lung-specific spatial compartment changes than localized breast cancer cells. We find similar brain microenvironment-specific compartment changes in prostate cancer cells that have the potential to metastasize to brain. These distal-site concordant compartment shifts often correspond to changes in gene expression in both cell lines and TCGA patient samples. However, we find that 3D genome changes do not always correlate with gene expression changes. Instead, particularly across the epithelial mesenchymal transition (EMT) spectrum, we find a set of genes that change expression without needing spatial reorganization and then a distinct set of epithelial and mesenchymal phenotype genes that change their spatial compartment without altered gene expression. We hypothesize that spatial genome organization facilitates immediate changes in gene expression in certain loci but can also poise certain regions for future activation in a way that favors EMT and adaptation to metastatic organ microenvironments.

RESULTS

Modeling the breast cancer lung metastasis system

To investigate whether breast cancer spatial genome organization showed rearrangements specific to a distant organ metastatic microenvironment, we needed to assemble genome spatial compartment data from different types of normal and cancerous cells for both breast and a secondary organ. Since most of breast cancer subtypes show a preference for lung metastasis, we first focused on this metastatic site (26). We selected a cohort of breast and lung cell lines, both normal and cancerous, for which publicly available genome organization (Hi-C) are available (Supplementary Table 1). A total of 15 cell lines with available data, including primary cells and cancer cell lines from both organs, were selected. A classification of the cells used in this study based on their malignancy status and other molecular signatures is shown in Fig. 1a. Briefly, for the breast cancer system we considered both HMEC and MCF10A as normal (non-tumorigenic) breast epithelial cells and placed them at the starting node of the model. Cancer cell lines were then divided into three subcategories based on breast cancer molecular signatures: luminal (ZR751, BT474, ZR7530, MCF7 and T47D), HER2-enriched (HCC1954 and SKBR3) and triple-negative (HCC70, BT549 and MDA-MB-231). We then assigned cell lines to “localized” or “metastatic” status based on the site from which they were originally collected in the patient (see details in Supplementary Table 1)(27-30). Specifically, within the luminal subtype, ZR751and BT474 were designated as localized cancers and the remaining three cell lines, ZR7530, MCF7, and T47D were used as models for ascites, lung, and lung metastatic breast cancers respectively. Similarly, for the HER2-enriched subcategory, HCC1954 and SKBR3 were considered localized and lung metastatic cancers, respectively. Finally, in the case of the triple-negative subtype, HCC70 and BT549 were used to model localized cancer whereas MDA-MB-231 represents lung metastatic cancer. For the lung portion of the system, HTBE primary cells represent normal lung epithelium and both the A549 and H460 cell lines represent localized lung cancer. The valuable MetMap study in mouse has measured the organs that can be colonized by human-derived cancer cell lines when they are injected into mice.(31) The cell lines we consider here may have different or additional

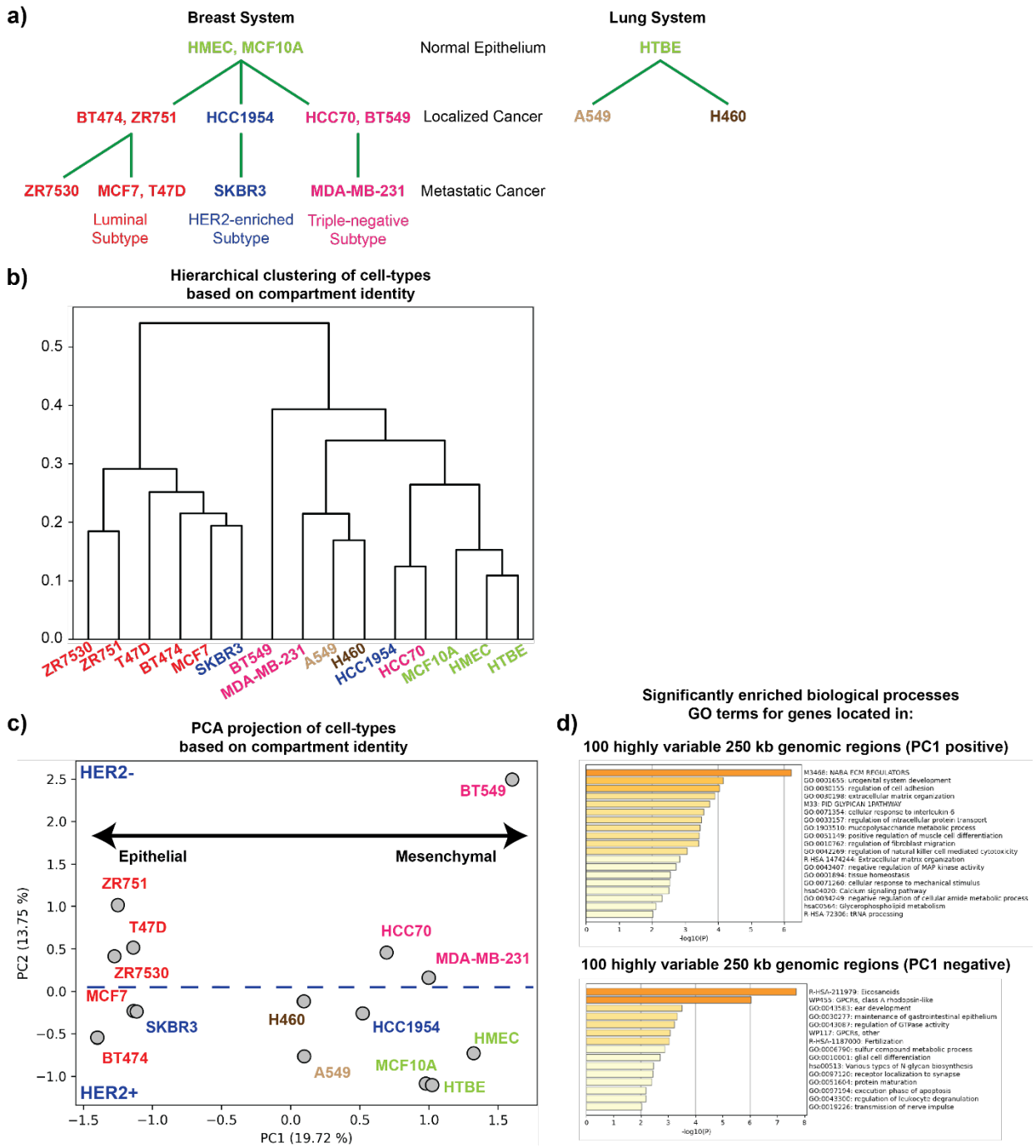

Fig 1. **Spatial compartment identity segregates cells by subtype and epithelial-mesenchymal status. a)** Primary cells and cell lines used in this study are arranged depending on their molecular subtypes and malignancy status based on prior available characteristics. Green: normal breast and lung epithelium, red: luminal subtype, blue: HER2-enriched subtype, pink: triple-negative subtype, brown: localized lung cancer. **b)** Hierarchical clustering of genome-wide compartment identity data for all breast and lung cell types at 250 kb resolution; colors as in **1a**. **c)** Principal component analysis of genome-wide compartment identity data for breast and lung cell types at 250 kb resolution; colors as in **1a**. The PC2 axis is divided into two subspaces that mostly segregate cell lines based on HER2 status (blue thick dotted line). **d)** GO term enrichment (biological processes) of the genes in genomic regions corresponding to the top 100 positive and 100 negative elements of the first eigenvector of genome-wide compartment identity PCA.

“potential” metastatic sites in this type of mouse model. But, it is known that species differences can lead to divergent organotropism in mouse models as compared to human patients.(32,33) Therefore, we use the site of collection as compared to the site of origin in the human patient as our reference point for primary and secondary site organs.

Subtype and epithelial-mesenchymal state-specific signatures are present in breast cancer compartmental organization

We compared the supervised ordering of the cell lines described above with the ordering derived from unsupervised analysis of genome-wide A/B compartment organization data. Compartment identity strengths were calculated using eigenvector decomposition on the 250 kb binned Hi-C data for each cell line (Methods). In this analysis a positive value signifies a euchromatic A compartment region, while a negative value represents a heterochromatic B compartment association.

When we applied hierarchical clustering on the genome-wide compartment data, we observed a grouping of cells that was congruent with the cell ordering constructed based on molecular and clinical features (Fig. 1b). For example, all the normal cell types, including both the breast and lung systems, were clustered together. Similarly, cells from the luminal subtype and the lung carcinoma cells were grouped in their respective clusters. However, we also observed a few interesting exceptions from the clustering result. MDA-MB-231, a breast cancer cell line isolated from lung pleural effusion, was grouped with the lung carcinoma cell lines, suggesting its compartment organization shares more features with lung cancer than do other breast cancer cell lines. We also found that HCC70 and HCC1954, which were both derived from localized cancer, clustered relatively close to the normal epithelial cells. Collectively, this unsupervised analysis shows that the different subtypes of cancer exhibit distinctive 3D genome organization features at the compartment level (Fig. 1b), as previous studies suggested (18).

Performing principal component analysis (PCA) on the genome-wide compartment data revealed similar trends (Fig. 1c). Interestingly, the ordering of cell types along the first principal component (PC1) captured a transition between epithelial and mesenchymal-like cells. Cells characterized as highly epithelial were found on the left extreme of the PC1 axis, and as we move towards the right along PC1, the cells acquire a mesenchymal phenotype. For example, MDA-MB-231 cells are derived from healthy cells of epithelial origin but have undergone an epithelial-mesenchymal transition and show a mesenchymal-like phenotype (32,33). In our Hi-C principal components analysis, this cell line has a highly positive PC1 value. In contrast, MCF7 cells maintain a more epithelial morphology and are found at the opposite end of the PC1 spectrum from MDA-MB-231. The organization of cell lines along PC1 from spatial compartmentalization matches quite well with the previously characterized epithelial-mesenchymal classification of these cell lines based on transcriptomic data (Supplementary Figure 1) (34,35). We noticed that the PC2 axis tends to segregate cell lines according to the presence or absence of HER2. As expected for an unsupervised analysis like PCA in a complex system, the separation is not perfect, as HER2 negative cell lines like MCF7 are plotted near HER2 positive cells. But, the trend of separation suggests that in breast cancer cell lines, the presence or absence of this receptor and its downstream signaling network has important effects on the genome organization at the compartment scale. To further characterize the genomic regions that showed distinctive compartment signatures across cell lines, we extracted lists of all genes that fell into genomic bins with the 200 most extreme eigenvector loadings for PC1: 100 strongest positive and 100 strongest negative bins. Gene ontology enrichment results for these gene lists coincide with the idea that PC1 represents the EMT axis: PC1 positive genes have functions related to cell adhesion (tenascin C, JAG1, ACTA2), extracellular matrix organization (APP, LAMA2, FGF2, BMP4) and positive regulation of cell migration, consistent with a mesenchymal phenotype, while PC1 negative genes show GO terms related to epithelial maturation (Fig. 1d) Overall, our analyses show that the genome organization is not only able to distinguish different subtypes of cancers but also exhibits changes along the epithelial-mesenchymal transition.

Breast cancer transcriptome profiles also exhibit epithelial-mesenchymal state-based segregation

We next analyzed the gene expression signatures that emerged across different subtypes of breast cancer (data sources described in Supplementary Table 1) and compared those with the patterns obtained from the compartmental analysis. We first performed PCA on the transcriptome profile of all the breast and lung system cells considered in this study (Fig. 2a). From the low dimensional PC projection result, we observed that similar breast cancer subtype cells distribute in the PC space based on their transcriptome profile. As in the compartment signature, MDA-MB-231 cells localize close to lung cancer cells A549 and H460, suggesting that this lung metastatic cell line has transcriptional similarities to lung cancer. We also noticed that along the largest PC axis (PC1), cells are ordered based on their epithelial-mesenchymal transition state. Performing GO term enrichment analysis using the genes from the regions with the top 100 most positive and 100 most negative eigenvector loadings for this PC1, we find that genes associated with PC1 positive values relate to epithelial organization, differentiation and negative regulation of cell migration, denoting a more epithelial phenotype (Fig. 2b). In contrast, PC1 negative genes relate to the mesenchymal functions with enriched terms for wound healing, hemostasis, cell movement, and ECM remodeling (Fig. 2b). Along with the unsupervised PCA, we also performed gene expression analysis in a supervised fashion using two different methods – gene set variation analysis (GSVA) and non-negative principal component analysis (nnPCA) (Fig. 2c and 2d, see Methods).

Briefly, for each cell type, an E (epithelial) score and an M (mesenchymal) score were calculated based on the expression levels of a curated set of 232 epithelial and 193 mesenchymal genes (35,36). A higher E score represents more epithelial characteristics whereas a higher M score corresponds to a mesenchymal phenotype. The GSVA and nnPCA transcriptome analyses recapitulate a similar ordering of cell lines according to epithelial-mesenchymal characteristics.

Distinct EMT gene sets are affected by spatial compartment changes vs. transcription changes, revealing multiple layers of EMT signature

Numerous studies have already demonstrated that cancer cells segregate along an EMT axis by transcriptomics (34-39). Given that spatial compartmentalization is often correlated with gene expression (22), a simple possible explanation for the EMT segregation of cells by compartment signatures could be that genes which change expression also change compartment. An alternative hypothesis would be that transcription and 3D genome organization change represent two independent aspects of the EMT transition. To evaluate these possibilities, we compared the genes obtained from the compartmental PCA analysis, the transcriptome PCA analysis, and the curated epithelial-mesenchymal gene set (Fig. 3a, gene lists found in Supplementary Table 2). We found that the genes in regions that change spatial compartmentalization have very little overlap with the sets of genes that most varied in transcription across these cell lines or were curated as differently expressed EMT genes (Fig. 3a). To further investigate this result, we examined the transcription status of genes that changed compartment across the EMT spectrum and, conversely, the compartment status of genes that changed transcription (Fig. 3b, c, and d). As expected, hierarchical clustering of the detected highly varying compartments (Fig 3b, top) and highly varying transcripts (Fig 3c and d, bottom) resulted in clear segregation of the cell types into similar groupings (i.e., HMEC and MCF10A together, BT474 and SKBR3 together with an opposing pattern). But, the genes which strongly changed compartments between cell lines showed no consistent patterns of correlated expression change (Fig. 3b, bottom). Likewise, very few compartment differences existed across cell lines at genes whose transcription segregated cells by EMT status (Fig. 3c and d, top). Notably, for both the transcriptome PCA analysis and the curated EMT gene set, the majority of the differentially expressed

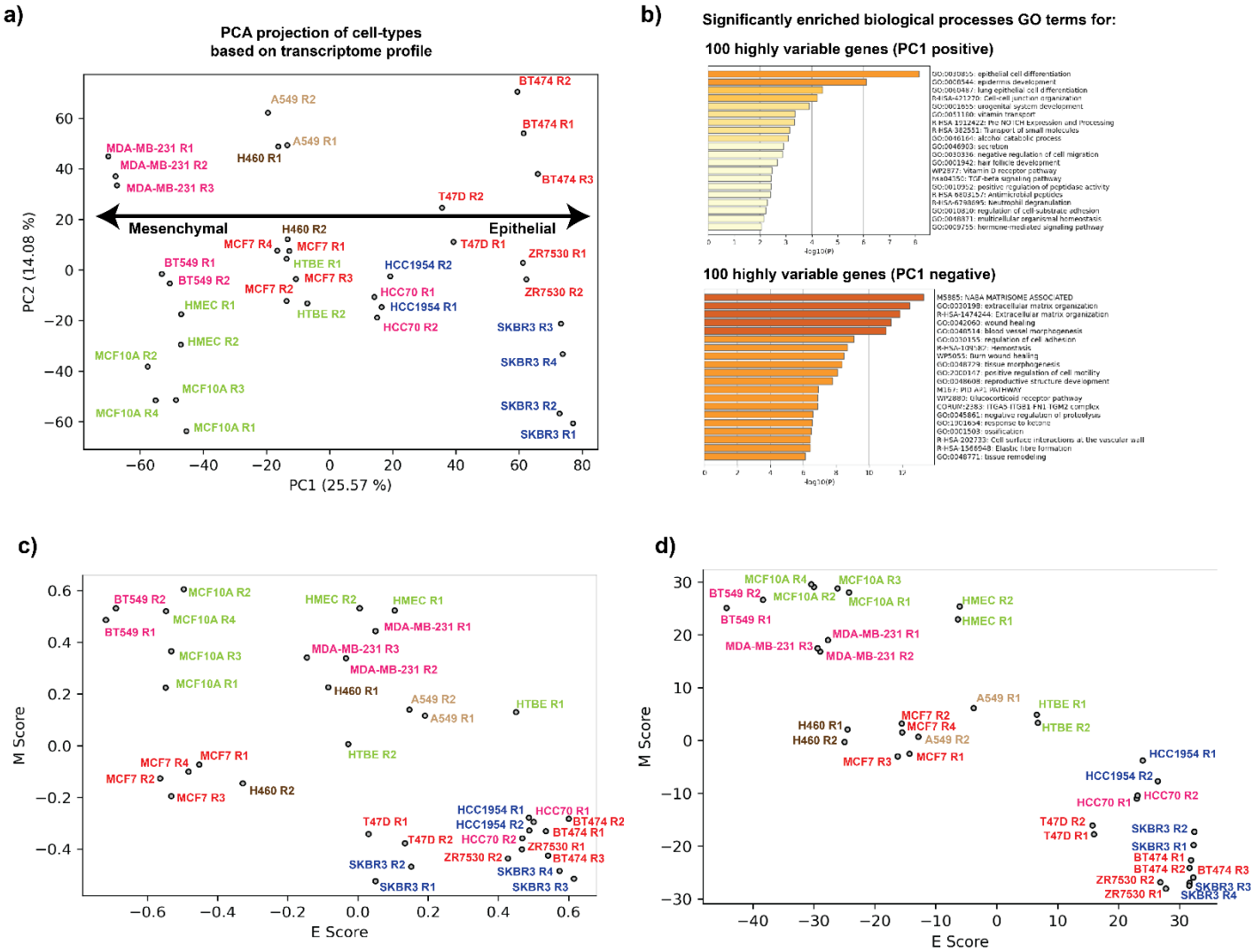

Fig 2. **Gene expression differences also segregate breast cancer cell lines along an EMT axis. a)** Principal component analysis of transcriptome data for breast and lung cell types in this study. Cell types are colored based on their molecular subtypes as represented in **Fig. 1a**. **b)** GO term enrichment (biological processes) of the genes corresponding to top 100 positive and 100 negative elements of the first eigenvector of the transcriptome profile PCA. **c, d**) Projection of breast and lung cell types on a curated epithelial and mesenchymal axis based on their gene expression using gene set variation analysis (GSVA) (**2c**) and non-negative principal component analysis (nnPCA) (**2d**) methods.

genes are found in the active A compartment. Overall, our results favor the hypothesis that compartmentalization and gene expression signatures are capturing two independent factors in EMT transition. The genes which differ most in expression across these cell lines are largely already accessible (A compartment) and therefore their expression is likely not controlled by spatial compartmentalization. Meanwhile, the additional EMT relevant genes that change compartment, but not transcription, may be now spatially poised for future gene regulation given additional environmental cues, as has been shown in other contexts (40-43). Overall, the set of genes that changes compartments does not have as clear epithelial and mesenchymal functional enrichment as the transcription change set of genes (as shown by the analysis of these gene sets presented in Figure 1d and Figure 2b). This could indicate that some of the compartment changes that occur with EMT are not directly gene regulatory but rather enable an overall conformational change of the chromatin that is needed for the alterations in physical cell state or to accomplish long distance gene regulation changes. Indeed, we observe specific cases where a compartment change region

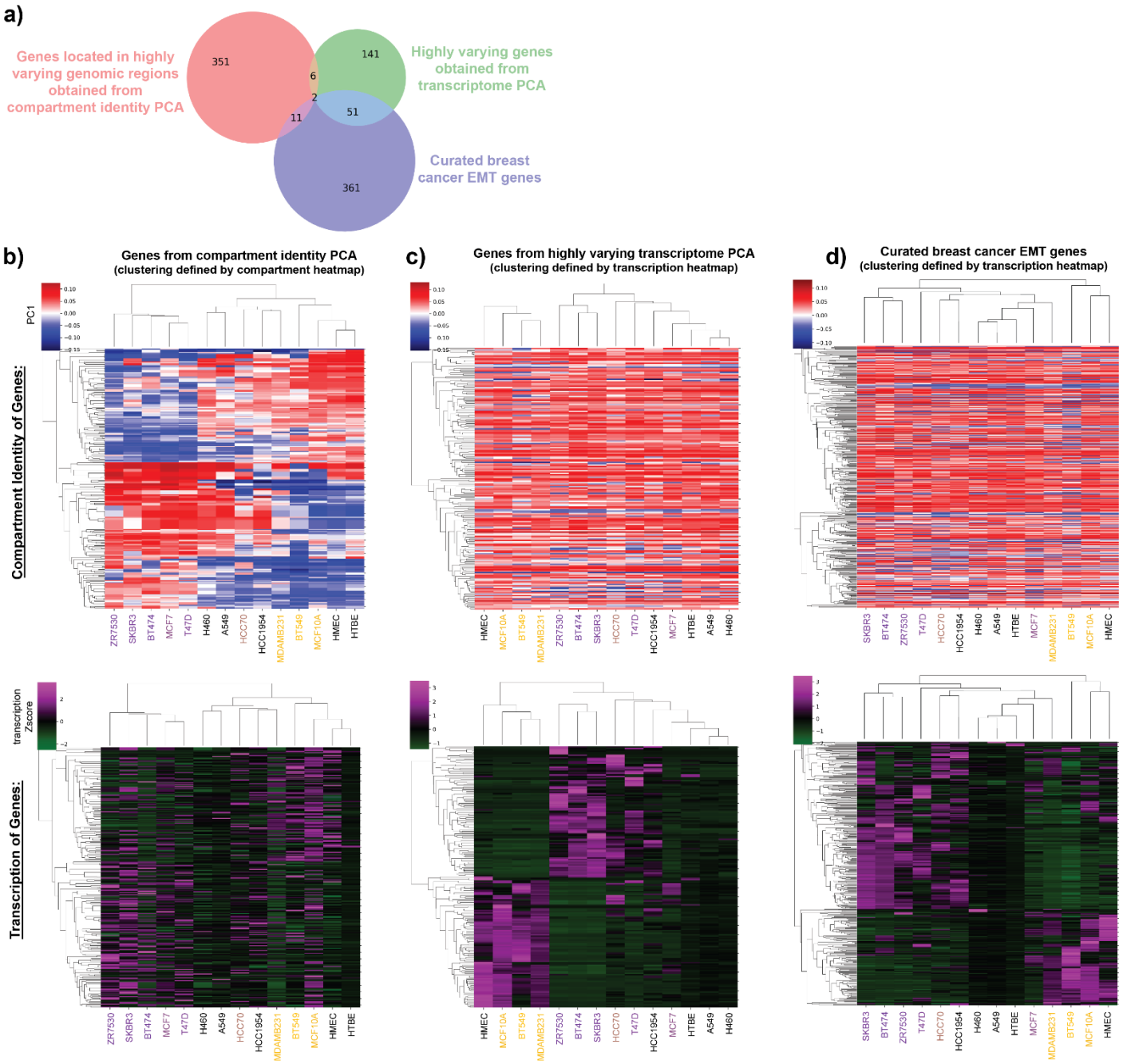

Fig 3. **Transcription and compartment changes capture distinct sets of EMT-related genomic regions. a**) Overlap analysis of the genes obtained from the compartmental analysis PC1 (red, same genes analyzed in Figure 1d), the transcriptome analysis PC1 (green, same genes analyzed in Figure 2b), and a curated breast cancer epithelial-mesenchymal gene set (blue). See Supplementary Table 2 for gene lists. **b-d)** Hierarchical clustering of compartment profiles (top row) and gene expression (bottom row) of the genes obtained from either **(b)** the compartmental analysis PC1, **(c)** transcriptome analysis PC1, or **(d)** the curated set of breast cancer epithelial and mesenchymal genes. For (b), the clustering order is determined by the compartment profile and then the same order of genes and cell lines is shown for the expression data. For (c) and (d), the clustering order is based on gene expression profiles and then the same order is shown for compartment profiles. Cell line names are colored according to position along the EMT axis (purple = more epithelial, yellow = more mesenchymal) as presented in Le et al., 2018 (34) and shown in Supplementary Figure 1a.

is located near but not at a gene expression change region (Supplementary Figure 1c). For example, TFF3 is in the set of genes that decreases in expression in M vs. E type breast cancer cell lines. While the gene itself remains in the A compartment in most all cell types, it lies near several regions that switch from A to B in M vs. E type cancer cell lines. Therefore, some of the discordant gene expression and compartment change regions may reflect the impact of compartment shifts on gene expression in neighboring regions. Notably, this same genomic region on chr21 showed striking compartment switches with cancer progression in prostate cancer as well.(19)

Lung metastatic breast cancer cell lines acquire lung-like genome architecture

To survive a metastatic environment, a cancer cell must express a repertoire of binding factors, extracellular matrix proteins, and ligands that positively interact with the secondary tissue microenvironment and foster colonization and growth. We hypothesized that for that to happen, the genome organization of the metastatic cell should change to resemble that of the cells at the secondary site, as a prelude of transcriptional activation. To evaluate this hypothesis, we compared the 3D genome organization of the metastatic and localized breast cancer cells with lung epithelial vs. lung cancer cells (Fig. 4a, 4b). As a baseline, we first compared the breast cancer compartment identity data with that of the normal breast epithelial cells. For each comparison, this resulted in four possible compartment identity combinations: AA (A compartment in both normal breast and breast cancer), AB (A compartment in normal breast and B compartment in breast cancer), BA (B compartment in normal breast and A compartment in breast cancer), and BB (B compartment in both normal breast and breast cancer). Similarly, the compartment identity of lung cancer cell lines was also compared with that of the normal lung cell. We then performed a cross-comparison between the breast and lung cancer comparisons, leading to a total of 16 compartment identity combinations (Fig. 4b). Of these, we were particularly interested in cases where the compartment state of healthy breast epithelial cells consistently differed from lung. It is notable that this set is relatively small to begin with (on the order of 200 250-kb bins), which emphasizes the overall similarities between epithelial cell types, whether they come from breast and lung. But for these regions that differ between healthy breast and lung, we next asked whether the breast cancer compartment would retain its breast identity or switch to match lung. In all the breast cancer cell lines derived from lung metastases, we found that more than half of these genomic regions switched compartments to match the lung compartment state (Supplementary Figure 2). We term these switches that make the breast cancer cell compartment identity more like the lung as “lung permissive genome organization changes”. They include: AB_BB (Genomic regions which are in the B compartment in both normal lung and localized lung cancer, and which change from A to B in the transition from normal breast to breast cancer) and BA_AA (Genomic regions which switch from B to A in breast cancer compared to normal breast epithelium, and which are in the A compartment in both normal lung and lung cancer). These combinations allow us to distinguish changes that could foster lung-specific-survival changes from cancer-specific traits such as AB_AB and BA_BA (compartments that change in the same direction from normal to cancer cells in both breast and lung). We then evaluated whether lung-permissive 3D genome changes occurred more frequently in lung metastatic breast cancer than in localized breast cancer.

To quantify the frequency of lung permissive changes overall, we calculated what fraction of the compartment alterations in breast cancer vs. normal breast tissue were also counted as lung permissive. We divided cells according to breast cancer subtype for this comparison. For the subtypes where several cell lines were available, we considered compartment identity combinations that were consistent between members of the group. We found that for luminal and triple-negative subtypes, metastatic breast cancer exhibits a higher level of lung permissive changes than localized cancer (Fig. 4c).

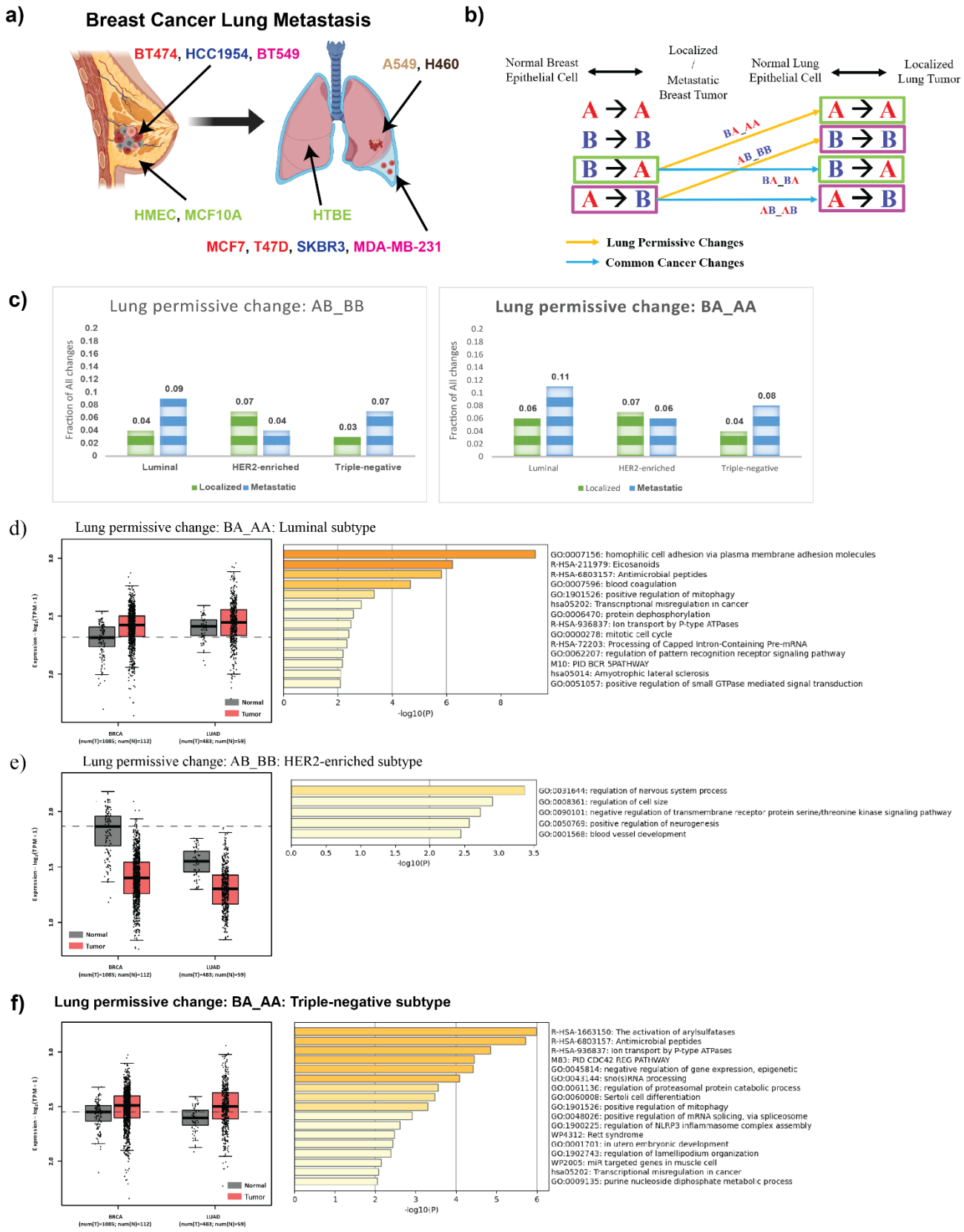

Fig 4. **a)** Normal breast epithelium (HMEC and MCF10A), localized breast cancer (BT474, HCC1954, and BT549), and

lung metastatic breast cancer (MCF7, T47D, SKBR3, and MDA-MB-231) cells are used to model different stages of breast cancer progression. To model diseased lung, normal lung epithelium (HTBE) and localized lung cancer (A549 and H460) cells are used. **b)** Schematic diagram representing breast cancer lung permissive changes calculation. First, for both breast and lung, cancer cell compartment identity is compared with the normal epithelial cell. This results in four possible compartment identity combinations: AA (A compartment in both normal and cancer), AB (A compartment in normal and B compartment in cancer), etc. Then, a cross-comparison between the breast and lung systems leads to 16 compartment identity combinations. Two specific combinations: AB_BB and BA_AA are defined as lung permissive changes (yellow arrows). **c)** Fraction of lung permissive changes shown by localized and metastatic cancers from different breast cancer subtypes. **d-f)** TCGA patient gene expression (left) and GO term enrichment (biological processes) (right) for genes from regions exhibiting **(d)** BA_AA lung permissive changes in the luminal metastatic breast cancer subtype, **(e)** AB_BB lung permissive changes in case of HER2-enriched metastatic breast cancer subtype, or **(f)** BA_AA lung permissive changes in case of triple-negative metastatic breast cancer subtype. TCGA data from BRCA and LUAD tumor vs. normal sets used and sample size indicated.

Genes affected by lung permissive compartment changes show lung-like expression levels cancer patients

After selecting the genomic regions that exhibit lung permissive changes in metastatic breast cancer cell lines, we extracted the underlying genes and used the GEPIA2 tool (44) to analyze the expression of these gene sets in TCGA breast and lung cancer patient data (Fig. 4d, 4e, and 4f; left panels). We found that in general, the gene expression levels change concordantly with the compartment. The results show that for the regions exhibiting lung-permissive changes, the switch in compartmental identity is associated with the gene expression change from normal to tumor. That is, for BA_AA regions, the breast tumor expression of these genes is higher than normal breast tissue expression, coinciding with higher expression in the lung (Fig. 4d). Conversely, for AB_BB regions, the breast tumor expression of genes in these regions is lower than for normal tissue, corresponding to lower expression in the lung (Fig. 4e). Notably, these lung-permissive compartment change genomic regions do not strongly overlap with the key compartment switch regions we identified by principal components analysis (Supplementary Table 3). Correspondingly, the enriched functions of genes in these lung permissive switch regions (Figure 4d-f) are different than the EMT-related functions found among the genes identified by PC1.

Organotropic 3D genome changes match target organ more than an unrelated organ

To verify that “lung permissive changes” truly occurred more than expected in lung metastatic cells, we used a different organ system- brain metastasis and brain cancer- as a negative control. Breast cancer can metastasize to brain(45,46), but the specific metastatic cell lines for which we have data were not derived from brain metastases. Therefore, we would expect fewer “brain like” compartment changes in metastatic breast cancer that metastasizes to lung. Before comparing percent permissive changes between organs, we developed an adjusted version of the calculation to account for differing baseline probabilities of compartment change at the distant organ site (Supplementary Figure 3 and Methods). The adjusted calculation still results in a higher level of lung permissive changes in lung metastatic breast cancer cell lines as compared to localized (Supplementary Figure 4d). For the brain comparison, we used the NHA cell line (normal human glia) and the SF9427 cell line (glioblastoma) as our reference models (Supplementary Figure 4a and 4b). We then calculated the adjusted level of brain permissive changes shown by localized and metastatic breast cancers (Supplementary Figure 4c). We found that the metastatic breast cancer cells used in this study showed lung-permissive changes more frequently than brain-permissive compartment changes (Supplementary Figure 4c and d). This is congruent with the fact that these metastatic breast cancer cell lines were all isolated from lung pleural effusions. We note that Hi-C data for brain metastatic breast cancer cell lines is not currently available for comparison.

Brain metastatic prostate cancer cell lines exhibit organotropic 3D genome changes toward brain identity

The analysis above suggests that lung metastatic breast cancer cell lines show more lung-permissive than brain-permissive genome organization changes. Next, we asked whether a cancer type that metastasizes to the brain exhibits brain-permissive compartment changes. For this purpose, we chose prostate cancer cell lines derived from localized (lymph node) and brain metastatic sites (LNCaP and DU145) and compared their genome architecture to the non-tumorigenic prostate epithelial cell line RWPE (Supplementary Figure 5a, b). We found that the brain metastatic cells showed more brain-permissive changes than the one derived from a lymph node (Supplementary Figure 5c). Also, particularly in the B to A direction, there were greater brain permissive changes in this prostate cancer system (Supplementary Figure 5c) than in the breast cancer system (Supplementary Figure 4c). This confirms that a cancer type that metastasized to the brain matched the brain genome compartmentalization profile more often than a cancer type that had not metastasized to the brain. We found that the genes in the switched compartment regions have functions that could relate to leaving the prostate epithelium and adapting to the brain environment (Supplementary Figure 5d). For example, the DU145 (brain metastatic) brain-permissive B to A switched genes were enriched for functions in regulation of synapse organization (including calcium channel CACNA2D3). Conversely, the regions that switch from A to B to match the brain B compartment and are enriched for keratinization and cell adhesion, suggesting a downregulation of prostatic epithelial properties.

DISCUSSION

During metastasis, breast cancer cells spread preferentially to certain organs, a series of events widely known as metastasis organotropism. This phenomenon, also shown by other types of cancer, may be explained, to some extent, by the 'seed and soil' hypothesis which states that cancer cells will be able to colonize a distant site if the cells have intrinsic ability to survive in the microenvironment of the secondary site. Though the process of metastasis is complex and involves other factors, the primary tumor microenvironment conveys survival capabilities to cancer cells that can potentially favor survival in other organs; for this to occur, the cancer cell must engage specific gene expression signatures. However, there are many open questions related to the rise of organ-specific survival traits. Are these gene signatures a result or consequence of genome organization changes? What is the role of genome rearrangements in metastasis organotropism for which there are no significant gene expression alteration? Does the genome organization of the metastatic cell become similar to that of the cells at the distant site? In this work, we have made a step toward answering these questions by analyzing compartmental organization data of breast cancer cell lines derived from lung metastasis in comparison to lung cell lines.

Unsupervised clustering of compartment organization data showed that compartment organization can distinguish cancer cell lines by subtype, providing the first indication that lung metastatic cell lines may show similarities to lung cell genome organization. For example, the triple-negative invasive metastatic breast cancer cell MDA-MB-231, clusters close to the lung carcinoma cells (A549 and H460). The clustering result supports the conclusion that the breast cancer subtypes are not only quite different at the gene expression level, but also at the level of genome organization. Furthermore, it also shows that the genome organization is highly consistent among luminal subtype cancers, irrespective of the tissue from which the cells were collected, while the other subtypes such as triple negative, were more heterogeneous in their structure. This is congruent with studies involving gene expression which have shown that the triple-negative breast cancer category is highly heterogeneous, encompassing six subtypes (47).

When we compared the compartmental organization changes shown by the localized and lung metastatic breast cancers to the compartment state of the lung, we found that, for luminal and triple-negative subtypes, metastatic cells exhibit higher percentages of lung permissive changes compared to localized cancers. Given that these cancer cell lines were derived from lung pleural effusions, we cannot tell whether these metastatic cells acquired a lung-like compartmental organization at the primary tumor site or later while adapting to the metastatic site. To answer this, we would need temporal genome organization data of metastatic progression, which is not currently available. Our current study is also limited by the use of cell

lines derived from multiple different patient sources and grown in culture. To validate and clarify our observations, future research would need to analyze the 3D genome structure of matched primary and metastatic tumors from the same human patient.

For the lung permissive chromatin conformation changes, we find that shifts toward the A compartment are favored compared to the B compartment. It may be possible that shifts toward the A compartment are predominant because both activation and gene repression are possible once a region is in the A compartment while a shift to the B compartment is not necessary to accomplish gene repression (40,48,49). Alternatively, this preference for shifts toward an A compartment could also be related to the general loss of heterochromatin previously observed in cancer (50). Indeed, we found that genes in these spatially reorganized regions, in patient samples, tend to change expression concordantly with their compartment changes. Overall, our results are consistent with the idea that genome compartmentalization can play a role in facilitating organ-specific metastasis.

While our gene expression and compartment analyses echo the current dogma that 3D genome compartments correlate with gene expression changes, further analysis revealed that gene expression and compartment signatures across breast cancer types surprisingly capture two orthogonal sets of genomic regions. We found that compartment profiles group breast cancer cell lines according to their epithelial vs. mesenchymal states (51). This observation becomes much more visible in the PCA transformed data as the cells are arranged along an epithelial-mesenchymal spectrum. Pathways of EMT such as the loss of epithelial identity and the gain of migratory ability can play important roles in facilitating metastasis, though it is important to note that EMT is only one of many complex processes involved in metastasis.(52) While epithelial vs. mesenchymal states have been clearly associated with gene expression, it is also becoming clear that these states exhibit characteristic genome architectures as well. Indeed, a recent study shows that genome organization changes during epithelial-mesenchymal transition in ovarian cancer system, by considering some representative epithelial and mesenchymal cell lines (53). Here, we extend this conclusion for another major cancer, the breast cancer system by including a spectrum of breast and lung cancer and normal cell lines. Across this cohort, we find not only that both 3D genome architecture and gene expression changes occur during EMT, but also that these changes happen in different regions. The 3D genome changes that most strongly capture the differences across the epithelial and mesenchymal spectrum occur at regions where gene expression is not notably distinct between epithelial and mesenchymal cells. Likewise, the signature gene expression changes occur in regions that tend to be in the A compartment across all cell types. This suggests that many genes key to EMT are regulated by local factors in the A compartment, while major spatial shifts in chromosome regions may alter long-distance regulation of genes in other regions, change the likelihood of future inducibility of genes, or represent alterations in the physical/mechanical properties of the cell during EMT, which can affect chromatin state (54,55). Overall, our results suggest that genome spatial compartment changes can help encode cancer cell states that may play a role in metastasis as well as enabling survival in a new organ context.

METHODS

Hi-C data processing

To process Hi-C sequencing reads, we used the HiC-Pro (https://github.com/nservant/HiC-Pro) pipeline with default parameters (56). The reads were aligned to the hg19 reference genome. Also, for each cell type, all the available reads from all replicates were merged together to improve the quality of contact data for further downstream processing. After obtaining the valid pair interactions, we binned the interactions at 250 kb resolution for each individual chromosome and normalized the matrices using the ICE technique (57). Following that, the number of normalized contacts of each chromosome was scaled to 1M using 'scaleMatrix.pl' script from the cworld-dekker (https://github.com/dekkerlab/cworld-dekker) tool suite. And finally, in order to get the compartmental organization for each individual chromosome, principal components analysis was performed on the normalized and scaled matrices using the

'matrix2compartment.pl' script from the cworld-dekker tool suite. The compartment analysis assigns A/B compartment identity to genomic regions based on the sign of the elements of the eigenvector (A-positive and B-negative).

RNA-seq data processing

The RNA-seq files were processed as per the steps described in the previous publication (40). Briefly, the BBDuk (https://github.com/kbaseapps/BBTools) tool was used first for adapter and quality trimming (58). After that, the reads were mapped to hg19 human reference genome using STAR aligner (RRID:SCR_004463) (59). Once aligned reads were generated, HTSeq-Counts (https://github.com/simon-anders/htseq) tool was used to perform gene level feature count (60). As the RNA-seq files were obtained from various sources and generated using different experimental techniques, they suffer from the batch effect which masks the biological signal. To remove the batch effect, ComBat-seq (https://github.com/zhangyuqing/ComBat-seq) tool was applied on the raw gene counts data (61).

Hierarchical clustering of genome-wide compartmental organization

To cluster different cell types based on their compartmental organization, hierarchical clustering was performed on the genome-wide compartment identity data. First, the pairwise similarity score between all the cells were calculated using Pearson's correlation. The similarity data was then further converted to distance data by subtracting from 1. And finally, hierarchical clustering was performed on the distance data using 'ward' linkage.

Gene set variation analysis (GSVA)

GSVA was performed to compute the enrichment scores for individual cell lines using RNA-seq data (62,63). A list of 232 epithelial signature genes, and a list of 193 mesenchymal signature genes were used to compute the E enrichment score (E score) and M enrichment score (M score) respectively (35,36,64).

Non-negative principal component analysis (nnPCA)

We performed nnPCA using the gene sets mentioned above. nnPCA provided higher resolutions for large number of samples in the EMT spectrum compared to GSVA according to previous studies (36,65). nnPCA determines the approximately orthogonal axes with non-negative coefficients (loadings) for features (genes). Variances of projections of data points (cell lines) onto these axes are maximized via an optimization method (66).

Correction for comparing organ permissive changes across different organs

It is possible that while comparing the host organ permissive changes across different secondary organs, the compartment switch ratio of different secondary localized cancers might include inherent bias to the comparison. To eliminate the bias, for the secondary organ systems, we first calculate the probability of a genomic region to be in a specific compartment in localized cancer given that the region belongs to the same compartment in normal cell (Supplementary Figure 3). For the A compartment, p(TA|NA) represents the probability of a region in A compartment in localized cancer cell given that region also belongs to A compartment in normal cell. Similarly, p(TB|NB) for the B compartment. If we have more than one cell line representing the secondary site localized cancer, the minimum probability across all the cells is considered. Once we have those, the probabilities are subtracted from one and the resultant values are multiplied with corresponding secondary organ specific percentages. For example, (1-p(TA|NA)) is multiplied to the percentage of BA_AA and (1-p(TB|NB)) to the AB_BB.

ACKNOWLEDGEMENTS

The authors would like to thank all the members of McCord lab for their fruitful suggestions in developing the research project. This work was supported by the National Institutes of Health NIGMS grant R35GM133557 to R.P.M and American Cancer Society postdoctoral fellowship 134060-PF-19-183-01-

CSM to R.S.M. T.H. is supported by the National Institutes of Health NIGMS grant R35GM149531. This work used Bridges2 at Pittsburgh Supercomputing Center and Expanse at San Diego Supercomputer Center through allocation BIO230105 (awarded to P.D.) from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

DISCLOSURE STATEMENT

The authors declare no competing interests exist.

DATA AVAILABILITY

Datasets used for this project were obtained from NCBI GEO, EMBL ENA, 4DN Data Portal and ENCODE repositories and their accession IDs are given in Supplementary Tables 1 and 2.

AUTHOR CONTRIBUTIONS

P.D. conceived the study in collaboration with R.P.M and R.S.M., performed data analyses, and prepared figures. R.S.M. performed the grouping and ordering of cancer cell lines into biological subtypes and interpreted the biological enrichment analysis. T.H. performed EMT gene signature analyses. R.P.M. supervised and guided the work. P.D. and R.P.M. wrote the manuscript with input from all authors.

SUPPLEMENTARY MATERIAL

The supplementary material associated with the paper can be found and accessed online.

REFERENCES

1. Fragomeni SM, Sciallis A, Jeruss JS. Molecular subtypes and local-regional control of breast cancer. Surgical Oncology Clinics **2018**;27:95-120
2. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. science **2011**;331:1559-64
3. Chen W, Hoffmann AD, Liu H, Liu X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. NPJ precision oncology **2018**;2:1-12
4. Savci-Heijink CD, Halfwerk H, Koster J, van de Vijver MJ. A novel gene expression signature for bone metastasis in breast carcinomas. Breast cancer research and treatment **2016**;156:249-59
5. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG*, et al.* Subtypes of breast cancer show preferential site of relapse. Cancer research **2008**;68:3108-14
6. Langley RR, Fidler IJ. The seed and soil hypothesis revisited—The role of tumor-stroma interactions in metastasis to different organs. International journal of cancer **2011**;128:2527-35
7. Obenauf AC, Massagué J. Surviving at a distance: organ-specific metastasis. Trends in cancer **2015**;1:76-91
8. Sharma S, Xing F, Liu Y, Wu K, Said N, Pochampally R*, et al.* Secreted Protein Acidic and Rich in Cysteine (SPARC) Mediates Metastatic Dormancy of Prostate Cancer in Bone. J Biol Chem **2016**;291:19351-63
9. Taichman RS, Patel LR, Bedenis R, Wang J, Weidner S, Schumann T*, et al.* GAS6 receptor status is associated with dormancy and bone metastatic tumor formation. PLoS One **2013**;8:e61873
10. Shiozawa Y, Berry JE, Eber MR, Jung Y, Yumoto K, Cackowski FC*, et al.* The marrow niche controls the cancer stem cell phenotype of disseminated prostate cancer. Oncotarget **2016**;7:41217-32
11. Dalla E, Sreekumar A, Aguirre-Ghiso JA, Chodosh LA. Dormancy in Breast Cancer. Cold Spring Harb Perspect Med **2023**;13

12. Padua D, Zhang XH-F, Wang Q, Nadal C, Gerald WL, Gomis RR*, et al.* TGFβ primes breast tumors for lung metastasis seeding through angiopoietin-like 4. Cell **2008**;133:66-77
13. Zhang XH-F, Jin X, Malladi S, Zou Y, Wen YH, Brogi E*, et al.* Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. Cell **2013**;154:1060-73
14. Sceneay J, Chow MT, Chen A, Halse HM, Wong CS, Andrews DM*, et al.* Primary tumor hypoxia recruits CD11b+/Ly6Cmed/Ly6G+ immune suppressor cells and compromises NK cell cytotoxicity in the premetastatic niche. Cancer research **2012**;72:3906-11
15. Gao Y, Bado I, Wang H, Zhang W, Rosen JM, Zhang XH-F. Metastasis organotropism: redefining the congenial soil. Developmental cell **2019**;49:375-91
16. McCord RP, Kaplan N, Giorgetti L. Chromosome conformation capture and beyond: toward an integrative view of chromosome structure and function. Mol Cell **2020**;77
17. Achinger-Kawecka J, Valdes-Mora F, Luu P-L, Giles KA, Caldon CE, Qu W*, et al.* Epigenetic reprogramming at estrogen-receptor binding sites alters 3D chromatin landscape in endocrine-resistant breast cancer. Nature communications **2020**;11:1-17
18. Kim T, Han S, Chun Y, Yang H, Min H, Jeon SY*, et al.* Comparative characterization of 3D chromatin organization in triple-negative breast cancers. Experimental & Molecular Medicine **2022**:1-16
19. San Martin R, Das P, Dos Reis Marques R, Xu Y, Roberts JM, Sanders JT*, et al.* Chromosome compartmentalization alterations in prostate cancer cell lines model disease progression. Journal of Cell Biology **2021**;221:e202104108
20. Barutcu AR, Hong D, Lajoie BR, McCord RP, van Wijnen AJ, Lian JB*, et al.* RUNX1 contributes to higher-order chromatin organization and gene regulation in breast cancer cells. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **2016**;1859:1389-97
21. Yang Y, Choppavarapu L, Fang K, Naeini AS, Nosirov B, Li J*, et al.* The 3D genomic landscape of differential response to EGFR/HER2 inhibition in endocrine-resistant breast cancer cells. Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms **2020**;1863:194631
22. Li H, Playter C, Das P, McCord RP. Chromosome compartmentalization: causes, changes, consequences, and conundrums. Trends Cell Biol **2024**
23. Schmitt AD, Hu M, Jung I, Xu Z, Qiu Y, Tan CL*, et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. Cell reports **2016**;17:2042-59
24. Vilarrasa-Blasi R, Soler-Vila P, Verdaguer-Dot N, Russiñol N, Di Stefano M, Chapaprieta V*, et al.* Dynamics of genome architecture and chromatin function during human B cell differentiation and neoplastic transformation. Nature communications **2021**;12:651
25. Owen JA, Osmanović D, Mirny LA. Design principles of 3D epigenetic memory systems. bioRxiv **2022**:2022.09. 24.509332
26. Bennett C, Carroll C, Wright C, Awad B, Park JM, Farmer M*, et al.* Breast Cancer Genomics: Primary and Most Common Metastases. Cancers **2022**;14:3046
27. Engel LW, Young NA. Human breast carcinoma cells in continuous culture: a review. Cancer Res **1978**;38:4327-39
28. Gazdar AF, Kurvari V, Virmani A, Gollahon L, Sakaguchi M, Westerfield M*, et al.* Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer **1998**;78:766-74
29. Wang J, Huang TY-T, Hou Y, Bartom E, Lu X, Shilatifard A*, et al.* Epigenomic landscape and 3D genome structure in pediatric high-grade glioma. Science Advances;7:eabg4126
30. Stone KR, Mickey DD, Wunderli H, Mickey GH, Paulson DF. Isolation of a human prostate carcinoma cell line (DU 145). International Journal of Cancer **1978**;21:274-81
31. Jin X, Demere Z, Nair K, Ali A, Ferraro GB, Natoli T*, et al.* A metastasis map of human cancer cell lines. Nature **2020**;588:331-6

32. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adélaïde J, Cervera N*, et al.* Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene **2006**;25:2273-84
33. Yin KB. The Mesenchymal-Like Phenotype of the MDA-MB-231 Cell Line. IntechOpen; 2011.
34. Le AV-P, Szaumkessel M, Tan TZ, Thiery J-P, Thompson EW, Dobrovic A. DNA methylation profiling of breast cancer cell lines along the epithelial mesenchymal spectrum—Implications for the choice of circulating tumour DNA methylation markers. International journal of molecular sciences **2018**;19:2553
35. Tan TZ, Miow QH, Miki Y, Noda T, Mori S, Huang RYJ*, et al.* Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. EMBO molecular medicine **2014**;6:1279-93
36. Panchy N, Watanabe K, Takahashi M, Willems A, Hong T. Comparative single-cell transcriptomes of dose and time dependent epithelial–mesenchymal spectrums. NAR Genomics and Bioinformatics **2022**;4:lqac072
37. Jung AR, Jung C-H, Noh JK, Lee YC, Eun Y-G. Epithelial-mesenchymal transition gene signature is associated with prognosis and tumor microenvironment in head and neck squamous cell carcinoma. Scientific Reports **2020**;10:3652
38. Cook DP, Vanderhyden BC. Transcriptional census of epithelial-mesenchymal plasticity in cancer. Science advances **2022**;8:eabi7640
39. 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. Nature communications **2018**;9:2028
40. Das P, San Martin R, McCord RP. Differential contributions of nuclear lamina association and genome compartmentalization to gene regulation. Nucleus **2023**;14:2197693
41. Stik G, Vidal E, Barrero M, Cuartero S, Vila-Casadesús M, Mendieta-Esteban J*, et al.* CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. Nature genetics **2020**;52:655-61
42. Leemans C, van der Zwalm MC, Brueckner L, Comoglio F, van Schaik T, Pagie L*, et al.* Promoter-intrinsic and local chromatin features determine gene repression in LADs. Cell **2019**;177:852-64.e14
43. Honda A, Hoeksema MA, Sakai M, Lund SJ, Lakhdari O, Butcher LD*, et al.* The Lung Microenvironment Instructs Gene Transcription in Neonatal and Adult Alveolar Macrophages. The Journal of Immunology **2022**:ji2101192
44. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. Nucleic acids research **2017**;45:W98-W102
45. Bachmann C, Schmidt S, Staebler A, Fehm T, Fend F, Schittenhelm J*, et al.* CNS metastases in breast cancer patients: prognostic implications of tumor subtype. Medical Oncology **2014**;32:400
46. Yoneda T, Williams PJ, Hiraga T, Niewolna M, Nishimura R. A Bone-Seeking Clone Exhibits Different Biological Properties from the MDA-MB-231 Parental Human Breast Cancer Cells and a Brain-Seeking Clone In Vivo and In Vitro. Journal of Bone and Mineral Research **2001**;16:1486-95
47. Wang D-Y, Jiang Z, Ben-David Y, Woodgett JR, Zacksenhaus E. Molecular stratification within triple-negative breast cancer subtypes. Scientific reports **2019**;9:1-10
48. Hakim O, Sung M-H, Voss TC, Splinter E, John S, Sabo PJ*, et al.* Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. Genome research **2011**;21:697-706

49. Vieux-Rochas M, Fabre PJ, Leleu M, Duboule D, Noordermeer D. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proceedings of the National Academy of Sciences **2015**;112:4672-7
50. Carone DM, Lawrence JB. Heterochromatin instability in cancer: from the Barr body to satellites and the nuclear periphery. 2013. Elsevier. p 99-108.
51. D'Amato NC, Ostrander JH, Bowie ML, Sistrunk C, Borowsky A, Cardiff RD*, et al.* Evidence for phenotypic plasticity in aggressive triple-negative breast cancer: human biology is recapitulated by a novel model system. **2012**
52. Celià-Terrassa T, Kang Y. How important is EMT for cancer metastasis? PLOS Biology **2024**;22:e3002487
53. Pang QY, Tan TZ, Sundararajan V, Chiu Y-C, Chee EYW, Chung VY*, et al.* 3D genome organization in the epithelial-mesenchymal transition spectrum. Genome Biology **2022**;23:1-31
54. Wang N, Zhang M, Chang Y, Niu N, Guan Y, Ye M*, et al.* Directly observing alterations of morphology and mechanical properties of living cancer cells with atomic force microscopy. Talanta **2019**;191:461-8
55. Scott AK, Rafuse M, Neu CP. Mechanically induced alterations in chromatin architecture guide the balance between cell plasticity and mechanical memory. Frontiers in Cell and Developmental Biology **2023**;11:1084759
56. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P*, et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome biology **2015**;16:1-11
57. Imakaev M, Fudenberg G, McCord RP, Naumova N, Goloborodko A, Lajoie BR*, et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat Methods **2012**;9
58. Bushnell B, Rood J, Singer E. BBMerge–accurate paired shotgun read merging via overlap. PloS one **2017**;12:e0185056
59. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S*, et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics **2013**;29:15-21
60. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. bioinformatics **2015**;31:166-9
61. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR genomics and bioinformatics **2020**;2:lqaa078
62. Deshmukh AP, Vasaikar SV, 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. Proceedings of the National Academy of Sciences **2021**;118:e2102050118
63. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics **2013**;14:1-15
64. Cursons J, Pillman KA, Scheer KG, Gregory PA, Foroutan M, Hediyeh-Zadeh S*, et al.* Combinatorial targeting by microRNAs co-ordinates post-transcriptional control of EMT. Cell systems **2018**;7:77-91. e7
65. Panchy N, Watanabe K, Hong T. Interpretable, scalable, and transferrable functional projection of large-scale transcriptome data using constrained matrix decomposition. Frontiers in Genetics **2021**;12:719099
66. Sigg CD, Buhmann JM. Expectation-maximization for sparse and non-negative PCA. 2008. p 960-7.

Supplementary Figure Legends

Supplementary Figure 1. Concordance between EMT classification of cell types by gene expression and Hi-C compartments. a) Breast cancer cell lines are arranged depending on their epithelial-mesenchymal status. The states are calculated based on their gene expression profile using the

method described in Tan et al. (Tan et al., 2014). This figure has been adapted from (Le et al., 2018), reused as consistent with CC-BY License. Stars indicate cell lines used in the current study. **b)** Scatter plot showing relation between compartment PC1 (from Fig. 1c) and gene expression EMT score (as in Supplementary Figure 1a) for breast cancer cell lines considered in this manuscript. **c)** Example of compartment changes near but not on top of a gene that changes expression across the EMT axis. TFF3 (yellow highlighted position) is downregulated in more mesenchymal like cancer cell lines (top 4) as compared to epithelial-like (bottom 4). The gene itself is in the A compartment (red, positive PC1 values) in all but 1 cell line, but many regions nearby switch to the B compartment in the top 4 vs. bottom 4 cell lines.

Supplementary Figure 2. Conservation vs. change in breast cancer cell compartment identity for regions that differ between breast and lung epithelial cells. a) Out of all genomic bins that are in the A compartment in normal breast epithelial cell lines but B compartment in normal lung epithelial cells, the graph shows the proportion that switch compartments to match lung (green) for each category of breast cancer vs. those that remain like breast (pink). b**)** Same as (a), but for genomic regions that are in the B compartment in normal breast epithelial cells and A compartment in normal lung.

Supplementary Figure 3. Calculation to adjust organ-permissive compartment switch calculation to account for different background levels of certain compartments. a) Calculation of the probability of a genomic region to be in a specific compartment in localized lung cancer given that the region belongs to a certain compartment in normal lung epithelial cell. For the A compartment, p(TA|NA) represents the probability of a region in A compartment in localized lung cancer cell given that region also belongs to A compartment in normal lung epithelial cell. Similarly, p(TB|NB) for the B compartment. **b)** Calculation of the probability of a genomic region to be in a specific compartment in localized brain cancer (SF9427) given that the region belongs to a certain compartment in normal brain cell (NHA).

Supplementary Figure 4 Comparison of Brain-Permissive and Lung-Permissive Changes in Lung Metastatic Breast Cancer. a) Normal human astrocyte (NHA) and glioblastoma (SF9427) cells are used for compartment comparisons to breast cancer changes. **b)** Schematic diagram representing breast cancer brain-permissive changes calculation. Detailed steps are mentioned in Fig. 4b. **c)** Adjusted (see Supplementary Figure 3 and Methods) levels of different brain permissive changes shown by localized and metastatic cancers from different breast cancer subtypes. d**)** Adjusted levels of different lung permissive changes shown by localized and metastatic cancers from different breast cancer subtypes.

Supplementary Figure 5. Prostate cancer cells that metastasize to brain show increased brain-permissive compartment changes at neuronal-related genes. a) Normal epithelial (RWPE), localized prostate cancer (LNCaP), and brain metastatic prostate cancer (DU145) Hi-C datasets are used for compartment comparisons to glioblastoma/normal brain changes. **b)** Schematic diagram representing prostate cancer brain-permissive changes calculation. Detailed steps are mentioned in Fig. 4b. **c)** Adjusted (see Supplementary Figure 3 and Methods) levels of different brain permissive changes shown by localized and metastatic prostate cancer. **d)** Gene Ontology term enrichment for genes in regions in each category of comparisons. Note brain-relevant terms such as regulation of synapse organization in genes switched toward the A compartment in cancer that match brain A compartment and keratinization in genes switched toward the B compartment in cancer that match brain B compartment.

Supplementary Table 1. Classifications of cell line models used and sources of all publicly available Hi-C and RNA-seq data used in the analyses. Information about cell lines derived from the following sources: ATCC, Engel and Young 1978, Gazdar et al., 1998, Cellosaurus.org, Wang et al. 2021, and Stone et al., 1978.

Supplementary Table 2. Lists of genes obtained from compartmental PCA, transcriptomics PCA and curated epithelial-mesenchymal gene set. Compartmental_PC1_Positive: genes found within the top 100 highly varying genomic regions in compartmental PCA analysis (positive PC1); Compartmental_PC1_Negative: genes found within the top 100 highly varying genomic regions in

compartmental PCA analysis (negative PC1); RNAseq_PC1_Positive: top 100 highly varying genes in transcriptomics PCA analysis (positive PC1); RNAseq_PC1_Negative: top 100 highly varying genes in transcriptomics PCA analysis (negative PC1); Curated_Epithelial: curated epithelial genes; Curated_Mesenchymal: curated mesenchymal genes.

Supplementary Table 3. Quantifying overlap between EMT and organotropic compartment switched genomic regions. Each entry represents the fraction of the top 100 positive (left) or negative (right) compartment PC1 regions which are also found in the set of genomic bins that show organotropic compartment switches in the categories shown.