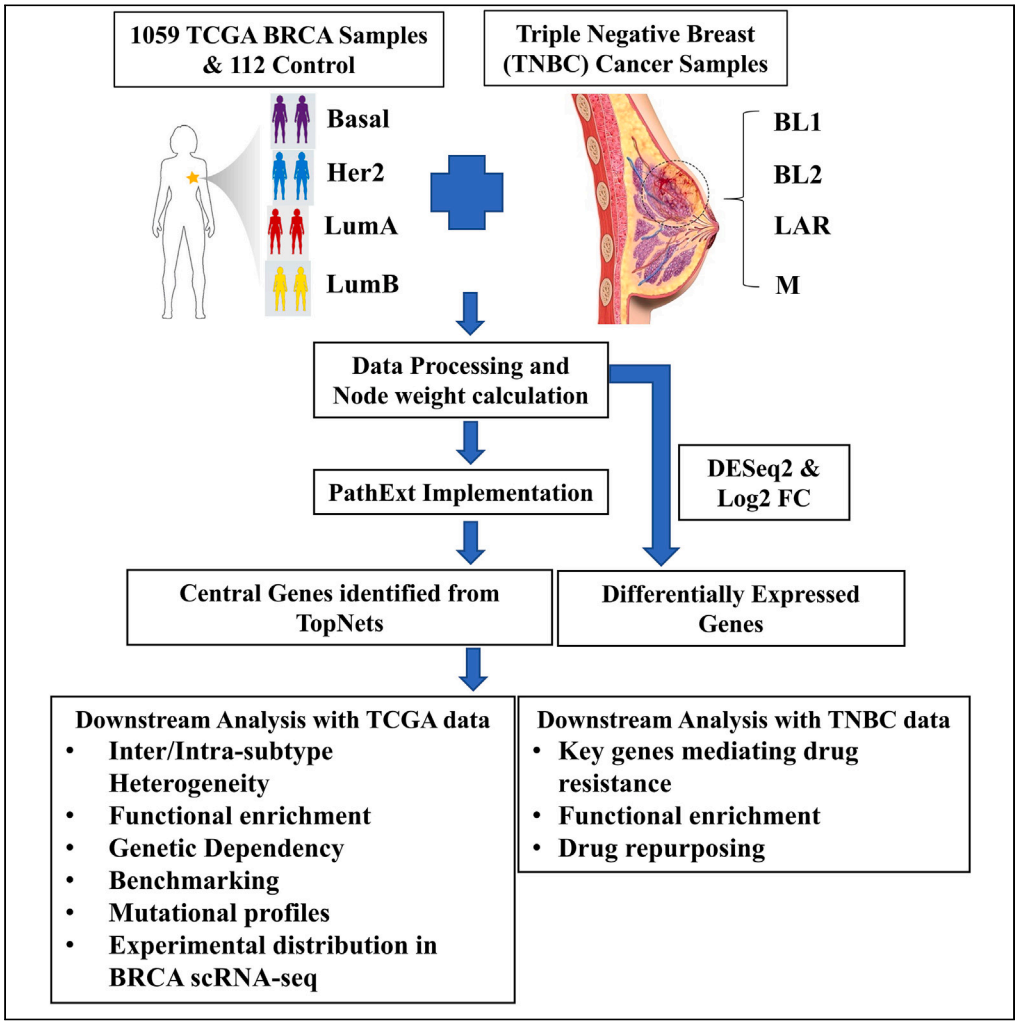
iScience

CellPress
OPEN ACCESS

Article

Network-based approach elucidates critical genes in BRCA subtypes and chemotherapy response in triple negative breast cancer



Piyush Agrawal, Navami Jain, Vishaka Gopalan, Annan Timon, Arashdeep Singh, Padma S. Rajagopal, Sridhar Hannenhalli

apiyush74@gmail.com (P.A.)
sridhar.hannenhalli@nih.gov (S.H.)

Highlights

Applied our network-based tool PathExt to TCGA 1,059 BRCA tumors and 112 normal samples

Revealed subtype-specific biological processes compared to standard DEGs methods

Single-cell analysis unveils PathExt identified genes in tumor microenvironment

Identifies TNBC subtype-specific genes associated with resistance to chemotherapy

Agrawal et al., iScience 27, 109752
May 17, 2024 Published by Elsevier Inc.
https://doi.org/10.1016/j.isci.2024.109752

iScience

CellPress
OPEN ACCESS

Article

Network-based approach elucidates critical genes in BRCA subtypes and chemotherapy response in triple negative breast cancer

Piyush Agrawal,[1,4,*] Navami Jain,[2] Vishaka Gopalan,[1] Annan Timon,[3] Arashdeep Singh,[1] Padma S. Rajagopal,[1] and Sridhar Hannenhalli[1,*]

SUMMARY

Breast cancers (BRCA) exhibit substantial transcriptional heterogeneity, posing a significant clinical challenge. The global transcriptional changes in a disease context, however, are likely mediated by few key genes which reflect disease etiology better than the differentially expressed genes (DEGs). We apply our network-based tool PathExt to 1,059 BRCA tumors across 4 subtypes to identify key mediator genes in each subtype. Compared to conventional differential expression analysis, PathExt-identified genes exhibit greater concordance across tumors, revealing shared and subtype-specific biological processes; better recapitulate BRCA-associated genes in multiple benchmarks, and are more essential in BRCA subtype-specific cell lines. Single-cell transcriptomic analysis reveals a subtype-specific distribution of PathExt-identified genes in multiple cell types from the tumor microenvironment. Application of PathExt to a TNBC chemotherapy response dataset identified subtype-specific key genes and biological processes associated with resistance. We described putative drugs that target key genes potentially mediating drug resistance.

INTRODUCTION

Breast cancer (BRCA) is one of the leading cancers among women worldwide.[1] In 2023, 300,500 are expected to be diagnosed and 43,700 are expected to die from the disease [https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html]. Triple negative breast cancers (TNBCs), defined clinically by the lack of estrogen receptor (ER), progesterone receptor (PR), or human epidermal growth factor receptor (HER2) on immunohistochemistry[2] are especially aggressive with limited therapeutic options.[3] Gene expression is used, primarily in the translational context, to characterize BRCA subtypes: luminal A and luminal B (which overlap with hormone-receptor positive breast cancers), HER2-enriched (which overlaps with HER2-positive tumors), and basal-like (which only partially overlaps with TNBC).[4,5]

To identify potential targets for BRCA treatment, previous studies have relied on differentially expressed genes (DEGs). However, individual genes are subject to stochastic variability in gene expression, limiting their reproducibility and consistency.[6,7] Furthermore, transcriptional changes are often due to perturbations of key mediators (such as transcription factors, kinases, and other regulatory proteins) in a complex gene regulatory network. Identifying these potential mediators is more likely to provide mechanistic insights and therapeutic targets. Network-based models have been shown to improve potential target identification in various cancers.[8–12] For example, Choi et al.[12] have shown that a network-based model could identify potential targets and elucidated the underlying mechanisms of reprogramming basal-like cancer cells into luminal A cells. This analysis identified BCL11A and HDAC1/2 as optimal targets that can induce basal-like breast cancer reprogramming and endocrine therapy sensitivity. Likewise, Pe'er and Hacohen[8] outlined fundamental principles for constructing network models in cancer, emphasizing the integration of multi-omics data to capture complex molecular interactions. This integrative approach enables a more comprehensive understanding of the disease, facilitating the identification of key targets. Moreover, Mundi et al.[10] developed a transcriptome-based precision oncology platform that leverages network analysis to align patients with optimal therapies across a spectrum of treatment-resistant malignancies.

In contrast with DEGs, our previously developed network-based tool PathExt[13] (https://github.com/NarmadaSambaturu/PathExt) identifies differentially active pathways in a knowledge-based network and the corresponding central genes in the subnetwork (called TopNet) composed of the differentially active paths. Central genes identified by PathExt offer a more robust and mechanistic view of transcriptional changes across conditions compared to DEGs, explaining differential expression of downstream genes.

[1]Cancer Data Science Lab, National Cancer Institute, NIH, Bethesda, MD, USA
[2]Stanford University, Stanford, CA, USA
[3]University of Pennsylvania, Philadelphia, PA, USA
[4]Lead contact
*Correspondence: apiyush74@gmail.com (P.A.), sridhar.hannenhalli@nih.gov (S.H.)
https://doi.org/10.1016/j.isci.2024.109752

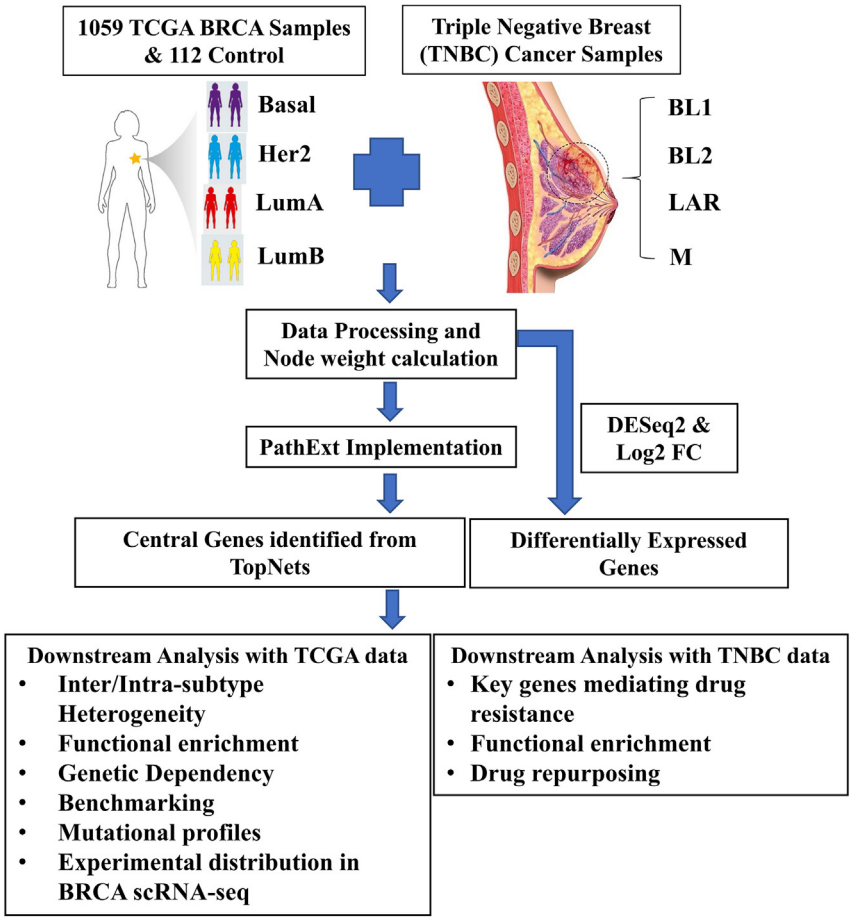


Figure 1. Architecture of the study

PathExt was implemented on the TCGA-BRCA dataset to identify subtype-specific most frequent central mediating genes as well as DEGs for comparison. Central genes were used for various analyses such as identifying enriched biological processes, genetic dependencies of the cell, identifying cell types mediating global transcriptomic response in single cell data, recapitulating breast cancer driver genes, FDA-approved breast cancer targets, comparison with previous network-based methods and DEGs. PathExt was also implemented on the TNBC dataset to identify key genes associated with non-response to neoadjuvant chemotherapy treatment in subtype-specific manner.

We applied PathExt framework to the TCGA BRCA transcriptional data (tumor and normal samples)[14] to identify key genes mediating the global transcriptional changes in each BRCA subtype. Against multiple benchmark datasets, PathExt-identified genes substantially outperform DEGs in recapitulating potential driver genes and perform comparably to another network-based approach, MOMA.[15] PathExt-identified genes are further validated by their effect on cellular viability, based on CRISPR knock-out data from the DepMap database,[16,17] as well as their greater-than-expected mutation frequencies in TCGA samples, in a subtype-specific manner (note that the PathExt relies only on the transcriptome data and does not utilize the mutational data). We further applied PathExt in a clinical trial of neoadjuvant chemotherapy in TNBC to identify key genes associated with treatment-resistant tumors. Lastly, based on computational drug screening, we propose potential therapeutic strategies for targets identified by PathExt. Overall, PathExt refined prior characterization of inter-tumor heterogeneity and identified more consistent genes associated with BRCA subtypes, as well as candidate genes to study resistance to neoadjuvant chemotherapy in TNBC.

RESULTS

Overview of the workflow

The overall workflow of the study is shown in Figure 1. Given a BRCA transcriptomic profile and control samples, as well as a knowledge-based protein interaction network, PathExt identifies key genes likely to mediate the observed global transcriptomic changes in the specific sample relative to the control samples. We separately identify genes mediating up-regulation (based on Activated TopNet) as well as down-regulation (based on Repressed TopNet) of gene expression. We applied PathExt to 1,059 TCGA BRCA transcriptomic samples across 4 BRCA subtypes using 112 healthy controls to identify top 100 key mediator genes in each sample (STAR Methods). We further identified the

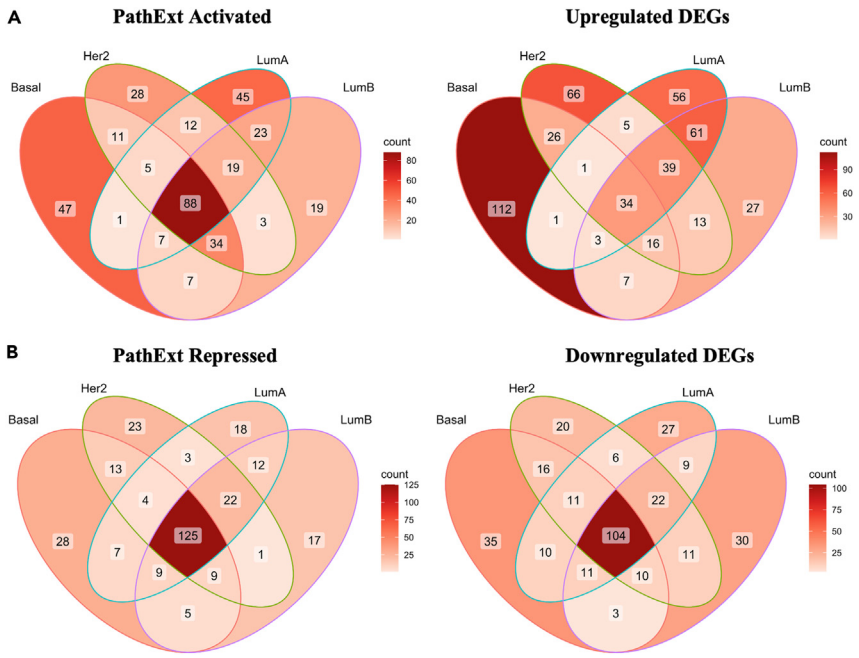


Figure 2. Gene overlap analysis across BRCA subtypes

Venn diagram showing gene overlap among central genes in various subtypes identified in (A) PathExt Activated TopNets and Upregulated DEGs and for (B) PathExt Repressed TopNets and Downregulated DEGs.

top 200 genes most frequently identified across samples in each cancer subtype as a key mediator. As a control, we followed an analogous procedure to identify the top 200 most frequent up and downregulated DEGs in each subtype. Complete lists of genes and their frequencies are provided in the Tables S1–S4. We performed a series of downstream analyses to assess the relative merits of these identified genes in terms of their functional roles in BRCA. We additionally analyzed a TNBC dataset to identify central genes associated with nonresponse to neoadjuvant chemotherapy in various TNBC subtypes.

PathExt reveals putative BRCA subtype-specific and shared mechanisms more consistently than DEGs

The following analyses are based on the PathExt-identified top 200 most frequent central genes in Activated and Repressed TopNets for each subtype (Table S5) and the analogously calculated top 200 most frequent upregulated and downregulated DEGs (Table S6). These top 200 genes are selected based on the frequency of each gene across samples within each subtype.

Global inter-sample transcriptional heterogeneity confounds clinical research. By focusing on key upstream mediators, rather than the downstream affected genes, PathExt promises to identify genes that are more likely to be shared across samples. Consistent with this expectation, we observed that PathExt genes are much more frequently represented within subtype-specific samples than DEGs, suggesting that PathExt may better reveal shared mechanisms across tumors. For example, in the basal subtype, top Activated PathExt genes *CDC20* and *TTK* were key genes in 175/177(~99%) tumors, while the most frequently upregulated DEG—*CT83* was the top DEG in only 117/177(~66%) tumors. Likewise, in the HER2 subtype, out of 80 samples, PathExt identified *ERBB2* (encoding HER2) as the central gene in 42 samples (>50%) whereas it was differentially upregulated in only 11 patients (~15%), underscoring the potential of PathExt in capturing functionally relevant genes. Identification of such broadly relevant genes will provide insight into the mechanisms governing various biological and cellular processes, disease progression and potential therapeutic targets.

PathExt genes are also much more frequently represented than DEGs when comparing across all breast cancer samples, as shown in Figure 2A for Activated TopNets and Figure 2B for Repressed TopNets. Direct comparison of genes prioritized by PathExt and DEGs also revealed substantial differences (Figures S1A and S1B).

We next identified the enriched biological processes associated with pan-subtype and subtype-specific genes. For the PathExt Activated TopNets, as expected, the common genes were associated mostly with cell cycle (Figure 3A, Table S7). However, subtype-specific genes were enriched for distinct processes that are largely supported by literature. For instance, the Basal subtype was associated with cell fate determination, regulation of T cell proliferation, neuron differentiation, interferon-gamma production, etc.[18–21] [Figure 3B]; enriched processes for other subtypes are summarized in Figures S2A and S2B (except LumB, where no enrichment was found), and the complete list of enriched processes is provided in the Tables S8–S10. Among top genes in Repressed TopNets, most subtypes share similar processes, including neuropeptide signaling pathway, cellular hormone metabolic pathway, muscle cell development, etc. (Figure 3C) whereas unique genes for the Basal subtype were enriched for the processes such as terpenoid metabolic process, liver development, etc. (Figure 3D). For remaining

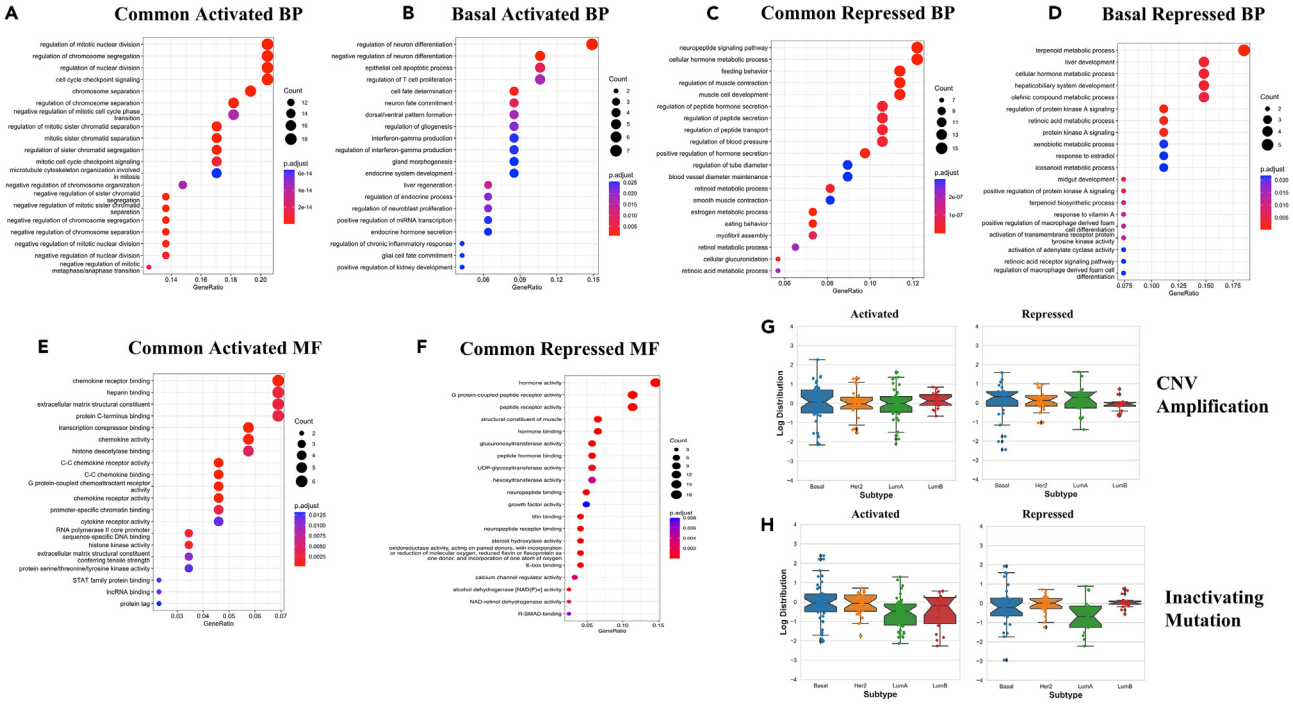


Figure 3. PathExt central genes are associated with essential biological processes, molecular functions and subtype-specific mutational properties

Top 20 enriched biological processes associated with activated common central genes among all the 4 BRCA subtypes (A); Top 20 enriched biological processes associated with activated genes uniquely in Basal (B); Top 20 enriched biological processes associated with repressed common central genes among all the 4 BRCA subtypes (C); and Top 20 enriched biological processes associated with repressed genes uniquely in Basal (D); Top 20 enriched molecular functions associated with activated common central genes among all the 4 BRCA subtypes (E); and, Top 20 enriched molecular functions associated with repressed common central genes among all the 4 BRCA subtypes (F); Boxplot representation of subtype-specific PathExt Activated and Repressed TopNets unique gene CNV amplification distribution (G); and subtype-specific PathExt Activated and Repressed TopNets unique gene inactivating mutation distribution (H). Here "Gene ratio" in the Figure 3(A–F) is the fraction of the input gene set in the given GO term.

subtypes, see Figures S2C–S2E, and the complete list of enriched processes is provided in the Tables S11–S15. In contrast, when analyzing pan-subtype and subtype-specific DEGs, distinct, but fewer, sets of processes were enriched. Notably, for many subtypes we didn't observe any significant enriched processes associated with up and downregulated DEGs; all results are provided in Figure S3, Tables S16–S21.

Enriched molecular functions for Activated TopNets of pan-subtype genes include chemokine binding, extracellular matrix structural constituent, and transcription corepressor binding (Figure 3E). Subtype-specific genes (except Her2, where no enrichment was found) also revealed subtype-specific molecular functions supported by literature[22–25] (Figures S4A–S4C; Tables S22–S25). Likewise, for the Repressed TopNets, molecular functions enriched among pan-subtype genes include hormone activity, glycosyltransferase activity, and titin binding. (Figure 3F). Enriched functions among subtype-specific PathExt genes (except LumB, where no enrichment was found) included somewhat related functions, e.g., G protein-coupled peptide receptor activity[26] (Figures S4D–S4F, Tables S26–S29). Pan-subtype and subtype-specific DEGs revealed processes only partly overlapping with PathExt (Figure S5, Tables S30–S34). Notably, for many subtypes, we did not observe any significant enriched molecular functions associated with up and downregulated DEGs.

Overall, PathExt reveals greater commonality across samples and across subtypes and identifies many shared and subtype-specific genes and processes mediating the global transcriptome shifts that are missed by the conventional DEG analysis, underscoring the complementary value of PathExt approach.

Next, we assessed the extent to which the subtype-specific genes identified by PathExt have any support for their subtype-specific functionality. We did this with respect to subtype-specific expression and mutational patterns of those genes. First, for every gene, in a given subtype, we computed the log-fold difference of that gene's expression in the subtype relative to other subtypes. Unsurprisingly, subtype-specific central activated genes show higher relative gene expression in the particular BRCA subtype, while central repressed genes exhibit lower relative expression in the BRCA subtype (Figure S6).

Given the potentially central roles of PathExt-identified genes, we assessed whether they exhibit subtype-specific enrichment for CNV amplifications (we refer to these as activating mutation). Note that we expect this for the central genes in both Activating and Repressed TopNets. We separately computed the frequency of activating (copy number amplification; CNA) and inactivating (copy number loss, nonsense, and frameshift indel) mutations in a subtype-specific manner. We compared the mutation frequency of the subtype-specific unique genes to the average frequency across all genes in the particular BRCA subtype and then normalized it based on the same quantity for the

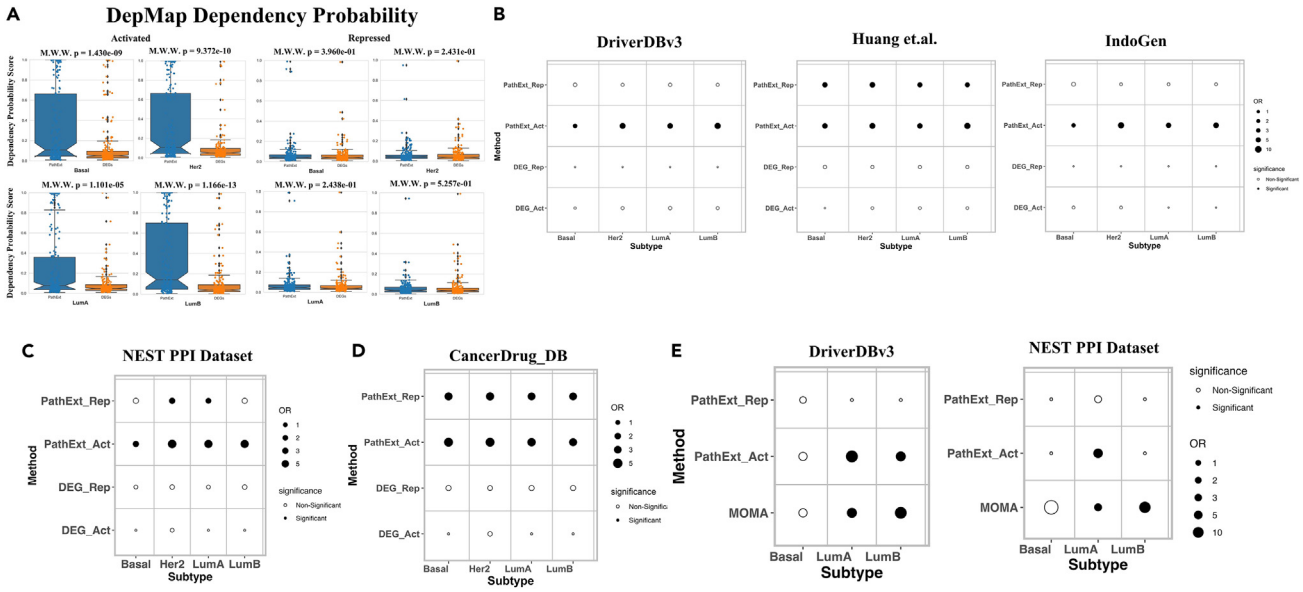


Figure 4. Evaluation of PathExt approach with DEGs and MOMA

Boxplot representation of gene dependency probability distribution by PathExt Activated TopNet genes and Upregulated DEGs and PathExt Repressed TopNet genes and downregulated DEGs in various BRCA subtypes (A); Fishers odds ratio (OR) showing that PathExt recapitulated BRCA specific driver genes from multiple datasets (B); NEST PPI mutated modules comprising a number of genes affecting various mechanisms (C); targets associated with FDA-approved BRCA drugs compared to DEGs with high statistical significance (D). PathExt showed comparable performance with MOMA (another network-based method) on various datasets (E).

other three subtypes. This provided us the subtype-specific relative log-normalized mutation frequency. While the signals were weak owing to the sparsity of mutations, we broadly observed a subtype-specific enrichment of activating mutations in subtype-specific genes (both from Activated and Repressed TopNets) (Figure 3G) and a depletion of inactivating mutations (Figure 3H). Supported by previously reported dominance of CNAs in BRCA genomic landscape[27] our results link CNAs to key genes affecting the global transcriptome in a BRCA subtype-specific fashion. Overall, our analyses support the role of key genes identified by PathExt in BRCA subtype-specific pathogenesis.

PathExt-identified genes perform better than DEGs in cell line, functional, and benchmarking tests

The key feature of PathExt is identification of genes that may not be overtly differentially expressed but may play a central role in mediating the tumor phenotype. First, we assess the extent to which subtype-specific PathExt genes exhibit subtype-specific function and clinical relevance. To do this, we utilized the publicly available DepMap dataset reporting cellular viability upon genome-wide CRISPR-Cas9 knock out (KO) across hundreds of cell lines, providing a dependency probability score for each gene in each cell line, where a higher score indicates greater essentiality or dependency[17]; by default, a score >0.5 indicates dependency. Critically, one can obtain dependency for each gene in BRCA subtype-specific cell lines. For each gene-subtype pair, we computed the gene's average dependency probability scores across the cell lines derived from the specific BRCA subtype. We then compared the score distribution for PathExt genes with that for DEGs. As shown in Figure 4A, PathExt Activated genes exhibited significantly greater dependency in subtype-specific cell lines compared to upregulated DEGs. Interestingly, gene dependency of the Repressed TopNets and downregulated DEGs was very low suggesting that these genes do not affect cell survival. Our conclusions drawn from Figure 4A do not change when we compare the fraction of essential genes (dependency probability score >0.5) between PathExt and DEGs based on Fisher's test (Figure S7). Depmap gene dependency probability value for the central genes (when available) of PathExt and DEGs in various subtypes is provided in the Tables S35–S38.

Next, we assessed how significantly PathExt and DEGs recapitulated previously compiled driver genes in three 3 publicly available datasets: (i) DriverDBv3[28] (including 313 BRCA driver genes), (ii) IntoGen[29] (including 99 BRCA driver genes), and (iii) Huang et al. dataset[30] (including 500 BRCA driver genes). For each benchmark gene set, we assessed the significance of overlap (using Fisher exact test) with the top 200 PathExt genes or DEGs in activated or repressed scenarios. As shown in Figure 4B, in all benchmarks, PathExt Activated TopNet central genes show significant overlap with the BRCA drivers, while upregulated DEGs do not.

Integrating protein interaction networks and tumor mutation profiles, Zheng et al. have reported a map of 395 protein systems (NEST) that are recurrently mutated in one or more cancer types likely under somatic mutation selection.[31] PathExt genes are associated significantly with several NEST systems, including cell cycle, nucleosome and ribosome, cytoplasm and extracellular space, and signaling systems. More specifically, PathExt Activated TopNet central genes from all four subtypes showed significant overlap with NEST genes while central genes from Repressed TopNets showed significant overlap for subtypes Her2 and LumA. In sharp contrast, none of the DEGs (up or downregulated) from any of the subtypes showed significant overlap with the NEST systems (Figure 4C).

Next, we compiled an additional translational benchmark gene set comprising 474 unique target genes of 46 FDA-approved BRCA drugs from CancerDrugs_DB.[32] Again, we found that these genes exhibit significantly higher overlap with the top 200 Activated PathExt genes compared to DEGs (Figure 4D).

Finally, we compared PathExt with MOMA (multi-omics master-regulator analysis),[15] another network-based approach which integrated transcription and mutational data in the context of inferred regulatory networks and previously reported 407 master regulator (MR) proteins across 20 TCGA cohorts. These MR proteins are further grouped into 24 pan-cancer master regulator blocks. Specifically, the study has reported 39, 92, and 23 MR proteins for basal, LumA and LumB subtypes, respectively. We compared the overlap of MOMA MR proteins with the DriverDBv3 driver cancer gene dataset and NEST PPI dataset with that of PathExt identified TopNet genes. For a fair comparison, we selected the same number of genes (based on frequency) which MOMA had for each subtype. Based on the Fisher test, PathExt genes exhibit a comparable overlap with the benchmark sets relative to MOMA (Figure 4E).

Overall, PathExt identified genes show greater cell dependency, recapitulate previously identified cancer drivers, BRCA specific mutated modules, and BRCA FDA-approved drug targets far more effectively than the conventional DEG approach and performs comparably to a recent network-based multi-omics approach MOMA.

The aforementioned analyses were based on unpaired data that we refer to as the first approach of DEG computation (See STAR Methods), where BRCA samples for each subtype were compared with common pool of healthy control samples to estimate the node weight or infer differential expression. We repeated the analysis in a paired fashion where only the donor-matched samples were used and each BRCA sample was compared with the corresponding matched control from the same donor. The central genes and DEGs were computed in the similar fashion as done for unpaired analysis and various analysis were performed. The paired analysis significantly recovered the top genes identified in the unpaired analysis (See Table S39). List of top central genes and DEGs along with their frequencies is provided in the Tables S40–S43, respectively.

We identified the enriched biological processes associated with pan-subtype and subtype-specific genes. For the PathExt Activated TopNets, as observed in unpaired analysis, the common genes were associated mostly with cell cycle (Figure S8A, Table S44). However, subtype-specific genes were enriched for distinct processes in comparison to what we observed in unpaired analysis. For instance, basal subtype was largely associated with the processes associated with chemokine signaling and cell cycle (Figure S8B); top enriched processes for other subtypes are summarized in Figures S8C–S8E, and the complete list of enriched processes is provided in the Tables S45–S48.

Likewise, for Repressed TopNets, common genes among subtypes were enriched for homeostasis, lipid metabolism and cell proliferation and differentiation (Figure S9A and Table S49). Top enriched processes associated with unique genes in subtypes (except LumB, where no significant processes were observed) are summarized in Figures S9B–S9D whereas complete list is provided in Tables S50–S52.

In contrast, when analyzing pan-subtype and subtype-specific DEGs, we didn't observe significant processes associated with upregulated DEGs except in LumA subtype (Figure S10A). However, for downregulated genes, distinct, but fewer, sets of processes were revealed. All results for up and downregulated DEGs are provided in Figures S10B–S10F and Tables S53–S58.

Next, we compared the performance of the central genes and DEGs identified in the paired analysis in various benchmark datasets. First, we looked at the DepMap dependency score and observed that PathExt Activated genes exhibited significantly greater dependency in subtype-specific cell lines compared to upregulated DEGs. As expected, gene dependency of the Repressed TopNets and downregulated DEGs were very low (Figure S11). We further assessed the biological value of the PathExt and DEG identified top genes against multiple cancer driver benchmark datasets specific to BRCA, NEST BRCA PPI module and list of target genes of FDA-approved BRCA drugs. We observed that PathExt performs better in all benchmark datasets compared to DEGs (Figure S12). Overall, our conclusions based on the aforementioned unpaired analyses hold true in paired analysis.

We also compared PathExt results with the DEGs computed using DESeq2 tool[33] (second approach) on the TCGA-BRCA dataset in both paired and unpaired manner to characterized DEGs. DEGs were selected based on LogFC and adjusted p-adjusted value (FDR) cutoff of 0.05 in a subtype-specific manner. Top 200 genes with highest positive and negative LogFC values were considered as up and downregulated DEGs respectively. Findings are discussed in the following texts for unpaired and paired cases.

Unpaired analysis

DESeq2 was implemented on unpaired samples in subtype-specific manner to characterize DEGs (Table S59). Based on LogFC and FDR cutoff, we characterized top 200 most up and downregulated genes in each subtype (Table S60).

Firstly, we performed gene enrichment analysis and identified the enriched biological processes associated with pan-subtype and subtype-specific genes. For the upregulated DEGs, the common genes were associated mostly with "lipid associated and metabolic processes" (Figure S13A and Table S61). However, subtype-specific genes were enriched for distinct processes, for example, basal subtype was enriched for only one significant process i.e., "tyrosine metabolic process", and Lum-B was mainly enriched for "humoral immune response" (Figures S13B and S13C). No significant processes show up in Her2 and Lum-A subtype. Likewise, for downregulated genes, we observed significant enriched processes only for basal and Lum-B subtypes where both subtypes were largely associated with the processes associated with "cell cycle" and "checkpoint signaling" (Figures S13D and S13E and Tables S62 and S63). No significant processes were enriched with common downregulated genes and Her2 and Lum-A subtype-specific genes.

Next, we compared the performance of the PathExt central genes and DEGs identified in the paired analysis in various benchmark datasets. First, we looked at the DepMap dependency score and observed that PathExt Activated genes exhibited significantly greater

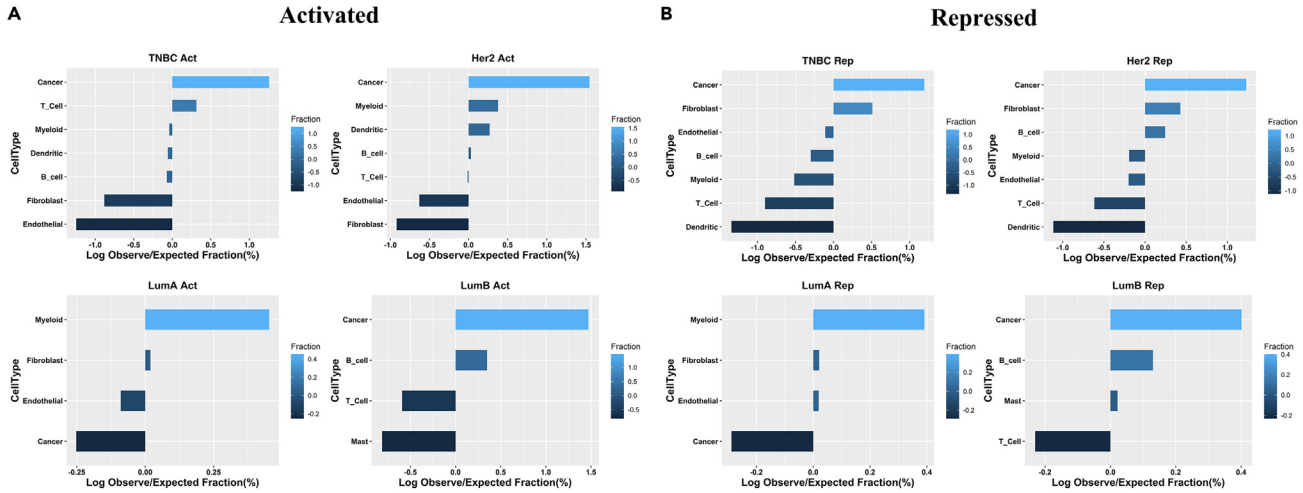


Figure 5. Gene expression distribution of PathExt central genes in individual cell type

Gene expression distribution of top200 PathExt (A) activated central genes and (B) repressed central genes in individual cell types in each BRCA subtype. Within each subtype-specific scRNA-seq dataset, we first identified in each cell type the genes that are expressed in that cell type with an above-mean expression. We then examined the fractions of PathExt genes expressed in each cell type (observed frequency) relative to background expectation across all genes (expected frequency). Finally, Log (observed/expected) value was computed for the central genes across cell types.

dependency in subtype-specific cell lines compared to upregulated DEGs. As expected, gene dependency of the Repressed TopNets and downregulated DEGs were very low (Figure S14A).

We further assessed the biological value of the PathExt and DEG identified top genes against cancer driver benchmark datasets specific to BRCA, NEST BRCA PPI module and list of target genes of FDA-approved BRCA drugs. We observed that PathExt performs better in all benchmark datasets compared to DEGs (Figures S14B–S14D).

Paired analysis

In the case of paired analysis, DESeq2 identified only a few genes with significant FDR value of 0.05, except the LumA subtype (Table S64). The number of significant DEGs with LogFC (both positive and negative) was 49 in basal, 12 in Her2, 774 in LumA, and 37 in Lum-B subtypes. Disregarding the subtypes with very few DEGs and focusing on LumA, we selected top 200 up and downregulated genes for LumA subtype (Table S65) and looked for the enriched biological processes. We observed that upregulated genes were largely associated with processes such as "B cell activation" and "phagocytosis" (Figure S15A, Table S66), whereas downregulated genes were enriched for processes such as "signal release", "protein secretion", and "regulation of acute inflammatory response", etc. (Figure S15B, Table S67).

Overall, we observed that even DESeq2 identified DEGs did not do as well as PathExt on multiple benchmarks. Furthermore, in case of paired analysis, except in case of LumA, DESeq2 did not identify enough DEGs, likely due to smaller sample size. This result further reflects the utility of PathExt which is applicable on the datasets with fewer samples, indeed, even a single foreground sample.

PathExt leverages single-cell data to identify cell-specific mediators in tumor microenvironment

Breast cancer heterogeneity is attributable to both transcriptional variation in cancer cells as well as variation in cellular composition in the tumor microenvironment. We therefore analyzed BRCA subtype-specific single cell transcriptomic data by Qian et al.[34] to chart the distribution of expression of the top 200 PathExt-identified genes (identified using bulk TCGA-BRCA RNA-seq data) across different cell types in the tumor microenvironment (STAR Methods). Within each subtype-specific dataset we first identified in each cell type the genes that are expressed in that cell type with an above-mean expression; note that a gene can be expressed in multiple cell types, but not all. We then examined, in a subtype-specific manner, the fractions of PathExt genes expressed in each cell type relative to background expectation across all genes.

In the Qian et al. dataset, the top genes from the Activated TopNets are expressed most frequently in the malignant cells, except in LumA, where top genes are most enriched in myeloid cells and depleted in malignant cells (Figure 5A). Additional cell types show enriched expression for top activated genes as well as—T cells in TNBC, myeloid cells in Her2, and B cells in LumB. Repressed central genes show similar broad trends except that in three subtypes central repressed genes are also enriched in fibroblasts, consistent with the enrichment of extracellular matrix processes among these genes (Figure 5B). We have provided the observed and expected frequencies of the central genes (Activated and Repressed TopNets) expressed in the cell types in the Table S68.

PathExt identifies potential mediators of resistance to neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone/paclitaxel in TNBC

TNBC is the most aggressive BRCA subtype and presents a major therapeutic challenge. Although TNBC is a clinically defined subtype, it substantially overlaps with the basal subtype defined based on transcriptional profile. Gene expression has been used to further classify TNBC breast cancers.[35] Here, we aim to apply PathExt to identify key genes potentially mediating therapeutic resistance. We used publicly available data from a phase II trial of neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone/paclitaxel therapy, where pathologic complete response (PCR) was the measure of response.[36,37] We specifically focused on the 138 TNBC samples classified into 4 subtypes (BL1-47, BL2-27, LAR-21, and M-43), further classified as responders or non-responders. We applied PathExt to identify central genes (activated and repressed) associated with non-responders (in each sample independently), using samples from responders as the control, and ranked genes based on number of samples of a specific TNBC subtype in which they were detected among the top 100 central genes. Complete lists of PathExt identified TNBC subtype-specific genes and their frequencies in Activated and Repressed TopNets associated with non-responders are provided in the Tables S69 and S70, respectively.

First, we selected the top 20 most frequent central genes in Activated TopNets from each TNBC subtype yielding 60 non-redundant genes, classified in 5 parts: Common—the genes present in at least 2 of the 4 TNBC subtypes among the top 20 genes, and 4 gene sets comprising genes uniquely found in a subtype among the top 20. Figure S16 shows the subtype-specific frequencies of these genes; note that a gene, say *TGFB1*, may be among the top 20 uniquely in BL1 and yet may have a high frequency (although not among top 20) in another subtype (M). As can be seen, most subtypes show relatively high specificity among the top genes. Figure S17 shows a similar figure for top 20 central genes in Repressed TopNets in each subtype. Interestingly, we observed that the gene *TGFB1* was present among top20 central genes in both Activated and Repressed TopNets of BL1 subtype. This observation underscores an important distinction between PathExt and differential expression-based approach. Although the gene appears as a central gene in the Repressed TopNet, it may not be downregulated itself but instead is situated along several downregulated Paths in the network. This can happen if a gene positively regulates one set of gene and processes and at the same time negatively regulates another set of genes and processes. This is true for many signaling molecules including TGFB.[37]

Next, we examined the top 100 most frequent central genes from each subtype and identified genes common among all subtypes and unique to a given subtype. For the activated TopNets, 5 genes—*FOXA1*, *CTNNB1*, *JUN*, *FOS*, and *ALB* were found in all TNBC subtypes and 49, 55, 61, and 48 genes were unique to BL1, BL2, LAR, and M subtypes, respectively. Common genes were broadly associated with developmental processes (Figure 6A and Table S71). Multiple studies have shown that *FOXA1* functions as prognostic marker in various subtypes including TNBC where it can either co-express with androgen receptor (*AR*)[38] or it can transcriptionally silence *SOD2* and *IL6*.[39] Likewise, *CTNNB1* has been shown to be involved in Wnt signaling pathway, and it has been shown that level of *CTNNB1* in WNT/CTNNB1 signaling is elevated in TNBC and has poor clinical prognoses.[40,41] *c-JUN* activation has been associated with many cellular processes and one of the examples is the activation of *JNK* (c-Jun N-terminal kinase). And the high level of *JNK* has been observed to be elevated in TNBC.[42] Likewise, role of *FOS* and *ALB* has also been reported in TNBC.[43,44]

Interestingly, subtype-specific genes (based on top 100) were enriched for distinct processes (Figures 6B–6E) and were also supported by literature to a large extent. For example, BL1 was enriched for regulation of apoptotic signaling, leukocyte chemotaxis, and wound healing, all of which have been shown to associate with immune evasion and therapy resistance[45–47]; Likewise, BL2 was mainly associated with humoral immune response and immune cell migration.[48,49] We observed enrichment of chemokines encoding genes, for example, *CXCL8*, *CXCL1*, etc. in BL2 which are associated with directing immune cell migration necessary to mount and delivering an effective anti-tumor immune response.[50] LAR subtype was highly enriched for metabolic processes (hormone metabolic and vitamin D metabolic) and lipid modification.[51,52] Gong et al. classified the TNBC samples into 3 categories based on their metabolic features and their further analysis reveals that inhibiting lactate dehydrogenase could enhance the response to anti-PD1 immunotherapy[53] and lastly, M subtype was associated with protein modification (protein autophosphorylation and peptidyl-threonine modification), signaling (GPCR signaling pathway, mitotic spindle checkpoint signaling), and negative regulation of nuclear division[54–56] (Figures 6B–6E). Complete lists of the enriched common and unique processes in each subtype are provided in the Tables S72–S75.

Likewise, in the case of Repressed TopNets, 6 genes—*STAT3*, *TGFB1*, *TNF*, *EP300*, *TP53*, and *FOXA1* were common in all the subtypes and were mainly associated with the transcription associated processes. 61 genes were unique in the BL1 subtype and were associated with the immune system and function. BL2 with 45 unique genes was enriched for lipid-related processes, LAR with 40 unique genes was associated with signaling processes, and lastly, M subtype with 55 unique genes was enriched with cell division processes. Complete lists of the enriched common and unique processes in each subtype for Repressed TopNets are provided in the Tables S76–S80 and Figures S18A–S18E. Overall, these results reveal genes potentially mediating resistance across all TNBC subtypes and also distinct processes mediating resistance in each subtype.

Next, we assessed the extent to which PathExt-identified genes can help differentiate responders from non-responders. For this, we identified 21 genes common among the top 200 activated genes in all four TNBC subtypes. However, for comparison with differential expression-based approach, when we analogously selected upregulated DEGs there were only 13 genes that were among topmost upregulated genes in all four subtypes (Table S81). The average gene expression of these genes (21 from PathExt or 13 from DEGs) were used as features to build a support vector machine (SVM) machine learning models and performed 5-fold cross-validation on the original dataset i.e., GSE41998. This yielded, as expected, high accuracy for both PathExt genes (AUROC of 0.76) as well as DEG genes (AUROC of 0.75). However, when, we tested the performance of the aforementioned model (trained and tested using GSE41998 dataset) on an additional independent dataset

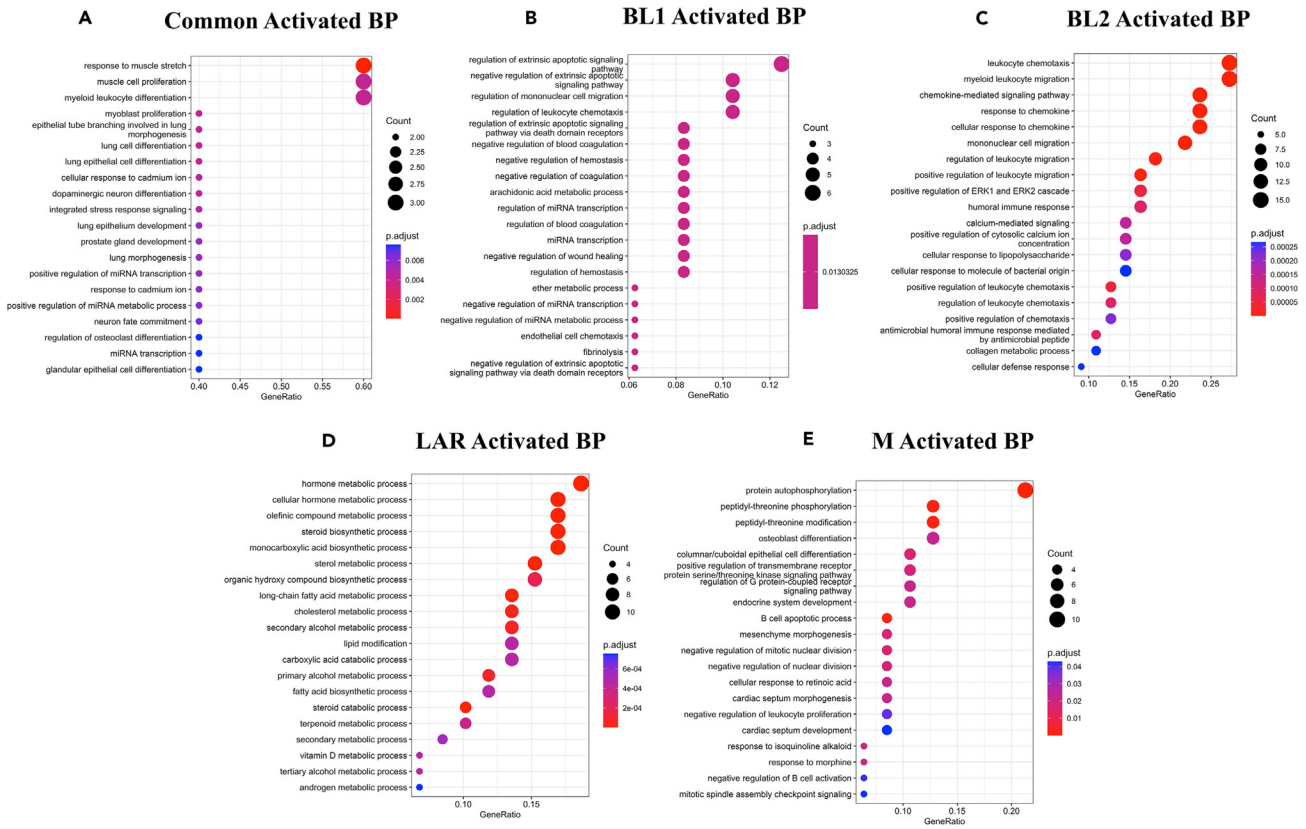


Figure 6. Top central PathExt activated central genes in TNBC non-responders are associated with essential biological processes

Top 20 enriched biological processes associated with common central genes among all the 4 TNBC subtypes (A). Top 20 enriched biological processes associated with unique BL1 (B); BL2 (C); LAR (D); and M (E) subtype central genes.

GSE163882[57] and observed that PathExt identified genes' expression discriminates responders and non-responders with high AUROC of 0.72 and AUPRC of 0.79 [Figure 7A] compared to DEGs, which achieved an AUROC of 0.65 and AUPRC of 0.75 [Figure 7B]. All the training and testing were done using only the samples classified as TNBC. Overall, the results underscore the generalizability of PathExt-identified genes across cohorts. Graphical representation of PathExt workflow in characterizing central genes associated with resistance to chemotherapy and developing prediction model to classify responders and non-responders is shown in Figure S19.

As a community resource, we perform drug repurposing analysis and provide potential drugs targeting the genes potentially mediating resistance in TNBC as currently only poly (ADP-ribose) polymerase (PARP) inhibitors and immune checkpoint inhibitors have been approved for TNBC treatment.[58] We selected top 100 most frequent central genes present in the Activating TopNets of TNBC subtypes and selected the 28 genes present in at least 3 TNBC subtypes as they represent higher confidence targets from computational perspective and are more likely to have broader applicability. Among these targets was the gene *ESR1* which encodes estrogen receptor and ligand-activation transcription factor, and its absence along with progesterone and HER2 receptor is a classification criterion for TNBC. However, a very recent work showed that immunotherapy combined with anti-estrogen therapy exhibited a dramatic anti-tumor effect in estrogen receptor negative tumors, acting on the tumor microenvironment[59] consistent with potential key role of *ESR1*.

Out of the aforementioned 28 selected targets, 3 were excluded due to unsuitability for docking due to their structural properties. Therefore, we proceeded with 25 genes for virtual drug screening.

First, we looked for the database/repository which already provides information of the drugs against the potential target, based on drug repurposing studies. One such drug repurposing database is CLUE[60] (https://clue.io/repurposing-app), hosted by BROAD Institute. It is a curated, annotated and highly optimized open-access repository of >6,000 compounds, many of which are FDA approved. We explored the drug-target data provided in the CLUE database and found that out of our PathExt-identified 25 target genes, 15 have at least one approved drug targeting the gene in this database. Table 1 shows the targets for which at least one drug is mapped in the CLUE data (we have provided up to 3 approved drugs mapped to each target). For the remaining 10 targets, we performed virtual screening by molecular docking using drugs undergoing clinical trials or are FDA approved (STAR Methods). Autodock Vina was used for the molecular docking experiment and based on the binding free energy, we provided top 3 mapped drugs for each gene in Table S82. Figure S20 provides an example of docking results for one of the target genes FOXA1 with the drug (1R,2S)-ephedrine. Image of other docked targets along with the ligand with lowest binding energy is shown in Figure S21.

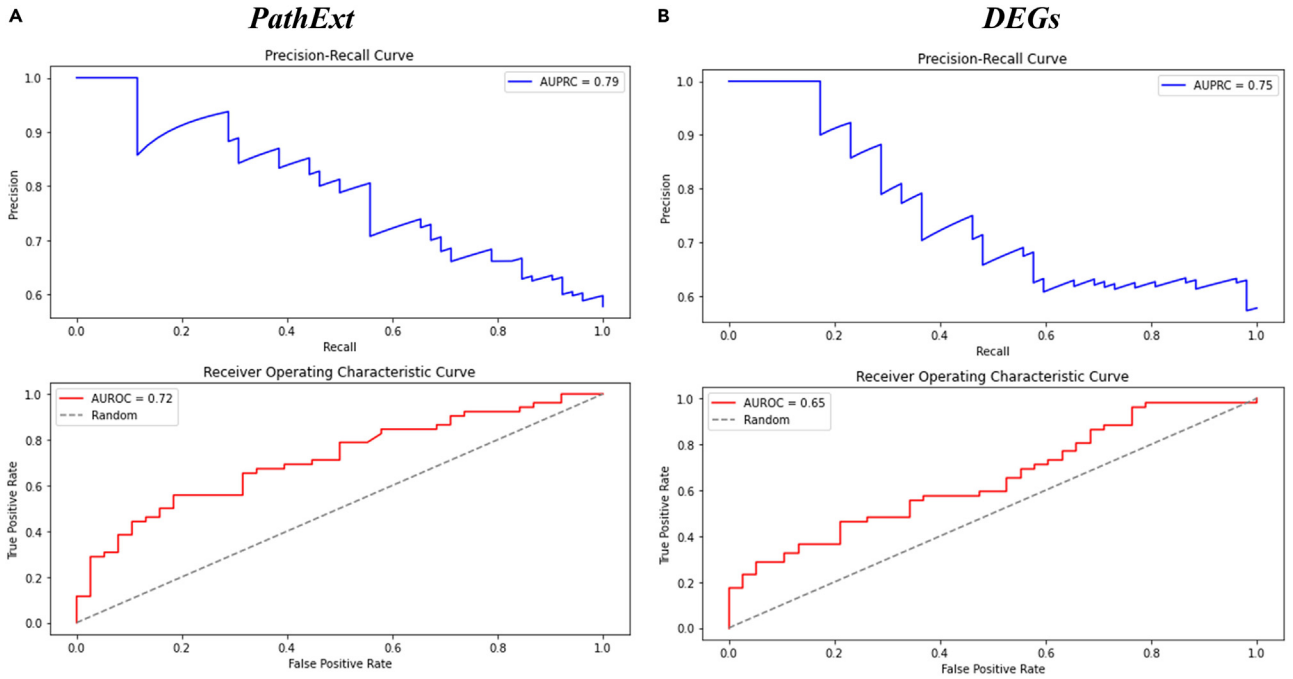


Figure 7. Performance of SVM model on an independent dataset

Average gene expression of 21 central genes from PathExt Activated TopNets and 13 upregulated DEGs were used as feature to build SVM model. PathExt discriminate responder and nonresponder significantly with high AUPRC and AUROC (A) compared to DEGs (B).

We also performed the similar analysis in a subtype-specific manner. We selected top 10 targets from each subtype and mapped them with the drugs already reported in the CLUE database. List of the target genes and the drugs mapped into the CLUE database is provided in the further text in Table 2. For the targets that did not map to CLUE database, virtual screening was performed. Some of the targets were excluded as they were not suitable for the docking due to their structural properties. Result for the remaining targets is provided in the Table S83. Image of the targets along with the ligand with lowest binding energy post-docking is shown in Figure S22.

Table 1. List of the targets from the Activated TopNets and the drugs mapped in CLUE database

Target	Top potential drugs
ALB	Erythromycin-estolate, Erythromycin-ethylsuccinate, Iodipamide
CALM1	Chlorpromazine, Trifluoperazine, Loperamide
EGFR	Brigatinib, Olmutinib, Gefitinib
ESR1	Hexestrol, Danazol, Gestrinone
JUN	Ephedrine-(racemic), Irbesartan, Vinblastine
ADCY2	Forskolin
APP	Curcumin
CCND1	Arsenic-trioxide
CCR5	Maraviroc
CREB1	Adenosine-phosphate, Naloxone
CTNNB1	Urea
EP300	Curcumin
FOS	Ephedrine-(racemic)
IL6	Ibudilast
STAT3	Acitretin, Niclosamide

This table comprises of those targets for which at least one drug was mapped in CLUE database. Maximum of 3 drugs are provided for each target gene.

Table 2. List of the targets from the TNBC subtype-specific Activated TopNets and the drugs mapped in CLUE database

BL1 Subtype		BL2 Subtype		LAR Subtype		M Subtype	
Target	Top potential drugs	Target	Top potential drugs	Target	Top potential drugs	Target	Top potential drugs
STAT3	Acitretin, Niclosamide	JUN	Ephedrine-(racemic), Irbesartan, Vinblastine	RXRA	Acitretin, Adapalene, Bexarotene	CTNNB1	Urea
ESR1	Hexestrol, Danazol, Gestrinone	FOS	Ephedrine-(racemic)	POMC	Loperamide	CCND1	Arsenic-trioxide
CTNNB1	Urea	CTNNB1	Urea	ERBB4	Vandetanib, Afatinib, Dacomitinib	ESR1	Hexestrol, Danazol, Gestrinone
EP300	Curcumin	STAT3	Acitretin, Niclosamide	GAL	Nadide, Lovastatin	JUN	Ephedrine-(racemic), Irbesartan, Vinblastine
PTGS2	Sulfasalazine, Thioctic-acid, Icosapent						
JAK2	Fedratinib, Baricitinib, Peficitinib						
JUN	Ephedrine-(racemic), Irbesartan, Vinblastine						

This table comprises of those targets for which at least one drug was mapped in CLUE database. Maximum of 3 drugs are provided for each target gene.

DISCUSSION

Tumor heterogeneity in breast cancer, particularly TNBC, remains an ongoing hurdle for identifying therapeutic targets with broad applicability. By identifying central mediators of gene expression changes, PathExt is designed to refine expression variance and can be used in multiple translational contexts.

PathExt achieves its goal by first identifying critical pathways in a pre-defined knowledge-based gene network, i.e., those that exhibit significantly different activity in a specific transcriptomic sample compared to controls, then identifying key mediators of the critical pathways. In doing so PathExt relies less on differential expression of individual genes (which are highly variable[7]) and more on network relationships. Unlike differential expression, which relies on sufficient sampling, PathExt can be applied to a single sample, making it applicable for clinical scenarios where large cohorts are rarer. By identifying key genes in a sample-wise fashion (as opposed to a cohort for differential expression), PathExt directly accounts for inter-sample heterogeneity. By focusing on genes mediating the global transcriptome, PathExt shows far greater commonality across BRCA subtypes across multiple benchmark sets compared to DEGs. PathExt also performs comparably to a previous network-based approach—MOMA.

We note that the central genes mediating the Activated TopNet (comprising paths with significantly higher activity in the condition of interest relative to control) are not necessarily differentially upregulated (Figure S23). Likewise, central genes mediating the Repressed TopNet (comprising paths with significantly lower activity in the condition) are not necessarily differentially downregulated (Figure S24); in particular, many of these genes may indeed have repressive effects on other genes while themselves being upregulated. This explains the somewhat counterintuitive observation that the key genes in both the Activated and the Repressed TopNets exhibit elevated inactivating mutations in cancer (Figure 3H). Furthermore, the cell types expressing the genes identified based on bulk sequencing are not immediately clear. Our single-cell analysis of PathExt-identified genes points to the role of tumor microenvironment in oncogenesis, where not only the malignant but various immune compartments may play a key role in mediating the global gene expression, as profiled by bulk sequencing.

As shown previously, PathExt recapitulates the genes previously associated with BRCA more frequently compared to DEGs. For instance, *ERBB2* gene (which encodes for HER2) was identified as a top central gene in over 50% (42/80) of the patients by PathExt compared to fewer than 15% patients (11/80) by DEG.

Besides revealing an overall greater commonality in the key genes across samples and BRCA subtypes, PathExt also reveals shared genes and pathways between specific BRCA subtypes that are consistent with known BRCA subtype-specific biology recently reviewed by Nolan et al.[27] For instance, basal and Her2 like-tumors show low and medium expression of LumA signatures respectively, and we observed that top 10 most frequent LumA genes (except *TTK, NEK2,* and *BIRC5*) were among the top genes in very few basal tumors and in nearly half of the Her2 tumors. Equally importantly, PathExt contributes to the currently limited knowledge of subtype-specific keys genes and their associated biological processes. For instance, Tan et al. recently showed upregulation of neuronal genes preferentially in TNBC associated with neural crest and glial development.[21] Encouragingly, PathExt identified cell-fate determination and regulation of T cell proliferation among the top enriched processes uniquely in basal subtype (which substantially overlaps with TNBC). Likewise, Hartman et al. have shown high expression of *HER2* leads to secretion of interleukin-6 (*IL6)* which further activates *STAT3* ultimately contributing to tumorigenesis.[61] Uniquely in Her2 BRCA, PathExt reveals biological processes associated with IL6 response and inflammation. Consistent with reports linking aging with luminal breast cancer,[62–64] PathExt identified "aging" as the top enriched process specifically in LumA BRCA.

PathExt application to TNBC response revealed 5 genes —*FOXA1*, *CTNNB1*, *JUN*, *FOS*, and *ALB* associated with non-responsive Activated TopNets in all four TNBC subtypes. These genes were broadly enriched for development, signaling, and transcription. One of the top enriched processes we observed associated with resistance was "cell fate determination" which was also reviewed by O'Reilly et al. as one of the factors associated with chemoresistance in TNBC.[18] *FOXA1* upregulation in basal-like cell line MDA-MB-231 leads to increased drug resistance.[65] Likewise, *CTNNB1* is associated with Wnt signaling pathways, known to be associated with BRCA.[42,66,67] There is literature support for other genes as well.[42,67] For the Repressed TopNets, 6 genes were common among all the subtypes, viz. *STAT3*, *TGFB1*, *TNF*, *EP300*, *TP53*, and *FOXA1*. Interestingly, *FOXA1* is revealed as a key gene in both Activated and Repressed TopNets across all subtypes. Dai et al. have shown that downregulation of *FOXA1* leads to increased malignancy and cancer stemness by suppressing *SOD2* and *IL6*,[39] underscoring a pleiotropic role of *FOXA1*. Another resistance-associated gene PathExt revealed is *EP300*, a known modulator of paclitaxel resistance and stemness.[68] Interestingly, the study we used for our analysis included paclitaxel treatment after neoadjuvant chemotherapy.

Lastly, we characterized a few potential therapeutic targets associated with resistance in TNBC and proposed potential drugs against them based on two approaches. First, we mapped the targets in the CLUE database and reported the drugs associated with it. Secondly, for the targets not covered by CLUE, we performed virtual screening via molecular docking experiments and proposed top 3 potential hits based on binding free energy. One of the examples of such an experiment is the docking of FOXA1 (PDB ID: 7vox, chain A)[69] with the compound (1R,2S)-ephedrine. The compound shows good binding energy of −4.1 kcal/mol and was found to interact mostly with the charged residues (Ser, Met, Arg, Lys, and Tyr). We also compared the binding similarity of other anticancer drugs such as erlotinib, imatinib, lapatinib, etc. with ephedrine and observed that other compounds exhibit higher free energy, however, similar interacting residues within the receptor (Table S84). Investigated the literature, we found a recent study has reported the anti-cancer property of ephedrine in TNBC cell lines, MDA-MB-231, and MCF-10.[70,71] Anti-cancer properties of other proposed compounds such as sorafenib and bicalutamide in TNBC is also supported by recent studies.[72,73]

Overall, PathExt is complementary to the conventional DEG approach and reveals common and BRCA subtype-specific key genes and processes, as well as gene mediating chemotherapy response in TNBC. Although in this work we have applied PathExt to two contexts, namely, BRCA subtypes, and therapy response in TNBC, PathExt has more general application to identifying key genes associated with any clinical feature associated with the transcriptomic samples. For instance, using deconvolution, one can quantify various cell types, such as cytotoxic T cells, in each BRCA sample, and apply PathExt to identify key differential mediator genes between samples with high and low cytotoxic T cells. Lastly, as a community resource, we have provided potential drugs that are either approved or undergoing trials targeting key genes revealed by our analyses.

Limitations of the study

As the current work is based on TCGA data, future work requires analysis of more cohorts to establish the success of PathExt in identifying central genes in each BRCA subtypes. Also, at the time of the study, only one single-cell RNA-sequencing dataset was available where few samples were present for subtypes such as luminal-A and luminal-B, therefore, we need more such datasets with sufficient samples in each subtype to validate our findings. Lastly, we proposed key drug targets and potential drugs against them based on *in silico* analysis, however, validation of such finding is important using *in vitro* experiments.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Data collection
  - Identification of sample-specific central genes using PathExt
  - Identification of sample specific differentially expressed genes
  - Gene ontology enrichment analysis
  - Cell-line specific genetic dependency
  - Comparison of PathExt and DEGs genes with benchmarks gene sets
  - Breast cancer subtype single cell data analysis
  - Identifying key genes associated with neoadjuvant treatment in TNBC subtypes
  - Drug repurposing study
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109752.

ACKNOWLEDGMENTS

This work utilized the computational resources of the NIH HPC Biowulf cluster. Authors are also thankful to the other lab members for their valuable suggestions. We are also thankful to Dr. Stanley Lipkowitz, M.D., Senior Investigator at NCI, NIH for his excellent suggestions for improving the manuscript.

This work was supported by the Intramural Research Program of the National Cancer Institute, NIH.

AUTHOR CONTRIBUTIONS

P.A. and N.J. downloaded and processed the data. V.G. downloaded and processed the single cell data. P.A., A.T., A.S., and S.H. perform the analysis. P.A. and S.H. perform the statistical analysis. P.A., P.S.R., and S.H. wrote the manuscript. P.A. and S.H. supervised the study. All authors read the article and approved the submitted version.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 18, 2023
Revised: March 18, 2024
Accepted: April 12, 2024
Published: April 16, 2024

REFERENCES

1. Siegel, R.L., Miller, K.D., Wagle, N.S., and Jemal, A. (2023). Cancer statistics, 2023. CA A Cancer J. Clin. *73*, 17–48. https://doi.org/10.3322/CAAC.21763.
2. Onitilo, A.A., Engel, J.M., Greenlee, R.T., and Mukesh, B.N. (2009). Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin. Med. Res. *7*, 4–13. https://doi.org/10.3121/CMR.2009.825.
3. Lehmann, B.D., Colaprico, A., Silva, T.C., Chen, J., An, H., Ban, Y., Huang, H., Wang, L., James, J.L., Balko, J.M., et al. (2021). Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. Nat. Commun. *12*, 6276. https://doi.org/10.1038/S41467-021-26502-6.
4. Bernard, P.S., Parker, J.S., Mullins, M., Cheung, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. *27*, 1160–1167. https://doi.org/10.1200/JCO.2008.18.1370.
5. Bastien, R.R.L., Rodríguez-Lescure, Á., Ebbert, M.T.W., Prat, A., Munárriz, B., Rowe, L., Miller, P., Ruiz-Borrego, M., Anderson, D., Lyons, B., et al. (2012). PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. BMC Med. Genom. *5*, 44. https://doi.org/10.1186/1755-8794-5-44.
6. Crow, M., Lim, N., Ballouz, S., Pavlidis, P., and Gillis, J. (2019). Predictability of human differential gene expression. Proc. Natl. Acad. Sci. USA *116*, 6491–6500. https://doi.org/10.1073/PNAS.1802973116/-/DCSUPPLEMENTAL.
7. Cui, W., Xue, H., Wei, L., Jin, J., Tian, X., and Wang, Q. (2021). High heterogeneity undermines generalization of differential expression results in RNA-Seq analysis. Hum. Genom. *15*, 7. https://doi.org/10.1186/S40246-021-00308-5.
8. Pe'er, D., and Hacohen, N. (2011). Principles and strategies for developing network models in cancer. Cell *144*, 864–873. https://doi.org/10.1016/J.CELL.2011.03.001.
9. Pai, S. (2022). Network Approaches for Precision Oncology. Adv. Exp. Med. Biol. *1361*, 199–213. https://doi.org/10.1007/978-3-030-91836-1_11.
10. Mundi, P.S., Dela Cruz, F.S., Grunn, A., Diolaiti, D., Mauguen, A., Rainey, A.R., Guillan, K., Siddiquee, A., You, D., Realubit, R., et al. (2023). A Transcriptome-Based Precision Oncology Platform for Patient-Therapy Alignment in a Diverse Set of Treatment-Resistant Malignancies. Cancer Discov. *13*, 1386–1407. https://doi.org/10.1158/2159-8290.CD-22-1020.
11. Hofree, M., Shen, J.P., Carter, H., Gross, A., and Ideker, T. (2013). Network-based stratification of tumor mutations. Nat. Methods *10*, 1108–1115. https://doi.org/10.1038/NMETH.2651.
12. Choi, S.R., Hwang, C.Y., Lee, J., and Cho, K.H. (2022). Network Analysis Identifies Regulators of Basal-Like Breast Cancer Reprogramming and Endocrine Therapy Vulnerability. Cancer Res. *82*, 320–333. https://doi.org/10.1158/0008-5472.CAN-21-0621.
13. Sambaturu, N., Pusadkar, V., Hannenhalli, S., and Chandra, N. (2021). PathExt: a general framework for path-based mining of omics-integrated biological networks. Bioinformatics *37*, 1254–1262. https://doi.org/10.1093/BIOINFORMATICS/BTAA941.
14. Koboldt, D.C., Fulton, R.S., McLellan, M.D., Schmidt, H., Kalicki-Veizer, J., McMichael, J.F., Fulton, L.L., Dooling, D.J., Ding, L., Mardis, E.R., et al. (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70. https://doi.org/10.1038/NATURE11412.
15. Paull, E.O., Aytes, A., Jones, S.J., Subramaniam, P.S., Giorgi, F.M., Douglass, E.F., Tagore, S., Chu, B., Vasciaveo, A., Zheng, S., et al. (2021). A modular master regulator landscape controls cancer transcriptional identity. Cell *184*, 334–351.e20. https://doi.org/10.1016/J.CELL.2020.11.045.
16. Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564–576.e16. https://doi.org/10.1016/J.CELL.2017.06.010.
17. Chiu, Y.C., Zheng, S., Wang, L.J., Iskra, B.S., Rao, M.K., Houghton, P.J., Huang, Y., and Chen, Y. (2021). Predicting and characterizing a cancer dependency map of tumors with deep learning. Sci. Adv. *7*, eabh1275. https://doi.org/10.1126/SCIADV.ABH1275.
18. O'Reilly, E.A., Gubbins, L., Sharma, S., Tully, R., Guang, M.H.Z., Weiner-Gorzel, K., McCaffrey, J., Harrison, M., Furlong, F., Kell, M., and McCann, A. (2015). The fate of chemoresistance in triple negative breast cancer (TNBC). BBA Clin. *3*, 257–275. https://doi.org/10.1016/J.BBACLI.2015.03.003.
19. Oshi, M., Asaoka, M., Tokumaru, Y., Yan, L., Matsuyama, R., Ishikawa, T., Endo, I., and Takabe, K. (2020). CD8 T cell score as a prognostic biomarker for triple negative breast cancer. Int. J. Mol. Sci. *21*, 6968–6983. https://doi.org/10.3390/IJMS21186968.
20. Singh, S., Kumar, S., Srivastava, R.K., Nandi, A., Thacker, G., Murali, H., Kim, S., Baldeon, M., Tobias, J., Blanco, M.A., et al. (2020). Loss of ELF5-FBXW7 stabilizes IFNGR1 to promote the growth and metastasis of triple-negative breast cancer through interferon-γ signalling. Nat. Cell Biol. *22*, 591–602. https://doi.org/10.1038/S41556-020-0495-Y.
21. Tan, R., Li, H., Huang, Z., Zhou, Y., Tao, M., Gao, X., and Xu, Y. (2020). Neural Functions Play Different Roles in Triple Negative Breast Cancer (TNBC) and non-TNBC. Sci. Rep. *10*, 3065. https://doi.org/10.1038/S41598-020-60030-5.
22. Perreault, A.A., Sprunger, D.M., and Venters, B.J. (2019). Epigenetic and transcriptional profiling of triple negative breast cancer. Sci. Data *6*, 190033. https://doi.org/10.1038/SDATA.2019.33.
23. Fiscon, G., Pegoraro, S., Conte, F., Manfioletti, G., and Paci, P. (2021). Gene network analysis using SWIM reveals interplay between the transcription factor-encoding genes HMGA1, FOXM1, and MYBL2 in triple-negative breast cancer. FEBS Lett. *595*, 1569–

1586. https://doi.org/10.1002/1873-3468.14085.

24. Gudjonsson, T., Rønnov-Jessen, L., Villadsen, R., Rank, F., Bissell, M.J., and Petersen, O.W. (2002). Normal and tumor-derived myoepithelial cells differ in their ability to interact with luminal breast epithelial cells for polarity and basement membrane deposition. J. Cell Sci. *115*, 39–50. https://doi.org/10.1242/JCS.115.1.39.
25. Hu, Y., Xie, Q., Wu, X., Liu, W., Li, D., Li, C., Zhao, W., Chen, L., Zheng, Z., Li, G., and Guo, J. (2022). Tension of plus-end tracking protein Clip170 confers directionality and aggressiveness during breast cancer migration. Cell Death Dis. *13*, 856. https://doi.org/10.1038/S41419-022-05306-6.
26. Asano, S., Ono, A., Sakamoto, K., Hayata-Takano, A., Nakazawa, T., Tanimoto, K., Hashimoto, H., and Ago, Y. (2023). Vasoactive intestinal peptide receptor 2 signaling promotes breast cancer cell proliferation by enhancing the ERK pathway. Peptides *161*, 170940. https://doi.org/10.1016/J.PEPTIDES.2023.170940.
27. Nolan, E., Lindeman, G.J., and Visvader, J.E. (2023). Deciphering breast cancer: from biology to the clinic. Cell *186*, 1708–1728. https://doi.org/10.1016/J.CELL.2023.01.040.
28. Liu, S.H., Shen, P.C., Chen, C.Y., Hsu, A.N., Cho, Y.C., Lai, Y.L., Chen, F.H., Li, C.Y., Wang, S.C., Chen, M., et al. (2020). DriverDBv3: a multi-omics database for cancer driver gene research. Nucleic Acids Res. *48*, D863–D870. https://doi.org/10.1093/NAR/GKZ964.
29. Martínez-Jiménez, F., Muiños, F., Sentís, I., Deu-Pons, J., Reyes-Salazar, I., Arnedo-Pac, C., Mularoni, L., Pich, O., Bonet, J., Kranas, H., et al. (2020). A compendium of mutational cancer driver genes. Nat. Rev. Cancer *20*, 555–572. https://doi.org/10.1038/S41568-020-0290-X.
30. Xi, J., Yuan, X., Wang, M., Li, A., Li, X., and Huang, Q. (2020). Inferring subgroup-specific driver genes from heterogeneous cancer samples via subspace learning with subgroup indication. Bioinformatics *36*, 1855–1863. https://doi.org/10.1093/BIOINFORMATICS/BTZ793.
31. Zheng, F., Kelly, M.R., Ramms, D.J., Heintschel, M.L., Tao, K., Tutuncuoglu, B., Lee, J.J., Ono, K., Foussard, H., Chen, M., et al. (2021). Interpretation of cancer mutations using a multiscale map of protein systems. Science *374*, eabf3067. https://doi.org/10.1126/SCIENCE.ABF3067.
32. Pantziarka, P., Capistrano I, R., De Potter, A., Vandeborne, L., and Bouche, G. (2021). An Open Access Database of Licensed Cancer Drugs. Front. Pharmacol. *12*, 627574. https://doi.org/10.3389/FPHAR.2021.627574.
33. Love, M.I., Huber, W., and 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.
34. Qian, J., Olbrecht, S., Boeckx, B., Vos, H., Laoui, D., Etlioglu, E., Wauters, E., Pomella, V., Verbandt, S., Busschaert, P., et al. (2020). A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res. *30*, 745–762. https://doi.org/10.1038/S41422-020-0355-0.
35. Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., and Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. *121*, 2750–2767. https://doi.org/10.1172/JCI45014.
36. Horak, C.E., Pusztai, L., Xing, G., Trifan, O.C., Saura, C., Tseng, L.M., Chan, S., Welcher, R., and Liu, D. (2013). Biomarker analysis of neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone or Paclitaxel in early-stage breast cancer. Clin. Cancer Res. *19*, 1587–1595. https://doi.org/10.1158/1078-0432.CCR-12-1359.
37. Tzavlaki, K., and Moustakas, A. (2020). TGF-β Signaling. Biomolecules *10*, 487. https://doi.org/10.3390/BIOM10030487.
38. Guiu, S., Mollevi, C., Charon-Barra, C., Boissière, F., Crapez, E., Chartron, E., Lamy, P.J., Gutowski, M., Bourgier, C., Romieu, G., et al. (2018). Prognostic value of androgen receptor and FOXA1 co-expression in non-metastatic triple negative breast cancer and correlation with other biomarkers. Br. J. Cancer *119*, 76–79. https://doi.org/10.1038/S41416-018-0142-6.
39. Dai, X., Cheng, H., Chen, X., Li, T., Zhang, J., Jin, G., Cai, D., and Huang, Z. (2019). FOXA1 is Prognostic of Triple Negative Breast Cancers by Transcriptionally Suppressing SOD2 and IL6. Int. J. Biol. Sci. *15*, 1030–1041. https://doi.org/10.7150/IJBS.31009.
40. Khramtsov, A.I., Khramtsova, G.F., Tretiakova, M., Huo, D., Olopade, O.I., and Goss, K.H. (2010). Wnt/beta-catenin pathway activation is enriched in basal-like breast cancers and predicts poor outcome. Am. J. Pathol. *176*, 2911–2920. https://doi.org/10.2353/AJPATH.2010.091125.
41. Weeks, S.E., Kammerud, S.C., Metge, B.J., AlSheikh, H.A., Schneider, D.A., Chen, D., Wei, S., Mobley, J.A., Ojesina, A.I., Shevde, L.A., and Samant, R.S. (2021). Inhibiting β-catenin disables nucleolar functions in triple-negative breast cancer. Cell Death Dis. *12*, 242. https://doi.org/10.1038/S41419-021-03531-Z.
42. Xie, X., Kaoud, T.S., Edupuganti, R., Zhang, T., Kogawa, T., Zhao, Y., Chauhan, G.B., Giannoukos, D.N., Qi, Y., Tripathy, D., et al. (2017). c-Jun N-terminal kinase promotes stem cell phenotype in triple-negative breast cancer through upregulation of Notch1 via activation of c-Jun. Oncogene *36*, 2599–2608. https://doi.org/10.1038/ONC.2016.417.
43. Gao, W., Li, M., and Zhang, Y. (2021). Fibrinogen/Albumin Ratio (FAR) in Patients with Triple Negative Breast Cancer and Its Relationship with Epidermal Growth Factor Receptor Expression. OncoTargets Ther. *14*, 5403–5415. https://doi.org/10.2147/OTT.S339973.
44. Wu, W., Warner, M., Wang, L., He, W.W., Zhao, R., Guan, X., Botero, C., Huang, B., Ion, C., Coombes, C., and Gustafsson, J.A. (2021). Drivers and suppressors of triple-negative breast cancer. Proc. Natl. Acad. Sci. USA *118*, e2104162118. https://doi.org/10.1073/PNAS.2104162118.
45. SenGupta, S., Hein, L.E., Xu, Y., Zhang, J., Konwerski, J.R., Li, Y., Johnson, C., Cai, D., Smith, J.L., and Parent, C.A. (2021). Triple-Negative Breast Cancer Cells Recruit Neutrophils by Secreting TGF-β and CXCR2 Ligands. Front. Immunol. *12*, 659996. https://doi.org/10.3389/FIMMU.2021.659996.
46. Chen, J.Q., and Russo, J. (2009). ERalpha-negative and triple negative breast cancer: molecular features and potential therapeutic approaches. Biochim. Biophys. Acta *1796*, 162–175. https://doi.org/10.1016/J.BBCAN.2009.06.003.
47. Fulda, S., and Debatin, K.M. (2006). Extrinsic versus intrinsic apoptosis pathways in anticancer chemotherapy. Oncogene *25*, 4798–4811. https://doi.org/10.1038/SJ.ONC.1209608.
48. Wu, D. (2005). Signaling mechanisms for regulation of chemotaxis. Cell Res. *15*, 52–56. https://doi.org/10.1038/SJ.CR.7290265.
49. Garaud, S., Buisseret, L., Solinas, C., Gu-Trantien, C., De Wind, A., Van Den Eynden, G., Naveaux, C., Lodewyckx, J.N., Boisson, A., Duvillier, H., et al. (2019). Tumor infiltrating B-cells signal functional humoral immune responses in breast cancer. JCI Insight *5*, e129641. https://doi.org/10.1172/JCI.INSIGHT.129641.
50. Ozga, A.J., Chow, M.T., and Luster, A.D. (2021). Chemokines and the immune response to cancer. Immunity *54*, 859–874. https://doi.org/10.1016/J.IMMUNI.2021.01.012.
51. Sun, X., Wang, M., Wang, M., Yu, X., Guo, J., Sun, T., Li, X., Yao, L., Dong, H., and Xu, Y. (2020). Metabolic Reprogramming in Triple-Negative Breast Cancer. Front. Oncol. *10*, 428. https://doi.org/10.3389/FONC.2020.00428.
52. Blasiak, J., Pawlowska, E., Chojnacki, J., Szczepanska, J., Fila, M., and Chojnacki, C. (2020). Vitamin D in Triple-Negative and BRCA1-Deficient Breast Cancer-Implications for Pathogenesis and Therapy. Int. J. Mol. Sci. *21*, 3670. https://doi.org/10.3390/IJMS21103670.
53. Gong, Y., Ji, P., Yang, Y.S., Xie, S., Yu, T.J., Xiao, Y., Jin, M.L., Ma, D., Guo, L.W., Pei, Y.C., et al. (2021). Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. Cell Metabol. *33*, 51–64.e9. https://doi.org/10.1016/J.CMET.2020.10.012.
54. Kohale, I.N., Yu, J., Zhuang, Y., Fan, X., Reddy, R.J., Sinnwell, J., Kalari, K.R., Boughey, J.C., Carter, J.M., Goetz, M.P., et al. (2022). Identification of Src Family Kinases as Potential Therapeutic Targets for Chemotherapy-Resistant Triple Negative Breast Cancer. Cancers *14*, 4220. https://doi.org/10.3390/CANCERS14174220.
55. Singh, A., Nunes, J.J., and Ateeq, B. (2015). Role and therapeutic potential of G-protein coupled receptors in breast cancer progression and metastases. Eur. J. Pharmacol. *763*, 178–183. https://doi.org/10.1016/J.EJPHAR.2015.05.011.
56. Deng, N., Chen, K., Fan, H., and Jin, F. (2022). The synergistic effect of CDKN2B-AS1 and SPC25 on triple-negative breast cancer. Ann. Transl. Med. *10*, 783. https://doi.org/10.21037/ATM-22-2900.
57. Chen, J., Hao, L., Qian, X., Lin, L., Pan, Y., and Han, X. (2022). Machine learning models based on immunological genes to predict the response to neoadjuvant therapy in breast cancer patients. Front. Immunol. *13*, 948601. https://doi.org/10.3389/FIMMU.2022.948601.
58. Li, Y., Zhan, Z., Yin, X., Fu, S., and Deng, X. (2021). Targeted Therapeutic Strategies for Triple-Negative Breast Cancer. Front. Oncol. *11*, 731535. https://doi.org/10.3389/FONC.2021.731535.
59. Kajihara, N., Ge, Y., and Seino, K.I. (2023). Blocking of oestrogen signals improves anti-tumour effect regardless of oestrogen receptor alpha expression in cancer cells. Br. J. Cancer *129*, 935–946. https://doi.org/10.1038/S41416-023-02381-0.

60. Corsello, S.M., Bittker, J.A., Liu, Z., Gould, J., McCarren, P., Hirschman, J.E., Johnston, S.E., Vrcic, A., Wong, B., Khan, M., et al. (2017). The Drug Repurposing Hub: a next-generation drug library and information resource. Nat. Med. *23*, 405–408. https://doi.org/10.1038/NM.4306.
61. Hartman, Z.C., Yang, X.Y., Glass, O., Lei, G., Osada, T., Dave, S.S., Morse, M.A., Clay, T.M., and Lyerly, H.K. (2011). HER2 overexpression elicits a proinflammatory IL-6 autocrine signaling loop that is critical for tumorigenesis. Cancer Res. *71*, 4380–4391. https://doi.org/10.1158/0008-5472.CAN-11-0308.
62. Hofstatter, E.W., Horvath, S., Dalela, D., Gupta, P., Chagpar, A.B., Wali, V.B., Bossuyt, V., Storniolo, A.M., Hatzis, C., Patwardhan, G., et al. (2018). Increased epigenetic age in normal breast tissue from luminal breast cancer patients. Clin. Epigenet. *10*, 112. https://doi.org/10.1186/S13148-018-0534-8.
63. Yau, C., Fedele, V., Roydasgupta, R., Fridlyand, J., Hubbard, A., Gray, J.W., Chew, K., Dairkee, S.H., Moore, D.H., Schittulli, F., et al. (2007). Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers. Breast Cancer Res. *9*, R59. https://doi.org/10.1186/BCR1765.
64. Benz, C.C. (2008). Impact of aging on the biology of breast cancer. Crit. Rev. Oncol. Hematol. *66*, 65–74. https://doi.org/10.1016/J.CRITREVONC.2007.09.001.
65. Kumar, U., Ardasheva, A., Mahmud, Z., Coombes, R.C., and Yagüe, E. (2021). FOXA1 is a determinant of drug resistance in breast cancer cells. Breast Cancer Res. Treat. *186*, 317–326. https://doi.org/10.1007/S10549-020-06068-5.
66. van Schie, E.H., and van Amerongen, R. (2020). Aberrant WNT/CTNNB1 Signaling as a Therapeutic Target in Human Breast Cancer: Weighing the Evidence. Front. Cell Dev. Biol. *8*, 25. https://doi.org/10.3389/FCELL.2020.00025.
67. Tanaka, S., Ishii, T., Sato, F., Toi, M., and Itou, J. (2020). Eribulin mesylate-induced c-Fos upregulation enhances cell survival in breast cancer cell lines. Biochem. Biophys. Res. Commun. *526*, 154–157. https://doi.org/10.1016/J.BBRC.2020.03.042.
68. Asaduzzaman, M., Constantinou, S., Min, H., Gallon, J., Lin, M.L., Singh, P., Raguz, S., Ali, S., Shousha, S., Coombes, R.C., et al. (2017). Tumour suppressor EP300, a modulator of paclitaxel resistance and stemness, is downregulated in metaplastic breast cancer. Breast Cancer Res. Treat. *163*, 461–474. https://doi.org/10.1007/S10549-017-4202-Z.
69. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The Protein Data Bank. Nucleic Acids Res. *28*, 235–242. https://doi.org/10.1093/NAR/28.1.235.
70. Abdulkarim Alharbi, S., Eldin Ahmed Abdelsalam, K., Asad, M., Alrouji, M., Ahmed Ibrahim, M., and Almuhanna, Y. (2024). Assessment of the anti-cancer potential of Ephedra foeminea leaf extract on MDA-MB-231, MCF-7, 4 T1, and MCF-10 breast cancer cell lines: Cytotoxic, apoptotic and oxidative assays. Saudi Pharmaceut. J. *32*, 101960. https://doi.org/10.1016/J.JSPS.2024.101960.
71. Bensam, M., Rechreche, H., Abdelwahab, A.E., Abu-Serie, M.M., and Ali, S.M. (2023). The role of Algerian Ephedra alata ethanolic extract in inhibiting the growth of breast cancer cells by inducing apoptosis in a p53-dependent pathway. Saudi J. Biol. Sci. *30*, 103650. https://doi.org/10.1016/J.SJBS.2023.103650.
72. Liu, C.Y., Su, J.C., Huang, T.T., Chu, P.Y., Huang, C.T., Wang, W.L., Lee, C.H., Lau, K.Y., Tsai, W.C., Yang, H.P., et al. (2017). Sorafenib analogue SC-60 induces apoptosis through the SHP-1/STAT3 pathway and enhances docetaxel cytotoxicity in triple-negative breast cancer cells. Mol. Oncol. *11*, 266–279. https://doi.org/10.1002/1878-0261.12033.
73. Ávalos-Moreno, M., López-Tejada, A., Blaya-Cánovas, J.L., Cara-Lupiañez, F.E., González-González, A., Lorente, J.A., Sánchez-Rovira, P., and Granados-Principal, S. (2020). Drug Repurposing for Triple-Negative Breast Cancer. J. Personalized Med. *10*, 200–234. https://doi.org/10.3390/JPM10040200.
74. Agrawal, P., Sambaturu, N., Olgun, G., and Hannenhalli, S. (2022). A Path-Based Analysis of Infected Cell Line and COVID-19 Patient Transcriptome Reveals Novel Potential Targets and Drugs Against SARS-CoV-2. Front. Immunol. *13*, 918817. https://doi.org/10.3389/FIMMU.2022.918817.
75. Lehmann, B.D., Jovanović, B., Chen, X., Estrada, M.V., Johnson, K.N., Shyr, Y., Moses, H.L., Sanders, M.E., and Pietenpol, J.A. (2016). Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS One *11*, e0157368. https://doi.org/10.1371/JOURNAL.PONE.0157368.
76. Koschützki, D., and Schreiber, F. (2008). Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene Regul. Syst. Biol. *2*, 193–201. https://doi.org/10.4137/GRSB.S702.
77. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141. https://doi.org/10.1016/J.XINN.2021.100141.
78. Hippen, A.A., Falco, M.M., Weber, L.M., Erkan, E.P., Zhang, K., Doherty, J.A., Vähärautio, A., Greene, C.S., and Hicks, S.C. (2021). miQC: An adaptive probabilistic framework for quality control of single-cell RNA-sequencing data. PLoS Comput. Biol. *17*, e1009290. https://doi.org/10.1371/JOURNAL.PCBI.1009290.
79. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science *313*, 1929–1935. https://doi.org/10.1126/SCIENCE.1132939.
80. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., and Hutchison, G.R. (2011). Open Babel: An open chemical toolbox. J. Cheminf. *3*, 33. https://doi.org/10.1186/1758-2946-3-33.
81. Gan, J.H., Liu, J.X., Liu, Y., Chen, S.W., Dai, W.T., Xiao, Z.X., and Cao, Y. (2023). DrugRep: an automatic virtual screening server for drug repurposing. Acta Pharmacol. Sin. *44*, 888–896. https://doi.org/10.1038/S41401-022-00996-2.
82. Eberhardt, J., Santos-Martins, D., Tillack, A.F., and Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J. Chem. Inf. Model. *61*, 3891–3898. https://doi.org/10.1021/ACS.JCIM.1C00203.
83. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. *30*, 2785–2791. https://doi.org/10.1002/JCC.21256.
84. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/J.CELL.2021.04.048.
85. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/JOSS.03021.
86. M Rosa, G.J., Gelfand, A.E., Diggle, P.J., Fuentes, M., Guttorp, P., Daniel Commenges, E., Lumley, T., Tan, M.T., Tian, G., Ng Alessandra Mattei, K.W., et al. (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. Biometrics *67*, 678–679. https://doi.org/10.1111/J.1541-0420.2011.01616.X.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Breast cancer Bulk RNA-seq	TCGA[14]	https://portal.gdc.cancer.gov/
Breast Cancer Single cell RNA-seq	Qian et al.[34]	http://blueprint.lambrechtslab.org
TNBC microarray dataset 1	Horak et al.[36]	GEO: GSE41998
TNBC RNAseq	Chen et al.[48]	GEO: GSE163882
3D Structures of the target genes	RCSB-PDB[66]	https://www.rcsb.org/
FDA approved/undertrials drugs for docking	DrugBank[42]	https://go.drugbank.com/
Depmap Dependency Probability Score	DepMap Portal[16]	https://depmap.org/portal/download/all/
Significant mutated BRCA Module from NEST PPI database	Zheng et al.[31]	https://idekerlab.ucsd.edu/nest/
BRCA driver genes from DriverDBv3 database	Liu et al.[28]	https://ngdc.cncb.ac.cn/databasecommons/database/id/7081
BRCA driver genes from IntOGen database	Jimenez et al.[29]	https://www.intogen.org/search
BRCA driver genes	Huang et al.[30]	https://github.com/JianingXi/DriverSub
Curated list of FDA approved BRCA drugs and their targets	Pantziarka et al.[32]	https://www.anticancerfund.org/en/cancerdrugs-db
Software and algorithms		
PathExt	Sambaturu et al.[13]	https://github.com/NarmadaSambaturu/PathExt
ClusterProfiler 4.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
miQC	Bioconductor	https://bioconductor.org/packages/release/bioc/html/miQC.html
DESeq2	Bioconductor	https://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html
Obabel	PyData	https://pypi.org/project/openbabel/
AutoDock Vina	MGLTools	https://vina.scripps.edu/downloads/
ggplot2	Bioconductor	https://ggplot2.tidyverse.org/
Pandas	PyData	https://pandas.pydata.org/
Numpy	NumPy	https://numpy.org/
seaborn	PyData	https://seaborn.pydata.org/
matplotlib	PyData	https://pypi.org/project/matplotlib/
ComplexHeatmap	Bioconductor	https://bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
dplyr	CRAN	https://cran.r-project.org/web/packages/dplyr/index.html
ggpubr	CRAN	https://cran.r-project.org/web/packages/ggpubr/index.html
ggVennDiagram	CRAN	https://cran.r-project.org/web/packages/ggVennDiagram/readme/README.html
Support Vector Classifier	scikit-learn	https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Piyush Agrawal (apiyush74@gmail.com, piyush.agrawal@nih.gov).

Materials availability

This study did not generate any new and unique reagents.

Data and code availability

This paper analyzes existing, publicly available data. The accession numbers for the datasets are listed in the key resources table.

All the codes used in running the PathExt tool and dataset analysis are deposited at GitHub (https://github.com/hannenhalli-lab/PathExt_BRCA). DOIs are listed in the key resources table.

Further inquiries can be directed to the corresponding authors.

METHOD DETAILS

Data collection

Transcriptomic profiles for 1059 primary BRCA tumors and 112 normal adjacent samples were downloaded from TCGA dated April 19, 2021.[14] 1059 BRCA tumor samples were classified into 4 subtypes – Basal (192 samples), Her2 (80 samples), Luminal-A (574 samples) and Luminal-B (213 samples). When looked at the paired data (tumor and normal adjacent tissue was taken from the same patient), the statistics changed as follows; Basal (18 samples), Her2 (9 samples), Luminal-A (62 samples) and Luminal-B (23 samples).

Gene expression was jointly quantile normalized and was used as an input for the PathExt.[74] The analysis was performed in both paired and unpaired manner. In case of unpaired analysis, we used the whole TCGA-BRCA dataset whereas in case of paired analysis, we used only the paired data. In case of unpaired data, mean average expression of 112 normal adjacent samples were taken as "control" and each patient was considered as individual "case". However, in case of paired analysis, normal adjacent of each case was considered as control. Using these case and control in unpaired and paired analysis, PathExt based central genes and DEGs were characterized and used further for various analysis.

We also analyzed the TNBC transcriptomic dataset by Horak et al.[36] where they look for the biomarkers associated with the response to the neoadjuvant doxorubicin/cyclophosphamide followed by Ixabepilone. In a follow up study, Lehmann et al. classified 138 of the above TNBC samples into 4 TNBC subtypes (see[75] for classification details) - BL1 (47 samples), BL2 (27 samples), LAR (21 samples), and M (43 samples). We applied PathExt to each TNBC subtype separately to identify central genes mediating resistance (non-response).

Identification of sample-specific central genes using PathExt

PathExt method has been described in detail in our previous publications.[13,74] Here we provide a very brief sketch. Based on a set of control transcription profiles (e.g., normal adjacent breast samples for a BRCA sample), we first compute weight for each node in a given knowledge-based protein interaction network, such that the node weight represents the extent of upregulation (Activated) or downregulation (Repressed) of the gene in the foreground sample relative to the control. Computation of the node (gene) weight in the current study is modified from the original PathExt paper and is computed in the following manner.

In the original manuscript log (fold change) was used as the node weight. While log (fold change) is a reasonable choice for node weight, given the dependence of the magnitude of log(fold change) on the expression level, we instead computed the *expected* log (fold change) for a given control expression and then used the difference of *observed* and *expected* log(fold change) as the node weight. To compute the expected absolute log (fold change), we regressed absolute log (fold change) values across all genes against the gene expression in control samples using Loess fit, implemented in R.

More formally, Node weight for a given gene 'g', is computed by the following equation:

$$Node\ Weight\ (g)\ =\ E\_LogFC(g\_e)\ -\ O\_LogFC(g\_e)$$

Where, 'g_e' be the expression of g in control; 'E_LogFC(g_e)' is the expected log_fold for g_e as determined by the Loess fit; and 'O_LogFC(g)' is log of the ratio of the expression of g in case versus control samples.

For Activated TopNets the fold change was computed in the Cases relative to the Control, and for Repressed TopNets the fold change was computed in the Control relative to the Cases. Following node weight computation (for Activated or Repressed TopNet), where higher node weight represent greater changes, we compute the edge weight such that high weights of the two adjacent nodes map to a low edge weight (edge weight = 1/Sqrt(X,Y), where are X and Y are the node weight of the two adjacent nodes for the edge. Next, given the edge weights, we compute shortest paths between every two nodes in the networks, where short (low-cost) paths indicate the most perturbed paths.

Once PathExt identifies significantly perturbed paths in the network and finally in the subnetwork composed of such significantly perturbed paths (Activated or Repressed TopNet), it then identifies the top central genes based on betweenness centrality measure.[76] These central genes represent key mediators of global transcription changes (upregulation for Activated TopNet and downregulation for Repressed TopNet) in a particular sample. Having computed the top 100 most central genes for each BRCA sample we compute the number of sample (of a given BRCA subtype) in which a gene was among the top central genes, and then selected the top 200 most frequent genes to represent subtype-specific key genes; this was done independently for Activated and Repressed TopNets.

Analogous procedure was applied to identify key genes corresponding to non-responders relative to responders among the TNBC BRCA cohort.

Identification of sample specific differentially expressed genes

For comparison, we identified DEGs in two different approaches. In the first approach, in each sample, for each gene, we computed log2 FC = (Expression of gene in Case/Expression of gene in Control). In each sample, we selected the top 100 genes with highest Log2 FC as the up-regulated and the bottom 100 as the downregulated. We did not perform significance testing as this was based on a single sample, however, we note that in all cases top100 genes had logFC > 2 and bottom 100 genes had log FC < -2. After obtaining the top (or bottom) 100 genes in each sample individually, for each gene, we computed the frequency with which it occurred in the top100 (or bottom100) list across samples of a subtype. The above procedure was performed independently for each BRCA subtype and top 200 genes (up and downregulated) were selected for the analysis.

In the second approach, we computed subtype specific DEGs using DESeq2 tool.[33] Subtype-specific samples were taken as "Treated" class and all 112 normal samples were taken as "Control" class for computing subtype-specific DEGs. Based on LogFC value and adjusted p-value (FDR) <0.05, top200 up and down regulated genes were selected. In case of paired analysis, subtype specific paired Control and Treated samples were considered while computing DEGs. Analogous procedure was applied to identify key DEGs corresponding to non-responders relative to responders among the TNBC BRCA cohort.

Gene ontology enrichment analysis

To identify enriched biological processes and molecular functions in various subtypes, we used the Clusterprofiler 4.0 package,[77] where the foreground was the top200 central genes (PathExt or DEGs) and the background was the default background used by the package. The database used was the human database and the minimum gene size and maximum gene size was set as 10 and 200 respectively to ensure specific terms. We then used the function "simplify" with cutoff value 0.8 to remove the redundant terms. Default parameters were used otherwise. In the case of TNBC responder and non-responders, we used similar parameters except foreground and background gene list. In this case, the foreground was top100 genes and the background was a customized list of 10994 genes, those that were detected in the cohort [Table S85].

Cell-line specific genetic dependency

From the DepMap database,[16] we obtained the gene dependency probability scores for 16708 genes in 44 breast cancer cell lines, further categorized into BRCA subtypes. For each gene, we then obtained the average dependency probability score across the cell lines in a subtype-specific manner. As recommended, genes with average dependency probability score >= 0.5 were recorded as essential for comparing PathExt and DEG genes.

Comparison of PathExt and DEGs genes with benchmarks gene sets

We compiled various previously published datasets of potential BRCA driver genes, approved FDA drugs for BRCA and assessed their overlap with the PathExt central genes or DEGs. We used Fisher exact test to assess the statistical significance of overlap and reported Odds Ratio (>1 indicates greater than expected overlap).

Breast cancer subtype single cell data analysis

We used a single cell RNA-seq dataset for the 4 BRCA subtypes by Qian et al.[34] Read count matrices of scRNA-seq data (obtained using 10X v2 sequencing) breast cancer tumors were downloaded from http://blueprint.lambrechtslab.org. Author-supplied annotations were used to label each cell. The miQC package[78] was used to remove dead cells, with a probability threshold of 0.5 (i.e., a probability of at least 0.5 that the cell is not dead) used to retain high-quality cells for downstream analyses. After filtering the dataset, we retained only those cell types having a minimum of 50 cells. This provided 7 cell type data for Basal and Her2 subtypes, and 4 cell type data for LumA & LumB subtypes. Next, we computed the z-score of gene expression for the PathExt top200 central genes in each cell (using all cells to estimate the mean and the standard deviation) and then took the average z-score value of the cells in a given cell type. Next, we selected the genes with z-score > 0 in each cell type and were considered as 'expressed' in that cell type. Next, we computed the fraction (%) of such expressed genes in every cell type and normalized the fraction across all cell types to get 'Observed frequency' in a cell type. We also computed the 'Expected frequency' in a similar manner by considering all genes. Finally, Log(Observed/Expected) value was computed for the central genes across cell types.

Identifying key genes associated with neoadjuvant treatment in TNBC subtypes

PathExt was applied to a previously published cohort of TNBC tumors having undergone neoadjuvant therapy with Doxorubicin/Cyclophosphamide followed by Ixabepilone or Paclitaxel drug, and further classified into responders and non-responders[36] where responders include patients with complete or partial response, whereas non-responders include patients with stable or progressive disease. TNBC is further subdivided into 4 subtypes, BL1, BL2, LAR and M,[75] and hence, we applied PathExt in a subtype-specific manner to identify central genes among non-responders relative to responders. List of the GSE41998 samples used in the current analysis along with its TNBC classification in the Table S86.

Drug repurposing study

We selected the top 100 most frequent central genes present in Activated TopNets from each TNBC subtype and then further selected genes present in at least 3 subtypes. This resulted in 29 genes out of which 3 genes were not suitable for analysis because of their structural

properties. We also created another dataset comprising of top 10 most frequent genes in Activated TopNets in a subtype-specific fashion to provide subtype-specific targets and drugs targeting these genes. Based on the CMap database (using CLUE platform),[79] we first identified the drugs targeting these genes. We consider only those drugs which are either approved or undergoing trials. For the remaining unmapped targets, we performed virtual screening using drugs which are either FDA approved or are under clinical trials. The 3D structures of the targets were downloaded from the RCSB-PDB[69] and further refined using Open Babel software.[80] Next, we used online server DrugRep[81] to locate various binding pockets on the protein surface and selected the pocket with the largest volume covering most active site residues. Identification of active sites and residues facilitate the docking experiment and help in identifying drugs with improved binding. Virtual Screening was performed using AutoDock Vina software.[82] The software requires a receptor and ligand file in a specific file format ("pdbqt" file format), which was prepared using MGLTools software.[83] Post virtual screening, based on docking free energy and root mean square deviation (RMSD), top 3 drugs for each target were proposed. Further, experimental work is required to prove the efficacy of these drugs.

QUANTIFICATION AND STATISTICAL ANALYSIS

PathExt[13] was used to compute central genes. Likewise, DESeq2[33] was used to compute differential expression. Single-cell analysis was performed using the Seurat[84] pipeline. All biological function analyses were performed using the *enrichGO* function in clusterProfiler 4.0[77] R package. Figures are generated using the Python package seaborn[85] and the ggplot2[86] R packages. We performed Fisher's exact test and computed the odd's ratio to observe significant overlap of PathExt central genes and DEGs with various benchmarking datasets used in the study. We also performed Wilcoxon-test (two-sided) to compare PathExt central genes with DEGs in multiple analyses. The specific test is mentioned in each context.