Integrated transcriptomic meta-analysis provides novel insights into breast cancer molecular subtypes

Akhilesh Kumar Bajpai[1,2], Sravanthi Davuluri[1,2], Kavitha Thirumurugan[1], Kshitish K. Acharya[2,3,4*]

[1]Structural Biology Lab, Centre for Biomedical Research, School of Bio Sciences & Technology (SBST), Vellore Institute of Technology (VIT) University, Vellore - 632014, Tamil Nadu, India.

[2]Shodhaka Life Sciences Pvt. Ltd., Electronic city, Phase I, Bengaluru (Bangalore) - 560100, Karnataka, India (www.shodhaka.com).

[3]Biological data Analyzers' Association (BdataA), Electronic City, Phase I, Bengaluru (Bangalore) - 560100, Karnataka, India (www.startbioinfo.com/BdataA)

[4]Insitute of Bioinformatics and Applied Biotechnology (IBAB), Phase I, Electronic City, Bengaluru (Bangalore) - 560 100, Karnataka, India (www.ibab.ac.in).

*Corresponding author

Kshitish K Acharya

Phone: 91-80-65316565

Fax: 91-80-28528904

Email: juneindia@gmail.com

Abstract

Breast cancer is the most common cancer in women worldwide. There are four major breast cancer subtypes (luminal A, luminal B, HER2-enriched and triple-negative/TNBC). TNBC is the most aggressive form with the worst prognosis. However, the differences among the subtypes have not been completely established at the molecular level, thereby limiting therapeutic and diagnostic options for TNBC. We performed a meta-analysis of microarray and RNA-sequencing data to identify candidate genes with an expression-based association in each molecular subtype of breast cancer. The protein interaction network of the candidate genes was analyzed to discover functionally significant gene clusters and hub genes. Potential therapeutic candidates for TNBC were explored through gene-miRNA interactions. We identified 316, 347, 382, and 442 candidate genes in luminal A, luminal B, HER2 and TNBC subtypes, respectively. A total of 135 (26 up- and 109 down-regulated) candidate genes were shared by all four subtypes. Functional analysis of the candidate genes indicated up-regulation of 'cell cycle' and 'p53 signaling' pathways and down-regulation of multiple signaling pathways. COL10A1 was found to be highly up-regulated in all subtypes. It may be a good target for research towards multiple types of applications, including therapeutics. KIF4A, a commonly up-regulated X-chromosome gene was significantly associated with the survival of breast cancer patients. Protein interaction network and centrality analysis revealed that low-moderately differentially regulated genes play an important role in functional cascades across proteins in breast cancer subtypes and may be potential candidates for therapeutics. Targeting FN1 (fibronectin 1), the key up-regulated hub gene by miR-1271 5p may be an important molecular event to be targeted for potential therapeutic application in TNBC.

Keywords: Breast cancer; protein interaction network; integrated analysis; biomarker; microarray; RNA-sequencing.

Introduction

Breast cancer is the most common cancer in women, accounting for 24% of all the cancer cases, and with an estimated number of ~2 million cases worldwide. It is responsible for ~600,000 cancer related deaths and accounts for 15% of all cancer related mortalities in women worldwide (Torre et al. 2016). Based on the expression of signature genes, breast cancer has been classified into four major molecular subtypes: luminal A, luminal B, Her2-enriched (HER2), and basal-like/triple-negative breast cancer (TNBC) (Perou et al. 2000; Sorlie et al. 2001). Luminal A is the most common sub-type (~50% cases) followed by luminal B (~20% cases), HER2 and TNBC (~15% each). Among these, luminal A has the best prognosis and TNBC has the worst. TNBC is the most aggressive sub-type, which is responsible for poorer survival and higher incidences of distant metastasis (Anders and Carey 2009). There are limited targeted therapy options for TNBC due to the lack of expression of estrogen (ER), progesterone (PR), and HER2 receptors (Bianchini et al. 2016; Kumar et al. 2017).

High-throughput technologies, such as microarray and RNA-sequencing have been used extensively to study the mechanisms of various diseases including breast cancer. Multiple studies have used these large-scale datasets to identify better prognostic and therapeutic markers in breast cancer subtypes, understand the underlying disease mechanisms, identify novel metastatic genes, and predict metastatic outcomes. With the availability of multi-omics high-throughput data, it has been possible to explore the cellular systems as a network of interacting molecules, including proteins, miRNAs, and lncRNAs etc. For instance, a study by Hsu et al., (Hsu et al. 2019) identified six novel immunoglobulin genes as biomarkers for better prognosis of triple-negative breast cancer using publicly available microarray studies. Similarly, Dong et al., (Dong et al. 2018) identified key genes and pathways in triple-negative breast cancer based on an integrated bioinformatics approach. A study by Guo et al., (Guo et al. 2017) identified and validated nine novel TNBC associated biomarkers by integrating gene expression data and protein interactions. Furthermore, microarray gene expression data was used to discover novel genes associated with lymph-node metastasis in TNBC subtype (Mathe et al. 2015). Also, the authors correlated the gene expression changes with miRNA and confirmed the relationship between specific miRNAs and lymph-node metastasis. Yu et al., used gene co-expression and protein interaction networks to predict 7 putative genes in breast cancer (Yue et al. 2016). The

3

authors showed that the 7 gene signature could be a useful predictor of disease prognosis in breast cancer patients. A recent integrated network analysis and machine learning approach was able to accurately classify TNBC from non-TNBC based on a 16 gene signature (Naorem et al. 2019). Comprehensive transcriptomic and genomic analyses have identified novel molecular TNBC subtypes and subtype specific genes. Liu et al., (Liu et al. 2016) classified 165 TNBC tumors into four distinct subtypes, including immunomodulatory, luminal androgen receptor, mesenchymal-like, and basal-like and immune suppressed subtypes based on gene expression microarray data. Similar TNBC subtypes were identified by another study based on RNA and DNA profiling analyses of 198 tumors (Burstein et al. 2015). Although substantial research has been performed on breast cancer, including different molecular subtypes, there is a gap in understanding the molecular basis of each subtype. In addition, more focused studies are required to identify better therapeutic targets for the aggressive TNBC subtype.

In the current study, we have performed meta-analysis of microarray and RNA sequencing data to identify candidate genes associated with four molecular subtypes of breast cancer (luminal A, luminal B, HER2 and TNBC). Furthermore, protein interaction network analysis of the candidate genes was explored to identify functionally important molecules in each subtype. In addition, miRNA-gene interactions were explored to identify potential therapeutic targets for TNBC subtype.

Methods

Microarray and RNA sequencing data collection and processing

Literature search was performed using selected search engines, as indicated before (Bajpai et al., 2011). The microarray data corresponding to breast cancer molecular subtypes and breast normal tissues were downloaded from Gene Expression Omnibus (GEO) repository (Barrett et al. 2013). The raw data files of cancer and normal samples were pre-processed using *MAS5* algorithm (Hubbell et al. 2002) in a R-based environment to obtain absolute calls ('P', 'A', and 'M') of each gene in each sample. The normalized files with absolute calls were processed using a meta-analysis approach reported earlier (Acharya et al. 2010) to obtain genes with consistently transcribed status ('P') along with their reliability scores. Briefly, the method assigns a reliability score of +2 or -2 for a gene to be 'transcribed' or 'dormant (A)', respectively in each sample/hybridization. The score for each gene was then added across multiple

4

samples/hybridizations to obtain a cumulative reliability score, in each subtype as well as normal breast tissue. The cumulative reliability score was then used to assign the final expression status (transcribed/dormant) for the genes. Genes with 'transcribed' status having a reliability score of at least 20 (present in at least 10 samples/hybridizations) were considered further. Differentially expressed genes in each breast cancer subtype compared to normal breast samples were obtained by using the GEO2R function implemented in GEO repository. The GEO2R function uses the normalized data submitted by the authors and applies Linear Models for Microarray data ('*limma*') (Ritchie et al. 2015) algorithm to identify differentially expressed genes between the groups. Genes with a |log2fold ≥2| and FDR (B&H) adjusted *P* value of <0.01 in at least 2 independent studies were considered further.

The Cancer Genome Atlas (TCGA) RNA sequencing based raw read count data for the breast cancer molecular subtypes and normal breast samples were downloaded from Firebrowse cancer data portal (http://firebrowse.org/). Only primary tumor samples were considered for the analysis. Differentially expressed genes between each breast cancer molecular subtype and paired normal breast tissues were obtained using R based *edgeR* package (Robinson et al. 2010). Genes with |log2fold ≥2| and FDR (B&H) *P* value of <0.01 were considered to be significant.

Identification of candidate genes associated with breast cancer molecular subtypes and functional analysis

Meta-analysis of microarray and RNA-seq data was performed to identify candidate genes associated with breast cancer subtypes. Briefly, genes having both 'transcribed' and 'up-regulated' status in cancer (henceforth referred to as 'up-regulated') and 'transcribed in normal breast' and 'down-regulated' status in cancer (henceforth referred to as 'down-regulated') based on microarray technology were considered as a preliminary-set of genes associated with each breast cancer subtype. Among these, genes in agreement with TCGA RNA-sequencing data were considered as the final set of candidate genes in each subtype. Based on the fold change obtained from TCGA-RNA-seq data, the candidate genes were classified at 3 levels of differential expression: low (log2 fold ≤3), moderate (log2 fold between ≥3.5 and ≤5 fold) or high (log2 fold ≥5.5).

The functional analysis, including Gene Ontology (GO), pathway, and chromosome enrichment analysis of the candidate genes was performed using DAVID (Huang da et al. 2009).

Annotations with a *P*-value <0.05 were considered significant. The significant GO terms from each subtype were further submitted to REVIGO (Supek et al. 2011) to remove the redundancy and obtain a more meaningful set of annotations. Heatmaps of the commonly regulated candidate genes were constructed using the ClustVis tool (Metsalu and Vilo 2015) based on TPM values in each molecular subtype.

Network analysis of the candidate genes associated with each breast cancer subtype

The known protein interactions (i.e., those with 'experimental' and 'database' evidence) among the differentially expressed candidate genes in each breast cancer subtype were obtained from the STRING database (Szklarczyk et al. 2019), as suggested to be suitable by a recent study (Bajpai et al., 2020) with a medium confidence threshold (confidence score, 0.4). The interactions were visualized and analyzed using the Cytoscape network analysis tool (Shannon et al. 2003), and MCODE tool (Bader and Hogue 2003) was used to identify protein clusters in each network. Significantly enriched (*P* <0.05) annotations for the selected clusters were obtained using the DAVID analysis tool (Huang da et al. 2009). Using CytoHubba (Chin et al. 2014), network statistics, such as node degree, betweenness and closeness centrality were determined. Significance analysis of the network topological features between different groups of differentially regulated genes (low, moderate and high) was performed using the two-tailed, unequal variance t-test. The miRNAs for the target genes were predicted using TargetScan (Agarwal et al. 2015), miRanda (Betel et al. 2008), and miRDB (Wong and Wang 2015) and the ones identified by all the three tools were considered further. The expression of miRNAs was verified using the literature as well as dbDEMC (Yang et al. 2017) and UALCAN (Chandrashekar et al. 2017) databases.

Results

Breast cancer and normal breast tissue samples used for derivation of candidate genes

A total of 8 microarray gene expression studies were identified from GEO repository corresponding to four different molecular subtypes of breast cancer (luminal A, luminal B, HER2, and TNBC). Among the eight studies identified, 6 had normal breast samples along with the cancer samples and were used for the derivation of differentially expressed genes. Two of the studies (GSE45827 and GSE31448) contained samples for all the four molecular subtypes. Triple-negative breast cancer had the highest number of samples (465) followed by luminal A

(146), HER2 (95) and luminal B (83). The samples corresponding to only Affymetrix platforms
were used for the derivation of absolute calls because of their compatibility with the meta-
analysis approach used. Furthermore, additional normal breast samples collected from GEO
(data not shown) were used for the derivation of absolute calls.

The TCGA RNA sequencing data was obtained for a total of 572 samples across four breast
cancer molecular subtypes. Luminal A had highest number of samples (301) followed by luminal
B (150), TNBC (95) and HER2 (26). There were 90 samples corresponding to normal breast
tissue.

Candidate genes associated with breast cancer molecular subtypes
Six microarray studies were used to identify differentially expressed genes in four breast cancer
molecular subtypes compared to normal breast with a |log2 fold ≥ 2| and FDR adjusted (B&H) *P*
value of <0.01. Overall, the number of down-regulated genes obtained was more than that of the
up-regulated genes in most studies. Triple-negative breast cancer (TNBC) subtype had the
highest number of differentially expressed genes. Furthermore, 567 differentially expressed
genes in luminal A, 458 in luminal B, 584 in HER2, and 720 genes in TNBC were reproduced by
at least two microarray studies. Approximately, 80% of these genes also had 'Transcribed' status
in breast cancer (for up-regulated genes) or normal breast samples (for down-regulated genes),
and were considered as preliminary set of breast cancer-subtype associated candidate genes.
These included 499 differentially expressed genes in luminal A, 411 in luminal B, 515 in HER2,
and 638 genes in TNBC. RNA sequencing analysis identified a total of 10,938 genes to be
differentially expressed across four breast cancer subtypes with a |log2 fold ≥2| and FDR (B&H)
P-value of <0.01. Among these, 5,766 genes were up-regulated and 5,172 were down-regulated.
Triple negative breast cancer had maximum number of differentially expressed genes (2,950
genes).

A total of 1,487 candidate genes were identified based on the meta-analysis of microarray and
RNA sequencing technologies (common to both the technologies) across four molecular
subtypes of breast cancer. We identified 316 candidate genes (63 up- and 253 down-regulated) in
luminal A, 347 in luminal B (86 up- and 261 down-regulated), 382 in HER2 (101 up- and 281
down-regulated), and 442 genes in TNBC (199 up- and 243 down-regulated) subtypes.

Identification of COL10A1 as a potential diagnostic marker for breast cancer

The comparison of candidate genes among the subtypes showed 102 and 280 genes to be up- and down-regulated, respectively, in at least two molecular subtypes, whereas, 26 up- and 109 down-regulated genes were shared by all the four breast cancer subtypes (**Fig. 1**). In luminal A and luminal B subtypes, COL10A1 had the highest fold change (7.6-fold, $P = 2.19E-89$ and 7.4-fold, $P = 5.74E-63$, respectively), whereas, in HER2 and TNBC subtypes, MMP1 had the highest fold change for up-regulation (8.9-fold, $P = 1.23E-35$ and 8.2-fold, $P = 2.44E-42$). Among the top 10 up-regulated genes (based on fold change), COL10A1, COL11A1, and MMP11 were common across all four subtypes. While, COL10A1 and COL11A1 were up-regulated with at least 6 fold, MMP11 was up-regulated with more than 5.5 fold in all the subtypes. When the average fold was considered across all the four subtypes, COL10A1 had a highest mean fold of 7.1, followed by COL11A1 and MMP11 with 6.8 and 6.3 folds, respectively. In addition, the mRNA and protein expression of COL10A1 was confirmed to be significantly higher in breast cancer molecular subtypes compared to normal breast tissues (Chandrashekar et al. 2017). Thus, COL10A1 may be used as a potential diagnostic marker for breast cancer, irrespective of the subtype, whereas RAB26 and MMP1 may be specifically used as markers for ER/PR-positive and -negative breast cancers, respectively. Among the down-regulated genes, LEP and CA4 were in the top 5 in all four subtypes. Survival analysis indicated that among the commonly up-regulated genes, KIF4A was significantly associated with the survival of breast cancer patients, whereas, among the commonly down-regulated genes, significant association was observed for ADIPOQ, C2orf40, CD300LG, CIDEA, CLDN11, CNN1, and SDPR.
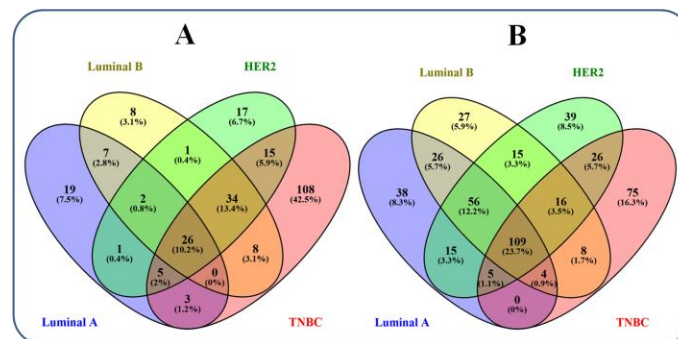

Fig. 1 Candidate genes shared across four breast cancer subtypes. A: up-regulated genes; B: down-regulated genes.

Functional analysis of the candidate genes associated with breast cancer molecular subtypes

The functional analysis of the candidate genes in each breast cancer subtype showed enrichment of multiple processes and pathways. Processes, such as cell division, cell proliferation, apoptosis, and chromosome segregation were found to be significantly enriched by the up-regulated genes in all the breast cancer subtypes. The down-regulated genes were significantly enriched for metabolic related processes, including cholesterol homeostasis, glucose homeostasis, and lipid metabolic process. Pathway enrichment analysis showed up-regulated genes to be significantly enriched for 'cell cycle' and 'p53 signaling' pathways in all four molecular subtypes (**Fig. 2**). The up-regulated genes in luminal A were specifically enriched for 'PI3K –Akt signaling', 'focal adhesion', and 'protein digestion and absorption' pathways. The up-regulated genes in TNBC were enriched for multiple pathways including, 'Fanconi anemia', 'HTLV-I infection', and 'homologous recombination' pathways. The down-regulated genes in each subtype were significantly enriched for multiple signaling pathways, such as 'PPAR', 'adipocyte', and 'MAPK' signaling pathways. Other commonly down-regulated pathways in all four subtypes include 'focal adhesion', 'Tyrosine metabolism', 'regulation of lipolysis', and 'drug metabolism-cytochrome P450'. Many pathways were specifically down-regulated in each breast cancer subtype. For example, 'glycolysis' and 'pyruvate metabolism' in luminal A, 'Leukocyte transendothelial migration', in luminal B, multiple signaling pathways in HER2, and 'ABC transporter', 'nitrogen metabolism', 'renin secretion', and 'axon guidance' in TNBC subtype (**Fig. 2**). Analysis of 26 commonly up-regulated genes showed 'cell cycle', 'p53 signaling,' and 'oocyte meiosis' pathways to be significantly enriched. The 109 commonly down-regulated genes were significantly enriched for multiple signaling pathways.

Chromosomal distribution showed significant enrichment of chromosome 1, 4, 8, 15, 17, and 20 for the up-regulated candidate genes in different molecular subtypes. While chromosome 15 and 17 were most significantly enriched for the up-regulated genes in luminal B ($P = 0.014$) and HER2 ($P = 0.015$) subtypes, respectively, chromosome 4 and 20 were most significantly enriched for luminal A ($P = 0.006$) and TNBC ($P = 0.022$) subtypes, respectively. For the down-regulated genes, chromosome 3 was most significantly enriched in three breast cancer subtypes (luminal B, $P = 0.018$; HER2, $P = 0.0009$; TNBC, $P = 0.028$) covering around 10% of the genes

in each, whereas, chromosome 11 was most significantly enriched for the down-regulated candidate genes in luminal A ($P = 0.029$) subtype (covering ~10% of the genes). As expected, there were no genes from Y chromosome in any subtype.
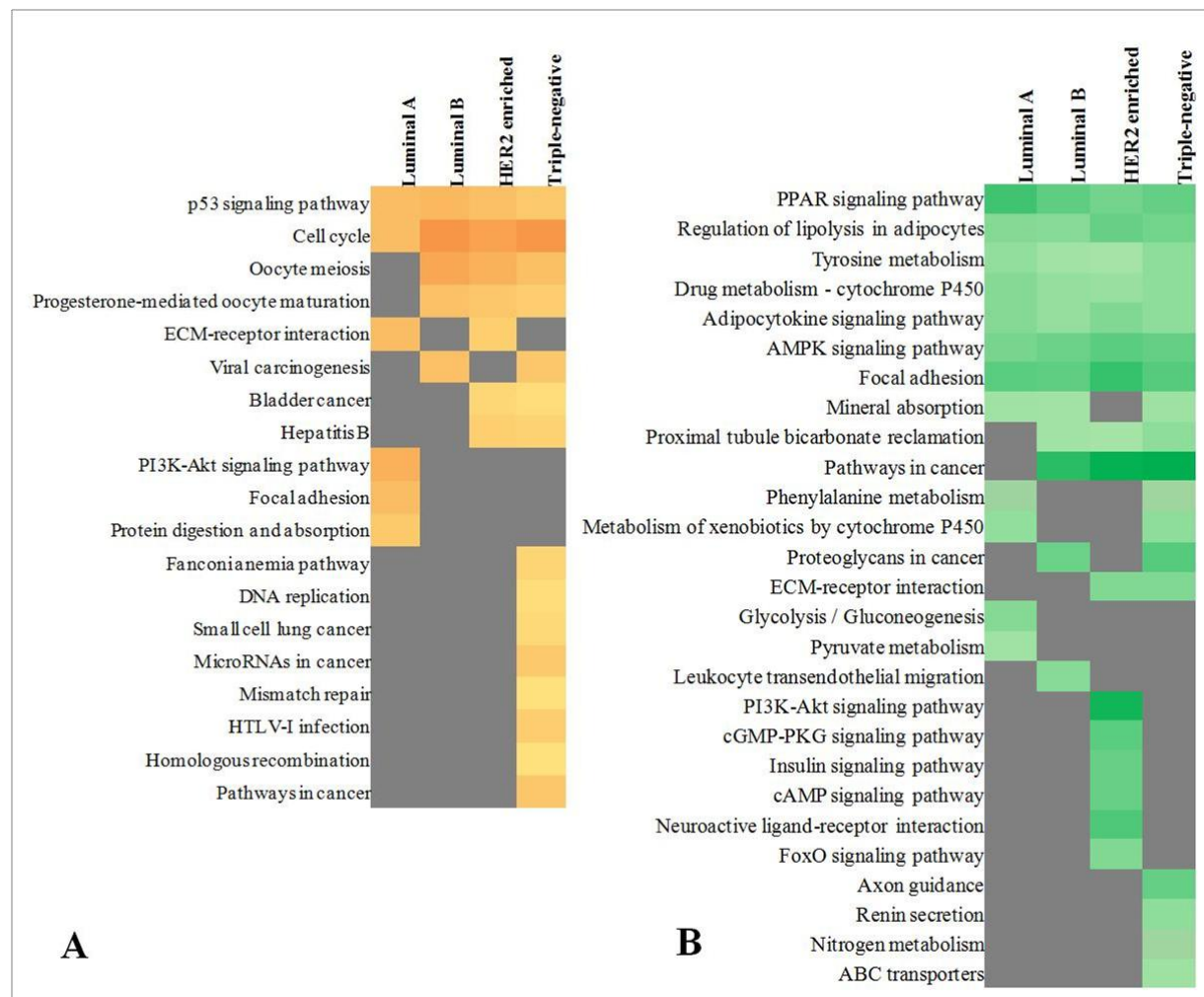

Fig. 2 KEGG pathways significantly enriched (p <0.05) by (A) up-regulated and (B) down-regulated genes in breast cancer subtypes. Orange/Green box: presence of pathway; dark grey box: absence of pathway. The percentage of genes involved in each pathway in each subtype is represented by the intensity of orange/green color (dark color: higher percentage; light color: lower percentage).

Protein interaction network analysis of candidate genes associated with breast cancer molecular subtypes

Protein-protein interaction network was constructed among the differentially expressed candidate genes for each breast cancer subtype. The up-regulated genes were found to be more tightly

connected than the down-regulated genes. Overall, highly up- or down-regulated genes had comparatively lesser number of interactions. The protein interaction networks along with the analysis of top 3 clusters in each subtype have been described below.

The protein interaction network revealed that 143 candidate genes in luminal A were connected with 220 edges (**Fig. 3**). Cluster analysis of the network identified a total of 10 clusters containing 42 proteins, of which 16 proteins (38%) were covered in the top 3 clusters (**Fig. 3**). The cluster 1 was most significantly enriched for 'chemokine signaling' and 'cytokine-cytokine receptor interaction' pathways, whereas, cluster 2 represented processes, such as 'muscle contraction and 'actin filament organization', and cluster 3 was significantly enriched for 'cell cycle', and 'p53 signaling' pathways (**Fig. 3**). Furthermore, analysis of the protein interaction network using identified EGFR to be the key hub molecule, followed by PPARγ and CDK1 based on betweenness centrality. EGFR and PPARγ were down-regulated, whereas, CDK1 was up-regulated in luminal A subtype.

Among the 347 candidate genes in luminal B, 172 were found to be connected with 393 edges (**Fig. 3**). Cluster analysis of the protein interaction network identified 12 clusters with 64 nodes, 50% of which were covered by the top 3 clusters (**Fig. 3**). Cluster 1 was highly connected (91 edges) with 14 up-regulated nodes and was significantly enriched for cell cycle related processes and pathways. Based on functional annotations, genes in cluster 1 of luminal B and cluster 3 of luminal A were involved in similar processes and pathways. Cluster 2 had 11 down-regulated genes and was significantly enriched for 'chemokine signaling'. Cluster 3 contained 7 up-regulated genes connected with 21 edges, and was enriched for processes, such as 'microtubule based movement', 'antigen processing and presentation', 'vesicle mediated transport', and 'cytokinesis'. Based on the betweenness centrality, PPARγ was found to be the key node, followed by EZH2 and CAV1. PPARγ and CAV1 were down-regulated, whereas, EZH2 was up-regulated in luminal B subtype.
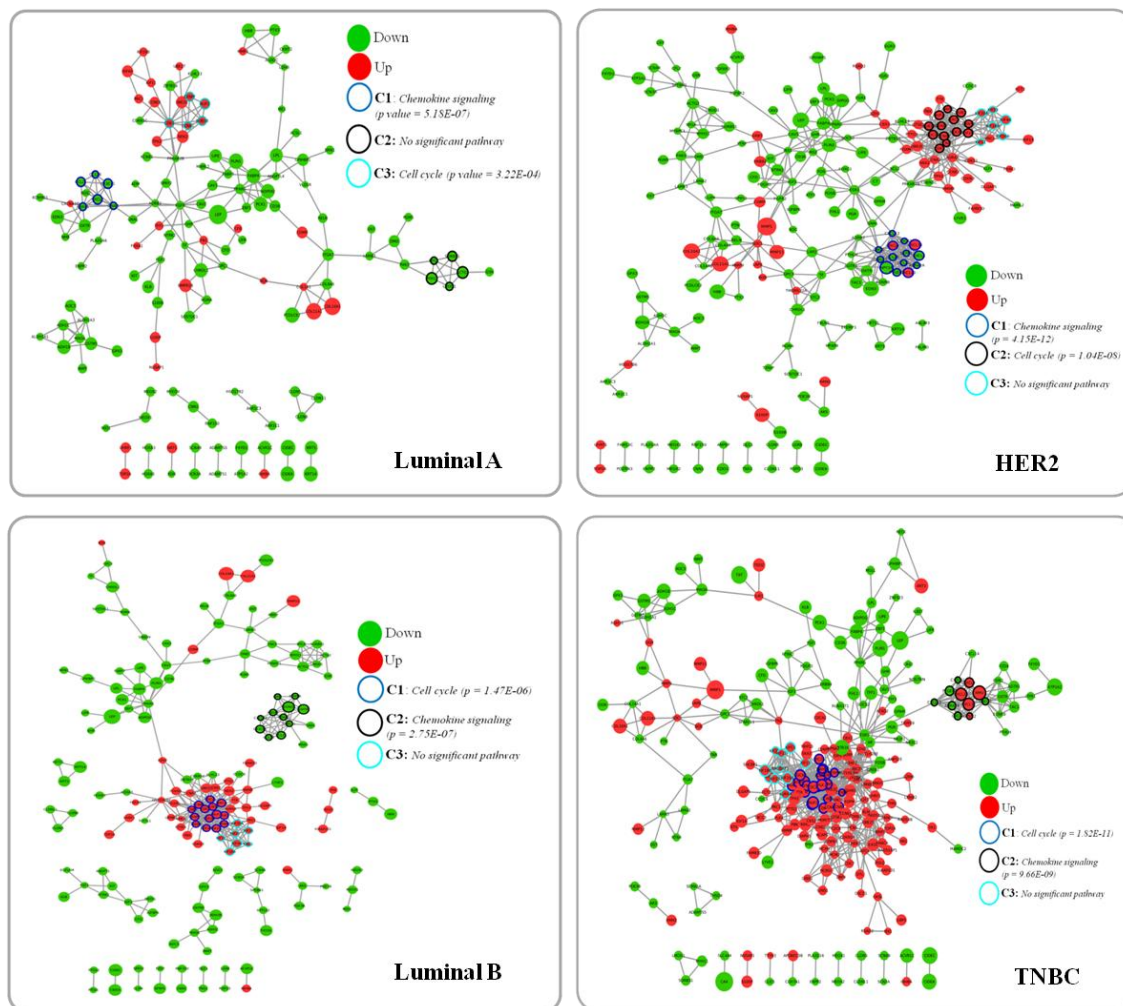
Fig. 3 Protein interaction network of candidate genes in breast cancer subtypes. C1: Cluster1, C2: Cluster 2, C3: Cluster 3. Node size indicates the increasing fold change. Most significantly enriched pathway for each cluster is mentioned.

Among the 382 differentially expressed candidate genes in HER2 subtype, 212 genes were connected with 542 edges (**Fig. 3**). The analysis of the protein interaction network identified 13 clusters with 83 nodes, 33 of which (15%) were represented by the top 3 clusters (**Fig. 3**). Cluster 1 having 14 genes was highly connected with 91 edges and was most significantly enriched for 'chemokine signaling pathway'. Cluster 2 containing 13 up-regulated genes was significantly enriched for 'cell cycle' and 3 other pathways. Cluster 3 having 6 up-regulated genes was enriched for processes related to 'microtubule-based movement', 'cell cycle' and 'antigen processing and presentation'. ESR1, a down-regulated gene in HER2 subtype, was identified as

the key node based on betweenness centrality measure. Among the up-regulated genes, CDK1 and SDC1 were among the top ten hub molecules.

A total of 256 genes were found to be connected with 973 edges in triple-negative breast cancer protein interaction network (**Fig. 3**). The cluster analysis of the network resulted in a total of 18 clusters, highest among all the subtypes. Cluster 1 contained 23 up-regulated highly connected (253 edges) genes and was significantly enriched for 'cell cycle' and 5 other pathways. Custer 2 contained 5 up-regulated and 6 down-regulated genes and was significantly enriched for 'chemokine signaling', 'cytokine-cytokine receptor interaction', and 'Toll-like receptor signaling' pathways. Cluster 3 contained 10 up-regulated genes, all of which were associated with 'microtubule-based movement'. ESR1, a down-regulated gene with a fold change of 4.2 was identified as the top hub gene in TNBC subtype, whereas, among the up-regulated genes, FN1 (fibronectin-1) was the key hub molecule with highest betweenness centrality measure.

Identification of functionally important genes in breast cancer subtypes

The protein interaction network analysis across breast cancer subtypes identified functionally important molecules based on multiple topological features (**Fig. 4**). The results show that the average node degree of the low (log2 fold ≤3) and moderately differentially regulated genes (log2 fold between ≥3.5 & ≤5) was significantly higher compared to that of highly differentially regulated genes (log2 fold ≥5.5) in all breast cancer subtypes, except in luminal A (**Fig. 4A**). Similar trend was observed, when the percentage of nodes with a degree of at least 5 were compared. In case of luminal B, HER2 and TNBC subtypes, 38, 48, and 58% of the moderately differentially regulated genes, respectively were found to be interacting with at least 5 proteins, whereas, <25% of the highly differentially regulated genes in these subtypes had a degree of at least 5 (**Fig. 4B**). When global network topological features, such as betweenness and closeness centrality were explored, overall, highly differentially regulated genes had significantly lower score compared to the low and moderately differentially regulated genes (**Fig. 4C and D**). For example, EGFR in luminal A, PPARγ in luminal B, and ESR1 in HER2 and TNBC subtypes had a fold change of less than 5, but had highest betweenness centrality score. However, COL10A1, which had a very high fold change (> 6) in all the breast cancer subtypes, interacted with only 4 genes in any subtype and its betweenness centrality score was zero. Similarly, MMP11 was up-regulated with a fold change of more than 5.5 in all the four subtypes; however, it interacted with

13

only 1 gene in three subtypes, and had no interaction in luminal A subtype. LEP, a highly down-regulated gene in all the four subtypes (≥ 5.8 fold) interacted with a maximum of 6 proteins and had a low betweenness centrality score. These findings suggest that the genes differentially expressed with a low-moderate fold change play an important role in the information flow across the proteins in breast cancer subtypes and may be potential therapeutic targets.
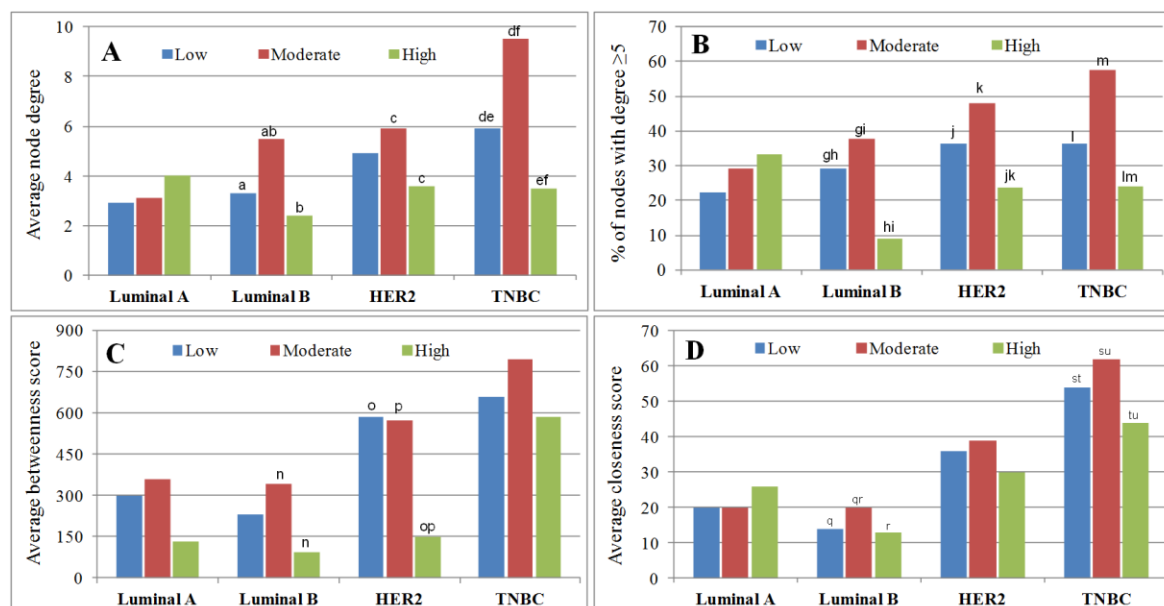

Fig. 4 Topological features of the protein interaction network across breast cancer subtypes for low (log2fold ≤ 3), moderate (log2fold between ≥ 3.5 & ≤ 5), and high (log2fold ≥ 5.5) deregulated genes. A & B: Node degree; C: Betweenness centrality; D: Closeness centrality. The bars sharing same alphabetical letter are significantly different from each other with a p value <0.05. Only significant values are indicated. **a:** 0.0096; **b:** 0.0003; **c:** 0.0193; **d:** 0.0077; **e:** 0.0219; **f:** 4.2E-06 ; **g:** 0.0043; **h:** 0.0002; **i:** 1.5E-06; **j:** 0.0001; **k:** 2.5E-05; **l:** 0.0298; **m:** 0.0029; **n:** 0.0280; **o:** 0.0029; **p:** 0.0095; **q:** 0.0195; **r:** 0.0188; **s:** 0.0307; **t:** 0.0493; **u:** 0.0006.

Interaction between FN1 and miR-1271-5p may be explored as a potential therapeutic option for TNBC

The FN1 gene was found to be up-regulated in TNBC subtype with a log 2 fold of 2.5, and had the highest betweenness among the up-regulated genes in TNBC. Hence, we explored its therapeutic potential using mRNA-miRNA interactions. Thirteen miRNAs were predicted to target FN1 gene based on three different methods (TargetScan, miRanda, and miRDB) (**Fig. 5**). Of these, miR-200c-3p, miR-200b-3p, and miR-1-3p were also found to be reported in

miRTarBase (Chou et al. 2018), which stores experimentally proven interactions. Furthermore, among the 13 miRNAs, 12 have been reported in context of TNBC subtype in literature. However, there was no literature evidence indicating a direct association of hsa-miR-1271-5p with TNBC, although its association with breast cancer has been reported in few studies (Du and Liu 2018; Liu et al. 2019). Additionally, hsa-miR-1271-5p is down-regulated in breast cancer tissues compared to normal breast tissues based on the evidences from literature (Du and Liu 2018; Liu et al. 2019) as well as miRNA expression databases. Thus, miR-1271-5p mimics can be explored as a potential therapeutic option for TNBC subtype.
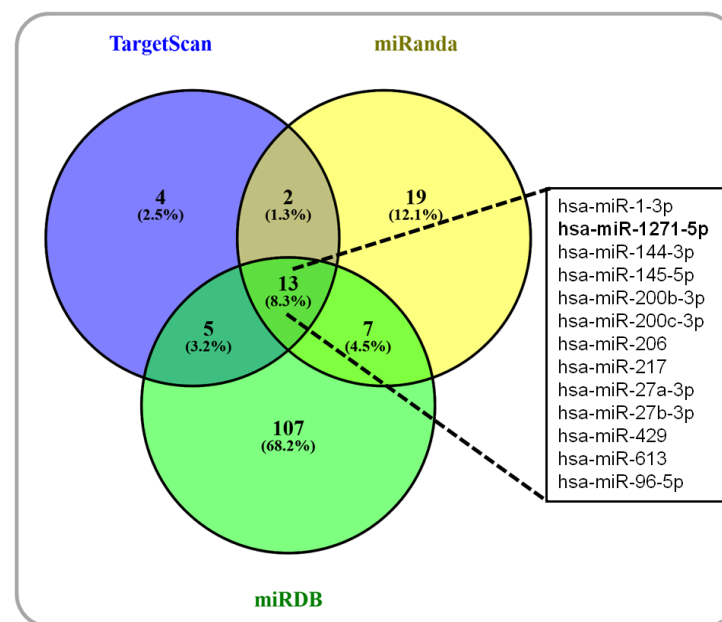

Fig. 5 miRNAs targeting FN1 gene. The box provides a list of miRNAs targeting FN1 gene, common to all the three prediction methods (TargetScan, miRanda and miRDB). Novel miRNA targeting FN1 gene in TNBC subtype is highlighted in bold.

Discussion

Meta-analysis of transcriptomic data is known to provide novel insights into the molecular details of cancer (Davuluri et al., 2021). In the current study, we have performed meta-analysis of publicly available microarray and RNA-sequencing data across four major molecular subtypes of breast cancer (luminal A, luminal B, HER2 and TNBC). Although RNA sequencing data contributed comparatively higher number of genes; microarray studies provided higher sample size, particularly for HER2 and TNBC subtypes. We identified candidate genes strongly

15

associated with each molecular subtype as well as genes commonly differentially expressed in all four subtypes. The functional analysis of the candidate genes in each subtype showed up-regulation of 'cell cycle' and 'p53 signaling' pathways, corroborating with findings from other studies involving breast and other cancers (Peng et al. 2017; Tian et al. 2018). Furthermore, the current study showed down-regulation of multiple signaling pathways (PPAR, AMPK, and adipocytokine signaling) in breast cancer subtypes. AMPK signaling has been reported to prompt cells to produce energy at the expense of growth and motility (Zhao et al. 2017). Hence, decreasing AMPK activity may induce proliferation and metastasis of tumor cells. Multiple studies have reported that activation of AMPK signaling leading to apoptosis of breast cancer cells (Lin et al. 2010; Lin et al. 2017; Song et al. 2017).

Furthermore, enrichment analysis showed chromosome 1 to be significantly enriched by the up-regulated genes in two subtypes, and chromosome 3 by the down-regulated genes in three subtypes. Study by Soloviev et al. (Soloviev et al. 2013) has reported significant overexpression of 26 genes mapped to chromosome 1q in breast cancer compared to normal breast tissues. Another study (Orsetti et al. 2006) identified 30 candidate genes from chromosome 1 showing significant overexpression correlated to copy number increase using microarray. Multiple genes mapping to chromosome 3 display low expression in breast cancer tissues, generally due to the loss of chromosome 3 regions (Al Sarakbi et al. 2009; Silva et al. 2006; Thompson et al. 2014; Zia et al. 2007).

The genes, MMP11, COL11A1, and COL10A1 were up-regulated in four breast cancer subtypes with a log 2 fold value of >5.5. The role of MMP11 has been well established in multiple cancers including breast cancer (Pang et al. 2016; Zhang et al. 2016). Few studies have also shown that MMP11 can be used as a biomarker for metastatic breast cancer (Gonzalez de Vega et al. 2019; Roscilli et al. 2014). Similarly, COL11A1 has also been studied as a marker in breast cancer (Freire et al. 2014) as well as other cancers (Garcia-Pravia et al. 2013; Wu et al. 2014). The role of COL10A1 has been explored in multiple cancers (Chapman et al. 2012) and recently in breast cancer subtypes (Zhang et al. 2020). This gene encodes an alpha chain of type X collagen and it is a component of extracellular matrix. Mutations in COL10A1 are associated with Schmid type metaphyseal chondrodysplasia (Warman et al. 1993) and Japanese type spondylometaphyseal dysplasia (Ikegawa et al. 1998). Chapman et al., (Chapman et al. 2012) reported elevated

16

expression of COL10A1 in diverse tumor types and restricted expression in human normal tissues. The authors stated that COL10A1 can be used as a novel target for diagnosis and treatment of multiple cancer types. In addition, increased expression of stromal COL10A1 correlates with poor pathological response in ER+/HER2+ breast tumors (Brodsky et al. 2016). Nevertheless, the current study shows that COL10A1 is highly up-regulated in all breast cancer subtypes; therefore, it may be used as a diagnostic marker for breast cancer, irrespective of the subtype.

The degree and betweenness are important measures for identifying functionally important genes/proteins in a protein interaction network (Yu et al. 2007). While degree, which indicates the number of connected nodes, is a local network property; betweenness centrality is a global network property and indicates the nodes important for information flow across the complete network. Joy et al. (Joy et al. 2005), found that the proteins with high betweenness score are more likely to be essential in a yeast protein interaction network. Further, a mammalian transcription network clearly displays a high degree and betweenness score for the tumor-suppressor and proto-oncogenes (Potapov et al. 2005). In the current study, we showed that low-moderately differentially expressed genes had high betweenness score and may be important for information flow across the protein interaction network in breast cancer subtypes than the highly-deregulated genes. Hence, they can be better therapeutic targets.

Fibronectin 1(FN1) gene was found to be the key up-regulated 'hub gene' in TNBC subtype based on the highest betweenness centrality score with a fold change of 2.5. It encodes a fibronectin protein, which is found on the cell surface and in extracellular matrix and is known to be involved in cell adhesion, motility, and wound healing (Li et al. 2017b; Xiao et al. 2018; Yoshimura et al. 2018). Being central to the protein interaction network, suppression of FN1 may disrupt the information flow across multiple proteins, which may affect multiple processes and pathways, and hence targeting FN1 may help in tumor inhibition. In addition, its involvement in cell adhesion and migration process makes it a potential candidate drug target. Recent studies have also shown its association with multiple neoplasms. High stromal expression of FN1 has been shown to be associated with aggressive pancreatic carcinoma (Hu et al. 2019), although the study did not find any association with overall survival of patients. A study by Yoshimura et al. (Yoshimura et al. 2018) showed that miR-99a-5p promotes cell invasion in

17

ovarian cancer by up-regulating fibronectin and vitronectin. Another study (Xiao et al. 2018) showed that high FN1 expression in esophageal carcinoma is closely associated with poor prognosis. The high stromal content of the protein facilitates tumor cell metastasis by promoting morphological changes and improving the motility of cancer cells. Other cancers where high FN1 expression has been reported include lung adenocarcinoma (Li et al. 2017b), prostate cancer (Petrini et al. 2017), and hepatocellular carcinoma (Liu et al. 2017). The role of FN1 has also been reported in breast cancer detection, development, and metastasis (Hong et al. 2014; Jeon et al. 2015; Moon et al. 2016), including TNBC and other breast cancer subtypes (Suman et al. 2016). Study by Wang et al., (Wang et al. 2011) showed that miR-1 suppresses migration and invasion of laryngeal carcinoma by negatively regulating FN1 gene. In the current study, we identified FN1 gene to be targeted by miR-1271-5p based on three different prediction methods. Multiple studies have reported the anti-cancer role of miR-1271 in various cancers (Li et al. 2017a; Wang et al. 2018; Xie et al. 2018). In addition, a few studies have reported its tumor suppressor activity in breast cancer via targeting SNAI2 (Liu et al. 2019) and SPIN1 (Du and Liu 2018) genes. However, the relationship between miR-1271-5p and FN1 has not been reported in breast cancer till date, although their interaction has been experimentally validated in glioma (Gong et al. 2017). Our results indicate that the interaction between FN1 and miR-1271-5p can be exploited as a potential therapeutic option for TNBC. However, *in vitro* and *in vivo* studies are required to confirm the role of miR-1271-5p in the suppression of FN1 gene in TNBC subtype.

The current study had a few limitations. While categorizing the genes into different regulation-based groups (low, moderate and high), the fold change threshold was considered based on the minimum and maximum fold change values of the candidate genes in breast cancer subtypes. However, the fold change threshold for deriving such genes may vary depending on the condition of interest. For instance, in the current study, genes with |log2fold <3| were considered as lowly deregulated. However, same criteria may not hold true for other cancer types or conditions, where the fold difference between the test and control conditions is not considerably high. Furthermore, the functional importance of the low-moderately deregulated genes in breast cancer subtypes need to be confirmed experimentally.

Conclusions

Meta-analysis of microarray and RNA sequencing data identified candidate genes associated with four major subtypes of breast cancer. Functional analysis of these genes showed up-regulation of 'cell cycle' and 'p53 signaling' pathways, and down-regulation of 'PPAR', 'Adipocytokine' and 'AMPK' signaling, 'drug metabolism', 'tyrosine metabolism', 'regulation of lipolysis' and 'focal adhesion' pathways in all four subtypes of breast cancer. While multiple chromosomes were found to be significantly enriched by the up-regulated genes, chromosome 3 had a significantly higher share for the down-regulated genes in majority of the subtypes. COL10A1 was found to be significantly up-regulated in all breast cancer subtypes with a high fold (mean log 2 fold of 7.1) and may be used as a potential diagnostic marker for breast cancer, irrespective of the molecular subtype. Among the commonly up-regulated candidate genes, the analysis showed an association of KIF4A with the survival of breast cancer patients. Significant association was observed for seven of the commonly down-regulated candidate genes. Protein interaction network analysis of the candidate genes identified distinct clusters of proteins and hub molecules in each subtype. Moreover, the network analysis showed that low-moderately deregulated genes control the information flow across proteins in breast cancer subtypes, and hence may be better therapeutic targets. MiR-1271-5p may have a role in the negative regulation of FN1 gene, an up-regulated and functionally important hub gene in TNBC. Hence, synthetic miR-1271 mimics may be used as a potential therapeutic option for treating triple-negative breast cancer. However, experimental validations are required to confirm the regulatory role of miR-1271-5p in TNBC.

19

Acknowledgements

Conflicts of interest/Competing interests

AKB and SD were supported by Shodhaka Life Sciences Pvt. Ltd. KKA received financial support only from IBAB, and is a scientific adviser for Shodhaka LS Pvt. Ltd. This company supports basic research and the affiliation with this company in no way alters the pure academic nature of the work being reported.

Author contributions

AKB: Conceptualization, Methodology, Data curation, Validation, Formal analysis, Investigation, Visualization, Writing - Original Draft; **SD:** Data curation, Validation, Visualization, Investigation, Writing - Original Draft; **KT:** Supervision, Writing - Review & Editing; **KKA:** Conceptualization, Methodology, Supervision, Project administration, Writing - Review & Editing.

20

References

Acharya KK et al. (2010) A novel tissue-specific meta-analysis approach for gene expression predictions, initiated with a mammalian gene expression testis database BMC Genomics 11:467 doi:1471-2164-11-467 [pii] 10.1186/1471-2164-11-467

Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs Elife 4 doi:10.7554/eLife.05005

Al Sarakbi W, Sasi W, Jiang WG, Roberts T, Newbold RF, Mokbel K (2009) The mRNA expression of SETD2 in human breast cancer: correlation with clinico-pathological parameters BMC Cancer 9:290 doi:1471-2407-9-290 [pii] 10.1186/1471-2407-9-290

Anders CK, Carey LA (2009) Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer Clin Breast Cancer 9 Suppl 2:S73-81 doi:6358726R5114J615 [pii] 10.3816/CBC.2009.s.008

Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks BMC Bioinformatics 4:2

Bajpai, A., Davuluri, S., Haridas, H. et al. In search of the right literature search engine(s). Nat Prec (2011). https://doi.org/10.1038/npre.2011.2101.3

Bajpai AK, Davuluri S, Tiwary K, Narayanan S, Oguru S, Basavaraju K, Dayalan D, Thirumurugan K, Acharya KK. Systematic comparison of the protein-protein interaction databases from a user's perspective. J Biomed Inform. 2020 Mar;103:103380. doi: 10.1016/j.jbi.2020.103380. Epub 2020 Jan 28.

Barrett T et al. (2013) NCBI GEO: archive for functional genomics data sets--update Nucleic Acids Res 41:D991-995 doi:gks1193 [pii] 10.1093/nar/gks1193

Betel D, Wilson M, Gabow A, Marks DS, Sander C (2008) The microRNA.org resource: targets and expression Nucleic Acids Res 36:D149-153 doi:gkm995 [pii] 10.1093/nar/gkm995

Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L (2016) Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease Nat Rev Clin Oncol 13:674-690 doi:nrclinonc.2016.66 [pii] 10.1038/nrclinonc.2016.66

Brodsky AS et al. (2016) Identification of stromal ColXalpha1 and tumor-infiltrating lymphocytes as putative predictive markers of neoadjuvant therapy in estrogen receptor-positive/HER2-positive breast cancer BMC Cancer 16:274 doi:10.1186/s12885-016-2302-5 10.1186/s12885-016-2302-5 [pii]

Burstein MD et al. (2015) Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer Clin Cancer Res 21:1688-1698 doi:1078-0432.CCR-14-0432 [pii] 10.1158/1078-0432.CCR-14-0432

Chandrashekar DS, Bashel B, Balasubramanya SAH, Creighton CJ, Ponce-Rodriguez I, Chakravarthi B, Varambally S (2017) UALCAN: A Portal for Facilitating Tumor Subgroup Gene Expression and Survival Analyses Neoplasia 19:649-658 doi:S1476-5586(17)30179-3 [pii] 10.1016/j.neo.2017.05.002

Chapman KB, Prendes MJ, Sternberg H, Kidd JL, Funk WD, Wagner J, West MD (2012) COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature Future Oncol 8:1031-1040 doi:10.2217/fon.12.79

Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY (2014) cytoHubba: identifying hub objects and sub-networks from complex interactome BMC Syst Biol 8 Suppl 4:S11 doi:1752-0509-8-S4-S11 [pii] 10.1186/1752-0509-8-S4-S11

Chou CH et al. (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions Nucleic Acids Res 46:D296-D302 doi:4595852 [pii] 10.1093/nar/gkx1067

Davuluri S, Bajpai AK, Thirumurugan K, Acharya KK. The molecular basis of gender disparities in smoking lung cancer patients. Life Sci. 2021 Feb 15;267:118927. doi: 10.1016/j.lfs.2020.118927. Epub 2020 Dec 28.

Dong P, Yu B, Pan L, Tian X, Liu F (2018) Identification of Key Genes and Pathways in Triple-Negative Breast Cancer by Integrated Bioinformatics Analysis Biomed Res Int 2018:2760918 doi:10.1155/2018/2760918

Du HY, Liu B (2018) MiR-1271 as a tumor suppressor in breast cancer proliferation and progression via targeting SPIN1 Eur Rev Med Pharmacol Sci 22:2697-2706 doi:10.26355/eurrev_201805_14966

Freire J, Dominguez-Hormaetxe S, Pereda S, De Juan A, Vega A, Simon L, Gomez-Roman J (2014) Collagen, type XI, alpha 1: an accurate marker for differential diagnosis of breast carcinoma invasiveness in core needle biopsies Pathol Res Pract 210:879-884 doi:S0344-0338(14)00225-8 [pii] 10.1016/j.prp.2014.07.012

Garcia-Pravia C et al. (2013) Overexpression of COL11A1 by cancer-associated fibroblasts: clinical relevance of a stromal marker in pancreatic cancer PLoS One 8:e78327 doi:10.1371/journal.pone.0078327 PONE-D-13-17106 [pii]

Gong J, Wang ZX, Liu ZY (2017) miRNA1271 inhibits cell proliferation in neuroglioma by targeting fibronectin 1 Mol Med Rep 16:143-150 doi:10.3892/mmr.2017.6610

Gonzalez de Vega R et al. (2019) MMP-11 as a biomarker for metastatic breast cancer by immunohistochemical-assisted imaging mass spectrometry Anal Bioanal Chem 411:639-646 doi:10.1007/s00216-018-1365-3 10.1007/s00216-018-1365-3 [pii]

Guo J, Gong G, Zhang B (2017) Screening and identification of potential biomarkers in triple-negative breast cancer by integrated analysis Oncol Rep 38:2219-2228 doi:10.3892/or.2017.5911

Hong H et al. (2014) Pigment epithelium-derived factor (PEDF) inhibits breast cancer metastasis by down-regulating fibronectin Breast Cancer Res Treat 148:61-72 doi:10.1007/s10549-014-3154-9

Hsu HM et al. (2019) Six novel immunoglobulin genes as biomarkers for better prognosis in triple-negative breast cancer by gene co-expression network analysis Sci Rep 9:4484 doi:10.1038/s41598-019-40826-w 10.1038/s41598-019-40826-w [pii]

Hu D, Ansari D, Zhou Q, Sasor A, Said Hilmersson K, Andersson R (2019) Stromal fibronectin expression in patients with resected pancreatic ductal adenocarcinoma World J Surg Oncol 17:29 doi:10.1186/s12957-019-1574-z 10.1186/s12957-019-1574-z [pii]

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources Nat Protoc 4:44-57 doi:nprot.2008.211 [pii] 10.1038/nprot.2008.211

Hubbell E, Liu WM, Mei R (2002) Robust estimators for expression analysis Bioinformatics 18:1585-1592

Ikegawa S, Nishimura G, Nagai T, Hasegawa T, Ohashi H, Nakamura Y (1998) Mutation of the type X collagen gene (COL10A1) causes spondylometaphyseal dysplasia Am J Hum Genet 63:1659-1662 doi:S0002-9297(07)61610-3 [pii] 10.1086/302158

Jeon M, Lee J, Nam SJ, Shin I, Lee JE, Kim S (2015) Induction of fibronectin by HER2 overexpression triggers adhesion and invasion of breast cancer cells Exp Cell Res 333:116-126 doi:S0014-4827(15)00067-1 [pii] 10.1016/j.yexcr.2015.02.019

Joy MP, Brock A, Ingber DE, Huang S (2005) High-betweenness proteins in the yeast protein interaction network J Biomed Biot chnol 2005:96-103 doi:S1110724304406019 [pii] 10.1155/JBB.2005.96

Kumar C, Rani N, Velan Lakshmi PT, Arunachalam A (2017) A comprehensive look of poly(ADP-ribose) polymerase inhibition strategies and future directions for cancer therapy Future Med Chem 9:37-60 doi:10.4155/fmc-2016-0113

Li L, Qu YW, Li YP (2017a) Over-expression of miR-1271 inhibits endometrial cancer cells proliferation and induces cell apoptosis by targeting CDK1 Eur Rev Med Pharmacol Sci 21:2816-2822

Li Y et al. (2017b) alpha1-antitrypsin promotes lung adenocarcinoma metastasis through up-regulating fibronectin expression Int J Oncol 50:1955-1964 doi:10.3892/ijo.2017.3962

Lin JN et al. (2010) Resveratrol modulates tumor cell proliferation and protein translation via SIRT1-dependent AMPK activation J Agric Food Chem 58:1584-1592 doi:10.1021/jf9035782

Lin Z et al. (2017) Mevastatin blockade of autolysosome maturation stimulates LBH589-induced cell death in triple-negative breast cancer cells Oncotarget 8:17833-17848 doi:14868 [pii] 10.18632/oncotarget.14868

Liu BW, Yu ZH, Chen AX, Chi JR, Ge J, Yu Y, Cao XC (2019) Estrogen receptor-alpha-miR-1271-SNAI2 feedback loop regulates transforming growth factor-beta-induced breast cancer progression J Exp Clin Cancer Res 38:109 doi:10.1186/s13046-019-1112-4 10.1186/s13046-019-1112-4 [pii]

Liu YR et al. (2016) Comprehensive transcriptome analysis identifies novel molecular subtypes and subtype-specific RNAs of triple-negative breast cancer Breast Cancer Res 18:33 doi:10.1186/s13058-016-0690-8 10.1186/s13058-016-0690-8 [pii]

Liu ZC et al. (2017) Epidermal growth factor and tumor necrosis factor alpha cooperatively promote the motility of hepatocellular carcinoma cell lines via synergistic induction of fibronectin by NF-kappaB/p65 Biochim Biophys Acta Gen Subj 1861:2568-2582 doi:S0304-4165(17)30266-0 [pii] 10.1016/j.bbagen.2017.08.010

Mathe A, Wong-Brown M, Morten B, Forbes JF, Braye SG, Avery-Kiejda KA, Scott RJ (2015) Novel genes associated with lymph node metastasis in triple negative breast cancer Sci Rep 5:15832 doi:srep15832 [pii] 10.1038/srep15832

Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap Nucleic Acids Res 43:W566-570 doi:gkv468 [pii] 10.1093/nar/gkv468

Moon PG et al. (2016) Fibronectin on circulating extracellular vesicles as a liquid biopsy to detect breast cancer Oncotarget 7:40189-40199 doi:9561 [pii] 10.18632/oncotarget.9561

Naorem LD, Muthaiyan M, Venkatesan A (2019) Integrated network analysis and machine learning approach for the identification of key genes of triple-negative breast cancer J Cell Biochem 120:6154-6167 doi:10.1002/jcb.27903

Orsetti B et al. (2006) Genetic profiling of chromosome 1 in breast cancer: mapping of regions of gains and losses and identification of candidate genes on 1q Br J Cancer 95:1439-1447 doi:6603433 [pii] 10.1038/sj.bjc.6603433

Pang L, Wang DW, Zhang N, Xu DH, Meng XW (2016) Elevated serum levels of MMP-11 correlate with poor prognosis in colon cancer patients Cancer Biomark 16:599-607 doi:CBM601 [pii] 10.3233/CBM-160601

Peng J et al. (2017) Expression profile analysis of long noncoding RNA in ER-positive subtype breast cancer using microarray technique and bioinformatics Cancer Manag Res 9:891-901 doi:10.2147/CMAR.S151120 cmar-9-891 [pii]

Perou CM et al. (2000) Molecular portraits of human breast tumours Nature 406:747-752 doi:10.1038/35021093

Petrini I et al. (2017) ED-B fibronectin expression is a marker of epithelial-mesenchymal transition in translational oncology Oncotarget 8:4914-4921 doi:13615 [pii] 10.18632/oncotarget.13615

Potapov AP, Voss N, Sasse N, Wingender E (2005) Topology of mammalian transcription networks Genome Inform 16:270-278 doi:162270 [pii]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies Nucleic Acids Res 43:e47 doi:gkv007 [pii] 10.1093/nar/gkv007

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data Bioinformatics 26:139-140 doi:btp616 [pii] 10.1093/bioinformatics/btp616

Roscilli G et al. (2014) Circulating MMP11 and specific antibody immune response in breast and prostate cancer patients J Transl Med 12:54 doi:1479-5876-12-54 [pii] 10.1186/1479-5876-12-54

Shannon P et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks Genome Res 13:2498-2504 doi:10.1101/gr.1239303 13/11/2498 [pii]

Silva J et al. (2006) Analysis of the candidate tumor suppressor Ris-1 in primary human breast carcinomas Mutat Res 594:78-85 doi:S0027-5107(05)00406-9 [pii] 10.1016/j.mrfmmm.2005.07.017

Soloviev M, Esteves MP, Amiri F, Crompton MR, Rider CC (2013) Elevated transcription of the gene QSOX1 encoding quiescin Q6 sulfhydryl oxidase 1 in breast cancer PLoS One 8:e57327 doi:10.1371/journal.pone.005732 PONE-D-12-28133 [pii]

Song W, Yan CY, Zhou QQ, Zhen LL (2017) Galangin potentiates human breast cancer to apoptosis induced by TRAIL through activating AMPK Biomed Pharmacother 89:845-856 doi:S0753-3322(16)32857-8 [pii] 10.1016/j.biopha.2017.01.062

Sorlie T et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications Proc Natl Acad Sci U S A 98:10869-10874 doi:10.1073/pnas.191367098 98/19/10869 [pii]

Suman S, Basak T, Gupta P, Mishra S, Kumar V, Sengupta S, Shukla Y (2016) Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer J Proteomics 148:183-193 doi:S1874-3919(16)30346-3 [pii] 10.1016/j.jprot.2016.07.033

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms PLoS One 6:e21800 doi:10.1371/journal.pone.0021800 PONE-D-11-04111 [pii]

Szklarczyk D et al. (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets Nucleic Acids Res 47:D607-D613 doi:5198476 [pii] 10.1093/nar/gky1131

Thompson PA et al. (2014) Loss of LRIG1 locus increases risk of early and late relapse of stage I/II breast cancer Cancer Res 74:2928-2935 doi:74/11/2928 [pii] 10.1158/0008-5472.CAN-13-2112

Tian T et al. (2018) Identification of long non-coding RNA signatures in triple-negative breast cancer Cancer Cell Int 18:103 doi:10.1186/s12935-018-0598-8 598 [pii]

Torre LA, Siegel RL, Ward EM, Jemal A (2016) Global Cancer Incidence and Mortality Rates and Trends--An Update Cancer Epidemiol Biomarkers Prev 25:16-27 doi:1055-9965.EPI-15-0578 [pii] 10.1158/1055-9965.EPI-15-0578

Wang F, Song G, Liu M, Li X, Tang H (2011) miRNA-1 targets fibronectin1 and suppresses the migration and invasion of the HEp2 laryngeal squamous carcinoma cell line FEBS Lett 585:3263-3269 doi:S0014-5793(11)00669-7 [pii] 10.1016/j.febslet.2011.08.052

Wang M, Gao W, Lu D, Teng L (2018) MiR-1271 Inhibits Cell Growth in Prostate Cancer by Targeting ERG Pathol Oncol Res 24:385-391 doi:10.1007/s12253-017-0254-y 10.1007/s12253-017-0254-y [pii]

Warman ML et al. (1993) A type X collagen mutation causes Schmid metaphyseal chondrodysplasia Nat Genet 5:79-82 doi:10.1038/ng0993-79

Wong N, Wang X (2015) miRDB: an online resource for microRNA target prediction and functional annotations Nucleic Acids Res 43:D146-152 doi:gku1104 [pii] 10.1093/nar/gku1104

Wu YH, Chang TH, Huang YF, Huang HD, Chou CY (2014) COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer Oncogene 33:3432-3440 doi:onc2013307 [pii] 10.1038/onc.2013.307

Xiao J et al. (2018) Expression of fibronectin in esophageal squamous cell carcinoma and its role in migration BMC Cancer 18:976 doi:10.1186/s12885-018-4850-3 10.1186/s12885-018-4850-3 [pii]

Xie F et al. (2018) MiR-1271 negatively regulates AKT/MTOR signaling and promotes apoptosis via targeting PDK1 in pancreatic cancer Eur Rev Med Pharmacol Sci 22:678-686 doi:10.26355/eurrev_201802_14293

Yang Z, Wu L, Wang A, Tang W, Zhao Y, Zhao H, Teschendorff AE (2017) dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers Nucleic Acids Res 45:D812-D818 doi:gkw1079 [pii] 10.1093/nar/gkw1079

Yoshimura A et al. (2018) Exosomal miR-99a-5p is elevated in sera of ovarian cancer patients and promotes cancer cell invasion by increasing fibronectin and vitronectin expression in neighboring peritoneal mesothelial cells BMC Cancer 18:1065 doi:10.1186/s12885-018-4974-5 10.1186/s12885-018-4974-5 [pii]

Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics PLoS Comput Biol 3:e59 doi:06-PLCB-RA-0302R2 [pii] 10.1371/journal.pcbi.0030059

Yue Z, Li HT, Yang Y, Hussain S, Zheng CH, Xia J, Chen Y (2016) Identification of breast cancer candidate genes using gene co-expression and protein-protein interaction information Oncotarget 7:36092-36100 doi:9132 [pii] 10.18632/oncotarget.9132

Zhang M, Chen H, Wang M, Bai F, Wu K (2020) Bioinformatics analysis of prognostic significance of COL10A1 in breast cancer Biosci Rep 40 doi:222087 [pii] 10.1042/BSR20193286

Zhang X, Huang S, Guo J, Zhou L, You L, Zhang T, Zhao Y (2016) Insights into the distinct roles of MMP-11 in tumor biology and future therapeutics (Review) Int J Oncol 48:1783-1793 doi:10.3892/ijo.2016.3400

Zhao H, Orhan YC, Zha X, Esencan E, Chatterton RT, Bulun SE (2017) AMP-activated protein kinase and energy balance in breast cancer Am J Transl Res 9:197-213

Zia MK, Rmali KA, Watkins G, Mansel RE, Jiang WG (2007) The expression of the von Hippel-Lindau gene product and its impact on invasiveness of human breast cancer cells Int J Mol Med 20:605-611