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

Charlie George Barker[1], Eirini Petsalaki[1*], Girolamo Giudice[1], Emmanuel Nsa Ekpenyong[1], Chris Bakal[2], Evangelia Petsalaki[1*]

[1]European Molecular Biology Laboratory-European Bioinformatics Institute, Hinxton CB10 1SD, UK

[2]Institute of Cancer Research, 237 Fulham Road, London, SW3 6JB, UK

[*]correspondence should be addressed to Evangelia Petsalaki, petsalaki@ebi.ac.uk

Abstract

The morphology of breast cancer cells is often used as an indicator of tumour 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 metastase. However, the exact regulatory systems that define these changes remain poorly characterised. In this study, we employ a network-systems approach to integrate imaging data and RNA-seq expression data. By constructing a cell-shape signaling network from shape correlated gene expression modules and their upstream regulators, we found central roles for development pathways such as Wnt and Notch as well as evidence for the fine control of NFkB signaling by numerous kinase and transcriptional regulators. Further analysis of our network implicates the small GTPase, Rap1 as a potential mediator between the sensing of mechanical stimuli and regulation of NFkB activity. Overall, our analysis provides mechanistic information on the interplay between cell signaling, gene regulation and cell morphology and our approach is generalisable to other cell phenotypes.

Introduction

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 tumour progression through the modulation of migration, invasion and overall tissue structure (Baskaran et al., 2020; Krakhmal et al., 2015). The unique mechanical properties of the tumour tissue, primarily driven by changes in cell shape and the extra-cellular matrix, are hypothesised 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 tumour organisation have been found to be a factor in modulating the intra-cellular signaling state through pathways able to integrate mechanical stimuli from the extra-cellular environment (Miralles et al., 2003; Olson and Nordheim, 2010; Orsulic et al., 1999; Zheng et al., 2009). 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 worse prognosis in patients (Roche, 2018).

Breast cancer is the most common cancer among women, and in most cases treatable with a survival rate of 99% among patients with a locally contained tumour. However, among those patients presenting with a metastatic tumour this rate drops to 27% (Siegel et al., 2019). During the development of breast cancer tumours, cells undergo progressive transcriptional and morphological changes that can

ultimately lead towards EMT and subsequent metastasis (Feng et al., 2018; Lee et al., 2015; 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 tumour 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 employing these morphology determinant pathways as potential therapeutic targets and predictors of prognosis.

Signaling and transcriptomic programs are known to be modulated by external cues in the contexts of embryonic development (Wozniak and Chen, 2009), stem-cell maintenance (Bergert et al., 2020; De Belly et al., 2020) and angiogenesis (Chatterjee, 2018). Numerous studies have flagged NFkB as a focal point for mechano-transductive pathways in various contents (Cowell et al., 2009; Ishihara et al., 2013; Shrum et al., 2009; Tong and Tergaonkar, 2014), but gaps in our knowledge remain as to how these pathways may interact and affect breast cancer development. Sero and colleages studied the link between cell shape in breast cancer and NFkB activation by combining high-throughput image analysis of breast cancer cell lines with network modelling (Sero et al., 2015). They found a relationship between cell shape, mechanical stimuli and cellular responses to NFkB and hypothesised that this generated a negative feedback loop, where 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 NFkB 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 authors rely on a list of pre-selected transcription factors 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; Horton et al., 2016; Robertson et al., 2015; Rolfe et al., 2014). Furthermore, how exactly extra-cellular mechanical cues are 'sensed' by the cell and passed on to NFkB in breast cancer is not 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 identify an unbiased, data-derived cellular signaling network, specific to the regulation of cell shape beyond NFkB, by considering functional co-expression modules and cell signaling processes rather than individual genes. To this end, 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 (Figure 1A). By organizing expression data into context-specific modules we leverage the transcriptome's propensity to be regulated in regulons, thereby aiding the inference of signaling activity. This, combined with the use of feature-correlated modules, allows for the analysis of complex phenotypes, defined by multiple features. The resulting signaling network was validated using treatments with kinase inhibitors (Subramanian et al., 2017) and revealed further roles for NFkB in modulating mechanotransduction in response to morphological changes in breast cancer.

Results

Identification of gene co-expression modules correlated with cell shape features

We first sought to identify gene co-expression modules that are relevant to the regulation of cell shape. To this end, we used weighted correlation network analysis (WGCNA) (Langfelder and Horvath, 2008) on RNA-seq expression data from 13 breast cancer and one non-tumorigenic epithelial breast cell lines to identify gene co-expression modules correlated with 10 specific cell shape variables (Sailem and Bakal, 2017) (Methods). These described the size, perimeter and texture of the cell and the nucleus (n = 75,653). Of 102 gene co-expression modules (Figure S1A), 34 were significantly correlated (P<0.05; Student's T-

test, Pearson Correlation) with one of 8 cell shape features (Figure 1B).

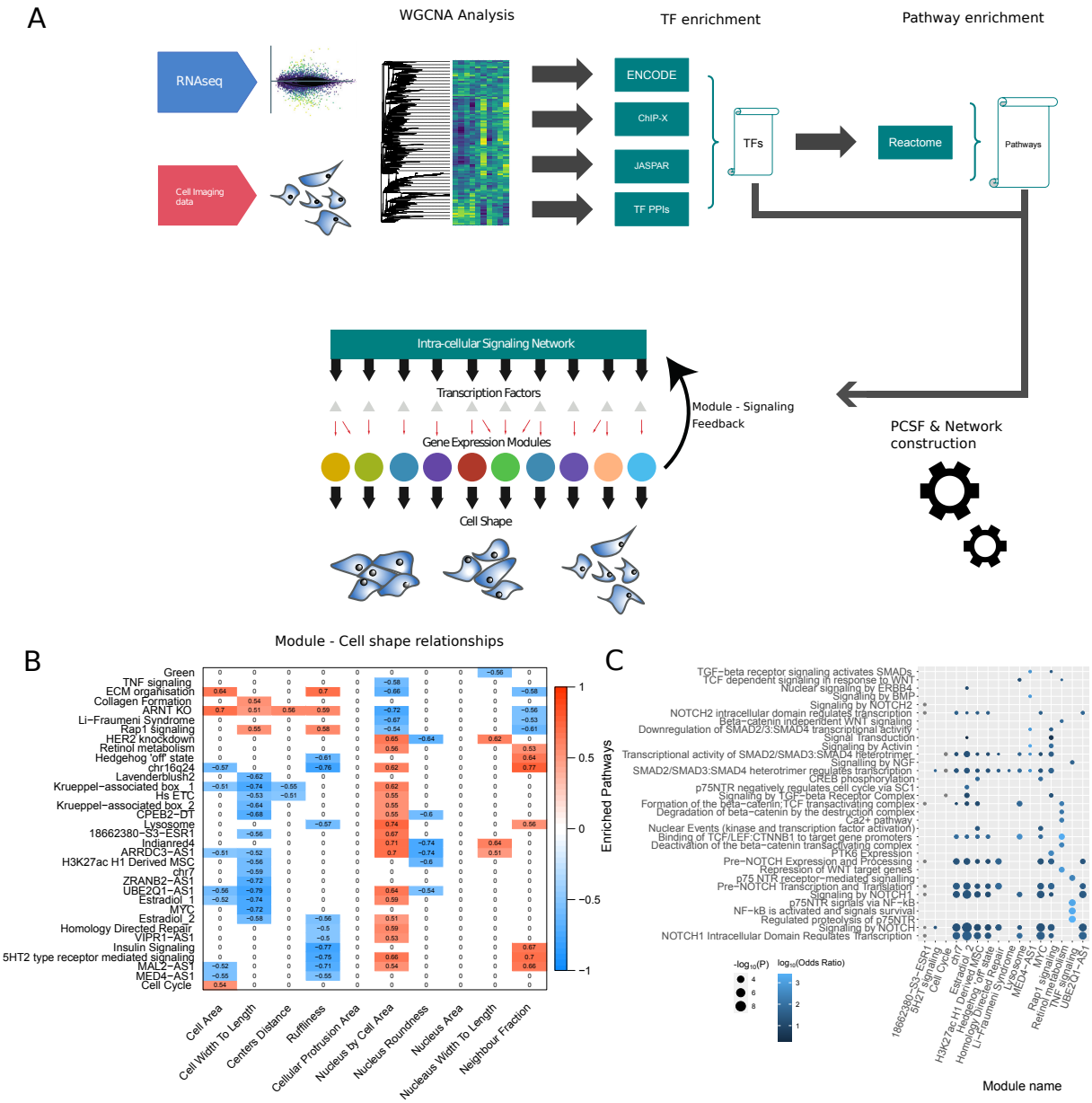

Figure 1: **A**. Schematic illustrating the steps involved in phenotype-specific network construction. In the first panel, gene expression modules are identified by integrating cell shape variables (derived from imaging data) with RNAseq data from breast cancer cell lines. These gene expression modules 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. This constructed network is represented by the schematic of the different components that form our phenotype specific network. **B**. Heatmap of significantly correlated gene expression eigengenes with cell shape features. Non-significant interactions were set to 0 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 supplementary table 1) 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 in supplementary table 3). 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 -log10(P) and the colour indicates a log 10 transformation of the odds ratio.

We used EnrichR 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, while the 'Insulin Signaling' module is also enriched for cell-cell communication and the 'ECM organisation' module is also enriched in terms such as axon guidance and EPH-Ephrin signaling (supplementary table 1). Modules that are most correlated with all features are the 'ARNT KO' module, 'ARRDC3-AS1' module and the 'ECM organisation' module (Figure S1B).

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

To link these expression modules to the intra-cellular 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 (Han et al., 2018), to be significantly enriched (P<0.1; Fisher's exact test) in our modules (supplementary table 2). 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), and 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 gene expression modules. For this analysis, we also include ENCODE and ChEA Consensus TFs from ChIP-X (Lachmann et al., 2010), DNA binding preferences from JASPAR (Stormo, 2013; Wasserman and Sandelin, 2004), 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 (supplementary table 3) 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 many identified gene expression modules were regulated by common signaling pathways, with 6 modules sharing pathways associated with downstream signaling and regulation of Notch (Figure 1C).

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 8 morphological features including area, ruffliness, protrusion area and neighbour frequency and performed differential expression analysis between the identified clusters (Figure 2A, Figure S2A). Cluster A is more heterogeneous in its morphology, containing the non-tumorigenic 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 with the exception of HCC1954, which was clustered morphologically with luminal subtypes while being characterised 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 ± SD]), higher ruffliness (0.235 ± 0.12) and lower neighbour fraction (0.258 ± 0.22). The luminal-like cluster had a higher nuclear/cytoplasmic area (0.186 ± 0.1; P<0.001), lower ruffliness (0.213 ± 0.14; P<0.001), and a higher neighbour fraction (0.338 ± 0.26; P<0.001, One-way ANOVA; Tukey HS, n = 75,653). The neighbour fraction feature corresponds to the fraction of the cell membrane that is in contact with neighbouring cells. The lower number of cell-cell contacts in basal-like breast cancer cell lines are indicative of more mesenchymal features associated with worse prognosis due to metastasis. Increased cell-cell contacts in both the luminal-like cluster and the more heterogeneous cluster A correspond to 'cobblestone' epithelial morphology. Interestingly, these groups closely aligned with the expression of the cell adhesion protein, N-cadherin (Figure 2A), the expression for which is closely associated with a migratory and metastatic phenotype (Shih and Yamada, 2012).

Using the identified groups of cell lines in the previous step, differential expression analysis and transcription factor activity analysis was used to study gene regulation signatures specific to cell line morphological clusters. The results are shown in Figure 2B-C, with gene set enrichment analysis showing

upregulation of genes involved in the extra-cellular matrix, collagens, integrins and angiogenesis in the basal-like cluster (Figure S2B). Significantly enriched terms (P<0.05) in downregulated genes include 'fatty acid and beta-oxidation' and 'ERBB network pathway'. In the genes upregulated in the luminal-like cluster, we observed enrichment of terms such as 'hallmark-oxidative phosphorylation'. Downregulated genes were enriched in 'integrin-1 pathway', 'core matrisome' and genes linked to 'hallmark epithelial-mesenchymal transition and migration' (Figure S2C). 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' (Figure S2D).

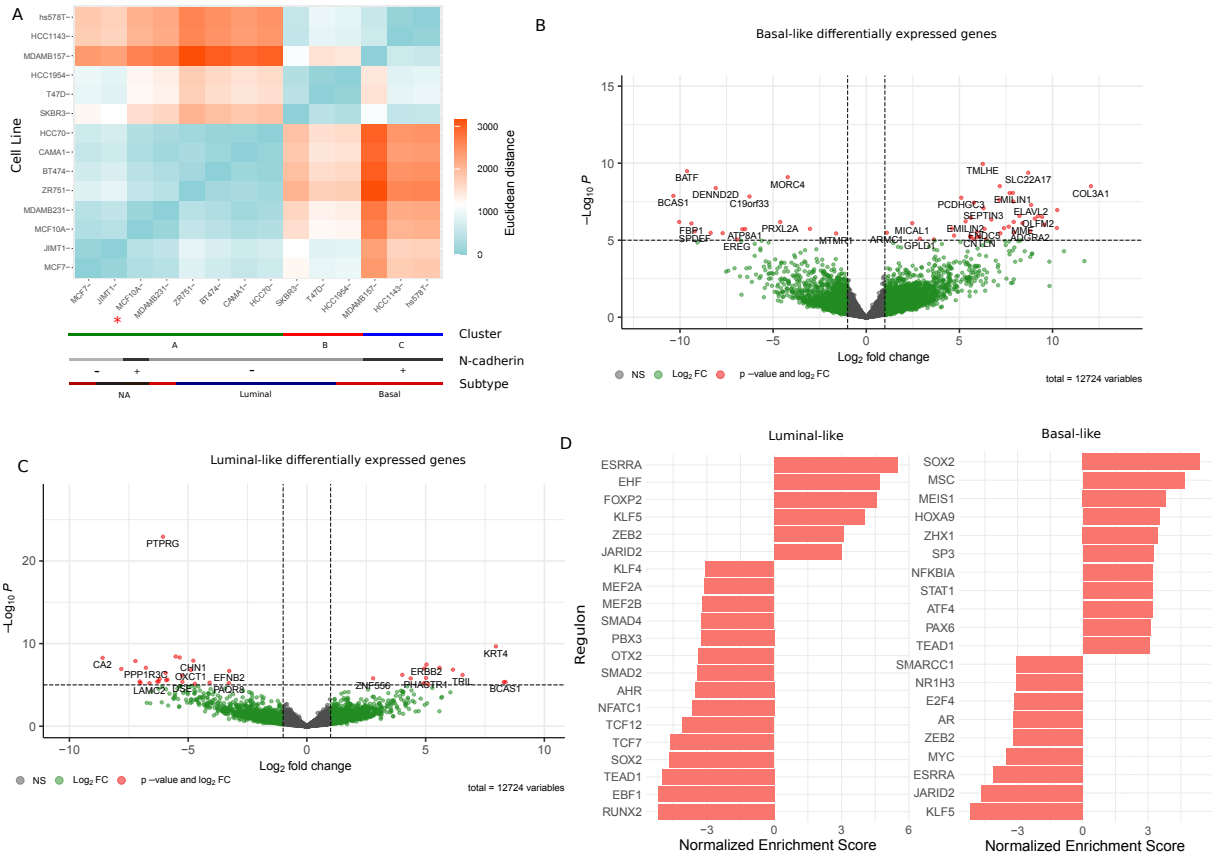

Figure 2: **A.** Heatmap of Euclidean distance between cell lines for shape features to illustrate clusters arising from k-means method. The coloured lines on the bottom show the assigned cluster and the cadherin expression and assigned canonical cancer subtype. **B & C.** Volcano plots showing significantly differentially expressed genes for basal-like and luminal-like morphological clusters. **D.** TF activity for morphological clusters C and B (basal-like and luminal-like respectively) as derived from RNAseq.

We also calculated the differential expression for the WGCNA gene expression modules and found distinct patterns of expression between luminal-like and basal-like clusters of cell lines (Figure S2E). Among these, the Rap1 Signaling module is upregulated in basal-like clusters and downregulated in luminal-like clusters. This is consistent with the fact that this gene expression module is negatively correlated with neighbour fraction, a feature that is observed to decrease in mesenchymal-like cell shapes. 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 NFkB) and ECM organisation module (enriched in focal adhesion proteins - Figure S2E).

To link the observed gene expression differences to cell signaling we used the tool DOROTHEA (Garcia-Alonso et al., 2019) to calculate transcription factor activities, as their modulation is one of the main results of cell signaling processes. We corroborated that the heterogenous B/L group had significantly activated Myc levels. In the luminal-like cluster, ESRRA (estrogen-related receptor α) is the most significantly overrepresented regulome, followed by EHF, KLF5 and ZEB2. Under-represented regulomes

include KLF4, SMAD4, SMAD2, SOX2 and RUNX2. For the basal-like cluster, the regulome with the highest normalised enrichment score is SOX2, as well as Musculin and HOXA9. Downregulated regulomes include ZEB2, Myc, ESRRA and KLF5 (Figure 2D).

Assembly of a data-driven cell shape regulatory network

To link the results of the above analysis into one coherent picture of the signaling regulation of cell shape in breast cancer, we used the Prize Collecting Steiner Forest (PCSF) algorithm (Akhmedov et al., 2017). This is an approach that aims to maximise 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 transcription factors 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.

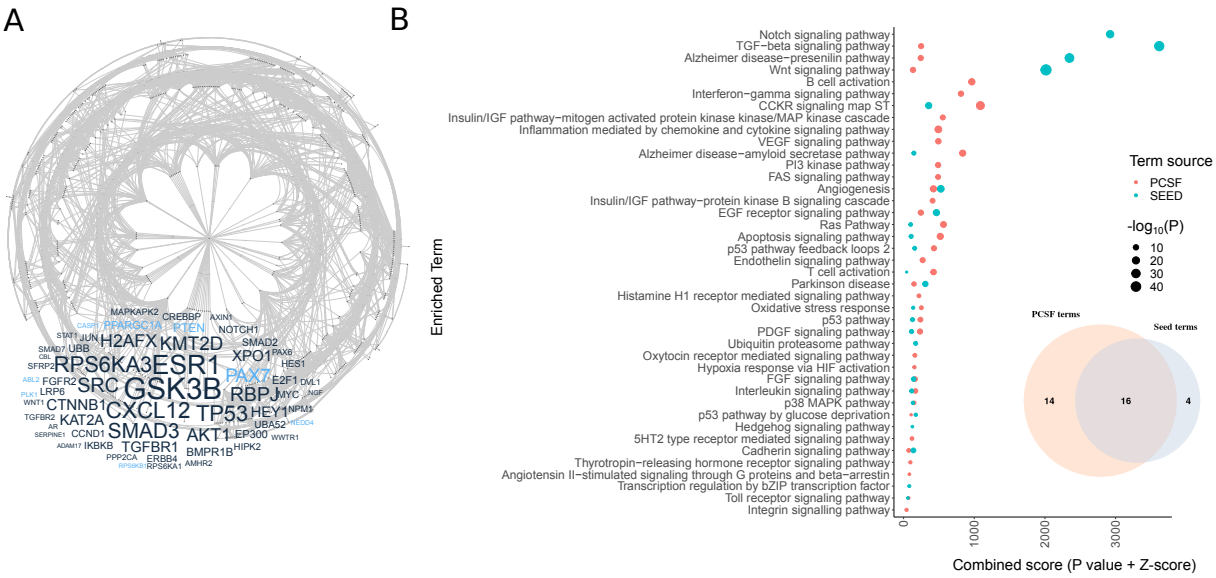

Figure 3: **A.** Network derived from integrating enriched pathways and transcription factor regulons within cell shape correlated gene expression modules. Below is a word cloud illustrating the top 50 nodes with highest betweenness centrality in the network, with those nodes belonging to the original seeds coloured black, and those genes included by the PCSF step coloured light blue. **B.** Gene set enrichment (Panther DB 2016) of original seed nodes and the nodes included by PCSF with the dot size indicating the level of significance (-Log10P) of the term enrichment. Inset is a Venn diagram showing the overlap in enriched terms between the seed nodes and the terms included by PCSF.

The resulting network of 691 nodes included 97.11% of the genes identified by our analysis (Figure 3A). The new proteins that were included by the PCSF algorithm to maximise prize collection showed gene set enrichment of common terms relative to the original prizes (including Wnt, EGF, Angiogenesis, Ras, Cadherin and TGF-beta pathways), but also included are some new terms (VEGF, Integrin and Endothelin pathways) (P<0.001; Figure 3B). The WGCNA expression modules were included in the network as super-nodes linked with the network through edges describing significant correlations between transcription factors regulons (as defined by TRRUST) and the gene content of the modules.

Studying the network properties of our PCSF-derived regulatory network we find that the degree distribution is typical for a biological network, with the degree exponent gamma lying between those of the databases Reactome and BioGrid, implying a scale-free degree distribution (Figures S3A-C). The proteins in the network can be ranked by betweenness centrality to disseminate them based on network impor-

tance. Nodes with high centrality lie between many paths and can control information flow. Proteins with the highest centrality are primarily prizes (GSK3b, ESR1, p53, SMAD3 - Figure S3D) indicating that the PCSF solution was not achieved by the inclusion of new hub proteins that are not of interest to our analysis. That being said, a small minority of high centrality nodes were not in the original prizes, including PAX7, PTEN and PPARGC1a.

Heat diffusion of functionally 'hot' TFs reveals differentially activated processes in the cell shape regulatory network

As transcription factor activity remains the most reliable indicator of signaling that can be extracted from transcriptomics data (Szalai and Saez-Rodriguez, 2020), we applied heat diffusion on the network to identify regions of the network on which differentially regulated transcription factors have an effect. The algorithm Random Walk with Restart (RWR (Tong et al., 2008)) was used to diffuse from activated and inactivated transcription factors in our network reflected by the normalized enrichment scores of transcription factors identified by DOROTHEA (Garcia-Alonso et al., 2019) (Methods).

The hottest super-node in both luminal and basal diffusions was the gene expression module, Rap1 Signaling (Figure 4A and B), a module which is correlated with several cell shape variables (neighbour frequency, ruffliness, 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 S4A-B) we can see that the source of this module's heat is from the transcription factors JARID2 and RUNX2 in luminal-like cell lines, and JARID2 for basal-like.

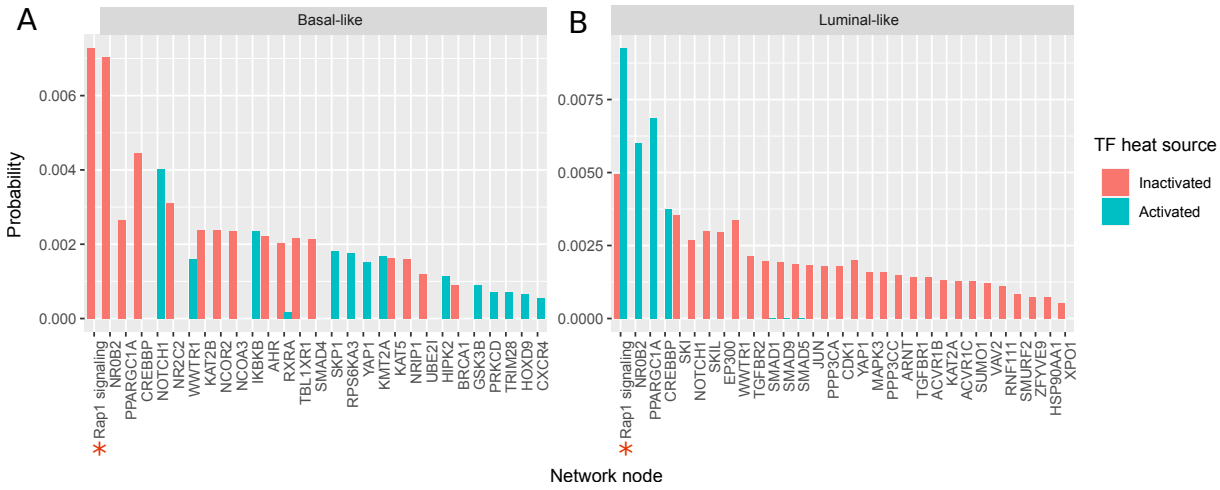

Figure 4: Bar Plot showing heat diffusion in predicted cell shape network from activated and inactivated transcription factors in basal-like cell lines **(A)** and luminal **(B)**. The y axis is a steady state probability (or the 'heat' of the nodes in the network after the diffusion) over the graph imposed by the starting seeds, ordered by size. Red bars represent heat diffused from transcription factor seeds that are predicted to be in-activated, and blue bars show heat diffused from transcription factor seeds that are predicted to be activated. Red stars along the x axis indicate supernodes that represent gene expression modules.

Specific proteins that were hot after performing the heat diffusion in basal-like cell lines include the orphan nuclear receptor NR0B2. Individual RWR found that heat was diffused to this node from 3 seed transcription factors: AR, ESRRA and NR1H3. Other hot proteins were SMAD4, which is regulated by TGF-beta, IKBKB, which is an activator of NFkB and YAP1. GO term enrichment of 'hot' proteins in basal-like cell lines found significantly enriched terms such as Wnt signaling pathway, Angiogenesis related signaling and TGF-beta signaling pathway (P<0.01; Figure S4C). For luminal-like cell lines, NR0B2 is also significantly hot (as a result of ESRRA activity) as well as transcriptional co-activator PPARGC1A. GO terms enriched in the luminal heat diffusion include many shared with basal (TGF-b, Wnt, CCKR signaling and angiogenesis), but also include new terms such as HIF activation, p53 pathway and Toll receptor

signaling pathway (P<0.01; Figure S4D).

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

To validate our network, we used an independent dataset 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 dataset (Subramanian et al., 2017). Here, they measured morphological changes in the breast cancer cell line HS578T in response to various small molecule kinase inhibitors using high through-put imaging techniques (Hamilton et al., 2007). 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. Figure 5A illustrates 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. 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, and 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 (SKBR3, MCF7 and non-tumorigenic mammary cell line MCF10A) with results with limited statistical significance (S5A-C) suggesting a bias in our network towards certain breast cancer cell lines (see Discussion).

Interestingly, there is greater variance in the effect size for kinase inhibitors targeting proteins contained within the predicted regulatory network than those outside. We hypothesised 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 5B 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 gene expression modules 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 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. For variables that were not correlated to any gene expression module, we see visibly lower correlation coefficients and insignificant associations (Spearman 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.

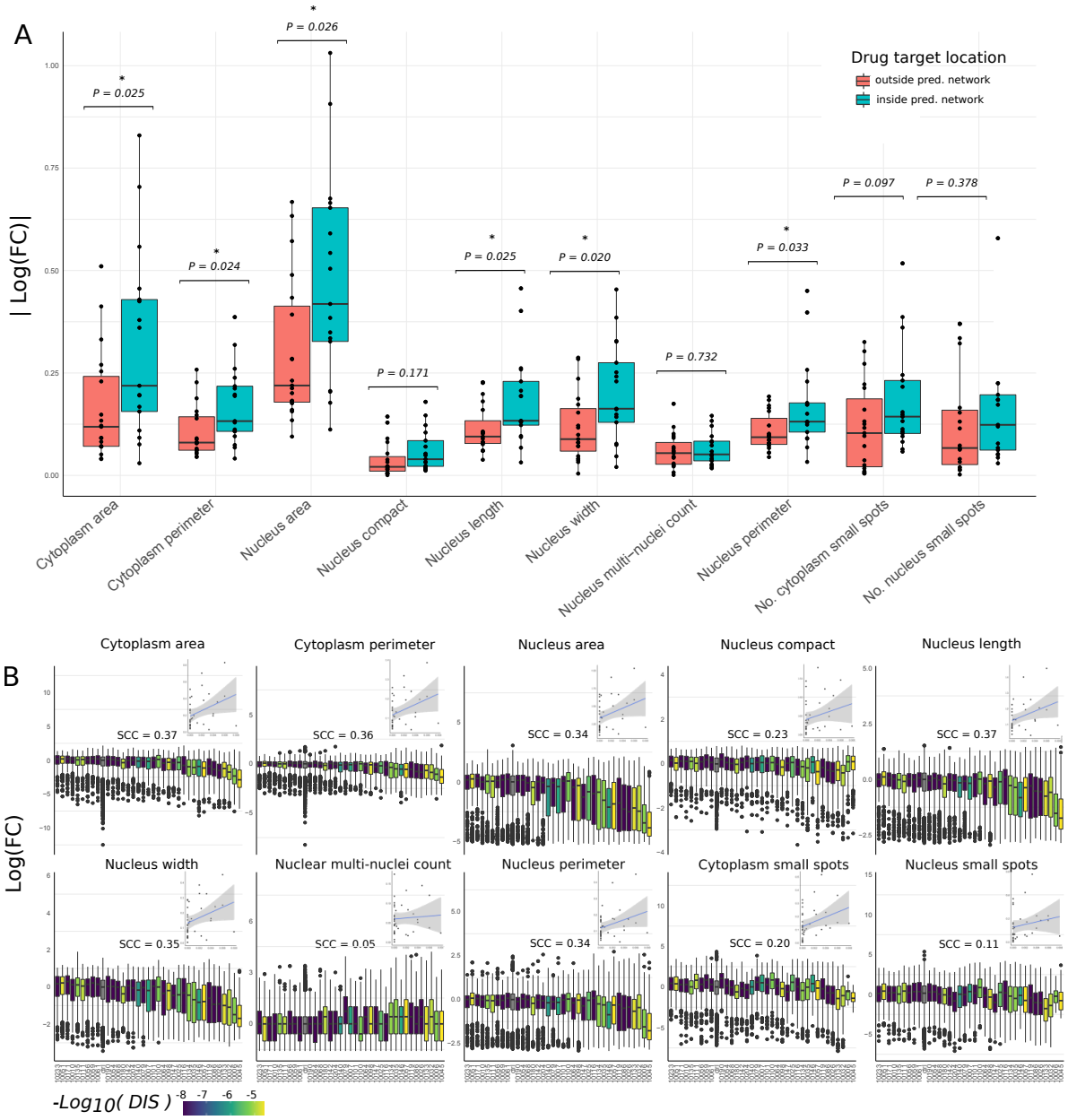

Figure 5: **A.** Box plots showing the absolute log(10) fold changes after treatment with a drug relative to a control for each cell shape variable. The drugs are grouped depending on whether they target kinases within the predicted regulatory network (blue) and those targeting other kinases not predicted to be associated with cell shape (red). P values (Welch Two Sample t-test) are showing with stars 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 LINCs ID shown on the x axis) and coloured based on the drug influence score (DIS) and each data-point represents a single cell. Inset 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 correlation coefficients are shown above the inset plots.

Discussion

We present a method that uses transcriptomics and phenotypic data to derive a concise sub-network describing the signaling involved in the regulation of cell shape. This analysis recovered known processes like 'adherens junction proteins', 'cadherin' and 'integrin' as well as pathways responsible for the regula-

tion of cell shape in development, such as Wnt (Kadzik et al., 2014; Wildwater et al., 2011), TGF-beta (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). Cadherin is a cell-cell adhesion protein whose various subtypes contribute to distinct cell morphologies and invasion modes (Eslami Amirabadi et al., 2019). Integrins are transmembrane receptors that regulate cell extra-cellular matrix (ECM) adhesion (Filer and Buckley, 2013) as well as other signaling processes affecting cancer prognosis (Taherian et al., 2011).

Importantly, this analysis also sheds light on processes with less characterised associations with cell shape in cancer. We found that the Rap1 signaling module, for example, is significantly correlated with cell shape, and the most differentially expressed module between Luminal-like and Basal-like cell line clusters. We found that it was upregulated in basal-like cell lines while downregulated in luminal, consistent with its negative correlation with neighbour fraction; a cell shape feature most contributing to the 'cobblestone' like features of an epithelial and non-metastatic cell type. This gene expression module was also significantly hot in the RWR analysis, being at the network confluence of 2 activated transcription factors (JARID2 and RUNX2). Heat diffusion by this method is undirected which is preferable as it means one can use RWR to inform on the downstream effects of activated seeds, but also on possible upstream activity that might be driving seed activity. Rap1 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 (Zhang et al., 2017). It has also been illustrated that Rap1 is able to sense mechanical stress acting on a cell and modulate signaling accordingly in other tissues (Ke et al., 2019). Specifically, Rap1 has been shown to modulate and activate NFkB activity in response to TNF- α stimulation in mesenchymal stem cells (Zhang et al., 2015) and modulate migration and adhesion (Yu et al., 2016). Rap1 is able to regulate IKKs (IκB kinases) in a spatio-temporal manner (Ohba et al., 2003), and is crucial for IkBK to be able to phosphorylate NFkB subunit p65 to make it competent (Teo et al., 2010).

Previous studies have established a connection between the NFkB signaling pathway and regulation of cell shape in breast cancer (Sailem and Bakal, 2017; Sero et al., 2015). Our findings also illustrate the significance of this pathway in the regulation of cell morphology, with multiple NFkB regulators and transcriptional co-activators being flagged in our results. Some morphology correlated gene expression modules were significantly differentially expressed between cell shape subtypes with the ARNT KO module being significantly upregulated in basal-like cell shapes relative to luminal. We also found this gene expression module to have the highest total correlation with all of the morphological features (Figure S1B), indicating a strong association with cell shape. By studying terms enriched in this module from the EnrichR library, we find both 'TNF- α signaling via NFkB' to be enriched as well as genes downregulated during AhR nuclear translocator (ARNT) shRNA KO. Signaling by TNF- α is able to activate NFkB, a transcription factor known to control the expression of many EMT related genes (Pires et al., 2017) which has shown to be more sensitive to TNF- α stimulation in mesenchymal-like cellular morphologies than epithelial. This was hypothesised to generate a negative feedback which reinforces a metastatic phenotype of breast cancer cells (Sero et al., 2015). Here we observe also that an ARNT KO/TNF- α module is upregulated in basal-like cell lines, consistent with these findings. ARNT is a protein shown to be involved in regulating tumour 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 NFkB signaling with the activated form possibly interfering with the action of activated p65 (Øvrevik et al., 2014). Our findings that the upregulation of a gene expression module that is associated with ARNT knockdown further gives credence to NFkB being positively regulated in mesenchymal-like cell morphologies.

Through heat diffusion of activated transcription factors, we also observe IKBKB (an upstream activator of NFkB) to be hot in basal-like cell lines, consistent with our knowledge that it is activated by Rap1 signaling (Teo et al., 2010). Additionally, several NFkB transcriptional coactivators are significantly 'hot' after RWR simulations from activated transcription factors. The orphan nuclear receptor NR0B2 was also identified by the RWR simulation, being at the network confluence of active transcription factors ESRRA, NR1H3 and AR. Despite being poorly studied, it has been observed to be a coregulator of NFkB (Kim et al., 2001) and has already been identified as a critical regulator of cancer (especially breast cancer (Zhang et al., 2011)). A similar story is told with PPARGC1A, an interacting partner of NFkB subunit p65 (Alvarez-Guardia et al., 2010) as well as a key metabolic re-programmer which we found to be hot in basal-like cell lines through the differentially activated transcription factor, ESRRA. These findings provide evidence that NFkB is modulated by both phosphorylation, spatial-temporal location and transcriptional co-activation

for the cell to sense mechanical stimuli and modulate cell morphology and signaling in breast cancer.

Another term repeatedly enriched in various steps was 'angiogenesis' and its regulator, VEGF. It is well characterised how the development of a vasculature in a tumour is important for cancer survival and development (Madu et al., 2020) as it is required for the sufficient supply of oxygen and nutrients for continued growth. However, here we find that VEGF and other regulators of angiogenesis are implicated in the regulation of cellular morphology, since they are included in our network through enrichment in both the Rap1 module and within proteins added to our regulatory network when we perform PCSF. Other studies have observed VEGF expression to be correlated with breast cancer cell migration (Su et al., 2006) and the knockout of its gene can modulate cell shape (Kiso et al., 2018). The crosstalk and combination of VEGF's function of modulating cell shape to facilitate metastasis as well as encouraging vascular growth could be coordinated to maximise the likelihood of an invasive tumour entering circulation, however more experimental work is required to delineate this relationship. Other modules are also shown to discriminate between luminal and basal cell morphologies, however, interestingly many could not be annotated with useful terms (only poorly annotated lncRNAs and chromosomal locations). These modules include some of those with greatest combined correlation with cell shape features (Figure S1B). This illustrates that many processes and expression patterns associated with the regulation of cell shape are of novel function and are currently poorly understood.

A draw-back of the approach we present here is the apparent bias of our network to HS578T as opposed to MCF7, MCF10A and SKBR3 upon validation with kinase inhibitors. HS578T is the most morphologically distinct (Figure 2A) of all the breast cancer cell lines available for our validation, belonging to the basal-like cluster of cell shapes while SKBR3 belongs to luminal-like and MCF7 and MCF10 belong to the heterogeneous cluster. This illustrates that our methodology exhibits a bias towards the extremities of our observed phenotypes (in this case, the more cell-lines with a more mesenchymal appearance). This could be due to a number of reasons. More highly activated pathways in extremes of phenotypes could cause an exaggerated effect upon disruption by kinase inhibition relative to those cell lines whose phenotypes are more typical. Another explanation is that the 280 differentially expressed genes identified in basal-like cell lines contribute more than the 170 in luminal-like to our network construction. The effect of this would be that the resulting network is biased towards describing regulation of cell shape in basal breast cancer relative to other subtypes. While this should be taken into consideration when interpreting the network's results, it still provides valuable insight into the regulation of the cell shape of the breast cancer cell lines that are more susceptible to metastasis, which is more interesting in terms of disease prognosis than other morphologies. Another potential drawback is the fact that the RNA-seq and morphological variables were obtained in 2D culture and they may differ in more physiological conditions. Nevertheless, it has been shown that cell shape still encodes metastatic potential in 2D culture (Wu et al., 2020), providing evidence for the utility of our method and findings.

Aside from the biological findings of this study, we illustrate a method 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 ((Ben-Hamo et al., 2014, p.; Guan et al., 2012; Padi and Quackenbush, 2018; Soul et al., 2015)) our approach infers transcription factor activities from the expression data and uses these as an anchor to infer upstream signaling networks relevant to the regulation of our phenotypes. Transcription factor activities can represent the outcome of a signal transduction process compared to 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; Wang et al., 2019). Here, using context-specific gene expression modules, we produced a network connecting the genes of interest from diverse analyses and used a heat-diffusion algorithm to further focus on signaling proteins of novel interest. Our method considers the interplay between the predicted function of gene expression modules themselves and the TFs that regulate them, in one contiguous network that provides a sketch of the wide variety of processes that regulate cell shape in breast cancer. By focusing on context-specific expression modules and the transcription factors that regulate them, we make the fullest use of gene expression data in understanding the links between cell morphology and intra-cellular signaling. By combining this with differential activated transcription factors between distinctive cell shapes we can reveal a clearer picture of how signaling regulates cell morphol-

ogy 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. In this case only heat diffusion and centrality measures were used on the derived network, but alternatively the network could be used to study signal-flow or be used as a basis for constructing more detailed and mechanistic models of cell signaling, able to predict any measured phenotype used in the WGCNA analysis. Thus, our method represents a data-driven, network-based approach compatible with many different multi-scale phenotypes that are driven by intracellular signaling.

Overall, our unbiased network-based method highlighted potential 'missing links' between sensing extracellular cues and transcriptional programmes 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.

Materials and Methods

Key Resources Table

Software/Databases	Description	Reference
WGCNA	Gene expression clustering	Langfelder and Horvath, 2008
Biomart	Mapping biological identifiers	Kinsella et al., 2011
EnrichR	Gene list enrichment analysis tools	Chen et al., 2013; Kuleshov et al., 2016
DESeq2	Statistical analysis of expression data	Love et al., 2014
FGSEA	Fast gene set enrichment analysis	Korotkevich et al., 2019
DOROTHEA	Regulon enrichment tool	Holland et al., 2020
PCSF	Subgraph identification, given prize nodes	Akhmedov et al., 2016
paxtoolsr	Tool to handle Bio-Pax pathways	Luna et al., 2016
OmnipathR	Tool to access Omnipath database	Türei et al., 2020
diffuseR	Heat diffusion algorithm	Tong et al., 2008
Expression Atlas	Gene expression data repository	Papatheodorou et al., 2020
Reactome	Pathway database	Jassal et al., 2020
TRRUST	Transcription factor database	Han et al., 2018
KEGG	Pathway database	Kanehisa et al.,2016
Panther	Pathway database	Mi et al., 2013
Wikipathways	Pathway database	Martens et al., 2021
MSigDB	Gene set database	Liberzon et al., 2015
Pathway Commons	Pathway database	Luna et al., 2016
HMS-LINCS	Drug Perturbation database	Stathias et al., 2020

Contact for Resource Sharing

petsalaki@ebi.ac.uk

Method Details

WGCNA analysis

Using Weighted correlation network analysis, we performed co-expression module identification using the R package WGCNA (Langfelder and Horvath, 2008). We used RNA-seq data from Expression Atlas

(in FPKM - E-MTAB-2770 and E-MTAB-2706) acquired from commonly used cancer cell lines of various cancer types (Papatheodorou et al., 2020). We collated 13 breast cancer and 1 non-tumorigenic cell line for which imaging data was available (BT474, CAMA1, T47D, ZR75.1, SKBR3, MCF7, HCC1143, HCC1954, HCC70, hs578T, JIMT1, MCF10A, MDAMB157 and MDAMB231 (Sero et al., 2015)). Using Ensembl-Biomart, we filtered genes to only include protein-coding genes (Kinsella et al., 2011) and genes whose FPKM was greater than 1, leaving a total of 15,304 genes.

We created a signed, weighted adjacency matrix using log2 transformed gene expression values and a soft threshold power (beta) of 9. We translated this adjacency matrix (defined by Eq.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).

$$Eq.1 \qquad a_{ij} = |(1 + cor(x_i, x_j))/2$$

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 (Sero et al., 2015). We correlated these features with module eigengenes using Pearson correlation and we tested these values for significance by calculating Student asymptotic p-values for given correlations. 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 (Chen et al., 2013; Kuleshov et al., 2016), and the signaling database Reactome (Jassal et al., 2020). Using the database TRRUST (Han et al., 2018), we identified TF regulons that significantly overlap (Exact Fisher's test, P<0.1) with the gene expression module contents. This was done separately for inhibitory and activatory expression regulons for each transcription factor, 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 (supplementary Table 1). As some clusters were very obscure, we utilized 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. (Kanehisa et al., 2016; Martens et al., 2021; Mi et al., 2013).

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, Figure 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 more than 0.5 counts per million in at least 8 cell lines were included. We calculated Log2 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., 2019), we performed gene set enrichment analysis of the differentially regulated proteins using the complete pathways gene set (Release 01 April 2020) from MSigDB (Liberzon et al., 2015) and the WGCNA gene expression modules identified in previous analysis. We calculated transcription factor regulon enrichment using the software DOROTHEA (Holland et al., 2020).

Network Generation

Using a Prize Collecting Steiner Forest (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 transcription factors significantly regulating the WGCNA modules using TRRUST (p<0.1), the differentially activated transcription factors identified by DOROTHEA (p<0.1), and the signaling proteins included in the REACTOME pathways that were enriched in transcription factors 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 (Stormo, 2013; Wasserman and Sandelin, 2004), 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).

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 1 and 0, 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 (Türei et al., 2016), extracted using the R package OmnipathR (Türei et al., 2020). 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-penalisation parameter which we set to 0.005.

We included the WGCNA modules themselves as super-nodes in the network, by adding incoming edges from the transcription factors contained within the regulatory network whose regulomes (as described in TRRUST (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 where there was significant correlation (|PCC| >0.5 & 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 gene expression module 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 (Chen et al., 2013).

Heat diffusion of functionally 'hot' TFs

We examined the potential effect of significantly activated (FDR <0.05), and deactivated TFs in different cell line clusters using heat diffusion in our generated network. We replaced edge weight with Resnik 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 1 and -1.

We used the differentially activated transcription factors identified using DOROTHEA (P<0.05) as seeds for a Random Walk with Restart (RWR) algorithm using the R package diffuseR (Tong et al., 2008). We judged a node to be significantly hot if its affinity score (or 'heat') relative to the inputted seeds was greater than the same node's affinity score with a random walk simulation performed with randomised seeds. We performed this randomised simulation 10,000 times, from which a p-value was determined to judge significance (P<0.1). We performed this heat diffusion by RWR for both luminal-like and basal-like morphological clusters on significantly activated and deactivated transcription factors separately, in addition to simulations where each seed was considered in isolation. We implemented these simulations with a restart probability of 0.95 and performed GO term enrichment on significantly hot nodes using EnrichR (Kuleshov et al., 2016) and the signaling database Panther (Mi et al., 2013).

Breast cancer cell morphology following kinase inhibitor treatment

We used small molecule kinase inhibition data from Harvard Medical School (HMS) Library of Integrated Network-based Cellular Signatures (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/`). This dataset is derived from the treatment of 6 cell lines with a panel of 105 small molecule kinase inhibitors. They measured textural and morphological variables following treatment by high-throughput image analysis (Hamilton et al., 2007; Haralick et al., 1973). We combined this assay with another dataset 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/). Using this dataset, we filtered for only molecule-binding target pairs with a binding class of 1 (representing a Kd <100nM affinity). Further to this, we removed molecules which had more than 5 targets with a Kd of 100nM. 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 10um drug treatment relative to the mean of the control set (untreated and DMSO treated cells). We did not present features which were functions of other features (such as ratios) as these were redundant. Correlations analysis was performed on the mean log fold change for a drug subtracted from the log fold change

The morphological data in the kinase inhibition screen was measured using two dyes (DRAQ5 and TMRE), the intensity of which we used to normalise 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 normalised 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 Target Affinity Spectrum 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:

$$Eq.2 \qquad \sum_{x=K\cap N} = PR(x) \times \frac{K \cap N}{min(|K|, |N|)}$$

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 having 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.

Quantification and Statistical Analysis

Statistical tests were performed in base R unless otherwise mentioned in the methods and p-value cutoffs are shown in parentheses after reporting an effect as significant. Weighted Pearson 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 3 cell line clusters (n = 75,653) and a Tukey honest significant differences 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 HS578T was treated with 37 kinase inhibitors (n = 23,128). Fisher's exact test was used to test significance of overlap between TRRUST regulons and identified gene expression modules (supplementary Table 2 shows the size of the overlap).

Enrichment of gene sets was performed by EnrichR, an enrichment library that utilises a hyper-geometric test to identify significantly enriched terms in a gene list. The pre-ranked 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 datasets.

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.

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 Cooks's distance and removes these genes from analysis.

Significance was determined for heat-diffusion by randomising seed nodes (preserving their values) 10,000 times and selecting only the non-seed nodes that were significantly hot relative to the randomised simulations (P<0.1).

Data and Software Availability

The complete R scripts used for this methodology are available on Gitlab; `https://gitlab.ebi.ac.uk/petsalakilab/phenotype_networks`.

Acknowledgements

We would like to thank EMBL-EBI for funding the project.

Author Contributions

CB: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing - Original Draft. **EiP**: Conceptualization, Methodology, Software. **GG**: Conceptualization, Supervision. ENE: Validation, Writing - Original Draft. **ChrB**: Writing - Review & Editing, Resources. **EP**: Writing - Review & Editing, Methodology, Conceptualization, Supervision, Project administration, Funding acquisition.

Competing interests

The authors declare no conflict of interest.

References

Akhmedov, M., Kwee, I., Montemanni, R., 2016. A divide and conquer matheuristic algorithm for the Prize-collecting Steiner Tree Problem. Comput. Oper. Res. 70, 18–25. https://doi.org/10.1016/j.cor.2015.12.015

Akhmedov, M., LeNail, A., Bertoni, F., Kwee, I., Fraenkel, E., Montemanni, R., 2017. A Fast Prize-Collecting Steiner Forest Algorithm for Functional Analyses in Biological Networks, in: Salvagnin, D., Lombardi, M. (Eds.), Integration of AI and OR Techniques in Constraint Programming, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 263–276. https://doi.org/10.1007/978-3-319-59776-8_22

Alvarez-Guardia, D., Palomer, X., Coll, T., Davidson, M.M., Chan, T.O., Feldman, A.M., Laguna, J.C., Vázquez-Carrera, M., 2010. The p65 subunit of NF-kappaB binds to PGC-1alpha, linking inflammation and metabolic disturbances in cardiac cells. Cardiovasc. Res. 87, 449–458. https://doi.org/10.1093/cvr/cvq080

Baskaran, J.P., Weldy, A., Guarin, J., Munoz, G., Shpilker, P.H., Kotlik, M., Subbiah, N., Wishart, A., Peng, Y., Miller, M.A., Cowen, L., Oudin, M.J., 2020. Cell shape, and not 2D migration, predicts extracellular matrix-driven 3D cell invasion in breast cancer. APL Bioeng. 4, 026105. https://doi.org/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. https://doi.org/10.1093/bioinformatics/btu199

Bergert, M., Lembo, S., Sharma, S., Russo, L., Milovanović, D., Gretarsson, K.H., Börmel, M., Neveu, P.A., Hackett, J.A., Petsalaki, E., Diz-Muñoz, A., 2020. Cell Surface Mechanics Gate Embryonic Stem Cell Differentiation. Cell Stem Cell. https://doi.org/10.1016/j.stem.2020.10.017

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. https://doi.org/10.1371/journal.pone.0036964

Brin, S., Page, L., 1998. The anatomy of a large-scale hypertextual Web search engine. Comput. Netw. ISDN Syst., Proceedings of the Seventh International World Wide Web Conference 30, 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Chatterjee, S., 2018. Endothelial Mechanotransduction, Redox Signaling and the Regulation of Vascular Inflammatory Pathways. Front. Physiol. 9. https://doi.org/10.3389/fphys.2018.00524

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128. https://doi.org/10.1186/1471-2105-14-128

Cooper, J., Giancotti, F.G., 2019. Integrin Signaling in Cancer: Mechanotransduction, Stemness, Epithelial Plasticity, and Therapeutic Resistance. Cancer Cell 35, 347–367. https://doi.org/10.1016/j.ccell.2019.01.007

Cowell, C.F., Yan, I.K., Eiseler, T., Leightner, A.C., Döppler, H., Storz, P., 2009. Loss of cell-cell contacts induces NF-kappaB via RhoA-mediated activation of protein kinase D1. J. Cell. Biochem. 106, 714–728. https://doi.org/10.1002/jcb.22067

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.

De Belly, H., Stubb, A., Yanagida, A., Labouesse, C., Jones, P.H., Paluch, E.K., Chalut, K.J., 2020. Membrane Tension Gates ERK-Mediated Regulation of Pluripotent Cell Fate. Cell Stem Cell. https://doi.org/10.1016/j.stem.2020.10.018

Eslami Amirabadi, H., Tuerlings, M., Hollestelle, A., SahebAli, S., Luttge, R., van Donkelaar, C.C., Martens, J.W.M., den Toonder, J.M.J., 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. https://doi.org/10.1007/s10544-019-0450-5

Euskirchen, G.M., Rozowsky, J.S., Wei, C.-L., Lee, W.H., Zhang, Z.D., Hartman, S., Emanuelsson, O., Stolc, V., Weissman, S., Gerstein, M.B., Ruan, Y., Snyder, M., 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res. 17, 898–909. https://doi.org/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 9. https://doi.org/10.3390/cancers9100134

Feng, Y., Spezia, M., Huang, S., Yuan, C., Zeng, Z., Zhang, L., Ji, X., Liu, W., Huang, B., Luo, W., Liu, B., Lei, Y., Du, S., Vuppalapati, A., Luu, H.H., Haydon, R.C., He, T.-C., Ren, G., 2018. Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis. 5, 77–106. https://doi.org/10.1016/j.gendis.2018.05.001

Filer, A., Buckley, C.D., 2013. 15 - Fibroblasts and Fibroblast-like Synoviocytes, in: Firestein, G.S., Budd, R.C., Gabriel, S.E., McInnes, I.B., O'Dell, J.R. (Eds.), Kelley's Textbook of Rheumatology (Ninth Edition). W.B. Saunders, Philadelphia, pp. 215–231. https://doi.org/10.1016/B978-1-4377-1738-9.00015-3

Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., Saez-Rodriguez, J., 2019. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 29, 1363–1375. https://doi.org/10.1101/gr.240663.118

Guan, Y., Gorenshteyn, D., Burmeister, M., Wong, A.K., Schimenti, J.C., Handel, M.A., Bult, C.J., Hibbs, M.A., Troyanskaya, O.G., 2012. Tissue-Specific Functional Networks for Prioritizing Phenotype and Disease Genes. PLOS Comput. Biol. 8, e1002694. https://doi.org/10.1371/journal.pcbi.1002694

Hamilton, N.A., Pantelic, R.S., Hanson, K., Teasdale, R.D., 2007. Fast automated cell phenotype image classification. BMC Bioinformatics 8, 110. https://doi.org/10.1186/1471-2105-8-110

Han, H., Cho, J.-W., Lee, Sangyoung, Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., Lee, Sungho, Kang, B., Jeong, D., Kim, Y., Jeon, H.-N., Jung, H., Nam, S., Chung, M., Kim, J.-H., Lee, I., 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. 46, D380–D386. https://doi.org/10.1093/nar/gkx1013

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. SMC-3, 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Holland, C.H., Tanevski, J., Perales-Patón, J., Gleixner, J., Kumar, M.P., Mereu, E., Joughin, B.A., Stegle, O., Lauffenburger, D.A., Heyn, H., Szalai, B., Saez-Rodriguez, J., 2020. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. Genome Biol. 21, 36.

https://doi.org/10.1186/s13059-020-1949-z

Horton, E.R., Astudillo, P., Humphries, M.J., Humphries, J.D., 2016. Mechanosensitivity of integrin adhesion complexes: role of the consensus adhesome. Exp. Cell Res. 343, 7–13. https://doi.org/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 β1/FAK axis. Oncotarget 6, 11530–11546.

Imamura, T., Hikita, A., Inoue, Y., 2012. The roles of TGF-β signaling in carcinogenesis and breast cancer metastasis. Breast Cancer Tokyo Jpn. 19, 118–124. https://doi.org/10.1007/s12282-011-0321-2

Ishihara, S., Yasuda, M., Harada, I., Mizutani, T., Kawabata, K., Haga, H., 2013. Substrate stiffness regulates temporary NF-κB activation via actomyosin contractions. Exp. Cell Res. 319, 2916–2927. https://doi.org/10.1016/j.yexcr.2013.09.018

Iván, G., Grolmusz, V., 2011. When the Web meets the cell: using personalized PageRank for analyzing protein interaction networks. Bioinformatics 27, 405–407. https://doi.org/10.1093/bioinformatics/btq680

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., May, B., Milacic, M., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2020. The reactome pathway knowledgebase. Nucleic Acids Res. 48, D498–D503. https://doi.org/10.1093/nar/gkz1031

Kadzik, R.S., Cohen, E.D., Morley, M.P., Stewart, K.M., Lu, M.M., Morrisey, E.E., 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. https://doi.org/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. https://doi.org/10.1093/nar/gkv1070

Ke, Y., Karki, P., Zhang, C., Li, Y., Nguyen, T., Birukov, K.G., Birukova, A.A., 2019. Mechanosensitive Rap1 activation promotes barrier function of lung vascular endothelium under cyclic stretch. Mol. Biol. Cell 30, 959–974. https://doi.org/10.1091/mbc.E18-07-0422

Kim, Y.S., Han, C.Y., Kim, S.W., Kim, J.H., Lee, S.K., Jung, D.J., Park, S.Y., Kang, H., Choi, H.S., Lee, J.W., Pak, Y.K., 2001. The orphan nuclear receptor small heterodimer partner as a novel coregulator of nuclear factor-kappa b in oxidized low density lipoprotein-treated macrophage cell line RAW 264.7. J. Biol. Chem. 276, 33736–33740. https://doi.org/10.1074/jbc.M101977200

Kinsella, R.J., Kähäri, A., Haider, S., Zamora, J., Proctor, G., Spudich, G., Almeida-King, J., Staines, D., Derwent, P., Kerhornou, A., Kersey, P., Flicek, P., 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database J. Biol. Databases Curation 2011, bar030. https://doi.org/10.1093/database/bar030

Kiso, M., Tanaka, S., Saji, S., Toi, M., Sato, F., 2018. Long isoform of VEGF stimulates cell migration of breast cancer by filopodia formation via NRP1/ARHGAP17/Cdc42 regulatory network. Int. J. Cancer 143, 2905–2918. https://doi.org/10.1002/ijc.31645

Kitano, H., 2002. Systems Biology: A Brief Overview. Science 295, 1662–1664. https://doi.org/10.1126/science.1069492

Kontomanolis, E.N., Kalagasidou, S., Pouliliou, S., Anthoulaki, X., Georgiou, N., Papamanolis, V., Fasoulakis, Z.N., 2018. The Notch Pathway in Breast Cancer

Progression. Sci. World J. 2018. https://doi.org/10.1155/2018/2415489

Korotkevich, G., Sukhov, V., Sergushichev, A., 2019. Fast gene set enrichment analysis. bioRxiv 060012. https://doi.org/10.1101/060012

Krakhmal, N.V., Zavyalova, M.V., Denisov, E.V., Vtorushin, S.V., Perelmuter, V.M., 2015. Cancer Invasion: Patterns and Mechanisms. Acta Naturae 7, 17–28.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., Monteiro, C.D., Gundersen, G.W., Ma'ayan, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, W90-97. https://doi.org/10.1093/nar/gkw377

Kumar, S., Jang, I., Kim, C.W., Kang, D.-W., Lee, W.J., 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. https://doi.org/10.1038/srep36461

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., Ma'ayan, A., 2010. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinforma. Oxf. Engl. 26, 2438–2444. https://doi.org/10.1093/bioinformatics/btq466

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. https://doi.org/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. https://doi.org/10.1038/cddis.2013.244

Lee, M.-H., Wu, P.-H., Gilkes, D., Aifuwa, I., Wirtz, D., 2015. Normal mammary epithelial cells promote carcinoma basement membrane invasion by inducing microtubule-rich protrusions. Oncotarget 6, 32634–32645. https://doi.org/10.18632/oncotarget.4728

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, A., Trairatphisan, 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, 1–10. https://doi.org/10.1038/s41540-019-0118-z

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Luna, A., Babur, Ö., Aksoy, B.A., Demir, E., Sander, C., 2016. PaxtoolsR: pathway analysis in R using Pathway Commons. Bioinforma. Oxf. Engl. 32, 1262–1264. https://doi.org/10.1093/bioinformatics/btv733

Madu, Chikezie O., Wang, S., Madu, Chinua O., Lu, Y., 2020. Angiogenesis in Breast Cancer Progression, Diagnosis, and Treatment. J. Cancer 11, 4474–4494. https://doi.org/10.7150/jca.44313

Martens, M., Ammar, A., Riutta, A., Waagmeester, A., Slenter, D.N., Hanspers, K., A. Miller, R., Digles, D., Lopes, E.N., Ehrhart, F., Dupuis, L.J., Winckers, L.A., Coort, S.L., Willighagen, E.L., Evelo, C.T., Pico, A.R., Kutmon, M., 2021. WikiPathways: connecting communities. Nucleic Acids Res. 49, D613–D621. https://doi.org/10.1093/nar/gkaa1024

Mi, H., Muruganujan, A., Thomas, P.D., 2013. PANTHER in 2013: modeling the evolution of

gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. 41, D377-386. https://doi.org/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. https://doi.org/10.1016/s0092-8674(03)00278-2

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., Speed, T., Spellman, P.T., DeVries, S., Lapuk, A., Wang, N.J., Kuo, W.-L., Stilwell, J.L., Pinkel, D., Albertson, D.G., Waldman, F.M., McCormick, F., Dickson, R.B., Johnson, M.D., Lippman, M., Ethier, S., Gazdar, A., Gray, J.W., 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell 10, 515–527. https://doi.org/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. https://doi.org/10.1093/emboj/cdg087

Olson, E.N., Nordheim, A., 2010. Linking actin dynamics and gene transcription to drive cellular motile functions. Nat. Rev. Mol. Cell Biol. 11, 353–365. https://doi.org/10.1038/nrm2890

Orsulic, S., Huber, O., Aberle, H., Arnold, S., Kemler, R., 1999. E-cadherin binding prevents beta-catenin nuclear localization and beta-catenin/LEF-1-mediated transactivation. J. Cell Sci. 112, 1237–1245.

Øvrevik, J., Låg, M., Lecureur, V., Gilot, D., Lagadic-Gossmann, D., Refsnes, M., Schwarze, P.E., Skuland, T., Becher, R., Holme, J.A., 2014. AhR and Arnt differentially regulate NF-κB signaling and chemokine responses in human bronchial epithelial cells. Cell Commun. Signal. CCS 12, 48. https://doi.org/10.1186/s12964-014-0048-8

Padi, M., Quackenbush, J., 2018. Detecting phenotype-driven transitions in regulatory network structure. Npj Syst. Biol. Appl. 4, 1–12. https://doi.org/10.1038/s41540-018-0052-5

Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A.M.-P., George, N., Fexova, S., Fonseca, N.A., Füllgrabe, A., Green, M., Huang, N., Huerta, L., Iqbal, H., Jianu, M., Mohammed, S., Zhao, L., Jarnuczak, A.F., Jupp, S., Marioni, J., Meyer, K., Petryszak, R., Prada Medina, C.A., Talavera-López, C., Teichmann, S., Vizcaino, J.A., Brazma, A., 2020. Expression Atlas update: from tissues to single cells. Nucleic Acids Res. 48, D77–D83. https://doi.org/10.1093/nar/gkz947

Piran, Mehran, Karbalaei, R., Piran, Mehrdad, 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. https://doi.org/10.3390/biom10060850

Pires, B.R.B., Mencalha, A.L., Ferreira, G.M., de Souza, W.F., Morgado-Díaz, J.A., Maia, A.M., Corrêa, S., Abdelhay, E.S.F.W., 2017. NF-kappaB Is Involved in the Regulation of EMT Genes in Breast Cancer Cells. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0169622

Resnik, P., 1995. Using information content to evaluate semantic similarity in a taxonomy, in: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1, IJCAI'95. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 448–453.

Robertson, J., Jacquemet, G., Byron, A., Jones, M.C., Warwood, S., Selley, J.N., Knight, D., Humphries, J.D., Humphries, M.J., 2015. Defining the phospho-adhesome through the phosphoproteomic analysis of integrin signalling. Nat. Commun. 6, 6265.

https://doi.org/10.1038/ncomms7265

Roche, J., 2018. The Epithelial-to-Mesenchymal Transition in Cancer. Cancers 10. https://doi.org/10.3390/cancers10020052

Rolfe, R.A., Nowlan, N.C., Kenny, E.M., Cormican, P., Morris, D.W., Prendergast, P.J., 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. https://doi.org/10.1186/1471-2164-15-48

Sailem, H.Z., 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. https://doi.org/10.1101/gr.202028.115

Sero, J.E., Sailem, H.Z., Ardy, R.C., Almuttaqi, H., Zhang, T., Bakal, C., 2015. Cell shape and the microenvironment regulate nuclear translocation of NF-κB in breast epithelial and tumor cells. Mol. Syst. Biol. 11. https://doi.org/10.15252/msb.20145644

Sharma, S., Petsalaki, E., 2019. Large-scale datasets uncovering cell signalling networks in cancer: context matters. Curr. Opin. Genet. Dev., Cancer Genomics 54, 118–124. https://doi.org/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. https://doi.org/10.1242/jcs.103861

Shrum, C.K., Defrancisco, D., Meffert, M.K., 2009. Stimulated nuclear translocation of NF-kappaB and shuttling differentially depend on dynein and the dynactin complex. Proc. Natl. Acad. Sci. U. S. A. 106, 2647–2652. https://doi.org/10.1073/pnas.0806677106

Siegel, R.L., Miller, K.D., Jemal, A., 2019. Cancer statistics, 2019. CA. Cancer J. Clin. 69, 7–34. https://doi.org/10.3322/caac.21551

Song, Y., Washington, M.K., Crawford, H.C., 2010. Loss of FOXA1/2 Is Essential for the Epithelial-to-Mesenchymal Transition in Pancreatic Cancer. Cancer Res. 70, 2115–2125. https://doi.org/10.1158/0008-5472.CAN-09-2979

Soul, J., Hardingham, T.E., Boot-Handford, R.P., Schwartz, J.-M., 2015. PhenomeExpress: A refined network analysis of expression datasets by inclusion of known disease phenotypes. Sci. Rep. 5, 8117. https://doi.org/10.1038/srep08117

Stathias, V., Turner, J., Koleti, A., Vidovic, D., Cooper, D., Fazel-Najafabadi, M., Pilarczyk, M., Terryn, R., Chung, C., Umeano, A., Clarke, D.J.B., Lachmann, A., Evangelista, J.E., Ma'ayan, A., Medvedovic, M., Schürer, S.C., 2020. LINCS Data Portal 2.0: next generation access point for perturbation-response signatures. Nucleic Acids Res. 48, D431–D439. https://doi.org/10.1093/nar/gkz1023

Stormo, G.D., 2013. Modeling the specificity of protein-DNA interactions. Quant. Biol. Beijing China 1, 115–130. https://doi.org/10.1007/s40484-013-0012-4

Su, J.-L., Yang, P.-C., Shih, J.-Y., Yang, C.-Y., Wei, L.-H., Hsieh, C.-Y., Chou, C.-H., Jeng, Y.-M., Wang, M.-Y., Chang, K.-J., Hung, M.-C., Kuo, M.-L., 2006. The VEGF-C/Flt-4 axis promotes invasion and metastasis of cancer cells. Cancer Cell 9, 209–223. https://doi.org/10.1016/j.ccr.2006.02.018

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., Lahr, D.L., Hirschman, J.E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I.C., Lam, D., Liberzon, A., Toder, C., Bagul, M., Orzechowski, M., Enache, O.M., Piccioni, F., Johnson, S.A., Lyons, N.J., Berger, A.H., Shamji, A.F., Brooks, A.N., Vrcic, A., Flynn, C., Rosains, J., Takeda, D.Y., Hu, R., Davison, D., Lamb, J., Ardlie, K., Hogstrom, L., Greenside, P., Gray,

N.S., Clemons, P.A., Silver, S., Wu, Xiaoyun, Zhao, W.-N., Read-Button, W., Wu, Xiaohua, Haggarty, S.J., Ronco, L.V., Boehm, J.S., Schreiber, S.L., Doench, J.G., Bittker, J.A., Root, D.E., Wong, B., Golub, T.R., 2017. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 171, 1437-1452.e17. https://doi.org/10.1016/j.cell.2017.10.049

Szalai, B., Saez-Rodriguez, J., 2020. Why do pathway methods work better than they should? FEBS Lett. n/a. https://doi.org/10.1002/1873-3468.14011

Taherian, A., Li, X., Liu, Y., Haas, T.A., 2011. Differences in integrin expression and signaling within human breast cancer cells. BMC Cancer 11, 293. https://doi.org/10.1186/1471-2407-11-293

Teo, H., Ghosh, S., Luesch, H., Ghosh, A., Wong, E.T., Malik, N., Orth, A., de Jesus, P., Perry, A.S., Oliver, J.D., Tran, N.L., Speiser, L.J., Wong, M., Saez, E., Schultz, P., Chanda, S.K., Verma, I.M., Tergaonkar, V., 2010. Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression. Nat. Cell Biol. 12, 758–767. https://doi.org/10.1038/ncb2080

Tong, H., Faloutsos, C., Pan, J.-Y., 2008. Random walk with restart: fast solutions and applications. Knowl. Inf. Syst. 14, 327–346. https://doi.org/10.1007/s10115-007-0094-2

Tong, L., Tergaonkar, V., 2014. Rho protein GTPases and their interactions with NFκB: crossroads of inflammation and matrix biology. Biosci. Rep. 34. https://doi.org/10.1042/BSR20140021

Türei, D., Korcsmáros, T., Saez-Rodriguez, J., 2016. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat. Methods 13, 966–967. https://doi.org/10.1038/nmeth.4077

Türei, D., Valdeolivas, A., Gul, L., Palacio-Escat, N., Ivanova, O., Gábor, A., Módos, D., Korcsmáros, T., Saez-Rodriguez, J., 2020. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. bioRxiv 2020.08.03.221242. https://doi.org/10.1101/2020.08.03.221242

Van Tubergen, E.A., Banerjee, R., Liu, M., Vander Broek, R., Light, E., Kuo, S., Feinberg, S.E., Willis, A.L., Wolf, G., Carey, T., Bradford, C., Prince, M., Worden, F.P., Kirkwood, K.L., D'Silva, N.J., 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. Off. J. Am. Assoc. Cancer Res. 19, 1169–1179. https://doi.org/10.1158/1078-0432.CCR-12-2927

Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat. Rev. Genet. 13, 227–232. https://doi.org/10.1038/nrg3185

Wang, B., Kumar, V., Olson, A., Ware, D., 2019. Reviving the Transcriptome Studies: An Insight Into the Emergence of Single-Molecule Transcriptome Sequencing. Front. Genet. 10. https://doi.org/10.3389/fgene.2019.00384

Wasserman, W.W., Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. Nat. Rev. Genet. 5, 276–287. https://doi.org/10.1038/nrg1315

Wildwater, M., Sander, N., Vreede, G. de, Heuvel, S. van den, 2011. Cell shape and Wnt signaling redundantly control the division axis of C. elegans epithelial stem cells. Development 138, 4375–4385. https://doi.org/10.1242/dev.066431

Wozniak, M.A., Chen, C.S., 2009. Mechanotransduction in development: a growing role for contractility. Nat. Rev. Mol. Cell Biol. 10, 34–43. https://doi.org/10.1038/nrm2592

Wu, P.-H., Gilkes, D.M., Phillip, J.M., Narkar, A., Cheng, T.W.-T., Marchand, J., Lee, M.-H.,

Li, R., Wirtz, D., 2020. Single-cell morphology encodes metastatic potential. Sci. Adv. 6, eaaw6938. https://doi.org/10.1126/sciadv.aaw6938

Wu, X., Chen, H., Parker, B., Rubin, E., Zhu, T., Lee, J.S., Argani, P., Sukumar, S., 2006. HOXB7, a Homeodomain Protein, Is Overexpressed in Breast Cancer and Confers Epithelial-Mesenchymal Transition. Cancer Res. 66, 9527–9534. https://doi.org/10.1158/0008-5472.CAN-05-4470

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. https://doi.org/10.1111/cas.13771

Yu, J.-L., Deng, R., Chung, S.K., Chan, G.C.-F., 2016. Epac Activation Regulates Human Mesenchymal Stem Cells Migration and Adhesion. STEM CELLS 34, 948–959. https://doi.org/10.1002/stem.2264

Zhang, Y., Chiu, S., Liang, X., Gao, F., Zhang, Z., Liao, S., Liang, Y., Chai, Y.-H., Low, D.J.H., Tse, H.-F., Tergaonkar, V., Lian, Q., 2015. Rap1-mediated nuclear factor-kappaB (NF- κ B) activity regulates the paracrine capacity of mesenchymal stem cells in heart repair following infarction. Cell Death Discov. 1, 1–11. https://doi.org/10.1038/cddiscovery.2015.7

Zhang, Y., Hagedorn, C.H., Wang, L., 2011. Role of Nuclear Receptor SHP in Metabolism and Cancer. Biochim. Biophys. Acta 1812, 893–908. https://doi.org/10.1016/j.bbadis.2010.10.006

Zhang, Y.-L., Wang, R.-C., Cheng, K., Ring, B.Z., Su, L., 2017. Roles of Rap1 signaling in tumor cell migration and invasion. Cancer Biol. Med. 14, 90–99. https://doi.org/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. https://doi.org/10.1111/j.1742-4658.2009.06986.x