Molecular subtyping and functional validation of TTK, TPX2, UBE2C, and LRP8 in sensitivity of TNBC to paclitaxel

Ramesh Elango,[1] Radhakrishnan Vishnubalaji,[1] Hibah Shaath,[1,2] and Nehad M. Alajez[1,2]

[1]Cancer Research Center, Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), PO Box 34110, Doha 00000, Qatar;
[2]College of Health & Life Sciences, Hamad Bin Khalifa University (HBKU), Qatar Foundation (QF), Doha, Qatar

Triple-negative breast cancer (TNBC) patients exhibit variable responses to chemotherapy, suggesting an underlying molecular heterogeneity. In the current study, we analyzed publicly available transcriptome data from 360 TNBC and 88 normal breast tissues, which revealed activation of nucleosome and cell cycle as the hallmarks of TNBC. Mechanistic network analysis identified activation of FOXM1 and ERBB2, and suppression of TP53 and NURP1 networks in TNBC. Employing Iterative Clustering and Guide-gene Selection (ICGS), Uniform Manifold Approximation and Projection (UMAP), and dimensionality reduction analyses, we classified TNBC into seven molecular subtypes, each exhibiting a unique molecular signature, including immune infiltration (CD19, CD8, and macrophages) and mesenchymal signature, which correlated with variable disease outcomes in a larger cohort (1,070) of BC. Mechanistically, depletion of *TTK, TPX2, UBE2C, CDCA7, MELK, NFE2L3, DDX39A,* and *LRP8* led to substantial inhibition of colony formation of TNBC models, which was further enhanced in the presence of paclitaxel. Our data provide novel insights into the molecular heterogeneity of TNBC and identified *TTK, TPX2, UBE2C,* and *LRP8* as main drivers of TNBC tumorigenesis.

INTRODUCTION

Despite worldwide initiatives in research and the significant advances made in the field, breast cancer (BC) remains the most common cancer, accounting for the highest number of cancer-related mortalities in women worldwide. Triple-negative BC (TNBC) represents 15% to 20% of invasive BCs and is characterized by the lack of expression of estrogen receptors (ERs), progesterone receptors (PRs), and lack of amplification of human epidermal growth factor receptor 2 (HER2).[1–3] TNBC has been shown to particularly affect women younger in age, and tumors tend to be larger in size with higher metastasis, relapse frequencies, poorer prognosis, and relatively worse outcomes in patients.[4–6] With the lack of available known targets in TNBC, patients generally do not benefit from endocrine therapy; therefore, surgery, radiotherapy, and chemotherapy remain the primary mode of treatment. Identifying additional treatment options for this subset of patients is essential; however, recent studies have demonstrated that due to the complex heterogeneity of TNBC, it is unlikely that there will be a "one fits all" solution, as emerging evidence shows that the term "TNBC" encompasses a wide range of BC subsets. Therefore, expanding our understanding on their classification and differentiation between subsets is essential in providing better-tailored therapies.

Studies on identifying invasive cell gene signatures in one TNBC cell line (SUM149) could not be verified in other TNBC cell lines in the same study due to the strong inter-tumor heterogeneity at the invasion front of different subtypes. This uniqueness of each biological sample presents challenges in biomarker discovery.[7] In fact, studies by Kim et al.,[8] as well as our own previous studies, show a degree of intra-tumor heterogeneity, with evidence for the pre-existence of subsets accountable to neoadjuvant chemotherapy (NAC) resistance when analyzing the transcriptomes of single cells within one patient. Heterogeneity even within cancer stem cells has been described by Wu et al.,[9] where deep single-cell RNA sequencing (RNA-seq) identified a high level of heterogeneity in alternative splicing patterns between cell populations that exhibited signs of stemness within the same cell line. Another study categorizes TNBC into six subtypes based on gene expression analysis, including two basal-like (BL1 and BL2), two mesenchymal (M and MSL), one immunomodulatory (IM), and one luminal androgen receptor (LAR) subtype based on distinct gene ontologies. Certain subtypes (LAR) were associated with decreased relapse-free survival.[10] These indications allow for better predictive approaches for better personal care. Furthermore, heterogeneity in TNBC has also been observed in the proteomic landscape,[11] further solidifying the existence of different subclasses within TNBC, emphasizing the need for a better understanding of the different classifications for a more efficient and tailored therapy plan for patients.

Our data analyzing the transcriptome from a total of 360 TNBC and 88 normal breast tissues showed a number of altered biological
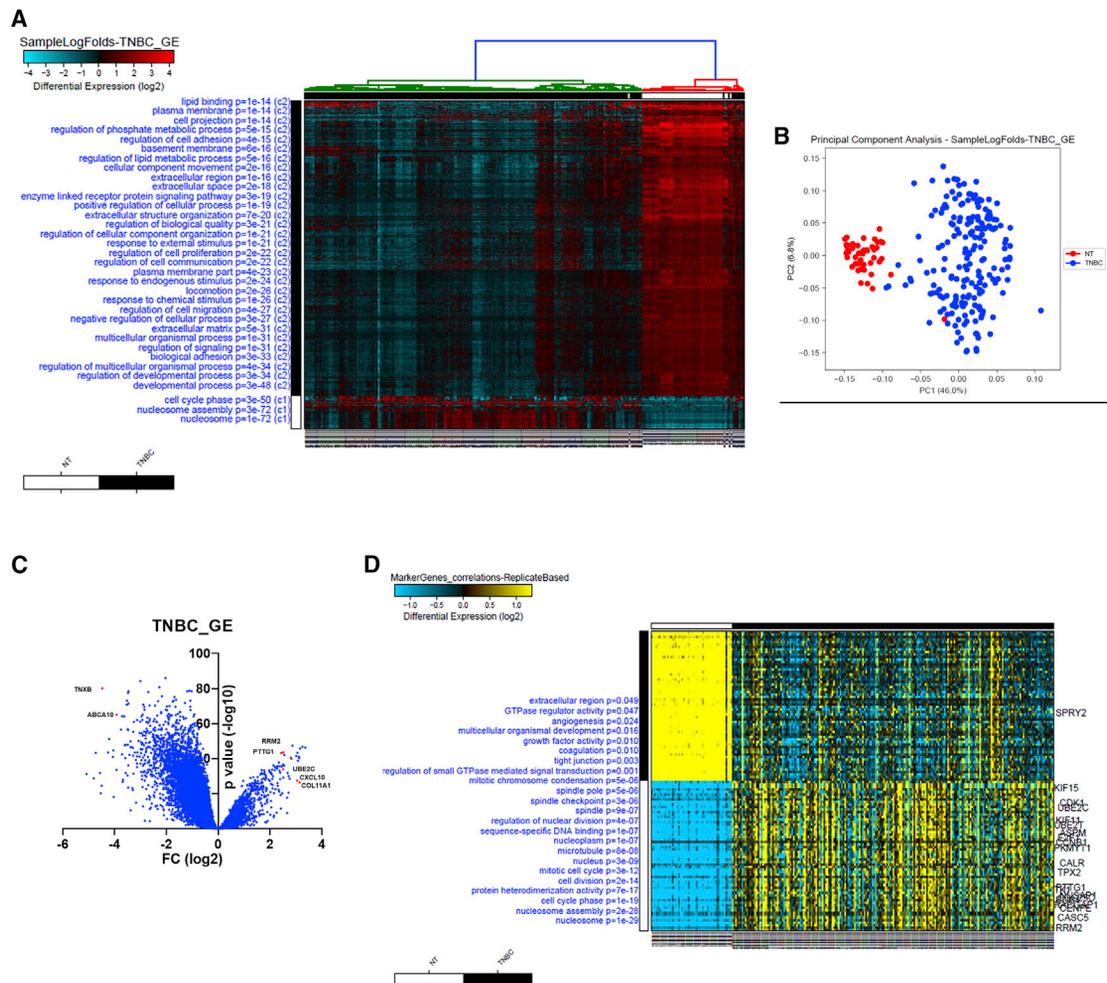
Figure 1. Transcriptional landscape in TNBC compared to adjacent normal tissue

(A) Hierarchical clustering of TNBC (n = 200) and adjacent normal tissue (n = 50) based on differentially expressed genes between the two groups. Each column represents one sample, and each row represents a gene. Expression level of each gene (log2) in a single sample is depicted according to the color scale. (B) Principal-component analysis (PCA) for the RNA transcriptome of TNBC (n = 200) and adjacent normal tissue (n = 50). (C) Volcano plot depicting the most differentially expressed genes in TNBC versus normal tissue. (D) Marker finder analysis depicting the list of genes that are selectively expressed in TNBC versus normal breast tissue.

processes in TNBC, as well as the activation of a number of mechanistic networks (MNs) including FOXM1 and ERBB2, while TP53 and NURP1 networks were mostly suppressed. A number of selected genes were validated for aberrant expression in our study group, and targeted depletion of *TPX2*, *UBE2C*, *CDCA7*, *MELK*, *NFE2L3*, *TTK*, *DDX39A*, and *LRP8* using RNAi led to substantial inhibition of colony formation in BT-549, MDA-MB-231, and HCC70 TNBC models, which was further augmented in the presence of paclitaxel (PTX). Using a number of computational algorithms, we unfolded extensive heterogeneity of TNBC at the transcriptome level. Subsets exhibiting unique molecular signatures were identified, including signatures with remarkable immune infiltration (CD19, CD8, and macrophages) and mesenchymal signature. Our data sheds light on the molecular heterogeneity of TNBC, subsequently leading to the activation of specific MNs unique to each subset. This will aid in better understanding

the mechanisms deployed in different TNBC subtypes for better biomarker discovery and more efficient methods of therapeutic intervention.

RESULTS

Transcriptome analysis of TNBC compared to normal tissue revealed enrichment in cell cycle and nucleosome assembly cellular processes

Initially, we identified the transcriptional portrait from 200 TNBC samples and 50 controls. Using Benjamini-Hochberg false discovery rate (FDR) multiple testing correction (p(corr) < 0.05) and 2.0 fold change (FC), we identified 323 upregulated and 2,797 downregulated genes that were differentially expressed between TNBC and normal tissue (Table S1). Hierarchical clustering based on differentially expressed genes revealed clear separation between the two groups (Figures 1A
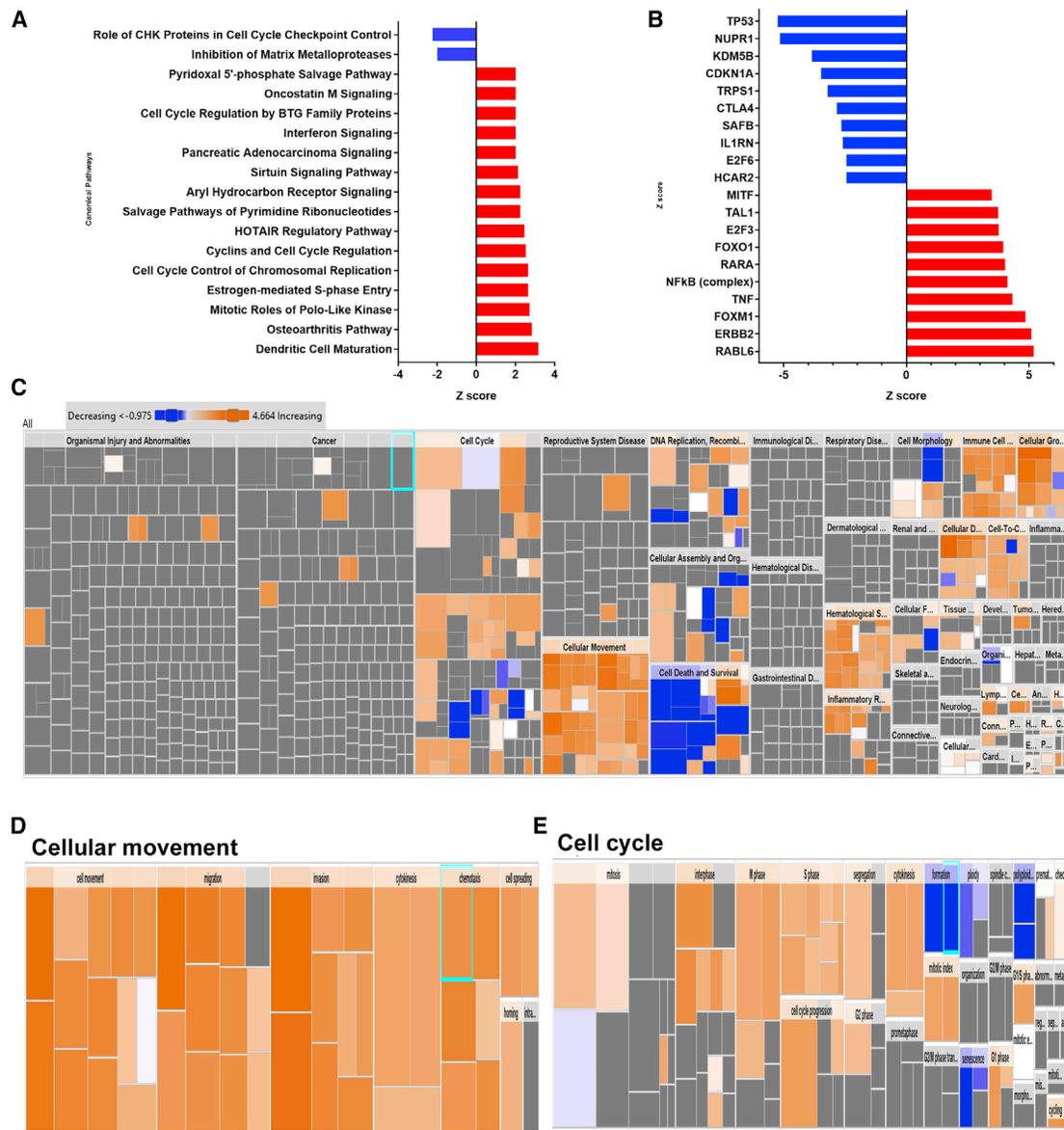
Figure 2. Downstream effector analysis of differentially expressed genes in TNBC and adjacent normal tissue

(A) Canonical pathway analysis depicting the most affected canonical pathways in TNBC and normal tissue. Z score correlates with the degree of enrichment. (B) Upstream analysis revealed enrichment and suppression of a number of functional categories in TNBC compared to adjacent normal tissue. (C) Tree map (hierarchical heatmap) depicting affected functional categories based on differentially expressed genes where the major boxes represent a category of diseases and functions. (D and E) Illustration of cellular movement (D) and cell cycle (E) functional categories are illustrated.

and 1B). Notably, enrichment in cellular processes related to cell cycle and nucleosome assembly were more prominent in TNBC tissue. A volcano plot (scatterplot) that shows statistical significance (log p value; y axis) versus magnitude of change (log FC; x axis) is depicted in Figure 1C. We subsequently employed the marker finder algorithm to predict markers that are associated with TNBC versus normal phenotype. Our data revealed enrichment in Gene Ontology (GO) terms associated with nucleosome assembly, cell cycle regulation, and microtubule among the top GO terms associated with TNBC (Figure 1D).

Ingenuity Pathway Analysis (IPA) revealed multiple activated MNs and functional categories in TNBCs

Canonical pathway analysis of the upregulated gene list in TNBC compared to normal tissue revealed activation of multiple pathways, including cell cycle regulation, mitotic role of polo like kinase, HOTAIR regulatory pathways, as well as interferon signaling, while processes related to regulation of the cell cycle by checkpoint kinase and inhibition of metalloproteases were inhibited (Figure 2A). Upstream regulator analysis (URA) revealed activation of RABL6,
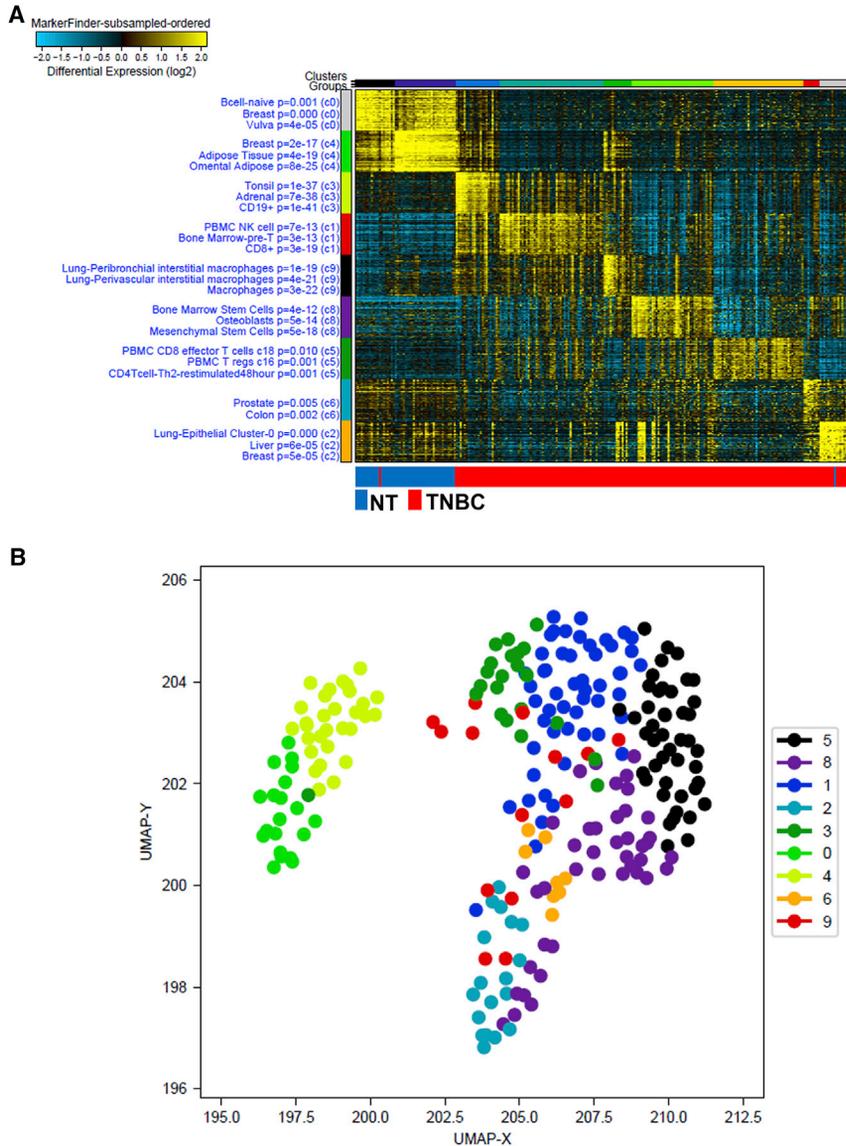
Figure 3. Iterative Clustering and Guide-gene Selection (ICGS) analysis revealed heterogeneity in TNBC

(A) Cell-type predictions and heatmap performed on TNBC (n = 200) and adjacent normal tissue (n = 50). (B) Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction analysis of TNBC and adjacent normal tissue.

TNBC and 50 normal samples to the ICGS2 algorithm and identified nine clusters with distinct gene signatures (Figure 3A). The color scale displays differential gene expression (log2). The first cluster was mostly for normal tissue (19 normal tissue [NT] and 1 TNBC), which was enriched in vulva (p = 4.0×10^{-5}), breast (p = 0.0004), and naive B cell (p = 0.0005). The second cluster was purely for normal tissue samples and was enriched in gene signatures of omental adipose (p = 8.1×10^{-25}), adipose tissue (p = 4.3×10^{-19}), and breast (p = 2.2×10^{-17}). The third cluster was purely TNBC and was enriched in signatures of CD19 (p = 1.2×10^{-41}), adrenal (p = 7.4×10^{-38}), and tonsil (p = 1.6×10^{-37}). Our data revealed a fourth cluster as the main cluster consisting of 52 TNBC samples. The gene signature from this cluster was consistent with CD8$^+$ (p = 3.8×10^{-19}), bone marrow-pre-T (p = 3.4×10^{-13}), and PBMC natural killer (NK) cell (p = 7.2×10^{-13}). An additional cluster consisting of 14 TNBC samples was identified with a gene signature indicative of macrophages (p = 3.9×10^{-22}), lung-perivascular interstitial macrophages (p = 4.2×10^{-21}), and lung-peribronchial interstitial macrophages (p = 1.4×10^{-19}). The fifth cluster consisted of 41 TNBC samples that resembled mesenchymal stem cells (p = 5.11×10^{-18}), osteoblasts (p = 5.4×10^{-14}), and bone marrow stem cells (p = 4.0×10^{-12}). The sixth cluster consisted of 45 TNBC samples and resembled CD4 Tcell-Th2-restimulated 48 hour (p = 0.0006), PBMC T regs c16 (p = 0.001), and PBMC CD8 effector T cells c18 (p = 0.009). The seventh cluster consisted of few TNBC samples (8) with gene signature resembling colon (p = 0.001) and prostate (p = 0.005). The last cluster consisted of 17 TNBC and 1 NT samples and had a signature resembling breast (p = 5.47×10^{-5}), liver (p = 6.9×10^{-5}), and lung-epithelial cluster-0 (p = 0.0001). Similar heterogeneity of TNBC was revealed using Uniform Manifold Approximation and Projection (UMAP) analysis (Figure 3B). Genes enriched in each cluster are indicated in Table S2.

We subsequently analyzed the transcriptome from each cluster compared to normal breast tissue using the upstream regulator and

ERBB2, FOXM1, TNF, NFKB, RARA, FOXO1, E2F3, TAL1, and MITF and suppression of TP53, NUPR1, KDM53, CDKN1A, TRPS1, CTLA4, SAFB, IL1RN, E2F6, and HCAR2 upstream networks in TNBC (Figure 2B). Concordantly, disease and function analysis on the upregulated genes revealed the most significant enrichment in pathways related to cell proliferation and movement (Figure 2C). Illustration of the cellular movement (Figure 2D) and cell cycle (Figure 2E) functional categories based on IPA analysis are found in Figure 2.

Heterogeneity of TNBC revealed using UMAP and Iterative Clustering and Guide-gene Selection 2 (ICGS2) algorithms

Clustering data presented in Figure 1 revealed variation in the transcriptome of various TNBC samples, suggesting the existence of heterogeneity among TNBC patients based on whole-transcriptome analysis. We subsequently subjected transcriptome data from 200

A

B

C
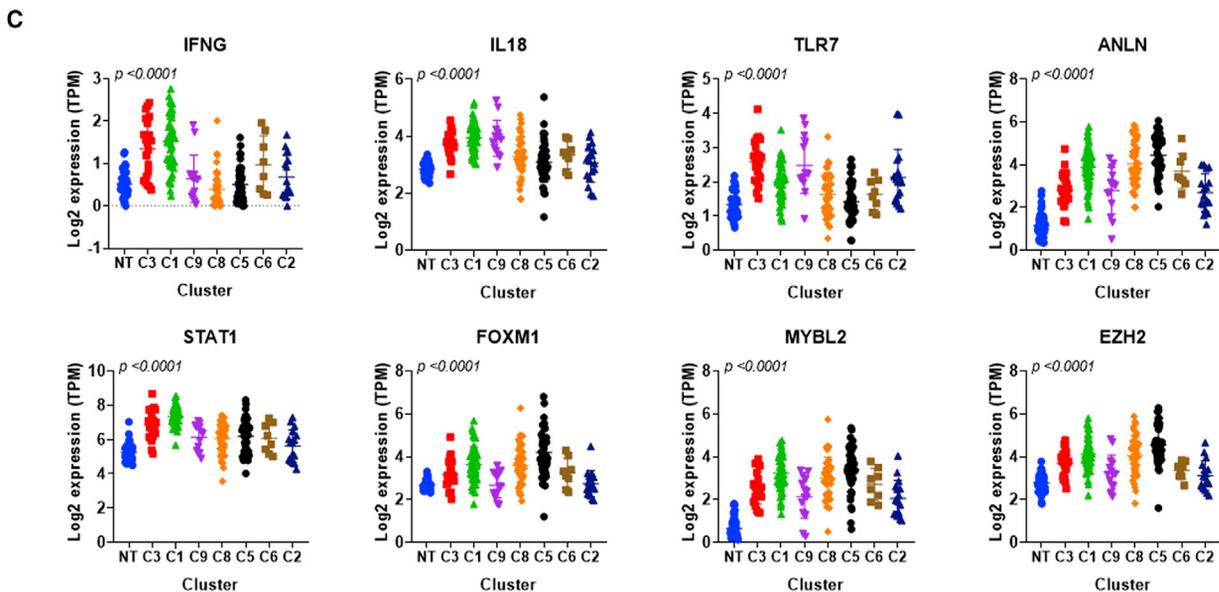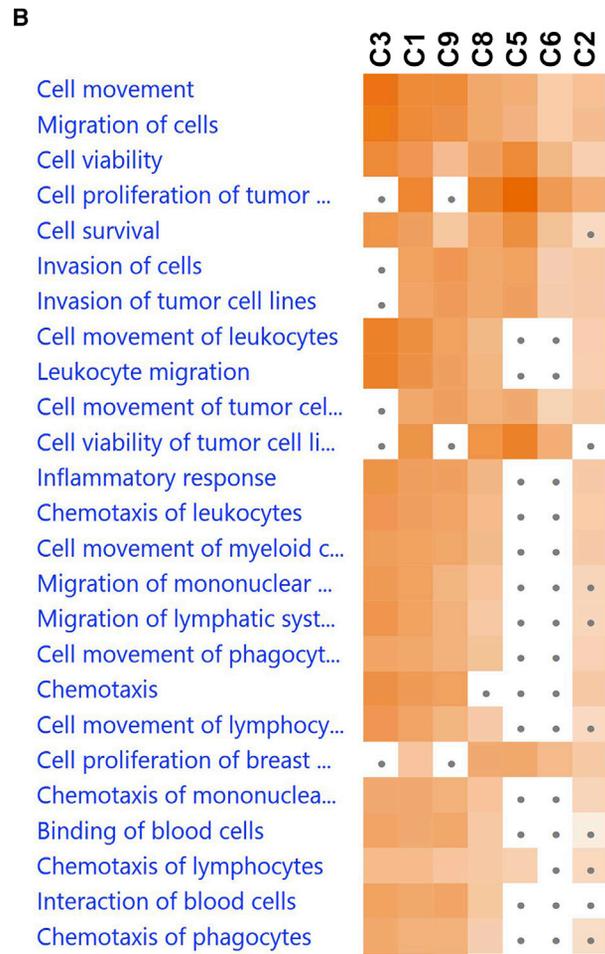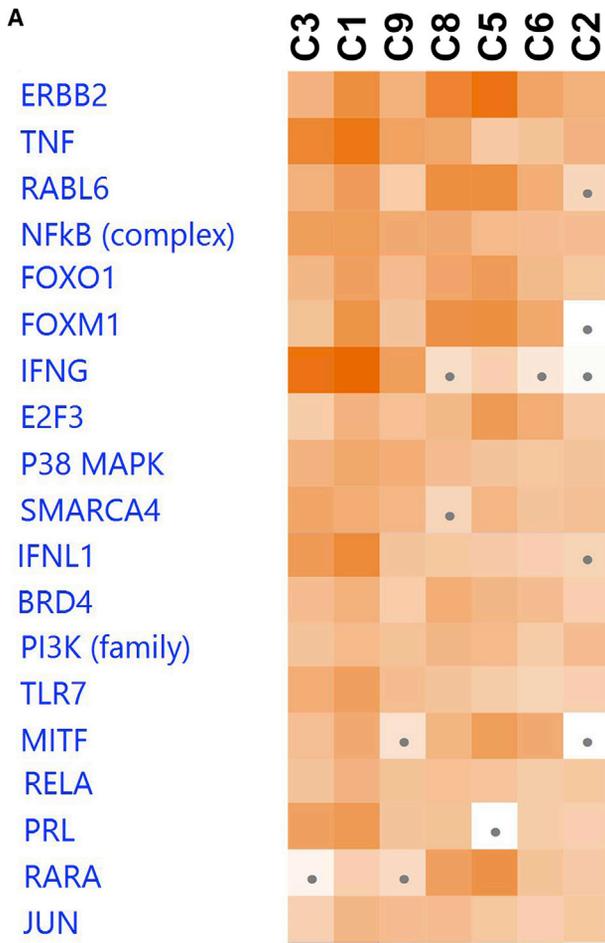
(legend on next page)

disease and functional analysis in IPA. Upregulated genes in each cluster are listed in Table S3. This comparative analysis revealed large similarities for clusters C3, C1, and C9 versus C8, C5, C6, and C2 (Figure 4A). The ERBB network was activated in all subtypes, but activity was higher in C1, C8, and C5. Tumor necrosis factor (TNF) activation was mostly seen in C3 and C1, while IFNG was predominantly activated in C3, C1, and C9. FOXM1 was activated in C8, C5, C6, and C1. The C3, C1, and C9 clusters were associated with enhanced cellular movement and migration (Figure 4B). Cell proliferation of tumor cells was absent in C3 and C9 clusters but was activated in C1, C8, C5, C6, and C2. Cell migration of leukocytes, chemotaxis, and inflammatory response was completely absent in C5 and C6 (Figure 4B). The expression of selected genes from the upstream regulator analysis in each cluster is shown in Figure 4c.

Prognostic value of the identified TNBC subtypes in BC

We subsequently subjected the upregulated gene signatures (compared to NT) derived from each cluster to overall survival (OS) and disease-free survival (DFS) analysis using the GEPIA2 database employing 1,070 BC patients. Our data revealed C3, and to lesser extent C9 and C1, to be associated with favorable OS and DFS (Figures 5A and 5B). C5, C8, and C6 were associated with unfavorable OS and DFS, while C2 was associated with unfavorable OS and favorable DFS (Figures 5A and 5B). Taken together, our data suggest BC subtypes associated with CD19, CD8, and macrophage signatures are associated with favorable clinical outcome.

Functional characterization of selected genes on TNBC viability and drug sensitivity

Eight genes were subsequently chosen for functional studies through exploration of the gene dependency database[12] and integration with differentially expressed genes in TNBC in the current study. TPX2 and UBE2C were upregulated in all seven clusters; CDCA7 in C1 and C5; MELK in C1, C3, C5, C6, and C8; TTK in C1, C3, C5, C6, and C8; and DDX39A and LRP8 in C1, C5, and C8; while NFE2L3 was upregulated in TNBC versus normal but not enriched in any TNBC subsets. The expression of the aforementioned genes was subsequently validated in a second cohort of 160 TNBC and 38 normal tissue, revealing significant upregulation in the validation cohort, which was concordant with discovery cohort data (Figure 6A). High expression of the same gene panel was observed in a panel of TNBC cell lines, suggesting their suitability as cell models to study the functions of those genes (Figure 6B). To provide additional biological insight into the significance of the identified genes in TNBC, we used a small interfering RNA (siRNA)-mediated silencing approach to suppress the expression of the aforementioned genes in the BT-549, MDA-MB-231, and HCC-70 TNBC models. siRNA-mediated targeting of the aforementioned genes led to substantial reduction in gene expression in the three TNBC models (Figure 6C). Knockdown of TPX2 has the most deleterious effects on colony-forming unit (CFU) potential of BT-549 TNBC model, followed by TTK, LRP8, UBE2C, CDCA7, MELK, DDX39A, and NFE2L3, respectively (Figures 6D and 6E). When combined with PTX, TPX2 has the most deleterious effects, followed by CDCA7, DDX39A, TTK, UBE2C, LRP8, NFE2L3, MELK, UBE2C, CDCA7, and MELK, respectively. Similar results were also observed using the MDA-MB-231 and HCC70 TNBC models (Figures 6F and 6G).

Knockdown of TPX2, TTK, and LRP8 affects TNBC cell cycle progression and enhances PTX drug sensitivity

As shown above, knockdown of TPX2, UBE2C, TTK, and LRP8 genes has significant inhibition of CFU of various TNBC models when combined with PTX. In order to gain more insight on the effects of targeting of the aforementioned genes on cell cycle and drug sensitivity, we conducted cell cycle analysis and acridine orange (AO)/ethidium bromide (EtBr) staining on TNBC cells depleted of TPX2, UBE2C, TTK, and LRP8 in the presence and absence of PTX. Representative data of cell cycle alterations are presented in Figure 8. TPX2, TTK, and LRP8 depletion alone led to an increase of apoptotic events in MDA-MB-231, which was further enhanced when combined with PTX (Figure 7A). UBE2C depletion has no significant effects on the cell cycle compared to control in the MDA-MB-231 model. Additionally, prominent reduction in G1 phase was observed in TTK-PTX and LRP8-PTX, while similar reduction was also seen in TPX-depleted cells in the presence or absence of PTX (Figures 7A and 7B). TPX2, TTK, UBE2C, and LRP8 depletion in BT-549 and HCC70 models as single agent or in combination with PTX induced cell death (Figures 7A and 7B).

We subsequently employed live dead AO/EtBr staining assay to confirm the mode of cell death and morphological changes in response to siRNA-mediated gene silencing. Concordant with cell cycle analysis, the cellular growth was reduced in all four siRNAs-transfected TNBC models (Figure 8). Cell shrinkage with less cytoplasm and nuclear ratio was observed in the MDA-MB-231 model endorsing the cell cycle findings. Overall, we observed a more profound reduction in cell growth and induction of cell death in TPX2 and TTK-depleted cells, which was further enhanced in the presence of PTX. Taken together, those data are concordant with CFU results implicating the aforementioned genes in TNBC biology.

DISCUSSION

While TNBC is classically characterized based on the lack of hormone receptor expression and HER2 (ERBB2) amplification, TNBC patients oftentimes respond differently to neoadjuvant and adjuvant chemotherapy, implying an underlying heterogeneity within the disease.[13]

Figure 4. Upstream regulator and functional annotation enrichment in TNBC subtypes

Upregulated genes from each TNBC subtype cluster were subjected to comparative upstream and functional annotation analysis using IPA. (A and B) Top affected upstream regulator (A) and functional categories (B) for each TNBC cluster are indicated. Color intensity corresponds to activation Z score. (C) The expression of selected genes from upstream regulator analysis in the indicated clusters. P values were calculated using ANOVA analysis.
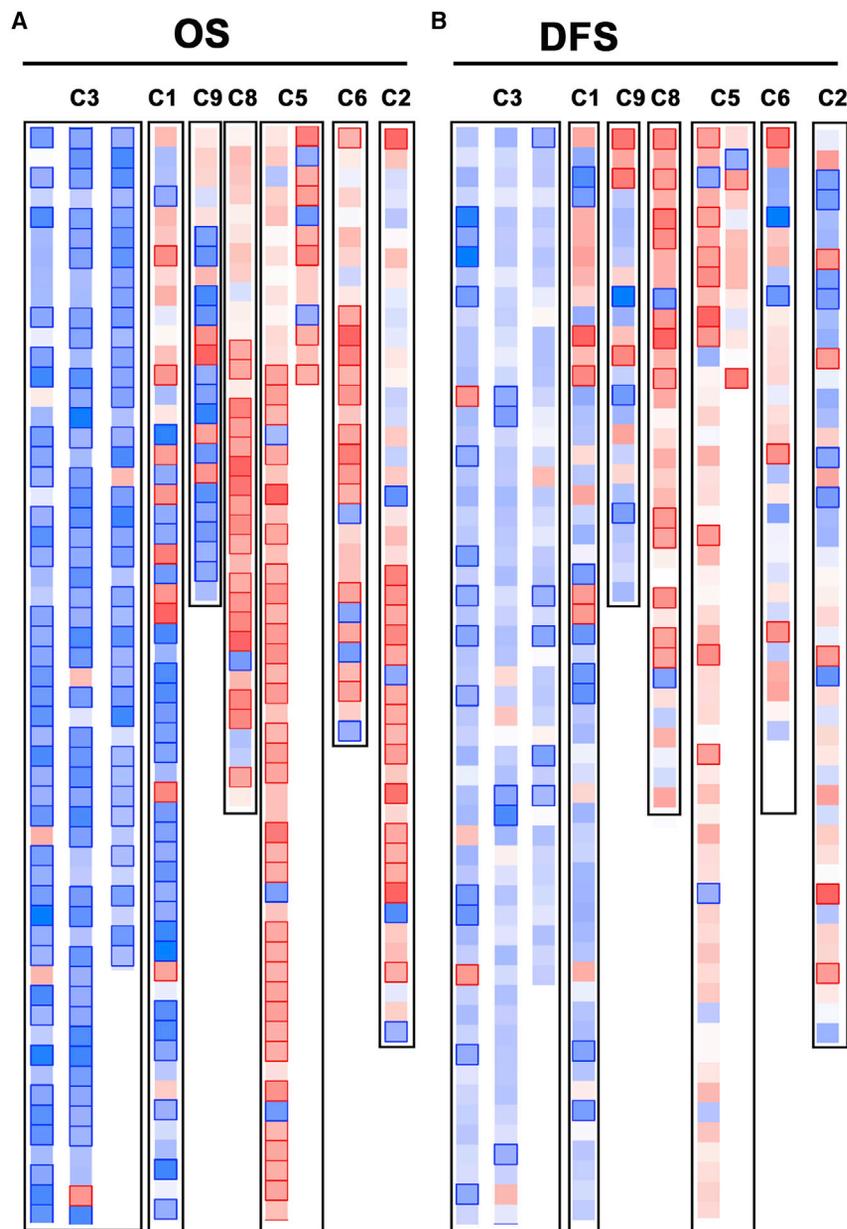
Figure 5. Prognostic value of gene signatures from the indicated TNBC molecular subtypes

(A and B) Survival heatmap for each TNBC cluster for overall survival (OS) and disease-free survival (DFS). Red color indicated HR >1, while blue color indicates HR <1. Squares with darker edges have the highest prognostic values.

pressed, and mesenchymal-like), although our data delineated TNBC heterogeneity into a total of seven groups. Our clustering analysis is based on transcriptome-derived gene signatures, delineating TNBC heterogeneity into a total of seven groups (four immune-enriched signatures [CD19, adrenal, and tonsil; CD8[+], bone marrow-pre-T, and PBMC NK cell; macrophages, lung-perivascular interstitial macrophages, and lung-peribronchial interstitial macrophages; and CD4Tcell-Th2, PBMC Tregs, and PBMC CD8 effector T cells]; mesenchymal signature enriched in genes related to mesenchymal stem cells, osteoblasts, and bone marrow stem cells; signature resembling colon and prostate; and cluster resembling breast, liver, and lung-epithelial cluster-0), while other studies sought to subtype TNBC using alternative approaches. Lehmann et al., as mentioned previously, categorizes TNBC into subtypes according to BL, mesenchymal-like and LAR subtypes, with the consideration of presence of different gene mutations in each subtype, while Burstein et al.[10] identified and confirmed four distinct TNBC subtypes based on a combination of basal and immune subtypes, i.e., LAR, mesenchymal (MES), BL immunosuppressed (BLIS), and BL immune-activated (BLIA).[15] In other studies, such as in Muranen et al.,[16] BC subsets were defined with regards to patient survival according to the St. Gallen 2013 criteria and the PAM50 gene expression signature, in addition to the presence of tumor CHEK2 mutations. While all methods of subtyping can overlap and are interchangeable, and while each strategy provides valuable prognostic significance in the treatment of TNBC patients, our approach revealed more refined subtypes than has been described before and provide deeper insight into enriched signaling pathways and functional categories for each subtype and their prognostic value for better development of more efficient therapies. The prognostic value of the identified signatures remains to be evaluated in independent TNBC cohorts.

In the current study, we employed multiple computational pipelines to dissect the alteration in signaling networks and functional annotations as well as tumor heterogeneity in TNBC based on RNA-seq analysis of a large cohort of 360 TNBC and 88 normal tissues. Our comparative analysis highlighted activation of nucleosome, cell cycle, and microtubule functional categories as the hallmarks of TNBC. Employing UMAP and ICGS2 algorithms, we identified seven subgroups of TNBC based on the expression of unique gene markers. The main cluster was indicative of CD8[+] infiltration, while an additional two clusters had mesenchymal phenotype or resembled Th2 and Treg CD4 cells. Our data is in agreement with Jiang and colleagues,[14] who classified the same patient cohort into four subtypes (LAR, IM, BL immune-sup-

The presence of immune infiltrating immune cells has been associated with response of TNBC to neoadjuvant and adjuvant chemotherapy. Denkert and colleagues[17] previously showed the percentage of intratumoral lymphocyte infiltration to be an independent prognostic
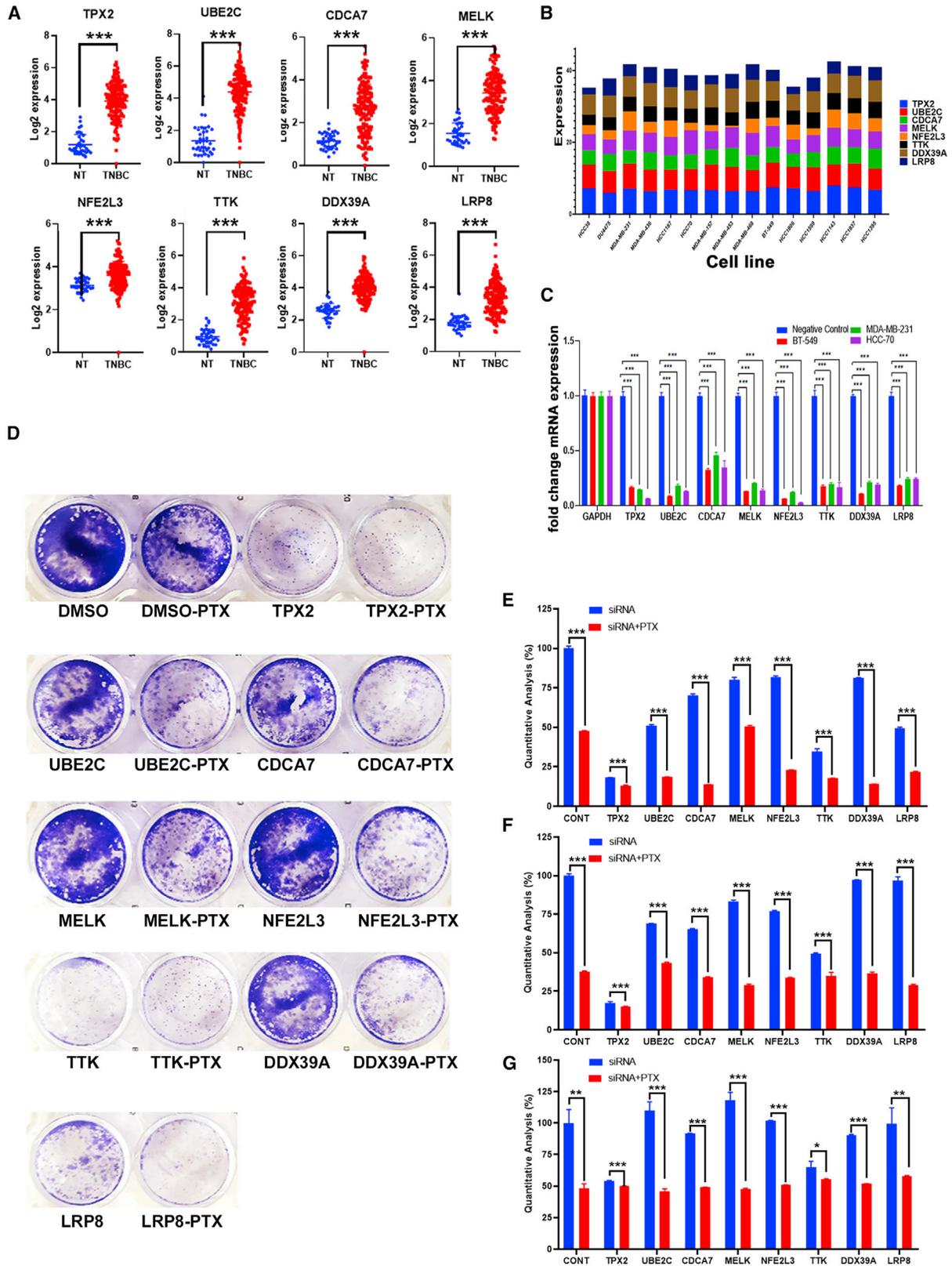
(legend on next page)

factor for pathologic complete response to NAC in BC patients. García-Teijido and colleagues[18] correlated immune infiltration with better response of TNBC patients to NAC.

Despite clear heterogeneity between TNBC samples employed in the current study, comparative analysis showed that our seven assigned clusters exhibit a certain degree of activated networks among clusters, notably the ERBB network, particularly activated in C1, C8, and C5. This is in agreement with a recent study, which shows heterogeneous ERBB activation in TNBC when analyzed at a single-cell level, otherwise undetected when analyzing in bulk.[19] This highlights the sensitivity of bioinformatics pipelines employed in our current study, allowing us to differentiate between distinct heterogeneous groups in the TNBC population. Being able to identify TNF-rich clusters as we see in C3 and C1 could be highly significant when considering clinical treatment options. TNF-related apoptosis-inducing ligand (TRAIL) receptor agonists have shown great promise with varying sensitivity in some subtypes of TNBC. Continuous treatment with MEDI3039 (TRAIL receptor agonist) proved particularly effective in basal B TNBC subtypes.[20] This shows the importance of identifying unique signatures in TNBC for better-tailored treatment options with maximum effectiveness.

Multiple studies, including our own, have highlighted the role of FOXM1 activation in TNBC and colorectal cancer (CRC).[21-24] Overexpression of FOXM1 and ERBB2 lead to genomic instability and uncontrolled cell division and malignancy, which are associated with poor prognosis and drug resistance in many cancers, including BCs.[25-29] In agreement with our data, Bollu et al.,[30] describe the relationship between FOXM1 and MELK, also shown to be upregulated in our patient cohort. Bollu et al.[30] describe a FOXM1-binding site within the MELK promoter region, which, upon knockdown of FOXM1, reduced MELK expression in p53 mutant TNBC cells. This highlights the importance of network analysis in deciphering how each of our aberrantly expressed genes affects downstream effector molecules and contributes to heterogeneity in TNBC, ultimately important for administering the appropriate therapies.

To provide a better insight into how some aberrantly expressed genes effect functionality, *TPX2*, *UBE2C*, *CDCA7*, *MELK*, *NFE2L3*, *TTK*, *DDX39A*, and *LRP8*, exhibiting various degrees of enrichment in each TNBC cluster, were targeted for siRNA-mediated depletion in three TNBC cell lines. Our data revealed depletion of those genes to inhibit CFU in all three TNBC models, suggesting a possible role for those genes in TNBC biology. Concordant with mechanistic data, elevated expression of *TPX2*, *UBE2C*, *MELK*, and *TTK* correlated with worse OS, while elevated expression of *TPX2*, *TTK*, and *CDCA7* correlated with worse refractory-free survival in an independent cohort of BC patients. Our data is in agreement with other reports implicating knockdown of *TPX2*, *TTK*, *UBE2C*, and *MELK* reduction in cellular migration, proliferation, and apoptosis and as favorable prognostic markers in bladder,[31] pancreatic,[32,33] and ovarian cancer cells.[34]

In conclusion, our data revealed molecular heterogeneity of TNBC and unfolded numerous MNs in TNBC and provides novel insights into TNBC biology and potential utilization of several targets as prognostic biomarkers and targets for therapeutic interventions.

MATERIALS AND METHODS
Data source and bioinformatics analysis
The transcriptome data were retrieved from 360 TNBC and 88 normal tissue from the Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/SRP157974) using the SRA toolkit v2.9.2 as previously described.[35] Two hundred TNBC and 50 control samples were used as the discovery cohort, while an additional 160 TNBC and 38 normal tissues from the same cohort were used as the validation cohort. Paired-end RNA-seq FASTQ files were subsequently pseudo-aligned to the human genome, and reads were counted using KALLISTO 0.42.1. Abundance data were subsequently subjected to ICGS, UMAP dimensionality reduction, principal-component analysis (PCA), and hierarchical clustering, as described before.[24,36,37] Employed algorithms combine multiple complementary subtype detection methods and hierarchical ordered partitioning and collapsing hybrid (HOPACH, sparse non-negative matrix factorization, cluster "fitness," support vector machine) to resolve rare and common cell states. ICGS2 identified cell clusters through a complex process of PageRank down-sampling, feature selection ICGS2, dimension reduction and clustering (sparse NMF, SNMF), cluster refinement (MarkerFinder algorithm), and finally cluster re-assignments using support vector machine (SVM). The MarkerFinder algorithm was subsequently applied to identify rigorously defined cell clusters with unique gene expression for downstream cell cluster assignment, which identified genes that are positively correlated with an idealized cluster-specific expression profile. Cell cluster assignment was finally achieved from the marker genes identified for sufficiently fitting clusters, based on the cells assigned to the specific SNMF.

Figure 6. Knockdown of selected genes reduces colony formation potential of TNBCs

(A) The expression of *TPX2*, *UBE2C*, *CDCA7*, *MELK*, *NFE2L3*, *TTK*, *DDX39A*, and *LRP8* was validated in a second cohort of 160 TNBC and 38 normal tissue. Transcriptome data were subjected to pseudoallignment using kallisto followed by gene abundance estimation and log2 transformation. Data are presented as dot plot with the corresponding p value indicated. (B) Expression of *TPX2*, *UBE2C*, *CDCA7*, *MELK*, *NFE2L3*, *TTK*, *DDX39A*, and *LRP8* in a number of TNBC cell lines based on cell line encyclopedia database. (C) qRT-PCR for the expression of *TPX2*, *UBE2C*, *CDCA7*, *MELK*, *NFE2L3*, *TTK*, *DDX39A*, and *LRP8* in BT-549, MDA-MB-231, and HCC70 transfected with targeting or scrambled siRNA. GAPDH was used as reference gene. Data are presented as mean ± SD, n = 6. ***p < 0.005. (D) Representative CFU for BT-549 cells on day 7 post-knockdown of the indicated genes alone or in combination with paclitaxel (20 nM). Wells are representative of two independent experiments for each treatment condition. (E–G) Quantitative analysis of the effect of gene silencing with and without paclitaxel (20 nM) on the ability of BT-549 (E), MDA-MB-231 (F), and HCC70 (G) CFU is shown. Data is presented as mean ± SD, n = 3.
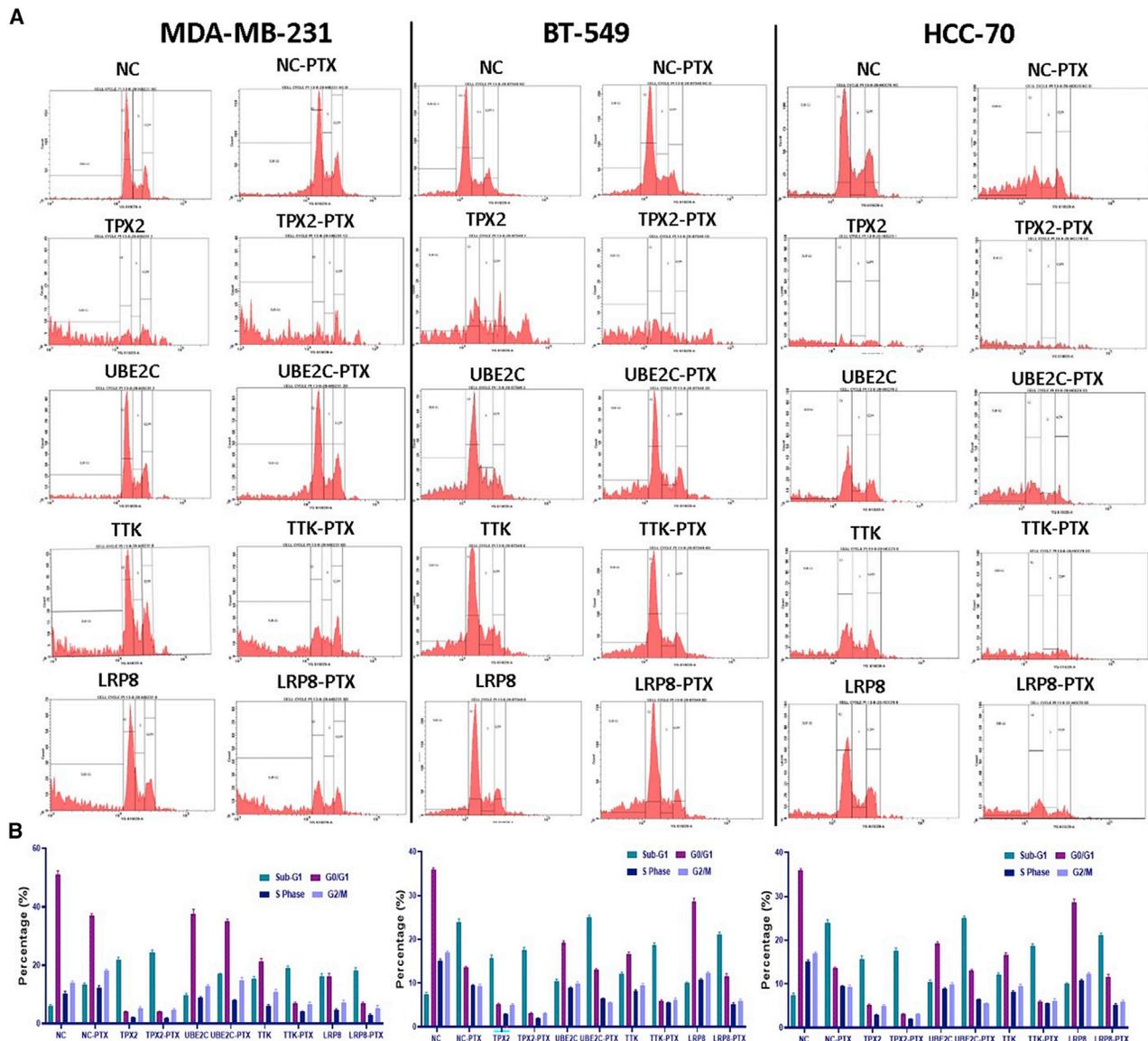
Figure 7. Cell cycle analysis of TNBC models in response to TPX2, UBE2C, TTK, and LRP8 depletion

(A) Histograms illustrate the changes in cell cycle of the indicated TNBC model post-knockdown of TPX2, UBE2C, TTK, or LRP8 as single agent or in combination with paclitaxel (20 nM). (B) Quantification of cell cycle distribution from (A) (n = 3).

Gene set enrichment analysis (GSEA) and modeling of gene interaction

Differentially expressed genes from the RNA-seq analysis were imported into the IPA software (Ingenuity Systems, USA) as we previously described.[38] Functional regulatory networks and canonical pathways were determined using URA, downstream effects analysis (DEA), MNs, and casual network analysis (CNA) prediction algorithms. IPA uses a precise database to paradigm functional regulatory networks from a list of individual genes and determines a statistical score, the Z score, for each network, according to the fit of the network to the set of focus genes. The biological functions assigned to each network are ranked according to the significance of that biological function to the network.[39]

Maintenance of cancer cell lines

Human TNBC (MDA-MB-231, BT-549 cell lines) were cultured in Dulbecco's modified Eagle's medium (DMEM), while HCC-70 were cultured in RPMI-1640. All culture media were supplemented with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin (penstrep). all were purchased from Thermo Scientific (Rockford, IL, USA). Cells were cultured as an adherent monolayer at 37°C under 5% CO_2 in a humidified incubator.
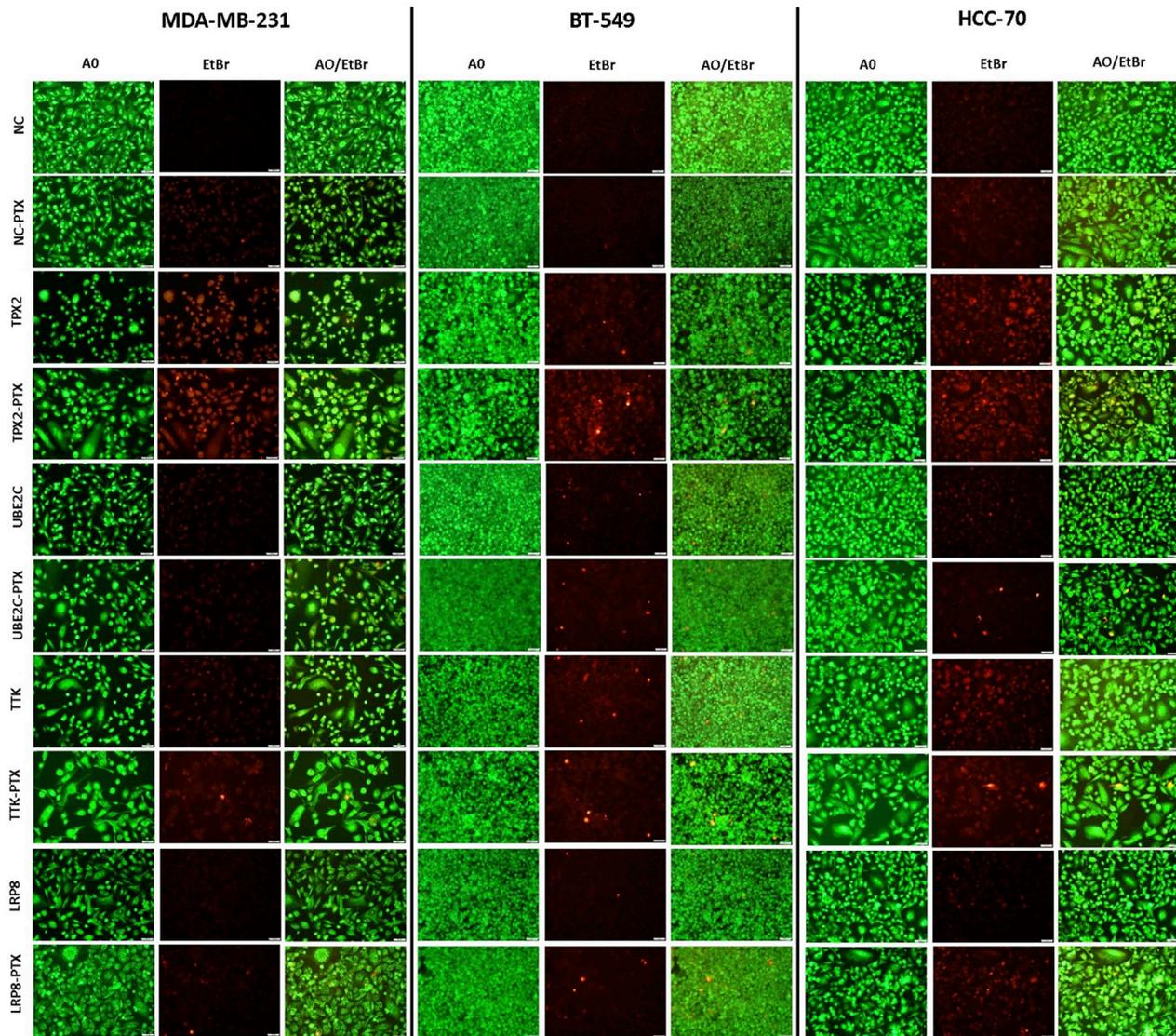
Figure 8. Dead-live staining of TNBC models in response to TPX2, UBE2C, TTK, and LRP8 depletion ± PTX

Representative fluorescence images for TNBC models post-siRNA-mediated knockdown of TPX2, UBE2C, TTK, and LRP8 alone or in combination with paclitaxel (20 nM). Cells were stained on day 4 with acridine orange/ethidium bromide to detect live (green) and dead cells (red; necrotic).

siRNA transfection in TNBC cells

To investigate the functional role of selected siRNAs in regulating BC biology, TNBC MDA-MB-231, BT-549, and HCC-70 cells (0.168×10^6 cells/mL) were transfected with the selected siRNAs (TPX2, UBE2C, CDCA7, MELK, NFE2L3, TTK, DDX39A, and LRP8) and negative control purchased from Ambion. Transfection was performed using a reverse transfection protocol as previously described.[38] In brief, siRNAs at a final concentration of 30 nM were diluted in 50 μL of Opti-MEM ($\neq$11058-021; Gibco, Carlsbad, CA, USA), and 1.5 μL of Lipofectamine 2000 (cat. no. 52758; Invitrogen) was diluted in 50 μL Opti-MEM. The diluted siRNAs and Lipofectamine 2000 were mixed and then incubated at ambient temperature for 20 min. One hundred microliters of transfection mixture were added to the 24-well tissue culture plate, and subsequently 300 μL of MDA-MB-231, BT-549, and HCC-70 (0.168×10^6 cells/mL) in transfection medium (Opti-MEM) were added to each well. Twenty hours later, the transfection medium (complete DMEM without antibiotics) was added into each well.

Colony-formation assay

The colony-forming ability of MDA-MB-231, BT-549, and HCC-70 cells transfected with TPX2, UBE2C, CDCA7, MELK, NFE2L3, TTK, DDX39A, and LRP8 siRNAs or siRNA-negative control in

Table 1. SYBR green primer sequences used in current study

No.	Names	Forward sequences	Reverse sequences
1	GAPDH	5′-GGAGCGAGATCCCTCCAAAAT-3′	5′-GGCTGTTGTCATACTTCTCATGG-3′
2	TPX2	5′-AGAAGAGGTGCTCTGAAGGC-3′	5′-CCAGCTGAAAAGGTTCCTGAACTA-3′
3	UBE2C	5′-GTTCCTGTCTCTCTGCCAACG-3′	5′-GTCTGATTCAGGGAAGGCAGAA-3′
4	CDCA7	5′-GGCTTTTCAGAAAGTGAGGTGC-3′	5′-CCTACAGCCTTCCCGAACTG-3′
5	MELK	5′-TTCTTAGGAACGCCGTACCAG-3′	5′-AAGCCACCTGTCCCAATAGTT-3′
6	NFE2L3	5′-GCAGAGAAACCTGACTGGGA-3′	5′-CCATCACTGATACTGCCTGGA-3′
7	TTK	5′-CTTTTCATTTCCCCAGCGCA-3′	5′-CCCGAGTTATCTGTAGTATCAGCA-3′
8	DDX39A	5′-AGTTGGAAGTGTCTCTTAGCAGC-3′	5′-AGCTGTGGATGGAAACGTAGG-3′
9	LRP8	5′-AAGTGTGTACCTGCCTCGTG-3′	5′-CGTCACCACAGTCGTCGTC-3′

the absence or presence of PTX was determined using a clonogenic assay as described before.[40] In brief, TNBC cells were transfected in a 24-well flat-bottom tissue culture plate, and 48 hr later, PTX was added at 20 nM final concentration. On day 7, the plates were washed and then stained with crystal violet and were subsequently scanned, and the number of colonies were observed under inverted microscope as we described before.[40]

Cell cycle analysis using flow cytometry

Cell cycle analyses were conducted on MDA-MB-231, BT-549, and HCC-70 cells post post transfection with siRNA targeting TPX2, UBE2C, TTK, and LRP8 as single agent or in combination with PTX as described before.[38] In brief, TNBC cells were transfected in a 6-well flat-bottom tissue culture plate, and 48 hr later, PTX was added at 20 nM final concentration. On day 4, floating cell population was collected and pooled with adherent trypsinized cells, followed by washing and fixing with 70% ethanol, and they were stored at 4° overnight. Before staining, cells were washed with PBS twice and incubated in RNase A (100 μg/mL) and propidium iodide (PI; 50 μg/mL) staining solution and then subjected to cell cycle analysis using BD LSRFortessa X-20 flow cytometer (BD Biosciences, CA, USA) at the FL3 channel.

Detection of cell death using fluorescence microscopy

The AO/EtBr fluorescence staining method was used to assess apoptosis/necrosis in post-knockdown genes or combination with PTX of TNBC cells, as we described before.[38] In brief, TNBC cells were transfected in a 24-well flat-bottom tissue culture plate, and 48 hr later, PTX was added at 20 nM final concentration. On day 4, TNBC cells were washed twice with PBS and subsequently stained with dual fluorescent staining solution containing 100 μg/mL AO and 100 μg/mL EtBr (AO/EtBr, Sigma Aldrich, St. Louis, MO, USA) for 2 min; subsequently, the cells were observed and imaged under an Olympus IX73 fluorescence microscope (Olympus, Tokyo, Japan). The differential uptake of AO/EtBr allows the identification of viable and non-viable cells. Principally, AO was used to visualize the number of cells that had undergone apoptosis, while EtBr-positive cells indicated necrotic cells.

Total RNA isolation, cDNA synthesis, and quantitative real-time PCR

The mRNA expression of TPX2, UBE2C, CDCA7, MELK, NFE2L3, TTK, DDX39A, and LRP8 in MDA-MB-231, BT-549 and HCC70 TNBC models transfected with targeting or scrambled control siRNAs was determined using quantitative real-time PCR as described before.[41] In brief, TNBC cells were transfected in a 6-well flat-bottom tissue culture plate. On day 4, total RNA was isolated from transfected siRNAs and negative control cells using a total RNA purification kit (Norgen Biotek, ON, Canada) according to the manufacturer's protocol. The concentrations and purity of extracted RNA was measured using NanoDrop 2000 (Thermo Fisher Scientific, DE, USA). Subsequently, 1,000 ng of total RNA was reverse transcribed using a high-capacity cDNA reverse transcript kit (Applied Biosystems, Foster City, CA, USA). Gene expression was quantified using PowerUP SYBR green master mix and a QuantStudio 7 flex real-time PCR system (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. The relative FC in mRNA expression was calculated using the $2-\Delta\Delta Ct$ method,[42] where the average of ΔCt values for the amplicon of interest were normalized to that of an endogenous gene (GAPDH), compared with control samples. The primer sequences used in this study are listed in Table 1. The primers were designed using Primer3 (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

Statistical and survival analysis

Statistical analyses and graphing were performed using GraphPad Prism 8.0 software (GraphPad, San Diego, CA, USA). The Benjamini-Hochberg FDR method was used for multiple testing corrections. For IPA analyses, a Z score ($-2.0 \geq Z \geq 2.0$) was considered significant. The log-rank test was used to compare the outcome between expression groups. Statistical analyses to compare specific gene expression and graphing were performed using Graphpad Prism 6.0 software (Graphpad Software, San Diego, CA, USA).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtm.2021.01.013.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

R.E. and R.V. performed lab experiments and wrote the first draft. H.S. wrote original draft of the manuscript and reviewed and edited it. N.M.A. conceptualized the study; performed bioinformatics analysis, project administration, and funding acquisition; and reviewed and finalized the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Foulkes, W.D., Smith, I.E., and Reis-Filho, J.S. (2010). Triple-negative breast cancer. N. Engl. J. Med. 363, 1938–1948.

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.

3. Venkitaraman, R. (2010). Triple-negative/basal-like breast cancer: clinical, pathologic and molecular features. Expert Rev. Anticancer Ther. 10, 199–207.

4. Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., and Narod, S.A. (2007). Triple-negative breast cancer: clinical features and patterns of recurrence. Clin. Cancer Res 13, 4429–4434.

5. Perou, C.M., Sørlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. (2000). Molecular portraits of human breast tumours. Nature 406, 747–752.

6. Hudis, C.A., and Gianni, L. (2011). Triple-negative breast cancer: an unmet medical need. Oncologist 16 (Suppl 1), 1–11.

7. Aw Yong, K.M., Ulintz, P.J., Caceres, S., Cheng, X., Bao, L., Wu, Z., Jiagge, E.M., and Merajver, S.D. (2020). Heterogeneity at the invasion front of triple negative breast cancer cells. Sci. Rep. 10, 5781.

8. Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell 173, 879–893.e13.

9. Wu, S., Zhang, H., Fouladdel, S., Li, H., Keller, E., Wicha, M.S., Omenn, G.S., Azizi, E., and Guan, Y. (2020). Cellular, transcriptomic and isoform heterogeneity of breast cancer cell line revealed by full-length single-cell RNA sequencing. Comput. Struct. Biotechnol. J. 18, 676–685.

10. 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.

11. Lawrence, R.T., Perez, E.M., Hernández, D., Miller, C.P., Haas, K.M., Irie, H.Y., Lee, S.I., Blau, C.A., and Villén, J. (2015). The Proteomic Landscape of Triple-Negative Breast Cancer. Cell Rep. 11, 990.

12. 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.

13. Di Cosimo, S., Appierto, V., Silvestri, M., Pruneri, G., Vingiani, A., Perrone, F., Busico, A., Folli, S., Scaperrotta, G., de Braud, F.G., et al. (2019). Targeted-Gene Sequencing to Catch Triple Negative Breast Cancer Heterogeneity before and after Neoadjuvant Chemotherapy. Cancers (Basel) 11, 1753.

14. Jiang, Y.Z., Ma, D., Suo, C., Shi, J., Xue, M., Hu, X., Xiao, Y., Yu, K.D., Liu, Y.R., Yu, Y., et al. (2019). Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. Cancer Cell 35, 428–440.e5.

15. Burstein, M.D., Tsimelzon, A., Poage, G.M., Covington, K.R., Contreras, A., Fuqua, S.A., Savage, M.I., Osborne, C.K., Hilsenbeck, S.G., Chang, J.C., et al. (2015). Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin. Cancer Res 21, 1688–1698.

16. Muranen, T.A., Blomqvist, C., Dörk, T., Jakubowska, A., Heikkilä, P., Fagerholm, R., Greco, D., Aittomäki, K., Bojesen, S.E., Shah, M., et al. (2016). Patient survival and tumor characteristics associated with CHEK2:p.I157T - findings from the Breast Cancer Association Consortium. Breast Cancer Res. 18, 98.

17. Denkert, C., Loibl, S., Noske, A., Roller, M., Müller, B.M., Komor, M., Budczies, J., Darb-Esfahani, S., Kronenwett, R., Hanusch, C., et al. (2010). Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer. J. Clin. Oncol. 28, 105–113.

18. García-Teijido, P., Cabal, M.L., Fernández, I.P., and Pérez, Y.F. (2016). Tumor-Infiltrating Lymphocytes in Triple Negative Breast Cancer: The Future of Immune Targeting. Clin. Med. Insights Oncol. 10 (Suppl 1), 31–39.

19. Cho, S.Y. (2019). Identification of ERBB Pathway-Activated Cells in Triple-Negative Breast Cancer. Genomics Inform. 17, e3.

20. Greer, Y.E., Gilbert, S.F., Gril, B., Narwal, R., Peacock Brooks, D.L., Tice, D.A., Steeg, P.S., and Lipkowitz, S. (2019). MEDI3039, a novel highly potent tumor necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL) receptor 2 agonist, causes regression of orthotopic tumors and inhibits outgrowth of metastatic triple-negative breast cancer. Breast Cancer Res. 21, 27.

21. Gu, Y., Wang, W., Wang, X., Xie, H., Ye, X., and Shu, P. (2019). Integrated network analysis identifies hsa-miR-4756-3p as a regulator of FOXM1 in Triple Negative Breast Cancer. Sci. Rep. 9, 13830.

22. Zanin, R., Pegoraro, S., Ros, G., Ciani, Y., Piazza, S., Bossi, F., Bulla, R., Zennaro, C., Tonon, F., Lazarevic, D., et al. (2019). HMGA1 promotes breast cancer angiogenesis supporting the stability, nuclear localization and transcriptional activity of FOXM1. J. Exp. Clin. Cancer Res. 38, 313.

23. Ring, A., Nguyen, C., Smbatyan, G., Tripathy, D., Yu, M., Press, M., Kahn, M., and Lang, J.E. (2018). CBP/β-Catenin/FOXM1 Is a Novel Therapeutic Target in Triple Negative Breast Cancer. Cancers (Basel) 10, 525.

24. Shaath, H., Toor, S.M., Nair, V.S., Elkord, E., and Alajez, N.M. (2019). Transcriptomic Analyses Revealed Systemic Alterations in Gene Expression in Circulation and Tumor Microenvironment of Colorectal Cancer Patients. Cancers (Basel) 11, 1994.

25. Song, B.N., and Chu, I.S. (2018). A gene expression signature of FOXM1 predicts the prognosis of hepatocellular carcinoma. Exp. Mol. Med. 50, e418.

26. Kim, I.M., Ackerson, T., Ramakrishna, S., Tretiakova, M., Wang, I.C., Kalin, T.V., Major, M.L., Gusarova, G.A., Yoder, H.M., Costa, R.H., and Kalinichenko, V.V. (2006). The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. Cancer Res. 66, 2153–2161.

27. Kalin, T.V., Wang, I.C., Ackerson, T.J., Major, M.L., Detrisac, C.J., Kalinichenko, V.V., Lyubimov, A., and Costa, R.H. (2006). Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. Cancer Res. 66, 1712–1720.

28. Caldwell, S.A., Jackson, S.R., Shahriari, K.S., Lynch, T.P., Sethi, G., Walker, S., Vosseller, K., and Reginato, M.J. (2010). Nutrient sensor O-GlcNAc transferase regulates breast cancer tumorigenesis through targeting of the oncogenic transcription factor FoxM1. Oncogene 29, 2831–2842.

29. Teh, M.T., Gemenetzidis, E., Chaplin, T., Young, B.D., and Philpott, M.P. (2010). Upregulation of FOXM1 induces genomic instability in human epidermal keratinocytes. Mol. Cancer 9, 45.

30. Bollu, L.R., Shepherd, J., Zhao, D., Ma, Y., Tahaney, W., Speers, C., Mazumdar, A., Mills, G.B., and Brown, P.H. (2020). Mutant P53 induces MELK expression by release of wild-type P53-dependent suppression of FOXM1. NPJ Breast Cancer 6, 2.

31. Li, F., Su, M., Zhao, H., Xie, W., Cao, S., Xu, Y., Chen, W., Wang, L., Hou, L., and Tan, W. (2019). HnRNP-F promotes cell proliferation by regulating TPX2 in bladder cancer. Am. J. Transl. Res. 11, 7035–7048.

32. Kaistha, B.P., Honstein, T., Müller, V., Bielak, S., Sauer, M., Kreider, R., Fassan, M., Scarpa, A., Schmees, C., Volkmer, H., et al. (2014). Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. Br. J. Cancer 111, 1780–1787.

33. Wang, X., Yin, L., Yang, L., Zheng, Y., Liu, S., Yang, J., Cui, H., and Wang, H. (2019). Silencing ubiquitin-conjugating enzyme 2C inhibits proliferation and epithelial-

mesenchymal transition in pancreatic ductal adenocarcinoma. FEBS J. *286*, 4889–4909.

34. Kohler, R.S., Kettelhack, H., Knipprath-Mészaros, A.M., Fedier, A., Schoetzau, A., Jacob, F., and Heinzelmann-Schwarz, V. (2017). MELK expression in ovarian cancer correlates with poor outcome and its inhibition by OTSSP167 abrogates proliferation and viability of ovarian cancer cells. Gynecol. Oncol. *145*, 159–166.

35. Leinonen, R., Sugawara, H., and Shumway, M.; International Nucleotide Sequence Database Collaboration (2011). The sequence read archive. Nucleic Acids Res. *39*, D19–D21.

36. Venkatasubramanian, M., Chetal, K., Schnell, D.J., Atluri, G., and Salomonis, N. (2020). Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF. Bioinformatics *36*, 3773–3780.

37. Shaath, H., Vishnubalaji, R., Elkord, E., and Alajez, N.M. (2020). Single-Cell Transcriptome Analysis Highlights a Role for Neutrophils and Inflammatory Macrophages in the Pathogenesis of Severe COVID-19. Cells *9*, 2374.

38. Elango, R., Vishnubalaji, R., Manikandan, M., Binhamdan, S.I., Siyal, A.A., Alshawakir, Y.A., Alfayez, M., Aldahmash, A., and Alajez, N.M. (2019).

Concurrent targeting of BMI1 and CDK4/6 abrogates tumor growth in vitro and in vivo. Sci. Rep. *9*, 13696.

39. Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523–530.

40. Vishnubalaji, R., Hamam, R., Yue, S., Al-Obeed, O., Kassem, M., Liu, F.F., Aldahmash, A., and Alajez, N.M. (2016). MicroRNA-320 suppresses colorectal cancer by targeting SOX4, FOXM1, and FOXQ1. Oncotarget *7*, 35789–35802.

41. Elango, R., Alsaleh, K.A., Vishnubalaji, R., Manikandan, M., Ali, A.M., Abd El-Aziz, N., Altheyab, A., Al-Rikabi, A., Alfayez, M., Aldahmash, A., and Alajez, N.M. (2020). MicroRNA Expression Profiling on Paired Primary and Lymph Node Metastatic Breast Cancer Revealed Distinct microRNA Profile Associated With LNM. Front. Oncol. *10*, 756.

42. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402–408.