Identification of Potential Genes in Her2+ Breast Cancer by Co-expression Network Analysis

CURRENT STATUS: UNDER REVIEW

Xinyang Li
Henan University of Science and Technology Affiliated First Hospital

ORCiD: https://orcid.org/0000-0002-8084-0450

Yukun Wang
Henan University of Science and Technology Affiliated First Hospital

Ziming Wang
Henan University of Science and Technology Affiliated First Hospital

Ge Yao
Henan University of Science and Technology Affiliated First Hospital

Jiahao Fan
Henan University of Science and Technology Affiliated First Hospital

Gaofeng Liang
Henan University of Science and Technology Affiliated First Hospital

Xinshuai Wang
Henan University of Science and Technology Affiliated First Hospital

xshuaiw@126.com Corresponding Author

10.21203/rs.3.rs-26134/v1

SUBJECT AREAS
- Epigenetics & Genomics
- Cancer Biology

KEYWORDS
- Her2+ breast cancer, weighted gene co-expression network analysis (WGCNA), bioinformatics analysis, molecular mechanism
Abstract

Background
Breast cancer (BC) is one of the most common malignancies in women all over the world. For patients with human epidermal growth factor receptor 2 positive (Her2+) BC, although the widespread use of anti-Her2 therapy has made survival increased, improved survival has led to an increased demand for better management and diminish toxic side effects of anticancer treatments. This study aimed to identify the potential biomarkers associated Her2 + BC in order to improve the overall management of BC treatment.

Results
Our research downloaded GSE54140 gene expression datasets from the Gene Expression Omnibus data sets, and used weighted gene co-expression network analysis (WGCNA) to develop a scale-free gene co-expression network to explore the associations between gene sets and clinical features. A total of 60 modules were analyzed, and found that the skyblue3 module was significantly related to Her2 + BC. The function of 93 genes in the skyblue3 module was annotated by DAVID bioinformatics tool, and it was demonstrated that the function of the module was mainly related to nuclear-transcribed mRNA catabolic process, cytosol, and oxidoreductase activity. Based on the WGCNA and Cytoscape software analysis, 9 hub genes (PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, STARD3, and NEUROD2) were identified. The Human Protein Atlas database detected that the protein level of PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, and STARD3 in tumor tissues were different from normal tissues. And survival analysis shows that PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival (P < 0.05).

Conclusions
In total, 9 candidate biomarkers were identified by comprehensive analysis, among which, the co-expansion of PGAP3, CRKRS, STARD3 and NEUROD2 related to ERBB2 may be associated with the occurrence of Her2 + BC. In addition, PPP1R1B, CRKRS and TCAP are related to drug resistance and adverse reactions in the treatment of Her2 + BC.

Background
Breast cancer (BC) is the most common cancer of women, which severely intimidates women's health [1]. Nearly 1.3 million women worldwide are pronounced with BC every year, and more than 400,000
Approximately 15% of breast cancer is the human epidermal growth factor receptor 2 positive (Her2 +) breast cancer, despite of the incessantly development of DFS and OS of Her2 + BC patients from available medicine [4], high rates of recurrence and metastasis are still the major cause of death [5]. In-depth study of Her2 + BC is warranted for treatment management. In recently years, numerous microarray profiling works have been executed in BC, and hundreds of differentially expressed genes have been collected. Nevertheless, the difference analysis neglects the communication between genes so that there is no respectable biomarker for clinical application in BC. At present, the scale-free gene co-expression network is developed by weighted gene co-expression network analysis (WGCNA), to explore the associations between gene sets and clinical manifestation, to identify the modules and the hub genes of extremely relevant genes in each module [6]. The relative commonality of nodes whose extent enormously eclipses the general level is the most compelling feature of scale-free networks. Play particular purposes in the network, the nodes with the highest degree are termed hubs. An authoritative practice for constructing free-scale gene co-expression networks is co-expression analysis. Therefore, WGCNA is impeccable for pinpointing key genes and modules that affect phenotypic traits. In this work, we acquired the gene expression data from the gene expression omnibus (GEO) datasets, and pick out remarkably germane gene modules linked to Her2 + BC according to the WGCNA algorithm to seek hub genes that may become targets for diacrisis and therapeutics.

Results

Data processing

In this paper, the GSE54140 (15,612 genes) gene expression matrix was procured applying the R language function, and then after screening, standardization and probe transformation, the genes (11,709 genes) with the top 25% of the variance were completely preferred for the construction of co-expression network.

Co-expression network construction and key modules identification

The co-expression network needs to make the appropriate weighting coefficient β to ensure that the
network comply a scale-free distribution, in this section, the power of β value of 3 was decided to develop a co-expression network, as shown in Figure 1. We then calculated the TOM of each gene pair, 60 modules were displayed by hierarchical clustering according to degree of TOM's difference as shown in Figure 2. Each module had a group of coordinately expressed genes with high TOM, meaning that the intrinsic genes may participate in the same biological process. In order to distinguish the various modules, each module was assigned a unique color. 400 genes were randomly selected to make a TOM heat map. As shown in Figure 3, the darker the color, the higher the TOM, which suggests the high degree of independence between modules and the relative independence of gene expression in each module. The association analysis between tumor characteristics and co-expression modules was carried out to identify the correlation between MEs and tumor characteristics. As shown in Figure 4, Module-trait relationships heat map, the eigengene of the skyblue3 module (93 genes) were undoubtedly correlated with Her2+ BC (cor=0.74, P=3e-12). By calculating the correlation coefficient of GS and MM in skyblue3 module, and performing cluster analysis on the traits and modules, the significant correlation between skyblue3 module and Her2+ BC was determined, and the credibility of the results was confirmed, as shown in Figure 5 and 6. Consequently, further analysis was performed in module skyblue3 gene.

**Functional enrichment analysis of genes in modules**

The names of all genes in the skyblue3 module were uploaded to DAVID bioinformatics tool. For the BP, the genes were mainly denriched in nuclear-transcribed mRNA catabolic process, response to organophosphorus, positive regulation of hepatocyte proliferation, regulation of microtubule-based process, et al, as shown in Table 1. The genes in the CC group were mainly concentrated in cytosol, cytoplasm, ribosome, cytosolic large ribosomal subunit, et al, as shown in Table 2. For the MF, the genes were mainly enriched in oxidoreductase activity, protein binding, actin filament binding, aldehyde dehydrogenase [NAD(P)+] activity, et al, as shown in Table 3.

**Hub genes identification and analysis**

Our research caught that the skyblue3 module had powerful correlation with Her2+ BC, which suggests that hub genes may exists in skyblue3 module. Then we used Cytoscape software to
visualize the network of the skyblue3 module which was connected into the module through WGCNA calculation. Using MCOD function and correlation degree analysis, the MCODE score > 59.5 was presented, with the hub genes that degree > 74 as the hub gene, represented by red dots, as shown in Figure 7. The hub genes in the skyblue3 module included PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, STARD3, and NEUROD2. Moreover based on The Human Protein Atlas database, the protein levels of PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, and STARD3 in tumor tissues were different from normal tissues, as shown in Figure 8. Then, all the hub genes underwent survival analysis by cBioPortal datasets. As shown in Figure 9, PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival.

Discussion

BC is a life-threatening disease for females, especially Her2 + BC with rapid tumor progression, easy recurrence and metastasis and poor prognosis. Despite of the great progress in Her2 + BC treatment options in the past decades, much more remains to be done, more cancer-driving genes need to be identified. Therefore it is critical to find more of the potential genes involved in the development and progression of Her2 + BC.

In this work, we used WGCNA to dynamically study genes co-expression in Her2 + BC, Luminal-A BC, and Luminal-B BC, and to explore the related modules and hub genes. We concluded that the skyblue3 module had the highest correlation with Her2 + BC. Skyblue3 module related to Her2 + BC has well defined functions including nuclear-transcribed mRNA catabolic process, cytosol, and oxidoreductase activity. In the genes of the skyblue3 module, PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, STARD3, and NEUROD2 were regarded as the hub genes.

ERBB2, generally named Her2, is amplified or overexpressed in 20–30% of invasive breast carcinomas [7]. Moreover, amplification and overexpression of Her2 gene has also been detected in ovarian and gastric cancer [8–10]. Her2 integrates firmly to other epidermal growth factor receptor family members to assemble a heterodimer, which reinforce kinase-mediated activation of downstream signaling pathways, such as PI3K/AKT activation and RET signaling, so as to facilitate the increment and metastasis of cancer cells [11]. As a cell surface related protein, the most critical treatment for
Her2-positive cancer is anti-Her2 strategy, but the subsequent Her2 resistance is a new clinical puzzler [12]. Now, an increasing number of teams have found that many oncogenes co-amplify with the Her2 gene.

PGAP3, CRKRS, STARD3 and NEUROD2 were co-amplified with the Her2 gene. PGAP3, glycosylated phosphatidylinositol (GPI)-specific phospholipase members, is principally involved in the protein distribution and pack [13]. This group confirmed that concurrent amplification of copy number variation at PGAP3 and Her2 loci were detected in BC tissues [14]. Since Her2 gene is significantly associated with the occurrence and invasion of BC, the amplification of PGAP3 gene may affect the effect of Her2 gene on BC. CRKRS, cyclin-dependent kinase 12, also known as CDK12, which phosphorylates RNA polymerase II and modulates transcription factors [15]. Recent studies confirmed that genomic stability needs to be maintained by Cdk12 regulating DNA repair genes by suppressing intron polyadenylation [16–18]. This group study found that CRKRS could induce drug resistance in breast cancer endocrine therapy, which was related to its silent activation of mitogen-activated protein kinase signaling pathway, leading to the loss of estrogen receptor dependence [19]. In breast cancers, CDK12, as a carcinogene, is constantly co-amplified with the ERBB2 oncogene, which suggests that the tumor is prone to invasion and metastasis and poor prognosis [20, 21]. This group study found that CDK12 can not only regulate RNA processing and DNA repair, but also the overexpression of CDK12 can enhancement the oncogenicity of BC cells [22]. STARD3-related protein located on the cell membrane of late endosomes may be involved in fostering the carriage or assignment of cholesterol and sphingolipid to the intracellular membrane compartments and the catabolism of steroid hormones [23, 24]. Recent publication confirmed that the expression level of STARD3 protein is forcefully associated with Her2 amplification, and the high expression of STARD3 may boost BC invasion by increasing membrane cholesterol and intensifying oncogene signal [25]. NEUROD2-related protein is part of the family of neurogenic basic helix-loop-helix proteins, which affect the adjust and control of glutamate and GABA genes [26]. Miangela's research discovered that the copy number of Her2 and NEUROD2 increased in male BC tissue, and merely NEUROD2 amplification appear to have an independent prognostic effect [27].
PPP1R1B and TCAP are related to anti-Her2 resistance and treatment. PPP1R1B encodes multiple transcripts and two experimentally-documented proteins Darpp-32 and t-Darpp [28]. Recent publications suggested that the protein encoded by PPP1R1B was overexpressed in BC, esophagus cancer, lung cancer, et al [29, 30]. Professional team proved that the resistance of trastuzumab, a key drug in anti-Her2 therapy, was related to the regulation of IGF-1R and AKT signal pathways by t-Darpp [31, 32]. The protein encoded by TCAP gene, widely distributed in cardiomyocytes, is mainly involved in cardiac conduction and striated muscle contraction [33]. Although doxorubicin and anti-Her2 targeting therapy, trastuzumab, are conventional drugs for Her2 + BC, their cardiotoxicity is a thorny clinical issue. The results of Daniel’s study suggest that the variation of TCAP gene may have modifying effects in cardiomyopathy [34], so whether targeting this gene can shorten the cardiac toxic and side effects of chemotherapy patients.

CISD3, a member of the CDGSH domain-containing family, codes mitochondrial inner NEET protein (MiNT) [35]. Details about biological characteristics of MiNT are relatively indigent, but its high expression in many cancers indicates that it plays a unique role in cancer [36].

The protein levels PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, and STARD3 in tumor tissues were different from normal tissues. And survival analysis shows that PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival. These results indicate that these genes may be tumorigenic genes in BC.

Conclusion
In summary, our study used a systems biology based WGCNA approach to determine the module highly related to Her2 + BC, whose function was mainly concentrated in nuclear-transcribed mRNA catabolic process, cytosol, and oxidoreductase activity. Although most of the hub genes accentuated in this study have been reported in previous reports, there has been no comprehensive and all-round analysis. Undoubtedly, ERBB2 plays an important role in the occurrence and development of BC. In addition, we found that PGAP3, CRKRS, STARD3 and NEUROD2 were associated with co-amplification of ERBB2, PPP1R1B may mediate anti-ERBB2 drug resistance by activating IGF-1R and AKT signal pathway, and TCAP may be related to cardiomyopathy caused by doxorubicin or Trastuzumab. As
endocrine therapy is the most important treatment for hormone receptor positive breast cancer. It has been found that CRKRS mediates endocrine drug resistance by affecting mitogen-activated protein kinase signal pathway. Hub genes such as PGAP3 and STARD3 were highly correlated with Her2 + BC development. However, PPP1R1B, CRKRS, and TCAP may be brought new insights in BC study in treatment target. Further investigation about these genes is warranted.

Materials And Methods

Microarray data and data pre-processing

GSE54140 data was downloaded from the GEO datasets (www.ncbi.nlm.nih.gov/gds). Microarray Data from GSE54140 involved 21 Her2-positive BC, 20 Luminal-A BC, 22 Luminal-B BC, 32 BRCA1-mutated BC, 21 BRCA1-non-mutated BC and 15 BRCA1-unscreened BC. In this search, 63 samples including Her2+ BC, Luminal-A BC and Luminal-B BC were preferred for investigation. Rstudio (3.6.2 version; www.rstudio.com) backed by R language platform (3.6.2 version; www.r-project.org) and relevant software packages were developed to arithmetic data. Before WGCNA, this study pre-processed the microarray Data: the probe name was transformed to the gene name based on the platform information, and the microarray data were standardized through the R function. After standardization and probe switch, the dataset accommodating 15,612 genes was further processed, and the top 25% of genes with the highest degree of variation (11,709 genes) were screened by analysis of variance for co-expression network analysis.

Co-expression network construction

After pre-processing, the selected expression data profiles were developed to a gene co-expression network applying WGCNA package in R (www.cran.r-project.org/web/packages/WGCNA/index.html). The gradient method was employed to weigh the connection strength between each pair of nodes and assess the independence and average degree connectivity of the assorted modules with different power values (range 1-20). In this work, the value whose independence degree was greater than 0.8 for the first time was seems as the appropriate power value ($\beta$) to ensure a scale-free network. The hierarchical clustering dendrograms were developed utilizing correlation coefficients between genes, and genes with analogous expression profiles were divided into the same gene module. In the cluster
dendrograms, different branches mean different gene modules, and different colors refer to different modules. Module indicates a gene set with high topological overlap. Subsequently, the soft threshold power was applied to translate the adjacency matrix into topological overlap measure (TOM), and 400 genes were randomly distributed to make a TOM heat map to validate the high independence between modules and the relative independence of gene expression in each module.

**Module and clinical trait association analysis**

Module Eigengenes (MEs) were described as the first principal component of each gene module, and the expression of MEs was considered as a representative of all genes in a given module. The Pearson correlation coefficient and P value of MEs and clinical trait was computed by WGCNA arithmetic to appraise the potential association between gene modules and clinical traits. Then, farther dissect the module with the highest correlation coefficient. Gene significance (GS) and module significance (MS) were applied to study the module expression patterns related to sample types. The correlation coefficients of different types of samples were defined as GS, and the average of the absolute GS values of all genes in the module was defined as MS. Module membership (MM) was used to appraise the relevancy between genes and modules, and the hub genes in target modules were further screened.

**Enrichment analysis of module**

The function enrichment analysis of modules was developed by Gene Ontology (GO) pathway enrichment analysis employing the DAVID bioinformatics tool (version 6.8; www.david.ncifcrf.gov/) to explore the biological function. Go term enrichment analysis refer to biological process (BP), cellular component (CC), and molecular function (MF). P < 0.05 as the threshold was considered statistically significant.

**Hub genes identification and analysis**

The corresponding heat map was procured by assaying the correlation between each module and the clinical traits of the sample, and the module with the highest correlation was imported into Cytoscape software (version 3.7.2; www.cytoscape.org/) for analysis and visualization. MCOD, a plug in Cytoscape, was applied to screen the module. Module with highest MCODE score (53.9) were
presented, and then the hub genes that degree > 74 were identified by analysis. The Human Protein Atlas database (www.proteinatlas.org) was employed to identify the protein expression of hub genes by comparing normal tissue and cancer tissue. Subsequently, to validate the hub genes, we took the Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016) from cBioPortal database (www.cbioportal.org/), consisting of 2,509 breast cancers. Upload hub gene name to survival analysis of cBioPortal to determine the importance of genes in biological processes.

**Statistical method**

This work ran R language software and WGCNA for statistical calculation. WGCNA was used to structure free-scale gene co-expression networks to determine the communication between genes, thereby enabling the identification of modules (clusters) of highly correlated genes, and the hub gene in each module. P < 0.05 was considered statistically significant. Hypergeometric test was used for enrichment analysis, and Kaplan-Meier statistical method was used for survival analysis.

**Abbreviations**

BC: breast cancer; Her2+: human epidermal growth factor receptor 2 positive; GEO: gene expression omnibus; WGCNA: weighted gene co-expression network analysis; TOM: topological overlap measure; MEs: module eigengenes; GS: gene significance; MS: module significance; MM: module membership; GO: gene ontology; KEGG: kyoto encyclopedia of genes and genomes; BP: biological process; CC: cellular component; MF: molecular function; GPI: glycosylated phosphatidylinositol;

**Declarations**

**Acknowledgments**

Thanks for all the authors who provided the technical help for the analysis.

**Authors’ contributions**

XL and XW wrote and revised the manuscript; XL was the main contributor to this manuscript; XW critically revised and corrected the manuscript; YW, ZW, JF, GL and GY Analysis and interpretation of relevant data; All authors read and approved the final manuscript.

**Funding**

Not applicable.
Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors have consented for the publication.

Competing interests

The authors declare no potential conflicts of interest.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394-424.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics. 2018. 2018;68(1):7-30.
3. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin. 2017;67(6):439-48.
4. Ahmad A. Breast Cancer Statistics: Recent Trends. Adv Exp Med Biol. 2019;1152:1–7.
5. Fan W, Chang J, Fu P. Endocrine therapy resistance in breast cancer: current status, possible mechanisms and overcoming strategies. Future Med Chem. 2015;7(12):1511-9.
6. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
7. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134-50.
8. Aznab M, Maleksabet D, Khazaeei S, Khazaeei M, Rezaei M. The Role of Human Epidermal Growth Factor Receptor (HER2/neu) in the Prognosis of Patients with Gastric Cancer. Asian Pac J Cancer Prev. 2019;20(7):1989-94.
9. Maennling AE, Tur MK, Niebert M, Klockenbring T, Zeppernick F, Gattenlohner S, et al. Molecular Targeting Therapy against EGFR Family in Breast Cancer: Progress and Future Potentials. Cancers (Basel). 2019;11(12).

10. Kim H, Seo S, Kim K. Prognostic significance of Human epidermal growth factor receptor-2 expression in patients with resectable gastric adenocarcinoma. 2019;17(1):122.

11. Jung DH, Bae YJ, Kim JH, Shin YK, Jeung HC. HER2 Regulates Cancer Stem Cell Activities via the Wnt Signaling Pathway in Gastric Cancer Cells. Oncology. 2019;97(5):311–8.

12. Sandoo A, Kitas GD, Carmichael AR. Breast cancer therapy and cardiovascular risk: focus on trastuzumab. Vasc Health Risk Manag. 2015;11:223–8.

13. Howard MF, Murakami Y, Pagnamenta AT, Daumer-Haas C, Fischer B, Hecht J, et al. Mutations in PGAP3 impair GPI-anchor maturation, causing a subtype of hyperphosphatasia with mental retardation. Am J Hum Genet. 2014;94(2):278–87.

14. Singh RR, Patel KP, Routbort MJ, Aldape K, Lu X, Manekia J, et al. Clinical massively parallel next-generation sequencing analysis of 409 cancer-related genes for mutations and copy number variations in solid tumours. Br J Cancer. 2014;111(10):2014–23.

15. Greenleaf AL. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. Transcription. 2019;10(2):91–110.

16. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev. 2011;25(20):2158–72.

17. Bosken CA, Farnung L, Hintermair C, Merzel Schachter M, Vogel-Bachmayr K, Blazek D, et al. The structure and substrate specificity of human Cdk12/Cyclin K. Nat
18. Dubbury SJ, Boutz PL, Sharp PA. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. Nature. 2018;564(7734):141–5.

19. Iorns E, Martens-de Kemp SR, Lord CJ, Ashworth A. CRK7 modifies the MAPK pathway and influences the response to endocrine therapy. Carcinogenesis. 2009;30(10):1696–701.

20. Kauraniemi P, Barlund M, Monni O, Kallioniemi A. New amplified and highly expressed genes discovered in the ERBB2 amplicon in breast cancer by cDNA microarrays. Cancer Res. 2001;61(22):8235–40.

21. Comprehensive molecular portraits. of human breast tumours. Nature. 2012;490(7418):61–70.

22. Tien JF, Mazloomian A, Cheng SG, Hughes CS, Chow CC, Canapi T. L T, et al. CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. Nucleic Acids Res. 2017;45(11):6698–716.

23. Wilhelm LP, Wendling C, Vedie B, Kobayashi T, Chenard MP, Tomasetto C. STARD3 mediates endoplasmic reticulum-to-endosome cholesterol transport at membrane contact sites. 2017;36(10):1412–1433.

24. Wilhelm LP, Tomasetto C, Alpy F. Touche! STARD3 and STARD3NL tether the ER to endosomes. Biochem Soc Trans. 2016;44(2):493–8.

25. Vassilev B, Sihto H, Li S, Holtta-Vuori M, Ilola J, Lundin J, et al. Elevated levels of StAR-related lipid transfer protein 3 alter cholesterol balance and adhesiveness of breast cancer cells: potential mechanisms contributing to progression of HER2-positive breast cancers. Am J Pathol. 2015;185(4):987–1000.

26. Agrawal R, Garg A, Benny Malgulwar P, Sharma V, Sarkar C, Kulshreshtha R. p53 and miR-210 regulated NeuroD2, a neuronal basic helix-loop-helix transcription factor, is
downregulated in glioblastoma patients and functions as a tumor suppressor under hypoxic microenvironment. 2018;142(9):1817–1828.

27. Lacle MM, Moelans CB, Kornegoor R, van der Pol C, Witkamp AJ, van der Wall E, et al. Chromosome 17 copy number changes in male breast cancer. Cell Oncol (Dordr). 2015;38(3):237–45.

28. Avanes A, Lenz G, Momand J. Darpp-32 and t-Darpp protein products of PPP1R1B: Old dogs with new tricks. Biochem Pharmacol. 2019;160:71–9.

29. Alam SK, Astone M, Liu P, Hall SR, Coyle AM. DARPP-32 and t-DARPP promote non-small cell lung cancer growth through regulation of IKKalpha-dependent cell migration. 2018;1:43.

30. Belkhiri A, Zhu S, El-Rifai W. DARPP-32: from neurotransmission to cancer. Oncotarget. 2016;7(14):17631–40.

31. Lenz G, Hamilton A, Geng S, Hong T, Kalkum M, Momand J, et al. t-Darpp Activates IGF-1R Signaling to Regulate Glucose Metabolism in Trastuzumab-Resistant Breast Cancer Cells. Clin Cancer Res. 2018;24(5):1216–26.

32. Gu L, Waliany S, Kane SE. Darpp-32 and its truncated variant t-Darpp have antagonistic effects on breast cancer cell growth and herceptin resistance. PLoS One. 2009;4(7):e6220.

33. Francis A, Sunitha B, Vinodh K, Polavarapu K, Katkam SK, Modi S, et al. Novel TCAP mutation c.32C > A causing limb girdle muscular dystrophy 2G. PLoS One. 2014;9(7):e102763.

34. Serie DJ, Crook JE, Necela BM, Axenfeld BC, Dockter TJ, Colon-Otero G, et al. Breast Cancer Clinical Trial of Chemotherapy and Trastuzumab: Potential Tool to Identify Cardiac Modifying Variants of Dilated Cardiomyopathy. J Cardiovasc Dev Dis. 2017;4(2).
35. Tamir S, Paddock ML, Darash-Yahana-Baram M, Holt SH, Sohn YS, Agranat L, et al. Structure-function analysis of NEET proteins uncovers their role as key regulators of iron and ROS homeostasis in health and disease. Biochim Biophys Acta. 2015;1853(6):1294–315.

36. Lipper CH, Karmi O, Sohn YS, Darash-Yahana M, Lammert H, Song L, et al. Structure of the human monomeric NEET protein MiNT and its role in regulating iron and reactive oxygen species in cancer cells. Proc Natl Acad Sci U S A. 2018;115(2):272–7.

Tables

Table 1 BP enrichment analysis of genes in skyblue3 module

| Term     | Pathway ID  | Pathway description                                      | Count | P value  |
|----------|-------------|----------------------------------------------------------|-------|----------|
| BP       | GO:0000184  | nuclear-transcribed mRNA catabolic process               | 4     | 0.013983 |
| BP       | GO:0046683  | response to organophosphorus                             | 2     | 0.016807 |
| BP       | GO:2000347  | positive regulation of hepatocyte proliferation          | 2     | 0.025106 |
| BP       | GO:0032886  | regulation of microtubule-based process                   | 2     | 0.037425 |
| BP       | GO:0042493  | response to drug                                          | 5     | 0.039765 |
| BP       | GO:0046653  | tetrahydrofolate metabolic process                        | 3     | 0.041497 |
| BP       | GO:0014066  | regulation of phosphatidylinositol 3-kinase signaling     | 2     | 0.043074 |
| BP       | GO:0006081  | cellular aldehyde metabolic process                       | 2     | 0.045552 |
| BP       | GO:0006554  | lysine catabolic process                                  | 2     | 0.049590 |

BP, biological process; GO, gene ontology.

Table 2 CC enrichment analysis of genes in skyblue3 module
### Table 3 MF enrichment analysis of genes in skyblue3 module

| Term           | Pathway ID | Pathway description                        | Count | P value   |
|----------------|------------|---------------------------------------------|-------|-----------|
| MF            | GO:0016491 | oxidoreductase activity                     | 5     | 0.0095388 |
| MF            | GO:0005515 | protein binding                             | 47    | 0.0142282 |
| MF            | GO:0051015 | actin filament binding                      | 4     | 0.0174837 |
| MF            | GO:0004030 | aldehyde dehydrogenase [NAD(P)+ ] activity | 2     | 0.0246271 |
| MF            | GO:0070180 | large ribosomal subunit rRNA binding        | 2     | 0.0286731 |
| MF            | GO:0004028 | 3-chloroallyl aldehyde dehydrogenase activity | 2     | 0.0327025 |

MF, molecular function; GO, gene ontology.

**Figures**
Figure 1

Determination of soft-threshold power in the WGCNA. The left graph displays the scale-free fit index for various soft-thresholding powers. The right graph shows the mean connectivity for various soft-thresholding powers.
Hierarchical cluster dendrograms of all expressed genes based on topological overlap.
Figure 3

TOM heatmap between genes. The heat map describes the topological overlap matrix between 400 random genes. A darker red color indicates higher overlap.
Figure 4

Heatmap of the correlation between module eigengenes and clinical traits of breast cancer.

Each cell contains the correlation and p value of the corresponding module and character.
Figure 5

Scatterplot of gene significance of HER2+ BC versus intramodular module membership in the skyblue3 module.
Figure 6

Eigengene dendrogram representing the relationships between modules and HER2+ BC traits.
Visualization of weighted gene co-expression network analysis (WGCNA) network connections of the intramodular hub genes. The hub genes represented by red dots have the highest degree of association with other genes in the module.
Immunohistochemistry of the hub genes based on the human protein atlas. (A) Protein level of PGAP3 in normal tissue (staining: low; intensity: weak; quantity: > 75%). (B) Protein level of PGAP3 in breast cancer tissue (staining: high; intensity: strong; quantity: > 75%). (C) Protein level of PPP1R1B in normal tissue (staining: medium; intensity: moderate; quantity: > 75%). (D) Protein level of PPP1R1B in breast cancer tissue (staining: high; intensity: strong; quantity: > 75%). (E) Protein level of PNMT in normal tissue (staining: not detected; intensity: negative; quantity: none). (F) Protein level of PNMT in breast cancer tissue (staining: low; intensity: weak; quantity: > 75%). (G) Protein level of ERBB2 in normal tissue
(staining: medium; intensity: moderate; quantity: 75% - 25%). (H) Protein level of ERBB2 in breast cancer tissue (staining: high; intensity: strong; quantity: > 75%). (I) Proteins level of CISD3 in normal tissue (staining: low; intensity: weak; quantity: 75% - 25%). (J) Protein level of CISD3 in breast cancer tissue (staining: medium; intensity: moderate; quantity: > 75%).

(K) Protein level of CRKRS in normal tissue (staining: high; intensity: strong; quantity: > 75%). (L) Protein level of CRKRS in breast cancer tissue (staining: high; intensity: strong; quantity: > 75%). (M) Protein level of TCAP in normal tissue (staining: low; intensity: weak; quantity: 75% - 25%). (N) Protein level of TCAP in breast cancer tissue (staining: low; intensity: moderate; quantity: < 25%). (O) Protein level of STARD3 in normal tissue (staining: medium; intensity: moderate; quantity: > 75%). (P) Protein level of STARD3 in breast cancer tissue (staining: high; intensity: strong; quantity: > 75%).
Overall survival analysis of the hub genes in breast cancer based on cBioPortal database. Red line represents cases with alterations. Blue line represents cases without. (A) PGAP3; (B) PNMT; (C) ERBB2; (D) TCAP; (E) STARD3.