



Identification of phenotype-specific networks from paired gene expression–cell shape imaging data

Charlie George Barker, Eirini Petsalaki, Girolamo Giudice, et al.

Genome Res. 2022 32: 750-765 originally published online February 23, 2022

Access the most recent version at doi:[10.1101/gr.276059.121](https://doi.org/10.1101/gr.276059.121)

References This article cites 116 articles, 20 of which can be accessed free at:
<http://genome.cshlp.org/content/32/4/750.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Method

Identification of phenotype-specific networks from paired gene expression–cell shape imaging data

Charlie George Barker,¹ Eirini Petsalaki,¹ Girolamo Giudice,¹ Julia Sero,²
Emmanuel Nsa Ekpenyong,¹ Chris Bakal,³ and Evangelia Petsalaki¹

¹European Molecular Biology Laboratory-European Bioinformatics Institute, Hinxton CB10 1SD, United Kingdom; ²University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom; ³Institute of Cancer Research, London SW3 6JB, United Kingdom

The morphology of breast cancer cells is often used as an indicator of tumor severity and prognosis. Additionally, morphology can be used to identify more fine-grained, molecular developments within a cancer cell, such as transcriptomic changes and signaling pathway activity. Delineating the interface between morphology and signaling is important to understand the mechanical cues that a cell processes in order to undergo epithelial-to-mesenchymal transition and consequently metastasize. However, the exact regulatory systems that define these changes remain poorly characterized. In this study, we used a network-systems approach to integrate imaging data and RNA-seq expression data. Our workflow allowed the discovery of unbiased and context-specific gene expression signatures and cell signaling subnetworks relevant to the regulation of cell shape, rather than focusing on the identification of previously known, but not always representative, pathways. By constructing a cell-shape signaling network from shape-correlated gene expression modules and their upstream regulators, we found central roles for developmental pathways such as WNT and Notch, as well as evidence for the fine control of NF- κ B signaling by numerous kinase and transcriptional regulators. Further analysis of our network implicates a gene expression module enriched in the RAPI signaling pathway as a mediator between the sensing of mechanical stimuli and regulation of NF- κ B activity, with specific relevance to cell shape in breast cancer.

[Supplemental material is available for this article.]

The study of cancer has long been associated with changes in cell shape as morphology can be a reliable way to subtype cancer and predict patient prognosis (Wu et al. 2020). Recent research has implicated cellular morphology in more than just a prognostic role in cancer, with shape affecting tumor progression through the modulation of migration, invasion, and overall tissue structure (Krakhmal et al. 2015; Baskaran et al. 2020). The unique mechanical properties of the tumor tissue (primarily driven by changes in cell shape and the extracellular matrix) are hypothesized to contribute to the “stem cell niche” of cancer cells that enables them to self-renew as they do in embryonic development (Cooper and Giancotti 2019). Cell morphology and tumor organization have been found to be a factor in modulating the intracellular signaling state through pathways able to integrate mechanical stimuli from the extracellular environment (Orsulic et al. 1999; Miralles et al. 2003; Zheng et al. 2009; Olson and Nordheim 2010). The discovery of mechanosensitive pathways in various tissues has revealed a complex interplay between cell morphology and signaling (Kumar et al. 2016). Further studies have revealed that cell morphology can also be a predictor of tumorigenic and metastatic potential as certain nuclear and cytoplasmic features enhance cell motility and spread to secondary sites (Wu et al. 2020), aided by the epithelial-to-mesenchymal transition (EMT). This process is the conversion of epithelial cells to a mesenchymal phenotype, which contributes to metastasis in cancer and to worse prognosis in patients (Roche 2018).

Breast cancer is the most common cancer among women and, in most cases, is treatable, with a survival rate of 99% among pa-

tients with a locally contained tumor. However, among those patients presenting with a metastatic tumor, this rate drops to 27% (Siegel et al. 2019). During the development of breast cancer tumors, cells undergo progressive transcriptional and morphological changes that can ultimately lead toward EMT and subsequent metastasis (Lee et al. 2015b; Feng et al. 2018; Wu et al. 2020). Breast cancer subtypes of distinct shapes show differing capacities to undergo this transition. For example, long and protrusive basal breast cancer cell lines are more susceptible to EMT (Fedele et al. 2017) with fewer cell-to-cell contacts (Dai et al. 2015). Luminal tumor subtypes, on the other hand, are associated with good to intermediate outcomes for patients (Dai et al. 2015) and have a clear epithelial (or “cobblestone”) morphology with increased cell–cell contacts (Neve et al. 2006). It is evident that cell morphology plays significant roles in breast cancer, and a deeper understanding of the underlying mechanisms may offer possibilities for using these morphology-determinant pathways as potential therapeutic targets and predictors of prognosis.

Signaling and transcriptomic programs are known to be modulated by external physical cues in the contexts of embryonic development (Wozniak and Chen 2009), stem-cell maintenance (Bergert et al. 2021; De Belly et al. 2021), and angiogenesis (Chatterjee 2018). Numerous studies have flagged NF- κ B as a focal point for mechanotransductive pathways in various contexts (Cowell et al. 2009; Shrum et al. 2009; Tong and Tergaonkar 2014; Ishihara et al. 2019), but gaps in our knowledge remain as to how these pathways may interact and affect breast cancer development. Sero et al. (2015) studied the link between cell shape in breast cancer and NF- κ B activation by combining high-

Corresponding author: petsalaki@ebi.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.276059.121>. Freely available online through the *Genome Research* Open Access option.

© 2022 Barker et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

throughput image analysis of breast cancer cell lines with network modeling. They found a relationship between cell shape, mechanical stimuli, and cellular responses to NF- κ B and hypothesized that this generated a negative feedback loop, in which a mesenchymal-related morphology enables a cell to become more susceptible to EMT, thus reinforcing their metastatic fate. This analysis was extended by Sailem and Bakal (2017), who combined cell-shape features collected from image analysis with microarray expression data for breast cancer cell lines to create a shape-gene interaction network that better delineated the nature of NF- κ B regulation by cell shape in breast cancer. This approach was limited, as it only correlates single genes with cell shapes, thus relying on the assumption that a gene's expression is always a useful indicator of its activity (Vogel and Marcotte 2012). Furthermore, the investigators rely on a list of preselected transcription factors (TFs) of interest, and as such, the approach is not completely data-driven and hypothesis-free. Given our knowledge of the multitude of complex interacting signaling pathways in development and other contexts, it is safe to assume that there are many more players in the regulation of cancer cell morphology that have yet to be delineated (Bougault et al. 2012; Rolfe et al. 2014; Robertson et al. 2015; Horton et al. 2016). Furthermore, how exactly extracellular mechanical cues are "sensed" by the cell and passed on to NF- κ B in breast cancer is not clearly understood. From this it is clear that an unbiased approach is needed to identify novel roles for proteins in the interaction between cell shape and signaling.

Here we have developed a powerful network-based approach to bridge the gap between widely available and cheap expression data, signaling events, and large-scale biological phenotypes such as cell shape (Fig. 1A). Our study aims to identify a data-derived cellular signaling network, specific to the regulation of cell shape beyond NF- κ B, by considering functional coexpression modules and cell signaling processes rather than individual genes.

Results

Identification of gene coexpression modules correlated with cell-shape features

We first sought to identify gene expression modules (GEMs) that are relevant to the regulation of cell shape. To this end, we used weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath 2008) on bulk RNA-seq expression data from 13 breast cancer cell lines and one nontumorigenic epithelial breast cell line to identify gene coexpression modules correlated with 10 specific cell-shape variables (Methods; Sailem and Bakal 2017). These described the size, perimeter, and texture of the cell and the nucleus ($n=75,653$). Of 102 GEMs (Supplemental Fig. S1A), 34 were significantly correlated ($P<0.05$; Student's *t*-test, Pearson's correlation) (Supplemental Table S1; Supplemental Fig. S1B,C) with one of eight cell-shape features (Fig. 1B). A full list of the genes within the identified modules is presented in Supplemental Table S2.

We used Enrichr and their suite of gene set libraries (Kuleshov et al. 2016) to functionally annotate and label some of the modules using enrichment of genes contained within them. We found that the "RAP1 signaling" module is also enriched for terms such as VEGF signaling and hemostasis, whereas the "insulin signaling" module is also enriched for cell-cell communication, and the "ECM organization" module is also enriched in terms such as axon guidance and EPH-ephrin signaling (Supplemental Table S3). Modules that are most correlated with all features are the

"ARNTKO" module, "ARRDC3-AS1" module, and the "ECM organization" module (Supplemental Fig. S1B). Modules that could not be annotated with informative terms were designated as module "non annotated" (NA) 1, 2, 3, etc.

TF analysis of cell-shape gene coexpression modules reveals the signaling pathways that regulate them

To link these expression modules to the intracellular signaling network, we considered both the regulation of modules as transcriptional units as well as the signaling pathways that significantly regulate the identified regulons. Specifically, we first found 17 TF regulons, as defined in the database TRRUST v2 (Han et al. 2018), to be significantly enriched ($P<0.1$; Fisher's exact test) in our modules (Supplemental Table S4). We therefore consider these TFs as potentially relevant for the regulation of cell-shape features and their activity levels as a read-out of cell signaling activity in these cells. These TFs include the EMT antagonist FOXA1 (Song et al. 2010), as well as HOXB7 (Wu et al. 2006) and ZFP36 (Van Tubergen et al. 2013).

To extend this further, we sought to investigate the pathways responsible for regulating the identified TFs and, by extension, the GEMs. For this analysis, we also include ENCODE and ChEA Consensus TFs from ChIP-X (Lachmann et al. 2010), DNA binding preferences from JASPAR (Wasserman and Sandelin 2004; Stormo 2013), TF protein-protein interactions, and TFs from ENCODE ChIP-seq (Euskirchen et al. 2007) to get a more comprehensive picture of the pathways involved in regulation of cell morphology. Using the identified TFs (Supplemental Table S5) we then used Enrichr (Kuleshov et al. 2016) to perform a Reactome signaling pathway (Jassal et al. 2020) enrichment analysis. Results from this analysis showed that six modules shared pathways associated with downstream signaling and regulation of NOTCH (Fig. 1C). To ensure that our approach is not biased to any particular pathway, we repeated our approach on 1000 re-sampled GEMs, and created pathway-specific null distributions for each identified pathway. All pathways we identified from morphology-correlated modules had significantly lower *P*-values than randomized modules (FDR adjusted $P<0.05$). The only exceptions were one association with "signaling by NOTCH" and modules associated with "signal transduction," a spurious pathway containing the complete intracellular signaling system (Supplemental Table S6).

Clustering based on morphology reveals distinctive cell line shapes

To understand key differences in expression patterns and gene regulation between morphologically distinct breast cancer cell lines, we clustered them based on 10 morphological features including area, ruffiness, protrusion area, and neighbor frequency and performed differential expression analysis between the identified clusters (Fig. 2A; Supplemental Fig. S2A). Cluster A is more heterogeneous in its morphology, containing the nontumorigenic mammary epithelial cell line MCF-10A as well as cell lines from both luminal and basal breast cancer subtypes. Clusters B and C are more distinctly shaped, roughly composed of luminal and basal cell lines, respectively, except for HCC1954, which was clustered morphologically with luminal subtypes, while being characterized as basal. The basal-like cluster is most morphologically distinct from cluster A but also differs from the luminal-like cluster in that it has a lower nuclear/cytoplasmic area (0.133 ± 0.05 [mean \pm SD]), higher ruffiness (0.235 ± 0.12), and lower neighbor fraction

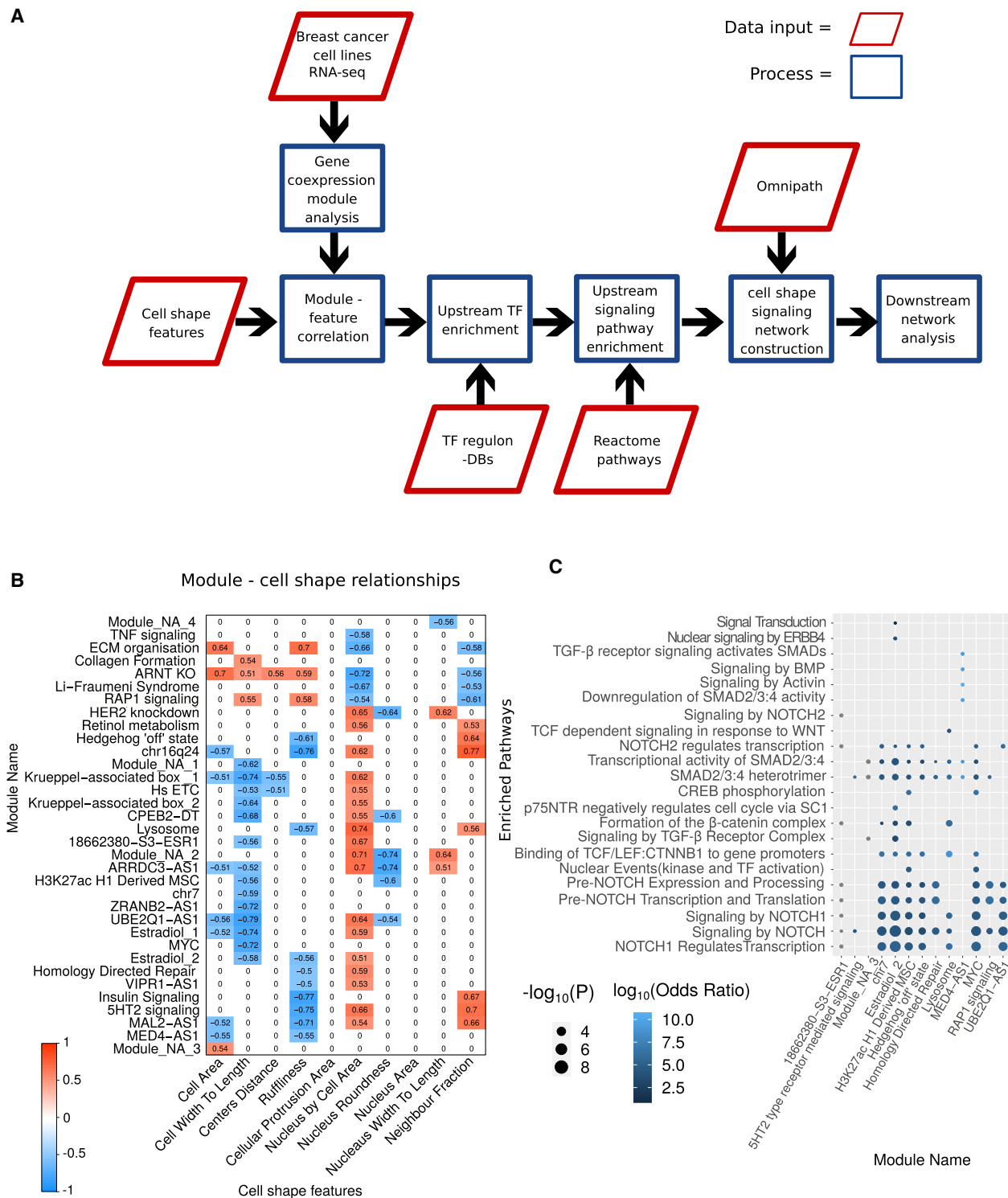


Figure 1. Overview of workflow and resultant gene expression modules and pathways. (A) Schematic illustrating the steps involved in phenotype-specific network construction. Gene expression modules are identified by integrating cell-shape variables (derived from imaging data) with RNA-seq data from breast cancer cell lines. These gene expression modules are correlated with specific cell-shape features to find morphologically relevant modules. Next, transcription factors (TFs) are identified whose targets significantly overlap with the contents of the expression modules. These TFs are used to identify pathways regulating the gene expression modules, which are then integrated to form a contiguous network using PCSF. (B) Heat map of significantly correlated gene expression module eigengenes with cell-shape features. Non significant interactions were set to zero for clarity. (C) Dot plot illustrating the enrichment of pathways among TFs found to regulate gene expression modules. The x-axis shows the module names (as defined by Supplemental Table S3), and the y-axis shows the signaling pathways found to be significantly ($P < 0.01$) enriched in the TFs that regulate the given module (as defined by Supplemental Table S5). The y-axis is arranged such that the terms with the highest combined odds ratio are at the bottom. Size of the dot represents the $-\log_{10}(P)$, and the color indicates a \log_{10} transformation of the odds ratio.

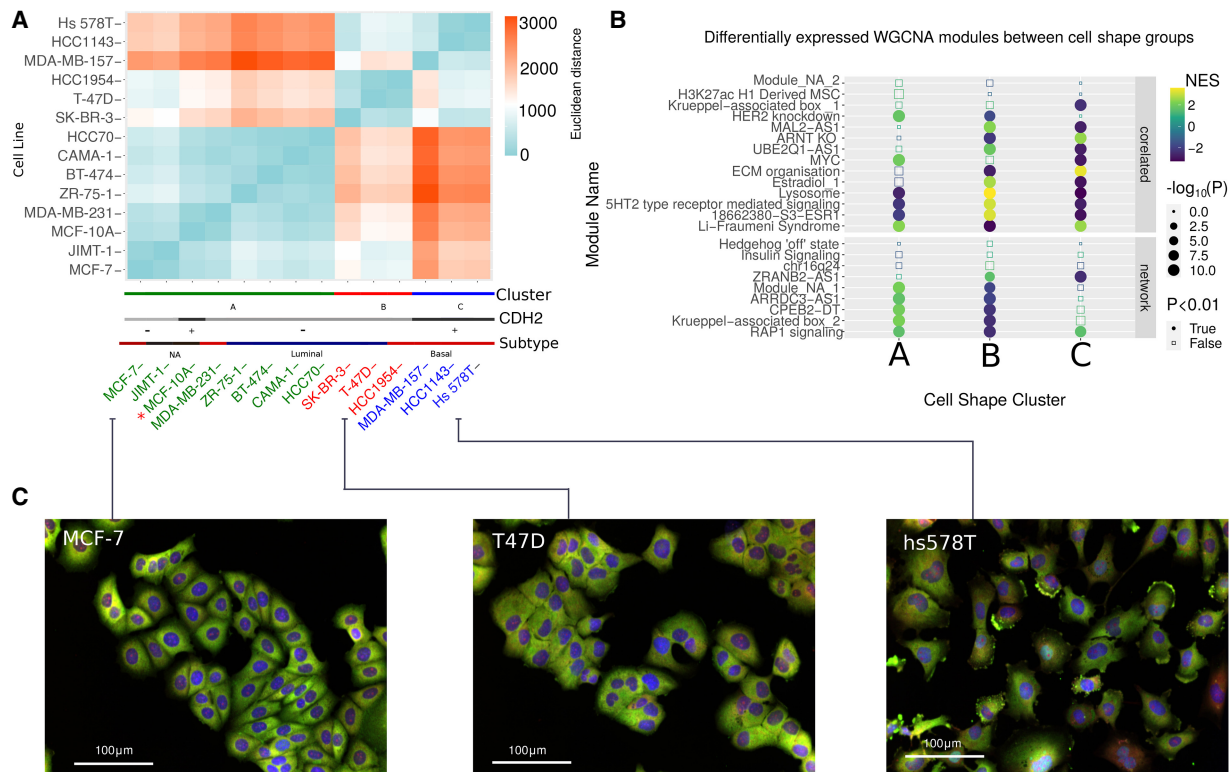


Figure 2. Clustering breast cancer cell lines into groups of similar morphology. (A) Heat map of Euclidean distance between cell lines for shape features to illustrate clusters arising from *k*-means method. The colored lines on the *bottom* show the assigned cluster and the cadherin expression and assigned canonical cancer subtype. (B) Dot plot showing the enrichment of gene expression modules in the different cell line clusters. Along the y-axis are the names of the clusters, faceted by whether they are included in the PCSF-derived regulatory network on the *bottom* and whether they are correlated with cell-shape variables, but not included in the network on the *top*. The x-axis shows the cell-shape clusters, with letters corresponding to the groups in A. (A) A heterogeneous mix of breast cancer subtypes, (B) luminal-like cell lines, (C) basal-like cell lines. Dots are colored based on the normalized enrichment value, with down-regulated modules in blue and the up-regulated modules in yellow. Size corresponds to significance [$-\log_{10}(P)$] with the shape illustrating which changes are significant (adjusted $P < 0.01$, Benjamini–Hochberg). (C) Images (see Methods) showing morphology of representative cell lines from each respective cluster. Colors indicate labeling with DAPI (blue), anti-p65 (green), and DHE (red).

(0.258 ± 0.22). The luminal-like cluster had a higher nuclear/cytoplasmic area (0.186 ± 0.1 ; $P < 0.001$), lower ruffiness (0.213 ± 0.14 ; $P < 0.001$), and a higher neighbor fraction (0.338 ± 0.26 ; $P < 0.001$, one-way ANOVA; Tukey HS, $n = 75,653$). The neighbor fraction feature corresponds to the fraction of the cell membrane that is in contact with neighboring cells. The lower number of cell–cell contacts in basal-like breast cancer cell lines is indicative of more mesenchymal features associated with worse prognosis owing to metastasis. Increased cell–cell contacts in both the luminal-like cluster and the more heterogeneous cluster A correspond to “cobblestone” epithelial morphology. We found that these groups are closely aligned with the expression of the cell adhesion protein cadherin 2 (CDH2, also known as N-cadherin) (Fig. 2A), the expression for which is closely associated with a migratory and metastatic phenotype (Shih and Yamada 2012). Representative images of the morphologically clustered cell lines are shown along with the clustering heat map in Figure 2A (complete data set of images provided online; <https://datadryad.org/stash/dataset/doi:10.5061/dryad.tc5g4>).

Using the identified groups of cell lines in the previous step, differential expression analysis and TF activity analysis were used to study gene regulation signatures specific to cell line morphological clusters. The results are shown in Supplemental Table S7, with gene set enrichment analysis showing up-regulation of genes involved in the extracellular matrix, collagens, integrins, and angio-

genesis in the basal-like cluster. Significantly enriched terms ($P < 0.05$) in down-regulated genes include “fatty acid and beta-oxidation” and “ERBB network pathway.” In the genes up-regulated in the luminal-like cluster, we observed enrichment of terms such as “hallmark-oxidative phosphorylation,” down-regulated genes were enriched in “integrin-1 pathway,” “core matrisome,” and genes linked to “hallmark epithelial–mesenchymal transition and migration.” For the remaining B/L group, the term with the highest normalized enrichment score was “targets of the transcription factor MYC” followed by terms associated with ribosomal RNA processing. Down-regulated terms include “cadherin signaling pathway” (Supplemental Table S7).

We also calculated the differential expression for the WGCNA GEMs and found distinct patterns of expression between luminal-like and basal-like clusters of cell lines (Fig. 2B). Among these, the RAP1 signaling module is up-regulated in basal-like clusters and down-regulated in luminal-like clusters. This is consistent with the fact that this GEM is negatively correlated with neighbor fraction, a feature that is observed to decrease in mesenchymal-like cell shapes (Dai et al. 2015). Other modules whose expression distinguishes basal-like from luminal-like include the MAL2-AS1 module (enriched in desmosome assembly), ARNT/KO module (enriched in TNF-signaling by NF- κ B), and ECM organization module (enriched in focal adhesion proteins) (see Supplemental Table S3).

To link the observed gene expression differences (Supplemental Fig. S2B,C) to cell signaling, we used the tool DOROTHEA (Garcia-Alonso et al. 2019) to calculate TF activities, as their modulation is one of the main results of cell signaling processes. We corroborated that the heterogeneous B/L group had significantly activated MYC levels. In the luminal-like cluster, estrogen-related receptor alpha (ESRRA) is the most significantly overrepresented regulome, followed by EHF, KLF5, and ZEB2. Underrepresented regulomes include KLF4, SMAD4, SMAD2, SOX2, and RUNX2. For the basal-like cluster, the regulome with the highest normalized enrichment score is SOX2, as well as MSC and HOXA9. Down-regulated regulomes include ZEB2, MYC, ESRRA, and KLF5 (Supplemental Fig. S2D).

Assembly of a data-driven cell-shape regulatory network

To integrate our data-driven GEMs with signaling pathways, we used the prize-collecting Steiner forest (PCSF) algorithm (Akhmedov et al. 2016). This is an approach that aims to maximize the collection of “prizes” associated with inclusion of relevant nodes, while minimizing the costs associated with edge-weights in a network. This allowed for the integration of the WGCNA modules, the Reactome pathways that regulate them, the TRRUST TFs, and the differentially expressed DOROTHEA regulons into a contiguous regulatory network describing the interplay between cell-shape and breast cancer signaling. The network used for this process was extracted from the database OmniPath (Türei et al. 2016) to provide a map of the intracellular signaling network described as a signed and directed graph. We incorporated identified GEMs into the network by interlinking them as nodes with the relevant TFs and signaling pathways.

The resulting network of 691 nodes included 97.11% of the genes identified by our analysis. The new proteins that were included by the PCSF algorithm to maximize prize collection showed gene set enrichment of common terms (Pathways from PANTHER) (Mi et al. 2013) relative to the original prizes (including WNT, EGF, angiogenesis, RAS, cadherin, and TGF β pathways) but also included are some new terms (VEGF, integrin, and endothelin pathways; $P < 0.001$) (Supplemental Fig. S3A).

Studying the network properties of our PCSF-derived regulatory network, we found that the degree distribution is typical for a biological network (Supplemental Fig. S3B–D). The proteins in the network can be ranked by betweenness centrality to disseminate them based on network importance. Nodes with high centrality lie between many paths and can control information flow. Proteins with the highest centrality are primarily prizes (GSK3B, ESR1, TP53, SMAD3) (Supplemental Fig. S3E), indicating that the PCSF solution was not achieved by the inclusion of new hub proteins that are not of interest to our analysis. Nevertheless, a small minority of high centrality nodes were not in the original prizes, implicating them as mediating the cross talk between pathways identified in Figure 1C. These include the proteins PAX7, PTEN, and PPARGC1A.

Small-molecule inhibitors targeting kinases in our network significantly perturb cell morphology

To validate our network, we used an independent data set to evaluate whether perturbing the function of kinases within our predicted network would produce a significant effect on morphological features. For this, we used the Broad Institute’s Library of Integrated Network-Based Cellular Signatures’ (LINCS) small-molecule kinase inhibitor data set (Subramanian et al.

2017). Here, they measured morphological changes in the breast cancer cell line Hs 578T in response to various small-molecule kinase inhibitors using high-throughput imaging techniques (Hamilton et al. 2007). The morphological variables measured in this data set are mostly analogous to the ones used to construct the network; however, there are some discrepancies that we used as negative controls to ensure our network was phenotype-specific.

We combined this with data from a target affinity assay (Moret et al. 2019) describing the binding affinities of small molecules to kinases. This enabled us to sort the kinase inhibitors into those that target proteins we predict regulate cell shape (through their inclusion in the PCSF-derived network) and those that do not. We found that there is a statistically significant ($n = 37$, Wald test $P < 0.05$) deviation from the control between drug treatments targeting kinases within the predicted network and those targeting other proteins for cytoplasmic area, cytoplasmic perimeter, nucleus area, nucleus length, nucleus width, and nucleus perimeter (Fig. 3A; Supplemental Fig. S4A). This difference is insignificant for features that were not correlated with gene expression modules in our initial analysis (such as number of small spots in the cytoplasm and nucleus, as well as nuclear compactness), indicating that our network is phenotype-specific to the features used in network generation. We also repeated this analysis in other cell lines (SK-BR-3, MCF-7, and nontumorigenic mammary cell line MCF-10A) with results with limited statistical significance (Supplemental Table S8). We additionally used a positive control in which the control cells had been treated with TNF-related apoptosis-inducing ligand (TRAIL) to ensure that the observed morphological effects were not caused by apoptotic factors (Supplemental Fig. S4B; Supplemental Table S8).

We found that there is greater variance in the effect size for kinase inhibitors targeting proteins contained within the predicted regulatory network than those outside. The individual effect on cell morphology for each drug is shown in Supplemental Figure S4, A and B. We hypothesized that it was the network properties of kinases within our network that dictated their effect on morphological features, with some targets being on the periphery of our predicted network and therefore having limited influence over the regulation of cell shape. To test this, we studied the extent to which the effect of a kinase inhibitor was correlated with the combined centrality of its targets as defined by our network. For this we used the centrality algorithm PageRank (Brin and Page 1998) and accounted for off-target effects of the kinase inhibitors using the Szymkiewicz–Simpson index (describing the overlap of a kinase inhibitor’s targets and the proteins that constitute the network) (Methods).

Figure 3B shows moderate correlations between target centrality and the effect size for each feature, illustrating that kinase inhibitors targeting proteins with high centrality in our network modulate cell shape more than inhibitors with peripheral targets. As with studying the effect of targeting kinases contained within our network versus those outside of it, this correlation is higher among morphological variables that are the same or similar to those cell-shape features correlated with GEMs used to construct the network. The correlation between combined centrality and drug absolute effect on cell area ($n = 37$) was moderate but significant for cytoplasm area, cytoplasm perimeter, nucleus area, nucleus length, nucleus half-width, and nucleus perimeter (with Spearman’s correlation coefficient between 0.34–0.37 for all of them, with $P < 0.05$). This correlation in change in morphological features with the centrality of the targeted kinases illustrates the relevance of our constructed network in regulating cell shape.

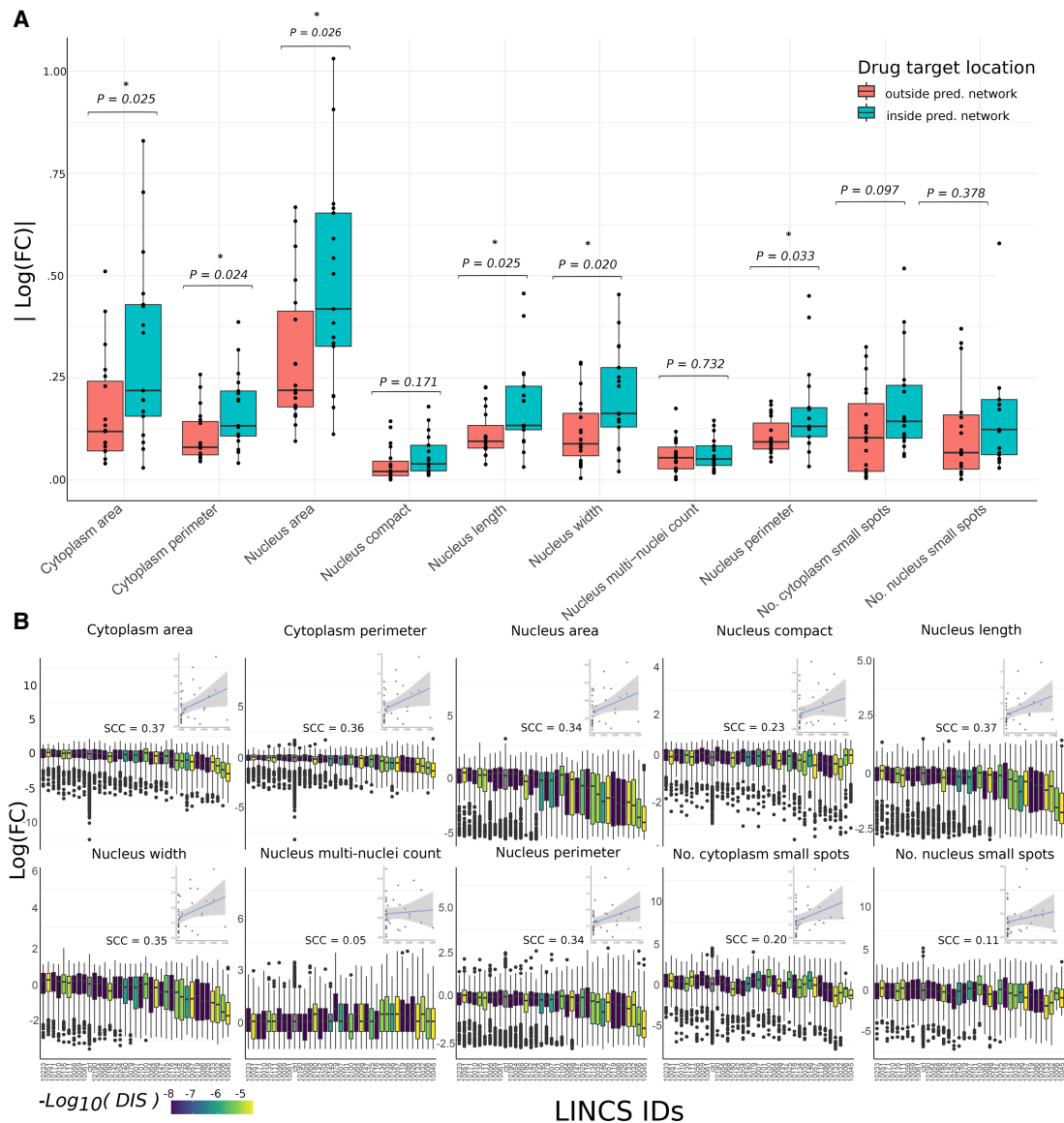


Figure 3. Effect of drug perturbation of derived network on breast cancer cell line morphology. (A) Box plots showing the absolute \log_{10} fold changes after treatment with a drug relative to a control for each cell-shape variable. The drugs are grouped by those targeting kinases within the predicted regulatory network (blue) and those targeting other kinases not predicted to be associated with cell shape (red). P -values (Welch two-sample t -test) are shown with asterisks indicating significance. (B) Bar plot showing the absolute difference in log fold changes of cell-shape variables after treatment with a drug relative to a control. Here, each drug is shown separately (with the LINC ID shown on the x -axis) and colored based on the drug influence score (DIS), and each data point represents a single cell. *Insets* are plots showing the correlation between this influence score and the difference between mean treated cells and mean control cells in each of the 10 measured cell-shape features for each drug. Spearman's correlation coefficients are shown *above* the *inset* plots.

For variables that were not correlated to any GEM, we see visibly lower correlation coefficients and insignificant associations (Spearman's correlation coefficients of 0.05–0.29, $P > 0.05$). These results illustrate that the topology of our network explains some of the variation in the effect of kinase inhibitors tested in a manner that is feature specific to the ones that were used to construct the network model.

Network propagation of activated TFs reveals differentially activated processes in the cell-shape regulatory network

As TF activity remains the most reliable indicator of signaling that can be extracted from transcriptomics data (Szalai and Saez-

Rodriguez 2020), we applied network propagation to identify sub-networks and nodes of which differentially regulated TFs have an effect. The algorithm random walk with restart (RWR) (Tong et al. 2008) was used to diffuse from activated and inactivated TFs in our network reflected by the normalized enrichment scores of TFs identified by DOROTHEA (Methods; Fig. 4A,B; Supplemental Table S9; Garcia-Alonso et al. 2019).

The most relevant supernode in both luminal and basal diffusions was the GEM, RAP1 signaling, a module that is correlated with several cell-shape variables (neighbor frequency, ruffiness, nuclear by cytosolic area, and cell width to length) and is enriched in members of the mechanosensitive RAP1 signaling pathway. By performing RWR diffusions on each of the seed nodes separately

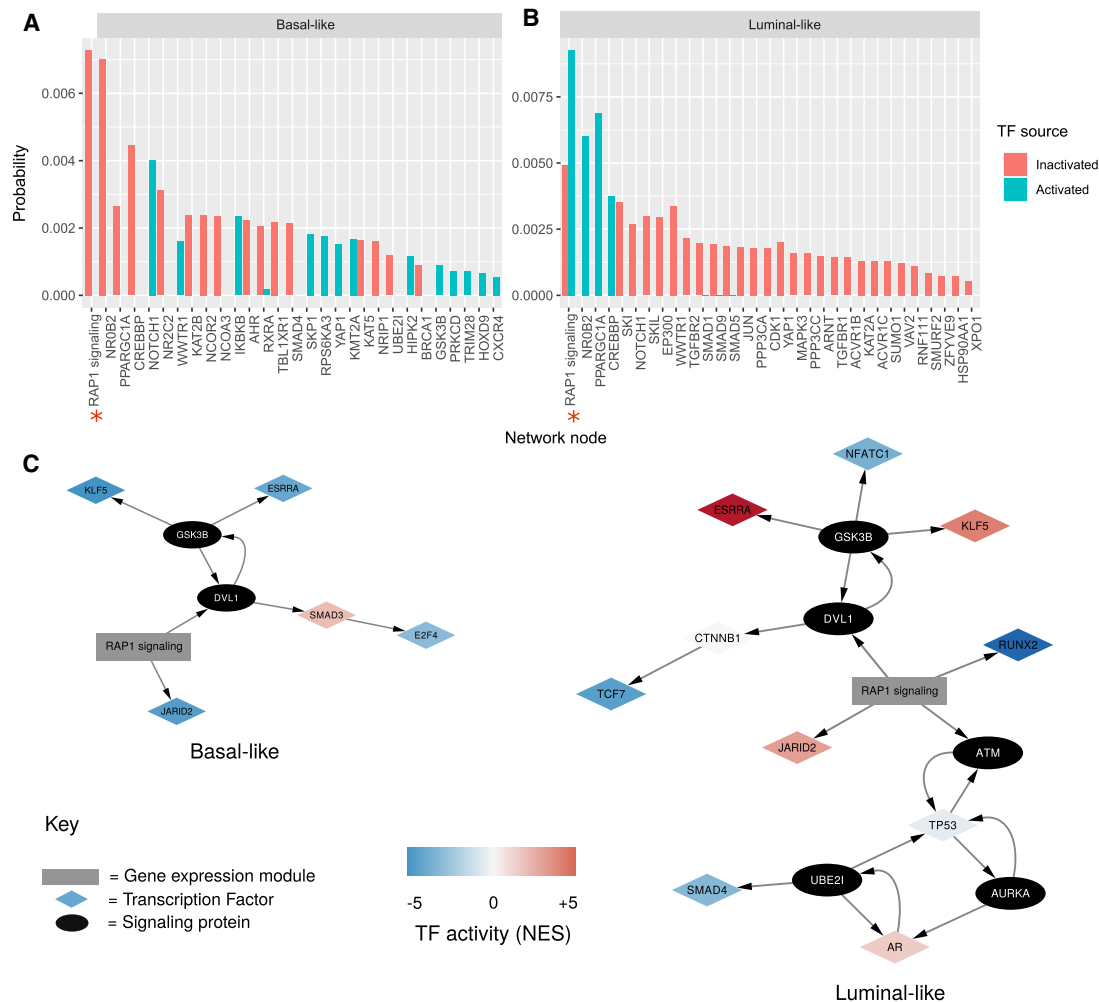


Figure 4. Network propagation of active TFs within cell-shape network. (A, B) Bar plot showing network propagation in a predicted cell-shape network from activated and inactivated TFs in basal-like cell lines (A) and luminal (B). The y-axis is a steady-state probability (or the “heat” of the nodes in the network after the diffusion) over the graph imposed by the starting seeds, ordered by size. Red bars represent propagation from TF seeds that are predicted to be inactivated, and blue bars show propagation from TF seeds that are predicted to be activated. Red asterisks along the x-axis indicate supernodes that represent gene expression modules. Only those nodes with combined probability > 0.0001 are shown, with the full results available in Supplemental Table S9. (C) Subnetworks illustrating the paths between activated TFs (in basal-like and luminal-like) and the “RAP1 signaling” gene expression module. TFs are shown as diamond-shaped nodes, with their color representing their activity. The “RAP1 signaling” gene expression module is shown as a gray rectangle. Signaling proteins are shown as black nodes.

(Supplemental Fig. S5A,B), we can see that the source of this module’s probability is mainly from the TFs JARID2 and RUNX2 in luminal-like cell lines and from JARID2 for basal-like. However, the TFs KLF5 and ESRRA in both morphological subtypes also contribute to the ranking of RAP1 signaling via GSK3B and DVL1 (Fig. 4C).

Specific proteins that were top ranked after performing the network propagation in basal-like cell lines include the orphan nuclear receptor NROB2. Individual RWR found three seed TFs responsible for this node’s high probability: AR, ESRRA, and NR1H3. Other proteins flagged by the propagation were SMAD4, which is regulated by TGF β ; IKKB, which is an activator of NF- κ B; and YAP1. For luminal-like cell lines, NROB2 is also significantly ranked from the network propagation (as a result of ESRRA activity), as well as transcriptional coactivator PPARGC1A and CREBBP.

The RAPI GEM correlates with known morphologically relevant TFs in both cell culture and clinical samples

To explore the significance of the RAP1 GEM in breast cancer, we measured its activity (Methods) in 78 BRCA cell lines. This enabled the correlation of its combined activity with the activity of known TFs predicted by DOROTHEA (Supplemental Fig. S6; Holland et al. 2020). We found that RAP1 GEM activity was significantly correlated (Kendall; $P < 0.01$, FDR adjusted) with the activity of 19 TFs. Among these are RUNX2 (consistent with the results from our network propagation), TEAD1 (TF mediating the function of YAP1/TAZ), and NFKB1. We also correlated TF activity using the same method on tumor samples extracted from The Cancer Genome Atlas (TCGA; <https://www.cancer.gov/tcga>). Using this publicly available data set, we studied 1090 BRCA tumors and performed differential expression on each sample. We found 40 TFs

significantly correlated (Kendall; $P < 0.01$, FDR adjusted) with RAPI GEM (Fig. 5A,B). The intersection of this analysis between in cell lines and the clinical data were the TFs: SP3, NFKB1, ZNF589, ZC3H8, HIF1A, STAT1, ZNF584, ZNF175, and KLF5.

We studied the expression of the module in tumor samples and compared different groups of clinically annotated morphological subtypes. The morphological subtype with the highest overall RAPI-GEM activity was metaplastic carcinoma, a subtype characterized by poorly cohesive sheets (Schwartz et al. 2013) and a high propensity to metastasize (Fig. 5C; Reddy et al. 2020). This morphological subtype has a distribution significantly greater ($P < 0.005$; two-sample Kolmogorov–Smirnov test) than the most frequently assigned morphological subtype (infiltrating duct carcinoma, NOS). This subtype is a common and homogenous breast cancer grouping characterized by its failure to exhibit morphological features that might allow it to be classified as anything more specific (Makki 2015).

Content of RAPI GEM and its network neighborhood shed light on signaling events relevant in the regulation of cell shape

To understand latent processes driven by components within our GEM, we also studied interactions between the Gene Ontology (GO) terms enriched within RAPI GEM (Fig. 5D). This revealed that, as well as RAPI signaling, the GEM is enriched in the AGE-RAGE signaling pathway and HIF1 signaling pathway (consistent with HIF1A's activity correlating highly with RAPI GEM in both cell line patient data). HIF1A is known to be regulated downstream from RAPI (Menon et al. 2012; Li et al. 2021), although not explicitly in breast cancer.

NF- κ B has been previously linked to the regulation of cell shape in breast cancer. To explore the interface of RAPI GEM with NF- κ B in terms of intracellular signaling, we identified a sub-network of our network responsible for mediating “information flow” between those two nodes using the algorithm maximum flow (Fig. 5E). By studying the flow of information from RAPI signaling, we can see that a LATS2/WWTR1/DVL1 (all of WNT signaling) lies between the target and source nodes with much of the flow being carried via these edges. This implicates YAP1/TAZ as being a key effector of the identified GEM. This finding is supported by TEAD1 (mediating gene expression of YAP1 and WWTR1/TAZ) being among the most highly correlated of TFs with RAPI GEM (Fig. 5B; Supplemental Fig. S6A).

Discussion

We present a method that uses transcriptomics and phenotypic data to derive a concise subnetwork describing the signaling involved in the regulation of cell shape. This analysis recovered known processes like “adherens junction proteins,” “cadherin” (Eslami Amirabadi et al. 2019), and “integrin” (Taherian et al. 2011; Filer and Buckley 2013), as well as pathways responsible for the regulation of cell shape in development, such as WNT (Wildwater et al. 2011; Kadzik et al. 2014), TGF β (Lee et al. 2013), and NOTCH (Kontomanolis et al. 2018). All of these pathways have previously been linked to the development of metastatic phenotypes in breast cancer cells (Imamura et al. 2012; Kontomanolis et al. 2018; Yin et al. 2018). Moreover, individual TFs identified include the known promoters of metastasis: SOX2 (Liu et al. 2018, 2; Li et al. 2019, 2), HOXA9 (Ko and Naora 2014), and ESRRA (Berman et al. 2017). Also, among these TFs were known regulators of cell shape and EMT, including KLF5

(Chen et al. 2015), ZEB2 (DaSilva-Arnold et al. 2019), and MYC (Cowling and Cole 2007; Lourenco et al. 2019).

Importantly, this analysis also sheds light on processes with less-characterized associations with cell shape in cancer. We found that a GEM enriched in RAPI signaling is significantly correlated with cell shape and is the most differentially expressed module between luminal-like and basal-like cell line clusters. We found that it was up-regulated in basal-like cell lines and down-regulated in luminal, consistent with its negative correlation with neighbor fraction, a cell-shape feature most contributing to the “cobblestone”-like features of an epithelial and nonmetastatic cell type. This GEM was also an important node in our identified signaling network, being at the network confluence of multiple activated TFs. We also showed this GEM to be expressed in patient data, with its activity being correlated with known developmental and morphologically related TFs, as well as those used to identify it in the network propagation analysis. In this way, our methodology uses cell line data for network construction and validation, but through our network approach, we focus on more general effects that can be tested and successfully validated in a wider breast cancer clinical context. Hence, we believe these results to be relevant in more general breast cancer applications but are also reflecting the inherent context specificity that exists in biology.

The namesake of our identified module, RAPI, is a small GTPase in the RAS-related protein family that has been shown to be involved in the regulation of cell adhesion and migration (Boettner and Van Aelst 2009; Zhang et al. 2017). Specifically, RAPI has been shown to modulate and activate NF- κ B activity in response to TNF stimulation in mesenchymal stem cells (Zhang et al. 2015) and to modulate migration and adhesion (Sawant et al. 2018). RAPI is able to regulate I κ B kinases (IKKs) in a spatio-temporal manner (Ohba et al. 2003) and is crucial for IKK to be able to phosphorylate the NF- κ B subunit RELA to make it competent (Teo et al. 2010). Here, we used our network-centric methodology to highlight a transcriptomic module, characterized in part by RAPI signaling, and this is a key node in our phenotype-specific signaling network. It is possible that our observations of the significance of RAPI are as a result of a more “direct” interaction between RAPI and the cytoskeleton. However, the transcriptomic module that we observed accounts for a much larger system-wide rewiring than simply the modulation of cytoskeletal proteins. This implies more complex transcriptional changes that are characteristic of a more robust breast cancer niche.

The RAPI signaling GEM identified in the network analysis represents a subset of the transcriptome observable among our analyzed cell lines. Although it is enriched in RAPI signaling, it is important to note that it represents a collection of latent biological processes rather than a single pathway assigned to it by gene set enrichment. From our network analysis, we hypothesize that it is able to interact with intracellular signaling pathways in order to modulate TF activity and consequently cell shape. Other pathways enriched in the expression module include the HIF1 signaling pathway, which is known to be activated by RAPI in melanoma (Lee et al. 2015a), but this has not been shown in breast cancer. Hypoxia inducible factor 1 (HIF1) is also of special relevance in tumorigenesis because hypoxia is one of the key stimuli that a cancer cell is able to process in order to determine its fate and maintain the cancer stem cell niche (Plaks et al. 2015). AGE-RAGE signaling was also enriched in our module of interest. The AGE-RAGE signaling pathway has recently been shown to overlap with the RAPI signaling pathway in cardiac fibroblasts to alter the expression of NF- κ B (Burr and Stewart 2021), although this cross talk has also not

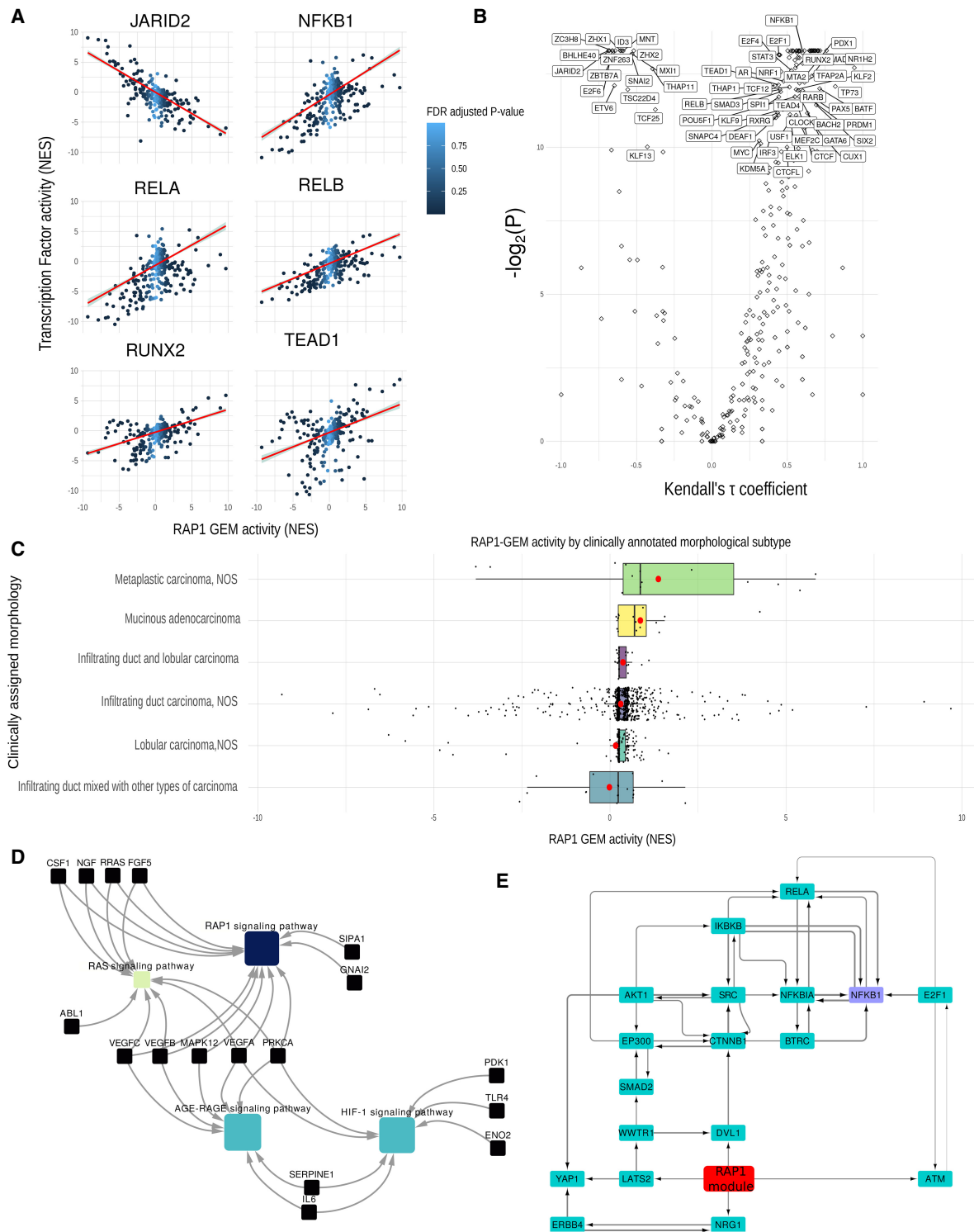


Figure 5. Expression of RAP1 gene expression module in further breast cancer cell lines and in clinical samples. (A) Plots showing the correlation between the RAP1 gene expression module activity (normalized enrichment score) (see Methods) and the activity (NES) of various TF (JARID2, NF- κ B1, RELA, RELB, RUNX2, and TEAD1). The line of best fit according to linear regression is shown in red, with the confidence interval in gray. Color of the points in the plot represents the FDR adjusted P -value of the RAP1 NES as calculated by DOROTHEA. (B) Volcano plot illustrating the correlation (Kendall's rank correlation) between activity of RAP1 gene expression module and TF activity, with Kendall's tau coefficient along the x -axis and $-\log_2(\text{FDR adjusted } P)$ along the y -axis. (C) Bar plot showing RAP1-GEM activity across different breast cancer samples, separated by clinically assigned morphology. The y -axis shows RAP1 gene expression module activity as calculated by DOROTHEA in NES. Mean values for each group are shown by a red dot. (D) Network showing gene set enrichments of the contents of RAP1 gene expression module. Genes are shown in pale blue, and pathways are shown by nodes whose color indicates significance of the associated term $[-\log_2(P)]$. (E) Subnetwork showing the top flow-carrying edges (99th percentile) calculated using the maximum-flow algorithm between RAP1 gene expression module and NFKB1.

been illustrated in breast cancer. Here, we observed genes of RAP1 signaling and AGE-RAGE functioning as a cohesive unit while also being correlated with NF- κ B activity. The combined enrichment of AGE-RAGE, HIF1, and RAP1 signaling is of particular interest because it implies a novel interaction between these three processes, common to all our cell lines, in a manner that has not been previously described in breast cancer.

We also observed that our GEM of interest is significantly correlated with NF- κ B in both clinical samples and cell culture. Other investigators have flagged the direct effect of RAP1 on the cytoskeleton and NF- κ B (Mun and Jeon 2012; Zhang et al. 2015), but here we go further, using our unbiased systems approach to link RAP1 signaling with multiple TFs and pathways. Based on known functions of RAP1, along with the functions of pathways that we found interact with it, we hypothesize that the identified transcriptomic unit is key in relaying information from a cell's physical environment to modulate and maintain the cancer stem cell niche (Roy Choudhury et al. 2019).

Previous studies have established a connection between the NF- κ B signaling pathway and regulation of cell shape in breast cancer (Sero et al. 2015; Sailem and Bakal 2017). Our findings also illustrate the significance of this pathway in the regulation of cell morphology, with multiple NF- κ B regulators and transcriptional coactivators being flagged in our results. Some morphology-correlated GEMs were significantly differentially expressed between cell-shape subtypes with the *ARNT* KO module being significantly up-regulated in basal-like cell shapes relative to luminal. We also found this GEM to have the highest total correlation with all of the morphological features, indicating a strong association with cell shape. By studying terms enriched in this module from the Enrichr library, we found "TNF-alpha signaling via NF- κ B" to be enriched as well as genes down-regulated during AHR nuclear translocator (*ARNT*) shRNA KO. Signaling by TNF is able to activate NF- κ B, a TF known to control the expression of many EMT related genes (Pires et al. 2017), which has shown to be more sensitive to TNF stimulation in mesenchymal-like cellular morphologies than epithelial. This was hypothesized to generate a negative feedback that reinforces a metastatic phenotype of breast cancer cells (Sero et al. 2015). Here we observed also that an *ARNT* KO/TNF module is up-regulated in basal-like cell lines, consistent with these findings. *ARNT* is a protein shown to be involved in regulating tumor growth and angiogenesis along with its binding partner aryl hydrocarbon receptor (AHR) (Huang et al. 2015). Previous studies have also shown its ability to modulate NF- κ B signaling with the activated form possibly interfering with the action of activated *RELA* (Øvrevik et al. 2014). Our findings that the up-regulation of a GEM that is associated with *ARNT* knockdown further give credence to NF- κ B being positively regulated in mesenchymal-like cell morphologies. Furthermore, the results of our network propagation yielded activators and transcriptional coactivators of NF- κ B (IKBKB [Teo et al. 2010], NROB2 [Zou et al. 2015], and CREBBP [Bhatt and Ghosh 2014]). These findings indicate that NF- κ B is modulated by both phosphorylation (through stimulation by TNF), spatial-temporal location (through RAP1), and transcriptional coactivation (through NROB2 and CREBBP) in breast cancer in a shape-dependent manner.

Aside from the biological findings of this study, we illustrate an approach for network analysis of a specific course-grained phenotype through expression, a notoriously poor (if cheap and widely available) proxy for gauging intracellular signaling (Piran et al. 2020). In contrast to existing methods that use gene expression as a direct proxy for signaling (Guan et al. 2012; Ben-Hamo et al.

2014; Soul et al. 2015; Padi and Quackenbush 2018), our approach infers TF activities from the expression data and uses these as an anchor to infer upstream signaling networks relevant to the regulation of our phenotypes. TF activities can represent the outcome of a signal transduction process compared with the expression profiles and are thus a better proxy for cell signaling activities of the cell (Szalai and Saez-Rodriguez 2020). Such an approach has been previously used, for example, by the tool CARNIVAL (Liu et al. 2019). However, this and other available tools neglect the propensity for the transcriptome to be regulated in a highly context-specific and modular structure (Kitano 2002; Sharma and Petsalaki 2019). Moreover, their reliance on annotated pathways to describe cell signaling undermines their ability to spot novel functional units specific to a given phenotype. Here, using context-specific GEMs, we produced a network connecting the genes of interest from diverse analyses and used a network propagation algorithm to further focus on signaling proteins of novel interest. Although there inevitably remains a level of bias stemming from the TF regulon and pathway annotations, our bottom-up approach seeks to identify unbiased latent modular structures within transcriptomic data first. This puts the emphasis on data-driven GEMs rather than on literature-derived regulons and pathways. This approach takes an important step toward reducing the bias associated with previously annotated pathways and allows the identification of important regulatory units and their function with respect to cell shape from a system biology point of view. Our network approach allows us to map the interface between two graphically presented systems in the cell: the transcriptome and intracellular signaling. Both can be easily combined with complex, multivariate phenotypic data, which here has revealed a clearer picture of how signaling regulates cell morphology in breast cancer.

The interoperability of this approach is obvious, with any number of continuous variables measured with gene expression able to be correlated with module eigengenes using WGCNA. Here, we used OmniPath as a base network, but other network-based representations of the cellular environment can be used based on the appropriate context. Thus, our method represents a data-driven, network-based approach compatible with many different multiscale phenotypes that are driven by intracellular signaling. Overall, our unbiased network-based method highlights potential "missing links" between sensing extracellular cues and transcriptional programs that help maintain the cancer stem cell niche and ultimately push breast cancer cells into EMT and metastasis. These represent starting points for further experimental studies to understand and therapeutically target the links between cell shape, cell signaling, and gene regulation in the context of breast cancer.

Methods

WGCNA analysis

Using weighted correlation network analysis, we performed coexpression module identification using the R package WGCNA (Langfelder and Horvath 2008). We used bulk RNA-seq data from ArrayExpress (in FPKM; E-MTAB-2770 and E-MTAB-2706) acquired from commonly used cancer cell lines of various cancer types and with the alignment performed to the NCBI Human Reference Genome GRCh38 (Papatheodorou et al. 2020). We collated 13 breast cancer and one nontumorigenic cell line for which imaging data were available (BT-474, CAMA-1, T-47D, ZR-75-1, SK-BR-3,

MCF-7, HCC1143, HCC1954, HCC70, Hs 578T, JIMT-1, MCF-10A, MDA-MB-157, and MDA-MB-231 (Sero et al. 2015). We acquired representative images of each cell line from Sero et al. (2015; <https://datadryad.org/stash/dataset/doi:10.5061/dryad.tc5g4>).

Cell imaging segmentation was performed using Acapella software (PerkinElmer) with an automated spinning disk confocal microscope. The presented images (Fig. 2) are taken from the above link, stained with DAPI (blue), anti-p65 (Abcam ab16502); Alexa Fluor 488 (anti-rabbit; Invitrogen) (green), and DHE (red). Using Ensembl-BioMart, we filtered genes to only include protein-coding genes (Kinsella et al. 2011) and genes whose FPKM was greater than one, leaving a total of 15,304 genes.

We created a signed, weighted adjacency matrix using \log_2 transformed gene expression values and a soft threshold power (beta) of nine. We translated this adjacency matrix (defined by Equation 1) into a topological overlap matrix (TOM; a measure of similarity), and the corresponding dissimilarity matrix (TOM – 1) was used to identify modules of correlated gene expression (minimum module size of 30). Jackknife cross-validation was used to assess the robustness of the identified modules to the removal of different cell lines from the analysis (Supplemental Fig. S1C), and all showed a high degree of conservation between resampled runs.

$$a_{ij} = |(1 + \text{cor}(x_i, x_j))/2|^\beta. \quad (1)$$

We took morphological variables referring to breast cancer cell lines from Sero et al. (2015), which include 10 significant features shown to be predictive of TF activation. We correlated these features with module eigengenes using Pearson's correlation, and we tested these values for significance by calculating Student asymptotic P -values for given correlations. Multiple hypothesis testing was performed using a permutation based procedure, whereby we recalculated the correlation matrix 1000 times with resampled data. We then generated null distributions for each ranked correlation statistic in our matrix and compared them with our real data of the same rank. We include in Supplemental Table S1 confidence intervals of our permutation-based multiple-correction procedure. For the modules that correlated with morphological features (Pearson correlation coefficient 0.5 and Student $P < 0.05$), we identified enriched signaling pathways using the R package Enrichr (Kuleshov et al. 2016) and the signaling database Reactome (Jassal et al. 2020). Reactome was used in preference to other pathway databases because of the more consistent inclusion of TFs within the annotated pathways. Using the database TRRUST v2 (accessed July 1, 2018) (Han et al. 2018), we identified TF regulons that significantly overlap (Fisher's exact test, $P < 0.1$) with the GEM contents. This was performed separately for inhibitory and activatory expression regulons for each TF, with regulatory relationships of unknown sign being used in the significance calculations for both.

We named gene expression clusters using significantly enriched terms identified by the Enrichr analysis (Supplemental Table S3). As some clusters were very obscure, we used the entire Enrichr list of libraries (<https://maayanlab.cloud/Enrichr/#stats> for full list) with precedence going to the signaling databases of KEGG, Reactome, PANTHER, and Wikipathways (accessed April 1, 2020) (Mi et al. 2013; Kanehisa et al. 2016; Martens et al. 2021). Some modules could not be assigned informative terms and so were named "not annotated" (NA).

Clustering and differential expression

Using the k -means algorithm, we classified the 14 breast cancer cell lines by the median values of each of their shape features ($k=3$) (see Supplemental Fig. S2A). We performed differential expression

analysis using the R package DESeq2 (Love et al. 2014). We filtered genes so that only protein-coding genes and those with >0.5 counts per million in at least eight cell lines were included. We calculated \log_2 fold changes with the cluster of interest as the numerator and the remaining cell lines acting as a control. Using the R package FGSEA (Korotkevich et al. 2021), we performed gene set enrichment analysis of the differentially regulated proteins using the complete pathways gene set (release April 1, 2020) from MSigDB (Liberzon et al. 2015) and the WGCNA GEMs identified in previous analysis. We calculated TF regulon enrichment using the software DOROTHEA (accessed April 1, 2020) (Holland et al. 2020).

Network generation

Using a PCSF algorithm, we generated a cell-shape regulatory network implemented through the R package PCSF (Akhmedov et al. 2016). For the prize-carrying nodes to be collected by the PCSF algorithm, we used the TFs significantly regulating the WGCNA modules using TRRUST v2 ($P < 0.1$), the differentially activated TFs identified by DOROTHEA ($P < 0.1$), and the signaling proteins included in the REACTOME pathways that were enriched in TFs identified ($P < 0.05$). We identified these pathways by using the TRRUST TFs identified in the previous steps, as well as ENCODE and ChEA Consensus TFs from ChIP-X (Lachmann et al. 2010), DNA binding preferences from JASPAR (Wasserman and Sandelin 2004; Stormo 2013), TF protein-protein interactions, and TFs from ENCODE ChIP-seq (Euskirchen et al. 2007). Using Enrichr, we identified pathways that were enriched in the identified TFs, and the proteins that were included in these pathways were extracted from Pathway Commons using the R package paxtoolsr (Luna et al. 2016). This was tested for bias to specific pathways by generating pathway-specific null distributions from 1000 resampled GEMs. Distributions of P -values for each Reactome pathway were generated, in which failed tests (because of no TF enrichment) were given a P -value of one. Results of this were corrected for multiple-hypothesis testing using FDR correction.

The "costs" associated with each edge in the regulatory network were the inverse of the number of sources linked to each regulatory connection scaled between one and zero, such that the more the number of citations for an edge, the lower the cost. For the base network used by the algorithm, we used the comprehensive biological prior knowledge database, OmniPath (accessed May 6, 2020) (Türei et al. 2021), extracted using the R package OmnipathR (Türei et al. 2016). We set each prize for significant TFs or signaling pathways to 100 and used a random variant of the PCSF algorithm with the result being the union of subnetworks obtained on each run (30 iterations) after adding random noise to the edge costs each time (5%). The algorithm also includes a hub-penalization parameter, which we set to 0.005. Other parameters include the tuning of node prices (set to one) and the tuning of trees in the PCSF output (40).

We included the WGCNA modules themselves as supernodes in the network by adding incoming edges from the TFs contained within the regulatory network whose regulomes (as described in TRRUST v2) (Han et al. 2018) significantly overlap (Fisher's exact test; $P < 0.1$) with the gene content of the module in question. We represented the respective cell-shape phenotypes as nodes in a similar fashion by including undirected edges from expression modules and phenotypes in which there was significant correlation ($|PCC| > 0.5$ and $P < 0.05$) between them. To account for expression modules' effect on upstream signaling, we added edges from the WGCNA modules back up to proteins that were themselves included within the modules. We set the edge weight of these to 1, such that any predicted activity of the GEM would be

translated directly into its constituent signaling proteins and thus account for feedback between cell-shape signaling networks and the context-specific expression modules identified in the first step. We identified enriched terms in the network using the 2016 release of the database PANTHER (Mi et al. 2013) and GSE package Enrichr (Kuleshov et al. 2016).

Network propagation of functional TFs

We examined the potential effect of significantly activated (FDR < 0.05) and deactivated TFs in different cell line clusters using network propagation in our generated network. We replaced edge weight with the Resnik best-match average (BMA) semantic similarity (Resnik 1995) between the biological process GO terms of the two interacting pairs, with the sign of the interaction being inherited from Omnipath (Türei et al. 2016). We then scaled the semantic similarity edge weights between one and minus one.

We used the differentially activated TFs identified using DOROTHEA ($P < 0.05$) as seeds for a RWR algorithm using the R package *diffuseR* (available at <https://github.com/dirmeier/diffuser>). We judged a node to be significantly ranked if its affinity score relative to the inputted seeds was greater than the same node's affinity score with a random walk simulation performed with randomized seeds. We performed this randomized simulation 10,000 times, from which a P -value was determined to judge significance ($P < 0.1$). We performed this propagation by RWR for both luminal-like and basal-like morphological clusters on significantly activated and deactivated TFs separately in addition to simulations in which each seed was considered in isolation. We implemented these simulations with a restart probability of 0.95. We generated a graphical representation of the network edges and TFs responsible for the ranking of RAP1 signaling by plotting all the shortest paths between RAP1 and the TFs that caused it to have a non-zero affinity score when each TF was considered in isolation.

Breast cancer cell morphology following kinase inhibitor treatment

We used single-cell, small-molecule kinase inhibition data from the Harvard Medical School (HMS) LINCS Center (Stathias et al. 2020), which is funded by NIH grants U54 HG006097 and U54 HL127365 (available from <https://lincs.hms.harvard.edu/mills-unpubl-2015/>, accessed August 1, 2020). This data set is derived from the treatment of six cell lines with a panel of 105 small-molecule kinase inhibitors. They measured textural and morphological variables following treatment by high-throughput image analysis (Haralick et al. 1973; Hamilton et al. 2007). We combined this assay with another data set from HMS-LINCS: a target affinity spectrum (TAS) for compounds in the HMS-LINCS small-molecule library measuring the binding assertions based on dose response affinity constants for particular kinase inhibitors (<https://lincs.hms.harvard.edu/db/datasets/20000/>, accessed August 1, 2020). Using this data set, we filtered for only molecule-binding target pairs with a binding class of one (representing a $K_D < 100$ nM affinity). Further to this, we removed molecules that had more than five targets with a K_D of 100 nM. Consequently, the remaining kinase inhibitors were relatively narrow spectrum, thus simplifying analysis of their phenotypic effect. We expressed these results as batch-specific log fold changes of 10- μ m drug treatment relative to the mean of the control set (untreated and DMSO-treated cells). Spearman's rank correlation was calculated between the drug target's network centrality and the absolute log fold change of the morphological variable. We also used the Kolmogorov–Smirnov statistic to assess significance between cell morphology after treat-

ment with drugs targeting kinases inside versus outside our predicted network. This was also repeated on other breast cancer cell lines and using a TRAIL control (Supplemental Table S8).

The morphological data in the kinase inhibition screen were measured using two dyes (DRAQ5 and TMRE), the intensity of which we used to normalize textural features and the measurement of cytoplasmic and nuclear small spots. We reported counts for small-nuclear or cytoplasmic spots as a mean of the individually normalized readings from both dyes. We considered a treatment perturbing our network if at least one of the kinase inhibitors targeted a protein that was represented by a node within the network.

Quantifying kinase inhibitor influence

We incorporate information from the TAS assay, as well as graph-based properties of kinase inhibitor targets, using the product of the Szymkiewicz–Simpson similarity (measured between the cell-shape network nodes and the drug targets) and the centrality of the targeted nodes in the predicted network with semantic similarity edge weights. The product of these generates, for a given kinase inhibitor, the statistic

$$\sum_{x \in K \cap N} PR(x) \times \frac{K \cap N}{\min(|K|, |N|)}, \quad (2)$$

where K is the set of kinases an inhibitor is predicted to target, N is the nodes of the network, and the function $PR()$ is the centrality of a particular node in the network as defined by the PageRank algorithm (Brin and Page 1998). This centrality measure has been shown to be effective in prioritizing proteins by relative importance in signaling or protein–protein interaction networks (Iván and Grolmusz 2011). We used this statistic as a measure of a kinase inhibitor's influence on cell shape.

Analysis of BRCA cell line and TCGA sample RNA-seq data

For the cell lines, we used RNA-seq data from the ArrayExpress (in FPKM; E-MTAB-2770 and E-MTAB-2706) (Papatheodorou et al. 2020). This was analyzed using DESeq2 (Love et al. 2014) per the methodology in the “Clustering and Differential Expression” section. Both TF and module activity were calculated using the algorithm VIPER (Alvarez et al. 2016). For patient data, the results shown here are based on data generated by The Cancer Genome Atlas (TCGA) Research Network (<https://www.cancer.gov/tcga>, accessed 01/04/21). For computational efficiency, we used gamma-Poisson models to predict differentially expressed genes from our samples using the package *glmGamPoi* (Ahlmann-Eltze and Huber 2021). We used the sample of interest as the numerator with the remaining tumor samples acting as a control. For quantifying correlation between RAP1 GEM and different TFs, we removed samples with insignificant activation of either the TF in question or RAP1 GEM (FDR adjusted P -value < 0.05). Correlation was quantified using the Kendall rank correlation coefficient. Differences in distributions of morphological subtypes was quantified by the Kolmogorov–Smirnov test.

Maximum-flow network analysis

For maximum-flow calculations, we used the Resnik BMA semantic similarity (Resnik 1995) as the maximum “carrying capacity” of an edge in the network. To visualize the optimized solution (as implemented by the R package *igraph*) (Csardi and Nepusz 2006), we selected only those edges in the 99th percentile of the flow-carrying edges in the network. Visualization was performed using the software Cytoscape (Shannon et al. 2003). Maximum flow was performed with the R package *igraph* (Csardi and Nepusz 2006).

Quantification and statistical analysis

Statistical tests were performed in base R (R Core Team 2021) unless otherwise mentioned in the Methods, and *P*-value cut-offs are shown in parentheses after reporting an effect as significant. A weighted Pearson's correlation with *t*-test for significance was used to correlate eigengenes and cell-shape features using the RNA package WGCNA. We used a one-way ANOVA test for comparing the means of the shape variables among the identified three cell line clusters ($n=75,653$) and a Tukey honest significant difference test to perform multiple-pairwise comparison among the means of the groups. The same tests were performed on the differences in 10 cell-shape variables when Hs 578T was treated with 37 kinase inhibitors ($n=23,128$). A Fisher's exact test was used to test significance of overlap between TRRUST regulons and identified GEMs (for size of the overlap, see Supplemental Table S4).

Enrichment of gene sets was performed by Enrichr, an enrichment library that uses a hypergeometric test to identify significantly enriched terms in a gene list. This tool (described by Chen et al. 2013) calculates a score combining the Fisher's exact test *P*-value of the enrichment with the *z*-score deviation from the expected rank. The preranked gene set enrichment algorithm FGSEA was used for the identification of enriched terms in the differentially expressed genes, allowing for accurate estimation of arbitrarily low *P*-values that occur in expression data sets.

Spearman rank correlation was used to measure the strength of the association between target network centrality and the measured effect of its perturbation by inhibition. Spearman was chosen because the centrality (combined with Szymkiewicz–Simpson) according to Equation 2 does not follow an exact normal distribution. Kendall rank correlation coefficient was used when calculating the correlation between TF activity and RAP1 GEM activity because confidence intervals for Spearman's r_s are less reliable and less interpretable than confidence intervals for Kendall's tau parameters. When trying to distinguish between many correlations of similar quality, this becomes more important. FDR adjustment for multiple testing correction was always used when multiple tests were performed in the same analysis.

Kolmogorov–Smirnov test was used to measure differences in distributions of clinically assigned tumor morphologies. This was because clinical groupings are mixed (i.e., infiltrating duct and lobular carcinoma), and others are characterized by an absence of features over their presence. This means that the assumption of normality required for a *t*-test is not fulfilled.

For differential expression analysis, the DESeq2 R package (Love et al. 2014) was used. DESeq2 fits negative binomial generalized linear models for each gene and uses the Wald test for significance testing. The package then automatically detects count outliers using Cook's distance and removes these genes from analysis.

Significance was determined for RWR network propagation by randomizing seed nodes (preserving their values) 10,000 times and selecting only the nonseed nodes that were significantly ranked relative to the randomized simulations ($P < 0.1$). Figures were presented using ggplot (Wickham 2009).

Software availability

The complete R scripts and data used for this methodology are available as Supplemental Code and at GitLab (https://gitlab.ebi.ac.uk/petsalakilab/phenotype_networks).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank the European Molecular Biology Laboratory (EMBL) and EMBL-EBI for funding the project. C.G.B. was funded by the EMBL International PhD Program. We also thank Bishoy Wadie and Vivian Robin for critical reading of the manuscript.

Author contributions: C.B. was responsible for conceptualization, methodology, software, validation, formal analysis, and writing the original draft. E.i.P. was responsible for conceptualization, methodology, and software. G.G. was responsible for conceptualization and supervision. J.S. was responsible for resources. E.N.E. was responsible for validation and writing the original draft. C.h.r.B. was responsible for writing, review, and editing and resources. E.P. writing, review, and editing; conceptualization; supervision; methodology; project administration; and funding acquisition.

References

- Ahlmann-Eltz C, Huber W. 2021. Glmgampoi: fitting γ -Poisson generalized linear models on single cell count data. *Bioinformatics* **36**: 5701–5702. doi:10.1093/bioinformatics/btaa1009
- Akhmedov M, Kwee I, Montemanni R. 2016. A divide and conquer mathematical algorithm for the prize-collecting Steiner tree problem. *Comput Oper Res* **70**: 18–25. doi:10.1016/j.cor.2015.12.015
- Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, Califano A. 2016. Network-based inference of protein activity helps functionalize the genetic landscape of cancer. *Nat Genet* **48**: 838–847. doi:10.1038/ng.3593
- Baskaran JP, Weldy A, Guarin J, Munoz G, Shpilker PH, Kotlik M, Subbiah N, Wishart A, Peng Y, Miller MA, et al. 2020. Cell shape, and not 2D migration, predicts extracellular matrix-driven 3D cell invasion in breast cancer. *APL Bioeng* **4**: 026105. doi:10.1063/1.5143779
- Ben-Hamo R, Gidoni M, Efroni S. 2014. Phenonet: identification of key networks associated with disease phenotype. *Bioinformatics* **30**: 2399–2405. doi:10.1093/bioinformatics/btu199
- Bergert M, Lembo S, Sharma S, Russo L, Milovanović D, Gretarsson KH, Börmel M, Neveu PA, Hackett JA, Petsalaki E, et al. 2021. Cell surface mechanics gate embryonic stem cell differentiation. *Cell Stem Cell* **28**: 209–216.e4. doi:10.1016/j.stem.2020.10.017
- Berman AY, Manna S, Schwartz NS, Katz YE, Sun Y, Behrmann CA, Yu JJ, Plas DR, Alayev A, Holz MK. 2017. ERR α regulates the growth of triple-negative breast cancer cells via S6K1-dependent mechanism. *Signal Transduct Target Ther* **2**: e17035. doi:10.1038/sigtrans.2017.35
- Bhatt D, Ghosh S. 2014. Regulation of the NF- κ B-mediated transcription of inflammatory genes. *Front Immunol* **5**: 71. doi:10.3389/fimmu.2014.00071
- Boettner B, Van Aelst L. 2009. Control of cell adhesion dynamics by Rap1 signaling. *Curr Opin Cell Biol* **21**: 684–693. doi:10.1016/j.ccb.2009.06.004
- Bougault C, Aubert-Foucher E, Paumier A, Perrier-Groult E, Huot L, Hot D, Duterque-Coquillaud M, Mallein-Gerin F. 2012. Dynamic compression of chondrocyte-agarose constructs reveals new candidate mechanosensitive genes. *PLoS One* **7**: e36964. doi:10.1371/journal.pone.0036964
- Brin S, Page L. 1998. The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* **30**: 107–117. doi:10.1016/S0169-7552(98)00110-X
- Burr SD, Stewart JA. 2021. Rap1a overlaps the AGE/RAGE signaling cascade to alter expression of α -SMA, p-NF- κ B, and p-PKC- ζ in cardiac fibroblasts isolated from type 2 diabetic mice. *Cells* **10**: 557. doi:10.3390/cells10030557
- Chatterjee S. 2018. Endothelial mechanotransduction, redox signaling and the regulation of vascular inflammatory pathways. *Front Physiol* **9**: 524. doi:10.3389/fphys.2018.00524
- Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**: 128. doi:10.1186/1471-2105-14-128
- Chen W-C, Lin H-H, Tang M-J. 2015. Matrix-stiffness-regulated inverse expression of Krüppel-like factor 5 and Krüppel-like factor 4 in the pathogenesis of renal fibrosis. *Am J Pathol* **185**: 2468–2481. doi:10.1016/j.ajpath.2015.05.019
- Cooper J, Giancotti FG. 2019. Integrin signaling in cancer: mechanotransduction, stemness, epithelial plasticity, and therapeutic resistance. *Cancer Cell* **35**: 347–367. doi:10.1016/j.ccell.2019.01.007

- Cowell CF, Yan IK, Eiseler T, Leightner AC, Döppler H, Storz P. 2009. Loss of cell–cell contacts induces NF- κ B via RhoA-mediated activation of protein kinase D1. *J Cell Biochem* **106**: 714–728. doi:10.1002/jcb.22067
- Cowling VH, Cole MD. 2007. E-cadherin repression contributes to c-Myc-induced epithelial cell transformation. *Oncogene* **26**: 3582–3586. doi:10.1038/sj.onc.1210132
- Csardi G, Nepusz T. 2006. The igraph software package for complex network research. *InterJournal Complex Syst* **1695**: 1–9.
- Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, Shi B. 2015. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res* **5**: 2929–2943.
- DaSilva-Arnold SC, Kuo C-Y, Davra V, Remache Y, Kim PCW, Fisher JP, Zamudio S, Al-Khan A, Birge RB, Illsley NP. 2019. ZEB2, a master regulator of the epithelial–mesenchymal transition, mediates trophoblast differentiation. *Mol Hum Reprod* **25**: 61–75. doi:10.1093/molehr/gay053
- De Belly H, Stubb A, Yanagida A, Labouesse C, Jones PH, Paluch EK, Chalut KJ. 2021. Membrane tension gates ERK-mediated regulation of pluripotent cell fate. *Cell Stem Cell* **28**: 273–284.e6. doi:10.1016/j.stem.2020.10.018
- Eslami Amirabadi H, Tuerlings M, Hollestelle A, SahebAli S, Luttge R, van Donkelaar CC, Martens JWM, den Toonder MJM. 2019. Characterizing the invasion of different breast cancer cell lines with distinct E-cadherin status in 3D using a microfluidic system. *Biomed Microdevices* **21**: 101. doi:10.1007/s10544-019-0450-5
- Euskirchen GM, Rozowsky JS, Wei C-L, Lee WH, Zhang ZD, Hartman S, Emanuelsson O, Stolc V, Weissman S, Gerstein MB, et al. 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. *Genome Res* **17**: 898–909. doi:10.1101/gr.5583007
- Fedele M, Cerchia L, Chiappetta G. 2017. The epithelial-to-mesenchymal transition in breast cancer: focus on basal-like carcinomas. *Cancers (Basel)* **9**: 134. doi:10.3390/cancers9100134
- 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. doi:10.1016/j.gendis.2018.05.001
- Filer A, Buckley CD. 2013. 15 - Fibroblasts and fibroblast-like synoviocytes. In *Kelley's textbook of rheumatology*, 2nd ed. (ed. Firestein GS, et al.), pp. 215–231. W.B. Saunders, Philadelphia.
- Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. 2019. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res* **29**: 1363–1375. doi:10.1101/gr.240663.118
- Guan Y, Gorenshsteyn D, Burmeister M, Wong AK, Schimenti JC, Handel MA, Bult CJ, Hibbs MA, Troyanskaya OG. 2012. Tissue-specific functional networks for prioritizing phenotype and disease genes. *PLoS Comput Biol* **8**: e1002694. doi:10.1371/journal.pcbi.1002694
- Hamilton NA, Pantelic RS, Hanson K, Teasdale RD. 2007. Fast automated cell phenotype image classification. *BMC Bioinformatics* **8**: 110. doi:10.1186/1471-2105-8-110
- Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, Yang S, Kim CY, Lee M, Kim E, et al. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* **46**: D380–D386. doi:10.1093/nar/gkx1013
- Haralick RM, Shanmugam K, Dinstein I. 1973. Textural features for image classification. *IEEE Trans Syst Man Cybern SMC-3*: 610–621. doi:10.1109/TSMC.1973.4309314
- Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, Joughin BA, Stegle O, Lauffenburger DA, Heyn H, et al. 2020. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol* **21**: 36. doi:10.1186/s13059-020-1949-z
- Horton ER, Astudillo P, Humphries MJ, Humphries JD. 2016. Mechanosensitivity of integrin adhesion complexes: role of the consensus adhesome. *Exp Cell Res* **343**: 7–13. doi:10.1016/j.yexcr.2015.10.025
- Huang C-R, Lee C-T, Chang K-Y, Chang W-C, Liu Y-W, Lee J-C, Chen B-K. 2015. Down-regulation of ARNT promotes cancer metastasis by activating the fibronectin/integrin β 1/FAK axis. *Oncotarget* **6**: 11530–11546. doi:10.18632/oncotarget.3448
- Imamura T, Hikita A, Inoue Y. 2012. The roles of TGF- β signaling in carcinogenesis and breast cancer metastasis. *Breast Cancer* **19**: 118–124. doi:10.1007/s12282-011-0321-2
- Ishihara Y, Kado SY, Hoepfer C, Harel S, Vogel CFA. 2019. Role of NF- κ B relB in aryl hydrocarbon receptor-mediated ligand specific effects. *Int J Mol Sci* **20**: 2652. doi:10.3390/ijms20112652
- Iván G, Grolmusz V. 2011. When the web meets the cell: using personalized PageRank for analyzing protein interaction networks. *Bioinformatics* **27**: 405–407. doi:10.1093/bioinformatics/btq680
- Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, et al. 2020. The reactome pathway knowledgebase. *Nucleic Acids Res* **48**: D498–D503. doi:10.1093/nar/gkz1031
- Kadzik RS, Cohen ED, Morley MP, Stewart KM, Lu MM, Morrisey EE. 2014. Wnt ligand/frizzled 2 receptor signaling regulates tube shape and branch-point formation in the lung through control of epithelial cell shape. *Proc Natl Acad Sci* **111**: 12444–12449. doi:10.1073/pnas.1406639111
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457–D462. doi:10.1093/nar/gkv1070
- Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011**: bar030. doi:10.1093/database/bar030
- Kitano H. 2002. Systems biology: a brief overview. *Science* **295**: 1662–1664. doi:10.1126/science.1069492
- Ko SY, Naora H. 2014. HOXA9 promotes homotypic and heterotypic cell interactions that facilitate ovarian cancer dissemination via its induction of P-cadherin. *Mol Cancer* **13**: 170. doi:10.1186/1476-4598-13-170
- Kontomanolis EN, Kalagasidou S, Pouliou S, Anthoulaki X, Georgiou N, Papamanolis V, Fasoulakis ZN. 2018. The notch pathway in breast cancer progression. *Sci World J* **2018**: 2415489. doi:10.1155/2018/2415489
- Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. 2021. Fast gene set enrichment analysis. bioRxiv doi:10.1101/060012
- Krakhmal NV, Zavyalova MV, Denisov EV, Vtorushin SV, Perelmuter VM. 2015. Cancer invasion: patterns and mechanisms. *Acta Naturae* **7**: 17–28. doi:10.32607/20758251-2015-7-2-17-28
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**: W90–W97. doi:10.1093/nar/gkw377
- Kumar S, Jang I, Kim CW, Kang D-W, Lee WJ, Jo H. 2016. Functional screening of mammalian mechanosensitive genes using *Drosophila* RNAi library–*Smarcd3/Bap60* is a mechanosensitive pro-inflammatory gene. *Sci Rep* **6**: 36461. doi:10.1038/srep36461
- Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**: 2438–2444. doi:10.1093/bioinformatics/btq466
- Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**: 559. doi:10.1186/1471-2105-9-559
- Lee J, Choi J-H, Joo C-K. 2013. TGF- β 1 regulates cell fate during epithelial–mesenchymal transition by upregulating survivin. *Cell Death Dis* **4**: e714–e714. doi:10.1038/cddis.2013.244
- Lee J-W, Ryu Y-K, Ji Y-H, Kang JH, Moon E-Y. 2015a. Hypoxia/reoxygenation-experienced cancer cell migration and metastasis are regulated by Rap1- and Rac1-GTPase activation via the expression of thymosin β -4. *Oncotarget* **6**: 9820–9833. doi:10.18632/oncotarget.3218
- Lee M-H, Wu P-H, Gilkes D, Aifuwa I, Wirtz D. 2015b. Normal mammary epithelial cells promote carcinoma basement membrane invasion by inducing microtubule-rich protrusions. *Oncotarget* **6**: 32634–32645. doi:10.18632/oncotarget.4728
- Li Z-R, Jiang Y, Hu J-Z, Chen Y, Liu Q-Z. 2019. SOX2 knockdown inhibits the migration and invasion of basal cell carcinoma cells by targeting the SRPK1-mediated PI3K/AKT signaling pathway. *Oncol Lett* **17**: 1617–1625. doi:10.3892/ol.2018.9810
- Li H, Liang J, Wang J, Han J, Li S, Huang K, Liu C. 2021. Mex3a promotes oncogenesis through the RAP1/MAPK signaling pathway in colorectal cancer and is inhibited by hsa-miR-6887-3p. *Cancer Commun* **41**: 472–491. doi:10.1002/cac2.12149
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**: 417–425. doi:10.1016/j.cels.2015.12.004
- Liu P, Tang H, Song C, Wang J, Chen B, Huang X, Pei X, Liu L. 2018. SOX2 promotes cell proliferation and metastasis in triple negative breast cancer. *Front Pharmacol* **9**: 942. doi:10.3389/fphar.2018.00942
- Liu A, Trairatphan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J. 2019. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. *NPJ Syst Biol Appl* **5**: 40. doi:10.1038/s41540-019-0118-z
- Lourenco C, Kalkat M, Houlahan KE, De Melo J, Longo J, Done SJ, Boutros PC, Penn LZ. 2019. Modelling the MYC-driven normal-to-tumour switch in breast cancer. *Dis Model Mech* **12**: dmm038083. doi:10.1242/dmm.038083
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Luna A, Babur Ö, Aksoy BA, Demir E, Sander C. 2016. PaxtoolsR: pathway analysis in R using Pathway Commons. *Bioinformatics* **32**: 1262–1264. doi:10.1093/bioinformatics/btv733

- Makki J. 2015. Diversity of breast carcinoma: histological subtypes and clinical relevance. *Clin Med Insights Pathol* **8**: 23–31. doi:10.4137/CPath.S31563
- Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, A. Miller R, Digles D, Lopes EN, Ehrhart F, et al. 2021. WikiPathways: connecting communities. *Nucleic Acids Res* **49**: D613–D621. doi:10.1093/nar/gkaa1024
- Menon J, Doebele RC, Gomes S, Bevilacqua E, Reindl KM, Rosner MR. 2012. A novel interplay between Rap1 and PKA regulates induction of angiogenesis in prostate cancer. *PLoS One* **7**: e49893. doi:10.1371/journal.pone.0049893
- Mi H, Muruganujan A, Thomas PD. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* **41**: D377–D386. doi:10.1093/nar/gks1118
- Miralles F, Posern G, Zaromytidou A-I, Treisman R. 2003. Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* **113**: 329–342. doi:10.1016/S0092-8674(03)00278-2
- Moret N, Clark NA, Hafner M, Wang Y, Loukine E, Medvedovic M, Wang J, Gray N, Jenkins J, Sorger PK. 2019. Cheminformatics tools for analyzing and designing optimized small-molecule collections and libraries. *Cell Chem Biol* **26**: 765–777.e3. doi:10.1016/j.chembiol.2019.02.018
- Mun H, Jeon TJ. 2012. Regulation of actin cytoskeleton by Rap1 binding to RacGEF1. *Mol Cells* **34**: 71–76. doi:10.1007/s10059-012-0097-z
- Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, et al. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**: 515–527. doi:10.1016/j.ccr.2006.10.008
- Ohba Y, Kurokawa K, Matsuda M. 2003. Mechanism of the spatio-temporal regulation of Ras and Rap1. *EMBO J* **22**: 859–869. doi:10.1093/emboj/cdg087
- Olson EN, Nordheim A. 2010. Linking actin dynamics and gene transcription to drive cellular motile functions. *Nat Rev Mol Cell Biol* **11**: 353–365. doi:10.1038/nrm2890
- Orsulic S, Huber O, Aberle H, Arnold S, Kemler R. 1999. E-cadherin binding prevents β -catenin nuclear localization and β -catenin/LEF-1-mediated transactivation. *J Cell Sci* **112**: 1237–1245. doi:10.1242/jcs.112.8.1237
- Øvrevik J, Låg M, Lecreur V, Gilot D, Lagadic-Gossman D, Refsnes M, Schwarze PE, Skuland T, Becher R, Holme JA. 2014. AhR and Arnt differentially regulate NF- κ B signaling and chemokine responses in human bronchial epithelial cells. *Cell Commun Signal* **12**: 48. doi:10.1186/s12964-014-0048-8
- Padi M, Quackenbush J. 2018. Detecting phenotype-driven transitions in regulatory network structure. *NPJ Syst Biol Appl* **4**: 16. doi:10.1038/s41540-018-0052-5
- Papatheodorou I, Moreno P, Manning J, Fuentes AM-P, George N, Fexova S, Fonseca NA, Füllgrabe A, Green M, Huang N, et al. 2020. Expression Atlas update: from tissues to single cells. *Nucleic Acids Res* **48**: D77–D83. doi:10.1093/nar/gkz947
- Piran M, Karbalaie R, Piran M, Aldahdooh J, Mirzaie M, Ansari-Pour N, Tang J, Jafari M. 2020. Can we assume the gene expression profile as a proxy for signaling network activity? *Biomolecules* **10**: 850. doi:10.3390/biom10060850
- Pires BRB, Mencialha AL, Ferreira GM, de Souza WF, Morgado-Díaz JA, Maia AM, Corrêa S, Abdelhay ESFW. 2017. NF- κ B is involved in the regulation of EMT genes in breast cancer cells. *PLoS One* **12**: e0169622. doi:10.1371/journal.pone.0169622
- Plaks V, Kong N, Werb Z. 2015. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell* **16**: 225–238. doi:10.1016/j.stem.2015.02.015
- R Core Team. 2021. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Reddy TP, Rosato RR, Li X, Moulder S, Piwnica-Worms H, Chang JC. 2020. A comprehensive overview of metaplastic breast cancer: clinical features and molecular aberrations. *Breast Cancer Res* **22**: 121. doi:10.1186/s13058-020-01353-z
- Resnik P. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence – Vol. 1, IJCAI'95*, pp. 448–453. Morgan Kaufmann Publishers, San Francisco.
- Robertson J, Jacquemet G, Byron A, Jones MC, Warwood S, Selley JN, Knight D, Humphries JD, Humphries MJ. 2015. Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. *Nat Commun* **6**: 6265. doi:10.1038/ncomms7265
- Roche J. 2018. The epithelial-to-mesenchymal transition in cancer. *Cancers (Basel)* **10**: 52. doi:10.3390/cancers10020052
- Rolfe RA, Nowlan NC, Kenny EM, Cormican P, Morris DW, Prendergast PJ, Kelly D, Murphy P. 2014. Identification of mechanosensitive genes during skeletal development: alteration of genes associated with cytoskeletal rearrangement and cell signalling pathways. *BMC Genomics* **15**: 48. doi:10.1186/1471-2164-15-48
- Roy Choudhury A, Gupta S, Chaturvedi PK, Kumar N, Pandey D. 2019. Mechanobiology of cancer stem cells and their niche. *Cancer Microenviron* **12**: 17–27. doi:10.1007/s12307-019-00222-4
- Sailem HZ, Bakal C. 2017. Identification of clinically predictive metagenes that encode components of a network coupling cell shape to transcription by image-omics. *Genome Res* **27**: 196–207. doi:10.1101/gr.202028.115
- Sawant K, Chen Y, Kotian N, Preuss KM, McDonald JA. 2018. Rap1 GTPase promotes coordinated collective cell migration in vivo. *Mol Biol Cell* **29**: 2656–2673. doi:10.1091/mbc.E17-12-0752
- Schwartz TL, Mogal H, Papageorgiou C, Veerapong J, Hsueh EC. 2013. Metaplastic breast cancer: histologic characteristics, prognostic factors and systemic treatment strategies. *Exp Hematol Oncol* **2**: 31. doi:10.1186/2162-3619-2-31
- Sero JE, Sailem HZ, Ardy RC, Almuttaqi H, Zhang T, Bakal C. 2015. Cell shape and the microenvironment regulate nuclear translocation of NF- κ B in breast epithelial and tumor cells. *Mol Syst Biol* **11**: 790. doi:10.15252/msb.20145644
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498–2504. doi:10.1101/gr.1239303
- Sharma S, Petsalaki E. 2019. Large-scale datasets uncovering cell signalling networks in cancer: context matters. *Curr Opin Genet Dev* **54**: 118–124. doi:10.1016/j.gde.2019.05.001
- Shih W, Yamada S. 2012. N-cadherin-mediated cell–cell adhesion promotes cell migration in a three-dimensional matrix. *J Cell Sci* **125**: 3661–3670. doi:10.1242/jcs.103861
- Shrum CK, Defrancisco D, Meffert MK. 2009. Stimulated nuclear translocation of NF- κ B and shuttling differentially depend on dynein and the dynactin complex. *Proc Natl Acad Sci* **106**: 2647–2652. doi:10.1073/pnas.0806677106
- Siegel RL, Miller KD, Jemal A. 2019. Cancer statistics, 2019. *CA Cancer J Clin* **69**: 7–34. doi:10.3322/caac.21551
- Song Y, Washington MK, Crawford HC. 2010. Loss of FOXA1/2 is essential for the epithelial-to-mesenchymal transition in pancreatic cancer. *Cancer Res* **70**: 2115–2125. doi:10.1158/0008-5472.CAN-09-2979
- Soul J, Hardingham TE, Boot-Handford RP, Schwartz J-M. 2015. PhenomeExpress: a refined network analysis of expression datasets by inclusion of known disease phenotypes. *Sci Rep* **5**: 8117. doi:10.1038/srep08117
- Stathias V, Turner J, Koleti A, Vidovic D, Cooper D, Fazel-Najafabadi M, Pilarczyk M, Terry R, Chung C, Umeano A, et al. 2020. LINC5 Data Portal 2.0: next generation access point for perturbation-response signatures. *Nucleic Acids Res* **48**: D431–D439. doi:10.1093/nar/gkz1023
- Stormo GD. 2013. Modeling the specificity of protein-DNA interactions. *Quant Biol* **1**: 115–130. doi:10.1007/s40484-013-0012-4
- Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, et al. 2017. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**: 1437–1452.e17. doi:10.1016/j.cell.2017.10.049
- Szalai B, Saez-Rodriguez J. 2020. Why do pathway methods work better than they should? *FEBS Lett* **594**: 4189–4200. doi:10.1002/1873-3468.14011
- Taherian A, Li X, Liu Y, Haas TA. 2011. Differences in integrin expression and signaling within human breast cancer cells. *BMC Cancer* **11**: 293. doi:10.1186/1471-2407-11-293
- Teo H, Ghosh S, Luesch H, Ghosh A, Wong ET, Malik N, Orth A, de Jesus P, Perry AS, Oliver JD, et al. 2010. Telomere-independent Rap1 is an IKK adaptor and regulates NF- κ B-dependent gene expression. *Nat Cell Biol* **12**: 758–767. doi:10.1038/ncb2080
- Tong L, Teragaonkar V. 2014. Rho protein GTPases and their interactions with NF κ B: crossroads of inflammation and matrix biology. *Biosci Rep* **34**: e00115. doi:10.1042/BSR20140021
- Tong H, Faloutsos C, Pan J-Y. 2008. Random walk with restart: fast solutions and applications. *Knowl Inf Syst* **14**: 327–346. doi:10.1007/s10115-007-0094-2
- Türei D, Korcsmáros T, Saez-Rodriguez J. 2016. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. *Nat Methods* **13**: 966–967. doi:10.1038/nmeth.4077
- Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Ivanova O, Gábor A, Módos D, Korcsmáros T, Saez-Rodriguez J. 2021. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol Syst Biol* **17**: e9923. doi:10.15252/msb.20209923
- Van Tubergen EA, Banerjee R, Liu M, Vander Broek R, Light E, Kuo S, Feinberg SE, Willis AL, Wolf G, Carey T, et al. 2013. Inactivation or loss of TTP promotes invasion in head and neck cancer via transcript stabilization and secretion of MMP9, MMP2, and IL-6. *Clin Cancer Res* **19**: 1169–1179. doi:10.1158/1078-0432.CCR-12-2927

- Vogel C, Marcotte EM. 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet* **13**: 227–232. doi:10.1038/nrg3185
- Wasserman WW, Sandelin A. 2004. Applied bioinformatics for the identification of regulatory elements. *Nat Rev Genet* **5**: 276–287. doi:10.1038/nrg1315
- Wickham H. 2009. *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York. <http://ggplot2.org>.
- Wildwater M, Sander N, de Vreede G, van den Heuvel S. 2011. Cell shape and Wnt signaling redundantly control the division axis of *C. elegans* epithelial stem cells. *Development* **138**: 4375–4385. doi:10.1242/dev.066431
- Wozniak MA, Chen CS. 2009. Mechanotransduction in development: a growing role for contractility. *Nat Rev Mol Cell Biol* **10**: 34–43. doi:10.1038/nrm2592
- Wu X, Chen H, Parker B, Rubin E, Zhu T, Lee JS, Argani P, Sukumar S. 2006. HOXB7, a homeodomain protein, is overexpressed in breast cancer and confers epithelial–mesenchymal transition. *Cancer Res* **66**: 9527–9534. doi:10.1158/0008-5472.CAN-05-4470
- Wu P-H, Gilkes DM, Phillip JM, Narkar A, Cheng TW-T, Marchand J, Lee M-H, Li R, Wirtz D. 2020. Single-cell morphology encodes metastatic potential. *Sci Adv* **6**: eaaw6938. doi:10.1126/sciadv.aaw6938
- Yin P, Wang W, Zhang Z, Bai Y, Gao J, Zhao C. 2018. Wnt signaling in human and mouse breast cancer: focusing on Wnt ligands, receptors and antagonists. *Cancer Sci* **109**: 3368–3375. doi:10.1111/cas.13771
- Zhang Y, Chiu S, Liang X, Gao F, Zhang Z, Liao S, Liang Y, Chai Y-H, Low DJH, Tse H-F, et al. 2015. Rap1-mediated nuclear factor- κ B (NF- κ B) activity regulates the paracrine capacity of mesenchymal stem cells in heart repair following infarction. *Cell Death Discov* **1**: 15007. doi:10.1038/cddiscovery.2015.7
- Zhang Y-L, Wang R-C, Cheng K, Ring BZ, Su L. 2017. Roles of Rap1 signaling in tumor cell migration and invasion. *Cancer Biol Med* **14**: 90–99. doi:10.20892/j.issn.2095-3941.2016.0086
- Zheng B, Han M, Bernier M, Wen J. 2009. Nuclear actin and actin-binding proteins in the regulation of transcription and gene expression. *FEBS J* **276**: 2669–2685. doi:10.1111/j.1742-4658.2009.06986.x
- Zou A, Lehn S, Magee N, Zhang Y. 2015. New insights into orphan nuclear receptor SHP in liver cancer. *Nucl Recept Res* **2**: 101162. doi:10.11131/2015/101162

Received July 29, 2021; accepted in revised form February 17, 2022.