Identification and Pharmacological Targeting of Treatment-Resistant, Stem-like Breast Cancer Cells for Combination Therapy

Jeremy Worley[1,10,†], Heeju Noh[1,†], Daoqi You[2,†], Mikko M. Turunen[1,†], Hongxu Ding[1,3], Evan Paull[1], Aaron T. Griffin[1], Adina Grunn[1], Mingxuan Zhang[1], Kristina Guillan[2], Erin C. Bush[1], Samantha J. Brosius[2], Hanina Hibshoosh[4,5], Prabhjot S. Mundi[1,4], Peter Sims[1], Piero Dalerba[4,5,6,7], Filemon S. Dela Cruz[2], Andrew L. Kung[2], and Andrea Califano[1,4,8,9,10]

[1] Department of Systems Biology, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, USA 10032

[2] Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[3] Department of Pharmacy Practice & Science, College of Pharmacy, University of Arizona, Tucson, Arizona, USA 85721

[4] Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, USA 10032

[5] Department of Pathology & Cell Biology, Columbia University Irving Medical Center, New York, USA 10032

[6] Columbia Stem Cell Initiative, Columbia University Irving Medical Center, New York, USA 10032

[7] Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, USA 10032

[8] Department of Biochemistry & Molecular Biophysics, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, USA 10032

[9] Department of Biomedical Informatics, Vagelos College of Physicians and Surgeons, Columbia University Irving Medical Center, New York, USA 10032

[10] J.P. Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY USA 10032

[†] These authors contributed equally

Correspondence to: kunga@mskcc.org and ac2248@cumc.columbia.edu

Conflict of Interest Disclosure

A.C. is founder, equity holder, and consultant of DarwinHealth Inc., a company that has licensed some of the algorithms used in this manuscript from Columbia University. Columbia University is also an equity holder in DarwinHealth Inc. US patent number 10,790,040 has been awarded related to this work, and has been assigned to Columbia University with Dr. Califano as an inventor. P.D. received stock and/or royalties from Oncomed Pharmaceuticals, Quanticel Pharmaceuticals and Forty Seven Inc., as a result of his acknowledgment as a co-inventor on patents licensed from the University of Michigan (US-07723112) and Stanford University (US-09329170, US-09850483, US-10344094, US-11130813), and related to: 1) the discovery of surface markers for the differential purification of cancer stem cell populations from human malignancies; 2) the use of single-cell genomics technologies for the identification of pharmacological targets expressed in cancer stem cell populations; 3) the combination of anti-CD47 and anti-EGFR monoclonal antibodies for the treatment of human colon cancer. P.D. recently owned stock of Eli Lilly and Company. P.Ds spouse is employed by Regeneron Pharmaceuticals Inc., and owns (or recently owned) stock of the following pharmaceutical companies: AbbVie, Amgen, AstraZeneca, Eli Lilly and Company, Gilead Sciences Inc., GlaxoSmithKline (GSK), Johnson & Johnson, Merck & Co., Novartis, Organon & Co., Pfizer, Teva Pharmaceutical Industries Ltd. and Viatris. A.L.K. is on the Scientific Advisory Board of Emendo Biotherapeutics, Karyopharm Therapeutics, Imago BioSciences, and DarwinHealth; is co-Founder and on the Scientific Advisory Board of Isabl; has equity interest in Imago BioSciences, Emendo Biotherapeutics and Isabl; and receives royalty income from Labcorp.

Abstract

Tumors frequently harbor isogenic yet epigenetically distinct subpopulations of multi-potent cells with high tumor-initiating potential—often called Cancer Stem-Like Cells (CSLCs). These can display preferential resistance to standard-of-care chemotherapy. Single-cell analyses can help elucidate Master Regulator (MR) proteins responsible for governing the transcriptional state of these cells, thus revealing complementary dependencies that may be leveraged via combination therapy. Interrogation of single-cell RNA sequencing profiles from seven metastatic breast cancer patients, using perturbational profiles of clinically relevant drugs, identified drugs predicted to invert the activity of MR proteins governing the transcriptional state of chemoresistant CSLCs, which were then validated by CROP-seq assays. The top drug, the anthelmintic albendazole, depleted this subpopulation *in vivo* without noticeable cytotoxicity. Moreover, sequential cycles of albendazole and paclitaxel—a commonly used chemotherapeutic —displayed significant synergy in a patient-derived xenograft (PDX) from a TNBC patient, suggesting that network-based approaches can help develop mechanism-based combinatorial therapies targeting complementary subpopulations.

Statement of significance

Network-based approaches, as shown in a study on metastatic breast cancer, can develop effective combinatorial therapies targeting complementary subpopulations. By analyzing scRNA-seq data and using clinically relevant drugs, researchers identified and depleted chemoresistant Cancer Stem-Like Cells, enhancing the efficacy of standard chemotherapies.

Introduction

Intratumor heterogeneity represents a major barrier in cancer treatment. Indeed, most tumors comprise co-existing, molecularly distinct subpopulations presenting non-overlapping drug sensitivities[1]. While some of the cells comprising them may represent genetically distinct subclones, a majority has emerged as representing the byproduct of pathophysiological epigenetic plasticity. In breast cancer (BRCA), for instance, there have been multiple reports of an isogenic Cancer Stem-like Cell (CSLC) subpopulation associated with differential expression of epigenetic regulators involved in controlling stemness programs, such as the BMI1, WNT, and NOTCH pathways[2-4]. CSLCs have been shown to display tumor-initiating capacity, expression of stem-cell markers, and resistance to common chemotherapeutics[5,6], such as paclitaxel—a microtubule inhibitor and antimitotic widely used in the treatment of multiple malignancies, including breast cancer. Indeed, while frequently leading to initial tumor shrinkage, treatment with this drug is often followed by relapse and resistance. Indeed, it has been suggested that chemotherapy resistant breast CSLCs may regenerate the full heterogeneity of the tumor, as confirmed by limiting dilution assays[7,8]. Multiple non-mutually exclusive mechanisms of chemotherapy resistance have been proposed for CSLCs in breast and other tumors, including upregulation of multi-drug transporters, increased DNA damage repair, and better scavenging of ROS[9-11]. Taken together, these data suggest that breast CSLCs pose a fundamental challenge to achieving durable remissions in BRCA, especially in Triple Negative Breast Cancer (TNBC), where chemotherapy remains a cornerstone of treatment.

To gain insight into the molecular heterogeneity of breast cancer and to predict the sensitivity of individual subpopulations to clinically relevant drugs, we generated single-cell RNA sequencing (scRNA-seq) profiles of malignant cells isolated from biopsies of seven metastatic breast cancer patients. To enrich for cells with a stem-like phenotype—or CSLCs for simplicity—which may include only a very small fraction of tumor cells, we used fluorescence-activated cell sorting (FACS), with antibodies selected to purify malignant cells with a phenotype analogous to that of stem/progenitor cells in the normal mammary epithelium[12]. See **Fig. 1A** for an illustrative graphical workflow of this process.

In previous studies, we have shown that highly sparse single scRNA-seq profiles, where >80% of the genes may produce no reads, can be transformed to fully populated protein activity profiles by the metaVIPER algorithm[13]—the single-cell adaptation of the extensively validated VIPER algorithm[14]. This is accomplished by measuring the activity of each regulatory and signaling

protein based on the expression of its entire repertoire of transcriptional targets, akin to using a highly multiplexed, tissue-specific gene reporter assay. As a result, the most differentially active VIPER-inferred proteins are also enriched for Master Regulator (MR) proteins representing mechanistic determinants, via their target genes, of the associated transcriptional state.

MetaVIPER analysis of single cells isolated from the seven metastatic breast cancer patients accrued to the study—including five hormone receptor-positive (HR+) and two triple-negative (TNBC) tumors—effectively separated cells with a more stem-like vs. more differentiated transcriptional state, using a *stemness score* (*SS*) based on both established breast cancer stemness markers and CytoTRACE analysis[15]. Consistent with expectations, cells with the highest score (i.e., most stem-like) emerged as the most resistant to *in vivo* treatment with paclitaxel, while those with the lowest score (*i.e.*, most differentiated-like) were significantly depleted by the drug. This provided the molecular basis to identify and genetically/pharmacologically target candidate Master Regulators (MRs) of CSLC transcriptional state(s) identified by metaVIPER analysis.

We thus performed patient-by-patient analysis, using the VIPER algorithm to identify candidate MR proteins controlling the transcriptomic state of cells with the highest vs. lowest Stemness Score. Candidate MRs identified by the analysis were highly conserved across virtually all patients, independent of hormone receptor (HR) status, thus supporting the notion of a common CLSC MR signature. Indeed, >80% of the most significant VIPER-inferred activated and inactivated MRs were able to statistically significantly reprogram cells to a more differentiated or CSLC state, respectively, following their CRISPR-mediated silencing in a pooled CROP-seq[16] assay in cell lines comprising both subtypes. We thus leveraged the OncoTreat algorithm[17], which assesses the activity of MR proteins in drug vs. vehicle control-treated cells, to identify small molecule compounds capable of inverting the activity of the CLSC MR signature (MR-inverter drugs), thus potentially inducing differentiation or selective ablation. For this purpose, we leveraged gene expression profiles of BRCA cells—selected to faithfully recapitulate the CSLC MR signature—treated with a repertoire of 91 clinically relevant drugs, see **Fig. 1B** for an illustrative graphical workflow of these steps. Notably, OncoTreat-predicted drugs from either bulk[17-19] or single-cell profiles[20,21] have been extensively validated *in vivo* in prior studies.

Albendazole, a well-tolerated anthelmintic drug, emerged as the most statistically significant MR-inverter drug, yet at a concentration that was approximately ten-fold lower than its clinically tolerated dose; this was especially surprising since albendazole is not considered an anti-tumor

drug. Based on these results, this drug was selected for experimental validation *in vivo*. Mice from a TNBC PDX model were treated with either albendazole or vehicle control for 14 days and compared to paclitaxel-treated animals. In contrast to paclitaxel, which caused highly significant increase of the CSLC to differentiated cell ratio, albendazole treatment induced equally dramatic yet opposite effects, suggesting that alternating treatment with the two drugs may abrogate the tumor-initiating potential of paclitaxel-resistant cells, while also preventing uncontrolled tumor growth. The strong rationale for combination-based, sequential therapy was confirmed by a preclinical study, where treatment with multiple cycles of albendazole and paclitaxel displayed superior anti-tumor activity compared to the corresponding monotherapies, resulting in a statistically significant synergistic effect.

Results

Intratumor heterogeneity in human breast carcinomas: Since patient-derived breast cancer tissues vary widely in size, cellularity, necrotic fraction, stromal infiltration, and overall quality, we used FACS to purify malignant cells using appropriate antibody combinations. Single cells isolated from these tumors were then processed to generate plate-based scRNA-seq profiles using an approach that combines elements of Smart-seq2[22] and PLATE-seq[23] (see STAR methods). This procedure, which allows sorting individual cancer cells into separate wells filled with lysis buffer for RNA-seq profiling, is especially effective in enriching for relatively rare subpopulations from fresh tumor tissue, since it effectively supports FACS-based cell isolation while removing debris and dead cells that may otherwise degrade the performance of other platforms. It was thus preferred at the time, despite its higher cost and complexity.

Fresh samples were obtained from two metastatic TNBC and five metastatic HR^+ patients. To minimize post-resection transcriptional changes/drift, fresh samples were rapidly dissociated into a single-cell suspension (see STAR methods) and stained with DAPI, as well as α-EpCAM, α-CD49f, and Lin^- antibodies. EPCAM effectively distinguishes epithelial breast cancer cells from stromal subpopulations, whereas CD49f is known to be expressed at the highest levels in a subset of mammary epithelial cells acting as mammary repopulating units (MRUs) in transplantation assays[12] and has been previously used to enrich for breast cancer cells with stem-like properties[24-26]. Starting from primary malignant tissues, we sorted live ($DAPI^-$) epithelial ($EpCAM^+$) cells into two distinct batches, including: (1) a first batch of unselected cancer cells ($EPCAM^+$), representative of the full heterogeneity of the epithelial compartment, contributing ~25% of the total cells in the analysis and (2) a second batch of epithelial cells with a phenotype characteristic

of MRUs in the mouse mammary gland (EPCAM^{+}, CD49f^{high}), expected to be CSLCs-enriched[12,24-26], contributing the remaining ~75% of analyzed cells (**Fig. 1A, Suppl. Fig. S1**).

Copy Number Variation Analysis: After NGS library generation and sequencing (see STAR methods), we performed several data pre-processing steps to ensure that subsequent analyses would be restricted to high-quality cancer cells. This included inference of somatic copy number alteration (CNA) assessment, using the Trinity CTAT Project inferCNV algorithm (https://github.com/broadinstitute/inferCNV) to exclude confounding effects from normal cells in the tumor microenvironment. Compared to cells representative of normal breast epithelium, most of the cells isolated from the seven patients presented clearly aberrant CNA structure, consistent with the high cellularity of metastatic samples (**Suppl. Fig. S2**). Interestingly, no intratumor CNA heterogeneity was detected by the analysis, suggesting that, at least from a copy number alteration perspective, the cells in these samples were clonally identical. However, as expected, the analysis showed significant inter-tumor CNA heterogeneity across the seven patients, especially between TNBC and HR+ samples.

Protein activity-based analysis identifies a stem-like subpopulation: In addition to biological variation between tumors from different patients, substantial batch and biology-related effects may also challenge the analysis of single cells isolated from different samples. Batch effects can arise due to technical artifacts, such as changes in temperature or reagents between samples processed on different days, or liquid handling drift in multi-well plate assays. In addition, inter-patient CNA differences may also contribute to significant gene expression heterogeneity, which may confound the analysis. Indeed, while only a handful of genes in CNAs play a functional role in tumorigenesis, most of the genes in these amplicons may still produce substantial inter-patient bias at the gene expression level, even though the activity of their encoded proteins is ultimately buffered by the post-transcriptional autoregulatory logic of the cell. When combined with the high *gene dropout* rate of scRNA-seq profiles—where >75% of the genes may fail to be detected by even a single read—this limits the ability to perform detailed, quantitative analyses using traditional gene expression-based methodologies.

Various approaches to reduce noise and minimize gene dropout effects have been proposed[27,28]—such as metaCells[29] and imputation-based[30] methods—as well as normalization methods aimed at reducing batch effects[31,32]. These methodologies, however, may introduce artifacts that affect subsequent analyses. For instance, using metaCells may prevent identification of rare subpopulations, whose gene expression profile would be averaged with cells from

molecularly distinct subpopulations, while normalization may reduce biologically relevant differences between samples. Most critically, generating a comprehensive repertoire of candidate molecular determinants of tumor cell state, potentially associated with differential expression of only a handful of genes, is quite challenging if the expression of most genes is undetectable in individual cells. Transcriptional regulators, which are critical in maintaining cell state/identity, are especially affected by such *gene dropout* issues because they can be functionally active even when expressed at very low levels.

To address these challenges, we leveraged the PISCES single-cell analysis pipeline[33], which provides a systematic framework for protein activity-based analysis of single-cell data—from raw counts quality control to construction of gene regulatory networks, to the identification of MR proteins (see STAR methods). Specifically, PISCES leverages the metaVIPER[13] algorithm to measure a protein's differential activity based on the differential expression of its transcriptional targets, as inferred by the ARACNe[34] algorithm. These algorithms have been extensively validated, showing low false positive rates (in the 20% – 30% range)[14,35,36] and almost complete elimination of technical (i.e., non-biologically-relevant) batch effects. In particular, we have recently shown that metaVIPER protein activity measurements significantly outperform gene expression and even antibody-based measurements in single cells[20,37,38], including based on large-scale CITE-seq assays[33].

We used metaVIPER to infer protein activity of single cells isolated from breast cancer biopsies from the two TNBC and five HR+ patients described in the previous section. The relative tumor purity of metastases, combined with EPCAM-based flow cytometry sorting produced single cells that were virtually all tumor related, as shown by the inferCNV analysis (**Suppl. Fig. S2**). As a result, we used metaVIPER to integrate results from both a bulk-level ARACNe network—generated from the TCGA breast cancer cohort—as well as a network generated from the scRNA-seq profiles captured in this study (see STAR methods). This approach allows optimal dissection of tumor cell-specific interactions (from single-cell profiles), while still providing adequate coverage (from bulk profiles) of the transcriptional targets of regulatory proteins that are undetectable in single-cell profiles.

MetaVIPER computes the normalized enrichment score (NES) of a protein's targets in genes differentially expressed between each individual cell and a reference state, typically the centroid of the entire single-cell population (see STAR methods). As a result, positive and negative NES scores indicate higher and lower protein activity compared to the average of the single-cell

population, respectively. While VIPER is most effective in assessing the activity of regulatory proteins, we have shown that it can quantitate the differential activity of signaling proteins[14,39] and surface markers[37] with similar accuracy. As a result, we included 339 cell surface markers and 3,407 signaling proteins in the analysis (see STAR methods for selection criteria).

Due to the large-scale CNA differences detected by the analysis, inter-patient heterogeneity was highly dominant at the gene expression level, with almost each patient contributing to an independent cluster in a Principal Component Analysis (PCA) representation, using the 5,000 genes with the highest standard deviation (**Fig. 2A**). In contrast, since VIPER-inferred protein activity is robust to noise and resilient to technical artifacts that are inconsistent with the underlying regulatory network[13] (**Fig. 2B**), protein activity-based PCA analysis virtually eliminated inter-patient variability, except when biologically relevant (**Fig. 2C**). For instance, differences linked to HR status were captured by the second principal PCA component (y-axis), which accounts for 15% of cross-cell variability. Yet, the most significant source of variance, accounting for 31% of cross-cell variability, was captured by the first PCA component (x-axis), which could be associated with high vs. low stemness (**Fig. 2D**).

Cell stemness was assessed using two complementary metrics, including (a) the global activity of established breast CSLC markers and (b) CytoTRACE[15], an experimentally validated algorithm designed to infer stemness based on gene count signature analysis (**Fig. 2D**, see STAR methods). CytoTRACE was previously validated within a hematopoietic lineage context and is based entirely on assessing expressed gene counts (a rough measure of cell entropy) rather than specific knowledge of stem cell biology. As a result, it has shown limitations, for instance, in differentiating quiescent stem cells from cycling progenitor cells[15]. To address this issue we complemented and compared the CytoTRACE analysis with biologically-relevant insights derived from the VIPER-measured activity of 14 previously reported CSLC markers, including CD44+/CD24-[40], ITGA6 (CD49f)[26], BMI1[4], SALL4[41], NOTCH1[42], NOTCH2[42], KLF4[43], CTNNB1[44], ITGB3 (CD61)[45,46], ITGB1[47], PROM1 (CD133)[48], POU5F1 (OCT4)[49], SOX2[50], and KIT[51], resulting in a *consensus Stemness Score*, ranging from $SS = 0$ (most differentiated) to $SS = 1$ (most CSLC), shown as a color gradient in **Fig. 2D** (see STAR methods). Supporting the use of such consensus metric, the CytoTRACE and CSLC marker-based scores were highly correlated despite being assessed by completely independent methodologies (Spearman's $\rho = 0.43$, $p \leq 2.2\times10^{-16}$) (**Suppl. Fig. S3**).

Despite the potential noisy nature of single-cell data, the PCA plot region comprising CD49f^{high} cells was strongly associated with high activity of other established markers of stem-like function in mammary epithelial cells, such as BMI1[4] and NOTCH1/2[42], among several others, critically in both TNBC and HR+ derived cells (**Fig. 2E, Suppl. Fig. S4**). Consistent with the literature[48,52-54], activity of additional stemness markers such as PROM1, POU5F1, SOX2, and KIT was also more prominent in CD49f^{high} cells from TNBC patients (**Suppl. Fig. S4**). Differential activity of metabolic CSLC markers, such as ALDH1[55], was not detectable, likely because these enzymes are less related to transcriptional regulation.

In sharp contrast to VIPER-based analyses—and fully consistent with prior studies, see[20,37,38] for instance—the expression of genes encoding for these markers was mostly uninformative and failed to provide insight into CSLC characterization, because of the drastic gene dropout effect associated with scRNA-seq profiles (**Suppl. Fig. S5**). For instance, despite having a clear readout at the protein activity level, CD44, ITGB3, and SOX2 generated virtually no reads, thus preventing meaningful assessment of their differential expression, while expression of most other markers could not be associated to specific regions of the PCA plots.

Differential activity of subtype-specific markers was also evident for cells isolated from HR+ vs. TNBC patients, especially within the differentiated cell compartment. For instance, the activity of luminal markers, such as GATA3[56,57], FOXA1[58], the estrogen (ESR1) and progesterone (PGR) receptors, was markedly higher in differentiated HR+ derived cells (**Fig. 2F, Suppl. Fig. S6A**), while the activity of TNBC markers, such as FOXC1[59,60] and BCL11A[61], as well as basal cytokeratin (KRT17), and vimentin (VIM)[62,63], was higher in differentiated TNBC derived cells (**Fig. 2F**, **Suppl. Fig. S6B**). To provide an objective baseline we leveraged KRT19, an established marker of luminal differentiation, whose NUMB-mediated interaction with WNT/NOTCH pathways is well documented[64,65] and whose differential protein activity and differential gene expression could be effectively assessed in single cells. Indeed, differential expression of *KRT19* was highly consistent with metaVIPER-measured KRT19 activity (**Suppl. Fig. S7A-C**), confirming VIPER-based identification of luminal vs. basal cells. Compared to other cancers, such as colon cancer[65], KRT19 holds special relevance in breast cancer, where its attenuated expression is strongly associated with poor prognosis and stemness[64,65]; consistent with these findings, KRT19 activity was also significantly lower in the PCA region associated with highest stemness (**Fig. 2F**).

VIPER-inferred CSLCs are insensitive to paclitaxel: Rather than assessing self-renewal and multipotency as characteristics of *bona fide* CSLC state—still a rather controversial topic—we

focused on the more pragmatic and objective assessment of the differential sensitivity to paclitaxel by cells identified as CSLC by our analysis, which presents critical relevance to patient treatment. For this purpose, we analyzed single cells dissociated from PDX models established by transplantation of a human primary TNBC in the mammary fat pad of immunodeficient NOD/SCID/IL2R$\gamma^{-/-}$ (NSG) mice, which were treated with either vehicle control or paclitaxel for 14 days after reaching a tumor volume of 100 mm^3 (**Fig. 3A**; see STAR methods).

First, we assessed the fidelity of PDX-derived, single-cell subpopulations to those dissociated from human samples. Single-cell analysis of a vehicle control-treated mouse confirmed prior findings from patient-derived samples. Specifically, based on protein activity analysis with metaVIPER, the 1st principal component (PC1) was again associated with cell differentiation and significantly correlated with both CytoTRACE score (Spearman's $\rho = 0.65$, $p \leq 2.2\times10^{-16}$, **Suppl. Fig. S8A-B**) and with overall activity of the 14 CSLC markers ($\rho = 0.90$, $p \leq 2.2\times10^{-16}$, **Suppl. Fig. S8C**). More importantly, there was a highly significant overlap of proteins differentially active in cells with the highest vs. lowest Stemness Score in PDX vs. human samples, as evaluated by GSEA analysis (OncoMatch algorithm[18]) (NES = 7.97, $p = 1.6\times10^{-15}$). Finally, based on GSEA analysis of MSigDB hallmarks[66], genes encoding for proteins associated with the 1st PC were highly enriched in hallmarks associated with cell developmental processes such as epithelial-mesenchymal transition and myogenesis ($p = 3.4\times10^{-4}$ and $p = 1.9\times10^{-3}$, respectively) as well as PI3K-AKT-mTOR[67] ($p = 9.8\times10^{-4}$), KRAS[68] ($p = 2.0\times10^{-3}$), and P53[69] ($p = 2.0\times10^{-3}$) pathways (**Suppl. Table 1**).

Consistent with data from primary tumor tissues, differential expression of most CSLC markers in single cells isolated from PDX tissue was not informative or undetectable (**Suppl. Fig. S9**). However, at the protein activity level, the PCA regions with the highest activity of different CSLC markers—including CD49f, BMI1, CD44$^+$/CD24$^-$ and NOTCH1/2—were largely overlapping in both human and mouse samples (**Fig. 3B**). Putative CSLCs from PDX samples (i.e., with highest Stemness Score) also presented high activity and expression of the established quiescent breast CSLC marker BIRC5[70] (Spearman's $\rho = 0.53$, $p \leq 2.2\times10^{-16}$, **Suppl. Fig. S10**) and lower activity and expression of E2F family proteins ($\rho = -0.69$, $p \leq 2.2\times10^{-16}$, **Suppl. Fig. S11**), which transactivate genes for G1/S transition[71]. These differences were likely more evident in PDX samples because of faster growth kinetics, as compared to primary human tumors. These data suggest that CSLC are more quiescent than differentiated cells, thus providing additional rationale

for their paclitaxel resistance. Taken together, these data characterize the PDX as a high-fidelity model to study CSLC vs. differentiated cells[18].

Changes in CSLC vs. differentiated cell density following drug treatment were then assessed by computing the normalized ratio between the number of cells with the highest (SS $\geq$ 0.8, most CSLC) and lowest (SS $\leq$ 0.2, most differentiated) Stemness Score in paclitaxel vs. vehicle control-treated samples, see STAR methods. Paclitaxel treatment induced striking depletion of differentiated cells vs. CSLCs (**Fig. 3C**) ($p = 2.6\times10^{-4}$, by Fisher's exact test), thus confirming the expected paclitaxel resistance of CSLC compartment cells identified by VIPER analysis.

Since the PDX was derived from a TNBC tumor, the 2nd PC could not be associated with HR status, as shown instead across the original 7 patient-derived samples. Rather, GSEA analysis revealed enrichment in two key categories, including cellular responses to DNA damage and oxidative stress, two hallmarks of paclitaxel mechanism of action ($p = 1.9\times10^{-11}$ and $p = 2.4\times10^{-12}$, respectively, by GSEA) (**Suppl. Table 1**)[72-74]. Indeed, the cells that were least affected by the drug were those presenting both high stemness score and a low proliferative potential (upper right quadrant on the PCA plot). Yet, for any given value of the PC2 metagene, predicted CSLC were always less sensitive to treatment than their differentiated counterpart. Indeed, the density of cells with the highest stemness score was virtually unaffected by treatment.

MR Analysis of human breast cancer cells. VIPER analysis has been effective in identifying candidate MR proteins representing mechanistic determinants of cell state[75,76], as well as clinically validated biomarkers [77-81], see[82] for a recent perspective. Critically, we have shown that VIPER-inferred MRs are highly enriched in tumor-essential genes[75,76,83], such that their pharmacologic targeting can abrogate tumor viability *in vivo*[17-19]. Equally important, we have shown that genetic or pharmacologic targeting of MRs that are differentially active in molecularly distinct transcriptional states can effectively reprogram cells between these states[20,84,85]. This suggests that elucidating candidate MRs of breast CSLC state may help identify drugs that either selectively ablate paclitaxel-resistant cells or reprogram them to a paclitaxel-sensitive state, thus providing a rationale for combination therapy.

To discover the most conserved CSLC MRs across the available metastatic samples, we first leveraged metaVIPER to identify proteins whose transcriptional targets were most differentially expressed in the 20 cells with the highest vs. the 20 with the lowest Stemness Score in each individual patient, as well as in the PDX model, on an individual sample basis (see STAR

methods). As discussed, the most differentially active proteins are also those expected to be most likely to mechanistically regulate the cell state of interest, via their transcriptional targets. As previously shown[18,19], the PDX model was included in the analysis to help prioritize MR-inverter drugs that are conserved in a model that may be leveraged for drug validation *in vivo*.

As discussed, CytoTRACE was originally developed and validated only in a hematopoietic linage context[15]. As a result, for MR elucidation purposes, we decided to rely only on the differential activity of the 14 CSLC markers, including CD44+/CD24-, ITGA6, BMI1, SALL4, NOTCH1, NOTCH2, KLF4, CTNNB1, ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT (see STAR methods). Indeed, while the enrichment of breast CSLC and stem-related markers in differentially active protein was still significant when CSLC were predicted by CytoTRACE analysis (NES = 2.57, $p = 10^{-2}$), statistical significance increased substantially when relying only on the established CSLC markers (NES = 4.66, $p = 3.2\times10^{-6}$). Nevertheless, confirming that this choice has only minimal effects on MR analysis, statistically significantly MR proteins ($p \leq 10^{-3}$, Bonferroni corrected) were highly overlapping when CytoTRACE was included or excluded from the analysis ($p \leq 1.2\times10^{-44}$, by hypergeometric test).

Surprisingly, independent analysis of each patient and of the PDX model produced highly consistent MR predictions, including across HR+ and TNBC samples (**Fig. 4A, Fig. S12A-B**), suggesting that CSLC MR proteins are conserved independent of tumor HR status. This provided the rationale for the generation of a consensus CSLC MR signature, obtained by ranking all proteins by integrating their metaVIPER NES across all samples, using the weighted Stouffer's method (**Fig. 4B**, see STAR methods). Based on this analysis, in addition to the original 14 CSLC markers, other proteins broadly associated with stem cell processes—including ALDH family[86,87], ABC family[87], quiescent stem-cell markers (FGD5[88] and HOXB5[89]), embryonic diapause[90] and asymmetric cell division processes[91] (**Suppl. Table 2**)—also emerged as significantly enriched among the most differentially active proteins ($p = 2.0\times10^{-12}$) (**Suppl. Fig. S13**).

These results suggest that several of the most statistically significant differentially active proteins, not previously associated with breast CSLCs, may represent novel, *bona fide* MRs and potential biomarkers (**Fig. 4C and Fig. S14-18**, see also **Suppl. Table 3**), as later confirmed by CRISPR/Cas9-mediated KO (see next section). Among cell membrane-presented proteins, which may be leveraged for CSLC enrichment purposes, the analysis identified Integrin beta-8 (ITGB8) as the second most differentially active protein (after CD49f). ITGB8 was previously suggested as a marker of glioblastoma CSLCs[92] and was identified as a prime receptor binding a latent

complex of transforming growth factor beta 1 and beta 3 (TGF-β1/β3) in the extracellular matrix, responsible for activating TGF-β-associated signaling. Despite its role in tumor suppression in the early stages of tumorigenesis, TGF-β has been shown to prompt stem-like properties in advanced cancers and to increase chemotherapy resistance by promoting DNA damage response pathway activation[93-95].

MR Modularity Analysis: A key question in network-based analyses is whether—similar to what has been shown in other contexts[75,76,96]—candidate MRs may comprise hyper-connected, autoregulated modules providing coordinated, homeostatic cell state regulation. For this purpose, we assessed whether metaVIPER-inferred CSLC MRs were statistically significantly enriched in protein-protein and transcriptional interactions—as reported in PrePPI[97], STRING[98], and ARACNe-based networks—compared to an equivalent number of same-class proteins selected at random. The analysis revealed that the top 20 CSLC MRs form a highly hyperconnected module, with 67 MR-MR interactions, compared to only 13.2 detected on average in an equal size set of randomly selected proteins ($p = 6.6\times10^{-7}$). This supports the potential role of this module as a homeostatic On/Off switch controlling CSLC state (**Fig. 4D**), further suggesting that its inactivation may induce transition toward a more differentiated, paclitaxel-sensitive state.

CSLC MR validation by pooled, CRISPR-KO-mediated CROP-seq analysis: To validate the CSLC MRs inferred by these analyses, CRISPR droplet sequencing (CROP-seq) was used to assess whether KO of the 25 most significant MR of CSLC state (MR_{CSLC}, i.e., most active proteins in CSLC vs. differentiated cells) and 25 most significant MRs of differentiated state (MR_{DIFF}, i.e., most active proteins in differentiated cells vs. CSLCs) would induce reprogramming towards a more or less differentiated cell state, respectively. To optimally assess reprogramming, we selected two breast cancer cell lines that most effectively recapitulate the CSLC state, also assuming that all cell lines comprise differentiated cells. For this purpose, we assessed the enrichment of proteins in the consensus CSLC MR signature in proteins differentially active in each CCLE breast cancer cell line (based on bulk RNA-seq analysis), and ranked them from the one with the highest NES (HCC1143)—*i.e.*, most likely to be enriched in CSLCs—to the one with the most negative NES (VP229)—*i.e.*, most likely to be enriched in differentiated cells—(**Suppl. Fig. S19**). We then selected two of the most CSLC-enriched cell lines for CROP-seq assays, including HCC1143 (ranked No. 1) and HCC38 (ranked No. 3), which were also supported by literature evidence on CSLC content[99,100]. Single-cell analyses confirmed that both cell lines had substantial CSLC representation, compared to two of the most differentiated cell lines (MCF7 and

HCC2157), with HCC1143 presenting a greater fraction of differentiated cells compared to HCC38, potentially due to spontaneous differentiation in culture conditions (**Suppl. Fig. S20**).

The primary objective of CRISPR-Cas9-mediated gene knockout (CRISPR-KO) is to abrogate the function of the target protein. While it may reduce transcript copy number through mechanisms like nonsense-mediated decay, this effect is inconsistent and not generally detectable[101]. Therefore, we assessed KO efficiency based on VIPER-mediated analysis of the target protein in cells harboring the associated targeting guide RNAs (sgRNA) vs. intergenic control sgRNAs (see STAR methods). For each MR, we used 3 distinct sgRNAs and disregarded the effect of sgRNAs detected in < 10 cells. This allowed computing the effect of CRISPR/Cas9-mediated MR-KO on cell state, using the scRNA-seq profile of 10 or more cells containing the same targeting sgRNA, compared to cells harboring intergenic control sgRNAs. We then plotted the resulting effect on cell state reprogramming in HCC38 and HCC1143 cells by integrating across all positive and negative MRs of CSLC state (**Fig. 5A)**, as well as on an MR-by-MR basis (**Fig. 5B**). The expectation is that KO of positive and negative MRs will induce reprogramming towards a more or less differentiated state, respectively, as assessed by Stemness Score analysis. To avoid biasing the analysis, the MR directly targeted by a sgRNA in each cell was excluded from the Stemness Score assessment, such that only its downstream effectors were considered (see STAR methods).

Based on Stemness Score analysis and fully consistent with predictions, MR_{CSLC} KO induced significant shift of HCC38 cell state towards a differentiated state ($p = 1.2\times10^{-3}$, by Mann Whitney U Test). Given the small fraction of differentiated cells in this cell line (**Suppl. Fig. S20**), however, MR_{DIFF} KO did not induce significant shift towards a CSLC state. In contrast, both MR_{DIFF} KO and MR_{CSLC} KO induced significant reprogramming towards a CSLC ($p = 5.8\times10^{-4}$) and differentiated state ($p = 3.3\times10^{-2}$), respectively, in HCC1143 cells, which comprise a more balanced ratio of CSLC and differentiated cells (**Suppl. Fig. S20**). When enrichment in genes associated with stem cell process-related genes (i.e., not breast cancer-specific) was considered (see **Suppl. Table 3**) the same statistically significant trends were observed (**Suppl. Fig. S21**).

In summary, CROP-seq analysis produced highly consistent results in both cell lines, confirming the predicted role of most VIPER-inferred MRs. Note that the statistical significance of this analysis is quite underestimated, because both cell lines include a mixture of CSLC (low MR_{DIFF} and high MR_{CSLC}) and differentiated cells (high MR_{DIFF} and low MR_{CSLC}), while MR KO-mediated

effects can only be assessed in cells with high MR activity. As a result, the number of validated MRs is also likely to be underestimated.

The library-normalized differential abundance of sgRNA guides targeting positive MRs was not statistically significant compared to control sgRNAs (**Suppl. Fig. S22**), confirming that these MRs have no effect on cell viability or proliferation. In contrast, differential abundance of sgRNAs targeting negative MRs was significantly lower (**Suppl. Fig. S22**), suggesting that the latter—which includes cell proliferation and viability regulators—may include more essential proteins.

The contribution of each individual MR to cell state reprogramming was then analyzed and is shown in **Fig. 5B**. For the 25 MR_{CSLC} and 25 MR_{DIFF} tested in this assay, we only considered sgRNAs inducing effective MR KO, based on the above-described criteria. As a result, only 16 of 25 candidate MR_{SCLC} (BMPR1A, MTDH, ZNF131, MAML3, GON4L, ZNF24, SMAD5, KLF3, UBP1, SMAD1, TMF1, XBP1, MIER1, VEZF1, ETV3, ZNF566, underlined are statistically significant at $p \leq 0.05$, FDR corrected) and 9 of 25 candidate MR_{DIFF} (PCBD1, RUVBL2, HDGF, RPS3, RORC, ENY2, PEX14, THAP8, PARK7) could be evaluated in HCC38. Similarly, in HCC1143 cells, only 15 of 25 MR_{SCLC} (STAT3, BMPR1A, MTDH, ZNF131, GON4L, MYBL1, SMAD5, UBP1, NCOA1, SMAD1, TMF1, XBP1, VEZF1, ETV3, ZNF566) and 11 of 25 MR_{DIFF} (PCBD1, RUVBL2, HDGF, PRDX2, YBX1, RORC, LAMTOR5, ENY2, THAP8, HLX, PARK7) could be evaluated.

In summary, of 16 and 15 MR_{CSLC} tested one or both cell lines, 15 (94%) and 10 (67%) were validated in at least one or both cell lines ($p \leq 0.05$, FDR corrected), respectively. Similarly of 9 and 11 MR_{DIFF} tested one or both cell lines, 4 (44%) and 8 (73%) were validated in at least one or both cell lines ($p \leq 0.05$, FDR corrected), respectively.

CRISPR-mediated KO of the 5 most activated candidate MR_{SCLS} proteins, by VIPER analysis, identified 2 (BMPR1A and ZNF141) capable of inducing highly significant ($p \leq 8.0\times10^{-24}$ and $p \leq 3.5\times10^{-7}$, respectively for HCC38 and $p \leq 2.1\times10^{-24}$ and $p \leq 4.2\times10^{-6}$, respectively for HCC1143 after FDR correction) Stemness Score decrease in both cell lines, confirming their mechanistic role in CSLC state regulation. Among these, ZNF131 was the only one previously associated with essentiality in these cell lines (gene dependence score = -1.76 for HCC38 and –2.16 for HCC1143 by CERES [102], a copy-number correction method for computing gene essentiality). Indeed, ZNF131 KD-mediated centrosome fragmentation and cell viability decrease were previously reported in GBM[103]. This raises an important question related to the potential role of ZNF131 as

a CSLC-specific essential gene in breast cancer. Similarly, CRISPR-mediated KO of the 5 most inactive candidate MR_{DIFF} proteins, by VIPER analysis, identified PDBD1 capable of inducing statistically significant ($p \leq 0.022$ for HCC38 and $p \leq 7.4\times10^{-13}$ for HCC1143, FDR corrected) Stemness Score increase in both cell lines. Taken together, this confirms that VIPER-inferred MRs are highly enriched in mechanistic, causal determinants of CSLC state rather than pure gene/phenotype statistical associations.

Identification of drugs able to invert stem-like MR programs. The high validation rate of VIPER-inferred MRs in the CROP-seq analysis suggests that MR-inverter drugs capable of inhibiting and activating the most positive and negative MRs, respectively, should induce CSLC differentiation, thus increasing their sensitivity to chemotherapy. Indeed, MR-mediated reprogramming of cell state has already been validated in multiple contexts, from de-differentiation[84], to reprogramming[96,104] and trans-differentiation[85,105]. For this purpose, we leveraged the OncoTreat algorithm, which has proven highly effective in discovering MR-inverter drugs that were extensively validated *in vivo*, based on MR proteins inferred by VIPER analysis of both bulk[17-19] and single-cell profiles[20,21].

OncoTreat relies on perturbational RNA-seq profiles representing the response of cells—selected based on their ability to phenocopy the MR activity signature of interest—to treatment with multiple drugs and vehicle control. Perturbational profile analysis, using VIPER, allows measuring the differential activity of each MR in drug vs. vehicle control-treated cells thus providing a quantitative assessment of the activity inversion across the entire MR-signature. For this purpose, we used previously generated perturbational profiles in the BT20 BRCA cell line, which strongly recapitulates the consensus CSLC MR signature (6th most significant among 62 BRCA cell lines in CCLE, (NES = 7.3 by enrichment analysis), **Suppl. Fig. S22A-C**). Specifically, BT20 cells were profiled at 24h following treatment with 90 clinically relevant drugs, including FDA-approved, late-stage experimental oncology drugs (i.e., in Phase II and III clinical trials) and other selected drugs[23] (**Fig. 6A; Suppl. Tables 4,5**). Transcriptional profiles were generated using PLATE-seq[23]—a fully automated 96- and 384-well, microfluidic-based technology that is highly efficient and cost-effective—at an average depth of 2M reads. To optimize elucidation of drug mechanism of action (MoA), rather than activation of stress or death pathways, drugs were titrated at 1/10th of their EC_{50} concentration, based on 10-point dose response curves[17].

Analysis of proteins that were differentially active in drug vs. vehicle control-treated cells identified five protein clusters (M1 – M5) that were consistently activated or inactivated in response to

different drug subsets. These were significantly enriched in five main Gene Ontology (GO) pathways, including RNA splicing/Ribosome biogenesis (M1), Epigenetic modification/DNA methylation (M2), Cell cycle/Apoptosis (M3), Cellular response to steroid hormone stimulus and Stem cell population maintenance (M4), and Cell differentiation/Development (M5), respectively (**Fig. 6A; Suppl. Table 4**). Notably, drugs inducing activation or inversion (*i.e.*, positive or negative NES) of breast CSLC MRs had opposite effects on the M4/M5 vs. M1/M2/M3 modules. Specifically, M5 proteins, which were associated with differentiation and developmental processes, were significantly activated by the drugs inducing strongest inversion of CSLC MR activity. In contrast, the drugs predicted to further activate the CSLC MR signature induced activation of M4 proteins, associated with stem cell population maintenance.

Among the 17 statistically significant MR-inverter drugs predicted by OncoTreat ($p \leq 0.05$, FDR corrected), the anthelmintic drug albendazole emerged as the most significant one ($p = 4.0\times10^{-4}$) (**Fig. 6B; Suppl. Table 6**).

Albendazole validation in vivo: To experimentally validate albendazole's ability to deplete the CSLC compartment in breast cancer, we extended the protocol used to study paclitaxel in PDX models to assess the effect of 14-day treatment *in vivo* with albendazole vs. vehicle control treatment, at the single-cell level. For these *in vivo* studies, albendazole was used at 1/3rd of its maximum tolerated dose in mice, consistent with assessment of MR-inversion potential at low concentration. Although albendazole is not an oncology drug, it has been shown to inhibit growth of some cancer cell lines and of a murine carcinoma, reportedly by inducing oxidative stress[106-108]. Consistently, albendazole clustered separately from chemotherapeutic drugs (**Fig. 6A**), and its activity was associated with activation of cell differentiation pathways (**Fig. 6A**).

Consistent with the paclitaxel analysis, depletion of CSLC vs. differentiated cell compartment was computed by measuring the ratio between the number of cells with the highest ($SS \geq 0.8$) vs. lowest ($SS \leq 0.2$) stemness score in albendazole vs. vehicle control-treated samples, normalized to the subpopulations size (see STAR methods). Whereby paclitaxel had induced dramatic increase in this ratio, indicating relative depletion of the differentiated tumor cell compartment (**Fig. 3C**), albendazole had the opposite effect (**Fig. 7A**), producing equally significant relative depletion of the breast CSLC compartment ($p = 2.0\times10^{-4}$, by Fisher's exact test). When comparing albendazole to paclitaxel-treated tumors, relative changes in the density of the two compartments

were even more statistically significant ($p = 3.0\times10^{-12}$, by Fisher's exact test) (**Fig. 7B**), suggesting a highly complementary effect.

Albendazole synergizes with paclitaxel in a TNBC PDX model. Since albendazole and paclitaxel deplete complementary metastatic breast cancer cell compartments, it is reasonable to hypothesize that combining or alternating their administration may outperform either drug used as monotherapy. To test this hypothesis, we evaluated whether CSLC compartment depletion by repeated administration of albendazole would enhance the *in vivo* anti-tumor activity of paclitaxel.

A PDX line, established from a human primary TNBC, was implanted in the mammary fat pad of NSG mice. When tumors reached a volume of 100 mm^3, they were randomly enrolled to receive different treatments (paclitaxel monotherapy, albendazole monotherapy, albendazole + paclitaxel, and vehicle control) until six mice per arm were enrolled. Mice in the combination arms underwent two treatment cycles, separated by a 15-day drug holiday. Each cycle included albendazole-based sensitization for two weeks, starting at Day -13—defined as the day when a specific tumor reached a volume of 100 mm^3—followed by three paclitaxel treatments (Day 1, 8 and 15) (**Fig. 7C**). For monotherapy treatment, mice were treated for the same amount of time and on the same schedule with albendazole, paclitaxel, and vehicle control, independently.

Paclitaxel monotherapy significantly reduced relative tumor volume (TV), compared to vehicle control ($p = 0.0024$), while albendazole was indistinguishable from vehicle control ($p = 0.21$) (**Fig. 7D; Suppl. Fig. S23**). TV change was assessed from initiation of albendazole therapy (Day -13) through Day 49; during this period, the majority of vehicle control-treated animals (n = 5 of 6) required euthanasia, due to attaining the maximal allowed humane TV endpoint (median TV = 1543 mm^3). Additionally, compared to vehicle control, albendazole monotherapy showed no significant improvement in disease control ($p = 0.83$) or overall survival ($p = 0.63$) (**Suppl. Fig. S24**).

In sharp contrast, the albendazole + paclitaxel combination was associated with profound suppression of tumor growth, compared to both vehicle control ($p = 1.7\times10^{-4}$) and paclitaxel monotherapy ($p = 0.015$) (**Fig. 7E**). Drug synergy was further confirmed by Bliss independence analysis ($p = 9.0\times10^{-3}$) and translated into a statistically significant increase in overall survival ($p = 0.02$) (**Suppl. Fig. S24**).

Discussion

Despite remarkable therapeutic advances, the prognosis for metastatic breast cancer patients remains dismal. Among the most critical obstacles to achieving a permanent eradication of the disease is the heterogeneity of tumor cell response to therapy. Indeed, while many chemotherapies and targeted therapies may be highly effective on subpopulations that contribute to the bulk of the malignant tissue, the presence of drug-resistant subpopulations within the same tumor mass inevitably leads to relapse and poor outcome. The cellular heterogeneity associated with pre-existing differential drug sensitivity can be of a genetic origin, for instance due to mutations in the active site of the target protein[109] or to the presence of clonally distinct subpopulations with bypass or alternative mutations[110]. However, it is more often associated with the presence of epigenetically distinct transcriptional states with differential drug sensitivity—either pre-existing[1] or induced by cell adaptation[111,112]—some of which can plastically regenerate the full heterogeneity of the tumor[113]. This is especially relevant in the metastatic context, where tumors have already reached a high degree of heterogeneity, due to paracrine interaction differences at distinct distal sites. Consistently, progression to metastatic breast cancer dramatically reduces the probability of achieving complete and durable responses. Indeed, most metastatic breast cancer patients rapidly progress through multiple lines of anti-tumor treatment, and eventually end up receiving conventional chemotherapy, which typically provides only short-term control of the disease.

A growing body of evidence suggests that less differentiated breast cancer cells may be chemotherapy resistant, while retaining the ability to further differentiate and reconstitute the full heterogeneity of the tumor. These cells may thus play a key role in relapse to drug-resistant disease. Tumor cells with stem-like properties (CSLCs) and tumor initiating potential were first discovered in leukemia[114,115] and later reported also in solid tumors, such as gliomas[116,117], breast[118], and colon cancer[119]. As a result, the identification of novel therapeutic approaches to specifically target the CSLC compartment represents a potentially impactful area of investigation[120-122] and may help identify drugs that synergize with chemotherapy. Network-based, single-cell analysis of cells dissociated from metastatic breast cancer patients identified a well-defined transcriptional state controlled by an exceedingly conserved repertoire of MR proteins—including transcription factors and co-factors previously associated with mammary repopulation units and breast cancer stem cells—whose sensitivity to chemotherapy is dramatically reduced compared to differentiated breast cancer cells. Indeed—based on a consensus Stemness Score that combines both the CytoTRACE metric and the activity of 14 established BRCA stemness

marker proteins—there was highly significant association between cell stemness and chemotherapy resistance. This helped us identify a molecularly distinct subpopulation of chemotherapy resistant, poorly differentiated cells (CSLC for simplicity), based on the highly conserved repertoire of MR proteins that control their transcriptional state, across virtually all patients in the study. While this definition may encompass previously reported breast cancer stem cells, we use the term CSLC more broadly as it may also include an additional repertoire of undifferentiated, chemotherapy resistant progenitors. Thus, we make no claims that the CSLCs identified by our analysis represent *bona fide* tumor stem cells; rather, we show that they are chemotherapy resistant and would thus benefit from complementary therapeutic options. To enrich for CSLCs, we leveraged CD49f-based flow cytometry-based sorting of single cells dissociated from patient-derived samples. While CD49f is considered a marker of basal cells and is most highly expressed in a subset of cells from TNBC samples, previous results[24-26] and our analysis confirmed that CD49f is also differentially expressed in CSLCs from HR+ patients. Indeed, its expression gradient was significantly correlated with the activity of 14 previously reported BRCA CLSC markers across all patients in the study, independent of HR status, thus justifying its use in our study. Confirming the value and accuracy of the proposed protein activity assessment methodology, CD49f was identified as significantly differentially active by metaVIPER in cells dissociated from human samples (**Fig. 2E**), even though its encoding gene, *ITGA6*, could not be identified as differentially expressed (**Suppl. Fig. S5**). This is fully consistent with the fact that these cells were FACS sorted with and without the associated antibody and highlights the limitations introduced by gene dropout effects in scRNA-seq profiles.

Targeting the CSLC compartment may be accomplished by developing drugs that either preferentially kill these cells or reprogram them toward treatment-sensitive states. The latter strategy is supported by recent results in fields ranging from hematopoiesis, cancer, and diabetes[84,85,105,123] where genetic or pharmacologic targeting of MR proteins—as identified by network-based VIPER/metaVIPER analyses—effectively reprogrammed the cell's transcriptional state towards a different target state, thus also confirming their nature as mechanistic determinants of cellular state transitions. An additional advantage of these approaches is that metaVIPER analysis effectively removes technical artifacts (batch effects) and non-functional gene expression differences, for instance due to inter-tumor CNA heterogeneity[13,38], thus resulting in highly reproducible identification of MR proteins across samples from different patients.

To confirm mechanistic control of the CSLC state by metaVIPER-inferred MRs we performed pooled CRISPR/Cas9-mediated KO of candidate MRs in two cell lines, followed by scRNA-seq

profiling, using the CROP-seq methodology. As shown, following CRISPR/Cas9-mediated KO, the vast majority of positive and negative CSLC MRs identified by metaVIPER analysis induced statistically significant reprogramming towards either a more differentiated or a more CSLC state, respectively, thus confirming the algorithm's predictions. This includes four of the top five candidate MRs that had been previously nominated as potential players in CSLC biology but had not been experimentally validated, including STAT3[124,125], MTDH[126], ARID1A[127], BMPR1A[128], and ZNF131[103], the first two of which had been proposed as key (co-)regulators of breast CSLCs, through the JAK/STAT3 and NF-kB pathways, respectively[124-126]. These two pathways are not only crucial in immune and inflammatory response but also pivotal for crosstalk between tumor and immune cells, especially in tumor microenvironment[129]. Moreover, the downstream effectors of these signaling pathways are often linked to cell survival and self-renewal as well as tumor proliferation, invasion, and metastasis[130]. Of these five metaVIPER-nominated MRs, only MTDH failed to induce statistically significant reprogramming in HCC38 and HCC1143 cells.

With the possible exception of ZNF131, CRISPR/Cas9-mediated KO of positive CSLC MRs had virtually no effect on cell viability, confirming that cells were reprogrammed to a chemotherapy sensitive state and not selectively ablated. This supports the identification of the MR-inverter drugs via the OncoTreat algorithm, leading to the selection and *in vivo* experimental validation of the anthelmintic albendazole as a highly efficient mediator of CSLC reprogramming. Consistent with these findings, combination therapy with albendazole and paclitaxel resulted in more profound and durable responses, as compared to either monotherapy, leading to a statistically significant increase in overall survival of preclinical models.

Remarkably, since metaVIPER identified a CSLC transcriptional state (and associated MR signature) that was virtually identical across all the tissues and models in this study, irrespective of hormone receptor status, we anticipate that the synergy between albendazole and paclitaxel in a PDX model from a metastatic TNBC patient may also be conserved in HR+ tumors, potentially in combination with hormonal blockade therapy, and may thus be relevant to a large fraction of metastatic breast cancer patients, especially since albendazole is well tolerated.

In parasites, albendazole's mechanism of action is mediated by high-affinity binding to beta tubulin. While the binding is quite selective for parasite tubulin, the drug retains some tubulin-disrupting activity in cancer cells, even though no cytotoxicity is observed at clinically relevant concentrations. Consistently, there are a few tubulin-binding antineoplastic drugs in clinical trials—such as PTC596—that do not present the anti-mitotic cytotoxic effects of drugs such as

paclitaxel, which induce harmful myelosuppression. Indeed, no cytotoxic effects of albendazole were detected in this study, either *in vitro* or *in vivo*. While It has been hypothesized that drugs like PTC596 may work by modulating trafficking of CSLC proteins, like BMI-1, and DNA repair proteins, which may provide a partial rationale for albendazole's effect in CSLCs, and despite its highly reproducible effects *in vitro* and *in vivo*, the precise mechanism of action by which albendazole inverts the activity of CSLC MRs remains to be elucidated and will be the subject of future research. Notably, even though the study was limited to 90 drugs, it identified 17 as statistically significant candidates to reprogram CSLCs to a paclitaxel-sensitive state. As a result, we expect that extending this highly cost-effective approach to much larger drug/compound libraries may reveal even more potent agents.

Taken together, the data presented in this manuscript show that drugs targeting heterogeneous, drug-resistant subpopulations can be effectively identified by single-cell, network-based analyses and that non-oncology drugs may be effectively repurposed to enhance the therapeutic activity of anti-tumor agents, including chemotherapy.

Acknowledgements:

The results shown in this manuscript are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This study was supported by the NCI Outstanding Investigator Award R35CA197745, the NCI Office of Cancer Target Discovery and Development (CTD2) awards U01CA217858 and U01CA272610, as well as the NIH Shared instrumentation grants S10OD032433, S10OD012351 and S10OD021764, all to AC. This publication was also supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant Number UL1TR001873. These studies used the resources of the Herbert Irving Comprehensive Cancer Center Flow Cytometry Shared Resources funded in part through the NCI Center Grant P30CA013696. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Zhao, W., Dovas, A., Spinazzi, E.F., Levitin, H.M., Banu, M.A., Upadhyayula, P., Sudhakar, T., Marie, T., Otten, M.L., Sisti, M.B., et al. (2021). Deconvolution of cell type-specific drug responses in human tumor tissue with single-cell RNA-seq. Genome Med *13*, 82. 10.1186/s13073-021-00894-y.
2. de Sousa, E.M.F., and Vermeulen, L. (2016). Wnt Signaling in Cancer Stem Cell Biology. Cancers (Basel) *8*. 10.3390/cancers8070060.
3. Takahashi-Yanaga, F., and Kahn, M. (2010). Targeting Wnt signaling: can we safely eradicate cancer stem cells? Clin Cancer Res *16*, 3153-3162. 10.1158/1078-0432.CCR-09-2943.
4. Shimono, Y., Zabala, M., Cho, R.W., Lobo, N., Dalerba, P., Qian, D., Diehn, M., Liu, H., Panula, S.P., Chiao, E., et al. (2009). Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. Cell *138*, 592-603. 10.1016/j.cell.2009.07.011.
5. Clarke, M.F. (2019). Clinical and Therapeutic Implications of Cancer Stem Cells. N Engl J Med *380*, 2237-2245. 10.1056/NEJMra1804280.
6. Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W., et al. (2018). Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis *5*, 77-106. 10.1016/j.gendis.2018.05.001.
7. Raymond, E., Hanauske, A., Faivre, S., Izbicka, E., Clark, G., Rowinsky, E.K., and Von Hoff, D.D. (1997). Effects of prolonged versus short-term exposure paclitaxel (Taxol) on human tumor colony-forming units. Anticancer Drugs *8*, 379-385. 10.1097/00001813-199704000-00011.
8. Fillmore, C.M., and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res *10*, R25. 10.1186/bcr1982.
9. Nawara, H.M., Afify, S.M., Hassan, G., Zahra, M.H., Seno, A., and Seno, M. (2021). Paclitaxel-Based Chemotherapy Targeting Cancer Stem Cells from Mono- to Combination Therapy. Biomedicines *9*. 10.3390/biomedicines9050500.
10. Bai, X., Ni, J., Beretov, J., Graham, P., and Li, Y. (2018). Cancer stem cell in breast cancer therapeutic resistance. Cancer Treat Rev *69*, 152-163. 10.1016/j.ctrv.2018.07.004.
11. Diehn, M., Cho, R.W., Lobo, N.A., Kalisky, T., Dorie, M.J., Kulp, A.N., Qian, D., Lam, J.S., Ailles, L.E., Wong, M., et al. (2009). Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature *458*, 780-783. 10.1038/nature07733.
12. Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993-997. 10.1038/nature04496.
13. Ding, H., Douglass, E.F., Jr., Sonabend, A.M., Mela, A., Bose, S., Gonzalez, C., Canoll, P.D., Sims, P.A., Alvarez, M.J., and Califano, A. (2018). Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. Nat Commun *9*, 1471. 10.1038/s41467-018-03843-3.

14. Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet *48*, 838-847. 10.1038/ng.3593.
15. Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science *367*, 405-411. 10.1126/science.aax0249.
16. Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods *14*, 297-301. 10.1038/nmeth.4177.
17. Alvarez, M.J., Subramaniam, P.S., Tang, L.H., Grunn, A., Aburi, M., Rieckhof, G., Komissarova, E.V., Hagan, E.A., Bodei, L., Clemons, P.A., et al. (2018). A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. Nat Genet *50*, 979-989. 10.1038/s41588-018-0138-4.
18. Vasciaveo, A., Arriaga, J.M., de Almeida, F.N., Zou, M., Douglass, E.F., Picech, F., Shibata, M., Rodriguez-Calero, A., de Brot, S., Mitrofanova, A., et al. (2023). OncoLoop: A Network-Based Precision Cancer Medicine Framework. Cancer Discov *13*, 386-409. 10.1158/2159-8290.CD-22-0342.
19. Mundi, P.S., Dela Cruz, F.S., Grunn, A., Diolaiti, D., Mauguen, A., Rainey, A.R., Guillan, K., Siddiquee, A., You, D., Realubit, R., et al. (2023). A Transcriptome-Based Precision Oncology Platform for Patient-Therapy Alignment in a Diverse Set of Treatment-Resistant Malignancies. Cancer discovery *13*, 1386-1407. 10.1158/2159-8290.CD-22-1020.
20. Obradovic, A., Ager, C., Turunen, M., Nirschl, T., Khosravi-Maharlooei, M., Iuga, A., Jackson, C.M., Yegnasubramanian, S., Tomassoni, L., Fernandez, E.C., et al. (2023). Systematic elucidation and pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. Cancer Cell *41*, 933-949 e911. 10.1016/j.ccell.2023.04.003.
21. Obradovic, A., Tomassoni, L., Yu, D., Guillan, K., Souto, K., Fraser, E., Bates, S., Drake, C.G., Saenger, Y., Cruz, F.D., et al. (2022). Case Study of Single-cell Protein Activity Based Drug Prediction for Precision Treatment of Cholangiocarcinoma. bioRxiv 2022.02.28.482410.
22. Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc *9*, 171-181. 10.1038/nprot.2014.006.
23. Bush, E.C., Ray, F., Alvarez, M.J., Realubit, R., Li, H., Karan, C., Califano, A., and Sims, P.A. (2017). PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. Nat Commun *8*, 105. 10.1038/s41467-017-00136-z.
24. Lawson, J.C., Blatch, G.L., and Edkins, A.L. (2009). Cancer stem cells in breast cancer and metastasis. Breast Cancer Res Treat *118*, 241-254. 10.1007/s10549-009-0524-9.
25. Lawson, D.A., Bhakta, N.R., Kessenbrock, K., Prummel, K.D., Yu, Y., Takai, K., Zhou, A., Eyob, H., Balakrishnan, S., Wang, C.Y., et al. (2015). Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature *526*, 131-135. 10.1038/nature15260.
26. Vassilopoulos, A., Chisholm, C., Lahusen, T., Zheng, H., and Deng, C.X. (2014). A critical role of CD29 and CD49f in mediating metastasis for cancer-initiating cells isolated from a

Brca1-associated mouse model of breast cancer. Oncogene *33*, 5477-5482. 10.1038/onc.2013.516.

27. Tian, L., Dong, X., Freytag, S., Le Cao, K.A., Su, S., JalalAbadi, A., Amann-Zalcenstein, D., Weber, T.S., Seidi, A., Jabbari, J.S., et al. (2019). Benchmarking single cell RNA-sequencing analysis pipelines using mixture control experiments. Nat Methods *16*, 479-487. 10.1038/s41592-019-0425-8.
28. Wang, T., Li, B., Nelson, C.E., and Nabavi, S. (2019). Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics *20*, 40. 10.1186/s12859-019-2599-6.
29. Baran, Y., Bercovich, A., Sebe-Pedros, A., Lubling, Y., Giladi, A., Chomsky, E., Meir, Z., Hoichman, M., Lifshitz, A., and Tanay, A. (2019). MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. Genome Biol *20*, 206. 10.1186/s13059-019-1812-2.
30. Hou, W., Ji, Z., Ji, H., and Hicks, S.C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol *21*, 218. 10.1186/s13059-020-02132-x.
31. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol *20*, 296. 10.1186/s13059-019-1874-1.
32. Li, X., Wang, K., Lyu, Y., Pan, H., Zhang, J., Stambolian, D., Susztak, K., Reilly, M.P., Hu, G., and Li, M. (2020). Deep learning enables accurate clustering with batch effect removal in single-cell RNA-seq analysis. Nat Commun *11*, 2338. 10.1038/s41467-020-15851-3.
33. Obradovic, A., Vlahos, L., Laise, P., Worley, J., Tan, X., Wang, A., and Califano, A. (2021). PISCES: A pipeline for the Systematic, Protein Activity-based Analysis of Single Cell RNA Sequencing Data. bioRxiv.
34. Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics *7 Suppl 1*, S7. 10.1186/1471-2105-7-S1-S7.
35. Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat Genet *37*, 382-390. 10.1038/ng1532.
36. Basso, K., Saito, M., Sumazin, P., Margolin, A.A., Wang, K., Lim, W.K., Kitagawa, Y., Schneider, C., Alvarez, M.J., Califano, A., and Dalla-Favera, R. (2010). Integrated biochemical and computational approach identifies BCL6 direct target genes controlling multiple pathways in normal germinal center B cells. Blood *115*, 975-984. 10.1182/blood-2009-06-227017.
37. Obradovic, A., Chowdhury, N., Haake, S.M., Ager, C., Wang, V., Vlahos, L., Guo, X.V., Aggen, D.H., Rathmell, W.K., Jonasch, E., et al. (2021). Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. Cell *184*, 2988-3005 e2916. 10.1016/j.cell.2021.04.038.
38. Elyada, E., Bolisetty, M., Laise, P., Flynn, W.F., Courtois, E.T., Burkhart, R.A., Teinor, J.A., Belleau, P., Biffi, G., Lucito, M.S., et al. (2019). Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. Cancer discovery *9*, 1102-1123. 10.1158/2159-8290.CD-19-0094.

39. Piovan, E., Yu, J., Tosello, V., Herranz, D., Ambesi-Impiombato, A., Da Silva, A.C., Sanchez-Martin, M., Perez-Garcia, A., Rigo, I., Castillo, M., et al. (2013). Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. Cancer Cell *24*, 766-776. 10.1016/j.ccr.2013.10.022.
40. Sheridan, C., Kishimoto, H., Fuchs, R.K., Mehrotra, S., Bhat-Nakshatri, P., Turner, C.H., Goulet, R., Jr., Badve, S., and Nakshatri, H. (2006). CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. Breast Cancer Res *8*, R59. 10.1186/bcr1610.
41. Tatetsu, H., Kong, N.R., Chong, G., Amabile, G., Tenen, D.G., and Chai, L. (2016). SALL4, the missing link between stem cells, development and cancer. Gene *584*, 111-119. 10.1016/j.gene.2016.02.019.
42. BeLow, M., and Osipo, C. (2020). Notch Signaling in Breast Cancer: A Role in Drug Resistance. Cells *9*. 10.3390/cells9102204.
43. Yu, F., Li, J., Chen, H., Fu, J., Ray, S., Huang, S., Zheng, H., and Ai, W. (2011). Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. Oncogene *30*, 2161-2172. 10.1038/onc.2010.591.
44. Xu, X., Zhang, M., Xu, F., and Jiang, S. (2020). Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. Mol Cancer *19*, 165. 10.1186/s12943-020-01276-5.
45. Lo, P.K., Kanojia, D., Liu, X., Singh, U.P., Berger, F.G., Wang, Q., and Chen, H. (2012). CD49f and CD61 identify Her2/neu-induced mammary tumor-initiating cells that are potentially derived from luminal progenitors and maintained by the integrin-TGFbeta signaling. Oncogene *31*, 2614-2626. 10.1038/onc.2011.439.
46. Vaillant, F., Asselin-Labat, M.L., Shackleton, M., Forrest, N.C., Lindeman, G.J., and Visvader, J.E. (2008). The mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models of mammary tumorigenesis. Cancer Res *68*, 7711-7717. 10.1158/0008-5472.CAN-08-1949.
47. Barnawi, R., Al-Khaldi, S., Colak, D., Tulbah, A., Al-Tweigeri, T., Fallatah, M., Monies, D., Ghebeh, H., and Al-Alwan, M. (2019). beta1 Integrin is essential for fascin-mediated breast cancer stem cell function and disease progression. Int J Cancer *145*, 830-841. 10.1002/ijc.32183.
48. Brugnoli, F., Grassilli, S., Al-Qassab, Y., Capitani, S., and Bertagnolo, V. (2019). CD133 in Breast Cancer Cells: More than a Stem Cell Marker. J Oncol *2019*, 7512632. 10.1155/2019/7512632.
49. Wang, Y.J., and Herlyn, M. (2015). The emerging roles of Oct4 in tumor-initiating cells. Am J Physiol Cell Physiol *309*, C709-718. 10.1152/ajpcell.00212.2015.
50. Leis, O., Eguiara, A., Lopez-Arribillaga, E., Alberdi, M.J., Hernandez-Garcia, S., Elorriaga, K., Pandiella, A., Rezola, R., and Martin, A.G. (2012). Sox2 expression in breast tumours and activation in breast cancer stem cells. Oncogene *31*, 1354-1365. 10.1038/onc.2011.338.
51. Lennartsson, J., and Ronnstrand, L. (2012). Stem cell factor receptor/c-Kit: from basic science to clinical implications. Physiol Rev *92*, 1619-1649. 10.1152/physrev.00046.2011.
52. Fultang, N., Chakraborty, M., and Peethambaran, B. (2021). Regulation of cancer stem cells in triple negative breast cancer. Cancer Drug Resist *4*, 321-342. 10.20517/cdr.2020.106.

53. Abd El-Maqsoud, N.M., and Abd El-Rehim, D.M. (2014). Clinicopathologic implications of EpCAM and Sox2 expression in breast cancer. Clin Breast Cancer *14*, e1-9. 10.1016/j.clbc.2013.09.006.
54. Zhu, Y., Wang, Y., Guan, B., Rao, Q., Wang, J., Ma, H., Zhang, Z., and Zhou, X. (2014). C-kit and PDGFRA gene mutations in triple negative breast cancer. Int J Clin Exp Pathol *7*, 4280-4285.
55. Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., et al. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555-567. 10.1016/j.stem.2007.08.014.
56. Voduc, D., Cheang, M., and Nielsen, T. (2008). GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. Cancer Epidemiol Biomarkers Prev *17*, 365-373. 10.1158/1055-9965.EPI-06-1090.
57. Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., et al. (2007). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. Nat Cell Biol *9*, 201-209. 10.1038/ncb1530.
58. Metovic, J., Borella, F., D'Alonzo, M., Biglia, N., Mangherini, L., Tampieri, C., Bertero, L., Cassoni, P., and Castellano, I. (2022). FOXA1 in Breast Cancer: A Luminal Marker with Promising Prognostic and Predictive Impact. Cancers (Basel) *14*. 10.3390/cancers14194699.
59. Pan, H., Peng, Z., Lin, J., Ren, X., Zhang, G., and Cui, Y. (2018). Forkhead box C1 boosts triple-negative breast cancer metastasis through activating the transcription of chemokine receptor-4. Cancer Sci *109*, 3794-3804. 10.1111/cas.13823.
60. Wang, J., Xu, Y., Li, L., Wang, L., Yao, R., Sun, Q., and Du, G. (2017). FOXC1 is associated with estrogen receptor alpha and affects sensitivity of tamoxifen treatment in breast cancer. Cancer Med *6*, 275-287. 10.1002/cam4.990.
61. Khaled, W.T., Choon Lee, S., Stingl, J., Chen, X., Raza Ali, H., Rueda, O.M., Hadi, F., Wang, J., Yu, Y., Chin, S.F., et al. (2015). BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nat Commun *6*, 5987. 10.1038/ncomms6987.
62. Chen, M.H., Yip, G.W., Tse, G.M., Moriya, T., Lui, P.C., Zin, M.L., Bay, B.H., and Tan, P.H. (2008). Expression of basal keratins and vimentin in breast cancers of young women correlates with adverse pathologic parameters. Mod Pathol *21*, 1183-1191. 10.1038/modpathol.2008.90.
63. Tan, D.S., Marchio, C., Jones, R.L., Savage, K., Smith, I.E., Dowsett, M., and Reis-Filho, J.S. (2008). Triple negative breast cancer: molecular profiling and prognostic impact in adjuvant anthracycline-treated patients. Breast Cancer Res Treat *111*, 27-44. 10.1007/s10549-007-9756-8.
64. Saha, S.K., Kim, K., Yang, G.M., Choi, H.Y., and Cho, S.G. (2018). Cytokeratin 19 (KRT19) has a Role in the Reprogramming of Cancer Stem Cell-Like Cells to Less Aggressive and More Drug-Sensitive Cells. Int J Mol Sci *19*. 10.3390/ijms19051423.

65. Saha, S.K., Yin, Y., Chae, H.S., and Cho, S.G. (2019). Opposing Regulation of Cancer Properties via KRT19-Mediated Differential Modulation of Wnt/beta-Catenin/Notch Signaling in Breast and Colon Cancers. Cancers (Basel) *11*. 10.3390/cancers11010099.
66. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst *1*, 417-425. 10.1016/j.cels.2015.12.004.
67. Xia, P., and Xu, X.Y. (2015). PI3K/Akt/mTOR signaling pathway in cancer stem cells: from basic research to clinical application. Am J Cancer Res *5*, 1602-1609.
68. Tokumaru, Y., Oshi, M., Katsuta, E., Yan, L., Satyananda, V., Matsuhashi, N., Futamura, M., Akao, Y., Yoshida, K., and Takabe, K. (2020). KRAS signaling enriched triple negative breast cancer is associated with favorable tumor immune microenvironment and better survival. Am J Cancer Res *10*, 897-907.
69. Ghatak, D., Das Ghosh, D., and Roychoudhury, S. (2020). Cancer Stemness: p53 at the Wheel. Front Oncol *10*, 604124. 10.3389/fonc.2020.604124.
70. Siddharth, S., Das, S., Nayak, A., and Kundu, C.N. (2016). SURVIVIN as a marker for quiescent-breast cancer stem cells-An intermediate, adherent, pre-requisite phase of breast cancer metastasis. Clin Exp Metastasis *33*, 661-675. 10.1007/s10585-016-9809-7.
71. Bertoli, C., Skotheim, J.M., and de Bruin, R.A. (2013). Control of cell cycle transcription during G1 and S phases. Nat Rev Mol Cell Biol *14*, 518-528. 10.1038/nrm3629.
72. Branham, M.T., Nadin, S.B., Vargas-Roig, L.M., and Ciocca, D.R. (2004). DNA damage induced by paclitaxel and DNA repair capability of peripheral blood lymphocytes as evaluated by the alkaline comet assay. Mutat Res *560*, 11-17. 10.1016/j.mrgentox.2004.01.013.
73. McCormick, B., Lowes, D.A., Colvin, L., Torsney, C., and Galley, H.F. (2016). MitoVitE, a mitochondria-targeted antioxidant, limits paclitaxel-induced oxidative stress and mitochondrial damage in vitro, and paclitaxel-induced mechanical hypersensitivity in a rat pain model. Br J Anaesth *117*, 659-666. 10.1093/bja/aew309.
74. Ramanathan, B., Jan, K.Y., Chen, C.H., Hour, T.C., Yu, H.J., and Pu, Y.S. (2005). Resistance to paclitaxel is proportional to cellular total antioxidant capacity. Cancer Res *65*, 8455-8460. 10.1158/0008-5472.CAN-05-1162.
75. Paull, E.O., Aytes, A., Jones, S.J., Subramaniam, P.S., Giorgi, F.M., Douglass, E.F., Tagore, S., Chu, B., Vasciaveo, A., Zheng, S., et al. (2021). A modular master regulator landscape controls cancer transcriptional identity. Cell *184*, 334-351 e320. 10.1016/j.cell.2020.11.045.
76. Rajbhandari, P., Lopez, G., Capdevila, C., Salvatori, B., Yu, J., Rodriguez-Barrueco, R., Martinez, D., Yarmarkovich, M., Weichert-Leahey, N., Abraham, B.J., et al. (2018). Cross-Cohort Analysis Identifies a TEAD4-MYCN Positive Feedback Loop as the Core Regulatory Element of High-Risk Neuroblastoma. Cancer discovery *8*, 582-599. 10.1158/2159-8290.CD-16-0861.
77. Lassman, A.B., Wen, P.Y., van den Bent, M.J., Plotkin, S.R., Walenkamp, A.M.E., Green, A.L., Li, K., Walker, C.J., Chang, H., Tamir, S., et al. (2022). A Phase II Study of the Efficacy and Safety of Oral Selinexor in Recurrent Glioblastoma. Clin Cancer Res *28*, 452-460. 10.1158/1078-0432.CCR-21-2225.

78. Coutinho, D.F., Mundi, P.S., Marks, L.J., Burke, C., Ortiz, M.V., Diolaiti, D., Bird, L., Vallance, K.L., Ibanez, G., You, D., et al. (2022). Validation of a non-oncogene encoded vulnerability to exportin 1 inhibition in pediatric renal tumors. Med (N Y) *3*, 774-791 e777. 10.1016/j.medj.2022.09.002.
79. Sweet, K., Bhatnagar, B., Dohner, H., Donnellan, W., Frankfurt, O., Heuser, M., Kota, V., Liu, H., Raffoux, E., Roboz, G.J., et al. (2021). A 2:1 randomized, open-label, phase II study of selinexor vs. physician's choice in older patients with relapsed or refractory acute myeloid leukemia. Leuk Lymphoma, 1-12. 10.1080/10428194.2021.1950706.
80. Chari, A., Vogl, D.T., Gavriatopoulou, M., Nooka, A.K., Yee, A.J., Huff, C.A., Moreau, P., Dingli, D., Cole, C., Lonial, S., et al. (2019). Oral Selinexor-Dexamethasone for Triple-Class Refractory Multiple Myeloma. N Engl J Med *381*, 727-738. 10.1056/NEJMoa1903455.
81. Landsburg, D.J., Barta, S.K., Ramchandren, R., Batlevi, C., Iyer, S., Kelly, K., Micallef, I.N., Smith, S.M., Stevens, D.A., Alvarez, M., et al. (2021). Fimepinostat (CUDC-907) in patients with relapsed/refractory diffuse large B cell and high-grade B-cell lymphoma: report of a phase 2 trial and exploratory biomarker analyses. Br J Haematol *195*, 201-209. 10.1111/bjh.17730.
82. Califano, A., and Alvarez, M.J. (2017). The recurrent architecture of tumour initiation, progression and drug sensitivity. Nat Rev Cancer *17*, 116-130. 10.1038/nrc.2016.124.
83. Walsh, L.A., Alvarez, M.J., Sabio, E.Y., Reyngold, M., Makarov, V., Mukherjee, S., Lee, K.W., Desrichard, A., Turcan, S., Dalin, M.G., et al. (2017). An Integrated Systems Biology Approach Identifies TRIM25 as a Key Determinant of Breast Cancer Metastasis. Cell Rep *20*, 1623-1640. 10.1016/j.celrep.2017.07.052.
84. Arumugam, K., Shin, W., Schiavone, V., Vlahos, L., Tu, X., Carnevali, D., Kesner, J., Paull, E.O., Romo, N., Subramaniam, P., et al. (2020). The Master Regulator Protein BAZ2B Can Reprogram Human Hematopoietic Lineage-Committed Progenitors into a Multipotent State. Cell Rep *33*, 108474. 10.1016/j.celrep.2020.108474.
85. Talos, F., Mitrofanova, A., Bergren, S.K., Califano, A., and Shen, M.M. (2017). A computational systems approach identifies synergistic specification genes that facilitate lineage conversion to prostate tissue. Nature communications *8*, 14662. 10.1038/ncomms14662.
86. Toledo-Guzman, M.E., Hernandez, M.I., Gomez-Gallegos, A.A., and Ortiz-Sanchez, E. (2019). ALDH as a Stem Cell Marker in Solid Tumors. Curr Stem Cell Res Ther *14*, 375-388. 10.2174/1574888X13666180810120012.
87. Begicevic, R.R., and Falasca, M. (2017). ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. Int J Mol Sci *18*. 10.3390/ijms18112362.
88. Gazit, R., Mandal, P.K., Ebina, W., Ben-Zvi, A., Nombela-Arrieta, C., Silberstein, L.E., and Rossi, D.J. (2014). Fgd5 identifies hematopoietic stem cells in the murine bone marrow. J Exp Med *211*, 1315-1331. 10.1084/jem.20130428.
89. Chen, J.Y., Miyanishi, M., Wang, S.K., Yamazaki, S., Sinha, R., Kao, K.S., Seita, J., Sahoo, D., Nakauchi, H., and Weissman, I.L. (2016). Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature *530*, 223-227. 10.1038/nature16943.
90. Rehman, S.K., Haynes, J., Collignon, E., Brown, K.R., Wang, Y., Nixon, A.M.L., Bruce, J.P., Wintersinger, J.A., Singh Mer, A., Lo, E.B.L., et al. (2021). Colorectal Cancer Cells Enter a

Diapause-like DTP State to Survive Chemotherapy. Cell *184*, 226-242 e221. 10.1016/j.cell.2020.11.018.

91. Ting, S.B., Deneault, E., Hope, K., Cellot, S., Chagraoui, J., Mayotte, N., Dorn, J.F., Laverdure, J.P., Harvey, M., Hawkins, E.D., et al. (2012). Asymmetric segregation and self-renewal of hematopoietic stem and progenitor cells with endocytic Ap2a2. Blood *119*, 2510-2522. 10.1182/blood-2011-11-393272.
92. Guerrero, P.A., Tchaicha, J.H., Chen, Z., Morales, J.E., McCarty, N., Wang, Q., Sulman, E.P., Fuller, G., Lang, F.F., Rao, G., and McCarty, J.H. (2017). Glioblastoma stem cells exploit the alphavbeta8 integrin-TGFbeta1 signaling axis to drive tumor initiation and progression. Oncogene *36*, 6568-6580. 10.1038/onc.2017.248.
93. Barcellos-Hoff, M.H., and Akhurst, R.J. (2009). Transforming growth factor-beta in breast cancer: too much, too late. Breast Cancer Res *11*, 202. 10.1186/bcr2224.
94. Bellomo, C., Caja, L., and Moustakas, A. (2016). Transforming growth factor beta as regulator of cancer stemness and metastasis. Br J Cancer *115*, 761-769. 10.1038/bjc.2016.255.
95. Bhola, N.E., Balko, J.M., Dugger, T.C., Kuba, M.G., Sanchez, V., Sanders, M., Stanford, J., Cook, R.S., and Arteaga, C.L. (2013). TGF-beta inhibition enhances chemotherapy action against triple-negative breast cancer. J Clin Invest *123*, 1348-1358. 10.1172/JCI65416.
96. Laise, P., Turunen, M., Maurer, H.C., Curiel, A.G., Elyada, E., Schmierer, B., Tomassoni, L., Worley, J., Alvarez, M.J., Kesner, J., et al. (2021). Pancreatic Ductal Adenocarcinoma Comprises Coexisting Regulatory States with both Common and Distinct Dependencies. bioRxiv 2020.10.27.357269.
97. Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., et al. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature *490*, 556-560. 10.1038/nature11503.
98. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L.J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res *41*, D808-815. 10.1093/nar/gks1094.
99. Prat, A., Karginova, O., Parker, J.S., Fan, C., He, X., Bixby, L., Harrell, J.C., Roman, E., Adamo, B., Troester, M., and Perou, C.M. (2013). Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res Treat *142*, 237-255. 10.1007/s10549-013-2743-3.
100. Yamamoto, M., Taguchi, Y., Ito-Kureha, T., Semba, K., Yamaguchi, N., and Inoue, J. (2013). NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. Nat Commun *4*, 2299. 10.1038/ncomms3299.
101. Ishibashi, A., Saga, K., Hisatomi, Y., Li, Y., Kaneda, Y., and Nimura, K. (2020). A simple method using CRISPR-Cas9 to knock-out genes in murine cancerous cell lines. Sci Rep *10*, 22345. 10.1038/s41598-020-79303-0.
102. Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat Genet *49*, 1779-1784. 10.1038/ng.3984.

103. Ding, Y., Herman, J.A., Toledo, C.M., Lang, J.M., Corrin, P., Girard, E.J., Basom, R., Delrow, J.J., Olson, J.M., and Paddison, P.J. (2017). ZNF131 suppresses centrosome fragmentation in glioblastoma stem-like cells through regulation of HAUS5. Oncotarget *8*, 48545-48562. 10.18632/oncotarget.18153.
104. Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., et al. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318-325. nature08712 [pii] 10.1038/nature08712.
105. Dutta, A., Le Magnen, C., Mitrofanova, A., Ouyang, X., Califano, A., and Abate-Shen, C. (2016). Identification of an NKX3.1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. Science *352*, 1576-1580. 10.1126/science.aad9512.
106. Hou, Z.J., Luo, X., Zhang, W., Peng, F., Cui, B., Wu, S.J., Zheng, F.M., Xu, J., Xu, L.Z., Long, Z.J., et al. (2015). Flubendazole, FDA-approved anthelmintic, targets breast cancer stem-like cells. Oncotarget *6*, 6326-6340. 10.18632/oncotarget.3436.
107. Kim, U., Shin, C., Kim, C.Y., Ryu, B., Kim, J., Bang, J., and Park, J.H. (2021). Albendazole exerts antiproliferative effects on prostate cancer cells by inducing reactive oxygen species generation. Oncol Lett *21*, 395. 10.3892/ol.2021.12656.
108. Castro, L.S., Kviecinski, M.R., Ourique, F., Parisotto, E.B., Grinevicius, V.M., Correia, J.F., Wilhelm Filho, D., and Pedrosa, R.C. (2016). Albendazole as a promising molecule for tumor control. Redox Biol *10*, 90-99. 10.1016/j.redox.2016.09.013.
109. Jia, Y., Yun, C.H., Park, E., Ercan, D., Manuia, M., Juarez, J., Xu, C., Rhee, K., Chen, T., Zhang, H., et al. (2016). Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors. Nature *534*, 129-132. 10.1038/nature17960.
110. Dagogo-Jack, I., and Shaw, A.T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nature reviews. Clinical oncology *15*, 81-94. 10.1038/nrclinonc.2017.166.
111. Beltran, H., Rickman, D.S., Park, K., Chae, S.S., Sboner, A., MacDonald, T.Y., Wang, Y., Sheikh, K.L., Terry, S., Tagawa, S.T., et al. (2011). Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer discovery *1*, 487-495. 10.1158/2159-8290.CD-11-0130.
112. Goyal, Y., Busch, G.T., Pillai, M., Li, J., Boe, R.H., Grody, E.I., Chelvanambi, M., Dardani, I.P., Emert, B., Bodkin, N., et al. (2023). Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature. 10.1038/s41586-023-06342-8.
113. Neftel, C., Laffy, J., Filbin, M.G., Hara, T., Shore, M.E., Rahme, G.J., Richman, A.R., Silverbush, D., Shaw, M.L., Hebert, C.M., et al. (2019). An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. Cell *178*, 835-849 e821. 10.1016/j.cell.2019.06.024.
114. Diener, J., and Sommer, L. (2021). Reemergence of neural crest stem cell-like states in melanoma during disease progression and treatment. Stem Cells Transl Med *10*, 522-533. 10.1002/sctm.20-0351.
115. Pearce, D.J., Taussig, D., Simpson, C., Allen, K., Rohatiner, A.Z., Lister, T.A., and Bonnet, D. (2005). Characterization of cells with a high aldehyde dehydrogenase activity from cord blood and acute myeloid leukemia samples. Stem Cells *23*, 752-760. 10.1634/stemcells.2004-0292.

116. Alcantara Llaguno, S.R., and Parada, L.F. (2016). Cell of origin of glioma: biological and clinical implications. Br J Cancer *115*, 1445-1450. 10.1038/bjc.2016.354.
117. Chen, J., Li, Y., Yu, T.S., McKay, R.M., Burns, D.K., Kernie, S.G., and Parada, L.F. (2012). A restricted cell population propagates glioblastoma growth after chemotherapy. Nature *488*, 522-526. 10.1038/nature11287.
118. Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988. 10.1073/pnas.0530291100.
119. Dylla, S.J., Beviglia, L., Park, I.K., Chartier, C., Raval, J., Ngan, L., Pickell, K., Aguilar, J., Lazetic, S., Smith-Berdan, S., et al. (2008). Colorectal cancer stem cells are enriched in xenogeneic tumors following chemotherapy. PLoS One *3*, e2428. 10.1371/journal.pone.0002428.
120. Eyler, C.E., and Rich, J.N. (2008). Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis. J Clin Oncol *26*, 2839-2845. 10.1200/JCO.2007.15.1829.
121. Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A., and Lander, E.S. (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell *138*, 645-659. 10.1016/j.cell.2009.06.034.
122. Levina, V., Marrangoni, A.M., DeMarco, R., Gorelik, E., and Lokshin, A.E. (2008). Drug-selected human lung cancer stem cells: cytokine network, tumorigenic and metastatic properties. PLoS One *3*, e3077. 10.1371/journal.pone.0003077.
123. Son, J., Ding, H., Farb, T.B., Efanov, A.M., Sun, J., Gore, J.L., Syed, S.K., Lei, Z., Wang, Q., Accili, D., and Califano, A. (2021). BACH2 inhibition reverses beta cell failure in type 2 diabetes models. J Clin Invest *131*. 10.1172/JCI153876.
124. Jin, W. (2020). Role of JAK/STAT3 Signaling in the Regulation of Metastasis, the Transition of Cancer Stem Cells, and Chemoresistance of Cancer by Epithelial-Mesenchymal Transition. Cells *9*. 10.3390/cells9010217.
125. Galoczova, M., Coates, P., and Vojtesek, B. (2018). STAT3, stem cells, cancer stem cells and p63. Cell Mol Biol Lett *23*, 12. 10.1186/s11658-018-0078-0.
126. Liang, Y., Hu, J., Li, J., Liu, Y., Yu, J., Zhuang, X., Mu, L., Kong, X., Hong, D., Yang, Q., and Hu, G. (2015). Epigenetic Activation of TWIST1 by MTDH Promotes Cancer Stem-like Cell Traits in Breast Cancer. Cancer Res *75*, 3672-3680. 10.1158/0008-5472.CAN-15-0930.
127. Hiramatsu, Y., Fukuda, A., Ogawa, S., Goto, N., Ikuta, K., Tsuda, M., Matsumoto, Y., Kimura, Y., Yoshioka, T., Takada, Y., et al. (2019). Arid1a is essential for intestinal stem cells through Sox9 regulation. Proc Natl Acad Sci U S A *116*, 1704-1713. 10.1073/pnas.1804858116.
128. Wegleiter, T., Buthey, K., Gonzalez-Bohorquez, D., Hruzova, M., Bin Imtiaz, M.K., Abegg, A., Mebert, I., Molteni, A., Kollegger, D., Pelczar, P., and Jessberger, S. (2019). Palmitoylation of BMPR1a regulates neural stem cell fate. Proc Natl Acad Sci U S A *116*, 25688-25696. 10.1073/pnas.1912671116.
129. Fan, Y., Mao, R., and Yang, J. (2013). NF-kappaB and STAT3 signaling pathways collaboratively link inflammation to cancer. Protein Cell *4*, 176-185. 10.1007/s13238-013-2084-3.
130. Rinkenbaugh, A.L., and Baldwin, A.S. (2016). The NF-kappaB Pathway and Cancer Stem Cells. Cells *5*. 10.3390/cells5020016.

Methods

Lead contact

Further information and requests for resources, reagents, and code should be directed to and will be fulfilled by the Lead Contact, Andrea Califano (ac2248@columbia.edu).

Materials availability

This study did not generate new unique reagents.

Patient-derived tumor dissociation and staining: Fresh tumor fragments were acquired from tumor resections of 5 metastatic HR+ and 2 metastatic TNBC breast cancer patients, acquired under IRB AAAB2667. Fragments were quickly dissected to remove necrotic or calcified tissue and then minced, followed by single-cell dissociation using a Miltenyi gentleMACS Octo (Miltenyi, 130-096-427). The latter was performed at 37C using a human tumor dissociation kit (Miltenyi 130-095-929), as per the manufacturer's instructions with the following revisions. When running the gentleMACS, samples were checked after 30 minutes and removed if dissociation was completed. Otherwise, they were re-checked every 15 minutes for a maximum of 60min. At all steps, it is critical to work quickly and, whenever possible, on ice to avoid substantial transcriptional changes.

On removal from the gentleMACS, single-cell suspensions were filtered through a 100uM strainer (Miltenyi, #130-098-463). Cells were then pelleted, supernatant was removed, and red blood cells were lysed using RBC lysis buffer (Invitrogen, 50-112-9751). RBC lysis buffer was diluted in DMEM (Gibco, #11965084), and the cells were washed prior to resuspension in FACS stain buffer (BD, # 554656). Cells were stained as follows:

epitope	color	manufacurer	part #	ul per 100ul test
CD49f	APC	biolegend	313616	5
CD326 (Ep-CAM)	FITC	biolegend	369814	5
CD3	PE	BD	561803	20
CD45	PE	BD	103105	20
CD64	PE	BD	561926	20
CD16	PE	BD	561725	20
CD31	PE	biolegend	303105	5
CD140A	PE	biolegend	323505	5

Cells were stained for 20min on ice, then washed 2X in cold stain buffer. Cells were resuspended in 500-1000ul stain buffer and filtered prior flow cytometry-based sorting (FACS) for CD49f^{high}, CD326^{+}, DAPI^{-}, Lin^{-} . The number and purity of tumor cells obtained varies widely based on fragment size, cellularity, necrosis, calcification, RBC and adipocyte contamination, and other factors.

Plate-based scRNA-seq of patient-derived samples: At the time this study started, in 2015, there was no peer-reviewed plate-based scRNA-seq method that met our cost-per-cell and quality requirements. Consequently, we created a UMI-enabled, 3' scRNA-seq method that is described in detail at: *dx.doi.org/10.17504/protocols.io.s4hegt6*. In short, single cells are sorted directly into wells filled with hypotonic lysis buffer which contains RNAse inhibitor. Plates can be frozen at -80C for at least two weeks. After reverse transcription, wells are pooled into a single tube for NGS library generation. The resulting libraries were sequenced on a NextSeq 550 with approximately 0.5M-1M reads per well. The raw data were deposited in Gene Expression Omnibus[1] (GEO; accession number: GSE108540). Note that subsequent to this study, this method was deprecated in favor of mcSCRB-seq[2], which is more sensitive. Compared to droplet-based studies, this methodology provided much more effective elimination of dead cells and debris from necrotic tissue, which caused significant clogging and sequencing artifacts as well as higher-depth sequencing. Critically, the results obtained from the earlier technology were highly consistent with later technologies, including the 10X Genomics Chromium platform. Indeed, the initial results from human samples, including the Master Regulator proteins associated with BRCA stem-like progenitors, were fully recapitulated in single-cell profiles generated by Chromium 10X library generation from PDX models.

Patient-derived xenograft (PDX) model: Patient-derived xenografts were generated by implanting cells obtained from a triple-negative breast carcinoma (TNBC) patient. Tumor cell suspensions were mixed (1:1) with Matrigel (Corning) and implanted into the mammary fat pad of 6 to 8-week-old female NOD (NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ) SCID gamma (NSG) mice (Jackson Labs, Strain #005557) for scRNA-seq and therapeutic studies.

Preclinical studies in the PDX model: Paclitaxel was formulated in sterile saline and dosed at 25 mg/kg intravenously every week. Albendazole was formulated in 10% (v/v) DMSO (dimethylsulfoxide), 0.5% (w/v) carboxymethylcellulose, and 1% (v/v) Tween 80 in sterile water and dosed at 50 mg/kg IP 3 times weekly. The vehicle control mice were given 10% (v/v) DMSO (dimethylsulfoxide), 0.5% (w/v) carboxymethylcellulose, and 1% (v/v) Tween 80 in sterile water.

For therapeutic studies, tumor measurements were made biweekly using calipers. Treatment was initiated when tumor volume (TV) reached ~100 mm^3. TV is calculated using a modified ellipsoid formula: ½ (length X $width^2$). Mice were treated with albendazole for two weeks prior to initiating three weekly paclitaxel cycles (n=7 mice/arm). Mice were treated for up to 2 cycles of albendazole + paclitaxel therapy with a 2-week treatment break between cycles as well as with individual monotherapies and vehicle control. Mean tumor volume and tumor growth rates were compared among treatment groups, using a Tukey's honest significance test. Tumor growth rates were calculated assuming exponential kinetics[3] during the first cycle (from 1 to 18 days). Under this assumption, the growth rate (b) of each tumor model can be calculated by solving the following linear regression:

$$\text{Log}_{10}\ (\text{fold-increase in tumor volume}) = a + b \times \text{time} + \varepsilon,$$

where a is the log-scaled initial tumor volume, and ε denotes the error.

To support collection of sufficient tumor material for the analysis, treatments for single cell drug response studies were initiated when TV reached ~350-400 mm^3. Animals (n=2/arm) were treated for 15 days and tumors harvested 2 hours after administration of the last dose of vehicle or drug. Harvested tissue was enzymatically dissociated into a single cell suspension for downstream scRNA-seq using the Human Tumor Dissociation Kit (Miltenyi, 130-095-929) according to manufacturer's recommendations (other than the modifications discussed in the patient-derived sample preparation, above).

scRNA-seq profiles from PDX samples: Tumors were resected at the Columbia University Irving Medical Center (CUIMC) and stored in cold DMEM media on ice for transport from surgery to the lab. After gross necrotic tissue removal using a scalpel, we used a gentleMACS Octo (Miltenyi, 130-096-427) to dissociate minced samples to single-cell suspensions using a human tumor dissociation kit (Miltenyi, 130-095-929). The procedure was performed at 37C, as per the manufacturer's instructions, with the following change: samples were checked after 30min and the protocol was terminated if chunks of undissociated tissue were not visible. If a significant amount of undissociated sample remained, the program was run for 15min additional time. After dissociation, samples were filtered through a 100uM strainer, the cells pelleted, the supernatant removed, and red blood cells were lysed using RBC lysis buffer (Invitrogen, 50-112-9751). RBC lysis buffer was diluted in DMEM, and the cells were washed prior to mouse cell depletion using magnetic beads (Miltenyi, 130-104-694) following the manufacturer's instruction. The resulting

suspension of human tumor cells was adjusted in DMEM medium (Gibco, #11965084) to a concentration appropriate for loading into a 10X genomics Chromium controller (3' gene expression kit, #120267) with an expected output of 7,000 cells/sample. Samples were run in separate wells and libraries constructed according to the manufacturer's instructions. Libraries were sequenced on an Illumina Novaseq 6000, aiming for approximately 100,000 reads/cell. The raw PDX data were deposited in GSE226329.

Generation of scRNA-seq profiles from cell lines: To generate single-cell profiles of breast cancer cell lines, MCF7, HCC2157, HCC38, and HCC1143 cells were seeded into 15cm circular plates and incubated for 72 hours to achieve < 50% average confluence, followed by dissociation and scRNA-seq profile generation. Specifically, the resulting human tumor cell suspensions—HCC38 and HCC1143 in batch 1 and MCF7 and HCC2157 in batch 2—were diluted to the appropriate concentration for loading into a 10X genomics Chromium controller (3' gene expression kit, #120267) at an expected output of 5,000 cells per cell line. Within each batch, cells were pooled into the same well and libraries were constructed according to the manufacturer's instructions. Libraries were sequenced on an Illumina Novaseq 6000, aiming for approximately 100,000 reads/cell. Pooled scRNA-seq samples—i.e., sample1, including pooled reads from HCC38 and HCC1143 cells and sample2, including pooled reads from HCC2157 and MCF7 cells—were deposited in GEO (accession number: GSE241115).

Cell Culture Conditions and Media:

HCC1143 - RPMI 10%FBS + pen/strep

HCC38- RPMI 10%FBS + pen/strep

MCF7 - EMEM 10%FBS + 0.01 mg/ml Insulin + pen/strep

HCC2157 – DMEM 10%FBS + pen/strep

All cell lines were routinely tested for mycoplasma contamination. Cell lines were kept in a 37 °C humidity-controlled incubator with 5.0% CO2.

CROPSeq library design: For the CROPSeq screening[4], we designed a target gene list that included the top 25 predicted MR proteins (MR_{CSLC}) and bottom 25 predicted MR proteins (MR_{DIFF}) of the CSLC vs. differentiated cell states, for a total of 50 candidate MR genes in total. Each gene

was targeted with 3 sgRNAs designed using CRISPick[5,6]. 15 additional sgRNAs targeting intergenic genomic regions[7] were selected as negative controls for the assay.

sgRNA oligo synthesis and library cloning: Oligo libraries (165 oligos) were ordered from Twist-biosciences in following format:

CGATTTCTTGGCTTTATATATCTTGTGGAtttCGTCTCCCACCGNNNNNNNNNNNNNNNNNNNNNGTTTAGAGACGAAAGAGCTAAGCTGGAAACAGCATAGCAAG

Twist oligo pool was amplified by using following protocol:

1ul	Twist oligo library (1ng/ul)
5ul	10x Ex Taq buffer buffer
4 ul	dNTPs
0.375 ul	ExTaq
2ul	sgRNA_insert_dd_F (10uM)
2ul	sgRNA_insert_dd_R (10uM)
2.5ul	20xSYBR
33.125ul	H2O

98C 1 min

98C 10 sec

64C 30 sec

} done with qPCR, stopped before saturation.

72C 30 sec

72C 10 min

4C ∞

<u>Insert amplifying oligos:</u>

sgRNA_insert_dd_F: CGATTTCTTGGCTTTATATATCTTGTGGA

sgRNA_insert_dd_R: CTTGCTATGCTGTTTCCAGCTTAG

After PCR amplification, the insert was gel purified (GeneJet) and Golden-gate cloned into BsmBI-digested CROPseq-CaptureSeq-Guide-Puro-plasmid[8]. The Golden-gate cloned product was Isopropanol precipitated and large-scale electroporated into Lucigen Enduro competent cells. The bacterial colonies were scraped from 24,5cm x 24,5cm agar plates, so that the estimated library complexity was > 1000 colonies / sgRNA. The sgRNA library plasmid DNA was extracted by using NucleoBond Xtra Midi kit (Macherey-Nagel).

Lentiviral library packaging: 13 million 293T cells / plate were seeded in three 15cm dishes the night preceding the transfection assay. The following morning the viral transfections were conducted the following way:

- 22.1ug sgRNA-library containing CROPseq-CaptureSeq-Guide-Puro, or modified lenti-Cas9-sgHPRT1 (Addgene #196713)[9]. For this study the sgHPRT1 part was removed from the lenti-Cas9-sgHPRT1-vector.

- 16.6ug PsPAX2 (Addgene 12260)

- 5.5ug PMD2G (Addgene 8454).

- 1660ul of sterile H2O.

After mixing the plasmids, 110,6ul of Fugene HD (Promega) was added to the mix.

The transfection mixture was vortexed, and incubated 10 minutes in room temperature before adding dropwise to 293T cells. Altogether 3 x 15cm plates were transfected for sgRNA-library containing modified CROPseq-CaptureSeq-Guide-Puro and 1 x 15cm plates were transfected for modified lenti-Cas9-sgHPRT1.

The transfection mixture was removed the following day and virus was collected at 48h and 72h after initial transfections. To remove cellular debris, the virus containing supernatant was centrifuged 500 x g for 5min and filtered by using 0.45um PES filters (Millipore). The lentivirus was concentrated by using Lenti-X concentrator (Clontec), aliquoted and stored at -80C.

Generation of Cas9 expressing cell lines: Cas9 expressing cell lines were generated as follows: Concentrated modified lenti-Cas9-sgHPRT1 -lentivirus was transduced into breast cancer cell lines HCC1143 and HCC38 (in presence of 8ug/ml polybrene) at an estimated MOI = 0.3. The virus was removed the following day and Blasticidin was added to the cells. Blasticidin selection was continued as long as the control cells (non-transduced) were viable.

CROP-seq screening: sgRNA harboring lentiviruses were transduced into Cas9 expressing breast cancer cell lines HCC1143 and HCC38 (in 15cm plate-format) in presence of 8ug/ml polybrene, at an estimated MOI < 0.2. After 24h, the lentivirus containing media was removed, cells were washed with PBS, and puromycin-containing media (3ug/ml) was added to the cells for 48-96h until all control cells (not virus-infected) were dead. The cells were grown for 14 days post sgRNA transduction to allow potential reprogramming effects to take place, before the scRNA-seq. The CRISPR perturbed cell lines (HCC38, HCC1143) were adjusted to a concentration appropriate for loading into a 10X genomics (3' gene expression kit, #120267) with an expected output of 10,000 cells/each cell line. The 10x libraries were prepared based on manufacturer's instructions with feature barcode CRISPR screening protocol. Libraries were sequenced on an Illumina Novaseq 6000, aiming for approximately 100,000 reads/cell. The raw CROP-seq data were deposited in GSE241115.

Bliss independence analysis: We tested for pharmacological synergy using a multiplicative null model based on Bliss Independence. Briefly, the efficacy of each drug as a monotherapy was evaluated based on the reduction in PDX tumor growth at 25 days. A conservative empirical null model was generated by fitting a distribution generated by using every possible triplet of PDX treated with each monotherapy and vehicle control. An empirical left-tailed *p*-value was then computed for each combination-treated PDX model using the empirical null model; these *p*-values were integrated across independent biological replicate PDX models using Fisher's method to produce a single synergy *p*-value for each drug combination.

Kaplan-Meier analysis: Kaplan-Meier analyses comparing disease control and overall survival across treatment groups was performed using the log-rank test. The endpoint for disease control is defined as the time to TV doubling relative to TV at the initiation of paclitaxel therapy (defined

as Day 1). Overall survival is defined as time to mouse death, attainment of maximal allowed TV (humane endpoint), or development of moribund status requiring euthanasia. A Benjamini-Hochberg multiple testing correction was applied to relative TV comparisons and Kaplan-Meier analysis with adjusted p-values reported. Statistical significance was defined as $p<0.05$. All in vivo studies were performed in accordance with institutional guidelines and under protocol #16-08-011 approved by the Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Animal Care and Use Committee.

InferCNV: We applied inferCNV of the Trinity CTAT Project (https://github.com/broadinstitute/inferCNV) to the scRNA-seq data of patient tumor samples in order to estimate the copy number of variation of whole chromosomes and confirm the purity of tumor cells in a data set. For CNV reference data, we used the scRNA-seq data (GEO access number: GSE113196) of normal breast epithelial cells containing one basal and two liminal types in four different individuals from Nguyen et al., 2018[10]. For inferCNV, we used the order of human genes based on transcription start sites in GRCh38.

Protein activity analysis: Protein activity was computationally assessed using metaVIPER[11]. Briefly, metaVIPER is the single cell adaptation of the Virtual Inference of Protein-activity by Enriched Regulon analysis (VIPER)[12]. VIPER computes the normalized, rank-based enrichment score (NES) of a set of genes representing the transcriptional targets of each protein (regulon) in genes differentially expressed when comparing the cell state of interest to a reference cell state, generally the centroid of all the cells in the analysis. Thus, statistically significant positive and negative NES values identify either activated or inactivated proteins in the state of interest vs. reference state, while non-significant NES scores identify proteins that have not significantly changed activity. Unlike gene set enrichment analysis (GSEA), VIPER uses a probabilistic model to integrate the consensus between activated and inhibited targets, as well as targets with an unclear mode of regulation, and their differential expression. The use of context-specific networks, which can improve the fidelity of interactomes, is highly advantageous in the VIPER analysis with respect to the accurate inference of the protein activity. MetaVIPER, improves the analysis by allowing the use of multiple networks to define each protein's regulon set, this is important in single cell analyses where the optimal network may be cell specific and also helps prevent overfitting to a single network.

For patient-derived tumor samples, we first removed low-quality cells where the total read count (nReads) was greater than 35,000, the number of detected genes (nGenes) was either below 1,000 or above 4,000, and the mitochondrial RNA percentage (MT%) was greater than 25%. For PDX data, we filtered cells for MT% ≤ 16%, 6,000 ≤ nReads ≤ 85,000, and nGenes ≥ 1,000.

To assess protein activity from scRNA-seq data, raw gene expression data was normalized to $\log_2$ count per million (CPM). The data were then centered at the median expression and divided by the median absolute deviation (MAD) of each gene. Note that MAD was calculated excluding the zero counts for patient tumor data due to sparsity of single cell gene expression profiles. Genes with zero counts were set to N/A and were thus ignored by the R implementation of the VIPER algorithm used in this study.

Finally, during the metaVIPER analysis, we utilized breast cancer specific networks generated by analyzing both bulk samples and single cell RNA-seq profiles using the Adaptive Partitioning version of the Algorithm for the Reconstruction of Accurate Cellular Networks[13] (ARACNe-AP), as described in the next section.

Network inference by ARACNe-AP: Gene regulatory networks for human breast cancer were created using the extensively validated reverse-engineering algorithm ARACNe-AP[13]. Briefly, ARACNe-AP infers protein-gene regulatory interactions by first measuring the statistical significance of the mutual information between the gene expressions of each regulator protein and each candidate target and then by removing statistically significant candidate targets that violate the Data Processing Inequality (DPI), see[14] for additional details. After the analysis, each regulator protein is associated with a set of most directly regulated transcriptional targets (regulon). Since the accuracy of VIPER analyses does not further improve when regulons are larger than 40 genes[12], we conservatively retained only the 50 most statistically significant targets in each regulon. This allows to remove potential bias in VIPER NES assessment, since NES measured from larger regulons would be more statistically significant.

In this study, we inferred the regulatory interactions not only for transcriptional (co-)regulators but also for proteins indirectly involved in gene transcriptional regulation such as cell surface receptors, intracellular signaling molecules, and DNA-binding molecules. This is strongly supported by recent studies confirming that ARACNe-inferred regulons for these additional protein classes are highly effective in assessing their differential activity using the VIPER algorithm[15-17]. The candidate regulator proteins for these analyses (n = 8,519) were thus identified

based on the following Gene Ontology (GO) terms: transcription regulator activity (GO:0140110), transcription coregulator activity (GO:0003712), DNA-binding transcription activator activity (GO:0001216), regulation of gene expression (GO:0010468), signal transduction (GO:0007165) and cell surface (GO:0009986).

ARACNe-AP was then run using 200 bootstraps and a Bonferroni-corrected statistical significance threshold of $p = 0.05$ for protein-target interaction inference. The following datasets were used to generate complementary Breast Cancer networks: (a) RNA-seq profiles from samples in the TCGA-BRCA cohort[18] and (b) scRNA-seq profiles from human breast cancer cells isolated from PDX-derived samples. We did not use the single cells from the human samples because single cell networks must be generated on a sample-by-sample basis, to avoid major artifacts associated with batch effects, and the number of cells generated by single cell profiling of the patient-derived sample was too small to generate accurate, sample-specific ARACNe-inferred networks. As a result, single-cell networks were generated using scRNA-seq profiles from malignant human cells isolated from the PDX model using the ARACNe-AP algorithm, as described in the next paragraph.

To reduce gene dropout effects, single cell network inference from PDX-derived scRNA-seq profiles was performed by first converting individual scRNA-seq profiles to MetaCell[19,20] profiles. Additionally, to avoid batch effects, MetaCells were generated independently for each treatment condition (i.e., vehicle, paclitaxel and albendazole treatment). MetaCells are produced by combining the gene expression profiles of a “seed” cell, chosen at random, with those of *k* single cells with the closest gene expression profile (based on Spearman correlation). As discussed in the PISCES analysis pipeline[21], the per-MetaCell UMI count for optimal network inference should be $\geq$ 10,000. Based on the average single-cell UMI count in this study, we thus used $k = 9$ resulting in MetaCell profiles (*i.e.*, pseudo-bulk profiles) generated by merging the UMI counts of 10 individual cells in each MetaCell. Nearest neighbors were identified by gene expression-based Spearman’s correlation, after regressing out cell cycle contributions from the CPM-normalized and scaled data, using Seurat[22]. ARACNe-AP was then used to analyze the CPM-normalized expression produced by 150 MetaCells, with seed cells chosen at random from the entire population of each treatment conditions. Note that, to avoid mixing biological conditions, the nearest neighbor cells of each seed cell were selected within the same treatment condition.

For TCGA-based networks, we utilized the read count data normalized by Fragments Per Kilobase of transcript per Million mapped reads upper quartile (FPKM-UQ) in TCGA-BRCA project,

which were downloaded using TCGAbiolinks[23], an R package developed to retrieve datasets from the Genomic Data Commons[24] (GDC). Subtype-specific networks were generated by sub-selecting samples from the following BRCA subtypes: Luminal A (ER+/PR+/HER2-), Luminal B (ER+/PR+/HER2+), HER2-enriched (ER-/PR-/HER2+), and basal-like (ER-/PR-/HER2-), as defined in the Clinical Supplement data in GDC. Finally, we applied these 4 BRCA subtype networks as well as the PDX-derived single-cell network for the metaVIPER-based assessment of protein activity from patient samples, or only the latter network for protein activity assessment in PDX samples.

CytoTRACE Analysis: We utilized the CytoTRACE[25] R package downloaded from its official website (https://cytotrace.stanford.edu/). Briefly, CytoTRACE is a computational tool to infer a cell differentiation order, using the gene count signature that is defined as the geometric mean of top 200 genes correlating with the number of expressed genes over the cells in a dataset. CytoTRACE generates a relative score between 0 (the most differentiated) and 1 (the least differentiated) for individual cells. For patient-derived samples, we performed CytoTRACE analysis on the individual cells of each sample independently to account for potential germline differences that may create batch effects. For PDX-derived samples, which had a common germline background, we performed CytoTRACE using the single cells from all samples, including those treated with vehicle control, paclitaxel, and albendazole. For both analyses, raw counts were used as an input.

Marker-based Stemness Score: to generate a more biologically motivated, breast cancer specific assessment of cell stemness, we integrated the metaVIPER-based NES representing the differential activity of 14 previously established markers of CSLC state, including CD44+/CD24-[26], ITGA6 (CD49f)[27], BMI1[28], SALL4[29], NOTCH1[30], NOTCH2[30], KLF4[31], CTNNB1[32], ITGB3 (CD61)[33,34], ITGB1[35], PROM1 (CD133)[36], POU5F1 (OCT4)[37], SOX2[38], and KIT[39]. Note that the activity of the markers was inferred as described in Protein Activity Analysis above. In addition, the activity of CD44+/CD24- was calculated as the difference between the CD44 NES and the CD24 NES, such that greater NES values would indicate greater stemness. The integration of 14 maker's activities was done by standardizing the activities for individual markers then by averaging them. Consequently, the marker-based Stemness Score (MSS) that integrates a set of the stemness markers (M) can be summarized using the following equation:

$$MSS_j = \frac{1}{|M|} \sum_{m \in M} \frac{(p_{j,m} - \bar{p}_m)}{\sigma_m}$$

where $p_{j,m}$ is the protein activity of marker m in cell j, $\bar{p}_m$ is the average protein activity for marker m, and σ_m is a standard deviation of the protein activity of marker m.

Integrated Stemness Score: As discussed in the main text, there was highly statistically significant correlation between the CytoTRACE ($S_{cytotrace}$) and marker-based stemness (MSS) metrics, even though the methodology used to compute the two metrics were completely different and statistically independent. As a result, we proceeded to combine both metrics into a single integrated Stemness Score (ISS) by averaging the normalized score for each metric as follow.

$$ISS_j = (z_{j,cytotrace} + z_{j,MSS})/2$$

where $z_{j,cytotrace}=(S_{j,cytotrace} - \bar{S}_{cytotrace})/\sigma_{cytotrace}$ and $z_{j,MSS}=(MSS_j - \overline{MSS})/\sigma_{MSS}$ for cell j. Finally, ISS was linearly re-scaled to a score ranging from zero to one in Fig 2D and Fig3C, as follow.

$$ISS_j = (ISS_j - ISS_{min})/(ISS_{max} - ISS_{min})$$

Breast CSLC MR signature: To generate a consensus differential protein activity signature for breast CSLCs, we integrated the NES of each protein, as assessed by metaVIPER by comparing the 20 cells with the highest (most CSLC) and 20 cells with the lowest (most differentiated) marker-based stemness score across each patient-derived sample and the PDX-derived sample. The differential activity for each sample was computed using *viperSignature()* in VIPER[11,12] R package, with the setting of the protein activity profiles for 20 $cells_{CSLC}$ as a test and the protein activity profiles for 20 $cells_{DIFF}$ as a referential state. Then, the z-score output by *viperSignature()* was averaged over cells for each sample. Justifying the generation of a consensus protein activity signature, we noted that the most differentially active proteins, as assessed in each sample, were exceedingly conserved across all samples (**Suppl. Fig. S12B**). Subsequently, the differential activity of each protein was integrated across the 7 patients and the control PDX model, using weighted z-test (z_w):

$$z_w = \frac{\sum_k w_k z_k}{\sqrt{\sum_k w_k^2}}$$

where z_k is the differentially active protein for sample k and w_k is a weight assessed as the inverse of the standard deviation of sample k. Thus, the transcriptional regulator proteins with the most positive (activated) and negative (inactivated) integrated NES in z_w were selected as candidate MR proteins controlling the CSLC vs. differentiated states of breast cancer cells.

Stemness Marker Analysis: in addition to the above-discussed established CSLC markers in breast cancer, we also assessed stemness based on an additional set of markers previously associated with cell stemness (non-breast-cancer-specific). These include genes in the ALDH family[40,41] , ABC family[41], quiescent stem-cell markers (FGD5[42] and HOXB5[43]), embryonic diapause[44] and asymmetric cell division processes[45] (**Suppl. Table 3**). To identify genes involved in asymmetric cell division we used GO:0008356, while genes involved in embryonic diapause were selected from Rehman et al., 2021[44]. To assess enrichment of these gene sets in the breast CSLC signature, we used analytic rank-based enrichment analysis (aREA) algorithm introduced in the VIPER manuscript[12], which considers the sign (over or under-expressed). For this analysis the weight of each gene was set uniformly to one.

Assessing CSLC enrichment in CCLE cell lines: We downloaded the TPM-normalized expression data (CCLE_expression.csv as of 06/30/2022) for all breast cancer cell lines in the Cancer Cell Line Encyclopedia[46] from the Cancer Dependency Map (DepMap)[47]. Protein activity was assessed by metaVIPER, using the TCGA and single cell breast cancer-specific networks described above. To generate a signature for metaVIPER analysis, gene expression values were $\log_2$-scaled and normalized by subtracting the median value and dividing by the median absolute deviation (MAD) across all CCLE-BRCA samples. Note that MAD smaller than 0.01 was set to 0.01 to avoid denominator values approaching zero.

To assess the ability of each cell lines to recapitulate the CSLC MR signature differential activity, we used the OncoMatch algorithm[48,49] analysis, designed to assess the overlap between two differential protein activity signatures. Specifically, we measured the normalized enrichment score (NES) of the 25 most activated and 25 most inactivated proteins in the breast CSLC signature in proteins differentially active in each CCLE-BRCA cell line (Fig. S19A). Thus, statistically significant NES scores identify cell lines that recapitulate the CSLC MR signature while statistically significant negative NES score identify cell lines that recapitulate the signature of the

most differentiated BRCA cells. For this analysis, we only used the differential activity of transcriptional regulator proteins (n = 2088).

Single-cell Stemness Analysis of HCC38, HCC1143, MCF7, and HCC2157 cell lines. To identify candidate breast cancer cell lines for the study, we assessed the stemness of 62 breast cancer cell lines in CCLE based on the enrichment (NES) of their most differentially activated proteins in the CSLC MR signature. As also supported by literature evidence[50-52], we then selected HCC1143 (ranked 1) and HCC38 (ranked 3) as most enriched in the stem-like progenitor compartment and HCC2157 (ranked 49) and MCF7 (ranked 56) as basal and luminal cell lines most enriched in differentiated cells.

For efficiency purposes, we pooled HCC38 and HCC1143 cells and MCF7 and HCC2157 cells together in single-cell RNA sequencing. Cells were then demutiplexed using Demuxlet[53]. Briefly, this algorithm computes the statistical likelihood of a cell belonging to a specific sample based on the overlap of single-nucleotide polymorphisms (SNPs) between scRNA-seq reads and the referential Variant Call Format (VCF) data. For this analysis, we generated VCF files for each cell line from the CCLE raw data, as downloaded from the Sequence Read Archives[54] (SRA), see below. After removing poor quality cells (MT% ≥ 7.5%, nReads ≥ 200,000, nGenes ≤ 5,000 for the HCC38/HCC1143 pool and MT% ≥ 12%, nReads ≤ 10,000 or nReads ≥ 90,000, and nGenes ≤ 3,000, for the MCF7/HCC2157 pool), we ensured that single cells from the 4 cell lines were clearly separated into distinct clusters in a UMAP projection, with each cluster associated with a specific cell lines by Demuxlet. The integrated stemness score of each single cell was then computed in the same manner as described above for the patient and PDX-derived samples.

Variant Discovery Analysis: For variant discovery analysis, used to deconvolute the pooled scRNA-seq profiles, we downloaded the CCLE raw data (FASTQ files) for the HCC38, HCC1143, MCF7, and HCC2157 cell lines from SRA (accession IDs: SRR8615458, SRR8615819, SRR8615758, and SRR8615891). We generated BAM files, with read sequences aligned and mapped to the human genome reference (GRCh38; Homo_sapiens.GRCh38.dna_sm.primary_assembly.fa as of 12/12/2022) from Ensembl[55], using CellRanger[56], and used the GATK[57] (Genomic Analysis ToolKit) for variant calling analysis to generate the VCF files for each cell line. We applied *SplitNCigarReads*, a GATK subroutine, to split reads that contain Ns in their spliced alignment from the BAM files, before variant calling.

CROP-seq Data Analysis: We used CellRanger to analyze the raw scRNA-seq data produced by the CROP-seq assay, for feature barcode analysis. The minimum UMI thresholds to call cells harboring a specific sgRNA was determined based on a gaussian mixture model by CellRanger's CRISPR analysis for each sgRNA[58]. Cells with sgRNA UMI counts below the threshold were ignored as noise. Subsequently, poor quality cells were removed using the following criteria: For HCC38: MT% ≥ 9%, nReads ≥ 80,000, and nGenes ≤ 4,000; for HCC1143: MT% ≥ 10%, nReads ≥ 120,000, and nGenes ≤ 5,000). Cells with more than one sgRNA detected were excluded from the analysis.

To increase read depth and the number of detected genes in cells where a specific MR was knocked out, we used a pseudo-bulk[59] approach, in which UMI counts were aggregated from the cells detected with the same sgRNA. This was critical to support effective differential protein activity analysis to assess the stemness score associated with knock-out of each MR compared to cells harboring intergenic control sgRNAs.

The pseudo-bulk expressions were CPM-normalized and transformed by $\log_2$-scaling. Then, the differential gene expression between a target sgRNA and the intergenic sgRNAs was computed by subtracting the median value and dividing by the median absolute deviation (MAD) across the expression of pseudo-bulks harboring intergenic sgRNAs (n=15). Note that MAD smaller than 0.01 was set to 0.01 to avoid denominator values approaching zero. Then, differential protein activity, as induced by CRISPR/Cas9-mediated, KO was assessed by metaVIPER analysis of the differential gene expression signature, using the TCGA-BRCA bulk and PDX-derived single-cell networks. MetaVIPER-assessed protein activity was converted into z-score, using the *viperSignature()* function in the VIPER[12] R package. Specifically, *viperSignature()* was used to compute the z-score of the protein activity of the pseudo-bulk harboring a target MR sgRNA, compared to the protein activity of pseudo-bulks harboring intergenic sgRNAs (n=15), used as a reference control.

For each MR, effective KO was assessed by determining whether its activity in samples harboring the MR-targeting sgRNAs was lower than its average activity in all samples containing intergenic sgRNAs or sgRNAs targeting other MRs, by at least one standard deviation. Only MRs passing this test were further considered in the analysis. As a result, only 16 of 25 MR_{SCLC} (BMPR1A, MTDH, ZNF131, MAML3, GON4L, ZNF24, SMAD5, KLF3, UBP1, SMAD1, TMF1, XBP1, MIER1, VEZF1, ETV3, ZNF566) and 9 of 25 MR_{DIFF} (PCBD1, RUVBL2, HDGF, RPS3, RORC, ENY2,

PEX14, THAP8, PARK7) were further analyzed in HCC38 cells, respectively. Conversely, only 15 of 25 MR_{SCLC} (STAT3, BMPR1A, MTDH, ZNF131, GON4L, MYBL1, SMAD5, UBP1, NCOA1, SMAD1, TMF1, XBP1, VEZF1, ETV3, ZNF566) and 11 of 25 MR_{DIFF} (PCBD1, RUVBL2, HDGF, PRDX2, YBX1, RORC, LAMTOR5, ENY2, THAP8, HLX, PARK7) were evaluated in HCC1143 cells, respectively.

The stemness NES of the cell state induced was computed by weighted enrichment analysis of the top 25 activated and 25 inactivated MRs following MR KO in proteins differentially active and inactive in the patient-derived stemness MR signature, using the *aREA()* function in the VIPER package. Specifically, the top 25 activated MRs were weighted positively (+1), whereas the top 25 inactivated MR were weighted negatively (-1) in the aREA analysis. To avoid biasing the analysis, each knocked-out MR was excluded from the corresponding aREA analysis, such that only its downstream effectors were considered. The statistical significance of the change in Stemness Score (NES) in each group (MR_{CSLC} and MR_{DIFF}) was evaluated by Mann Whitney U Test, compared to the Stemness Score of the control group.

Next, to assess the reprogramming potential of individual MRs, we generated bootstrapped pseudo-bulk expression profiles for each sgRNA. In short, cells containing sgRNAs targeting the same MR, as well as all cells containing the pool of intergenic control sgRNAs, were resampled 100 times with replacement to generate pseudo-bulk profiles. The latter were then normalized and used for protein activity inference, as discussed above. The Stemness Score (NES) of each bootstrapped sample was computed and the overall differential Stemness Score was assessed by Mann Whitney U Test, by comparing the NES of bootstrapped MR-KO samples vs. controls (intergenic sgRNAs).

Cell fitness and gene dependency score: To assess cell fitness after CRISPR/Cas9-mediated MR KO, we first assessed the log_2-scaled fold change (log_2FC) of sgRNA abundance between the CROP-seq profiles and CRISPR library. Note that sgRNA counts were normalized by dividing the count of each sgRNA by the total sgRNA counts in the CROP-seq and CRISPR library, respectively. Then, we assessed MR essentiality by computing a gene dependency score, in which copy number-based bias was corrected in assessing cell fitness effects using the CERES[60] R package. Briefly, we utilized the CCLE copy number data, the gene annotation data from the consensus coding sequence[61] (CCDS), and bowtie[62] indices for human genome (hg19), as provided in the CERES manuscript, to compute a gene dependency score from the log_2FC values.

Note that we added the information of CCDS location for SMAD5 from Ensembl[55] GRCh37 in the gene annotation data as it was omitted in the reference data provided in CERES.

OncoTreat Analysis: To identify CSLC MR-inverter drugs, we used the extensively validated OncoTreat algorithm[48,49,63,64]. For each sample, the CSLC MR signature was assessed by metaVIPER analysis of the 20 cells with the highest vs. the 20 with the lowest stemness score in each of the 7 patient samples, using the *viperSignature*() function in the VIPER[11,12] R package. . Stemness was assessed by integrating the 14 established breast cancer CSLC markers and the CytoTRACE scores. Finally, the z-scores representing the differential protein activities across the 7 patient samples were integrated using the weighted z-test, as described above.

For the OncoTreat analysis we leveraged drug perturbation profiles generated by PLATE-seq profiling[65] of BT-20 cells at 6h and 24h following perturbation with 90 FDA-approved or late-investigational stage compounds, as described in[64]. Each compound was titrated at its 48h GI_{20} (the drug concentration at which 20% of the max inhibition of growth is achieved), as determined by 10-point dose response curves. To assess MR-inversion potential of a drug, we assessed the enrichment of the 25 positive CSLC MRs and 25 negative CSLC MRs in proteins that were inactivated and activated in drug vs. vehicle control treated cells. Drugs were then ranked based on their NES, with the top candidate MR-inverter drug having the most negative NES value. Consistent with previous OncoTreat publications[48,49,63,64], we only considered direct transcriptional regulators (n = 2,088) as candidate MRs.

Bi-clustering of BT-20 perturbation profiles: To identify drugs inducing similar protein activity-level effects in BT20 cells we performed bi-clustering of the perturbational profiles generated at 24h and 1/10 GI_{20} concentration, after gene expression was converted to differential protein activity by metaVIPER analysis, by comparing each treated sample to the pool of vehicle control treated samples. For this analysis, only statistically significant differentially expressed regulatory proteins (TFs/co-TFs) ($p < 10^{-5}$, n = 851) were considered. Bi-clustering was performed using complete hierarchical clustering method based on perturbational protein activity profiles Spearman's correlation.

GO enrichment analysis: Following bi-cluster analysis, 5 MR clusters were identified using the *cutree*() R function. These represent distinct subsets of MR proteins that are differentially activated or inactivated by specific drug subsets. To identify relevant biological pathways associated with each cluster, we performed enrichment analysis of biological pathway Gene Ontology (GO) terms, using the *enrichGO*() function in the clusterProfiler[66] R package (see Supplementary Table 4).

Drug Mechanism of actions in BT-20 data: Information on drug mechanism of actions (MoAs) of individual drugs in the analysis was obtained by manual analysis of the ChEMBL[67] and DrugCentral[68] databases (see Supplementary Table 5).

Visualizing the differential cell density following drug treatment *in vivo*: To visualize cell density changes in drug vs. vehicle control-treated PDX samples, across the entire cell state space, we computed it over a principal component (PC) projection of their protein activity profiles. Single cells treated with either vehicle control, paclitaxel, or albendazole were visualized in independent plots. To avoid bias due to differential cell counts across different treatments, we selected 2,500 cells from each sample at random. Statistical significance of the differential cell densities in drug vs. vehicle-treated samples was assessed by computing the mean and standard deviation of the 2-D kernel density of each sample, by estimating optimal kernel density from 30% of the cells selected 100 times at random from each sample. The 2-D kernel density was computed, using the *kde2d()* function in the MASS[69] R package, with bandwidth h = 0.01, n=50 grid points, and default bivariate normal kernel settings. The kernel density of each drug-treated sample was then converted to a z-score, using the statistical values from the control sample (i.e., vehicle treatment) as the null hypothesis. Therefore, a positive (or negative) z-score indicated an increased (or decreased) cell density in the drug-treated sample, compared to the control.

Hallmark enrichment analysis of the PDX data: To assess the enrichment of 50 MSigDB hallmarks[70] in the direction of the 1st and 2nd PCs of the PDX protein activity profiles, we first ranked the proteins based on the Spearman's correlation between their activity and both the 1st and 2nd PCs, reflecting the greatest variance axes in the data. As a result, proteins were ranked based on the correlation between the ordering of cells along each PC and protein activity. Gene

set enrichment analysis of the 50 cancer hallmark sets in each PC-specific ranked protein list was assessed, using the *aREA()* function in the VIPER R package. Cancer hallmark gene sets were obtained from the MsigDB database[70].

Assessing the Effect of Drug Treatment on the Ratio Between CSLCs and Differentiated Cells: To assess the effect of each drug treatment on cell stemness, we computed the ratio of most stem-like cells (stemness score > 0.8) to most differentiated cells (stemness score < 0.2) in drug vs. vehicle control-treated samples. Statistical significance was computed using the hypergeometric function (i.e., Fisher's Exact Test).

MR Modularity Assessment: For each of the top 20 candidate CSLC MRs, we collected its putative protein-protein interactions (PPIs) from the STRING[71] and PrePPI[72] databases and its putative regulatory interactions from ARACNe-AP[13] analysis. The analysis identified n = 67 molecular interactions in the candidate module comprising these MRs. To determine whether the modularity was significant, we assessed the average connectivity of 1,000 sets of 20 regulatory proteins selected at random, using the same approach to determine PPIs and regulatory interactions. The degree distribution of these random modules (size = 20 proteins) was then fit to a negative-binomial (NB) distribution model, using the *fitdistr()* function from the fitdistrplus[73] R package. The resulting parameters were mu = 13.2, which corresponds to the average expected number of interactions, and size = 5.80, corresponding to a dispersion of the NB distribution with *p-value* = 1, based on a Chi-squared goodness fit test. This null model was then used to assess the statistical significance of 20-MR modularity.

ADDITIONAL RESOURCES

scPLATE-seq protocol: **dx.doi.org/10.17504/protocols.io.s4hegt6**

References (Methods)

1 Clough, E. & Barrett, T. The Gene Expression Omnibus Database. *Methods Mol Biol* **1418**, 93-110 (2016). https://doi.org:10.1007/978-1-4939-3578-9_5

2 Bagnoli, J. W. *et al.* Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nat Commun* **9**, 2937 (2018). https://doi.org:10.1038/s41467-018-05347-6

3 Hather, G. *et al.* Growth rate analysis and efficient experimental design for tumor xenograft studies. *Cancer Inform* **13**, 65-72 (2014). https://doi.org:10.4137/CIN.S13974

4 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301 (2017). https://doi.org:10.1038/nmeth.4177

5 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016). https://doi.org:10.1038/nbt.3437

6 Sanson, K. R. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018). https://doi.org:10.1038/s41467-018-07901-8

7 Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890-903 e815 (2017). https://doi.org:10.1016/j.cell.2017.01.013

8 Tan, X. *et al.* Interrogation of genome-wide, experimentally dissected gene regulatory networks reveals mechanisms underlying dynamic cellular state control. *bioRxiv* **2021.06.28.449297** (2021).

9 Schmierer, B. *et al.* CRISPR/Cas9 screening using unique molecular identifiers. *Mol Syst Biol* **13**, 945 (2017). https://doi.org:10.15252/msb.20177834

10 Nguyen, Q. H. *et al.* Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity. *Nat Commun* **9**, 2028 (2018). https://doi.org:10.1038/s41467-018-04334-1

11 Ding, H. *et al.* Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat Commun* **9**, 1471 (2018). https://doi.org:10.1038/s41467-018-03843-3

12 Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838-847 (2016). https://doi.org:10.1038/ng.3593

13 Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233-2235 (2016). https://doi.org:10.1093/bioinformatics/btw216

14 Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-390 (2005). https://doi.org:10.1038/ng1532

15 Piovan, E. *et al.* Direct reversal of glucocorticoid resistance by AKT inhibition in acute lymphoblastic leukemia. *Cancer Cell* **24**, 766-776 (2013). https://doi.org:10.1016/j.ccr.2013.10.022

16 Elyada, E. *et al.* Cross-Species Single-Cell Analysis of Pancreatic Ductal Adenocarcinoma Reveals Antigen-Presenting Cancer-Associated Fibroblasts. *Cancer discovery* **9**, 1102-1123 (2019). https://doi.org:10.1158/2159-8290.CD-19-0094

17 Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* **184**, 2988-3005 e2916 (2021). https://doi.org:10.1016/j.cell.2021.04.038

18 Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113-1120 (2013). https://doi.org:10.1038/ng.2764

19 Aleksandar Obradovic, L. V., Pasquale Laise, Jeremy Worley, Xiangtian Tan, Alec Wang, Andrea Califano. PISCES: A pipeline for the Systematic, Protein Activity-based Analysis of Single Cell RNA Sequencing Data. *bioRxiv* (2021).
20 Baran, Y. *et al.* MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biol* **20**, 206 (2019). https://doi.org:10.1186/s13059-019-1812-2
21 Aleksandar Obradovic, V. O. P. V., Pasquale Laise, Jeremy Worley, Xiangtian Tan, Alec Wang, View ORCID ProfileAndrea Califano. PISCES: A pipeline for the Systematic, Protein Activity-based Analysis of Single Cell RNA Sequencing Data. *BioRxiv* (2021). https://doi.org:DOII: 10.1101/2021.05.20.445002
22 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048
23 Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71 (2016). https://doi.org:10.1093/nar/gkv1507
24 Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **375**, 1109-1112 (2016). https://doi.org:10.1056/NEJMp1607591
25 Gulati, G. S. *et al.* Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411 (2020). https://doi.org:10.1126/science.aax0249
26 Sheridan, C. *et al.* CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res* **8**, R59 (2006). https://doi.org:10.1186/bcr1610
27 Vassilopoulos, A., Chisholm, C., Lahusen, T., Zheng, H. & Deng, C. X. A critical role of CD29 and CD49f in mediating metastasis for cancer-initiating cells isolated from a Brca1-associated mouse model of breast cancer. *Oncogene* **33**, 5477-5482 (2014). https://doi.org:10.1038/onc.2013.516
28 Shimono, Y. *et al.* Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell* **138**, 592-603 (2009). https://doi.org:10.1016/j.cell.2009.07.011
29 Tatetsu, H. *et al.* SALL4, the missing link between stem cells, development and cancer. *Gene* **584**, 111-119 (2016). https://doi.org:10.1016/j.gene.2016.02.019
30 BeLow, M. & Osipo, C. Notch Signaling in Breast Cancer: A Role in Drug Resistance. *Cells* **9** (2020). https://doi.org:10.3390/cells9102204
31 Yu, F. *et al.* Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. *Oncogene* **30**, 2161-2172 (2011). https://doi.org:10.1038/onc.2010.591
32 Xu, X., Zhang, M., Xu, F. & Jiang, S. Wnt signaling in breast cancer: biological mechanisms, challenges and opportunities. *Mol Cancer* **19**, 165 (2020). https://doi.org:10.1186/s12943-020-01276-5
33 Lo, P. K. *et al.* CD49f and CD61 identify Her2/neu-induced mammary tumor-initiating cells that are potentially derived from luminal progenitors and maintained by the integrin-TGFbeta signaling. *Oncogene* **31**, 2614-2626 (2012). https://doi.org:10.1038/onc.2011.439
34 Vaillant, F. *et al.* The mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models of mammary tumorigenesis. *Cancer Res* **68**, 7711-7717 (2008). https://doi.org:10.1158/0008-5472.CAN-08-1949
35 Barnawi, R. *et al.* beta1 Integrin is essential for fascin-mediated breast cancer stem cell function and disease progression. *Int J Cancer* **145**, 830-841 (2019). https://doi.org:10.1002/ijc.32183
36 Brugnoli, F., Grassilli, S., Al-Qassab, Y., Capitani, S. & Bertagnolo, V. CD133 in Breast Cancer Cells: More than a Stem Cell Marker. *J Oncol* **2019**, 7512632 (2019). https://doi.org:10.1155/2019/7512632
37 Wang, Y. J. & Herlyn, M. The emerging roles of Oct4 in tumor-initiating cells. *Am J Physiol Cell Physiol* **309**, C709-718 (2015). https://doi.org:10.1152/ajpcell.00212.2015

38 Leis, O. *et al.* Sox2 expression in breast tumours and activation in breast cancer stem cells. *Oncogene* **31**, 1354-1365 (2012). https://doi.org:10.1038/onc.2011.338
39 Lennartsson, J. & Ronnstrand, L. Stem cell factor receptor/c-Kit: from basic science to clinical implications. *Physiol Rev* **92**, 1619-1649 (2012). https://doi.org:10.1152/physrev.00046.2011
40 Toledo-Guzman, M. E., Hernandez, M. I., Gomez-Gallegos, A. A. & Ortiz-Sanchez, E. ALDH as a Stem Cell Marker in Solid Tumors. *Curr Stem Cell Res Ther* **14**, 375-388 (2019). https://doi.org:10.2174/1574888X13666180810120012
41 Begicevic, R. R. & Falasca, M. ABC Transporters in Cancer Stem Cells: Beyond Chemoresistance. *Int J Mol Sci* **18** (2017). https://doi.org:10.3390/ijms18112362
42 Gazit, R. *et al.* Fgd5 identifies hematopoietic stem cells in the murine bone marrow. *J Exp Med* **211**, 1315-1331 (2014). https://doi.org:10.1084/jem.20130428
43 Chen, J. Y. *et al.* Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* **530**, 223-227 (2016). https://doi.org:10.1038/nature16943
44 Rehman, S. K. *et al.* Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell* **184**, 226-242 e221 (2021). https://doi.org:10.1016/j.cell.2020.11.018
45 Ting, S. B. *et al.* Asymmetric segregation and self-renewal of hematopoietic stem and progenitor cells with endocytic Ap2a2. *Blood* **119**, 2510-2522 (2012). https://doi.org:10.1182/blood-2011-11-393272
46 Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012). https://doi.org:10.1038/nature11003
47 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576 e516 (2017). https://doi.org:10.1016/j.cell.2017.06.010
48 Vasciaveo, A. *et al.* OncoLoop: A Network-Based Precision Cancer Medicine Framework. *Cancer Discov* **13**, 386-409 (2023). https://doi.org:10.1158/2159-8290.CD-22-0342
49 Alvarez, M. J. *et al.* A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat Genet* **50**, 979-989 (2018). https://doi.org:10.1038/s41588-018-0138-4
50 Prat, A. *et al.* Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* **142**, 237-255 (2013). https://doi.org:10.1007/s10549-013-2743-3
51 Yamaguchi, N. *et al.* Constitutive activation of nuclear factor-kappaB is preferentially involved in the proliferation of basal-like subtype breast cancer cell lines. *Cancer Sci* **100**, 1668-1674 (2009). https://doi.org:10.1111/j.1349-7006.2009.01228.x
52 Yamamoto, M. *et al.* NF-kappaB non-cell-autonomously regulates cancer stem cell populations in the basal-like breast cancer subtype. *Nat Commun* **4**, 2299 (2013). https://doi.org:10.1038/ncomms3299
53 Kang, H. M. *et al.* Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol* **36**, 89-94 (2018). https://doi.org:10.1038/nbt.4042
54 Katz, K. *et al.* The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res* **50**, D387-D390 (2022). https://doi.org:10.1093/nar/gkab1053
55 Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res* **50**, D988-D995 (2022). https://doi.org:10.1093/nar/gkab1049
56 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017). https://doi.org:10.1038/ncomms14049
57 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303 (2010). https://doi.org:10.1101/gr.107524.110

58 Xin, H. *et al.* GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. *Genome Biol* **21**, 188 (2020). https://doi.org:10.1186/s13059-020-02084-2

59 Murphy, A. E. & Skene, N. G. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat Commun* **13**, 7851 (2022). https://doi.org:10.1038/s41467-022-35519-4

60 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet* **49**, 1779-1784 (2017). https://doi.org:10.1038/ng.3984

61 Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res* **19**, 1316-1323 (2009). https://doi.org:10.1101/gr.080531.108

62 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009). https://doi.org:10.1186/gb-2009-10-3-r25

63 Obradovic, A. *et al.* Systematic elucidation and pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. *Cancer Cell* **41**, 933-949 e911 (2023). https://doi.org:10.1016/j.ccell.2023.04.003

64 Mundi, P. S. *et al.* A Transcriptome-Based Precision Oncology Platform for Patient-Therapy Alignment in a Diverse Set of Treatment-Resistant Malignancies. *Cancer discovery* **13**, 1386-1407 (2023). https://doi.org:10.1158/2159-8290.CD-22-1020

65 Bush, E. C. *et al.* PLATE-Seq for genome-wide regulatory network analysis of high-throughput screens. *Nat Commun* **8**, 105 (2017). https://doi.org:10.1038/s41467-017-00136-z

66 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021). https://doi.org:10.1016/j.xinn.2021.100141

67 Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **40**, D1100-1107 (2012). https://doi.org:10.1093/nar/gkr777

68 Ursu, O. *et al.* DrugCentral: online drug compendium. *Nucleic acids research* **45**, D932-D939 (2017). https://doi.org:10.1093/nar/gkw993

69 Ripley, W. N. V. a. B. D. *Modern Applied Statistics with S*. Fourth edn, (Springer, 2002).

70 Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015). https://doi.org:10.1016/j.cels.2015.12.004

71 Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic acids research* **49**, D605-D612 (2021). https://doi.org:10.1093/nar/gkaa1074

72 Zhang, Q. C., Petrey, D., Garzon, J. I., Deng, L. & Honig, B. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res* **41**, D828-833 (2013). https://doi.org:10.1093/nar/gks1231

73 Delignette-Muller, M. L. & Dutang, C. fitdistrplus: An R Package for Fitting Distributions. *Journal of Statistical Software* **64**, 1 - 34 (2015). https://doi.org:10.18637/jss.v064.i04

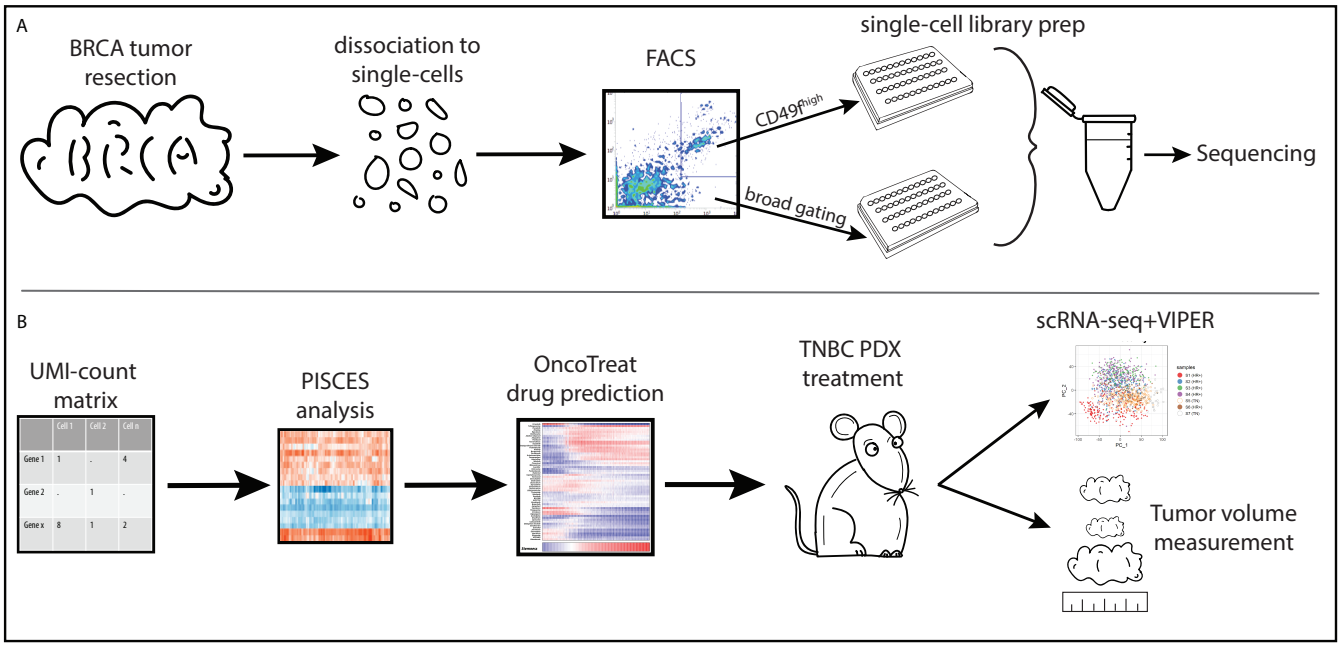


Figure 1. Overview of the workflow. **A.** The experimental workflow for generating scRNA-seq data from breast cancer cells from patient samples. FACS was used to enrich CSLCs. **B.** A systems biology approach to identifying a candidate drug targeting the CSLCs and subsequent experimental validations.

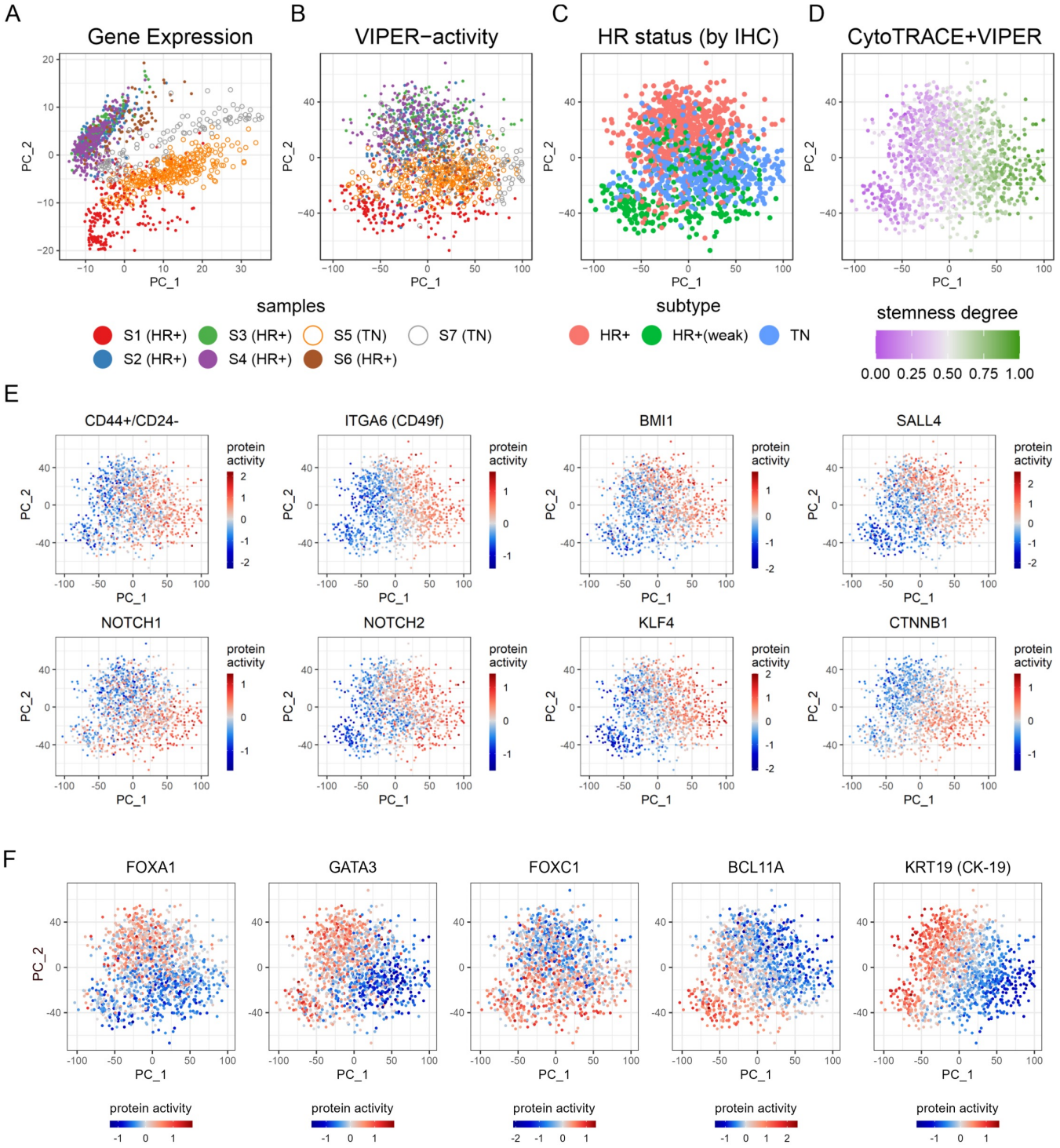


Figure 2. Analysis of scRNA-seq data for 7 breast cancer patient samples. Cells were clustered based on the first two principal components of the cell's gene expression (**A**) and the protein activity inferred by VIPER (**B-F**). In **A** and **B,** cells are colored according to the patient they came from. **C.** The breast cancer subclasses (HR+, weakly HR+, and TN) are shown. **D**. The degree of each cell's stemness is indicated using a green-grey-purple color gradient, corresponding to the degree of stemness from one (most stem-like) to zero (most differentiated). The stemness degree was estimated based on the combination of the CytoTRACE score and the protein activities of well-known stemness markers: CD44+/CD24-, ITGA6, BMI1, SALL4, NOTCH1, NOTCH2, KLF4, CTNNB1, ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT (see Methods). This stemness degree score was re-scaled to the range between 0 and 1. **E**. The VIPER-inferred protein activity (centered) of individual breast CLSC markers. From the highest to the lowest, activity is shown with a red-white-blue color gradient (white = mean). **F**. The VIPER-inferred protein activity of HR+ markers (FOXA1 and GATA3), TNBC makers (FOXC1 and BCL11A), and a differentiated-cell marker (KRT19).

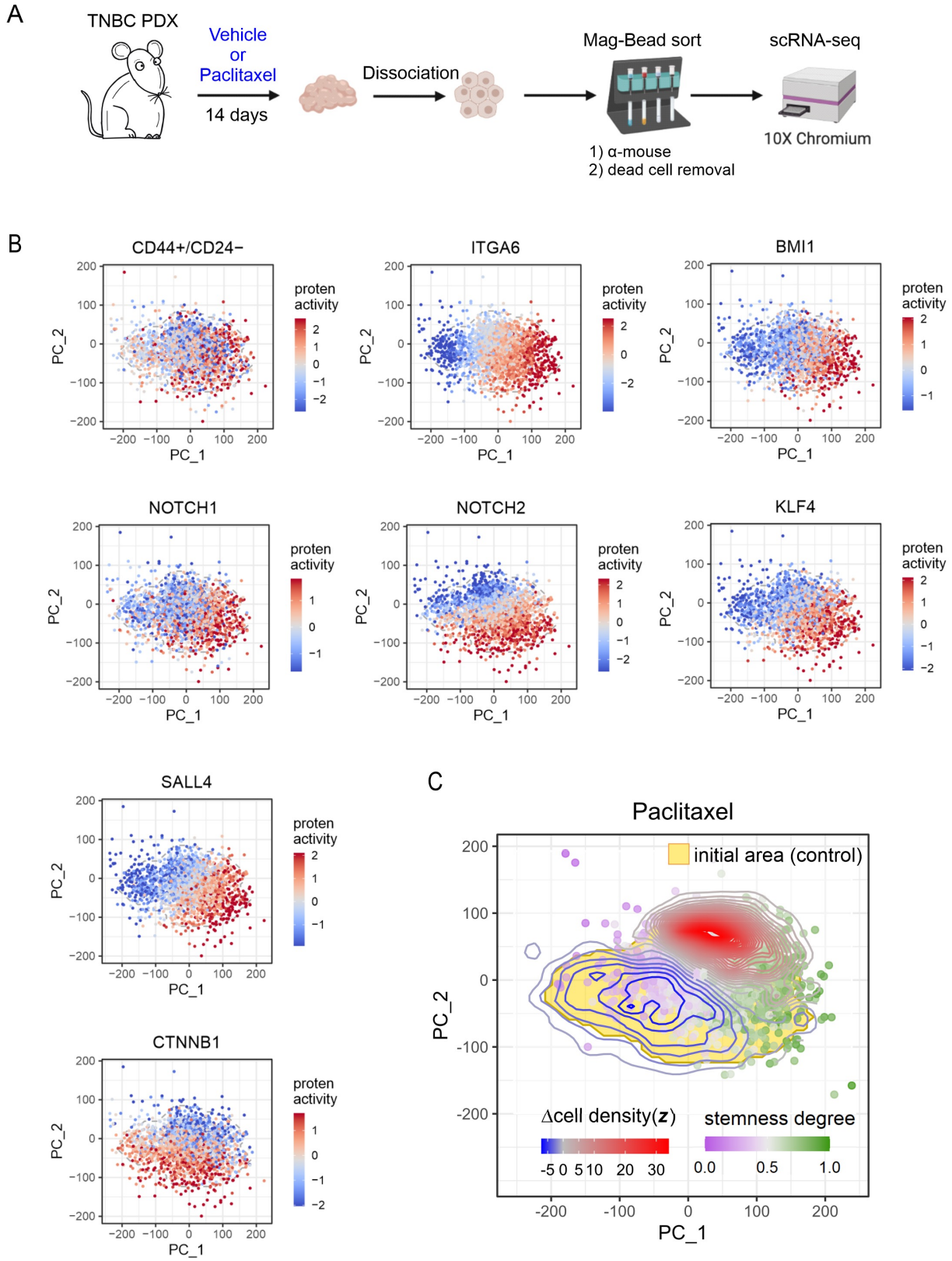


Figure 3. A. Workflow for scRNA-seq analysis of a TNBC PDX model. **B.** The VIPER-inferred activities of established breast CSLC markers in the vehicle control. **C.** Effect of Paclitaxel on the TNBC PDX cells. The cells were clustered based on the first two principal components of VIPER-inferred protein activity profiles under vehicle- and drug-treated conditions (see Methods). Based on the degree of stemness, cells were colored in a green-grey-purple color scheme (green: more stem-like cells, purple: more differentiated cells). The stemness degree was estimated by the combination of the CytoTRACE score and the protein activities of stemness markers (CD44+/CD24-, ITGA6, BMI1, SALL4, NOTCH1, NOTCH2, KLF4, CTNNB1, ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT) (see Methods). The estimated stemness degree score was rescaled to the range of 0-1. The area in yellow indicates a boundary of the cell cluster which 95% of cells in the control fall into. Cell density change (z-score) is shown with contour lines in the PDX sample treated with paclitaxel. The red and blue contour lines denote an increase or decrease, respectively, in cell densities under drug treatment compared to the control.

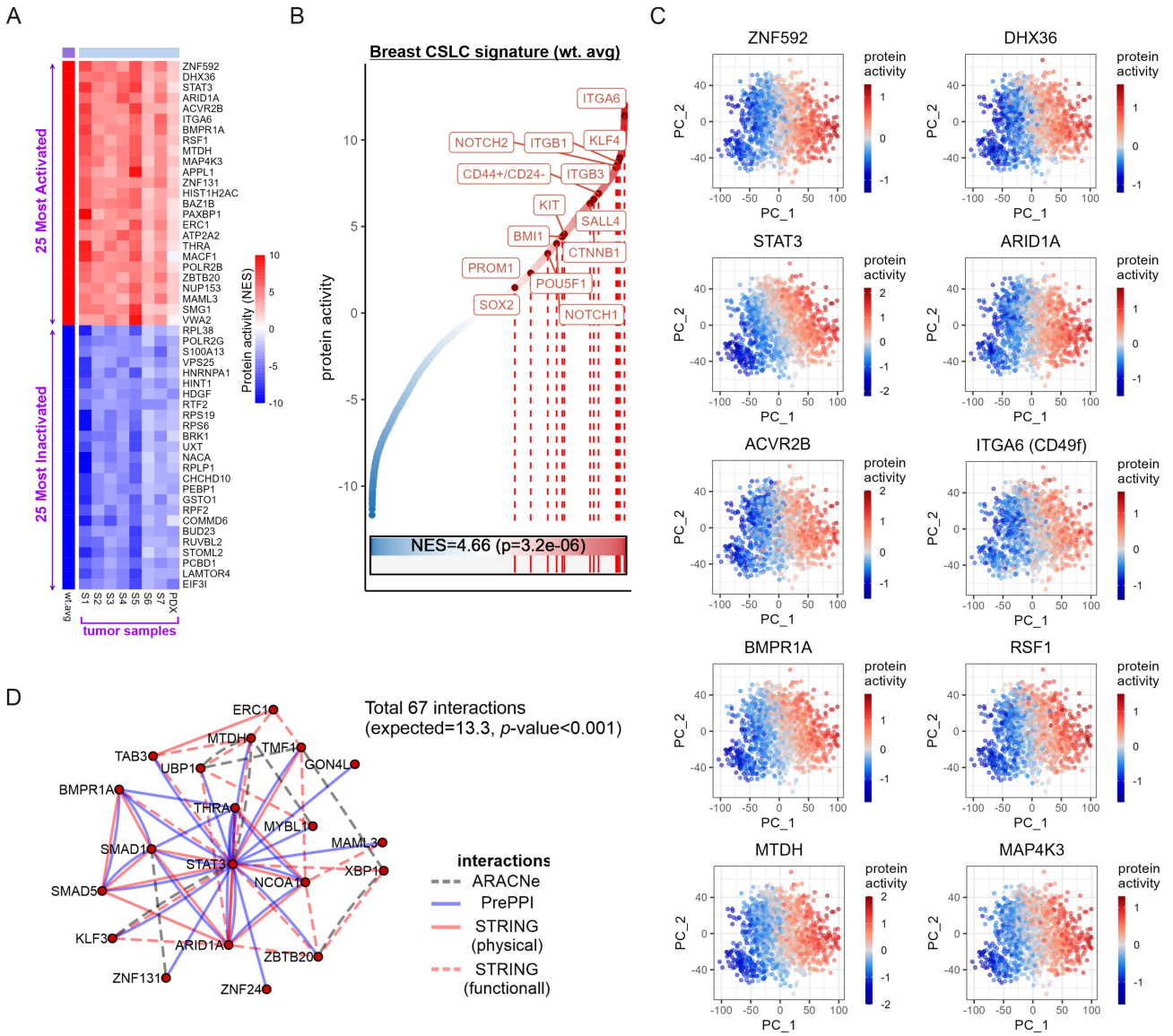


Figure 4. A. A heatmap showing the VIPER-inferred protein activity of the 25 most activated and the 25 most inactivated proteins in the breast CSLC signature and their activities in individual samples (7 patient samples and the PDX vehicle-treated sample). For each sample, differential protein activity from non-CSLCs to CSLC was computed using metaVIPER. The overall CSLC signature was obtained by the weighted average of the protein activities across samples. A larger positive (or negative) value in the signature means that the protein was more (or less) activated in CSLCs than in non-CSLCs. If there is little change in protein activity between non-CSLCs and CSLCs, the value approaches zero. Note that CSLCs and non-CSLCs were identified based on the average activity of the following CSLC markers in the sample: CD44+/CD24-, ITGA6, BMI1, SALL4, NOTCH1, NOTCH2, KLF4, CTNNB1, ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT. **B.** A waterfall plot displaying the sorted protein activities in the breast CSLC signature, in which the signatures of individual samples were integrated using weighted Stouffer's method. In this plot, the NES of the 14 breast CSLC markers is shown. **C.** Top 10 activated proteins in the identified signature and their protein activities in the patient data. **D.** Top 20 transcriptional regulators in the identified breast CSLC signature and their interactions identified by ARACNe, PrePPI, and STRING tools.

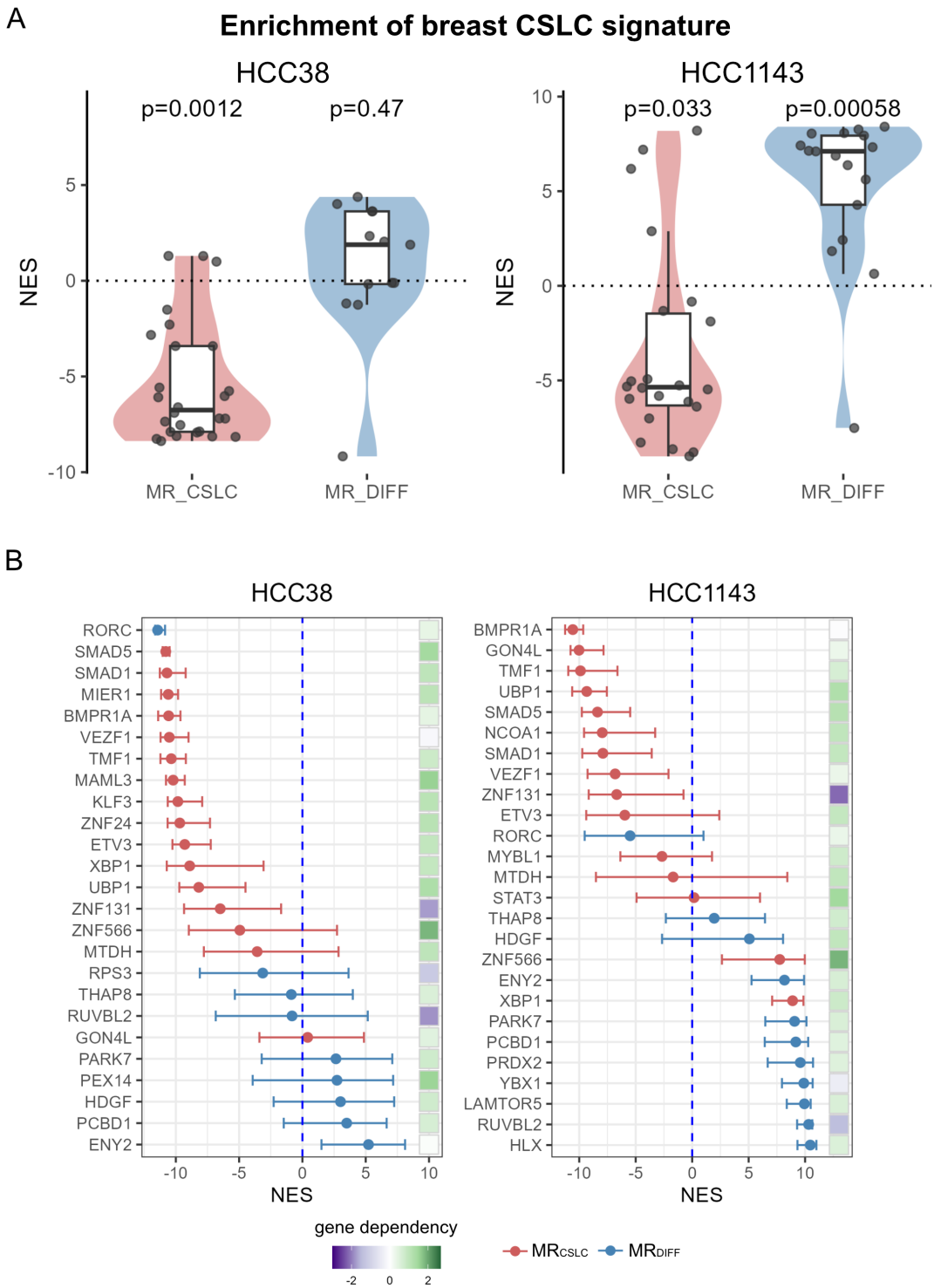


Figure 5. A. Cellular reprogramming after knocking out the top 25 most activated MRs (MR_{CSLC}) and the 25 most inactivated MRs (MR_{DIFF}) of breast CSLC signature, compared to the effects in the control sgRNA group for the breast cancer cell lines HCC38 and HCC1143. For each sgRNA, the VIPER-inferred protein activity profiles were generated from the pseudo-bulk expression of cells detected with the same sgRNA. The knockout (KO) efficiency was determined based on the threshold of one standard deviation below the target gene's mean protein activity. The enrichment score of the 50-MR set (MR_{CSLC} and MR_{DIFF}) was investigated in the protein activity profiles for each group of MR_{CSLC} and MR_{DIFF} (**A**) and for each sgRNA (**B**) to assess the effects of cellular reprogramming for both HCC38 and HCC1143 cell lines. To assess the reprogramming effects of each sgRNA, pseudo-bulk expressions were bootstrapped by resampling cells with the same sgRNA with replacement. Lower NES signifies greater differentiation. Error bars indicate the 1st and 3rd quartiles of NES for the reprogramming effects of multiple sgRNAs targeting the same MR in B. The effect of each MR on cell proliferation in the CROP-seq experiment is indicated by the gene dependency score, using a green-white-purple gradient where darker purple = greater dependence. White/green indicates no significant reduction in the proliferation rate when that MR is knocked out.

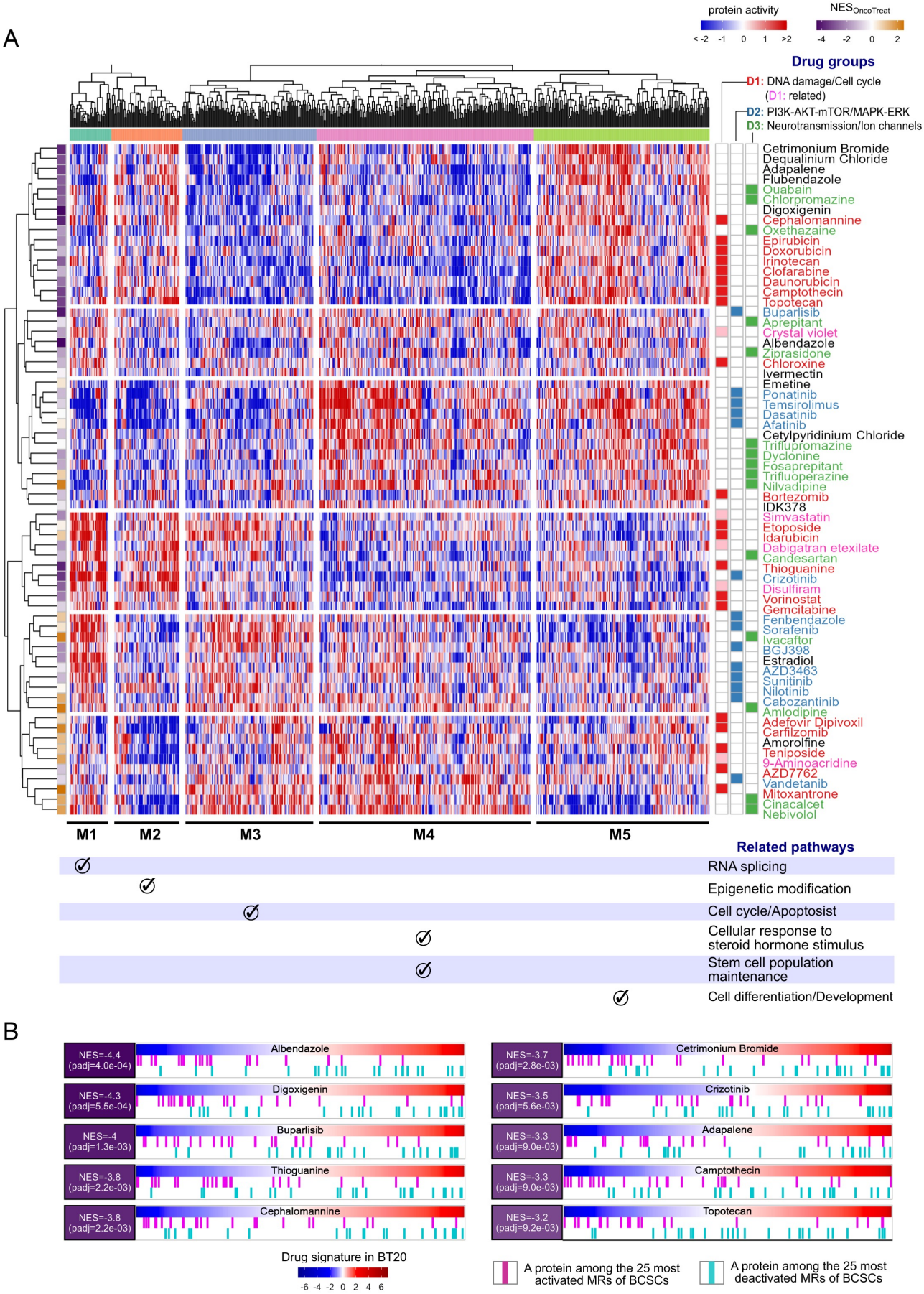


Figure 6 A. Bi-clustered drug perturbation profiles for the breast cancer cell line BT20. In the heatmap, the rows and columns are drug samples (24h, 1/10th EC20) and master regulator proteins ($FDR_{BCSC}<1\times10^{-5}$), respectively. The activated and inactivated proteins are shown in red and blue, and the protein activities with no change are shown in white. **B.** The enrichment plot of the top 10 drugs, predicted from the perturbation profiles with 24h treatment at 1/10th the drug's EC20 in BT20 using OncoTreat analysis. The magenta and turquoise bars denote the top 25 most activated proteins and the top 25 most inactivated proteins in the breast CSLC signature, respectively, which were derived from the 7 patient samples. In each plot, these 50 proteins (i.e. magenta and turquoise bars) were mapped to their corresponding activity in a drug sample.

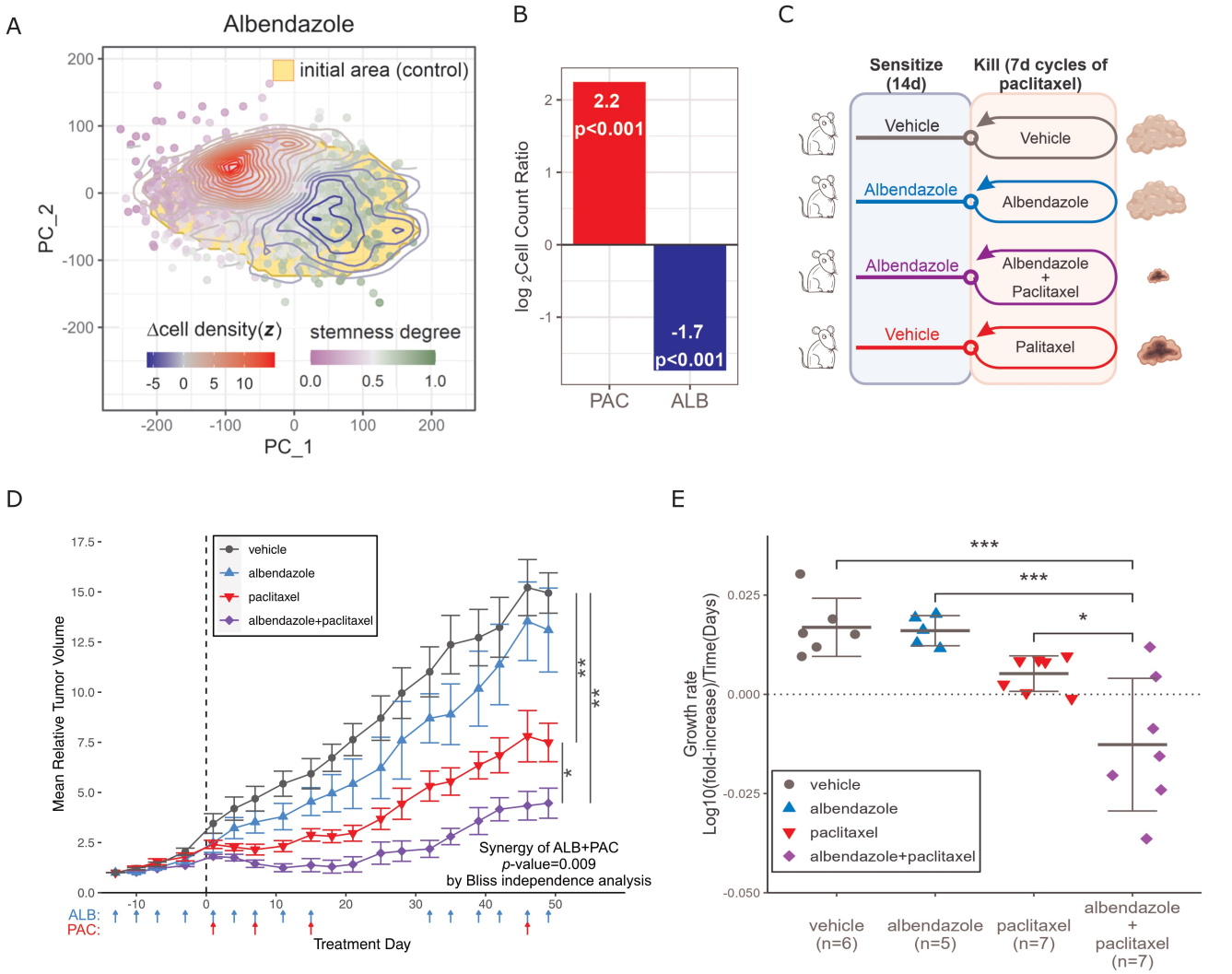


Figure 7. A. Analysis of scRNA-seq data showing the effect of albendazole on cells taken from a TNBC PDX model. The red and blue contour lines denote an increase or decrease, respectively, in cell densities under drug treatment compared to the vehicle control. The area in yellow indicates a boundary of the cell cluster into which 95% of cells in the control fall. **B.** The cell-count ratios between the stem-like (stemness score > 0.8) and differentiated (stemness score < 0.2) cells under the treatments of paclitaxel and albendazole compared to vehicle control. Based on Fisher's exact tests, the differences between the treatments are statistically significant (p-value < 0.001) compared to the ratio in the control. **C.** Schematic view of combination therapies used in the preclinical tests. **D.** Mean relative tumor volumes over time under individual therapeutic strategies. Biological replicates were averaged, and the error bar indicates one standard error of the mean. Mice were treated with albendazole 3 times weekly for two weeks (Day -13 to Day 0) before the start date of the combined drug therapy with paclitaxel to sensitize the tumor cells. Mice with albendazole monotherapy were treated for the same amount of time as those in the combination therapy (Day 0 to Day 49). **E.** A comparison of the tumor growth rates during the 1st cycle of drug treatments. Differences between mean growth rates were tested for statistical significance using Tukey's honest significance test (*p<0.05, **p<0.01, ***p<0.001). Tumor growth rates were calculated assuming exponential kinetics from 1 to 18 days. The error bar indicates one standard deviation from the mean growth rates.

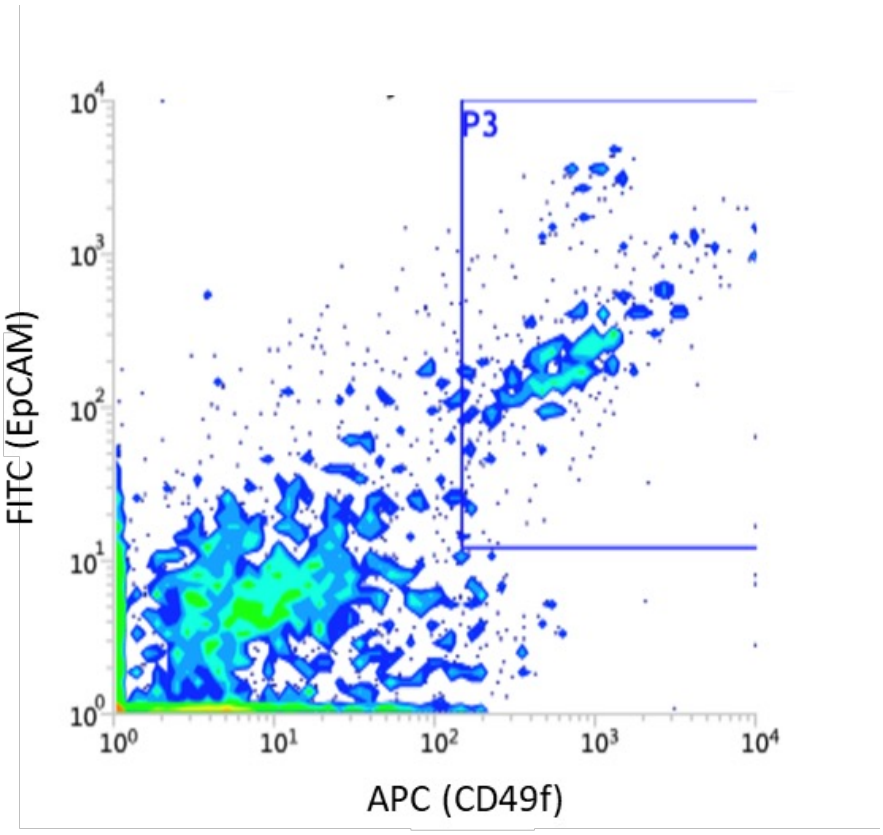


Figure S1. FACS of a human metastatic breast cancer tissue (patient S1). Malignant epithelial cells were isolated based on EPCAM$^+$ and CD49f^{high} and subjected to scRNA-seq.

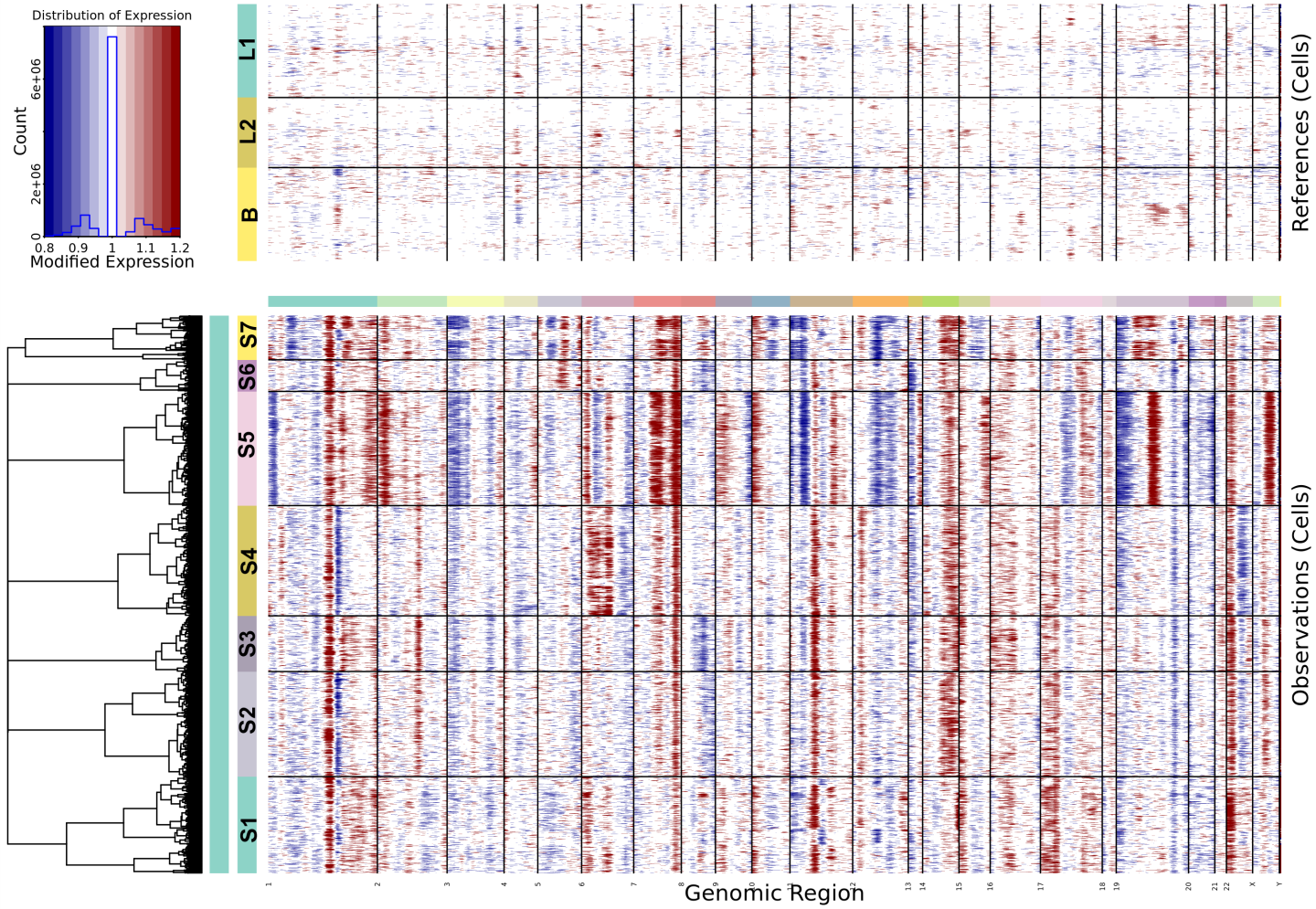


Figure S2. Inferred copy number variations in breast cancer patient samples (S1-S7) by inferCNV. Normal breast epithelial cells containing one basal (B) and two luminal (L1 and L2) subtypes from 4 individuals (GSE113196) were used as reference. In the patient samples, S1 and S2 were weakly HR+, S2, S3, and S4 were strongly HR+, and S5 and S7 were TNBCs, according to immunohistochemistry (IHC).

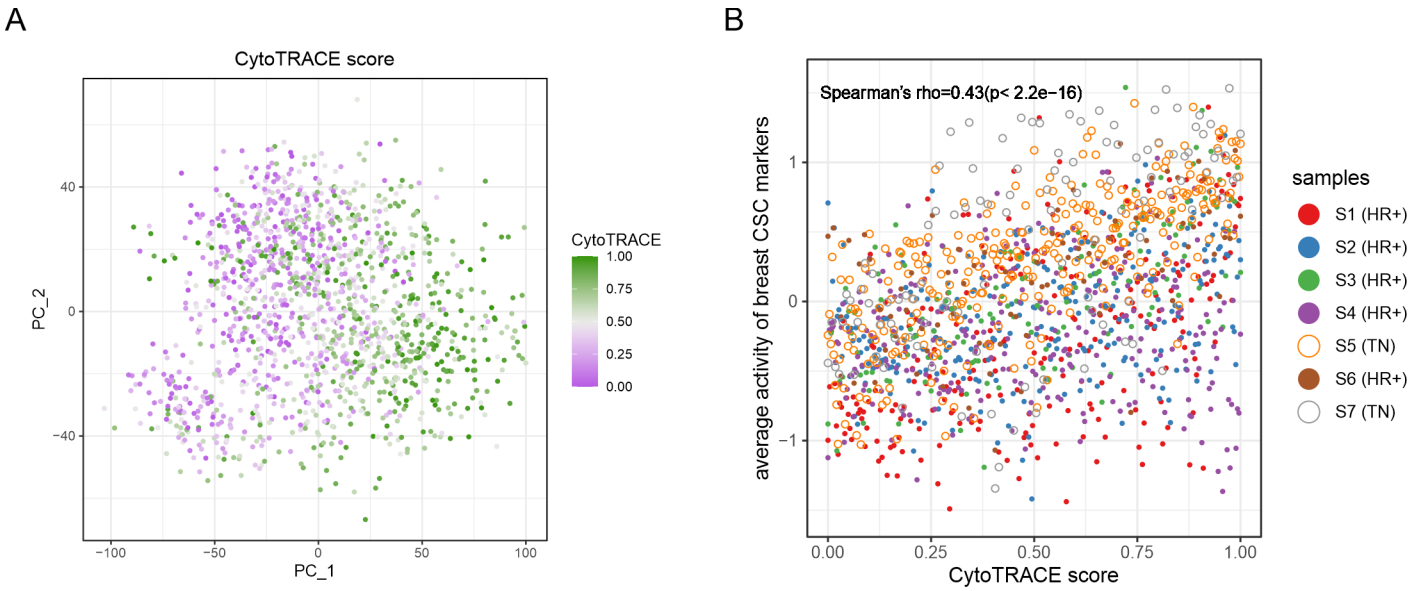


Figure S3. A. The CytoTRACE prediction in the patient scRNA-seq data. Cells were colored with a green-grey-purple gradient corresponding to the CytoTRACE score from one (least differentiated) to zero (most differentiated). **B.** The correlation between the CytoTRACE score and the average activity of breast CSC markers in the patient samples. The following proteins were considered as breast CSC markers in the patient data: CD44+/CD24-, ITGA6, BMI1, SALL4, NOTCH1, NOTCH2, KLF4, CTNNB1, ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT. The correlation between CytoTRACE and the maker's activity was statistically significant (Spearman's rho=0.43, *p*-value<0.001).

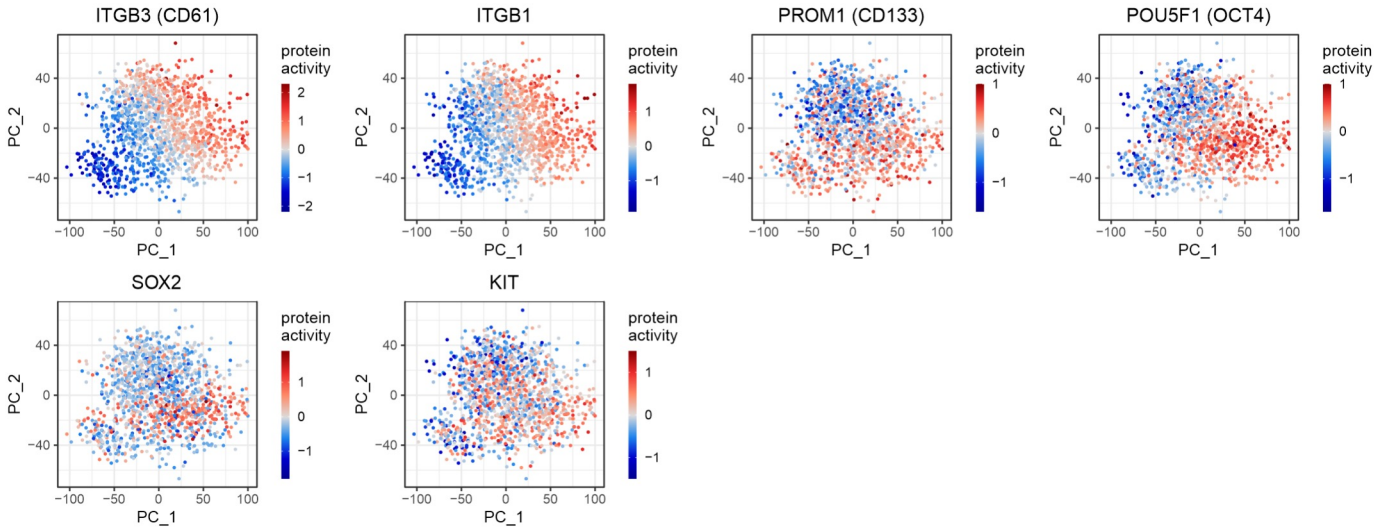


Figure S4. The inferred activity of several breast CSC markers, namely ITGB3, ITGB1, PROM1, POU5F1, SOX2, and KIT. Note that the VIPER-inferred activities were centered in this visualization.

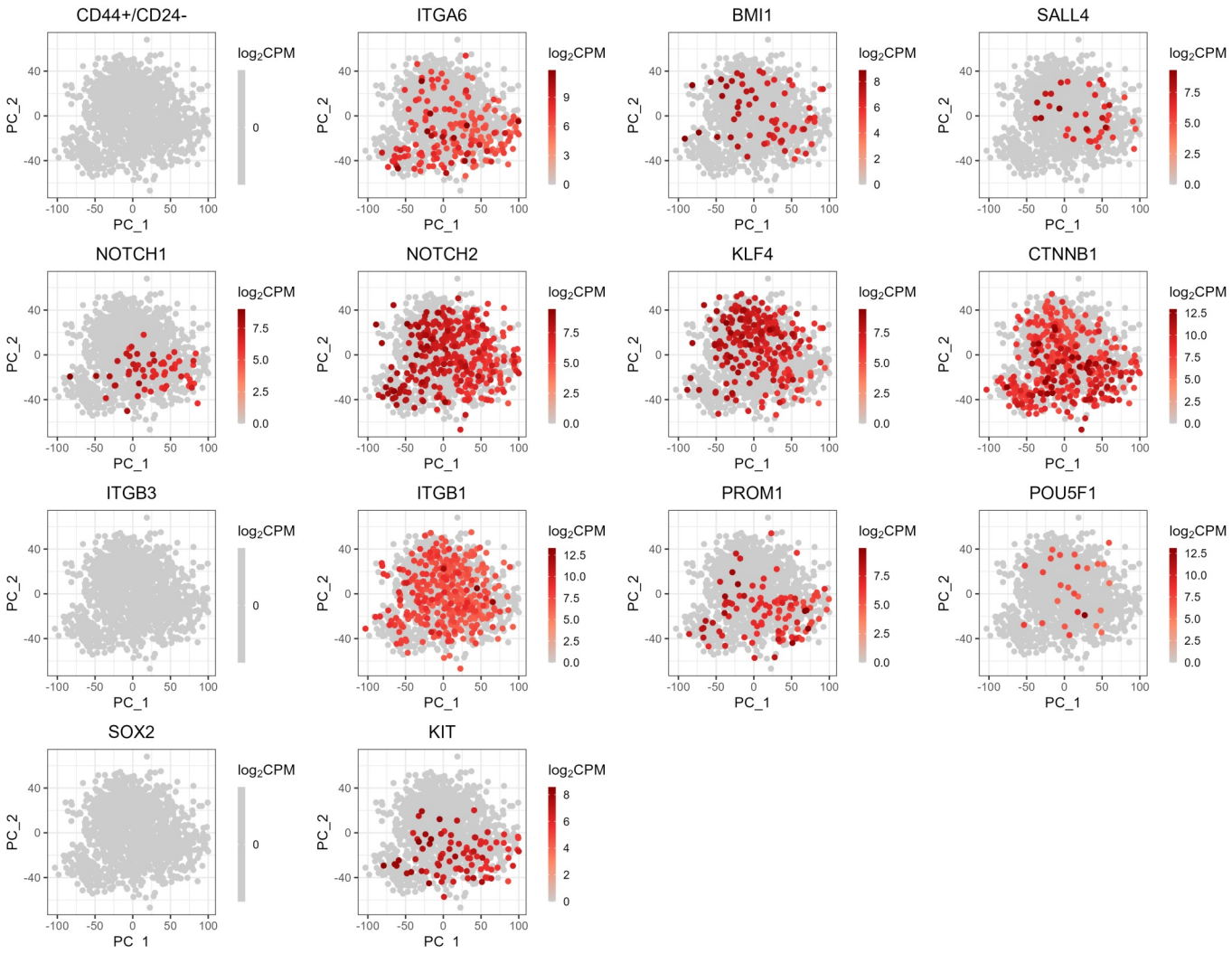


Figure S5. Gene expression (log_2CPM) of known breast CSLC markers. The same PCA mapping based on the protein activity was used for the comparison. Note that only the gene expression greater than the upper quantile is shown for visualization purposes.

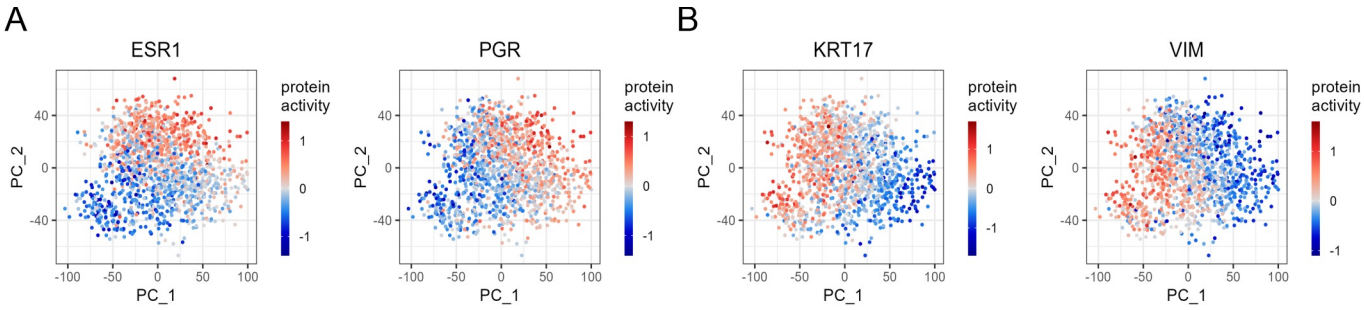


Figure S6. The inferred activity of breast cancer subtype markers. **A.** Two HRs (ESR1 and PGR). **B.** A basal-type marker (KRT17) and non-HR+ marker (VIM). While ESR1 and PGR activities are greater in HR+ patient samples, KRT17 and VIM activities are greater in TNBC samples. Note that the VIPER-inferred activities are centered in this visualization.

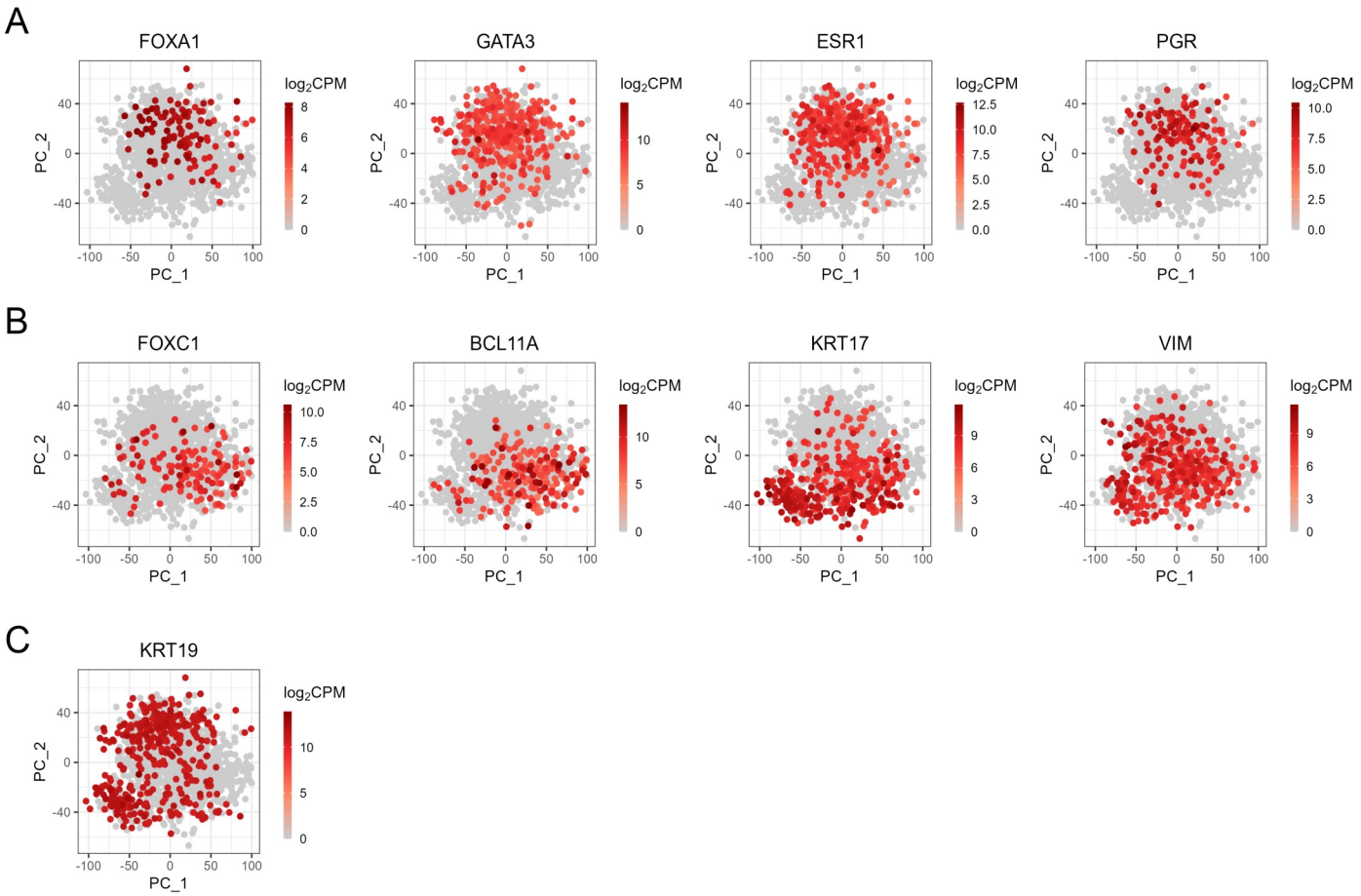


Figure S7. Gene expression (log_2CPM) of known breast cancer subtype markers in the patient samples: HR+ markers (**A**), TNBC-specific markers (**B**), and a differentiated cell marker (**C**). The same PCA-mapping based on the protein activity was used for the comparison. Note that only the gene expression greater than the upper quantile is shown for visualization purposes.

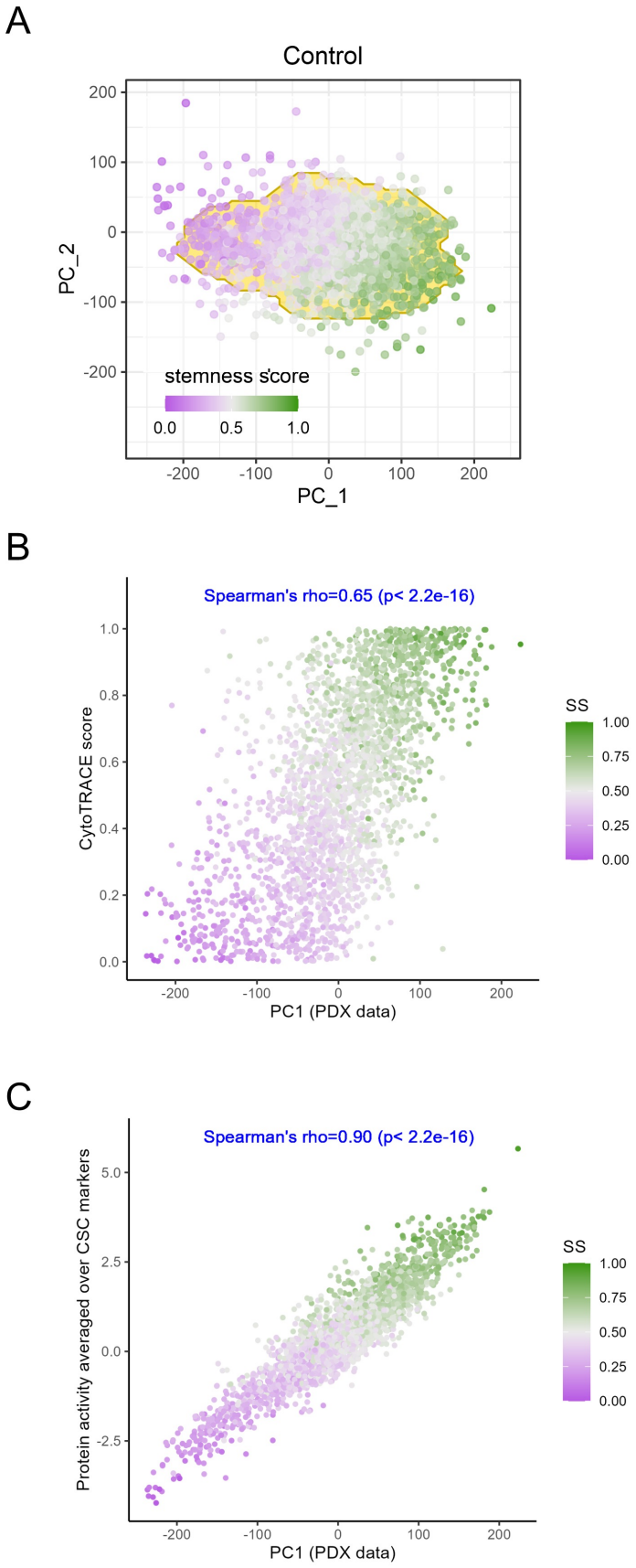


Figure S8. Protein activity-based cell clustering in the PDX control sample (i.e. vehicle treatment). The color of cells indicates the stemness score (SS) as described in Methods. **A.** The first two principal components (PCs) of the protein activity of cells. **B.** The correlation between the 1st PC and the CytoTRACE. **C.** The correlation between the 1st PC and the averaged activity of the 14 CSC markers.

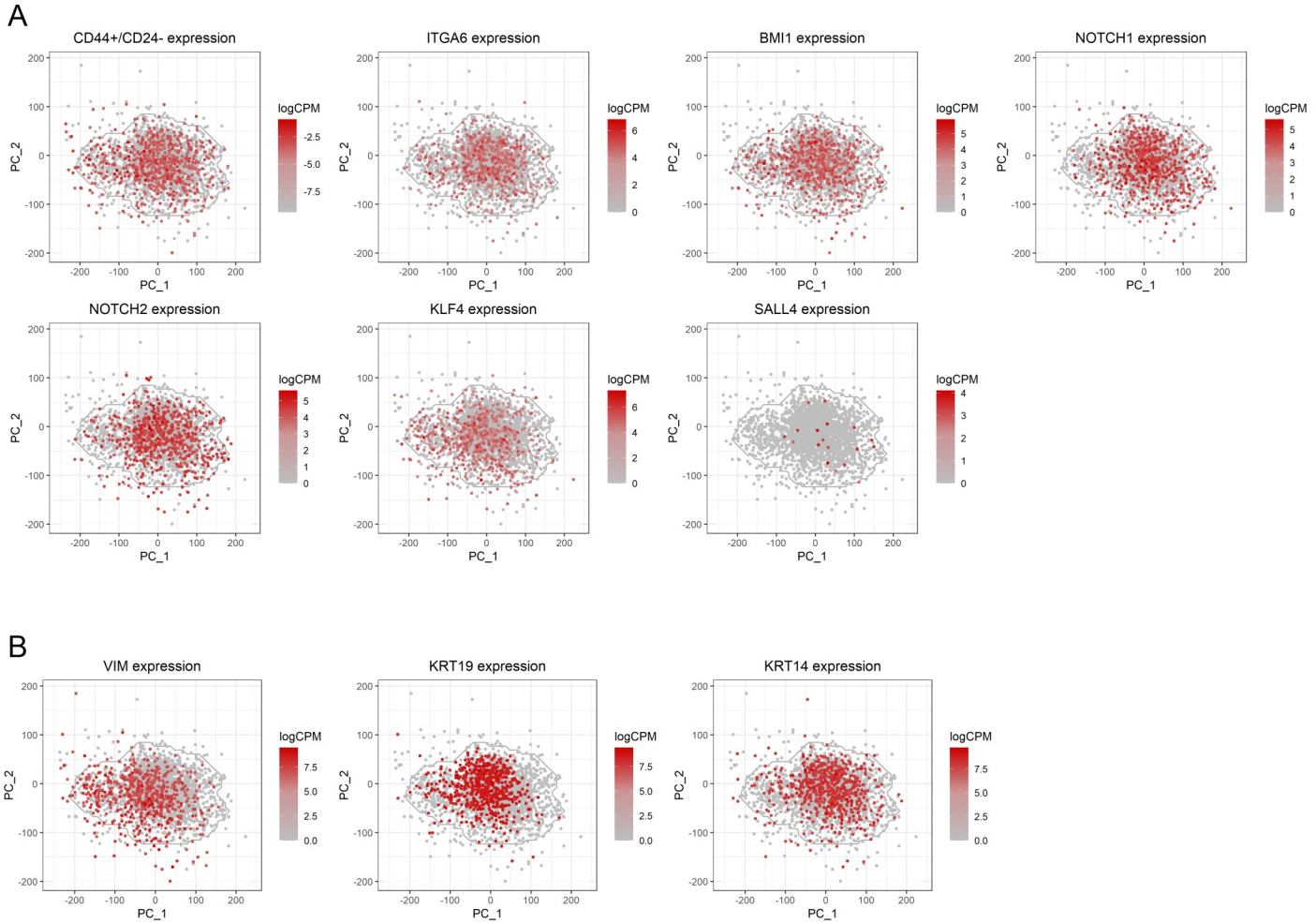


Figure S9. Gene expression (log_2CPM) of known breast CSC markers (**A**) and differentiated markers (**B**) in the PDX control sample. Protein activity-based PCA-mapping was used for the comparison. Note that only the gene expression greater than the upper quantile was shown for a visualization purpose.

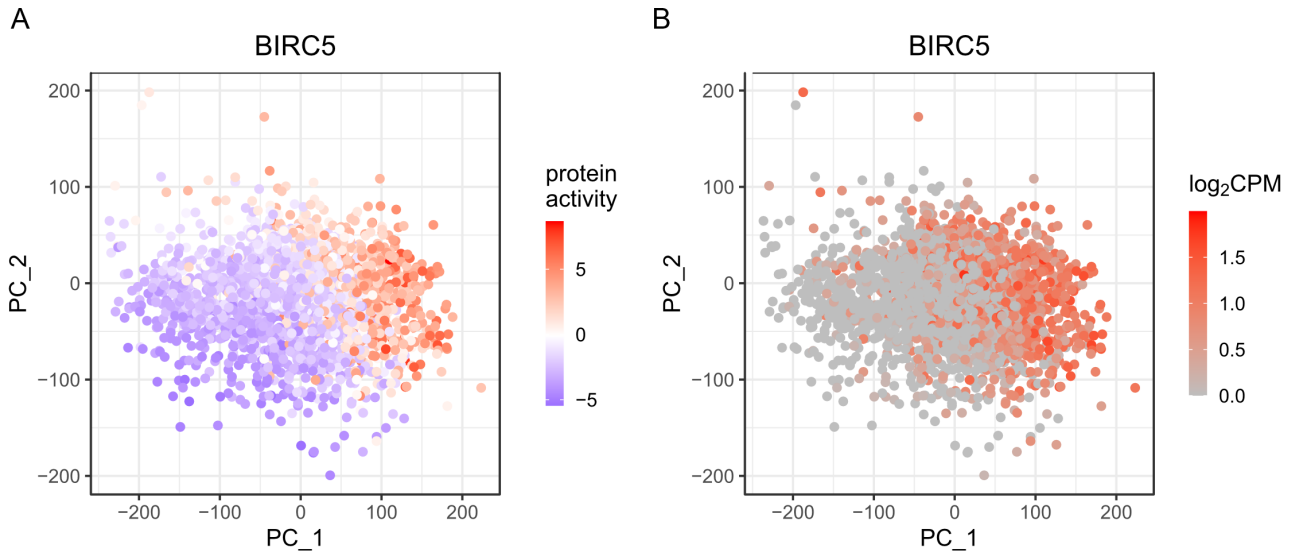


Figure S10. Protein activity (**A**) and gene expression (log_2CPM) (**B**) of BIRC5, a marker of quiescent-breast CSCs.

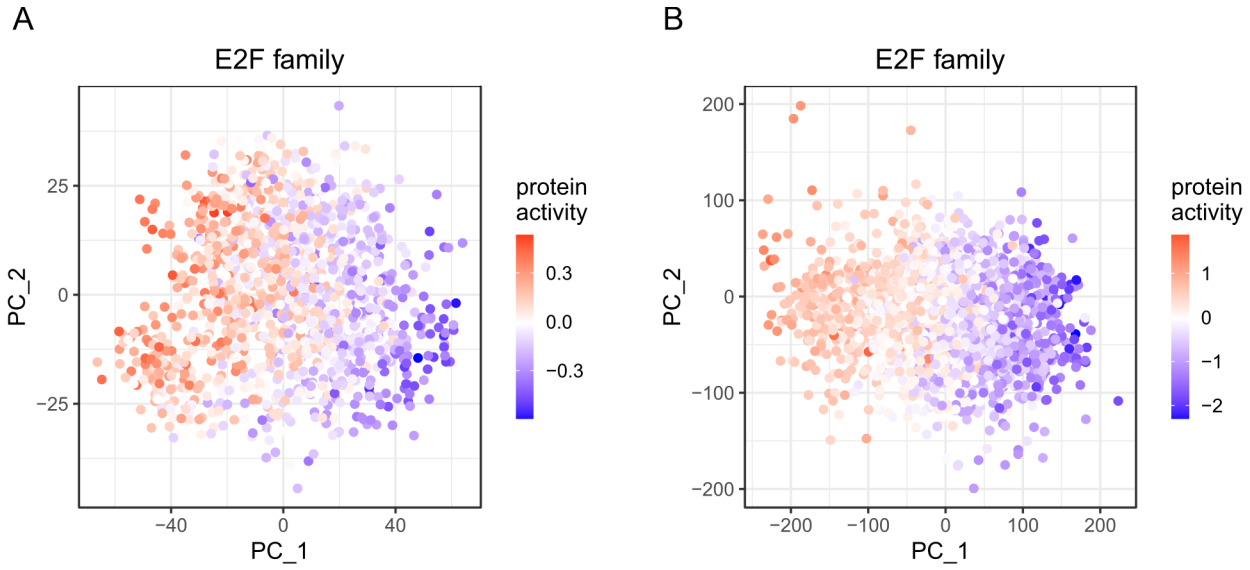


Figure S11. Protein activity of E2F family, a proliferative cell marker in the patient (**A**) and PDX control (**B**) data.

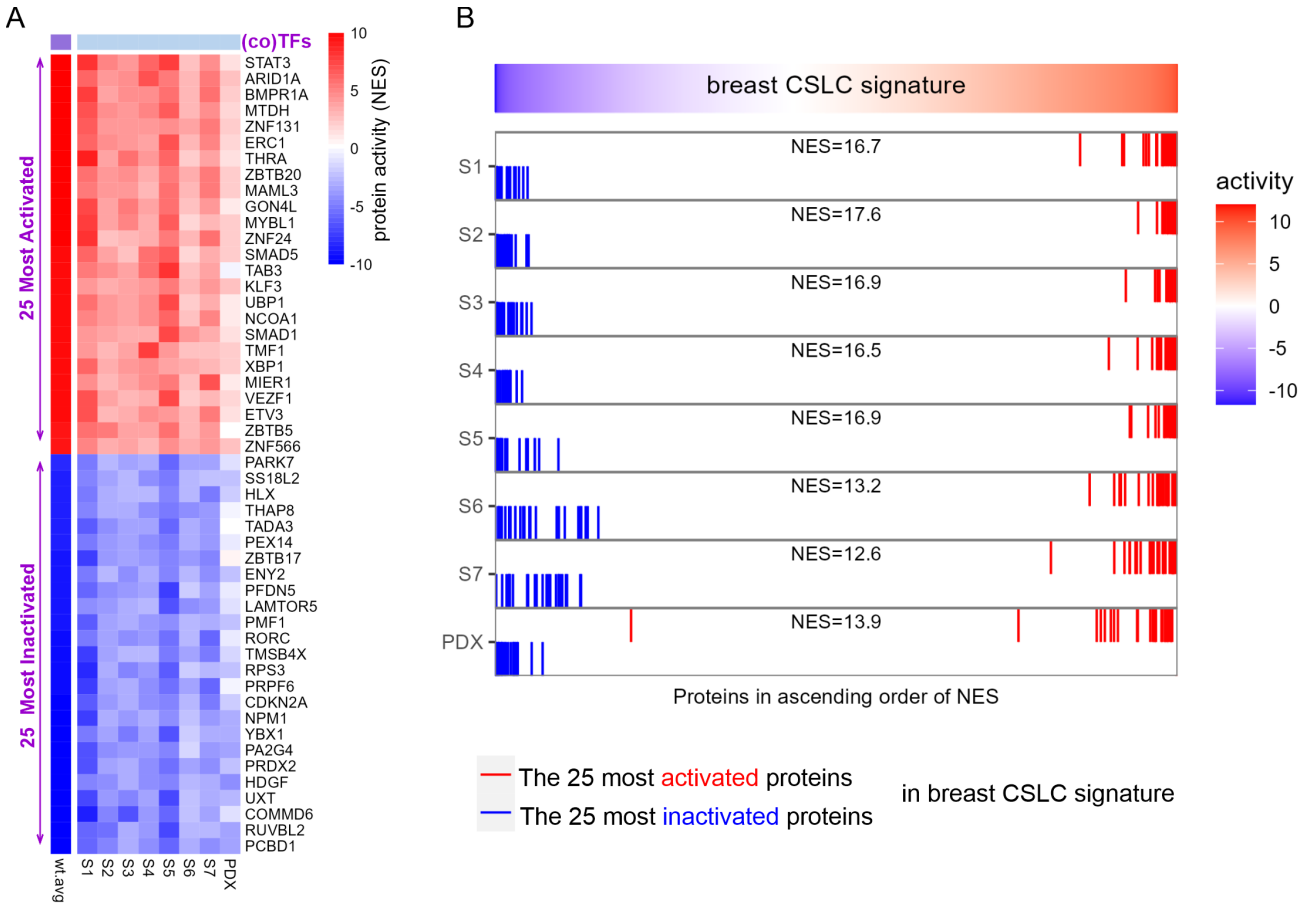


Figure S12. A. A heatmap exhibiting the activity of the 25 most activated and 25 most deactivated TFs/coTFs in the breast CSLC signature and their activities in individual samples (7 patient samples and the PDX vehicle-treated sample). **B.** An enrichment analysis plot of the 25 most activated and 25 most inactivated proteins of the CSLC signature in each individual sample. The consistency of the activity of these 50 proteins is observed in all samples, showing a significant enrichment (p-value< 1.0×10^{-16}).

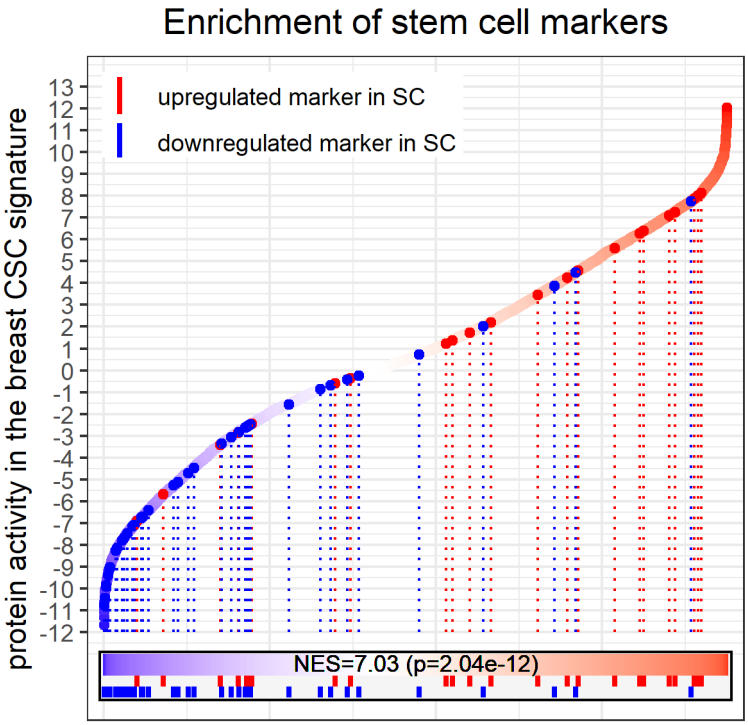


Figure S13. An waterfall plot of sorted protein activities in the breast CSLC signature, identified using both the patient and PDX control data. The NES of genes involved in general stem cell processes (supplementary data) is shown on the CSC signature.

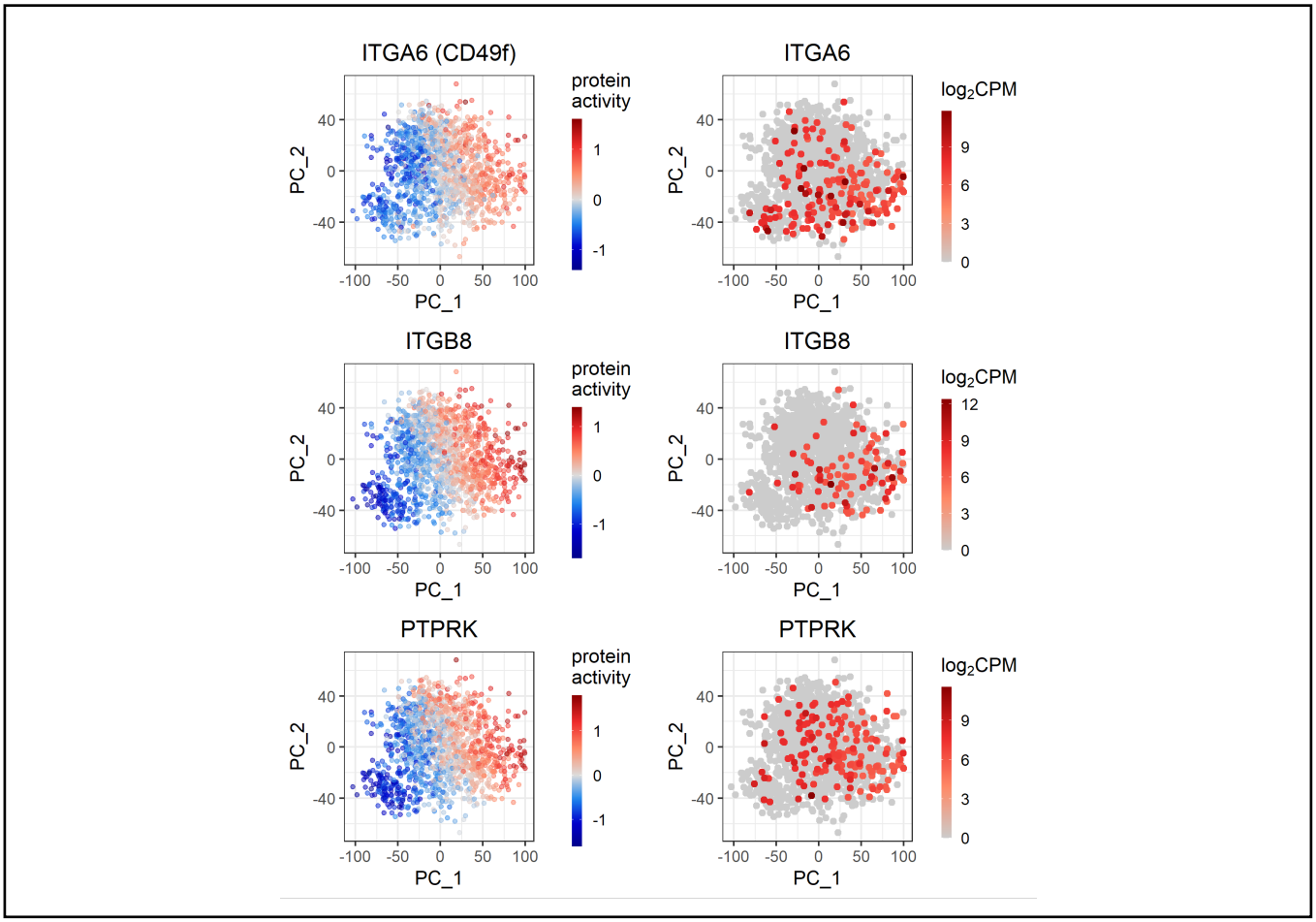


Figure S14. The VIPER-inferred activity (centered) and the gene expression (log_2CPM) of cell surface proteins among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile are shown for visualization purposes.

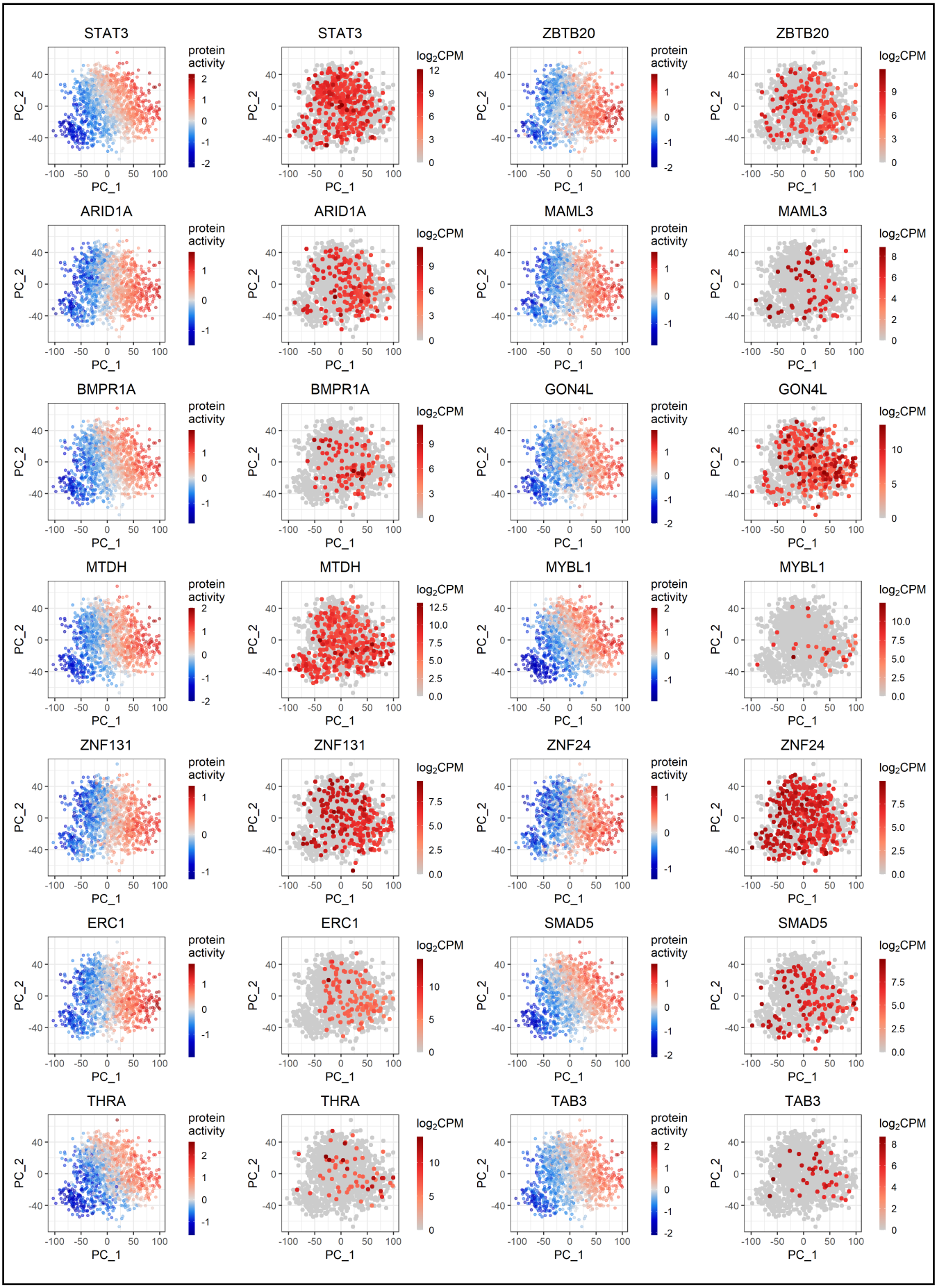


Figure S15-A. The VIPER-inferred activity (centered) and the gene expression (log2CPM) of transcription (co-)factors among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

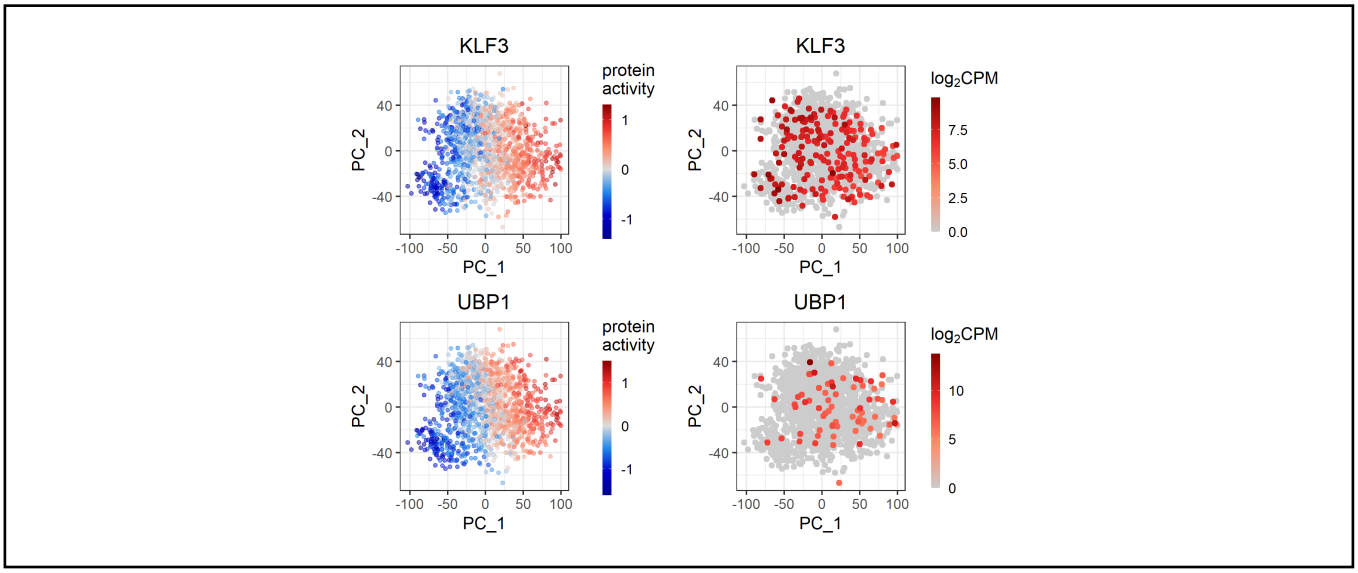


Figure S15-B (continued). The VIPER-inferred activity (centered) and the gene expression (log_2CPM) of transcription (co-)factors among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

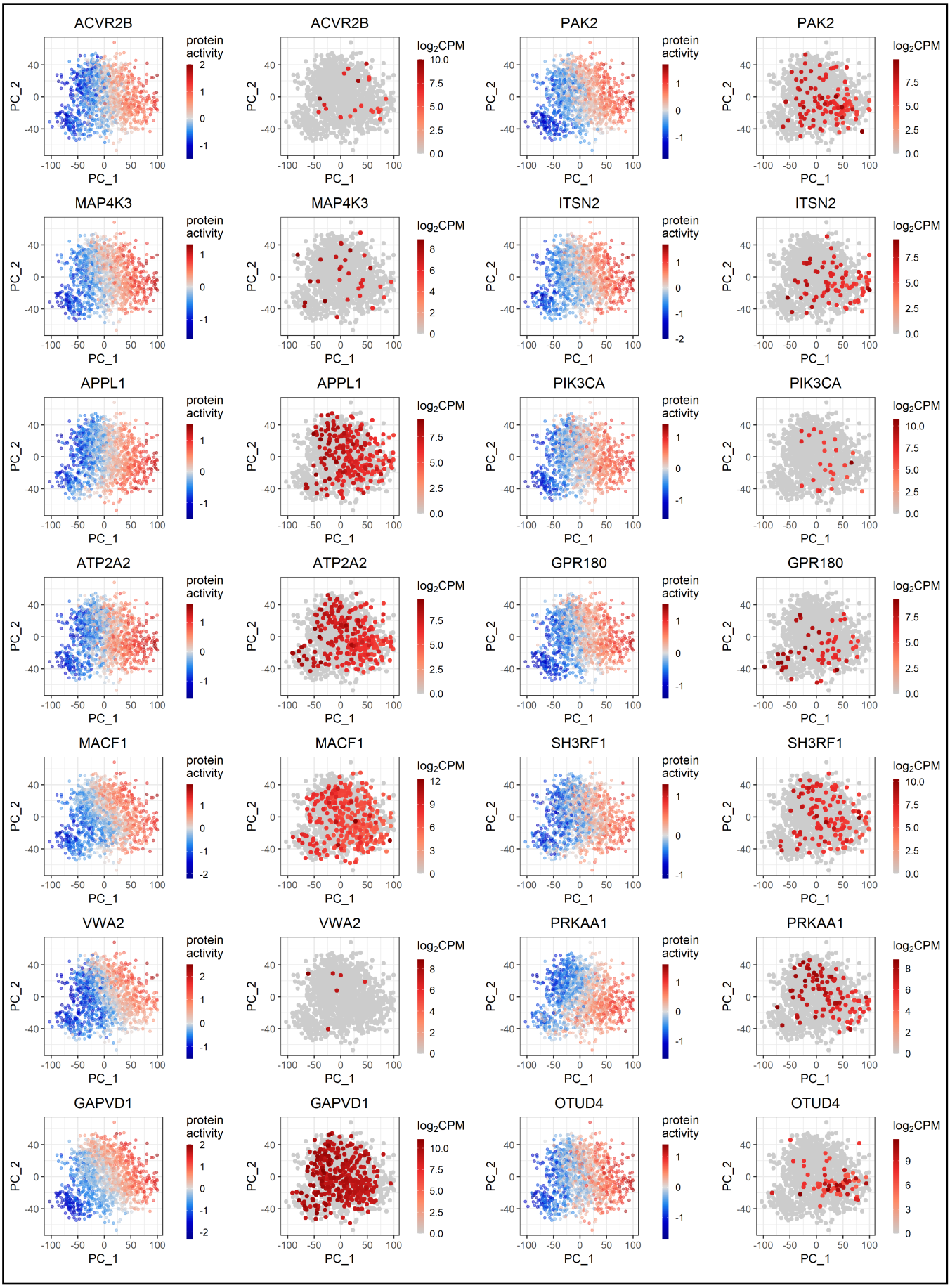


Figure S16-A. The VIPER-inferred activity (centered) and the gene expression ($\log_2$CPM) of signaling molecules among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

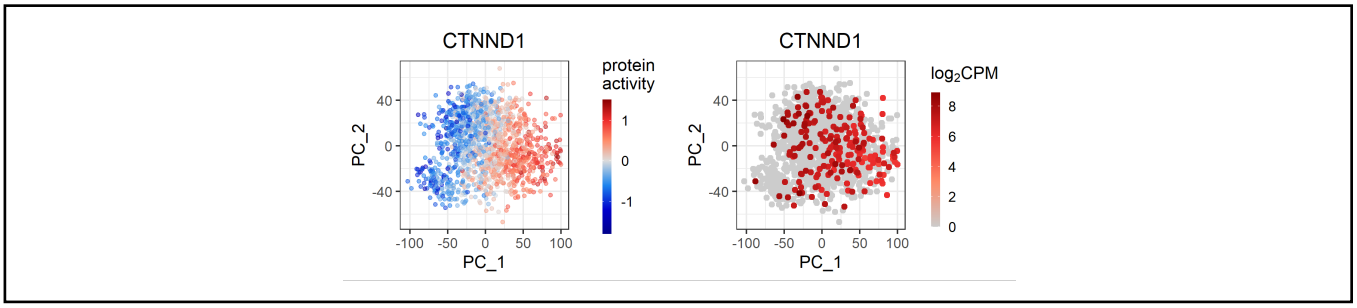


Figure S16-B (continued). The VIPER-inferred activity (centered) and the gene expression (log_2CPM) of signaling molecules among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

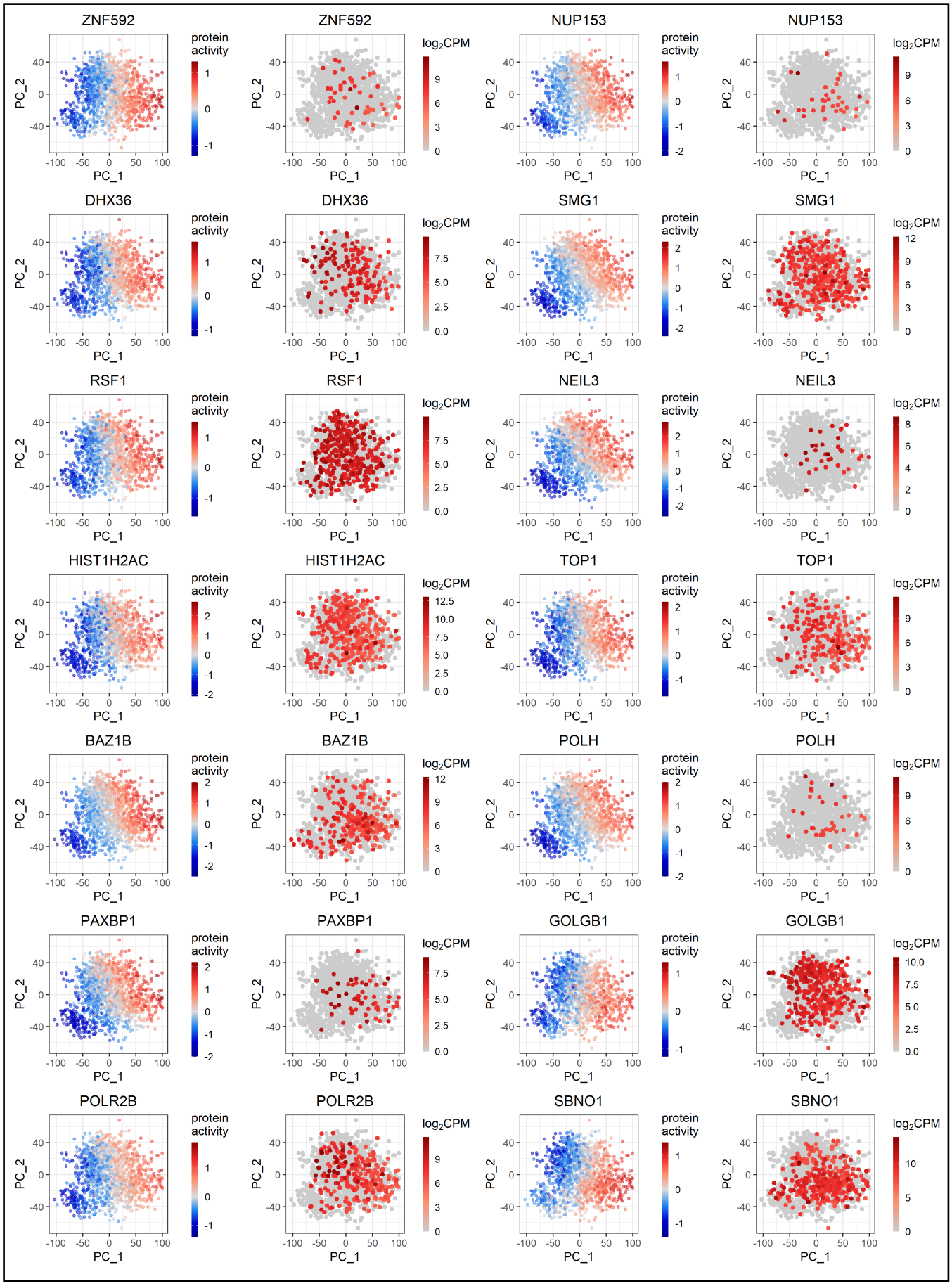


Figure S17-A. The VIPER-inferred activity (centered) and the gene expression (log_2CPM) of other proteins involved in transcriptional programs among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

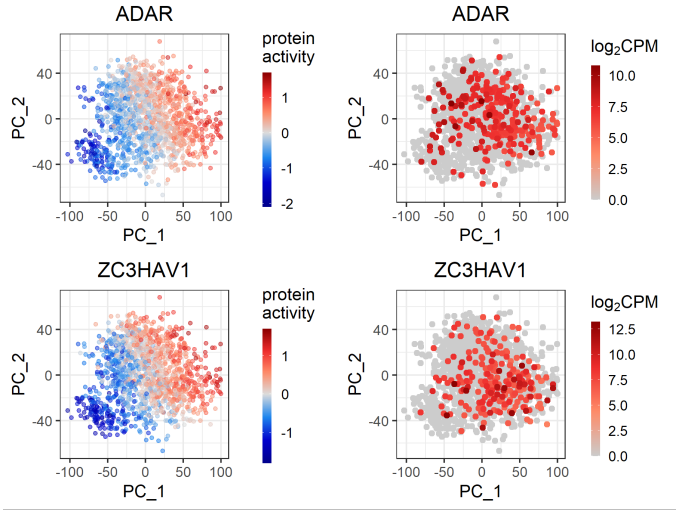


Figure S17-B (continued). The VIPER-inferred activity (centered) and the gene expression (log_2CPM) of other proteins involved in transcriptional programs among the top 50 most activated proteins in the breast CSCs. For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

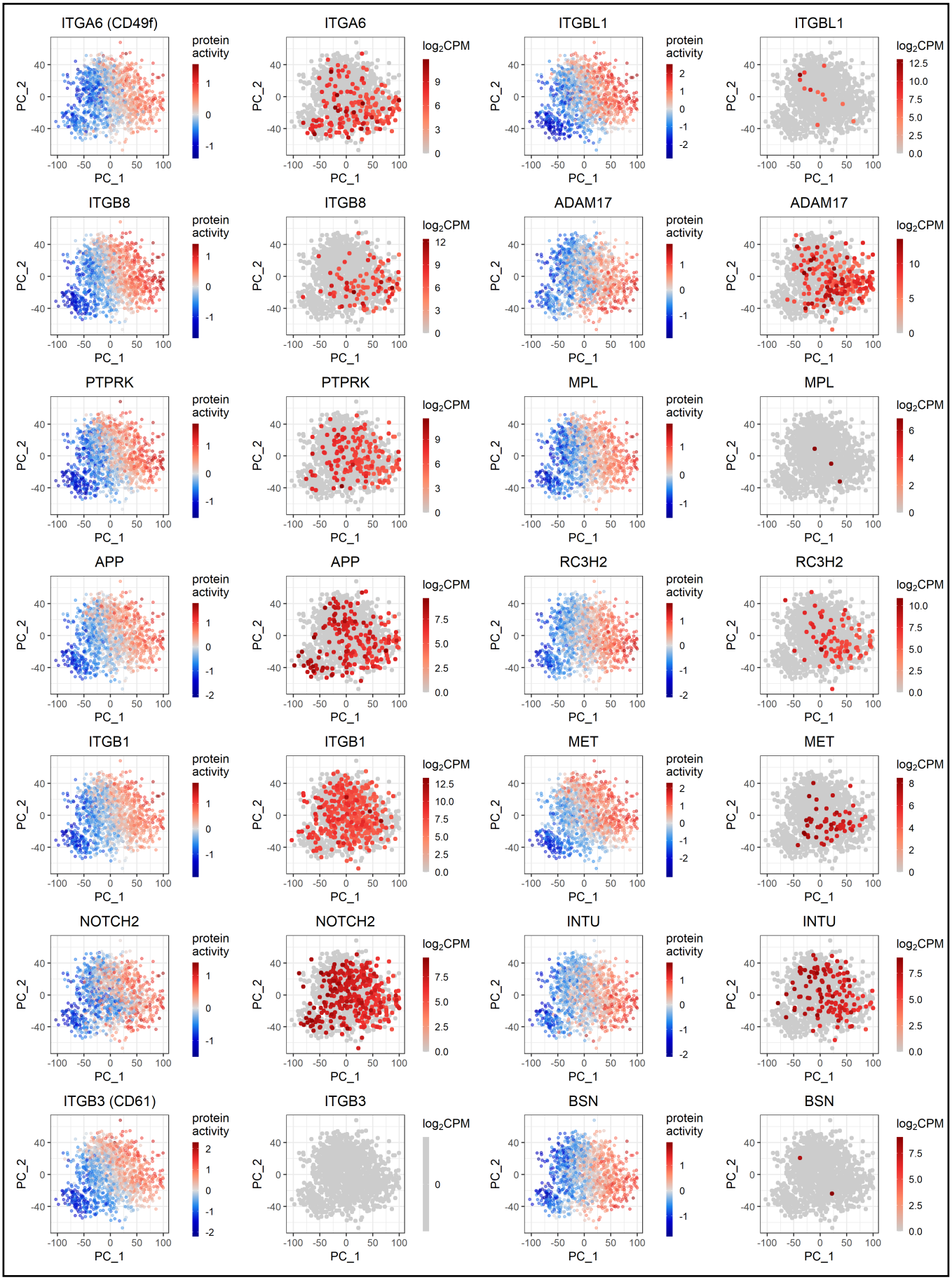


Figure S18-A. The VIPER-inferred activity (centered) and the gene expression ($\log_2$CPM) of cell-surface proteins among significantly activated proteins in the breast CSCs (Bonferroni adjusted *p*-value < 0.001). For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

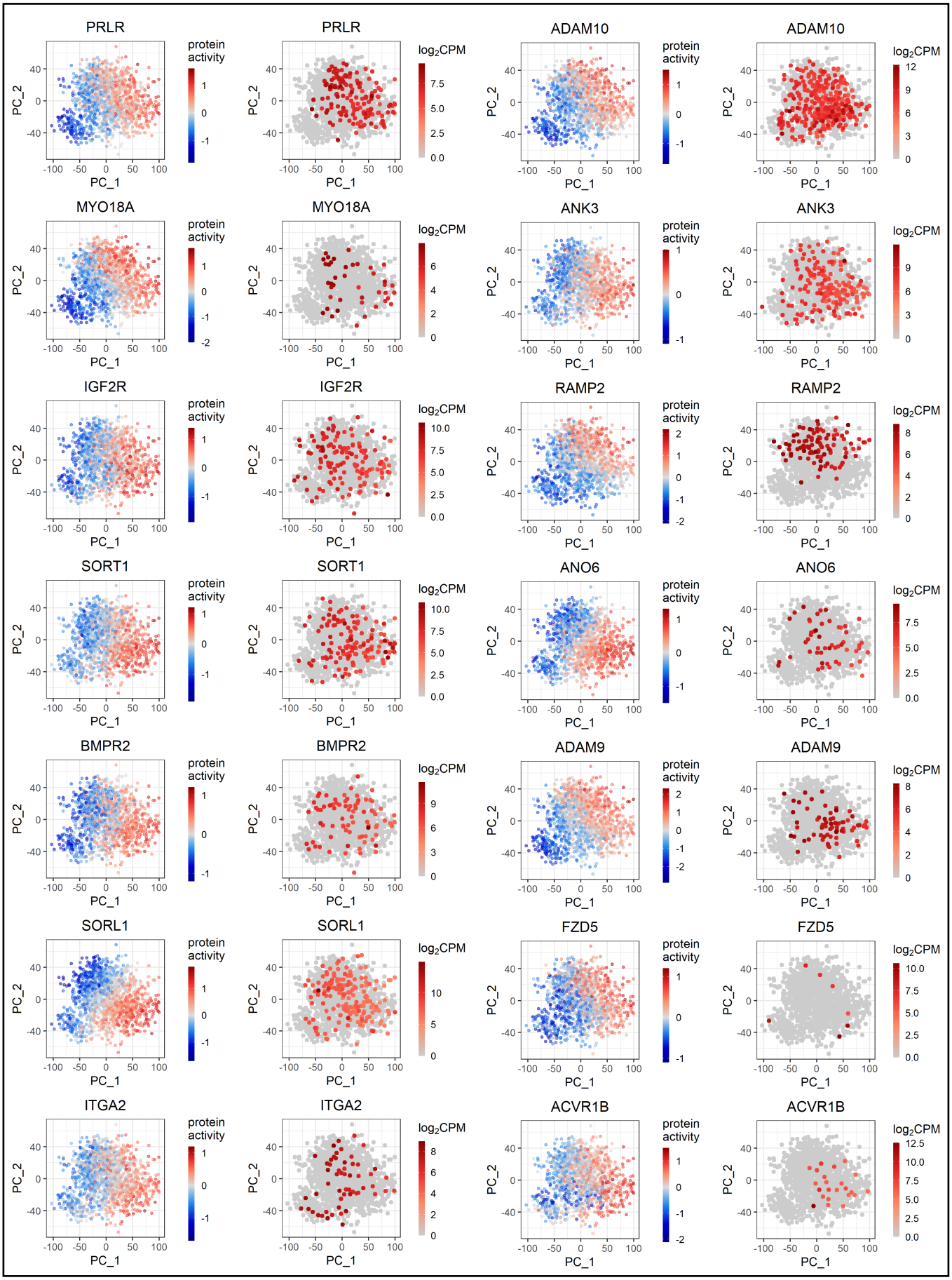


Figure S18-B (continued). The VIPER-inferred activity (centered) and the gene expression ($\log_2$CPM) of cell-surface proteins among significantly activated proteins in the breast CSCs (Bonferroni adjusted *p*-value < 0.001). For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

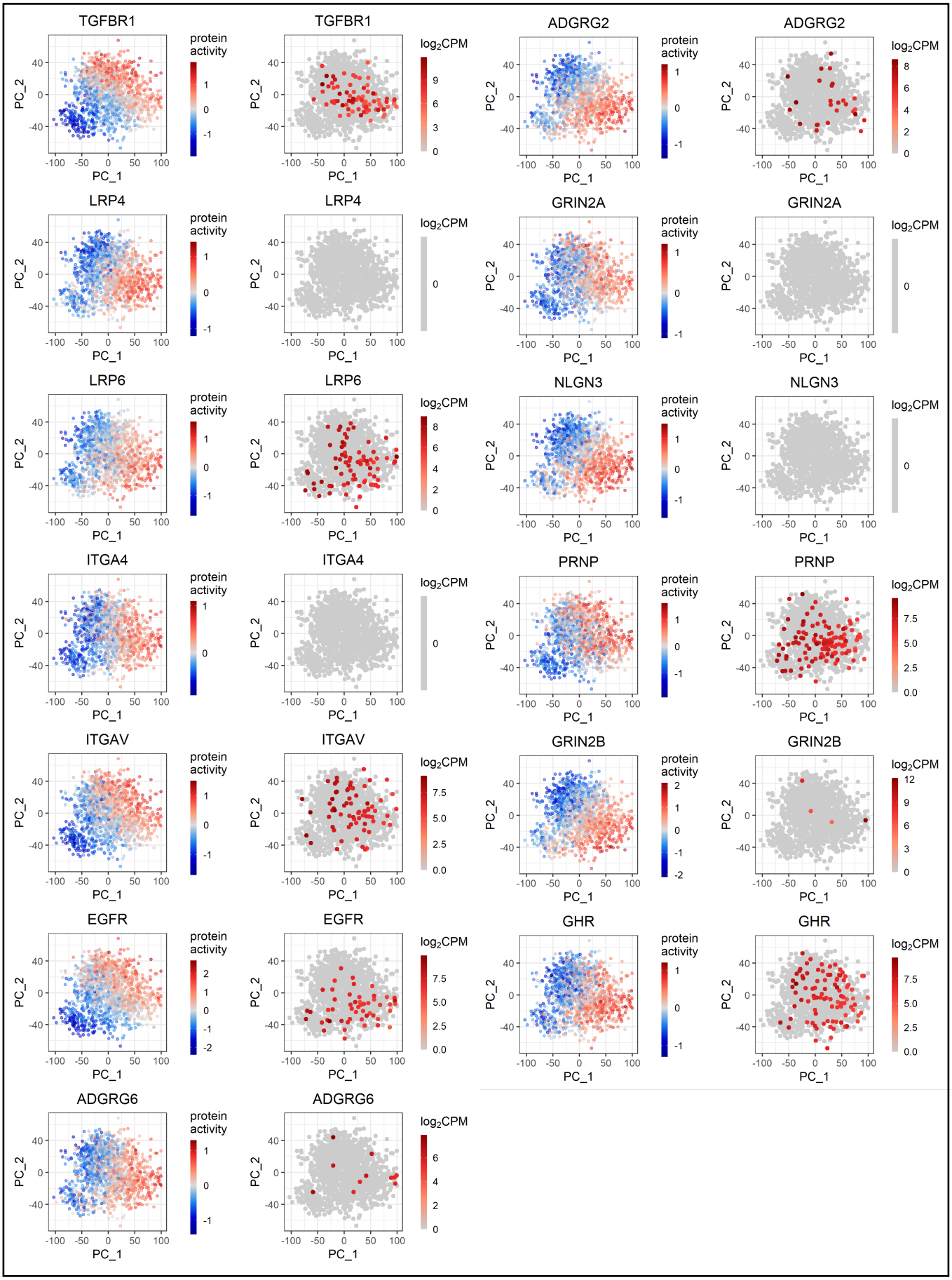


Figure S18-C (continued). The VIPER-inferred activity (centered) and the gene expression ($\log_2$CPM) of cell-surface proteins among significantly activated proteins in the breast CSCs (Bonferroni adjusted *p*-value < 0.001). For gene expression, only cells with expression greater than the upper quantile were shown for visualization purposes.

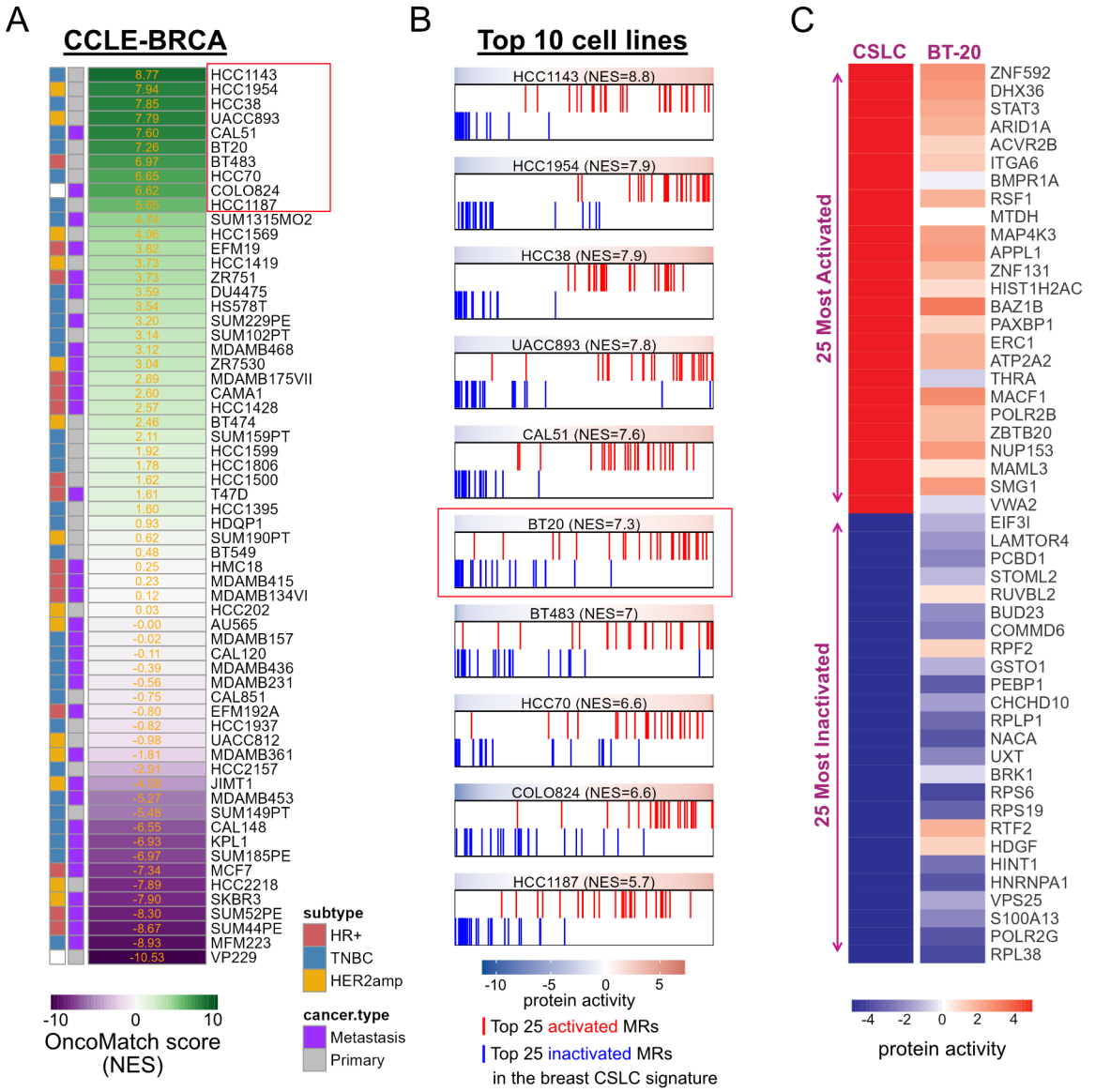


Figure S19. A. A heatmap illustrating the ranked list of breast cancer cell lines in CCLE, in a decreasing order of their OncoMatch score, which indicates the similarity score of the protein activity profiles between a cell line and the CSLC signature. Thus, the greater the OncoMatch score, the more stem-like cell properties. **B.** Enrichment analysis plots of the 25 most activated and the 25 most inactivated proteins in the breast CSLC signature for top 10 breast cancer cell lines by OncoMatch. The plot demonstrated a statistically significant enrichment (p-value$< 1\times10^{-16}$), between BT20 and the CSLC signature. **C.** A heatmap of the activity of the 25 most activated and the 25 most inactivated proteins in the breast CSLC signature and their activities in BT20.

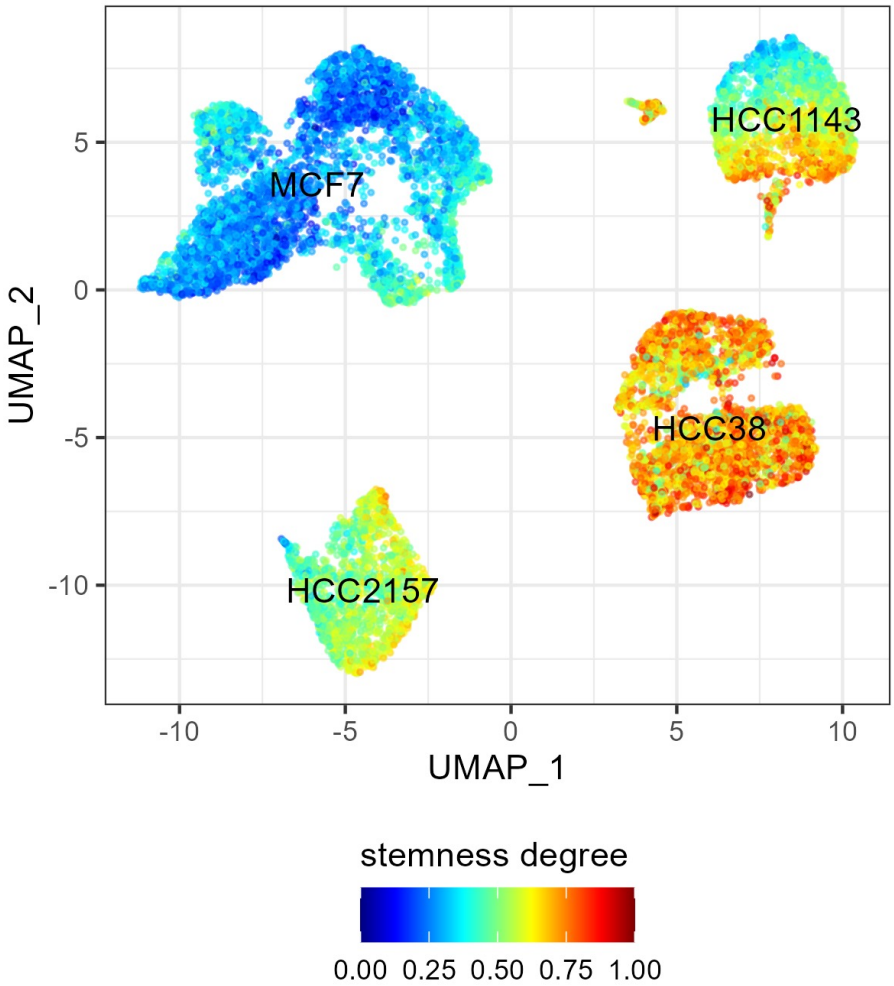


Figure S20. The UMAP projection of protein activity profiles of single cells for four breast cancer cell lines. HCC38 and HCC1143 are chosen for representing CSLC-rich cell lines, while HCC2157 and MCF7 are selected as well-differentiated cell lines (negative controls), based on our OncoMatch prediction as well as literature. The color of cells indicates the stemness degree calculated in the same manner as the previous (i.e. the weighted average of stemness marker activities and the CytoTRACE score). Thus, the greater score, the higher stemness in cells.

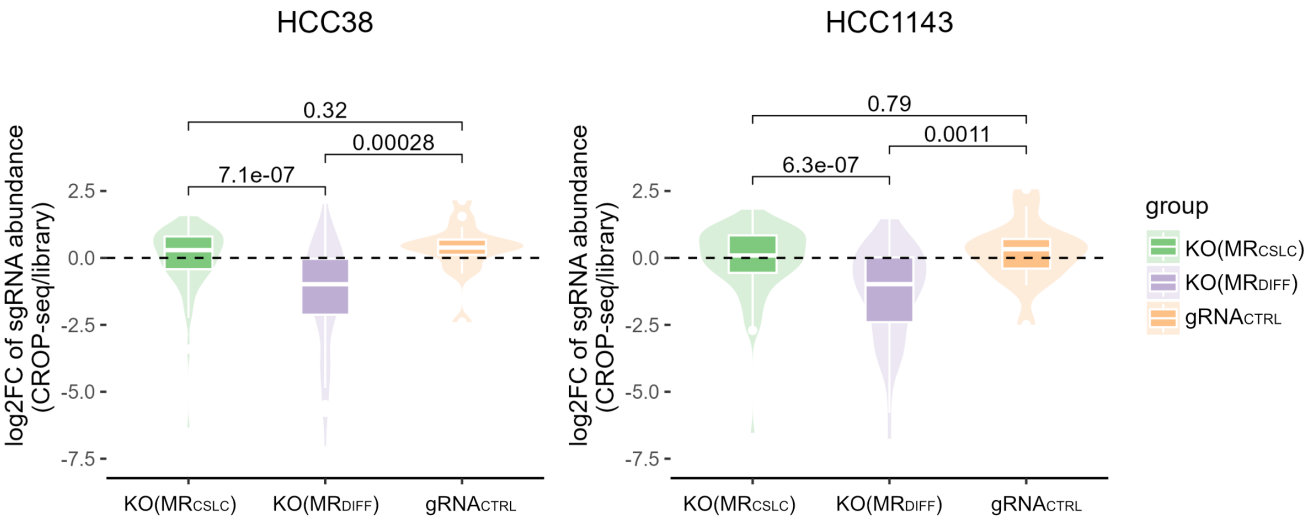


Figure S21. The log2 Fold Change (log2FC) of sgRNA abundance between CROP-seq and CRISPR library. sgRNA counts were normalized by dividing them by total sgRNA counts in each CROP-seq and CRISPR library. Unlike MR_{CSLC}, log2FC after MR_{DIFF} KOs is significantly diminished, implying many of MR_{DIFF} are responsible for cell fitness/proliferation.

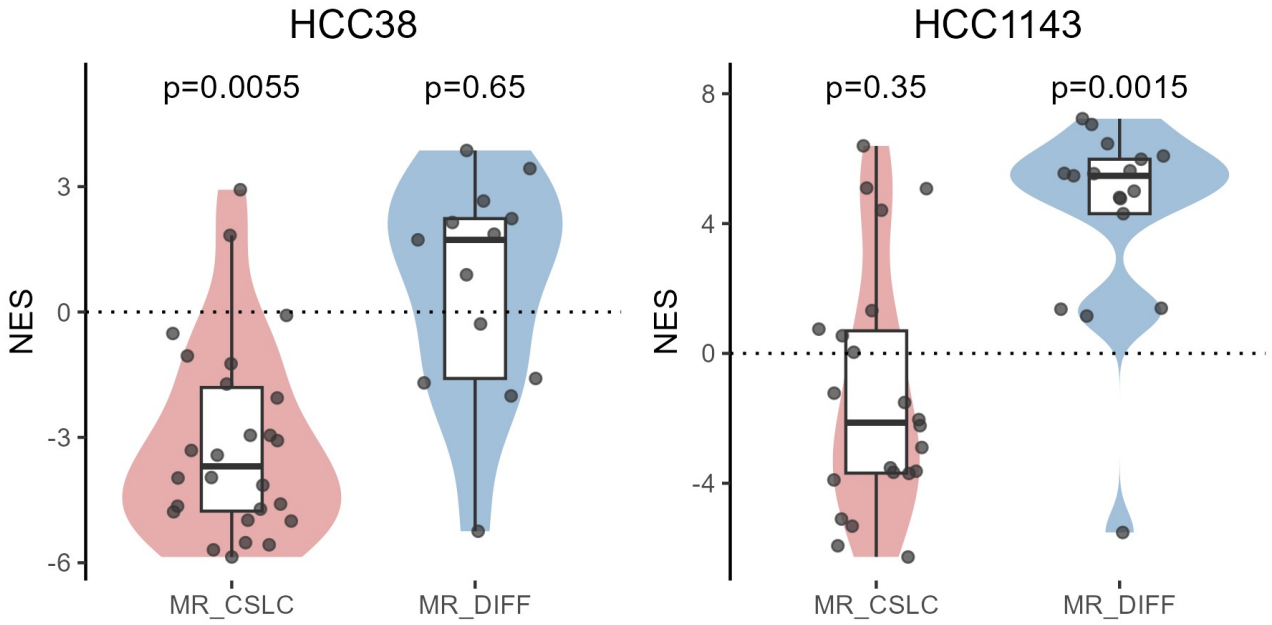
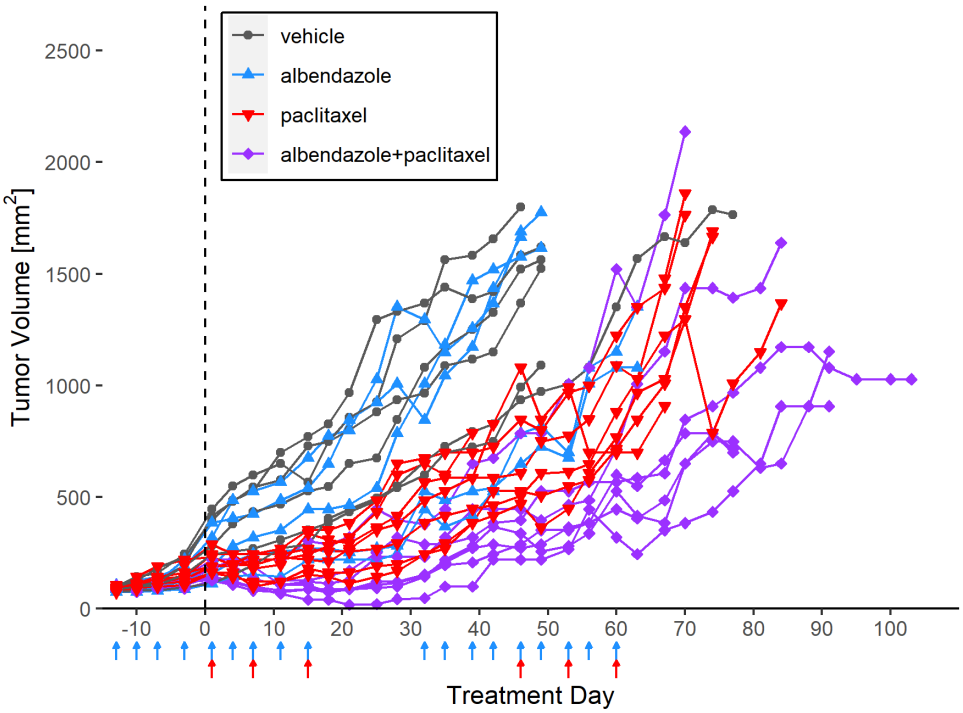


Figure S22. The enrichment of the gene set of stem cell process (supplementary table 2) between two groups (MR_{CSLC} and MR_{DIFF}). Similar to Fig.5A, the enrichment score (NES) of stem cell process genes were significantly diminished (p=5.5×10^{-3}) after MR_{CSLC} KO for HCC38, while NES was not significantly increased after MR_{DIFF} KO for the same cell line. On the contrary, for HCC1143, the increase of NES was more striking (p=1.5×10^{-3}) after MR_{DIFF} KO, while the NES change was not significant for the group of MR_{CSLC} KO.

A



B

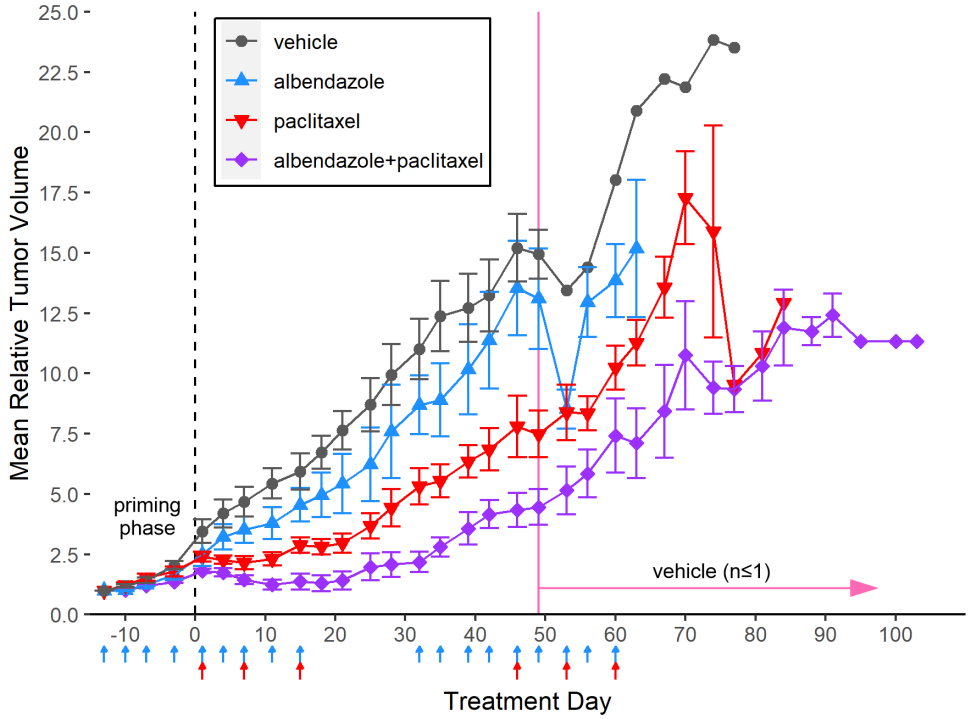


Figure S23. A. Spider plot of tumor volume measurements over time for individual mice. During a priming phase, mice were treated with albendazole 3 times weekly for two weeks before the start date of the combined drug therapy with paclitaxel, in order to sensitize the tumor cells. Mice with albendazole monotherapy were treated for the same amount of time as in the combination therapy. **B.** Mean relative tumor volumes over time. Tumor volumes are normalized by their volumes at Day-13. The error bar indicates one standard error of the mean. 5 out of 6 control mice (i.e. vehicle treatment) were euthanized before reaching Day 50.

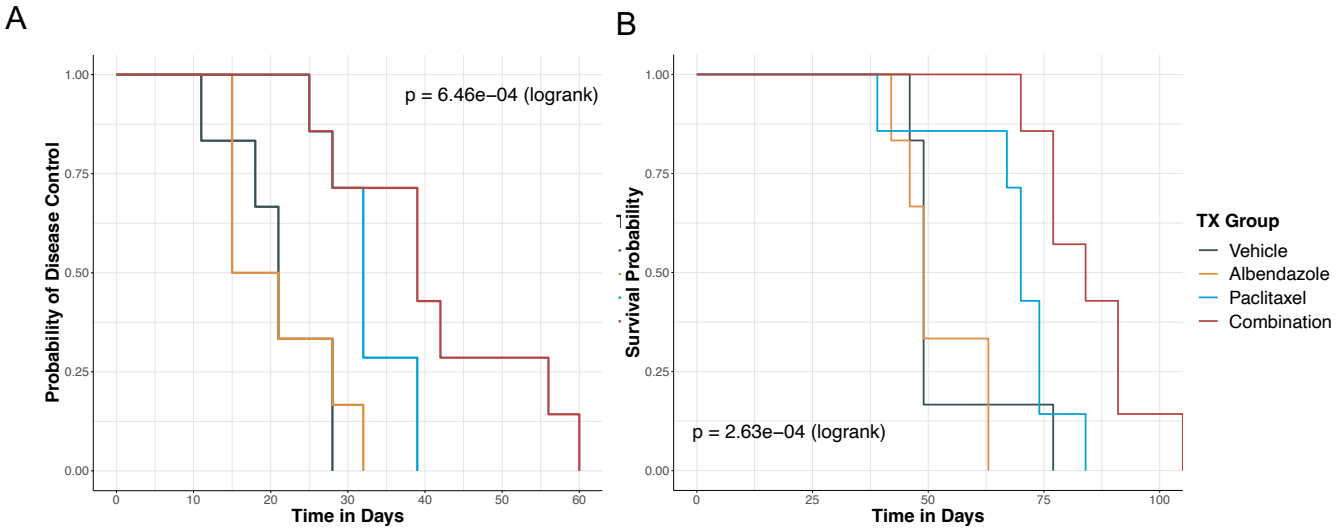


Figure S24. Kaplan-Meier analysis of the preclinical measurements for Disease Control (**A**) and survival (**B**) following treatment with albendazole and paclitaxel monotherapy vs. the combination.