



OPEN

## Simultaneous screening of overexpressed genes in breast cancer for oncogenic drivers and tumor dependencies

Adaobi Mofunanya<sup>1</sup>, Eleanor R. Cameron<sup>2</sup>, Christian J. Braun<sup>2</sup>, Frank Celeste<sup>1,3</sup>, Xiaoyu Zhao<sup>1,4</sup>, Michael T. Hemann<sup>2</sup>, Kenneth L. Scott<sup>5,6</sup>, Jinyu Li<sup>1</sup> & Scott Powers<sup>1,3,4</sup>✉

There are hundreds of genes typically overexpressed in breast cancer cells and it's often assumed that their overexpression contributes to cancer progression. However, the precise proportion of these overexpressed genes contributing to tumorigenicity remains unclear. To address this gap, we undertook a comprehensive screening of a diverse set of seventy-two genes overexpressed in breast cancer. This systematic screening evaluated their potential for inducing malignant transformation and, concurrently, assessed their impact on breast cancer cell proliferation and viability. Select genes including *ALDH3B1*, *CEACAM5*, *IL8*, *PYGO2*, and *WWTR1*, exhibited pronounced activity in promoting tumor formation and establishing gene dependencies critical for tumorigenicity. Subsequent investigations revealed that *CEACAM5* overexpression triggered the activation of signaling pathways involving  $\beta$ -catenin, Cdk4, and mTOR. Additionally, it conferred a growth advantage independent of exogenous insulin in defined medium and facilitated spheroid expansion by inducing multiple layers of epithelial cells while preserving a hollow lumen. Furthermore, the silencing of *CEACAM5* expression synergized with tamoxifen-induced growth inhibition in breast cancer cells. These findings underscore the potential of screening overexpressed genes for both oncogenic drivers and tumor dependencies to expand the repertoire of therapeutic targets for breast cancer treatment.

Genome-wide estimates of the number of genes overexpressed in specific cancer types range in the low hundreds to high hundreds<sup>1,2</sup>. Depending upon their expression pattern in cancers, these genes can be useful biomarkers for detection, prognosis, or prediction of treatment response. However, determining whether they promote tumorigenicity or evaluating their utility as potential therapeutic targets requires functional analysis. Overexpressed genes may be growth-inhibitory, particularly if their overexpression results from a dysregulated negative feedback loop. For example, the HPV oncoprotein E7 binds and induces proteolytic degradation of the Rb protein, which relieves the tumor suppressor gene p16<sup>ink4a</sup> from negative feedback control, leading to its upregulation<sup>3,4</sup>. Likewise, *BRAF-V600E* mutant melanomas and *KRAS*-mutant lung adenocarcinomas both overexpress several inhibitors of the mitogen-activated kinase pathway<sup>5,6</sup>. In these cases, the primary genetic alteration blocks the negative feedback exerted by the overexpressed inhibitory genes.

Single-gene, small-scale functional analysis of overexpressed genes has led to the generation of hundreds of proposed tumor-promoting genes<sup>7</sup>. At present, it is difficult to evaluate many of these studies since they often use unique experimental conditions to determine oncogenicity. More systematic approaches that survey the induced oncogenic phenotypes of multiple candidate genes have the advantage of providing several internal comparisons. Using cDNA-based or open-reading-frame (ORF) overexpression, systematic screening has identified oncogenic roles for *PVRL4* in breast cancer and proliferation of T-cells by *LTBR*<sup>8,9</sup>. In addition to these systematic screens for genes capable of promoting the tumorigenicity or proliferation, overexpression screens have also been used to identify *GATAD2B* and *PYGO2* as drivers of metastasis<sup>9,10</sup> as well as the chromatin-modifying complex SAGA as a driver of pluripotency<sup>11</sup>.

<sup>1</sup>Department of Pathology, Stony Brook Cancer Center, Stony Brook, NY 11794, USA. <sup>2</sup>Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. <sup>3</sup>Graduate Program in Genetics, Stony Brook University, Stony Brook, NY 11794, USA. <sup>4</sup>Graduate Program in Molecular and Cellular Biology, Stony Brook University, Stony Brook, NY 11794, USA. <sup>5</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. <sup>6</sup>Kenneth L. Scott is deceased. ✉email: scott.powers@stonybrook.edu

In this study, we set out to determine what percentage of overexpressed genes in breast cancer could promote tumorigenicity of premalignant mammary epithelial cells. We performed pooled screening with barcoded ORFs and measured the relative number of mammary epithelial cells expressing each individual ORF both before and after growth in three different conditions, including *in vivo* tumor formation. In parallel, we used shRNA screening to determine if the overexpression of these genes was selectively required in breast cancer cell lines. To our knowledge, this is the first report on concurrent screening for drivers and dependencies. Additionally, we used the proteomic method of reverse phase protein array (RPPA) profiling of protein levels and phosphorylation to gain valuable insights into the cancer biology and oncogenic mechanisms of our lead candidates.

## Results

### Selection of overexpressed genes and transfection of LentiORFs into both MCF10A and NMuMG

We selected genes based on their overexpression in breast cancer and their presence in a previously acquired Precision LentiORF expression library containing 3526 ORFs. We applied both outlier and cancer vs. normal differential analysis of breast cancer TCGA data using OncoPrint Research tools and chose the top 100 overexpressed genes for functional studies (Supplementary Table 1). We performed a pilot experiment with ten of the chosen ORFs and measured the ability of amplicon sequencing to detect the different ORFs after transfection into MCF10A cells, plasmid integration and selection, and one week of growth before amplifying the ORFs from genomic DNA using universal 5' and 3' ORF primers. Based on the uneven amplification of the ORFs, we decided to switch to a 24-bp barcoded ORF system which promised to yield more even amplification of the uniform-length and balanced nucleotide composition barcodes. We tagged each individual ORF with a unique barcode for next-generation sequencing and inserted the tagged-ORF into the pLenti6.3 R1R2 lentiviral expression vector using the production pipeline previously described<sup>12</sup>. We successfully transferred and sequence-verified seventy-two ORFs, and these barcoded-ORF expression vectors were transfected individually (to ensure an initial even representation) into the immortalized human mammary epithelial cell line MCF10A. MCF10A resemble primary breast epithelial cells in that they require substrate attachment, growth factors and hormones for growth, are contact-inhibited, and are not tumorigenic. They have two cancer-related genetic alterations: homozygous deletion of the *CDKN2A* locus and amplification of *MYC*, and this allows them to be transformed *in vitro* by mutant *PIK3CA* or *RAS* oncogenes, but acquiring the ability to form carcinomas *in vivo* requires several additional oncogenic alterations<sup>13</sup>. Mouse cells require fewer genetic alterations to form tumors, thus for *in vivo* screening we chose the normal murine mammary epithelial cell line NMuMG, which is heterozygous for a missense *TP53* mutant (R277C) and forms tumors when transfected with powerful oncogenes such as mutationally activated *RAS*<sup>14</sup>. In order to allow NMuMG cells to form tumors with weaker oncogenes, we overexpressed *MYC*, a procedure that we had used previously in our cDNA screen for tumor-promoting genes using *TP53* mutant hepatoblasts<sup>15</sup>. As we did with MCF10A, the seventy-two barcoded-ORF expression vectors were transfected individually into NMuMG-Myc cells.

Following selection of transfected cells with blasticidin, the individual transfectants were cultured briefly and then equal cell numbers of each were combined into a pool. Each pool was then divided into three separate cultures to generate biological replicates. After three days, there were sufficient cells from each replicate to use for isolation of genomic DNA, PCR amplification of barcodes, and amplicon sequencing to determine the baseline representation of the different ORF-transfectants. The MCF10A ORF-transfectant pool replicates were then placed into 3D culture (Matrigel) for three weeks to allow for spheroid formation, and in parallel grown for three weeks under standard 2D cell culture conditions (with passaging as needed to avoid confluence). Each of the NMuMG-Myc ORF-transfectant pool replicates were injected subcutaneously into immunodeficient mice and the mice were sacrificed after eight weeks and the tumors removed. None of the mice injected with control NMuMG-Myc empty vector transfectants formed tumors after eight weeks.

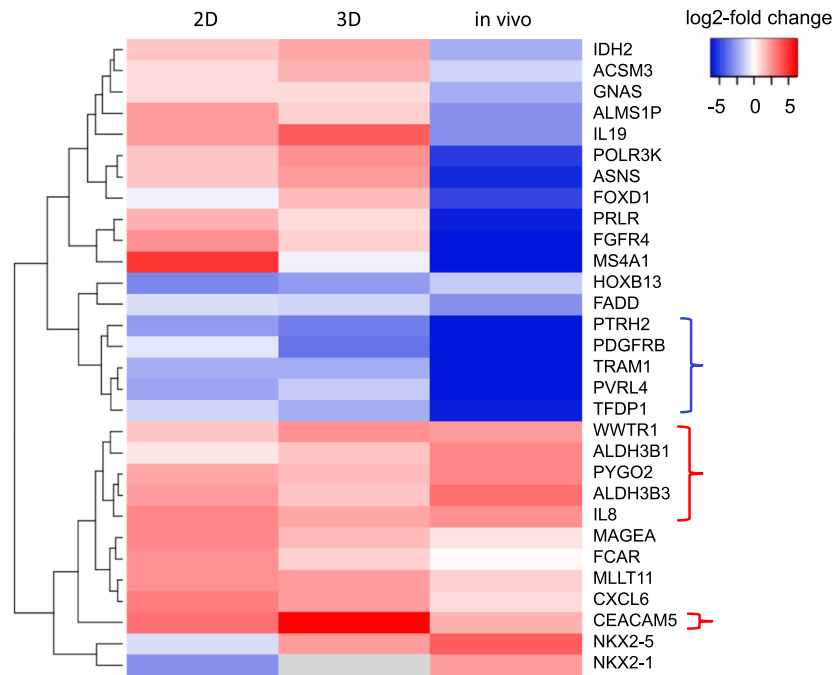
For each of these growth conditions, genomic DNA was isolated and the barcode read counts determined by amplicon sequencing. Following preprocessing, normalization, and log-transformation of the barcode read counts, the ORF frequencies in the three biological replicates of the initial time point were compared to the ORF frequencies in the latter time points.

### Overexpressed genes vary widely in their effects on growth and tumor formation

The values and statistical tests of log<sub>2</sub>-fold changes in the seventy-two ORF proportions for the three growth and/or tumor assays is tabulated in Supplementary Table 2. Filtering for statistical significance ( $p < 0.05$ ) reduced the number of ORFs to thirty, and we used a heatmap with hierarchical clustering to examine the overall patterns of log<sub>2</sub>-fold enrichment or depletion. One set of genes that clustered together showed enrichment in all three assays, and included *ALDH3B1*, *ALDH3B2*, *IL8*, *PYGO2*, and *WWTR1* (Fig. 1). *CEACAM5* also showed enrichment in all three assays, even though it was a member of a distinct but nearby cluster (Fig. 1). Interestingly, *CEACAM5* was also the strongest inducer of growth in 3D spheroids (Supplementary Table 2).

Several of these genes have been previously shown to either promote tumorigenicity or tumorigenic properties. These include the proinflammatory cytokine *IL8* which stimulates the growth of triple-negative breast cancer cells<sup>16</sup>; the transcriptional activator *WWTR1* which induces cancer stem-cell phenotypes in breast cancer<sup>17</sup>; and *CEACAM5* and *PYGO2* which have both been shown to promote metastasis when overexpressed<sup>9,18</sup>. Although not proven to be oncogenic, *ALDH3B1* and *ALDH3B2* are members of the aldehyde dehydrogenases family that are highly expressed in therapy-resistant cells in different cancer types<sup>19</sup>.

At the other end of the spectrum in Fig. 1, there is a cluster of genes with very high tumor-inhibitory activity which includes *PDGFRB*, *TRAM1*, *PVRL4*, *TFDP1*, and *PTRH2* (Fig. 1). Surprisingly, three of these genes have been shown to promote oncogenic properties when overexpressed<sup>20–23</sup>. Although our results with these genes



**Figure 1.** Heatmap visualization and hierarchical clustering of the growth phenotypes induced by overexpression of different ORFs. ORF library transfectants of MCF10A were grown in standard cell culture (2D) or suspension in matrigel (3D), and ORF library transfectants of NMuMG were screened in vivo. Log<sub>2</sub>-transformed changes in ORF proportions compared to initial library transfectants were determined by barcode-sequencing. Significance of log<sub>2</sub>-transformed changes were assessed using a Student's t-test (see text). The red brackets signify ORFs that showed enrichment in all three assays, whereas the blue bracket signifies the opposite.

appear to be contradictory, we note that extensively validated cancer genes (*RUNX1* and *NOTCH* family members) have been shown to have dual roles of either tumor suppression or survival/tumor promotion depending upon the specific cell type<sup>24,25</sup>. We think it is likely that some overexpressed genes will also have opposing roles that are cell-type dependent.

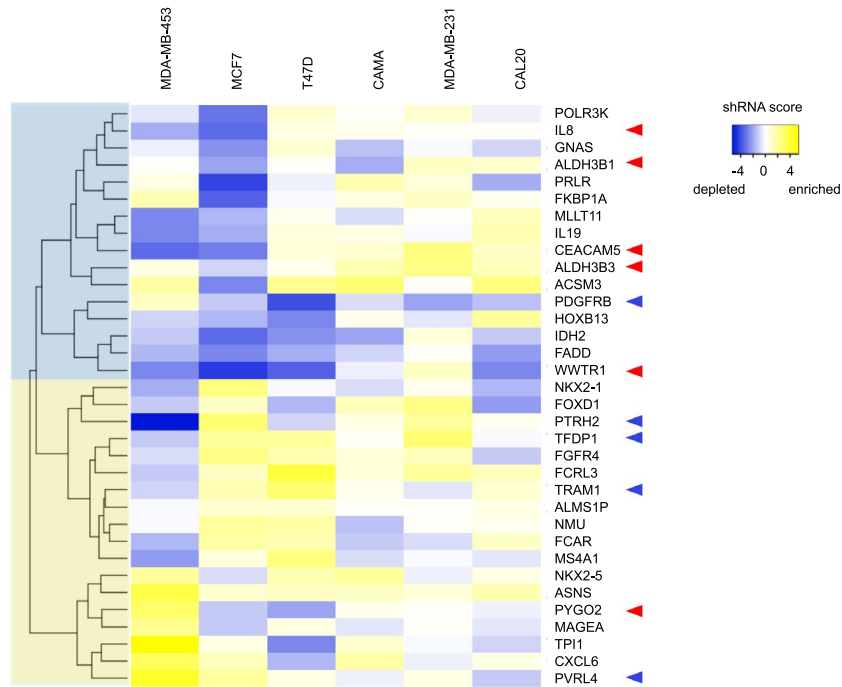
### Tumor dependencies revealed by shRNA screening

In parallel, we examined the tumor dependencies in five different breast cancer cell lines by pooled shRNA screening of the same set of seventy-two overexpressed genes. Focusing on the subset of thirty genes with significant results in the ORF screening, we used a heatmap with hierarchical clustering to examine the overall patterns of shRNA enrichment or depletion. Although there was pronounced variability between the cell lines, we detected two gene clusters, one showing slightly higher growth inhibition by shRNAs (shRNA depletion), and another showing slightly less growth inhibition (shRNA enrichment) (Fig. 2). Five out of six oncogenic genes from the ORF screen were in the cluster of genes showing slightly higher growth inhibition by shRNAs, and four out of five inhibitory genes from the ORF screen were in the other cluster of genes showing slightly less growth inhibition (Fig. 2). These results suggest that there is an association of growth and tumor-promoting ability with the capability of targeting shRNAs to inhibit growth and/or survival.

### Correlation between tumor promoting ability and tumor dependency

To quantitatively assess the relationship between growth and tumor-promoting ability and shRNA growth inhibition, we created a series of summary statistics for shRNA depletion for all five cell lines, including the minimum value, first quartile, and mean. The most significant correlation was between the first quartile summary shRNA statistic and tumor-forming ability. To switch to the more intuitive concept of tumor dependency, we used the inverse of the shRNA inhibition summary statistic. Plotting tumor-forming ability versus tumor dependency revealed a weak positive correlation ( $r=0.24$ ) which was not significant based on the standard cutoff of 0.05 ( $p < 0.108$ ), yet the trend is clear (Fig. 3).

The upper right section of the graph in Fig. 3 shows genes with concomitant high tumor-promoting activity and high tumor dependency and includes five of the six genes from the growth and tumor-promoting cluster (see Fig. 1). Also within this group is a tumor-promoting gene (*NKX2-1*) was not in the growth and tumor-promoting cluster because it failed to promote growth in cell culture (see Fig. 1). Conversely, a single gene from the growth and tumor-promoting cluster, *ALDH3B2*, showed low tumor dependency scores (Fig. 3). Whether these results indicate that concurrent screening for drivers and dependencies are an effective way to identify potential therapeutic targets is addressed in the discussion section.



**Figure 2.** Heatmap visualization and hierarchical clustering of gene inhibition mediated by shRNAs in a panel of five breast cancer cell lines. shRNA library transfectants of the five breast cancer cell lines were grown in standard cell culture for two weeks. Log<sub>2</sub>-transformed changes in shRNA proportions compared to initial library transfectants were determined by barcode-sequencing. Data was transformed into Z-scores for data visualization. Blue shading indicates slightly higher growth inhibition by shRNAs, whereas yellow shading indicates the opposite. Red arrows point to genes that promoted growth in all three growth assays per Fig. 1, blue arrows indicate genes that inhibited growth in all three assays.

### Confirmation that *CEACAM5* overexpression induces tumorigenicity

As pointed out earlier, the cell adhesion gene *CEACAM5* was the strongest inducer of growth in 3D spheroids—more than threefold greater than the next strongest inducer of growth in 3D spheroids (*IL19*) (Fig. 1, Supplementary Table 2). These findings suggest that the oncogenic mechanism of *CEACAM5* differs from that of other drivers. Based on this, we focused our follow-up studies on *CEACAM5*. We first tested the validity of the pooled screening result that *CEACAM5* overexpression promoted the tumorigenicity of NMuMG-Myc cells. NMuMG-Myc cells were transduced with a lentiORF *CEACAM5* expression construct and shown by quantitative RT-PCR to express high levels of *CEACAM5*. We determined that *CEACAM5* overexpression induced rapidly growing tumors following subcutaneous injection of  $5 \times 10^6$  cells. Over the same period of four weeks, orthotopic injection of  $5 \times 10^6$  cells into mammary fat pads produced slightly smaller tumors. Control NMuMG-Myc cells did not form tumors under either of these conditions (Supplementary Fig. 1).

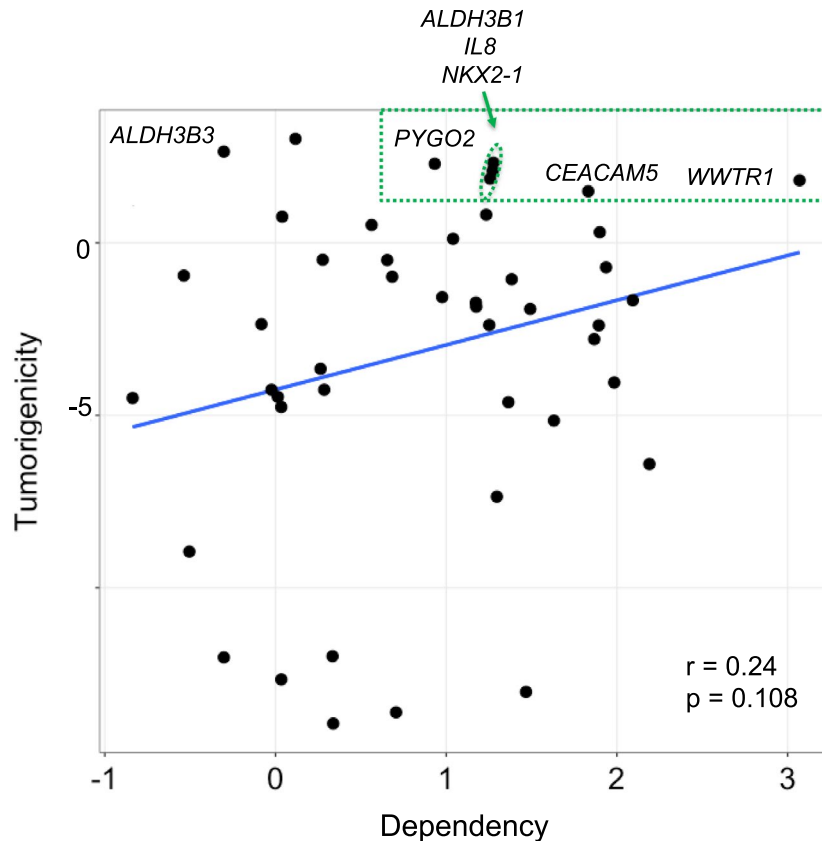
### shRNA mediated suppression of *CEACAM5* in reduces clonogenic growth in breast cancer cell lines and enhances tamoxifen-growth inhibition

To validate the pooled shRNA screening results indicating that some breast cancer cell lines are dependent on expression of *CEACAM5* for growth, we constructed two mir-30 based shRNAs plasmids targeting *CEACAM5* based on the two strongest shRNAs from the pooled screening results. Following lentiviral transfection, we determined that each one had strong on-target activity in suppressing *CEACAM5* expression (Fig. 4A). Both shRNAs suppressed the clonogenic growth of two breast cancer cell lines, T47D and MCF7, but neither shRNA was able to suppress the clonogenic growth of unrelated kidney epithelial 293 T cells (Fig. 4B).

We noted that high *CEACAM5* protein expression as ascertained by immunohistochemistry has been associated with resistance to endocrine therapy in breast cancer patients<sup>26</sup>. Based on this, we examined whether suppression of *CEACAM5* expression could modify growth-inhibitory effects of tamoxifen in the two ER-positive cell lines T47D and MCF7. Although suppression of *CEACAM5* expression did not modify the response of MCF7 (data not shown), it enhanced growth inhibition by tamoxifen in T47D (Fig. 5). This result suggests that it is possible that inhibition of *CEACAM5* could be of therapeutic benefit in treating breast cancer patients.

### Overexpression of *CEACAM5* obviates the need for insulin in defined medium and promotes spheroid growth by a unique mechanism

We tested the ability of *CEACAM5* overexpression to substitute for specific growth factors for cell culture growth. *CEACAM5* overexpressing cells were completely independent of insulin as a growth factor, and although less dependent upon EGF than control MCF10A cells, they still showed EGF dependence (Fig. 6A). To validate the

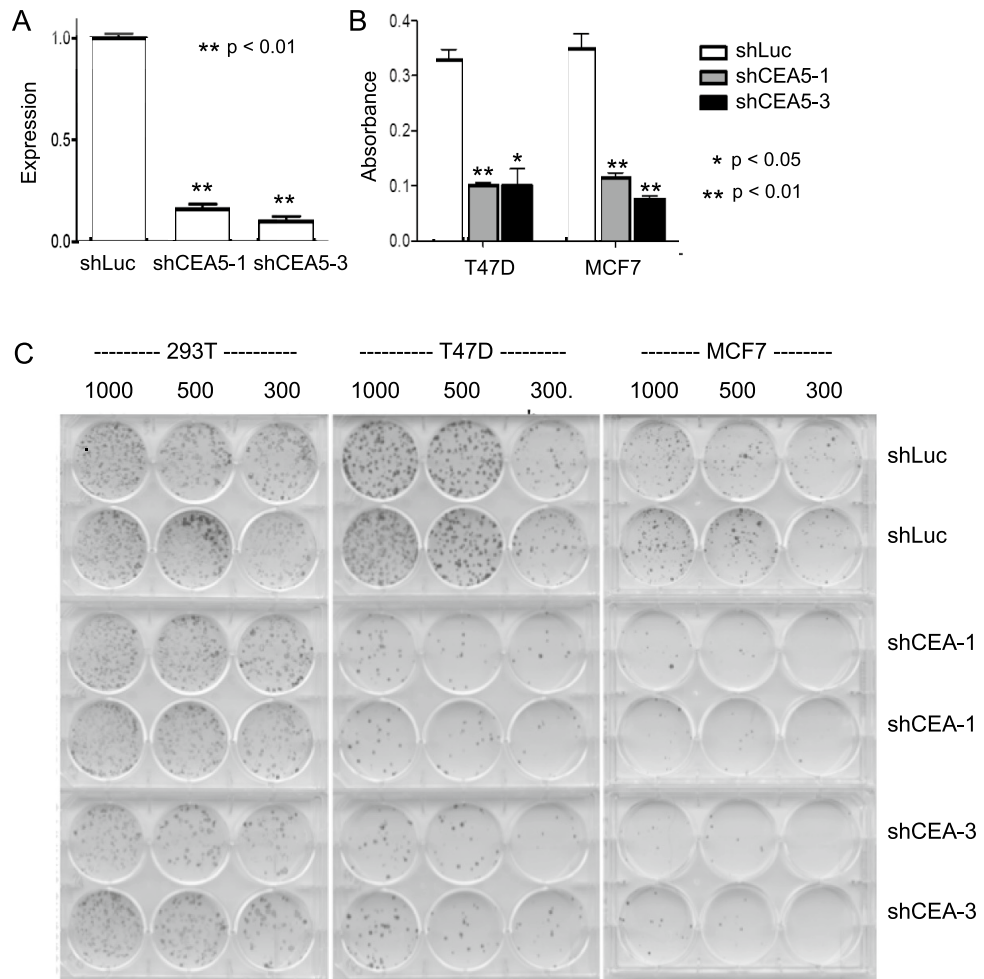


**Figure 3.** Relationship between induction of tumorigenicity and tumor dependency. Scatter plot of the tumor dependency scores (see text) on the x-axis, and tumorigenicity scores on the y-axis. The indicated genes correspond to ORFs that showed enrichment in all three assays (see Fig. 1) except for *NKX2-1* (see text). The fitted line represents the linear relationship between the two variables, which did not pass the significance threshold of  $p < 0.05$  but clearly indicated a trend.

ORF screening results that indicated that overexpression of *CEACAM5* greatly increased the growth of MCF10A cells in 3D culture, we examined MCF10A-*CEACAM5* after growth in a 3D matrigel matrix and determined that *CEACAM5* overexpression induced a modest but significant increase in the average diameter of spheroids formed in matrigel, as visualized by a time-series of frequency distributions of spheroid sizes (Fig. 6B). This modest increase could not explain the stronger growth predicted from the screening results. However, examination of spheroids after they were fixed, sectioned, and visualized by staining with DAPI revealed that *CEACAM5* overexpression increased growth significantly within the spheroid itself. Unlike the oncogenes *ERBB2*, *PIK3CA*, or *WWTR1* previously tested in the MCF10A spheroid formation assay, *CEACAM5* overexpression did not alter the spheroid morphology or prevent hollowing out of the inner mass of cells<sup>27</sup>, but instead induced a notable increase in growth in the spheroid periphery, increasing the number of epithelial cell layers (Fig. 6C).

#### RPPA analysis of the effects of *CEACAM5* overexpression on the functional proteome

To gain insight into the biochemical mechanisms by which *CEACAM5* overexpression induces tumorigenicity, we used the high-throughput reverse-phase protein analysis (RPPA) pipeline established at MD Anderson to examine how *CEACAM5* expression affected levels and post-translational modifications of several key proteins involved in growth control and other cancer-related phenotypes<sup>28</sup>. We transfected MCF10A cells with either the *CEACAM5* ORF expression construct, or *WWTR1*-, *ERBB2*-, *SCUBE3*-ORF expression constructs for comparison, along with empty vector as a control. Following selection for stable integration, we cultured the cells briefly in standard medium and prepared biological triplicates for RPPA analysis. Out of a total of 304 protein measurements, 93 were significantly different between the five groups (ANOVA  $p < 0.01$ ). To visualize differences more readily between the five groups, we used a more stringent significance cut-off ( $p < 0.001$ ) which reduced the number to 28. We used a heatmap with hierarchical clustering to examine the overall patterns of protein enrichment or depletion (Fig. 7). The bottom 10 proteins in the heatmap are expressed relatively higher in normal, empty-vector MCF10A cells, whereas the top 18 proteins are expressed in relatively lower levels in control MCF10A cells (Fig. 7). Within this top cluster is the middle branch of 5 proteins that are preferentially expressed in MCF10A-*CEACAM5* cells, most notably  $\beta$ -catenin (both total and phosphorylated), indicating strong activation of the canonical Wnt-signaling pathway (Fig. 7). The possibility that this underlies the mechanism by which MCF10A-*CEACAM5* spheroids form multiple epithelial cell layers is addressed in the discussion. Also



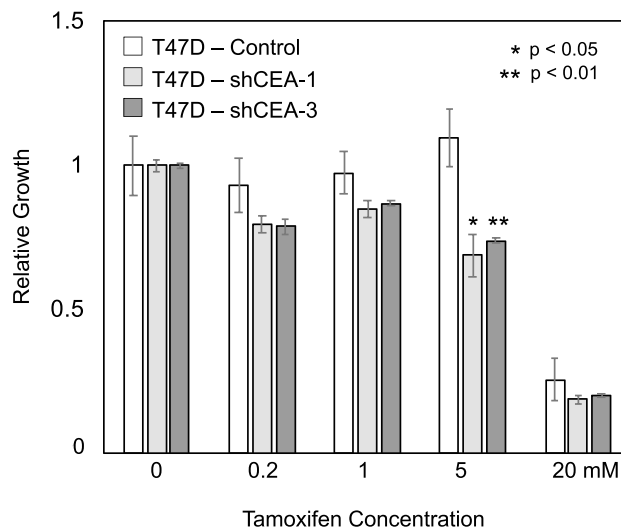
**Figure 3.** shRNA silencing of *CEACAM5* inhibits the clonogenic growth of breast cancer cell lines. Panel (A) Validation of on-target effects of shRNAs targeting *CEACAM5*. T47D cells were transfected with either a shRNA targeting luciferase (control) or two independent shRNAs targeting *CEACAM5* (Ceacam5-1 and Ceacam5-3). Following selection for stable transfectants, quantitative RT-PCR analysis of *CEACAM5* RNA was performed for expression levels of on RNA extracted from transfected T47D cells. GAPDH expression was used for normalization and results were represented as expression relative to T47D-luc. Panel (B) Quantification of clonal growth over 3 weeks in cell lines infected with shRNAs against *CEACAM5* (Ceacam5-1 and Ceacam5-3) relative to a shRNA against luciferase (control). Panel (C) Colony formation assay of 293 T, T47D and MCF7 cells infected with shRNA against luciferase (control) or *CEACAM5* (Ceacam5-1 and Ceacam5-3). The significance of the findings was assessed using a Student's t-test, while error bars depict the standard error of the means.

included in this group is CDK1, which in addition to its role in cell cycle progression also contributes to cancer progression<sup>29</sup>. Also notable is the activation of mTOR as indicated by increased phosphorylation of its canonical substrate, ribosomal subunit S6, a property it shares with WWTR1 (Fig. 7).

We reasoned that some of the proteins that were significantly altered at the  $p < 0.01$  level but not at the  $p < 0.001$  level might be relevant to how *CEACAM5* transforms cells. We did not analyze any protein that was not significantly different between control MCF10A and *CEACAM5*-MCF10 (t-test  $p < 0.05$ ). We detected increased phosphorylation of the RB1 protein, indicating activation of CDK4/6 signaling (Fig. 8). Additionally, we found four proteins with tumor suppressive properties that clustered together and were significantly reduced in *CEACAM5*-MCF10, including BECN, BAP1, PCDC4, and PMS2 (Fig. 8). The modest suppression of BCL2L1/Bim by overexpression of *CEACAM5* is consistent with the formation of empty spheroid lumen as discussed below.

## Discussion

The primary motivation behind this investigation was to ascertain the proportion of overexpressed genes in breast cancer that might function as oncogenic drivers and serve as potential tumor dependencies and therapeutic targets. Our analysis revealed that six out of the seventy-five genes scrutinized met these criteria. This observation,



**Figure 5.** Silencing of *CEACAM5* sensitizes T47 breast cancer cell line to growth inhibition by tamoxifen. Quantification of viability of T47D cells and T47D-*CEACAM5* knockdown derivatives after four days growth in standard medium with the indicated concentrations of tamoxifen.

considering the hundreds of genes overexpressed in breast cancer, implies the existence of a greater number of therapeutic targets than previously envisaged. However, it is noteworthy that five of the identified genes were previously shown to possess oncogenic properties in breast cancer, with *ALDH3B1* being the sole exception. The quest for genuinely novel targets may necessitate pooled screening conducted at a significantly larger scale.

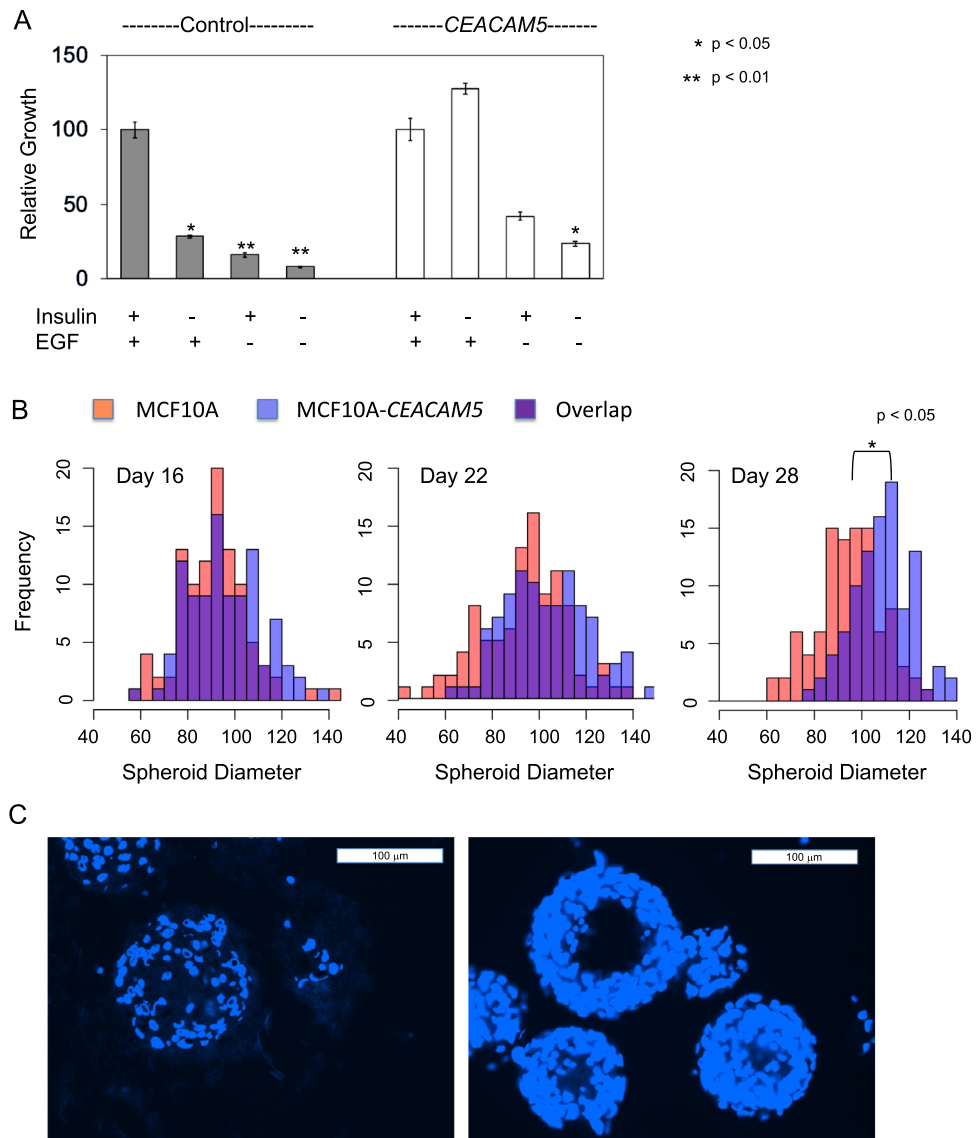
Conducting pooled screens presents notable technical challenges, irrespective of whether involving ORFs, shRNAs, or CRISPR technology. Rigorous quality control measures aimed at minimizing technical variability between biological replicates, genomic DNA purifications, and the generation of amplicons for DNA sequencing demand a high level of technical proficiency. Scaling such endeavors to encompass a broad scope surpasses the capabilities of individual laboratories. Although larger institutions have managed to scale screening efforts across numerous cell lines employing genome-wide shRNA or CRISPR/Cas9 libraries<sup>30,31</sup>, these initiatives have been confined to standard 2D cell culture conditions. Furthermore, parallel screening with ORFs has not been undertaken, and thus far, targets identified have not culminated in the development of new cancer therapeutics. It is our contention that concurrent screening for drivers and dependencies, particularly if incorporating in vivo assays, would be more effective at uncovering authentic cancer therapeutic targets.

A weakness of this study was not performing all three growth or tumorigenicity assays using the same cell line. MCF10A is an immortalized human mammary epithelial cell line with a homozygous deletion of the *CDK2NA* locus. NMuMG is an immortalized murine mammary epithelial cell line harboring a *Trp53* mutation. These two different genetic alterations both inactivate the p53 pathway, so the genetic contexts are similar. But the species are different. Our rationale for choosing MCF10A is that it is the most thoroughly studied normal mammary epithelial cell line, and being human it offered the potential of profiling of protein phosphorylation and protein levels by RPPA. However, normal human cells require six or more genetic alterations to become tumorigenic whereas their murine counterparts only require two<sup>32</sup>. From a practical standpoint, we chose to screen for tumor formation using NMuMG.

Another weakness of this study is the possibility that a given human gene cannot functionally substitute for their murine counterpart. This might explain why some genes promoted growth in MCF10A but not tumor formation in NMuMG, such as IL19 which promoted tumor formation in a previous study<sup>33</sup>.

The impetus for choosing *CEACAM5* for further study was our observation that it was the most potent gene at inducing growth of 3D spheroids. Additionally, although a previous report showed that *CEACAM5* overexpression tumor outgrowth at metastatic sites<sup>18</sup>, it did not show that *CEACAM5* overexpression could induce primary tumor growth and other oncogenic properties associated with pre-metastatic cancer. Our identification of an oncogenic role for *CEACAM5* in primary tumor growth aligns with its histological presence in early breast cancer stages<sup>34</sup>. Also aligning with a role in primary tumor growth is our finding that overexpression allowed for growth in minimal medium without added insulin.

We also found *CEACAM5* overexpression increased 3D spheroid growth of MCF10A cells by forming multiple layers of epithelial cells, while still forming a hollow lumen as normal MCF10A. The combination of these two effects are similar to the hyperproliferation of epithelial cells seen in some early precursor lesions of breast cancer<sup>35</sup>. A similar effect was observed in the colon of transgenic mice with multiple copies of the human locus containing *CEACAM5* and its close homolog *CEACAM6*. In addition to hyperproliferation, the colon of these mice showed increased crypt fission, a process that increases the number of epithelial cell layers<sup>36</sup>. By using RPPA proteomic profiling, we found that *CEACAM5* overexpression activated  $\beta$ -catenin signaling. Activation of  $\beta$ -catenin- signaling may be the mechanism by which *CEACAM5* overexpression causes multiple layers of epithelial cells to form in spheroids. This mechanism would be like the thickening of the intestinal mucosa due

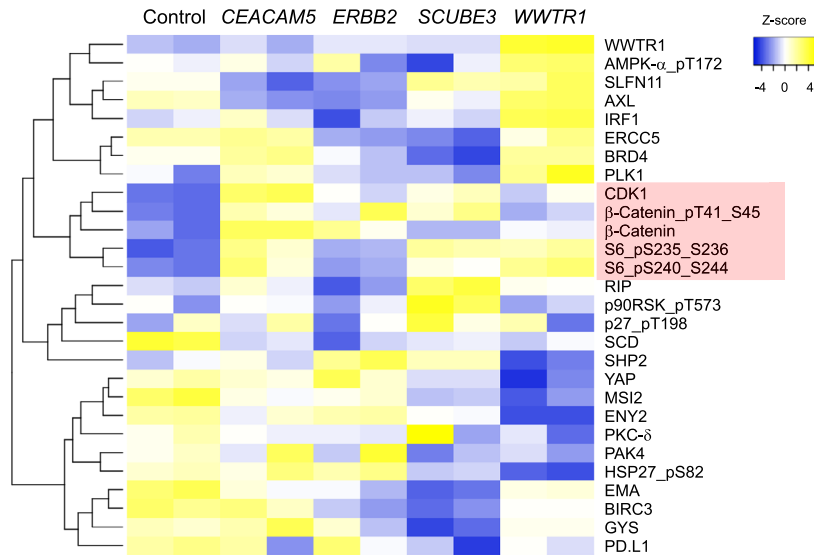


**Figure 6.** Effects of *CEACAM5* on growth-factor requirements and on 3D spheroid growth. Panel (A) Growth of control MCF10A compared to the *CEACAM5*-MCF10 in minimal medium with or without added growth factors (EGF, insulin). All values normalized were to growth with both EGF and insulin. The significance of the findings was assessed using a Student's t-test, while error bars depict the standard error of the means. Panel (B) Effect of *CEACAM5* overexpression on 3D growth in matrigel. MCF10a (red shading) and *CEACAM5*-MCF10A (blue shading) were cultured on matrigel. Acini structures were photographed on days indicated and their diameters were assessed using Leica Q500 MC Qwin software. The significance of the difference in the means of the diameters was assessed using a Student's t-test. Panel (C) Acini structures from MCF10a and MCF10a-*CEACAM5* grown in matrigel were fixed, embedded, and stained with DAPI.

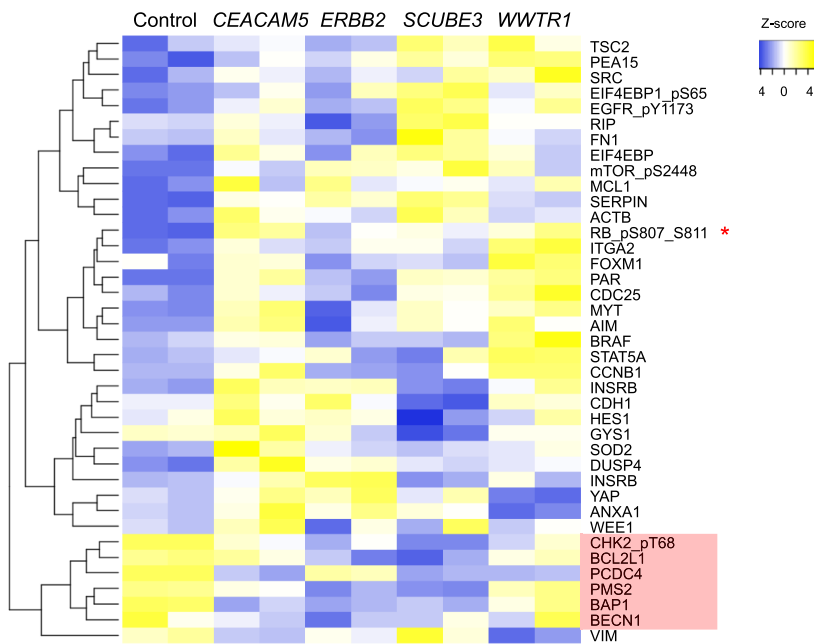
to the accumulation of additional layers of epithelial cells when the Wnt/ $\beta$ -catenin pathway is mutationally activated<sup>37</sup>. Furthermore, RPPA profiling unveiled only a subtle reduction in BCL2L1/Bim expression in MCF10A overexpressing *CEACAM5*. This finding aligns with the previous observation indicating that the formation of MCF10A spheroids with hollowed out interiors relies on Bim's role in initiating apoptosis<sup>38</sup>.

In conclusion, our findings suggest that expanding concurrent screening for drivers and dependencies to encompass thousands of genes could unveil novel therapeutic targets. However, the execution of such large-scale screening surpasses the capabilities of individual laboratories. Furthermore, our study highlights the efficacy of proteomic profiling using RPPA in elucidating the biochemical mechanisms underlying oncogenic phenotypes.





**Figure 7.** Oncogenes induce diverse alterations of protein levels and protein phosphorylation. Heatmap and hierarchical clustering of changes in log<sub>2</sub>-fold protein phosphorylation and total protein levels in different oncogene transformants of MCF10. Normalization was based on the 408 protein measurements made with each sample. The red shading highlights a cluster of proteins whose total levels or phosphorylation are higher in *CEACAM5*-transformants than they are in the other three transformants (*CDK1*,  $\beta$ -Catenin, phosphorylated  $\beta$ -Catenin) or tied for highest with *WWTR1*-transformants (two different phosphorylation sites for ribosomal protein S6).



**Figure 8.** Overexpression of *CEACAM5* induces phosphorylation of retinoblastoma protein and lowers protein levels of tumor suppressors. Heatmap and hierarchical clustering of relative changes in log<sub>2</sub>-fold protein phosphorylation and total protein levels in different oncogene transformants of MCF10. Normalization was based on the 408 protein measurements made with each sample. A red asterisk marks phosphorylated retinoblastoma protein (RB\_pS807\_S811). The red shading indicates a cluster of tumor suppressor proteins whose levels are lower in some or all the oncogene transformants.

## Methods

### Barcoded-ORF library and shRNA library generation

The barcoded-ORF library was constructed by Dr. Kenneth Scott's laboratory at Baylor College of Medicine with the human ORFeome version 8.1 Entry Clone Collection (hORFeome V8) as donors and pLenti6.3/V5 as destination vector using the Gateway recombination-based barcoding pipeline as described<sup>12</sup>. Pooling was done post-lentiviral transduction as described below. The shRNA library pool was designed and constructed at Dr. Michael Hemann's laboratory at the Koch Institute at MIT with six independently designed shRNA for each gene following published procedure using oligonucleotides bulk synthesized by LC Sciences (LC Sciences OligoMix) that were PCR amplified as described<sup>39</sup>. The amplified shRNAs were then batch cloned into the mir30-retroviral vector LMP and then 20 clones of the library were Sanger sequenced to validate the library.

### Pooled screening

The same set of ORFs were lentivirally transduced into both murine non-malignant NMuMG-Myc mammary epithelial cells and human MCF10A cells. To prepare lentivirus, each individual ORF plasmid was transfected separately along with helper plasmids into 293 T cells using jetPRIME transfection reagent (Polyplus Transfection) and virus was collected after 3 days. MCF10A and NMuMG-Myc cells were individually infected for 24 h, followed by addition of fresh media and 48 h before adding blasticidin to select for stable transfectants. After 8 days the individually transfected cells were then pooled and split into four biological replicates which were cultured for an additional 3 days before dividing them into the four conditions: (1) time zero, from which genomic DNA was isolated and used for PCR amplification of the barcodes; (2) 2D cell culture for 30 days (with periodic splitting to avoid reaching confluence); (3) 3D culture of single cells in matrigel for 30 days; and (4) aliquots comprised of  $2 \times 10^6$  cells were injected subcutaneously into nude mice for in vivo tumor formation for 8 weeks. Following genomic DNA isolation and PCR amplification of the barcodes, each PCR amplicon product was ligated to a unique index adaptor for construction of a multiplexed Illumina library using the TrueSeq DNA LT Sample Prep Kit (Illumina) for MiSeq sequencing. After mapping reads to the set of utilized barcode sequences, counts for each ORF were normalized by the number of total numbers of mapped reads per amplicon. Following log<sub>2</sub> transformation, the four values from the endpoint samples were compared to the corresponding initial values from the zero time point samples as described<sup>12</sup>.

For shRNA screening, the LMP shRNA plasmid library and helper plasmid were transfected into Bing packaging cells with Profection Mammalian Transfection System. After 48 h, the viral supernatant was collected and filtered. Breast cancer cell lines MCF7, Cal120, MDA-MB-453 were infected at a MOI of 0.3 using diluted viral supernatant supplemented with 8 µg/ml polybrene. 72 h after the infection, infected cells were selected using puromycin for 3 days. The cells were then split into separate cultures for generating quadruple replicates for zero time point cultures and quadruple replicates for two-weeks growth in vitro. The cells were collected and genomic DNA was extracted using E.Z.N.A Tissue DNA Kit (Omega Bio-Tek). PCR reactions was done using specific primers: half forward (5'-TAGTGAAGCCACAGATGTA) and bc1-reverse (5'-AAAGCGCATGTCCAGACTGCC) to amplify hairpin sequence. Each PCR amplicon was prepared for Illumina Miseq library using the TrueSeq DNA LT Sample Prep Kit (Illumina), where each amplicon was ligated to unique index Adaptors for MiSeq sequencing. Following normalization and log<sub>2</sub> transformation, the quadruple values from the endpoint samples were compared to the corresponding initial values from the zero time point samples as described<sup>39</sup>.

### Animals

One million mammary epithelial cells were trypsinized, resuspended in 100 µl DMEM and injected subcutaneously or orthotopically into mammary pads of four- to five-week-old female nude mice (NCR nu/nu; Envigo Inc., Wilmington, MA). Growth was followed over time by taking caliper measurements at indicated time points. Tumor volume was measured as  $0.52 \times \text{length} \times \text{width}^2$ . Prior to tumor excision, mice were euthanized by carbon dioxide asphyxiation. Tumors were excised six-eight weeks post injections or when one of the measurements reached 2 cm.

### Tissue culture and lentiviral infection

MCF10A, T47D, HEK293T, MCF7, MCF10A, NMuMG cells were obtained from the American Type Culture Collection (ATCC). HEK293T and T47D were cultured in DMEM with 10% fetal bovine serum (FBS) and 100 U/ml penicillin streptomycin. MCF10A cells were cultured in DMEM/F12 supplemented with 5% donor horse serum, EGF 20 ng/mL, insulin 10 µg/mL, hydrocortisone 100 µg/mL cholera toxin 10 ng/mL and 100 U/ml penicillin streptomycin. MCF7 and NMuMG-myc cells were cultured in DMEM with 10% fetal bovine serum (FBS), insulin 10 µg/mL and 100 U/ml penicillin streptomycin. To produce lentiviral particles, HEK293T cells were transiently transfected with transfer, envelope, and packaging plasmids using Profection Mammalian Transfection System (Promega). The viral supernatant was collected 24 h later. For lentiviral infection, MCF10A cells were grown in a 1:4 mixture of the growth medium and the viral supernatant for 24 h. Two days post infection, blasticidin (Invivogen) selection was performed for 8 days to select for positive clones and these cells were used for various assays. Colony formation assays were performed by plating 1000 cells in triplicate on 6-well plates. Medium was changed every 3 days. After 3 weeks culture, cells were methanol fixed and stained with 0.5% crystal violet. After pictures were taken, crystal violet was dissolved with 0.1% SDS overnight. Dissolved crystal violet staining was read at 595 nm using the Victor 3 machine (Perkin Elmer). 3D culture was performed as previously reported<sup>40</sup> using growth factor-reduced matrigel (BD Bioscience).

## RPPA and immunoblotting

RPPA was performed by the MD Anderson RPPA Core Facility using protein lysates. For each RPPA replicate, cells were washed with cold PBS twice and lysed with 100  $\mu$ L lysis buffer (1% Triton X-100, 50 mM HEPES, pH 7.4, 150 mM NaCl, 1.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 100 mM NaF, 10 mM Na pyrophosphate, 1 mM Na<sub>3</sub>VO<sub>4</sub> and 10% glycerol, containing freshly added protease and phosphatase inhibitors from Roche Applied Science Cat. # 05056489001 and 04906837001, respectively), and mixed with 4  $\times$  SDS sample buffer (40% Glycerol, 8% SDS, 0.25 M Tris-HCl, pH 6.8). Immunoblotting was carried out using antibodies against Ceacam5 (Sigma), Akt, phospho AKT (Cell Signaling and Actin (Abcam) or GAPDH (Santa Cruz) as loading control. Cells were lysed in lysis buffer (25 mM Tris, pH 7, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, 1% Triton X 100) and cocktail protease inhibitors (Roche) and centrifuged to clear the cell lysates. The cell lysates were then loaded on a 10% SDS-PAGE, and the proteins were subsequently transferred onto a nitrocellulose membrane. The membrane was blocked with 5% milk in TBST (20 mM Tris-HCl, pH7.5, 150 mM NaCl, 0.1% Tween-20) for 30 min and then incubated with antibodies in 5% milk in TBST overnight. After three extensive washes for 10 min each in TBST, the membrane was incubated with the secondary antibody, horseradish peroxidase (HRP)-conjugated anti-rabbit (Santa Cruz) or anti-mouse IgG antibody (Abcam) for 1 h, followed by three washes in TBST. The membrane was developed with the chemiluminescent kit (Biorad) and exposed to X-ray films (Kodak).

## Other methods

For real-time RT-PCR, total mRNA was isolated from the indicated cells using the RNeasy Mini Kit (Qiagen). cDNAs were synthesized using the cDNA Synthesis Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The transcript level of GAPDH was used as an internal control. The amplification steps consisted of 10 min at 50 °C and 5 min at 95 °C, followed by 40 cycle of denaturation for 10 s at 95 °C and annealing/extension for 30 s at 60 °C. All samples were analyzed in triplicate, and the relative gene expression was calculated according to the comparative threshold cycle ( $\Delta\Delta$ Ct) method.

Student's t-test was used to determine the significance of difference of means measurements, using the standard  $p < 0.05$  cutoff. Error bars represent the standard deviation of data.

For analysis of 3D Matrigel cultures, spheroids were photographed and their diameters were assessed using Leica Qwin Standard V2.6 software (<https://www.leica-microsystems.com/>). For immunofluorescence staining, acini were fixed in 4% paraformaldehyde, washed in PBS, and embedded in optimum cutting temperature (OCT) compound (Tissue-Tek) and then processed by regular histological procedures. 4', 6-diamidino-2-phenylindole (DAPI) (Molecular Probes) was used as DNA stain.

## Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 30 January 2024; Accepted: 6 June 2024

Published online: 09 June 2024

## References

- Bhattacharya, A. *et al.* Transcriptional effects of copy number alterations in a large set of human cancers. *Nat. Commun.* **11**(1), 715 (2020).
- Santarius, T., Shipley, J., Brewer, D., Stratton, M. R. & Cooper, C. S. A census of amplified and overexpressed human cancer genes. *Nat. Rev. Cancer* **10**(1), 59–64 (2010).
- Romagosa, C. *et al.* p16Ink4a overexpression in cancer: A tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene* **30**(18), 2087–2097 (2011).
- Wu, D. *et al.* MZF1 mediates oncogene-induced senescence by promoting the transcription of p16(INK4A). *Oncogene* **41**(3), 414–426 (2022).
- Moore, A. R., Rosenberg, S. C., McCormick, F. & Malek, S. RAS-targeted therapies: Is the undruggable drugged?. *Nat. Rev. Drug Discov.* **19**(8), 533–552 (2020).
- Pratilas, C. A. *et al.* V600EBRAF is associated with disabled feedback inhibition of RAF-MEK signaling and elevated transcriptional output of the pathway. *Proc. Natl. Acad. Sci.* **106**(11), 4519–4524 (2009).
- Liu, Y., Sun, J. & Zhao, M. ONGene: A literature-based database for human oncogenes. *J. Genet. Genom.* **44**(2), 119–121 (2017).
- Legut, M. *et al.* A genome-scale screen for synthetic drivers of T cell proliferation. *Nature* **603**(7902), 728–735 (2022).
- Lu, X. *et al.* An in vivo screen identifies PYGO2 as a driver for metastatic prostate cancer. *Cancer Res.* **78**(14), 3823–3833 (2018).
- Grzeskowiak, C. L. *et al.* In vivo screening identifies GATAD2B as a metastasis driver in KRAS-driven lung cancer. *Nat. Commun.* **9**(1), 2732 (2018).
- Naxerova, K. *et al.* Integrated loss- and gain-of-function screens define a core network governing human embryonic stem cell behavior. *Genes Dev.* **35**(21–22), 1527–1547 (2021).
- Tsang, Y. H. *et al.* Functional annotation of rare gene aberration drivers of pancreatic cancer. *Nat. Commun.* **7**, 1–10 (2016).
- Pires, M. M., Hopkins, B. D., Saal, L. H. & Parsons, R. E. Alterations of EGFR, p53 and PTEN that mimic changes found in basal-like breast cancer promote transformation of human mammary epithelial cells. *Cancer Biol. Ther.* **14**(3), 246–253 (2013).
- Termen, S., Tan, E. J., Heldin, C. H. & Moustakas, A. p53 regulates epithelial-mesenchymal transition induced by transforming growth factor beta. *J. Cell Physiol.* **228**(4), 801–813 (2013).
- Sawey, E. T. *et al.* Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by oncogenomic screening. *Cancer Cell* **19**(3), 347–358 (2011).
- Xiong, X. *et al.* CXCL8 in tumor biology and its implications for clinical translation. *Front. Mol. Biosci.* **9**, 723846 (2022).
- Tian, Q. *et al.* RICH1 inhibits breast cancer stem cell traits through activating kinases cascade of Hippo signaling by competing with Merlin for binding to Amot-p80. *Cell Death Dis.* **13**(1), 71 (2022).
- Powell, E. *et al.* A functional genomic screen in vivo identifies CEACAM5 as a clinically relevant driver of breast cancer metastasis. *NPJ Breast Cancer* **4**, 9 (2018).
- Dinavahi, S. S., Bazewicz, C. G., Gowda, R. & Robertson, G. P. Aldehyde dehydrogenase inhibitors for cancer therapeutics. *Trends Pharmacol. Sci.* **40**(10), 774–789 (2019).

20. Chia, L. *et al.* HMGA1 induces FGF19 to drive pancreatic carcinogenesis and stroma formation. *J. Clin. Invest.* **133**(6), 1601 (2023).
21. Pavlova, N. N. *et al.* A role for PVRL4-driven cell-cell interactions in tumorigenesis. *Elife* **2**, e00358 (2013).
22. Corpuz, A. D., Ramos, J. W. & Matter, M. L. PTRH2: An adhesion regulated molecular switch at the nexus of life, death, and differentiation. *Cell Death Discov.* **6**(1), 124 (2020).
23. Louie, M. C., Zou, J. X., Rabinovich, A. & Chen, H. W. ACTR/AIB1 functions as an E2F1 coactivator to promote breast cancer cell proliferation and antiestrogen resistance. *Mol. Cell Biol.* **24**(12), 5157–5171 (2004).
24. Goyama, S. *et al.* Transcription factor RUNX1 promotes survival of acute myeloid leukemia cells. *J. Clin. Invest.* **123**(9), 3876–3888 (2013).
25. Klinakis, A. *et al.* A novel tumour-suppressor function for the Notch pathway in myeloid leukaemia. *Nature* **473**(7346), 230–233 (2011).
26. Bartlett, J. M. *et al.* Mammostrat as an immunohistochemical multigene assay for prediction of early relapse risk in the tamoxifen versus exemestane adjuvant multicenter trial pathology study. *J. Clin. Oncol.* **30**(36), 4477–4484 (2012).
27. Ender, P. *et al.* Spatiotemporal control of ERK pulse frequency coordinates fate decisions during mammary acinar morphogenesis. *Dev. Cell* **57**(18), 2153–2167 (2022).
28. Akbani, R. *et al.* A pan-cancer proteomic perspective on the cancer genome atlas. *Nat. Commun.* **5**, 3887 (2014).
29. Yasukawa, M. *et al.* CDK1 dependent phosphorylation of hTERT contributes to cancer progression. *Nat. Commun.* **11**(1), 1557 (2020).
30. Tsherniak, A. *et al.* Defining a cancer dependency map. *Cell* **170**(3), 564–576 (2017).
31. Pacini, C. *et al.* A comprehensive clinically informed map of dependencies in cancer cells and framework for target prioritization. *Cancer Cell* **42**(2), 301–316 (2024).
32. Rangarajan, A., Hong, S. J., Gifford, A. & Weinberg, R. A. Species- and cell type-specific requirements for cellular transformation. *Cancer Cell* **6**(2), 171–183 (2004).
33. Hsing, C. H. *et al.* Upregulated IL-19 in breast cancer promotes tumor progression and affects clinical outcome. *Clin. Cancer Res.* **18**(3), 713–725 (2012).
34. Bechmann, M. B., Brydholm, A. V., Codony, V. L., Kim, J. & Villadsen, R. Heterogeneity of CEACAM5 in breast cancer. *Oncotarget* **11**(43), 3886–3899 (2020).
35. Allred, D. C., Mohsin, S. K. & Fuqua, S. A. Histological and biological evolution of human premalignant breast disease. *Endocr. Relat. Cancer* **8**(1), 47–61 (2001).
36. Chan, C. H., Camacho-Leal, P. & Stanners, C. P. Colorectal hyperplasia and dysplasia due to human carcinoembryonic antigen (CEA) family member expression in transgenic mice. *PLoS ONE* **2**(12), e1353 (2007).
37. Kretzschmar, K. & Clevers, H. Wnt/beta-catenin signaling in adult mammalian epithelial stem cells. *Dev. Biol.* **428**(2), 273–282 (2017).
38. Reginato, M. J. *et al.* Bim regulation of lumen formation in cultured mammary epithelial acini is targeted by oncogenes. *Mol. Cell Biol.* **25**(11), 4591–4601 (2005).
39. Meacham, C. E. *et al.* A genome-scale in vivo loss-of-function screen identifies Phf6 as a lineage-specific regulator of leukemia cell growth. *Genes Dev.* **29**(5), 483–488 (2015).
40. Debnath, J., Muthuswamy, S. K. & Brugge, J. S. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**(3), 256–268. [https://doi.org/10.1016/s1046-2023\(03\)00032-x](https://doi.org/10.1016/s1046-2023(03)00032-x) (2003).

## Acknowledgements

The authors thank members of the Powers and Hemann labs for discussions. The authors also thank the staff in the Biostatistics & Data Science Shared Resources of the Stony Brook Cancer Center.

## Author contributions

A.M. performed the majority of experiments and analyzed data. E.C., C.B., and M.H. designed and constructed the shRNA library. F.C. and X.Z. performed experiments. J.L. analyzed data. S.P. conceived of the study, performed experiments, analyzed data, and wrote the manuscript. All authors reviewed the manuscript. All authors have consented to publication.

## Funding

This research was supported by the NIH under Award Numbers R01CA217206 and U01CA168409, a Peter Rowley Breast Cancer Research Award from the New York State Department of Health, the Stony Brook Cancer Center, the Babylon Breast Cancer Coalition, the Carol M. Baldwin Breast Cancer Research Fund, and the Walk-for-Beauty Foundation Award. The Functional Proteomics RPPA Core facility is supported by MD Anderson Cancer Center Support Grant # 5 P30 CA016672-40. All studies utilizing mice were approved by and in accordance with Stony Brook University Institutional Animal Care and Use Committee and in accordance with ARRIVE guidelines (<https://arriveguidelines.org>). We acknowledge the Stony Brook Cancer Center Biostatistics & Bioinformatics Shared Resource for expert assistance with bioinformatic analysis. All data and generated materials (plasmids) are available upon request.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-64297-w>.

**Correspondence** and requests for materials should be addressed to S.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024