Identification of hub genes and important pathways in determination of breast cancer survival using bioinformatics approaches

Sepideh Dashti
Shaheed Beheshti University of Medical Sciences
Soudeh Ghafouri-Fard (s.ghafourifard@sbmu.ac.ir)
Shaheed Beheshti University of Medical Sciences https://orcid.org/0000-0002-0223-499X

Research article

Keywords: breast cancer, lncRNA, CCNA2, CDK1, MAD2L1 and CCNB2, bioinformatics

Posted Date: October 16th, 2019

DOI: https://doi.org/10.21203/rs.2.16078/v1

Abstract

Backgrounds Breast cancer is a highly heterogeneous disorder characterized by dysregulation of expression of numerous genes and cascades. The conventional pathologic classification of breast cancer is not sufficient for the prediction of breast cancer behavior and response to therapy.

Methods We have retrieved data of two microarray datasets (GSE65194 and GSE45827) from the NCBI Gene Expression Omnibus database (GEO). R package was used for identification of differentially expressed genes (DEGs), assessment of gene ontology (GO) and pathway enrichment evaluation. The DEGs were integrated to construct a protein-protein interaction (PPI) network. Next, hub genes were recognized using the Cytoscape software and lncRNA-mRNA co-expression analysis was performed to evaluate the potential roles of lncRNAs. The interactive information among DEGs and the PPI network was obtained using the STRING online database. Finally, the clinical importance of the obtained genes was assessed using Kaplan-Meier survival analysis.

Results After excluding the outliers from the GSE65194 and GSE45827 datasets and data normalization, 866 DEGs including 712 upregulated and 154 downregulated DEGs were detected between breast cancer and normal samples. Up-regulated DEGs were enriched in six pathways including 'Cell cycle', 'Oocyte meiosis' and 'Focal adhesion'. Down-regulated DEGs were enriched in five pathways including 'Peroxisome-proliferator-activated receptors (PPAR) signaling pathway', 'Metabolism of xenobiotics by cytochrome P450', 'Adipocytokine signaling pathway' and 'Cytokine-cytokine receptor interaction' pathways. CCNA2, CDK1, MAD2L1, and CCNB2 were significantly enriched in several biological pathways. These four genes showed strong expression in breast cancer samples as compared to normal breast tissue. We also identified 12 lncRNAs with a significant correlation with MAD2L1 and CCNB2 genes. MAD2L1, CCNA2, RAD51-AS1, and LINC01089 have the most prediction potential among all candidate hub genes.

Conclusion Our study offers a framework for recognition of the mRNA-lncRNA network in breast cancer and the detection of important pathways that could be used as therapeutic targets in this kind of cancer.

Background

Breast cancer is the second most frequent and the fifth cause of cancer-associated mortality [1]. This type of cancer is associated with dysregulation of several genes (including both coding and non-coding ones) and signaling pathways [2]. Breast cancer is a molecularly heterogeneous disorder which is classified to five subtypes including luminal A, luminal B, basal-like, HER2-enriched and normal-like. This classification is based on the presence/ abundance of estrogen receptor (ER), progesterone receptor (PR), HER2 and Ki67 [3, 4]. However, several recent studies have indicated significance of other genes and signaling pathways in determination of overall survival (OS) of patients [2, 5]. Among the recently appreciated genes in this regard are long non-coding RNAs (lncRNAs) [6]. These transcripts are involved in the regulation of fundamental cell survival pathways and have functional interactions with proteins and other non-coding RNAs that participate in the pathogenesis of breast cancer [7]. Identification of such networks is an important step towards design of targeted therapies in breast cancer.

In the current study, we have retrieved data of two microarray datasets (GSE65194 and GSE45827) from the NCBI Gene Expression Omnibus database (GEO). R package was used for identification of differentially expressed genes (DEGs), assessment of gene ontology (GO) and pathway enrichment evaluation. The DEGs were integrated to construct a protein–protein interaction (PPI) network. Next, hub genes were recognized using the Cytoscape software and lncRNA-mRNA co-expression analysis was performed to evaluate the potential roles of lncRNAs.

Methods

Figure 1 shows summary of the steps accomplished in bioinformatics strategy.

Gene expression profile data collection

Two gene expression profiles associated with breast cancer (GSE65194 and GSE45827) were obtained from the NCBI Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). A chip-based platform GPL570 (HG-U133_Plus_2) Affymetrix Human Genome U133 Plus 2.0 Array was applied for both datasets. The GSE65194 included 130 breast cancer samples and 11 normal breast tissue samples [8]. Similarly, the GSE45827 contained 130 tumor tissue specimen as well as 11 normal tissues samples [9].

Data Preprocessing and DEGs identification

All raw data files were subjected to quantile normalization and background correction using Robust Multichip Average (RMA) [10]. RMA is an effective tool in the affy Bioconductor package for both mRNA and lncRNA profiling data [11]. The linear models for microarray data (LIMMA) R package [12] in Bioconductor (http://www.bioconductor.org/) [13] were used to perform differentially expressed gene analysis (DEGA) between breast cancer and normal breast samples. The Student's t-test was applied and DEGs with false discovery rate (FDR) < 0.01 and a |log2FC (fold change)| > 2 were screened. We conducted a dimensional reduction analysis by performing Principal component analysis (PCA) [14] with the purpose of quality assessment using ggplot2 package of R software [15].

Functional enrichment analysis

To identify the role of DEGs in breast cancer, KEGG Pathway and GO function enrichment analysis in 3 functional ontologies namely biological process (BP), cellular component (CC) and molecular function (MF) were performed using the DAVID system (https://david.ncifcrf.gov/). The adjusted P < 0.05 was considered as statistically significant [16].

PPI Network Construction, cluster analysis and key gene identification

To predict interactive relationships among common DEGs encoding proteins, we constructed a PPI network using online STRING database (https://string-db.org/) [17]. The minimum interaction score > 0.4 was required to construct the PPI network. Cytoscape software version 3.7.1 (http://www.cytoscape.org/) was applied to visualize the PPI networks and analyze the hub genes [18]. We used Molecular COmplex DEtection (MCODE) algorithm (version 1.5.1) to find PPI subnetwork and the highly interconnected clusters within the PPI network. MCODE is a Cytoscape plug-in in which we set maximum depth = 100, node score = 0.2, and K-core = 2 as threshold parameters [19]. CytoHubba (version 1.6) [20] and CytoNCA (version 2.1.6) [21] are two other plug-in in which provide multiple algorithms to detect hub genes in the network. In addition, identified key genes were selected for additional expression analysis on 1104 cancer and 113 normal samples from the TCGA project in The Encyclopedia of RNA Interactomes (ENCORI) database (http://starbase.sysu.edu.cn/panCancer.php). Pearson correlation coefficient was assessed between hub genes. The correlation coefficients were also checked on TCGA dataset by using Gene Expression Profiling Interactive Analysis (GEPIA) database (http://gepia.cancer-pku.cn/).

Prediction of lncRNAs function

LncRNA-mRNA co-expression analysis was performed to evaluate the potential roles of lncRNAs. The full list of lncRNA genes with approved HUGO Gene Nomenclature Committee (HGNC) symbols was downloaded from (https://www.genenames.org/) [22]. The list of lncRNA gene names was compared to our dataset gene symbols and overlapped genes were chosen. Then, differentially expressed lncRNAs were selected according to (|logFC|) > 0.5 and the adjusted P-value < 0.01 cutoff criteria. The reason for application of easier selection criteria was the lower expression level of lncRNAs compared with mRNAs. Then, the Pearson correlation coefficient was calculated between the differentially expressed lncRNA and 2 key protein-coding genes that were obtained from the previous steps based on functional annotation and co-expression analysis (MAD2L1 and CCNA2) in our dataset. LncRNAs with correlation coefficients higher than 0.6 or lower than –0.6 were chosen as the lncRNAs that co-expressed with MAD2L1 and CCNA2.

Survival Analysis

Survival analysis was carried out on these candidate hub genes to check out their effects on breast cancer survival. Overall survival analysis was performed based on expression data from 6234 breast cancer patients by Kaplan Meier plotter (kmplot.com/) that can evaluate the effect of gene expression on survival in 21 cancer types [23]. The hazard ratio was calculated and the Pvalue was determined applying logrank tests.

Results

DEGs Screening

After removing the batch effects and performing data normalization, 866 DEGs including 712 upregulated and 154 downregulated DEGs were screened between breast cancer and normal samples from GSE65194 and GSE45827 according to |logFC| > 2 and FDR < 0.01 as cut-off criteria. We generated a MA plot to show the relationship between intensity and difference between tumor and normal data (Figure 2A). Moreover, to visualize the overall gene expression levels of the DEGs, a Volcano plot was created with log2 FC score and log10 P-values in R software (Figure 2B). The PCA plot was drawn to illustrate the spatial distribution of the samples. We found one sample from the normal group which is spatially far from other normal samples. As a consequence, we removed this sample (Figure 3A). Furthermore a heatmap was drawn to illustrate the correlation between samples (Figure 3B).

KEGG and GO enrichment analysis

To further examine the role of common DEGs in breast cancer, we performed GO and KEGG pathway enrichment analysis. We found 10 dysregulated pathways based on the adjusted P < 0.05. Up-regulated DEGs were enriched in six pathways including 'Cell cycle', 'Oocyte meiosis' and 'Focal adhesion'. Down-regulated DEGs were enriched in fivepathways including 'Peroxisome-proliferator-activated receptors (PPAR) signaling pathway', 'Metabolism of xenobiotics by cytochrome P450', 'Adipocytokine signaling pathway' and 'Cytokine-cytokine receptor interaction' pathways (Figure 4A). The results for each GO functional analysis are presented in Figure 4B, 4C, and 4D.

PPI Network Construction and Module analysis

The interactive information among DEGs and the PPI network was obtained using the STRING online database. Among the total of common DEGs, 866 DEGs (712 up-regulated and 154 downregulated) were filtered into the PPI network with 866 nodes and 10398 edges, at a combined score > 0.4. Finally, Genes with a combined score >0.9 were selected as key DEGs to be imported into Cytoscape. The Cytoscape software was applied to evaluate the interactive relationships between the candidate proteins. Afterward, two clusters consist of 65 nodes and 23 nodes were screened with a cut-off k-score = 12 depend on the MCODE scoring system (Figure 5).

The CytoNCA and the CytoHubba are two Cytoscape plug-in for centrality analysis and give us some insight into the most influential nodes or edges in a network. We ran CytoHubba application and extracted data from four calculations methods (EPC, MCC, MNC, and Stress). The top 100 nodes ranked by these four methods were selected. Moreover, four algorithms from CytoNCA application (Degree, Eigenvector, Betweenness, and Closeness) were employed and the top 100 nodes based upon these four approaches were obtained. Besides, a Venn diagram was created to identify the significant hub genes that are similar between all groups. Eventually, through overlapping analysis, we identified a list of 26 key genes most of them belonged to MCODE cluster 1. Since highly interconnected proteins in a network accumulate in a cluster, we chose only 20 genes from our list that belonged to cluster 1 (Table 1).

Key genes functional annotation and co-expression analysis

GO enrichment and KEGG pathway analysis on these 21 genes indicated that four pathways were enriched, including cell cycle, progesterone-mediated oocyte maturation, oocyte meiosis, and p53 signaling pathway. CCNA2, CDK1, MAD2L1, and CCNB2 were significantly enriched in some biological aspects such as

cell cycle, mitosis, nuclear division, M phase, cell cycle and progesterone-mediated oocyte maturation pathways. In particular, by checking the expression data of 1104 cancer and 113 normal samples from the TCGA project in ENCORI database, we found that these four genes showed strong expression in the breast cancer specimens as compared to their expression in normal breast tissue (Figure 6). Additionally, we calculated the Pearson correlation for these 20 candidate genes and found a strong and significant correlation between them. Interestingly, *CCNA2* and *MAD2L1* which are two important genes in the cell cycle pathway and some crucial biological processes related to cell division, were highly correlated genes with a correlation coefficient higher than 0.9 in our analysis (Figure 7A). Furthermore, these two genes correlation in TCGA dataset in the GEPIA database was consistent with our analysis (Figure 7B).

Identification of differentially expressed lncRNAs and co-expression analysis

After downloading the list of lncRNA genes from HGNC database, lncRNAs genes symbols were extracted from the GSE65194 and GSE45827. A total of 334 lncRNA probes were identified in these two datasets by using this approach. Finally, 159 lncRNAs probe ID with |logFC| > 0.5 and adjusted P value < 0.01 among 20 normal samples and 258 breast tissue samples were picked out. Among these lncRNAs, 69 lncRNAs were up-regulated and 90 lncRNAs were down-regulated in breast cancer. We calculated Pearson correlation coefficient between differentially expressed lncRNAs and MAD2L1 and CCNA2 based on their expression value. LncRNA with Pearson correlation coefficient ≥ 0.6 or ≤ -0.6 were selected as key lncRNA which co-expressed with MAD2L1 and CCNA2. Totally, 12 lncRNAs meet this criterion (Table 2).

Survival Analysis of candidate hub Genes

Associations between expression of candidate hub genes and OS of the breast cancer patients were evaluated using KM method to estimate the prognostic importance of the hub genes in our study. The results indicated that low expression of MAD2L1, CCNA2 and NCK1-DT lead to higher OS rate than high expression. Inversely, high expressions of MEG3, RAD51-AS1, PRINS, LINC01089, LINC02256, FUT8-AS1, LINC01279, CARMN, EPB41L4A-AS1, EIF3J-DT and TNFRSF14-AS1 result in a significantly longer OS time among patients with breast cancer. The results showed that MAD2L1, CCNA2, RAD51-AS1 and LINC01089 have the most prediction potential among all candidate hub genes (Table 2, Figure 8).

Discussion

In the present study, we used a bioinformatics strategy to identify key genes and signaling pathways in breast cancer pathogenesis with a focus on the role of lncRNAs and their interactions with proteins. Such interactions can be assessed using experimental approaches which are costly and laborious. Bioinformatics methods for such purpose fall into two groups: strategies that use sequence, structural data and physicochemical features, and methods that are based on network construction. The latter can provide the inherent characteristics of topological configuration of associated biological networks which is often disregarded by the former strategies [6]. In the present work, we used GPL570 which is a good platform to evaluate the functional roles of lncRNAs in tumorigenesis [11, 24]. We identified 712 upregulated and 154 downregulated DEGs between breast cancer and normal samples. Up-regulated DEGs were enriched in 'Cell cycle', 'Oocyte meiosis' and 'Focal adhesion'. A previous bioinformatics study using topological characteristics of genes in breast cancer has identified these pathways as hub subnetworks [25]. The role of these pathways has been acknowledged in the pathogenesis of another hormone related cancer namely prostate cancer [26]. We also detected down-regulated DEGs were enriched in 'PPAR signaling pathway', 'Metabolism of xenobiotics by cytochrome P450', 'Adipocytokine signaling pathway' and 'Cytokine-cytokine receptor interaction' pathways. PPARs are nuclear hormone receptors which participate in modulation of different aspects of tumorigenesis such as cell proliferation, survival and apoptosis [27]. Xenobiotic metabolizing enzymes are also involved in the tumorigenesis and response of cancer patients to therapeutic options. Integration of expression data of these genes with eQTL data and allele frequency data from the 1000 Genomes project has shown considerable inter-population differences in the related pathways which might influence cancer prognosis and response to treatment [28]. Adipocytokines can also influence cell proliferation and survival, and malignant phenotypes of breast cancer cells through regulation several cellular and molecular pathways thus aggravating survival of patients [29]. Cytokine signaling has important functions in formation, proliferation, and migration of breast cancer, thus modulating invasiveness, angiogenesis and metastatic potential of these cells [30].

Our *in silico* analyses revealed that CCNA2, CDK1, MAD2L1 and CCNB2 were significantly enriched in several biological pathways. These four genes showed strong expression in breast cancer samples as compared to their expression in normal breast tissue. Notably, these four genes have been among the top dysregulated genes in small cell lung cancer as revealed by GO, KEGG analysis and construction of PPI network [31]. Such similarity between these two different types of cancers implies fundamental role of these genes in the carcinogenesis process and potentiates them as therapeutic targets. MAD2L1 form a complex with the APC/C and CDC20 and subsequently stimulate the M-A checkpoint to halt the transition of cell at this stage in the presence of anomalous segregation of chromatin. Yet, over-expression of E2F1 in atypical cells affects the formation of the mentioned complex leading to cell cycle transition even in the presence of abnormal chromosomes [32]. CDK1/cyclin B is a maturation-promoting factor [33] and the checkpoint for G2/M transition [34, 35], so it is expected to be involved in the process of cell cycle regulation and tumorigenesis. We also identified 12 lncRNAs with significant correlation with MAD2L1 and CCNB2 genes. As expected from KEGG analysis, KM analysis indicated that low expression of MAD2L1, CCNA2 and NCK1-DT lead to higher OS rate than high expression. Inversely, high expressions of MEG3, RAD51-AS1, PRINS, LINC01089, LINC02256, FUT8-AS1, LINC01279, CARMN, EPB41L4A-AS1, EIF3J-DT and TNFRSF14-AS1 result in a significantly longer OS time among patients with breast cancer.

Conclusions

Our *in silico* method identified a number of hub genes and related lncRNAs which are possibly involved in the pathogenesis of breast cancer and patients' prognosis, so can be used as therapeutic targets or biomarkers for this malignancy.

Abbreviations

lncRNAs, Long non-coding RNAs

GEO: Gene Expression Omnibus;

GSE, GEO Series;

RMA: Robust Multichip Average;

LIMMA, Linear models for microarray data

DEG, Differentially expressed genes;

DEGA: Differentially expressed gene analysis;

KEGG: Kyoto Encyclopedia of Genes and Genomes;

GO: Gene ontology;

BP, Biological process;

CC, Cellular component;

MF, Molecular function;

PPI: Protein-protein interaction;

HGNC: The HUGO Gene Nomenclature Committee;

FDR: False discovery rate;

MA plot, Moving average plot;

FC, Fold change;

FDR, False discovery rate;

PCA, Principal component analysis;

STRING, Search tool for retrieval of interacting genes/proteins;

TCGA, *The Cancer Genome Atlas;*

ENCORI, The Encyclopedia of RNA Interactomes;

GEPIA, Gene Expression Profiling Interactive Analysis;

HR, Hazard ratio;

OS, Overall Survival;

KM, Kaplan Meier;

ER, Estrogen receptor;

PR, Progesterone receptor;

PPAR, Peroxisome-proliferator-activated receptors;

MAD2L1, Mitotic spindle assembly checkpoint protein;

CCNA2, Cyclin-A2;

CCNB1, Cyclin B1;

CDK1, Cyclin-dependent kinase 1;

eQTLs, *Expression quantitative trait loci;*

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets analyzed during the current study are available in the NCBI Gene Expression Omnibus database (GEO) repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE65194, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE45827]

Competing interests

The authors declare that they have no competing interests

Funding

The authors received no funding for this work.

Authors' contributions

SD conceived and designed the experiments, analyzed the data, prepared the figures and tables and drafted the work. SGF revised the work critically for important content, and was a major contributor in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

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: a cancer journal for clinicians* 2018, *68(*6):394–424.

2.Yang KD, Gao J, Luo M: *Identification of key pathways and hub genes in basal-like breast cancer using bioinformatics analysis. Oncotargets Ther* 2019, *12:*1319–1331.

3.Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al: Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology: official journal of the American Society of Clinical Oncology* 2009, *27(*8):1160–1167.

4.Prat A, Ellis MJ, Perou CM: *Practical implications of gene-expression-based assays for breast oncologists. Nature reviews Clinical oncology* 2012, *9(*1):48.

5.Feng YX, Spezia M, Huang SF, Yuan CF, Zeng ZY, Zhang LH, Ji XJ, Liu W, Huang B, Luo WP *et al: Breast cancer development and progression: Risk factors, cancer stem cells, signaling pathways, genomics, and molecular pathogenesis. Genes Dis* 2018, *5(*2):77–106.

6.Zhang H, Liang YC, Han SY, Peng C, Li Y: *Long Noncoding RNA and Protein Interactions: From Experimental Results to Computational Models Based on Network Methods. Int J Mol Sci* 2019, *20(*6).

7.Tuersong T, Li LL, Abulaiti Z, Feng SM: *Comprehensive analysis of the aberrantly expressed lncRNA-associated ceRNA network in breast cancer. Mol Med Rep* 2019, *19(*6):4697–4710.

8.Maire V, Némati F, Richardson M, Vincent-Salomon A, Tesson B, Rigaill G, Gravier E, Marty-Prouvost B, De Koning L, Lang G: *Polo-like kinase 1: a potential therapeutic option in combination with conventional chemotherapy for the management of patients with triple-negative breast cancer. Cancer research* 2013, *73(*2):813–823.

9.Gruosso T, Mieulet V, Cardon M, Bourachot B, Kieffer Y, Devun F, Dubois T, Dutreix M, Vincent-Salomon A, Miller KM: *Chronic oxidative stress promotes H2AX protein degradation and enhances chemosensitivity in breast cancer patients. EMBO molecular medicine* 2016, *8(*5):527–549.

10.Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: *Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics* 2003, *4(*2):249–264.

11.Zhang X, Sun S, Pu JKS, Tsang ACO, Lee D, Man VOY, Lui WM, Wong STS, Leung GKK: *Long non-coding RNA expression profiles predict clinical phenotypes in glioma. Neurobiology of disease* 2012, *48(*1):1–8.

12.Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: *limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research* 2015, *43(*7):e47-e47.

13.Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T: *Orchestrating high-throughput genomic analysis with Bioconductor. Nature methods* 2015, *12(*2):115.

14.Yeung KY, Ruzzo WL: *Principal component analysis for clustering gene expression data. Bioinformatics* 2001, *17(*9):763–774.

15.Wickham H: *ggplot2: elegant graphics for data analysis:* Springer; 2016.

16.Sherman BT, Lempicki RA: *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols* 2009, *4(*1):44–57.

17.Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P: *STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic acids research* 2018, *47(*D1):D607-D613.

18.Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: *Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research* 2003, *13(*11):2498–2504.

19.Bader GD, Hogue CW: *An automated method for finding molecular complexes in large protein interaction networks. BMC bioinformatics* 2003, *4(*1):2.

20.Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y: *cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC systems biology* 2014, *8(*4):S11.

21.Tang Y, Li M, Wang J, Pan Y, Wu F-X: *CytoNCA: a cytoscape plugin for centrality analysis and evaluation of protein interaction networks. Biosystems* 2015, *127*:67–72.

22.Braschi B, Denny P, Gray K, Jones T, Seal R, Tweedie S, Yates B, Bruford E: *Genenames. org: the HGNC and VGNC resources in 2019. Nucleic acids research* 2018, *47(*D1):D786-D792.

23.Nagy Á, Lánczky A, Menyhárt O, Győrffy B: *Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Scientific reports* 2018, *8(*1):9227.

24.Jiang L, Hong L, Yang W, Zhao Y, Tan A, Li Y: *Co-expression network analysis of the lncRNAs and mRNAs associated with cervical cancer progression. Archives of medical science: AMS* 2019, *15(*3):754.

25.Zhuang DY, Jiang L, He QQ, Zhou P, Yue T: *Identification of hub subnetwork based on topological features of genes in breast cancer. Int J Mol Med* 2015, *35(*3):664–674.

26.Fan ST, Liang ZM, Gao ZQ, Pan ZW, Han SJ, Liu XY, Zhao CL, Yang WW, Pan ZF, Feng WG: *Identification of the key genes and pathways in prostate cancer. Oncol Lett* 2018, *16(*5):6663–6669.

27.Gou Q, Gong X, Jin JH, Shi JJ, Hou YZ: *Peroxisome proliferator-activated receptors (PPARs) are potential drug targets for cancer therapy. Oncotarget* 2017, *8(*36):60704–60709.

28.Li Y, Steppi A, Zhou YD, Mao F, Miller PC, He MM, Zhao TT, Sun Q, Zhang JF: *Tumoral expression of drug and xenobiotic metabolizing enzymes in breast cancer patients of different ethnicities with implications to personalized medicine. Scientific Reports* 2017, *7.*

29.Li J, Han X: *Adipocytokines and breast cancer. Current problems in cancer* 2018, *42(*2):208–214.

30.Fasoulakis Z, Kolios G, Papamanolis V, Kontomanolis EN: *Interleukins Associated with Breast Cancer. Cureus* 2018, *10(*11).

31.Ni Z, Wang XT, Zhang TC, Li LL, Li JX: *Comprehensive analysis of differential expression profiles reveals potential biomarkers associated with the cell cycle and regulated by p53 in human small cell lung cancer. Exp Ther Med* 2018, *15(*4):3273–3282.

32.May KM, Paldi F, Hardwick KG: *Fission Yeast Apc15 Stabilizes MCC-Cdc20-APC/C Complexes, Ensuring Efficient Cdc20 Ubiquitination and Checkpoint Arrest. Current biology: CB* 2017, *27(*8):1221–1228.

33.Draetta G, Luca F, Westendorf J, Brizuela L, Ruderman J, Beach D: *Cdc2 protein kinase is complexed with both cyclin A and B: evidence for proteolytic inactivation of MPF. Cell* 1989, *56(*5):829–838.

34.Fisher D, Nurse P: *Cyclins of the fission yeast Schizosaccharomyces pombe. Seminars in cell biology* 1995, *6(*2):73–78.

35.Nasmyth K: *Viewpoint: putting the cell cycle in order. Science (New York, NY)* 1996, *274(*5293):1643–1645.

Tables

Table 1. Key differentially expressed genes acquired by centrality analysis.

			MCODE	Centrality	analysis	by	CytoNCA	Centrality	analysis	by	CytoHub
Gene	logFC	adj.P.Val	MCODE Score	Betweenness	Closeness	Degree	Eigenvector	EPC	MCC	MNC	Stress
CDK1	3.729327	3.03E-28	46.020339	6576.003525	0.551637	131	0.1437205	55.3	9.22E+13	158	35665
CCNB1	3.115417	1.59E-27	46.020339	3303.18478	0.534146	116	0.1417335	55.9	9.22E+13	133	21342
CCNA2	2.219921	2.57E-14	46.020339	2534.615585	0.52019	105	0.1342352	53.4	9.22E+13	122	16339
CDC20	2.756389	1.63E-14	46.020339	2758.724299	0.496036	103	0.1345755	51.8	9.22E+13	116	11604
MAD2L1	3.346366	4.29E-26	46.020339	1131.260005	0.48079	94	0.1333797	52.9	9.22E+13	110	72556
KIF11	3.207367	4.39E-28	46.020339	1464.21417	0.487751	92	0.1298102	50.4	9.22E+13	104	95940
CENPA	2.75168	8.64E-15	47.094949	1704.200531	0.481319	92	0.1282006	50.3	9.22E+13	100	74860
PCNA	2.453675	1.03E-37	43.857039	2862.973827	0.489385	92	0.1150848	47.9	9.22E+13	106	10826
EZH2	2.942149	1.48E-27	42.305272	3357.588136	0.511085	91	0.1046336	47.1	9.22E+13	112	24331
KIF23	2.687969	7.39E-20	46.020339	1657.240201	0.493799	89	0.1254939	49	9.22E+13	97	99424
TOP2A	4.595822	5.72E-29	46.020339	1122.678983	0.493243	88	0.1293264	51.4	9.22E+13	104	94808
UBE2C	3.062068	1.20E-22	46.020339	1455.353604	0.466951	88	0.1219277	48.8	9.22E+13	101	67434
BIRC5	2.792394	5.43E-15	46.020339	1678.399139	0.493799	88	0.1287475	48.9	9.22E+13	99	10828
KIF2C	2.389307	1.51E-15	46.94026	1182.010234	0.487751	88	0.1259678	48	9.22E+13	97	85072
RRM2	4.481084	1.30E-31	46.020339	1745.829126	0.497727	86	0.1228496	50.9	9.22E+13	101	11059
RACGAP1	2.400852	1.87E-23	46.020339	1361.211732	0.493799	86	0.1225891	49.8	9.22E+13	94	93686
KIF4A	2.206078	2.43E-13	46.94026	1008.406749	0.478689	80	0.1192849	46	9.22E+13	88	65850
KPNA2	3.428273	1.09E-42	46.880102	961.5679185	0.47454	78	0.111362	46.1	9.22E+13	86	91340
TYMS	2.878076	1.89E-21	47.774118	1340.60442	0.473514	76	0.1156782	48.1	9.22E+13	92	99082
RRM1	2.268661	2.13E-25	40.445411	1145.682149	0.45768	69	0.0981711	43.2	9.22E+13	80	69510

Abbreviations: logFC, log2 fold change; adj.P.Val, adjusted p-value; EPC, Edge Percolated Component; MCC, Maximal Clique Centrality; MNC, Maximum Neighborhood Component

Figures

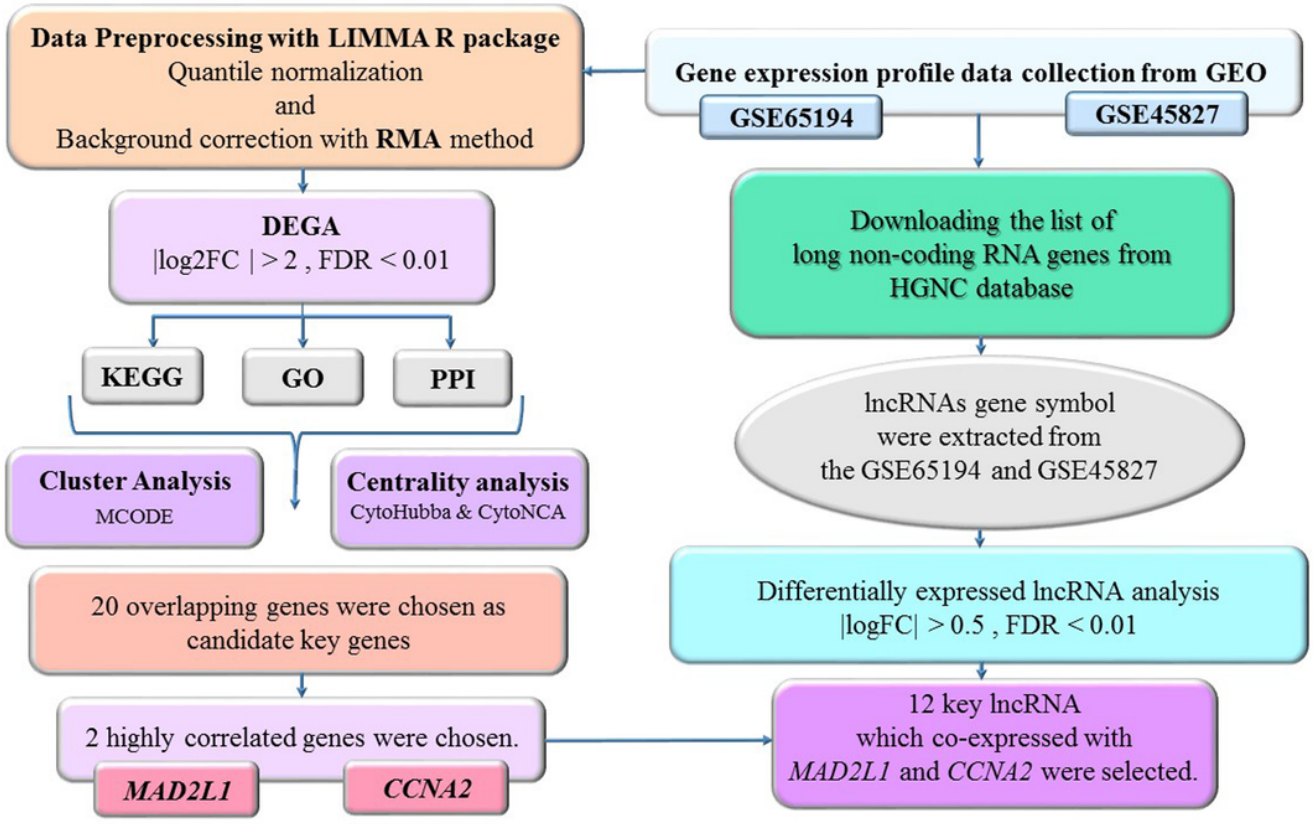

Figure 1

Study design flowchart. Abbreviations: GEO, Gene Expression Omnibus; RMA, Robust Multichip Average; DEGA, differentially expressed gene analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, gene ontology; PPI, protein-protein interaction; HGNC, The HUGO Gene Nomenclature Committee; FDR, false discovery rate

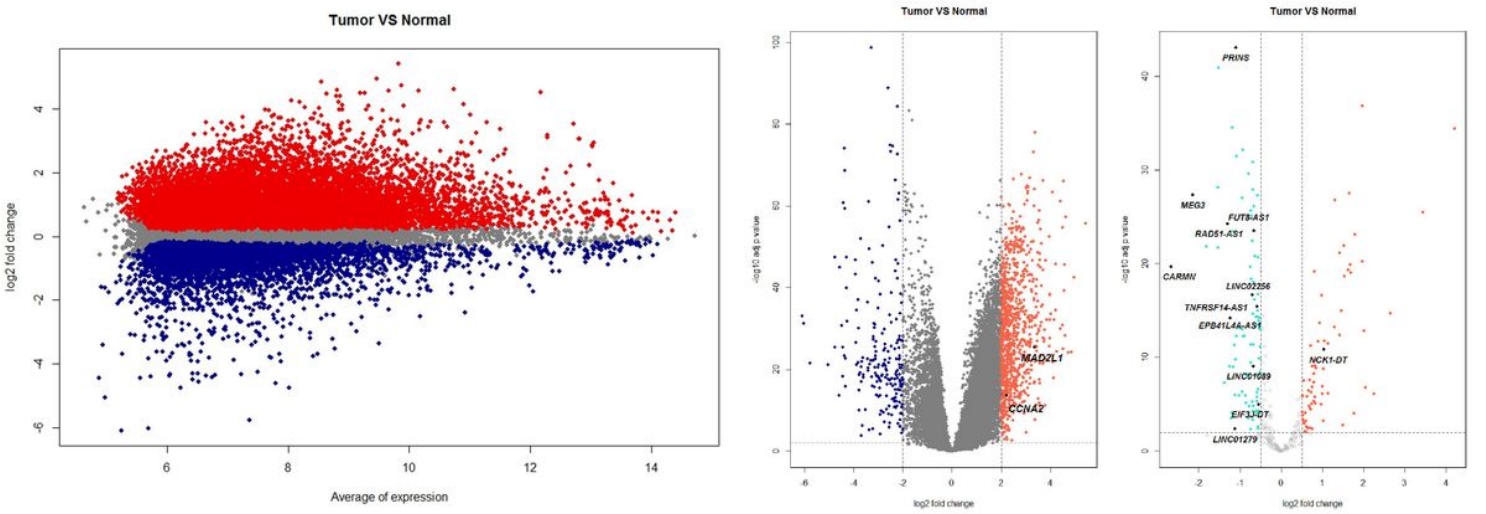

Figure 2

Differential expression between normal and tumor breast tissue. (A) MA plot to visualize the intensity of gene expression. Red and blue dots indicate all genes with significant up-regulation and significant down-regulation, respectively (FDR < 0.01). (B) Volcano plot of significant DEGs with |logFC| > 2. (C) Volcano plot of significant differentially expressed lncRNA with |logFC| > 0.5. Abbreviations: MA plot, moving average plot; DEG, differentially expressed genes; FC, fold change; FDR, false discovery rate

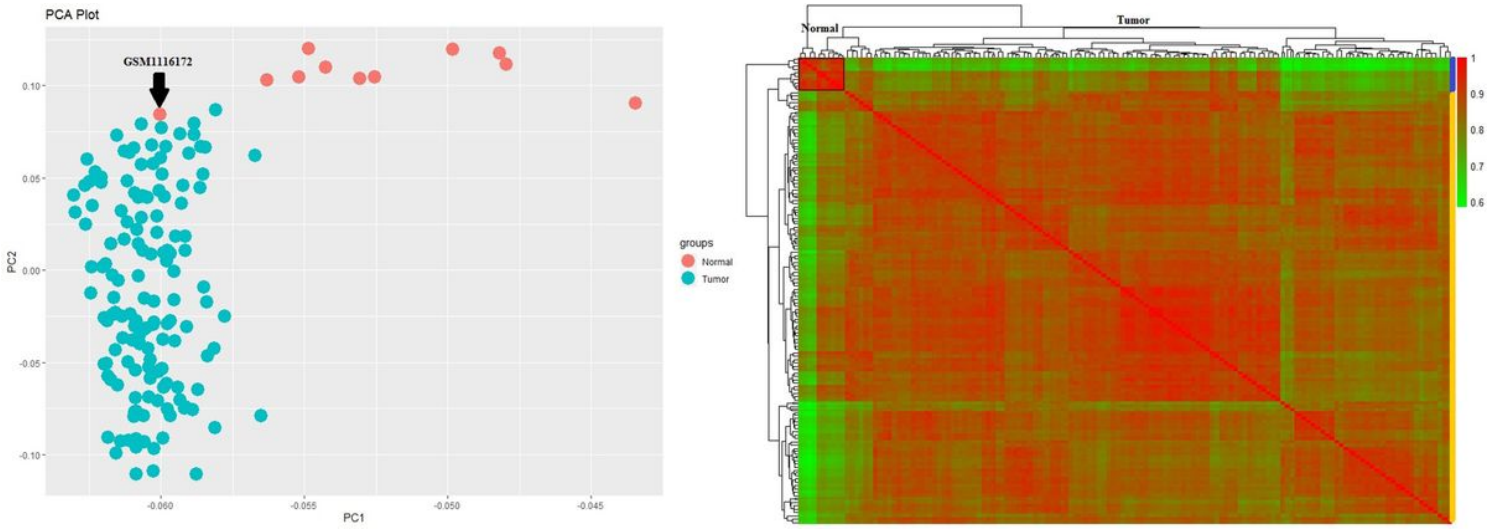

Figure 3

Sample enrichment visualization. (A) Principal component analysis (PCA) due to exploring the pattern of samples enrichment. GSM1116172 (a normal breast sample) was removed from further analysis in order to its wrong spatial enrichment. (B) Heatmap of DEGs for indicating the correlation between samples.

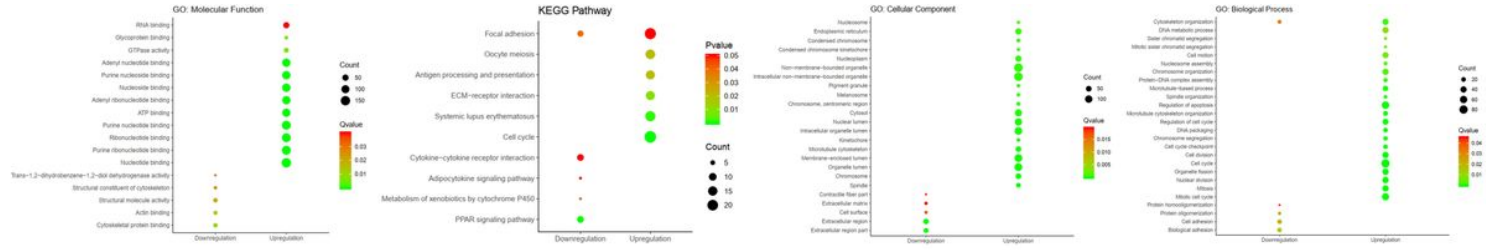

Figure 4

KEGG and GO enrichment analysis. (A) KEGG pathway. (B) GO for DEGs, Biological process. (C) GO for DEGs, Cellular component. (D) GO for DEGs, Molecular function. Abbreviations: KEGG, Kyoto encyclopedia of genes and genomes; GO, gene ontology; DEG, differentially expressed gene.

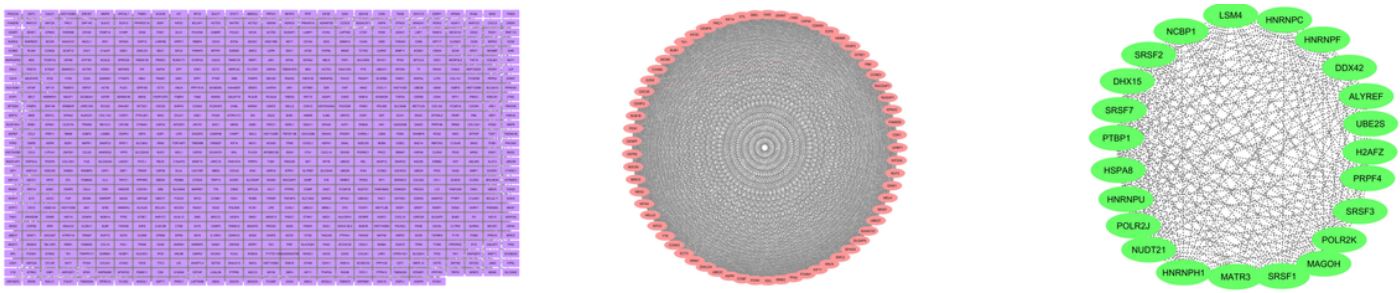

Figure 5

PPI Network Construction and Module analysis. (A) A PPI network with 866 nodes and 10398 edges using the STRING online database. (B) Cluster 1 containing 65 nodes and 1923 edges (C) cluster 2 containing 23 nodes and 206 edges. These 2 clusters had a cut-off k-score=12 depend on the MCODE scoring system. Abbreviations: PPI, protein–protein interaction; DEG, differentially expressed gene; STRING, search tool for retrieval of interacting genes/proteins;

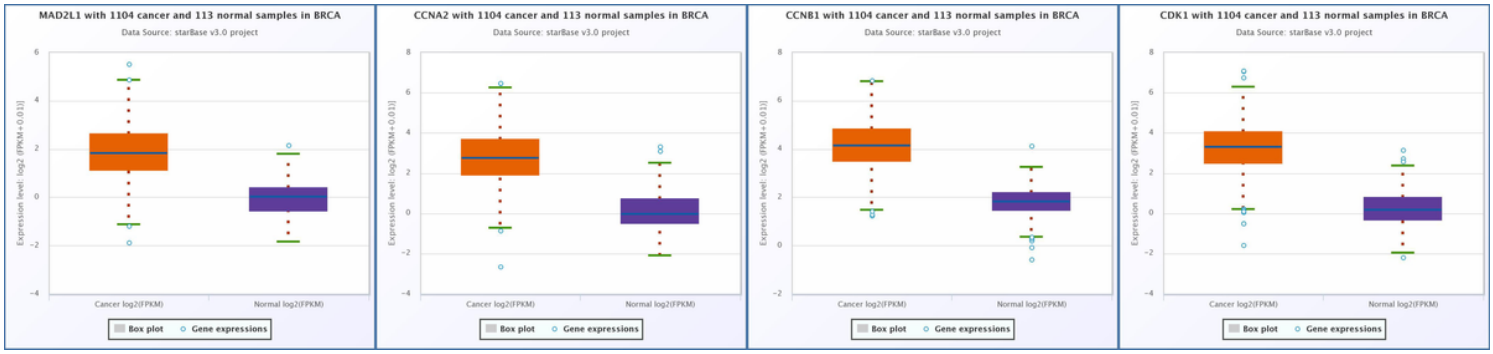

Figure 6

. Gene expression of 1104 cancer and 113 normal samples from the TCGA project in ENCORI database. (A) MAD2L1, (B) CCNA2, (C) CCNB1, (D) CDK1. Abbreviations: TCGA, The Cancer Genome Atlas; ENCORI, The Encyclopedia of RNA Interactomes; MAD2L1, mitotic spindle assembly checkpoint protein; CCNA2, Cyclin-A2; CCNB1, cyclin B1; CDK1, cyclin-dependent kinase 1.

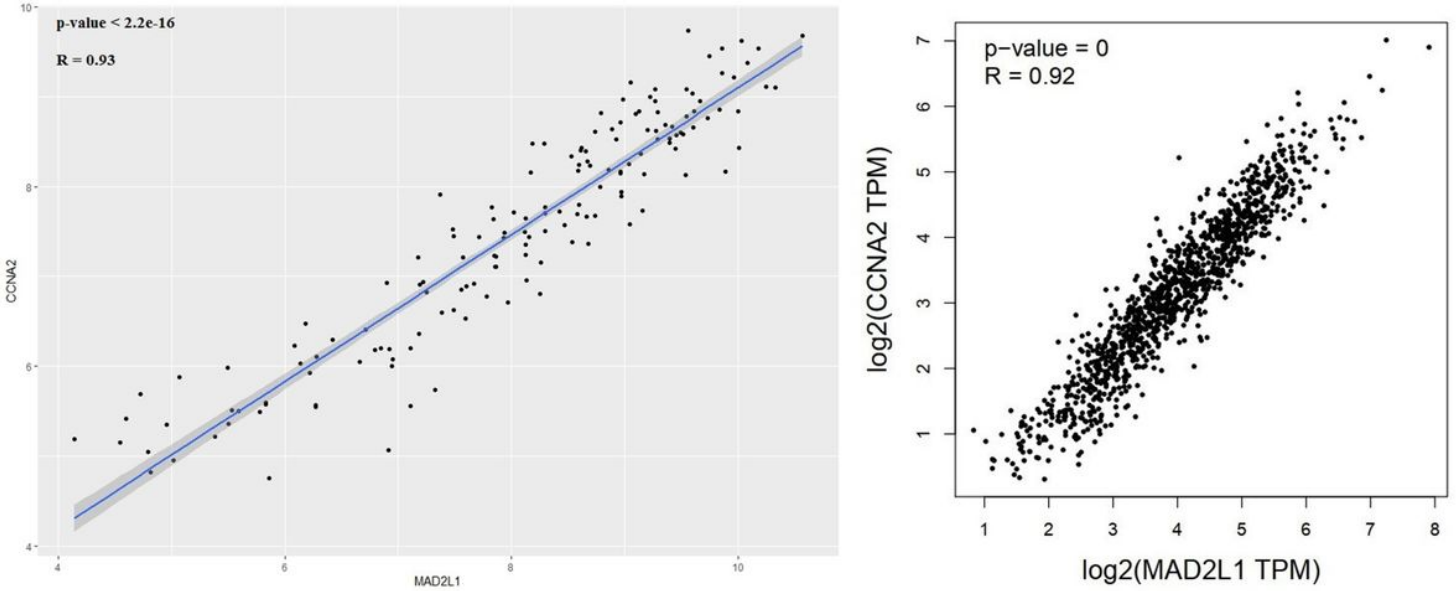

Figure 7

Pearson correlation coefficient analysis between MAD2L1 and CCNA2. (A) Based on GSE65194 and GSE45827, (B) Based on TCGA dataset in the GEPIA database. Abbreviations: R, Pearson's correlation coefficient; GEPIA, Gene Expression Profiling Interactive Analysis.

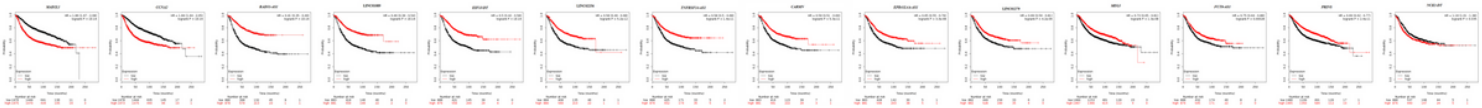

Figure 8

Survival analysis of candidate key genes. HR, hazard ratio.