Comprehensive network analysis of integrin $\alpha 11$ expressed in human breast cancer

Yun Fu

Shanghai First Maternity and Infant Hospital

Xiaoyin Sun

Shanghai First Maternity and Infant Hospital

Zhigang Zhuang (✉ zhuang_zg@163.com)

Shanghai First Maternity and Infant Hospital <https://orcid.org/0000-0002-7827-4402>

Research article

Keywords: TCGA database, breast cancer, co-expression, ITGA11

Posted Date: October 3rd, 2019

DOI: <https://doi.org/10.21203/rs.2.15539/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Breast cancer is the most common malignancy in women worldwide, but few genes have been reported to be involved in breast cancer development. **Methods:** We constructed a co-expression network to expand upon the knowledge of the various molecular biomarkers in breast cancer development. Transcriptome data for the tissues of all breast cancer and para-carcinoma types were retrieved from The Cancer Genome Atlas (TCGA) database. We performed a co-expression analysis of breast cancer transcriptome data using Pearson's correlation coefficient (PCC) for inter-sequence mutual rank (MR)-based cutoffs using 11 guide genes and employed Kaplan-Meier survival analyses to assess the prognostic values of hub genes. **Results:** A co-expression network centered on the Integrin $\alpha 11$ (ITGA11) of the extracellular matrix (including 11 guide genes and 72 edge genes) was extracted for breast cancer. The gene ontology (GO) analysis showed that various genes may also be included in the breast cancer network. Among them, CEMIP, COL11A1, CTHRC1, ITGA11, LUM and P4HA3 were negatively associated with the overall survival, and ADAMTS12 and LOXL1 were positively associated with the overall survival at early stage. However, there was no significant difference between the eight gene expressions and Disease-free survival. **Conclusions:** Our co-expression analysis includes the isolated transcriptome of breast cancer, which is a useful resource for breast cancer researchers, as it enables them to elucidate important and complex biological events to prevent and predict cancer.

Background

Breast cancer is one of the most common malignancies in women. There are approximately 1.2 million new breast cancer cases and 500,000 deaths worldwide each year[1]. The occurrence and development of breast cancer is a complex biological process involving multi-gene participation, multi-factor action, and multi-stage progression, which depends on the interaction and regulation of multiple genes, transcripts, and proteins[2]. A promising strategy for the treatment of breast cancer and other cancers is molecular-targeted therapy, which requires a detailed understanding of the specific molecular events and different pathological pathways involved in the development and proliferation of breast cancer. In order to better classify and characterize tumors, it is necessary to predict breast cancer by seeking its associated biomarkers, which may contribute to the design of multimodal treatment options for designated individuals, and may also improve clinical efficiency and clinical outcome assessments[3].

To study candidate biomarkers in cancer, it is important to systematically analyze the interaction and functions of genes that exhibit similar expression patterns. For breast cancer, one of the most effective methods is to examine a co-expression network. Analysis of gene co-expression networks can calculate the correlation between gene expression through large-sample expression profiles, and incorporate highly-related genes into biologically meaningful regulatory networks to visually reveal the relationship between functional molecules in biological systems. It can be used to discover key target genes with important biological functions and significance[4,5]. To date, TCGA program has collected and offered a large amount of transcriptome data on breast cancer, which is valuable for seeking comprehensive understanding of new developments in transcriptional pathways and mining biomarker gene.

In this study, we searched the TCGA database for transcriptome data from the tissues of all breast cancers and para-carcinomas involved in the development of breast cancer. We analyzed each of the guide genes and extracted a network of structural components of the extracellular matrix (ECM) (11 guide genes and 72 edge genes). All genes obtained from the co-expression network of this study may act as resources when exploring candidate biomarker genes in the further study of cancer prevention and control.

Methods

Microarray data

We searched the transcriptome data of breast cancer from the TCGA, including the gene profiles of all breast cancer and para-carcinoma tissues, for a total of 1,028 samples (see Table S1).

Gene ontology (GO) and expression coherence (EC)

The GO terminology was analyzed to obtain the genes associated with breast cancer and to classify their cellular components (CC), biological processes (BP), and molecular functions (MF). The GO terminology for the GO enrichment analysis and EC analysis was retrieved from DAVID (<http://david.abcc.ncifcrf.gov/>). For the GO enrichment analysis, the significant difference of GO enrichment in co-expressed genes was evaluated against a background set consisting of 15,438 genes, at least one of which was derived from a human co-expressed network dataset using hypergeometric distribution. Without multiple test correlations, the BP p -value <0.01 and MF p -value <0.1 were set as significance thresholds. EC analysis was performed as described by Aya K et al.[6]: Genes with the same GO term (containing 20–100 genes) were used as a set of predefined functional genes. To calculate the EC of the paired genes for each GO classification, Pearson's correlation coefficient (PCC) was used to measure expression similarity. A threshold of 0.722 corresponded to the median of all PCC values, while $>99\%$ of the random PCC distribution was derived from 1,000 random genes (yielding approximately $1,000 \times 999 \times 0.5$ gene pairs). To calculate the random EC of the GO category, the random gene set was sampled to the same size as the survey category.

Co-expression analysis and hub genes validation

To construct a co-expression network, we calculated the PCC values for all combinations of the 60,488 unique genes present in the 1,028 downloaded samples obtained from the TCGA database. We estimated a PCC threshold of 0.722, corresponding to the median of all PCC values, which was $>99\%$ th percentile of the random PCC distribution of 1,000 random genes (approximately $1,000 \times 999 \times 0.5$ gene pairs). According to Obayashi's report[7], we also calculated the mutual rank (MR) value between gene pairs as another value that indicated co-expression to further reduce the number of false positives. Only

gene pairs with an absolute MR value <10 were considered to be important linkages to the co-expression network. Calculations were performed with R/Bioconductor. To extract the breast cancer subnet dataset, we took two steps out of the guide gene to extract the vicinity of the network, as described by Mutwil et al. [8]. The network is illustrated using the Cytoscape program. Overall survival and disease free survival of genes from co-expression network in breast cancer based on Kaplan Meier-plotter(www.kmplot.com)[9]. The patients were stratified into high-level group and low-level group according to different expression ratio.

Results

Biological significance of expression similarity

We first used the GO terminology for EC analysis to examine the biological significance of the similarity in the transcriptome datasets of the 1,028 samples downloaded in this study (Table S1). EC analysis is a statistical analysis method used to detect whether the transcriptional profiles of genes belonging to a predefined functional gene set are interrelated[10]. EC scores can measure expression similarity within a predefined functional genome[10]; therefore, when most genes with the same GO term are co-expressed with each other, a higher EC score is obtained. In our EC analysis, EC scores (biological processes [BP], molecular function [MF], and cellular components [CC]) were higher in all three GO groups than in the random sampling (Fig 1). Among the three GO groups, CC showed the highest expression similarity, of which approximately 28.55% of the categories showed higher EC scores than random sampling at threshold EC score of 0.15, while approximately 25.86% and 25.62% did so in BP in MF, respectively (Fig 1 and Table S2).

Construction of the breast cancer network

Using the PCC method, the PCC threshold was 0.722, corresponding to median of all PCC values that were >99 th percentile of the random PCC distribution obtained for 1,000 random genes. We determined the total number of human breast cancers generated from the 1,028 sample data points that we downloaded. The final dataset of the network used in this study contained 40,750 genes (guide) and 209,928 gene pairs (edges). From the entire dataset of the human breast cancer co-expression network, we were able to construct a network around the gene of interest, which served as a guide gene, and used MR-based truncation to obtain genes with very close expression profiles.

The ECM, glycoprotein, signal transduction, and secretion play important roles in tissue development, cancer formation, and invasion. To test whether our co-expression network analysis could identify useful transcriptional networks related to human breast cancer development, we selected 11 genes involved in these pathways or functional processes to find which genes were co-expressed in all samples. After screening by cutoff and MF, a complex network of 72 genes was obtained (Fig 2) and a co-expression network centered on Integrin $\alpha 11$ (*ITGA11*) (guide gene; green circle in Fig 2) was constructed. The blue

circle represents the remaining 10 guide genes. As can be seen from Fig 2 and Table S3, the remaining 10 guide genes share a highly positive correlation with *ITGA11*. These genes were counted according to the UniProtKB keyword enrichment analysis (Table 1). As can be seen from the table, *ITGA11* shares similar functions with the 10 guide genes, and it can also be found in the same pathway.

GO analysis and KEGG pathway analysis of the breast cancer network

The good association between the 11 guide genes in the primary expression cluster indicates that the genes involved in breast cancer can be strictly co-regulated at the expression level; thus, the breast cancer development event may be an appropriate subject for co-expression analysis. The results of the GO analysis also support this (Fig 3). During biological process analysis, terms related to collagen catabolism (e.g. “collagen catabolic process”, “collagen fibril organization”, and “skeletal system development”), terminology related to the structural components of the ECM (e.g. “extracellular matrix organization”, “cell adhesion”, “extracellular matrix disassembly”, “cell–matrix adhesion”, and “integrin-mediated signaling pathway”) appear in the genes of breast cancer collagen catabolism and in the structural components of the ECM, respectively (Fig 3a). the cellular components of those genes obtained in the breast cancer networks were mainly associated with the ECM (e.g., “extracellular matrix”, “proteinaceous extracellular matrix”, “collagen trimer”, “extracellular region”, and “extracellular space”; Fig 3b). Additionally, The molecular functions of the obtained genes are primarily related to binding (e.g. “collagen binding”, “integrin binding”, “heparin binding”, and “calcium ion binding”; Fig 3c).

Furthermore, from the GO analysis (Fig 3), it seems various other genes can be included in the breast cancer network (white circles, Fig 2). This suggests how some genes that have not been reported in breast cancer can be mined from the existing data, and new interactions between these genes can also be identified from connections in the network (Fig 2). Cancer development is a very specific biological event; as such, co-expression network analysis may have the greatest success in identifying genetic interactions in breast cancer.

In addition, we performed a KEGG pathway analysis of the genes obtained from breast cancer co-expression networks. Through the above GO analysis, the pathway related to ECM receptor interaction, focal adhesion, and protein digestion and absorption is constant. We were surprised to find that pathways for infectious diseases, such as amoebiasis, were involved in the pathway enrichment of those genes obtained by co-expression analysis in this study (Table S4). All of these results support our co-expression analysis of gene predictions for breast cancer, which may be useful for subsequent studies and in the design of various medical treatments.

Identification and validation of hub genes

Based on the UniProtKB keyword enrichment analysis, 62 genes with high correlation with ITGA11 were identified as hub genes (Table 1). Survival analysis of hub genes were performed using Kaplan Meier plotter[9]. The patients were stratified into high-level group and low-level group according to different expression ratio. The customized cutoff-high and cutoff-low of the eight genes is ADAMTS12 (72:28), CEMIP (43:57), COL11A1 (51:49), CTHRC1 (25:75), ITGA11 (54:46), LOXL1 (50:50), LUM (73:27) and P4HA3(76:24). Among them, CEMIP, COL11A1, CTHRC1, ITGA11, LUM and P4HA3 were negatively associated with the overall survival, while ADAMTS12 and LOXL1 were positively associated with the overall survival at early stage (Fig 4). However, there was no significant difference between the eight gene expressions and Disease-free survival (Fig S1).

Discussion

The aim of this study was to identify the various gene interactions involved in the biological events associated with breast cancer using the 1,028 datasets from TCGA, specifically exploring human breast cancers and normal tissues. The combined application of these valuable microarray data allows us to build high-resolution networks (Fig 2) that contain useful information to detect gene regulatory networks in biological processes[7]. This suggests that we can perform co-expression screening using reliable network datasets and an effective human cancer database to track the genes involved in specific biological events[11]. In this study, 62 genes with high correlation with ITGA11 were identified as hub genes based on the UniProtKB keyword enrichment analysis. We searched for 11 important genes involved in breast cancer or in other cancer development (*COL8A1*, *COL10A1*, *COL11A1*, *COL12A1*, *ITGA11*, *AEBP1*, *CORIN*, *HSD17B6*, *THBS2*, *ANTXR1*, and *FN1*) to construct a gene co-expression network. Furthermore, among them, CEMIP, COL11A1, CTHRC1, ITGA11, LUM and P4HA3 were negatively associated with the overall survival, ADAMTS12 and LOXL1 were positively associated with the overall survival at early stage. The ECM receptor interaction pathway, the focal-adhesion pathway, and the PI3K–AKT pathway were shown to be involved in the progression of breast cancer[12,13].

ITGA11 is a member of the integrin family, which is involved in various processes that affect the biological behavior of cells, such as metastasis, embryogenesis, hemostasis, the immune response, tissue repair, cancer growth, tumor angiogenesis, and resistance to treatment[14,15]. However, its expression has been shown to be upregulated under malignant conditions, such as in non-small cell lung cancer, which has been suggested to be associated with cancer-cell growth[16,17]. In addition, the high expression of *ITGA11* is associated with the prognosis of breast cancer[18], glioblastoma[19], pancreatic cancer[20] and lung cancer[21]. Speculated from its gene ontology, *ITGA11* can participate in the biological process of collagen binding and cancer-cell differentiation, indicating that it is involved in the pathological process of breast cancer.

He et al's study[22] found that *COL1A1*, *COL3A1*, *COL4A1*, and *COL11A1* significantly enriched in the focal-adhesion and ECM receptor interaction pathways. In this study, different gene expressions in the blue circles, including *COL8A1*, *COL10A1*, *COL11A1*, *COL12A1*, and *ITGA11* in green circle, were identified

as key genes located upstream of these three signaling pathways. The first four genes were reported to be members of the collagen (COL) family and are associated with breast cancer.

FN1 is a glycoprotein that exists as a dimer on the cell surface and in the ECM, involved in cell adhesion, cell migration, wound healing, and cell metastasis[23]. FN1 is upregulated in breast cancer, and it has also been identified as a central protein in previously constructed protein–protein interaction (PPI) networks[24]. A prior gene regulatory network analysis of breast cancer showed that FN1 is upregulated in invasive breast cancer cell lines and is associated with the aggressive behavior of breast cancer cells[25]. FN1 is associated with five significantly enriched KEGG pathways, including “sticky spots”, “protein digestion and absorption”, and “ECM receptor interactions” (Table S3).

ANTXR1 is a transmembrane protein that is highly expressed in breast cancer cells[26]. ANTXR1 interacts with lipoprotein receptor-associated protein(LRP)6 and vascular endothelial growth factor receptor(VEGF)2, respectively, and regulates Wnt and VEGF signaling pathway[27,28]. In addition, ANTXR1 selectively expresses tumor vasculature and promotes tumor angiogenesis[29]. *ANTXR1* expression is predominantly observed in basal-like cell lines and a few luminal cell lines, which has a prognostic impact on ER α -negative/basal breast cancer. Further, Iyengar *et al.* recently confirmed that the adipocyte-derived carboxy terminal fragment of collagen VI α 3 activates AKT and β -catenin and promotes breast cancer progression by acting as *ANTXR1* expressed on cancer cells[30].

Conclusions

In summary, we have identified *COL8A1*, *COL10A1*, *COL11A1*, *COL12A1*, *ITGA11*, *AEBP1*, *CORIN*, *HSD17B6*, *THBS2*, *ANTXR1*, and *FN1* as biomarkers in breast cancer. Among them, CEMIP, COL11A1, CTHRC1, ITGA11, LUM and P4HA3 were negatively associated with the overall survival, and ADAMTS12 and LOXL1 were positively associated with the overall survival at early stage. As such, co-expression microarray data analysis can be an effective tool for identifying more prognostic markers and therapeutic targets in breast cancer.

List Of Abbreviations

TCGA: the Cancer Genome Atlas

PCC: Pearson’s correlation coefficient

MR: mutual rank

ITGA11: Integrin α 11

GO: gene ontology

ECM: extracellular matrix

EC: expression coherence

CC: cellular components

BP: biological processes

MF: molecular functions

COL: collagen

PPI: protein–protein interaction

LRP: lipoprotein receptor-associated protein

VEGF: vascular endothelial growth factor receptor

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The GO terminology for the GO enrichment analysis and EC analysis was retrieved from DAVID (<http://david.abcc.ncifcrf.gov/>).

Funding:

This work was supported by Shanghai Municipal Health Bureau (201640084), and the Fundamental Research Funds for the Central Universities (22120180407).

Authors' contributions:

Writing: Fu Yun

Data analysis and statistics: Sun Xiaoyin and Fu Yun

Editing: Zhuang Zhigang

Competing interests

All authors have no conflicts of interest.

References

1. Donepudi MS, Kondapalli K, Amos SJ, Venkanteshan P (2014) Breast cancer statistics and markers. *J Cancer Res Ther* 10 (3):506-511. doi:10.4103/0973-1482.137927
2. Mertins P, Mani DR, Ruggles KV, Gillette MA, Clauser KR, Wang P, Wang X, Qiao JW, Cao S, Petralia F, Kawaler E, Mundt F, Krug K, Tu Z, Lei JT, Gatz ML, Wilkerson M, Perou CM, Yellapantula V, Huang KL, Lin C, McLellan MD, Yan P, Davies SR, Townsend RR, Skates SJ, Wang J, Zhang B, Kinsinger CR, Mesri M, Rodriguez H, Ding L, Paulovich AG, Fenyo D, Ellis MJ, Carr SA, Nci C (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* 534 (7605):55-62. doi:10.1038/nature18003
3. Hainaut P, Plymoth A (2013) Targeting the hallmarks of cancer: towards a rational approach to next-generation cancer therapy. *Curr Opin Oncol* 25 (1):50-51. doi:10.1097/CCO.0b013e32835b651e
4. Geng R, Li Z, Yu S, Yuan C, Hong W, Wang Z, Wang Q, Yi Z, Fang Y (2018) Weighted gene co-expression network analysis identifies specific modules and hub genes related to subsyndromal symptomatic depression. *World J Biol Psychiatry*:1-23. doi:10.1080/15622975.2018.1548782
5. Gysi DM, Voigt A, Fragoso TM, Almaas E, Nowick K (2018) wTO: an R package for computing weighted topological overlap and a consensus network with integrated visualization tool. *BMC Bioinformatics* 19 (1):392. doi:10.1186/s12859-018-2351-7
6. Aya K, Suzuki G, Suwabe K, Hobo T, Takahashi H, Shiono K, Yano K, Tsutsumi N, Nakazono M, Nagamura Y, Matsuoka M, Watanabe M (2011) Comprehensive network analysis of anther-expressed genes in rice by the combination of 33 laser microdissection and 143 spatiotemporal microarrays. *PLoS One* 6 (10):e26162. doi:10.1371/journal.pone.0026162
7. Obayashi T, Kinoshita K (2009) Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res* 16 (5):249-260. doi:10.1093/dnares/dsp016
8. Mutwil M, Ruprecht C, Giorgi FM, Bringmann M, Usadel B, Persson S (2009) Transcriptional wiring of cell wall-related genes in Arabidopsis. *Mol Plant* 2 (5):1015-1024. doi:10.1093/mp/ssp055
9. Lanczky A, Nagy A, Bottai G, Munkacsy G, Szabo A, Santarpia L, Gyorffy B (2016) miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat* 160 (3):439-446. doi:10.1007/s10549-016-4013-7
10. Huang R, Wallqvist A, Covell DG (2006) Targeting changes in cancer: assessing pathway stability by comparing pathway gene expression coherence levels in tumor and normal tissues. *Mol Cancer Ther* 5 (9):2417-2427. doi:10.1158/1535-7163.MCT-06-0239
11. Wang Y, Zhang Y, Huang Q, Li C (2018) Integrated bioinformatics analysis reveals key candidate genes and pathways in breast cancer. *Mol Med Rep* 17 (6):8091-8100. doi:10.3892/mmr.2018.8895
12. Olex AL, Turkett WH, Fetrow JS, Loeser RF (2014) Integration of gene expression data with network-based analysis to identify signaling and metabolic pathways regulated during the development of osteoarthritis. *Gene* 542 (1):38-45. doi:10.1016/j.gene.2014.03.022

13. Chen J, Crawford R, Xiao Y (2013) Vertical inhibition of the PI3K/Akt/mTOR pathway for the treatment of osteoarthritis. *J Cell Biochem* 114 (2):245-249. doi:10.1002/jcb.24362
14. Bansal R, Nakagawa S, Yazdani S, van Baarlen J, Venkatesh A, Koh AP, Song WM, Goossens N, Watanabe H, Beasley MB, Powell CA, Storm G, Kaminski N, van Goor H, Friedman SL, Hoshida Y, Prakash J (2017) Integrin alpha 11 in the regulation of the myofibroblast phenotype: implications for fibrotic diseases. *Exp Mol Med* 49 (11):e396. doi:10.1038/emm.2017.213
15. Nam K, Son SH, Oh S, Jeon D, Kim H, Noh DY, Kim S, Shin I (2017) Binding of galectin-1 to integrin beta1 potentiates drug resistance by promoting survivin expression in breast cancer cells. *Oncotarget* 8 (22):35804-35823. doi:10.18632/oncotarget.16208
16. Zhu CQ, Popova SN, Brown ER, Barsyte-Lovejoy D, Navab R, Shih W, Li M, Lu M, Jurisica I, Penn LZ, Gullberg D, Tsao MS (2007) Integrin alpha 11 regulates IGF2 expression in fibroblasts to enhance tumorigenicity of human non-small-cell lung cancer cells. *Proc Natl Acad Sci U S A* 104 (28):11754-11759. doi:10.1073/pnas.0703040104
17. Navab R, Strumpf D, To C, Pasko E, Kim KS, Park CJ, Hai J, Liu J, Jonkman J, Barczyk M, Bandarchi B, Wang YH, Venkat K, Ibrahimov E, Pham NA, Ng C, Radulovich N, Zhu CQ, Pintilie M, Wang D, Lu A, Jurisica I, Walker GC, Gullberg D, Tsao MS (2016) Integrin alpha11beta1 regulates cancer stromal stiffness and promotes tumorigenicity and metastasis in non-small cell lung cancer. *Oncogene* 35 (15):1899-1908. doi:10.1038/onc.2015.254
18. Pan Y, Liu G, Yuan Y, Zhao J, Yang Y, Li Y (2017) Analysis of differential gene expression profile identifies novel biomarkers for breast cancer. *Oncotarget* 8 (70):114613-114625. doi:10.18632/oncotarget.23061
19. Schittenhelm J, Schwab EI, Sperveslage J, Tatagiba M, Meyermann R, Fend F, Goodman SL, Sipos B (2013) Longitudinal expression analysis of alphav integrins in human gliomas reveals upregulation of integrin alphavbeta3 as a negative prognostic factor. *J Neuropathol Exp Neurol* 72 (3):194-210. doi:10.1097/NEN.0b013e3182851019
20. Schnittert J, Bansal R, Mardhian DF, van Baarlen J, Ostman A, Prakash J (2019) Integrin alpha11 in pancreatic stellate cells regulates tumor stroma interaction in pancreatic cancer. *FASEB J* 33 (5):6609-6621. doi:10.1096/fj.201802336R
21. Fregni G, Quinodoz M, Moller E, Vuille J, Galland S, Fusco C, Martin P, Letovanec I, Provero P, Rivolta C, Riggi N, Stamenkovic I (2018) Reciprocal modulation of mesenchymal stem cells and tumor cells promotes lung cancer metastasis. *EBioMedicine* 29:128-145. doi:10.1016/j.ebiom.2018.02.017
22. He P, Zhang Z, Liao W, Xu D, Fu M, Kang Y (2016) Screening of gene signatures for rheumatoid arthritis and osteoarthritis based on bioinformatics analysis. *Mol Med Rep* 14 (2):1587-1593. doi:10.3892/mmr.2016.5423
23. Steffens S, Schrader AJ, Vetter G, Eggers H, Blasig H, Becker J, Kuczyk MA, Serth J (2012) Fibronectin 1 protein expression in clear cell renal cell carcinoma. *Oncol Lett* 3 (4):787-790. doi:10.3892/ol.2012.566

24. Liu X, Ma Y, Yang W, Wu X, Jiang L, Chen X (2015) Identification of therapeutic targets for breast cancer using biological informatics methods. Mol Med Rep 12 (2):1789-1795. doi:10.3892/mmr.2015.3565
25. Yang X, Jia M, Li Z, Lu S, Qi X, Zhao B, Wang X, Rong Y, Shi J, Zhang Z, Xu W, Gao Y, Zhang S, Yu G (2014) Bioinformatics analysis of aggressive behavior of breast cancer via an integrated gene regulatory network. J Cancer Res Ther 10 (4):1013-1018. doi:10.4103/0973-1482.137971
26. Chen D, Bhat-Nakshatri P, Goswami C, Badve S, Nakshatri H (2013) ANT XR1, a stem cell-enriched functional biomarker, connects collagen signaling to cancer stem-like cells and metastasis in breast cancer. Cancer Res 73 (18):5821-5833. doi:10.1158/0008-5472.CAN-13-1080
27. Hu K, Olsen BR, Besschetnova TY (2017) Cell autonomous ANT XR1-mediated regulation of extracellular matrix components in primary fibroblasts. Matrix Biol 62:105-114. doi:10.1016/j.matbio.2016.12.002
28. Verma K, Gu J, Werner E (2011) Tumor endothelial marker 8 amplifies canonical Wnt signaling in blood vessels. PLoS One 6 (8):e22334. doi:10.1371/journal.pone.0022334
29. Chaudhary A, Hilton MB, Seaman S, Haines DC, Stevenson S, Lemotte PK, Tschantz WR, Zhang XM, Saha S, Fleming T, St Croix B (2012) TEM8/ANT XR1 blockade inhibits pathological angiogenesis and potentiates tumoricidal responses against multiple cancer types. Cancer Cell 21 (2):212-226. doi:10.1016/j.ccr.2012.01.004
30. Alcala S, Martinelli P, Hermann PC, Heeschen C, Sainz B, Jr. (2019) The Anthrax Toxin Receptor 1 (ANT XR1) Is Enriched in Pancreatic Cancer Stem Cells Derived from Primary Tumor Cultures. Stem Cells Int 2019:1378639. doi:10.1155/2019/1378639

Table

Table 1. UniProtKB keyword enrichment analysis of the comprehensive co-expression network using *ITGA11* as bait

	Gene Name	UniProtKB Keywords (-log10(P-value), Fold Enrichment)				
		Extracellular matrix (27.29, 26.95)	Secreted (22.65, 5.94)	Signal (20.22, 3.54)	Glycoprotein (17.48, 3.18)	Cell adhesion (11.05, 9.87)
	ADAM metallopeptidase domain 12		✓	✓	✓	✓
2	ADAM metallopeptidase with thrombospondin type 1 motif 12	✓	✓	✓	✓	
	ADAM metallopeptidase with thrombospondin type 1 motif 2	✓	✓	✓	✓	
	AE binding protein 1		✓	✓	✓	
	AXL receptor tyrosine kinase			✓	✓	
	C1q and tumor necrosis factor related protein 3		✓	✓	✓	
	C1q and tumor necrosis factor related protein 6		✓	✓	✓	
	G protein-coupled receptor 1				✓	
	HtrA serine peptidase 1		✓	✓		
	NADPH oxidase 4				✓	
	SPARC/osteonectin, cwcw and kazal like domains proteoglycan 1	✓	✓	✓	✓	
	ST6 beta-galactoside alpha-2,6-sialyltransferase 2				✓	
	TIMP metallopeptidase inhibitor 2		✓	✓		
	WNT1 inducible signaling pathway protein 1		✓	✓	✓	✓
	Anthrax toxin receptor 1			✓	✓	
	Asporin	✓	✓	✓	✓	
	Cadherin 11			✓	✓	✓
	Cell migration inducing hyaluronan binding protein		✓	✓	✓	
	Cerebral endothelial cell adhesion molecule			✓	✓	✓
	Chondroitin sulfate synthase 3				✓	
	Collagen triple helix repeat containing 1	✓	✓	✓	✓	
	Collagen type I alpha 1 chain	✓	✓	✓	✓	
	Collagen type I alpha 2 chain	✓	✓	✓	✓	
	Collagen type III alpha 1 chain	✓	✓	✓	✓	
	Collagen type V alpha 1 chain	✓	✓	✓		
	Collagen type V alpha 2 chain	✓	✓	✓	✓	
	Collagen type VI alpha 3 chain	✓	✓	✓	✓	✓
	Collagen type VIII alpha 1 chain	✓	✓	✓		✓
	Collagen type VIII alpha 2 chain	✓	✓	✓		✓
	Collagen type X alpha 1 chain	✓	✓	✓		
	Collagen type XI alpha 1 chain	✓	✓	✓	✓	

Collagen type XII alpha 1 chain	✓	✓	✓	✓	✓
Corin, serine peptidase		✓		✓	
Fibroblast activation protein alpha		✓		✓	✓
Fibronectin 1	✓	✓	✓	✓	✓
Fin bud initiation factor homolog		✓	✓	✓	
Glucoside xylosyltransferase 2				✓	
Glycosyltransferase 8 domain containing 2				✓	
Hydroxysteroid 17-beta dehydrogenase 6			✓	✓	
Inhibin beta A subunit		✓	✓	✓	
Integrin subunit alpha 11			✓	✓	✓
Integrin subunit beta like 1		✓	✓	✓	
Leucine rich repeat containing 15			✓	✓	
Lumican	✓	✓	✓	✓	
Lysyl oxidase like 1		✓	✓		
Lysyl oxidase		✓	✓	✓	
Matrix metalloproteinase 11	✓	✓	✓		
Matrix metalloproteinase 14			✓		
Matrix metalloproteinase 2	✓	✓	✓	✓	
Matrix remodeling associated 5		✓	✓	✓	
Neurotrimin			✓	✓	✓
Nidogen 2	✓	✓	✓	✓	✓
Osteomodulin	✓	✓	✓	✓	✓
Periostin	✓	✓	✓	✓	✓
Plexin C1			✓	✓	
Prolyl 4-hydroxylase subunit alpha 3			✓	✓	
Scavenger receptor class F member 2			✓	✓	✓
Secreted frizzled related protein 2		✓	✓		
Secreted protein acidic and cysteine rich	✓	✓	✓	✓	
Thrombospondin 2			✓	✓	✓
Tumor necrosis factor superfamily member 4				✓	
Versican	✓	✓	✓	✓	

* Genes directly linked by *ITGA11* in the co-expression network.

Supporting Figure Legend

Fig S1. Disease free survival analysis of the eight hub genes in breast cancer based on Kaplan Meier-plotter. The patients were stratified into high-level group and low-level group according to different expression ratio. (A) ADAMTS12. (B) CEMIP. (C) COL11A1. (D) CTHRC1. (E) ITGA11. (F) LOXL1. (G) LUM. (H) P4HA3.

Figures

Figure 1

EC analysis of the 1,028-microarray data of human breast cancers (BC). BP, MF, and CC represent the biological process, molecular function, and cellular component groups of GO, respectively. The total number of categories in each GO group is indicated in parentheses. In the three GO groups (BP, blue; CC, red; and MF, gray), the EC score was calculated for each GO category (only gene numbers between 20 and 200) and compared with that of random sampling (orange) to estimate the statistically significant level.

a

b

c

Figure 3

The GO term analysis of BP, MF, and CC for the breast cancer network genes. The GO analysis of BP (A), CC (B), and MF(C) for the genes in the breast cancer network was performed. The minus logarithm of the p-value (x-axis) indicates the significance of the gene set belonging to predefined categories under the 15,438-gene background. The y-axis represents each GO category. The gene number and EC value of the category in the subnetwork are shown in brackets.

Figure 4

Overall survival of the eight hub genes in breast cancer based on Kaplan Meier-plotter. The patients were stratified into high-level group and low-level group according to different expression ratio. (A) ADAMTS12. (B) CEMIP. (C) COL11A1. (D) CTHRC1. (E) ITGA11. (F) LOXL1. (G) LUM. (H) P4HA3.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Table.S1sampleInfor1028cases.xlsx](#)
- [FigS1.tiff](#)
- [TableS2ECscoresofBP.docx](#)
- [Table.S4.xlsx](#)
- [Table.S3.xlsx](#)