Identification of New Key Genes in Breast Cancer by Co-expression Network Analysis

Xinyang Li
Henan University of Science and Technology Affiliated First Hospital 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 (xshuaiw@126.com)
Henan University of Science and Technology Affiliated First Hospital

Research

Keywords: breast cancer, weighted gene co-expression network analysis (WGCNA), bioinformatics analysis, molecular mechanism

Posted Date: April 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-23381/v1

Abstract

Background: Breast cancer is one of the most common malignancies in women all over the world. This study aimed to identify the potential biomarkers associated with the occurrence and development of breast cancer.

Results: Our research downloaded GSE54140 gene expression datasets and GPL10152 platform information from the Gene Expression Omnibus datasets, and used weighted gene co-expression network analysis (WGCNA) to construct 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 gene in tumor tissues was significantly higher than those in normal tissues. And survival analysis shows that PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival ($P < 0.05$).

Conclusion: A total of 9 candidate biomarkers were identified by comprehensive bioinformatics analysis, among which, the co-expansion of PGAP3 and CRKRS related to ERBB2 may be associated with the occurrence of breast cancer. In addition, PPP1R1B, CRKRS and TCAP are related to drug resistance and adverse reactions in the treatment of breast cancer.

Background

Breast cancer (BC) is the most common cancer in females, which is a serious threat to women's health [1]. Nearly 1.3 million women worldwide are diagnosed with breast cancer every year, and more than 400,000 people die from recurrence and metastasis of breast cancer [2, 3]. People with breast cancer have a great psychological and financial burden. Despite of the continuously improvement of DFS and OS of breast cancer patients from the effective treatment [4], the high rate of recurrence and metastasis is still the major cause of death in breast cancer patients [5]. In-depth study on the mechanism of invasion and metastasis of breast cancer is warranted. Exploring ‘precision therapy’ strategies for different breast cancer subtypes are still the aim and trend of breast cancer research development [4]. In recently years, many microarray profiling studies have been performed in breast cancer, and hundreds of differentially expressed genes have been obtained. However, the difference analysis ignores the interaction between genes so that there is no reliable biomarker for clinical use in breast cancer. Now, the weighted gene co-expression network analysis (WGCNA) is used to construct a scale-free gene co-expression network to explore the associations between gene sets and clinical features and to identify the modules (clusters) of highly related genes and the hub genes in each module [6]. The most salient characteristic of scale-free

networks is the relative commonness of nodes with a degree that greatly exceeds the average. The highest-degree nodes are often called ‘hubs’, and are thought to serve specific purposes in their networks. Co-expression analysis is a powerful technique to construct free-scale gene co-expression networks. Thus, WGCNA is ideal for the identification of gene modules and key genes that contribute to phenotypic traits. In the current study, we downloaded the gene expression microarray from the GEO datasets, and used the WGCNA algorithm to identify the highly correlated gene modules associated with breast cancer, and detected the hub genes that are potential diagnostic and treatment target biomarkers.

Materials And Methods

Microarray data and data pre-processing

Microarray Data of GSE54140 was downloaded from the Gene Expression Omnibus (GEO) datasets (www.ncbi.nlm.nih.gov/gds), which is a freely accessible GENE EXPRESSION database by NCBI. Microarray Data from GSE54140 included 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 study, 63 samples including HER2-positive BC, Luminal-A BC and Luminal-B BC were selected for analysis. Rstudio (3.6.2 version; www.rstudio.com) supported by R software platform (3.6.2 version; www.r-project.org) and relevant software packages were used to process the data. Before WGCNA, this study preprocessed the Microarray Data: the probe name was converted into the gene name using platform information, and microarray data were standardized through the R function. Finally, the microarray data with row name as sample name and column name as gene name was obtained. After standardization and probe summarization, the data set with 15,612 genes was further processed, and the top 25% most variant genes by analysis of variance (11,709 genes) were selected for the construction of co-expression network analysis.

Co-expression network construction

After pre-processing, the selected expression data profiles were constructed to a gene co-expression network using WGCNA package in R (www.cran.r-project. org/web/packages/WGCNA /index.html). The gradient method was used to calculate the connection strength between each pair of nodes and test the independence and the average degree of connectivity of the various modules with different power values (the power values ranged from 1 to 20). In the presented study, the appropriate power value (β) was selected when the degree of independence was 0.8 as the soft threshold parameter to ensure a scale-free network. The hierarchical cluster dendrograms was constructed by using the correlation coefficient between genes, and the genes with similar expression profiles were classified into the same gene module. Modules were identified as gene sets with high topological overlap. Different branches of the cluster dendrograms represented different gene modules, and different colors represented different modules. Subsequently, the soft threshold power was applied to transform the adjacency matrix into topological overlap measure (TOM), and 400 genes were randomly selected to make TOM heat map to prove the high

degree of independence between modules and the relative independence of gene expression in each module.

Module and clinical trait association analysis

Module Eigengenes (MEs) were defined 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 calculated by WGCNA algorithm to evaluate the potential correlation between gene modules and clinical traits. Subsequently, the module with the highest correlation coefficient was used for analysis. Gene Significance (GS) and Module Significance (MS) were used to calculate the expression patterns of modules related to sample types. GS was the correlation coefficients for different kinds of samples, and MS was the absolute value mean of GS of all genes in the module. Module membership (MM) was used to evaluate the degree of association between genes and modules and to filter hub genes in the target module.

Enrichment analysis of module

The enrichment analysis of module was conducted through Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis to explore the biological functions using the DAVID bioinformatics tool (version 6.8; www.david.ncifcrf.gov/). Go term enrichment analysis and KEGG pathway contain biological process (BP), cellular component (CC), molecular function (MF) and KEGG pathway. $P < 0.05$ as the threshold was considered statistically significant.

Hub genes identification and analysis

The corresponding heat map was obtained by analyzing 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 used to screen the module. Module with MCODE score > 59.5 were presented, and then the hub genes that degree > 74 were identified by analysis. The Human Protein Atlas database (www.proteinatlas.org) was used to detect the protein level of hub genes in tissues. Subsequently, to validate the hub genes, we used the Breast Cancer (METABRIC, Nature 2012 & Nat Commun 2016) from cBioPortal database (www.cbioportal.org/), which composed of 2,509 breast cancers samples/patients. Hub gene names were submitted in cBioPortal, and survival analysis of determining the importance of genes in biological processes was conducted.

Statistical method

In this study, R statistical software and WGCNA package were used for statistical calculation. WGCNA was used to construct free-scale gene co-expression networks to determine the relationships 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.

Results

Data processing

In this study, GSE54140 (15,612 genes) composed of gene expression data from 63 samples including HER2-positive BC, Luminal-A BC and Luminal-B BC was selected and conducted by the R software function, and then microarray data were standardized. Using platform file GPL10152, the probe name was converted into the gene name, finally, the microarray data with row name as sample name and column name as gene name and the top 25% most variant genes by analysis of variance (11,709 genes) were selected for the construction of co-expression network.

Co-expression network construction and key modules identification

The co-expression analysis was carried out to construct the co-expression network. Select the appropriate weighting coefficient β to ensure a scale-free network, in this study, the power of β value of 3 was selected to construct a co-expression network, as shown in Figure 1. We then calculated the TOM for each gene pair, 60 modules were displayed by hierarchical clustering according to degree of TOM's difference as shown in Figure 2. Each module contained a group of coordinately expressed genes with high TOM, and was potentially involved in shared biological processes. To distinguish the modules individually, each module was assigned a unique color. 400 genes were randomly selected to make TOM heat map, as shown in Figure 3, which shows 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) had significant correlation with HER2+ BC (cor=0.74, $P=3e^{-12}$). By calculating the correlation coefficient of GS and MM in skyblue3 module, and cluster analysis of traits and modules, the significant correlation between skyblue3 module and HER2+ BC was determined, and the credibility of the results was verified, as shown in Figure 5 and 6. Hence, genes in the skyblue3 module were selected for further analysis.

Functional enrichment analysis of genes in modules

All the name of genes in the skyblue3 module was submitted to DAVID bioinformatics tool. For the BP, the genes were mainly enriched in nuclear-transcribed mRNA catabolic process, response to organophosphorus, positive regulation of hepatocyte proliferation, regulation of microtubule-based process, response to drug, tetrahydrofolate metabolic process, regulation of phosphatidylinositol 3-kinase signaling, cellular aldehyde metabolic process, lysine catabolic process, as shown in Table 1. The genes in the CC group were mainly enriched in cytosol, cytoplasm, ribosome, cytosolic large ribosomal subunit, membrane, neuronal cell, 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, large ribosomal subunit rRNA binding, 3-chloroallyl aldehyde dehydrogenase activity, as shown in

Table 3. The KEGG pathway analysis was performed and showed that genes are mainly involved in Regulation of actin cytoskeleton (P > 0.05).

Hub genes identification and analysis

We found that the skyblue3 module had significant 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 and the intra-modular connectivity, which was calculated by WGCNA. Using MCOD function and correlation degree analysis, the MCODE score >59.5 were presented, and then the hub genes that degree > 74 were regarded as the hub gene represented by a red dot, 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 PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, and STARD3 gene in tumor tissues were significantly higher than those in normal tissues, as shown in Figure 8. Then, all the hub genes underwent survival analysis using cBioPortal datasets. As shown in Figure 9, PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival.

Discussion

Breast cancer is a life-threatening disease for females and one of the most frequently occurred malignancies in the world. Despite of the great progress in treatment options in the past decades, much more remains to be done, and 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 breast.

In this study, we used WGCNA analysis to dynamically study genes co-expression in HER2-positive BC, Luminal-A BC, and Luminal-B BC, and to explore the related modules and hub genes. We conclude that the skyblue3 module had the highest correlation with HER2-positive BC. Skyblue3 module related to HER2-positive BC has well defined functions including nuclear-transcribed mRNA catabolic process, cytosol, and oxidoreductase activity. Among the genes in the skyblue3 module, PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, STARD3, and NEUROD2 were regarded as the hub genes.

PGAP3 is a member of glycosylated phosphatidylinositol (GPI)-specific phospholipase, and involved in the lipid remodeling steps of GPI-anchor maturation and plays a role in protein sorting and trafficking [7]. Some studies have confirmed that using the ion proton plate system, PGAP3 was found to be amplified simultaneously in breast cancer samples, with specific amplification of the locus harboring ERBB2 and PGAP3 [8]. Considering the close connection between the development of breast cancer and ERBB2 gene, the variation of PGAP3 copy number that affects the function of ERBB2 may also be participated breast cancer oncogenesis. PPP1R1B is a protein coding multiple transcripts and two experimentally-documented proteins Darpp-32 and t-Darpp [9]. And the protein encoded by this gene has been confirmed to highly overexpressed in breast cancer, colon cancer, esophagus cancer, lung cancer, et al [10, 11]. Resistance to the anti-Her2 antibody Trastuzumab may be related to the overexpression of t-darpp/darpp-32 protein that activates the AKT pathway and ultimately leads to cell survival and apoptosis blocking

[12]. ERBB2, commonly referred to as HER2, encodes a member of the epidermal growth factor receptor family of receptor tyrosine kinases, it is amplified and/or overexpressed in 20-30% of invasive breast carcinomas [13]. Amplification and/or overexpression of this gene have also been reported in many other cancers including ovarian and gastric cancer [14-16]. ERBB2 binds tightly to other ligand-bound EGF receptor family members to form a heterodimer, stabilizing ligand binding and enhancing kinase-mediated activation of downstream signaling pathways, such as PI3K/AKT activation and RET signaling [17]. ErbB2 overexpression is tumor cells specific, and closely related to cancer cell proliferation, anti-apoptosis and motor ability. As a cell surface-associated protein, it is amenable to targeted inhibition by small molecules [18]. CISD3 is a member of the CDGSH domain-containing family, CISD3 codes mitochondrial inner NEET protein which resides inside the mitochondrial matrix [19]. And it may be a key player in the mitochondrial pathways such as regulating autophagy, apoptosis, and mitochondrial iron and reactive oxygen species homeostasis [20]. CRKRS, also named CDK12, encodes cyclin-dependent kinase 12 that phosphorylates RNA polymerase II, thereby acting as a key regulator of transcription [21]. Cdk12 maintains genomic stability by suppressing intron polygadenylation to regulate DNA repair genes [22-24]. Related studies show that endocrine therapy resistance in breast cancer is related to the activation of MAPK signal pathway by CRKRS silencing, which leads to the loss of endoplasmic reticulum dependence [25]. In breast cancers, CDK12 is frequently co-amplified with the HER2 (ERBB2) oncogene. Recent studies found direct correlations between CDK12 levels, DNAJB6 isoform levels and the migration capacity and invasiveness of breast tumor cells, and the team finally confirmed regulation by CDK12 modulates alternative last exon splicing of the DNA damage response activator ATM and a DNAJB6 isoform that influences cell invasion and tumorigenesis in xenografts, and suggests that CDK12 gene amplification can contribute to the pathogenesis of the cancer [26]. The protein encoded by TCAP gene were found in striated and cardiac muscle, it binds to the titin Z1-Z2 domains, it is a muscle assembly regulating factor and a substrate of titin kinase [27]. Among it is related pathways are cardiac conduction and striated muscle contraction. Doxorubicin and anti-HER2 targeting therapy of trastuzumab are frequently used breast cancer treatment agent, their most common adverse reaction is cardiac toxicity. Recent study reported that TCAP genetic variation may be associated with the development or progression of cardiomyopathy [28]. STARD3, StAR Related Lipid Transfer Domain Containing 3, is a protein coding gene. The encoded protein localizes to the membranes of late endosomes and may be involved in cholesterol exporting [29, 30]. It facilitates the transport/distribution of cholesterol and sphingolipid to the intracellular membrane compartments and the metabolism of steroid hormones [29, 31]. STARD3 gene is located to the minimal amplicon in HER2-positive breast cancers, it's overexpression leads to increased cholesterol biosynthesis and Src kinase activity in breast cancer cells, suggesting StARD3 over-expression may play a role in breast cancer aggressiveness through increasing membrane cholesterol and enhancing oncogenic signaling [32]. NEUROD2, Neuronal Differentiation 2, encodes a member of the NEUROD family of neurogenic basic helix-loop-helix proteins, which has an impact on the regulation of glutamatergic and GAB-Aergic genes [33]. NEUROD2 gene expression can induce transcription from neuron-specific promoters which contain a specific DNA sequence known as an E-box mediates calcium-dependent transcription activation [34].

The protein levels PGAP3, PPP1R1B, PNMT, ERBB2, CISD3, CRKRS, TCAP, and STARD3 gene in tumor tissues were significantly higher than those in normal tissues. And survival analysis shows that PGAP3, PNMT, ERBB2, TCAP, and STARD3 were negatively associated with the overall survival. These results suggest that these genes may be tumorigenic genes in breast cancer.

Conclusion

In summary, our study used a systems biology-based WGCNA approach to determine the module highly related to HER2 positive BC, whose function was mainly confined in nuclear-transcribed mRNA catabolic process, cytosol, oxidoreductase activity and actin cytoskeleton regulation. Though the majority of hub genes highlighted in this study have been previously reported, there is no comprehensive analysis of these genes. ERBB2 plays an important role in the occurrence and development of breast cancer. In addition, we found that PGAP3 and CRKRS were related to the co-amplification of ERBB2, PPP1R1B may mediate anti-ERBB2 drug resistance by activating AKT pathway, TCAP may be related to cardiomyopathy caused by doxorubicin or Trastuzumab, and STARD3 may contribute to ERBB2+ breast cancer aggressiveness through increasing membrane cholesterol and enhancing oncogenic signaling. As endocrine therapy is the most important treatment for hormone receptor positive breast cancer. It has been found that CRKRS can silence the activation of mitogen activated protein kinase (MAPK) signal pathway leading to endocrine therapy resistance. Hub genes such as PGAP3 and STARD3 are highly correlated with breast cancer development. However, PPP1R1B, CRKRS, and TCAP may be brought new insights in breast cancer study in treatment target. Further investigation about these genes is warranted.

Abbreviations

BC: breast cancer; 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.

Clinical significance

Our study reports several genes of breast cancer by Co-expression network analysis, they are associated with cancer’s development and patient’s prognosis and may serve as the therapeutic targets for the disease, especially in resistance and adverse reactions in the treatment of breast cancer.

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.

Authors' contributions

LX and WX wrote and revised the manuscript; LX was the main contributor to this manuscript; WX critically revised and corrected the manuscript; WY, WZ, FJ, LG and YG Analysis and interpretation of relevant data; All authors read and approved the final manuscript.

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, et al. 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 R L, Miller K D, Jemal A. Cancer statistics, 2018. 2018;68(1):7-30.
3. DeSantis C E, et al. Breast cancer statistics, 2017, racial disparity in mortality by state. CA Cancer J Clin. 2017;67(6):439-448.
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. Howard M F, 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.
8. Pan X, et al. Identification of the copy number variant biomarkers for breast cancer subtypes. 2019;294(1):95-110.

9. Avanes A, Lenz G, Momand J. Darpp-32 and t-Darpp protein products of PPP1R1B: Old dogs with new tricks. Biochem Pharmacol. 2019;160:71-79.
10. Alam S K, et al. DARPP-32 and t-DARPP promote non-small cell lung cancer growth through regulation of IKKalpha-dependent cell migration. 2018;1:43.
11. Belkhiri A, Zhu S, El-Rifai W. DARPP-32: from neurotransmission to cancer. Oncotarget. 2016;7(14):17631-40.
12. Gu L, Waliany S, Kane S E. 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.
13. Harbeck N, Gnant M. Breast cancer. Lancet. 2017;389(10074):1134-1150.
14. Aznab M, et al. 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):1989-1994.
15. Maennling A E, et al. Molecular Targeting Therapy against EGFR Family in Breast Cancer: Progress and Future Potentials. Cancers (Basel). 2019;11(12).
16. 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.
17. Jung D H, et al. HER2 Regulates Cancer Stem Cell Activities via the Wnt Signaling Pathway in Gastric Cancer Cells. Oncology. 2019;97(5):311-318.
18. Sandoo A, Kitas G D, Carmichael A R. Breast cancer therapy and cardiovascular risk: focus on trastuzumab. Vasc Health Risk Manag. 2015;11:223-8.
19. Tamir S, 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.
20. Lipper C H, 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):272-277.
21. Greenleaf A L. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. Transcription. 2019;10(2):91-110.
22. Blazek D, 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.
23. Bosken C A, et al. The structure and substrate specificity of human Cdk12/Cyclin K. Nat Commun. 2014;5:3505.
24. Dubbury S J, Boutz P L, Sharp P A. CDK12 regulates DNA repair genes by suppressing intronic polyadenylation. Nature. 2018;564(7734):141-145.
25. Iorns E, et al. CRK7 modifies the MAPK pathway and influences the response to endocrine therapy. Carcinogenesis. 2009;30(10):1696-701.
26. Tien J F, et al. CDK12 regulates alternative last exon mRNA splicing and promotes breast cancer cell invasion. Nucleic Acids Res. 2017;45(11):6698-6716.
27. Francis A, et al. Novel TCAP mutation c.32C>A causing limb girdle muscular dystrophy 2G. PLoS One. 2014;9(7):e102763.

28. Serie D J, 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).
29. Wilhelm L P, et al. STARD3 mediates endoplasmic reticulum-to-endosome cholesterol transport at membrane contact sites. 2017;36(10):1412-1433.
30. Alpy F, et al. STARD3 or STARD3NL and VAP form a novel molecular tether between late endosomes and the ER. J Cell Sci. 2013;126(Pt 23):5500-12.
31. Wilhelm L P, Tomasetto C, Alpy F. Touche! STARD3 and STARD3NL tether the ER to endosomes. Biochem Soc Trans. 2016;44(2):493-8.
32. Vassilev B, 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.
33. Agrawal R, et al. 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.
34. Wilke S A, et al. NeuroD2 regulates the development of hippocampal mossy fiber synapses. Neural Dev. 2012;7:9.

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	0.0139832
BP	GO:0046683	response to organophosphorus	2	0.0168073
BP	GO:2000347	positive regulation of hepatocyte proliferation	2	0.0251062
BP	GO:0032886	regulation of microtubule-based process	2	0.0374253
BP	GO:0042493	response to drug	5	0.0397658
BP	GO:0046653	tetrahydrofolate metabolic process	3	0.0414974
BP	GO:0014066	regulation of phosphatidylinositol 3-kinase signaling	2	0.0430742
BP	GO:0006081	cellular aldehyde metabolic process	2	0.0455526
BP	GO:0006554	lysine catabolic process	2	0.0495908

BP, biological process; GO, Gene Ontology.

Table 2 CC enrichment analysis of genes in skyblue3 module

Term	Pathway ID	Pathway description	Count	P value
CC	GO:0005829	cytosol	25	0.0033461
CC	GO:0005737	cytoplasm	31	0.0273077
CC	GO:0005840	ribosome	4	0.0321648
CC	GO:0022625	cytosolic large ribosomal subunit	3	0.0327510
CC	GO:0016020	membrane	16	0.0364893
CC	GO:0043025	neuronal cell	5	0.04262202

CC, cellular component; GO, Gene Ontology.

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

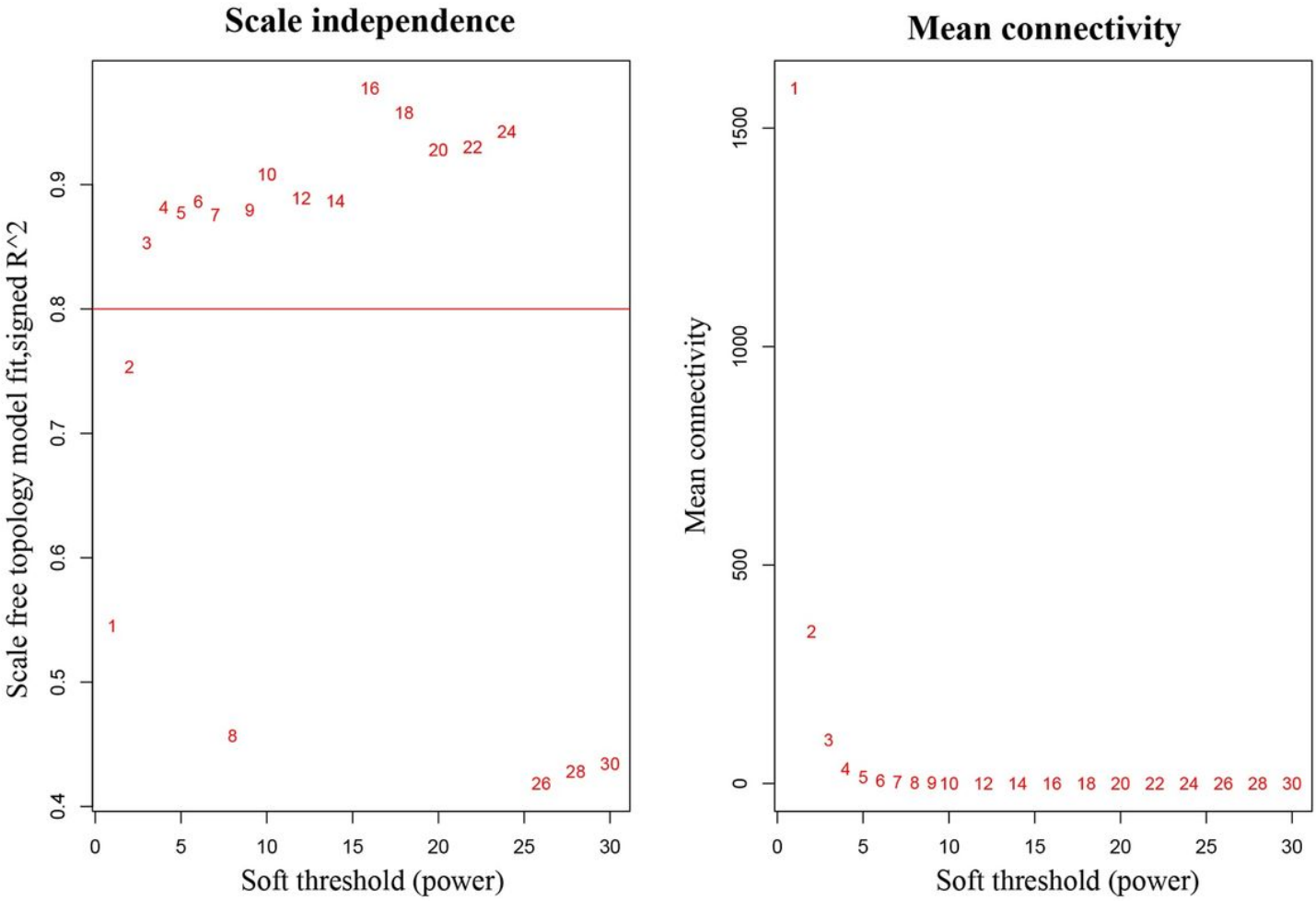

Figure 1

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.

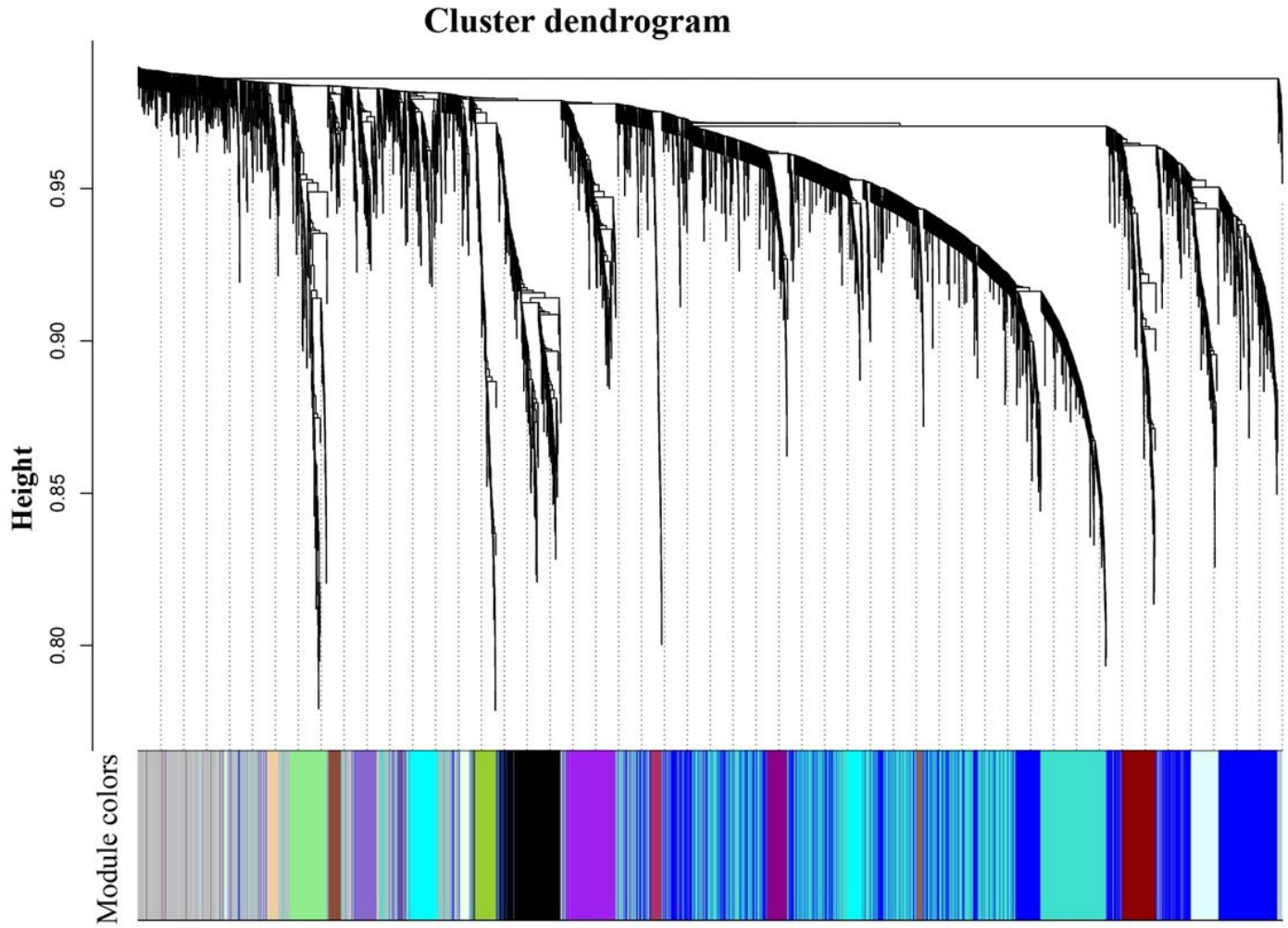

Figure 2

Hierarchical cluster dendrograms of all expressed genes based on topological overlap.

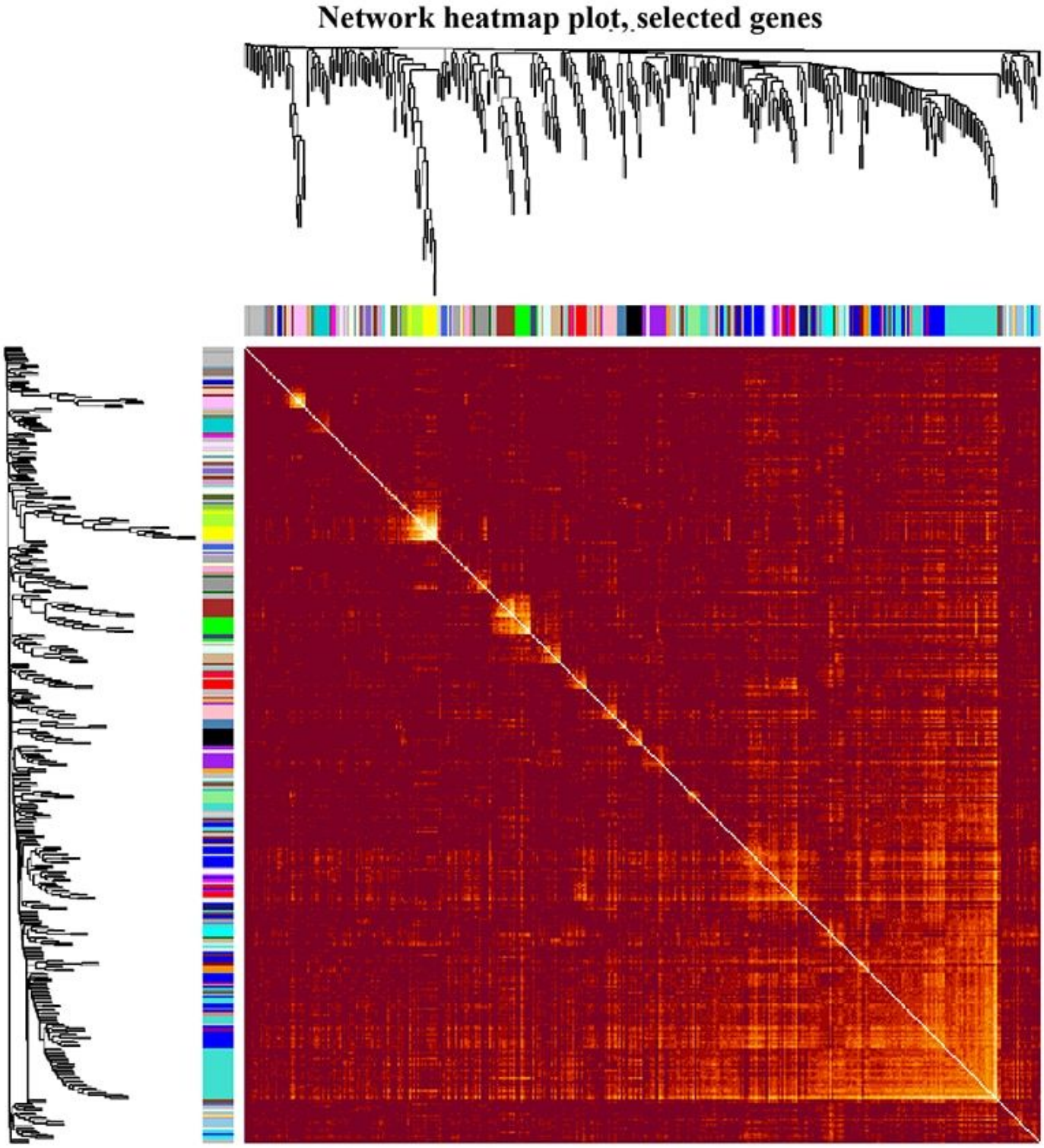

Figure 3

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

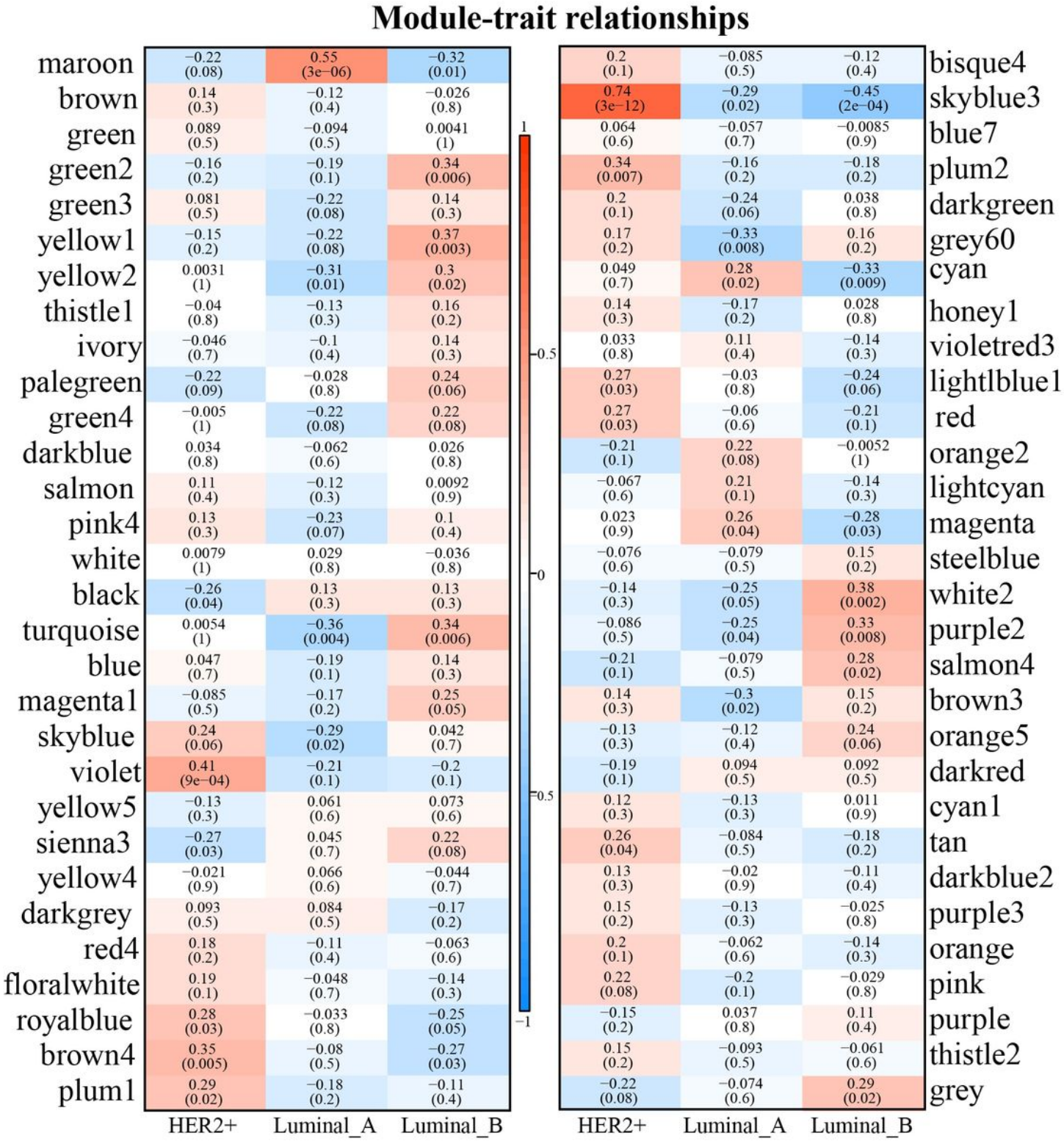

Figure 4

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.

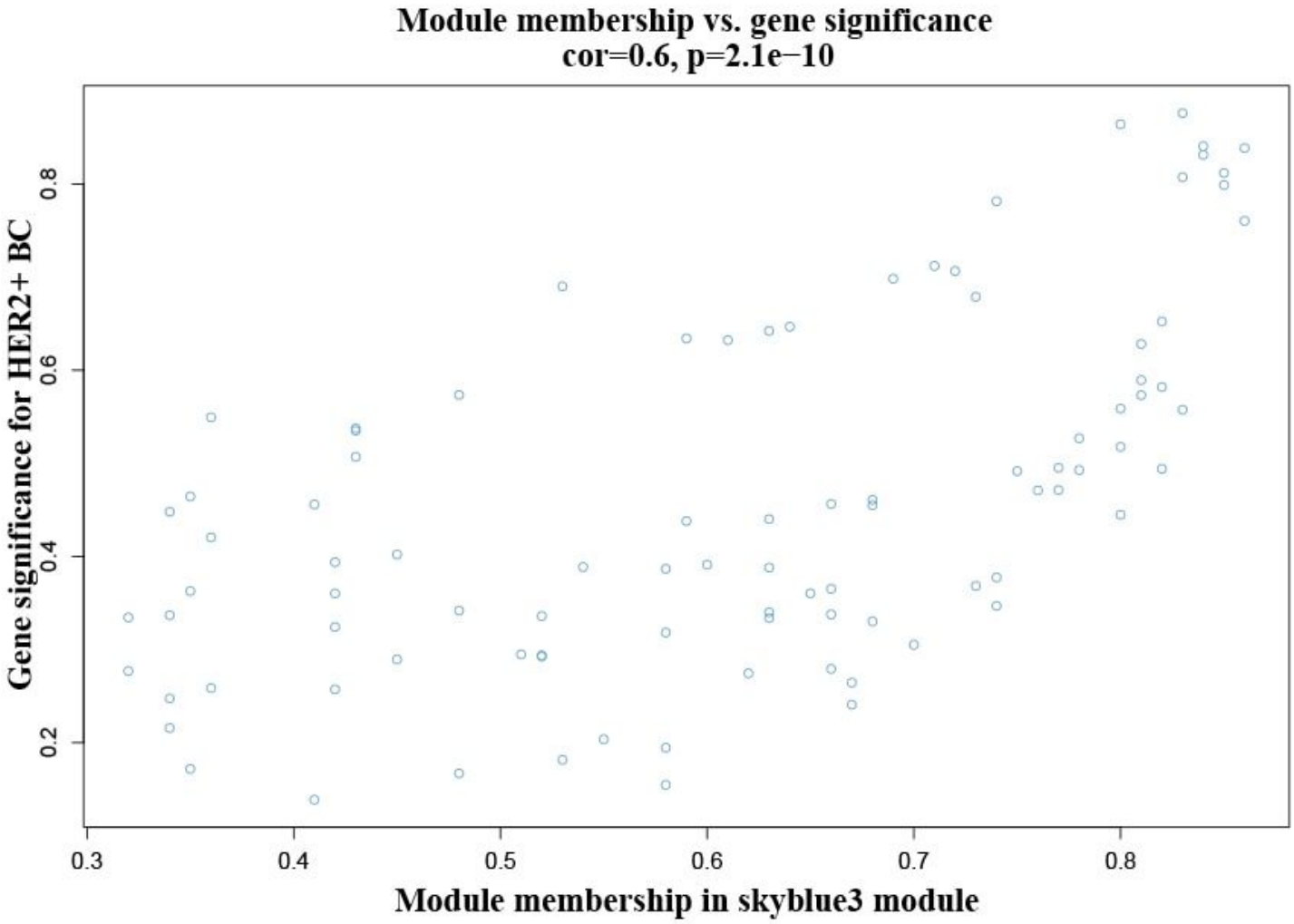

Figure 5

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

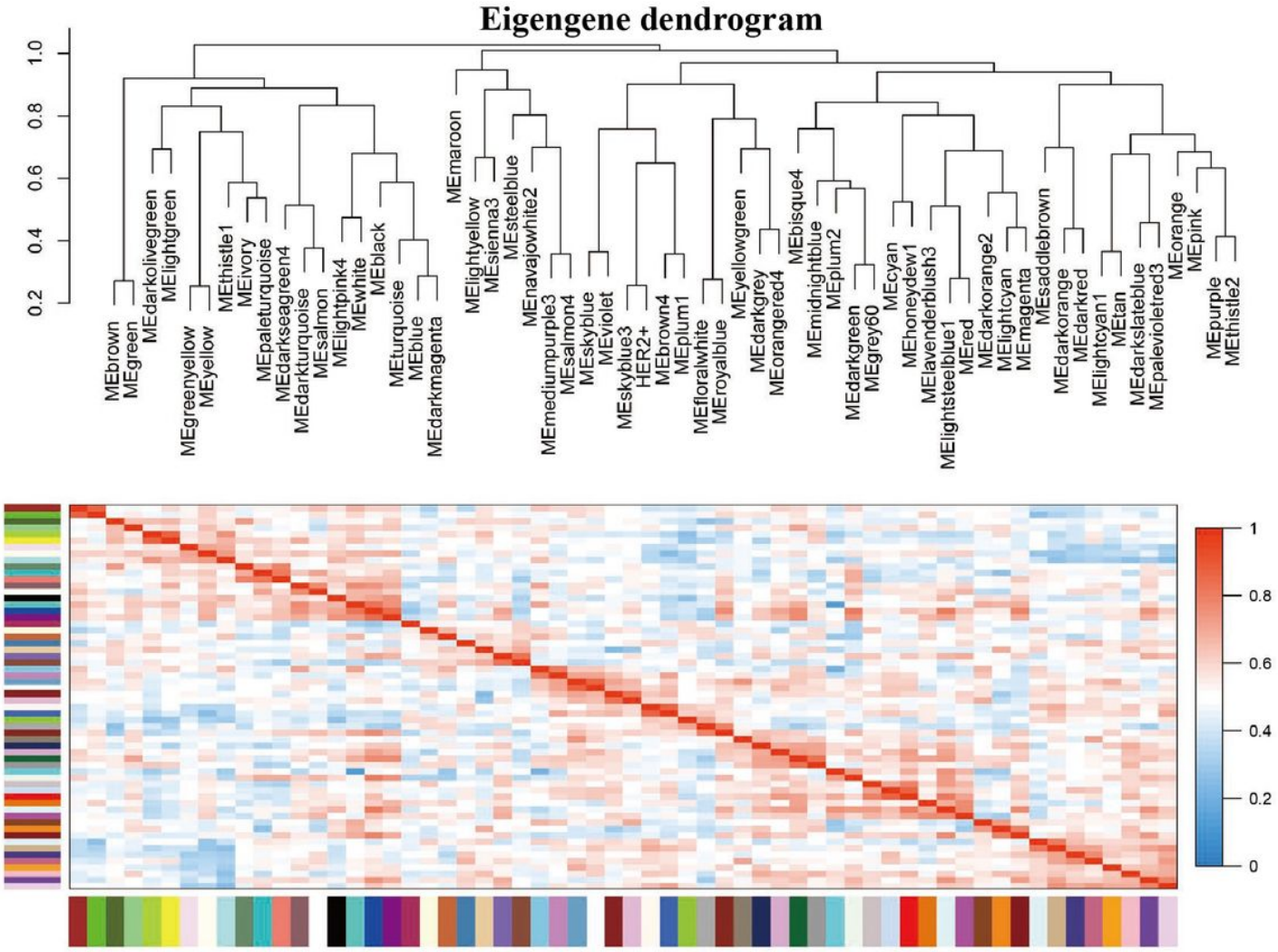

Figure 6

Eigengene dendrogram representing the relationships between modules and HER2+ BC traits.

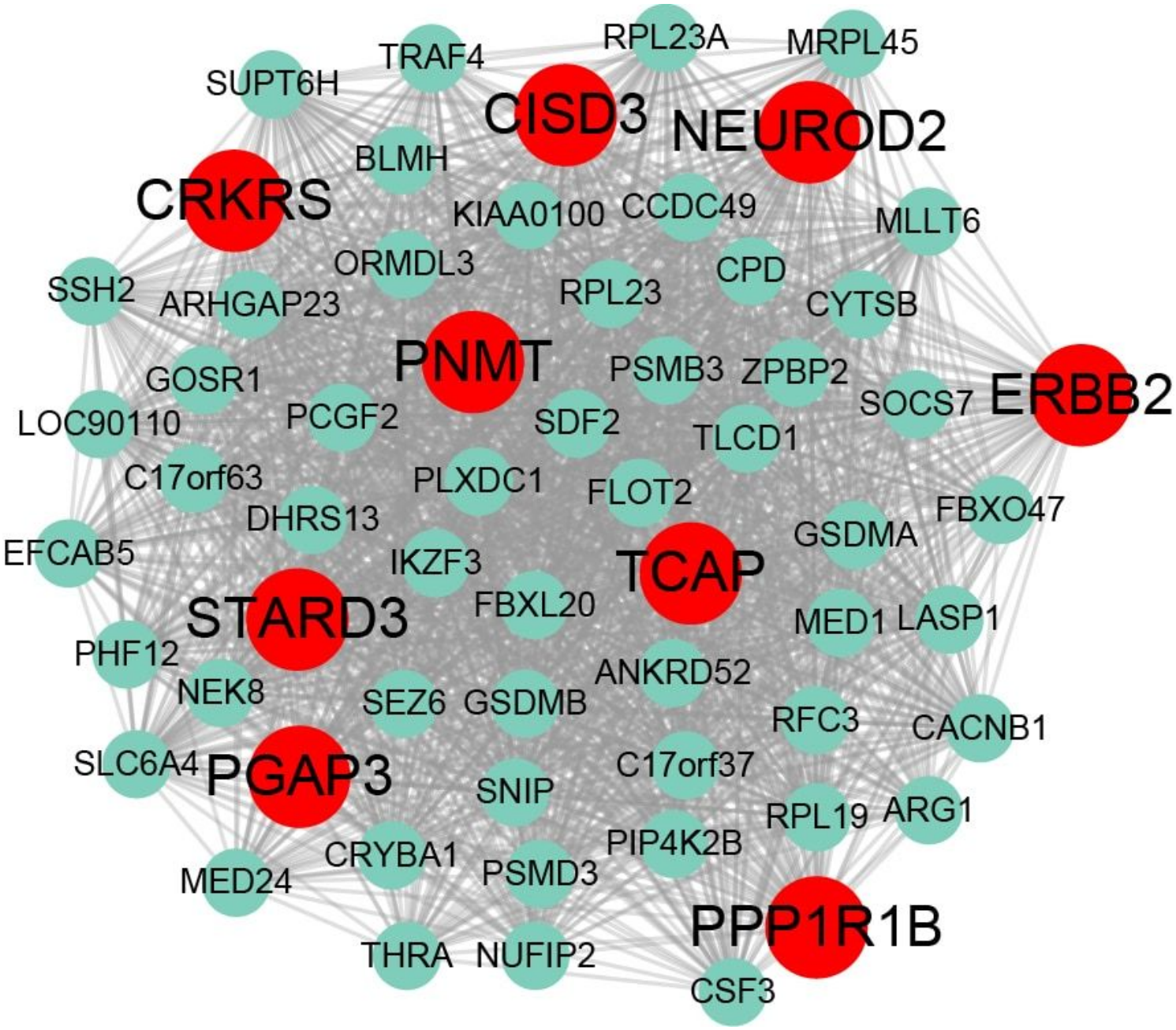

Figure 7

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.

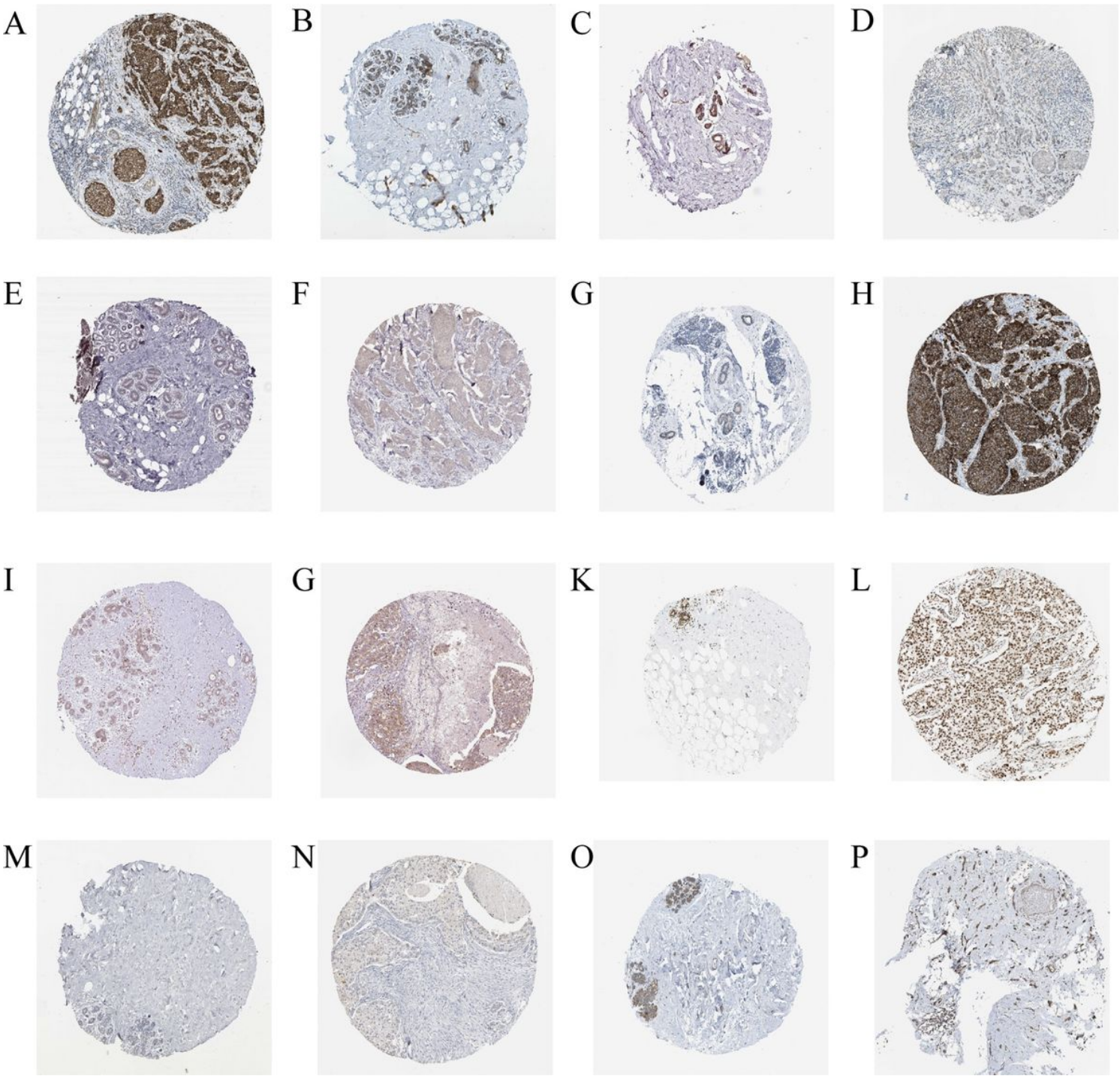

Figure 8

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%).

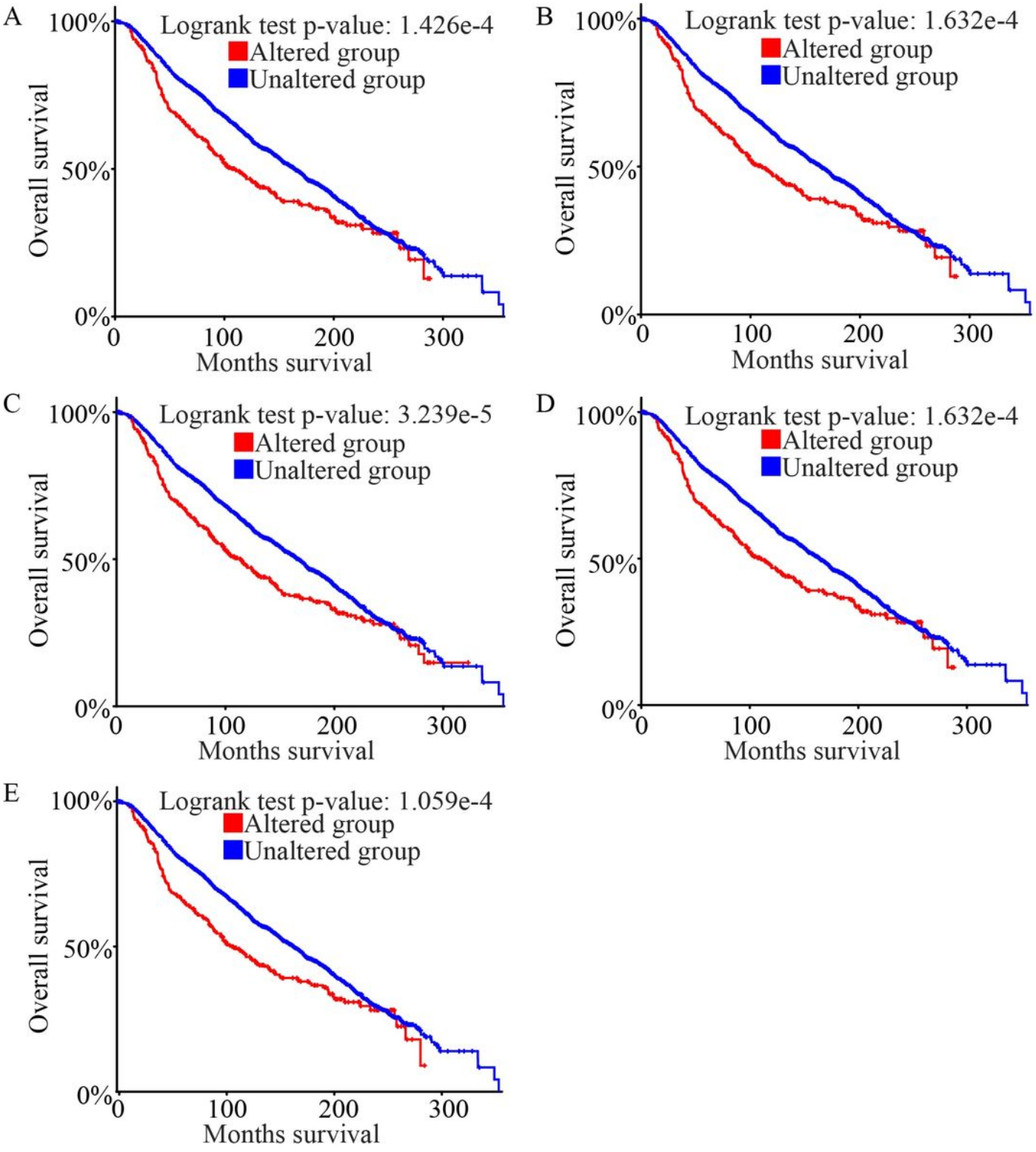

Figure 9

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.