Identification of a Lipid Metabolism-Associated Gene Signature Predicting Survival in Breast Cancer

Mingkai Gong
The Affiliated Hospital of Qingdao University
Xiangping Liu
The Affiliated Hospital of Qingdao University
Wu Yang
The Affiliated Hospital of Qingdao University
Hongming Song
The Affiliated Hospital of Qingdao University
Xian Zhao
The Affiliated Hospital of Qingdao University
Xiancheng Ai
The Affiliated Hospital of Qingdao University
Shuo Wang
The Affiliated Hospital of Qingdao University
Haibo Wang (hbwang66@126.com)
The Affiliated Hospital of Qingdao University

Research

Keywords: Bioinformatics, lipid metabolism associated genes, breast cancer, Lipid metabolism, TCGA

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-910589/v1

Version of Record: A version of this preprint was published at International Journal of General Medicine on December 1st, 2021. See the published version at https://doi.org/10.2147/IJGM.S343426.

Abstract

Background: Cancer metabolism and specifically lipid metabolism play an important role in breast cancer (BC) progression and metastasis. However, the role of lipid metabolism-associated genes (LMGs) in the diagnosis of breast cancer remains unknown.

Methods: The expression profiles and clinical follow-up information of BC were downloaded from The Cancer Genome Atlas (TCGA), and metabolic genes were downloaded from the Gene Set Enrichment Analysis (GSEA) dataset. Univariate cox regression and least absolute shrinkage and selection operator (LASSO) regression analyses were performed on the differentially expressed metabolism-related genes. Then, the formula of the metabolism-related risk model was composed, and the risk score of each patient was calculated. The breast cancer patients were divided into high-risk and low-risk groups with a cutoff of the median expression value of the risk score, and the prognostic analysis was also used to analyze the survival time between these two groups. Finally, we analyzed the expression, interaction and correlation among the lipid metabolism-associated genes risk model.

Results: The results from the prognostic analysis indicated that the survival was significantly poorer in the high-risk group than in the low-risk group in TCGA, single-sample gene set enrichment analysis (ssGSEA) shows it is plausible that lipid metabolism is highly correlated with tumor immunity.

Conclusion: Lipid metabolism-associated genes may become a new prognostic indicator predicting the survival of BC patients. The prognostic genes (n=16) may help provide new strategies for tumor therapy.

Background

Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer. Despite the dramatic improvement in breast cancer prognosis due to recent therapeutic advances such as more effective adjuvant and neo-adjuvant chemotherapies, together with more radical and safer surgery, advances in early diagnosis and treatment over the past decades, breast cancer prognosis is still very poor and the death rate is as high as 6.9%.[1] The exploration of potential biomarkers and regulatory mechanisms for early diagnosis and therapeutic targets of BC has important scientific significance and application values.

Cancer is basically a disease of abnormal cell growth and proliferation, in order to meet its needs, cells synthesize nucleic acids, proteins and lipids at an accelerated rate.[2] Unlimited cell proliferation is one of the most important characteristics of tumor cells, reprogrammed metabolism is considered a hallmark of cancer.[3, 4] Increased cancer cell proliferation requires the rapid synthesis of lipids for the generation of biological membranes and lipid provide cancer cells with energy during times of nutrient depletion, lipids are also important signalling molecules.[2, 5] The synthesis of lipids in healthy organisms is tightly controlled and responds to the nutrient's status of the cell. However, many human cancers display aberrant lipid metabolism.[5]

Next generation sequencing capabilities are revolutionizing approaches to all fields of medicine,recent advances in our understanding of tumor biology have uncovered a growing list of clinically relevant biomarkers.[6] However, the vast majority of studies have concentrated mainly on a single gene, and its predictive ability is insufficient compared with multiple biomarker-based models.[7] In clinical practice, the more accurately a patient's OS stage can be predicted, the sooner the clinician can make the clinical decision.

Lipid decomposition and anabolism of tumor cells, which are different from normal cells,have attracted more and more attention,and are also a new hotspot of tumor targeted therapy. In this study, we established a robust 16-gene prognostic signature for breast cancer by integrating TCGA datasets, which might complement classical clinical prognostic characteristics, and further aid clinicians in personalized treatment planning.

Results

Identification of prognostic LMGs in the breast TCGA cohort

Among 193 LMGs, 116 genes were significantly differentially expressed between BC samples and adjacent breast tissue samples (FDR < 0.05), 21 genes were statistically significantly related to OS in the univariate Cox regression model. Finally, a total of 18 intersection genes were selected as hub genes for further analyses (Fig. 2A-C). The PPI network of removing free nodes and Correlations between expression levels of genes are shown in (Fig. 2D,E). Mutational information of these 18 genes showed that amplifications, deep deletion were the most frequent mutation types (Fig. 2F). Three LMGs had mutation rates ≥ 4%, and LYPLA1 had the highest mutation rate.

Construction of a prognostic LMGs signature

According to LASSO, the optimal tuning parameter λ was identified based on a 1-SE (standard error) standard. As a result, 16 prognosis-related key LMGs were identified and integrated to construct a prognostic signature for BC (Fig. 3A,B). Figure 3C show the risk score distribution of patients and the survival status of patients in TCGA database. As we can see from PCA and t-SNE mappings (Fig. 3D,E), patients formed two distinct clusters. As the risk score increased, the patients' survival time decreased, and the death risk increased.patients in the high-risk group were more likely to die earlier than those in the low-risk group (Fig. 3F). The results from the Kaplan-Meier plot indicated that the survival was significantly poorer in the high-risk group than in the low-risk group (Fig. 3G). The prognostic efficiency of the 16-gene signature was assessed by time-dependent ROC curves. The area under the curve (AUC) values were 0.704 (1-year), 0.740 (3-year), and 0.728 (5-year) (Fig. 3H).

Independent prognostic value of the 16-gene signature

Samples have been chosen with well established clinical data. Among the 867 patients, we explored risk stratification for patients with different clinicopathologic factors, including risk parameters, age, histological grade and clinical stage. These variables indicated significant differences in univariate analysis and age, the stage, N classification, M classification showed significant differences in multivariable analysis. The risk score was significantly related to overall survival (OS) both in the univariate and multivariate Cox regression analyses (Fig. 4A,B and Table 1).

Table 1
Univariable and multivariable analyses for each clinical feature

Clinical feature	Number	Univariate Analysis			Multivariate Analysis		
		HR	95% CI	P value	HR	95% CI	P value
Risk Parameter(high-risk/low-risk)	427/419	2.979	2.287–3.881	< 0.001	2.521	1.927-3.300	< 0.001
Age(< 65/≥65)	626/241	2.270	1.568–3.287	< 0.001	2.386	1.632–3.487	< 0.001
Stage(I-II/III-IV)	659/208	2.478	1.719–3.572	< 0.001	1.876	1.087–3.236	0.024
T(I-II/III-IV)	741/126	1.667	1.100–2.530	0.016	0.923	0.552–1.545	0.762
N(0/1–3)	421/446	2.235	1.521–3.285	< 0.001	1.472	0.921–2.353	0.106
M(0/1–3)	851/16	6.121	3.361–11.151	< 0.001	2.343	1.183–4.638	0.015

Functional Enrichment Analyses

We performed GO and KEGG functional enrichment analyses on risk-related DEGs to analyze the relationship between biological functions and risk score. The results indicated that the 16 prognostic biomarkers mainly focused on immunoglobulin complex, lymphocyte mediated immunity and immune response-activating (Fig. 5A). KEGG functional enrichment analysis suggested that the 16 prognostic biomarkers were mainly related to Hematopoietic cell lineage, Cytokine-cytokine receptor interaction (Fig. 5B). To further explore relationship between the BC prognosis and immune status, we further used ssGSEA to quantify the infiltrating scores of immune cell and immunity-related functions. The enrichment score of aDCs, DCs, iDCs, Bcells, CD8 + Tcells, T helper cells, NK cells, TIL were higher in the low-risk group than in the high-risk group (adjusted p value < 0.05, Fig. 5C). Meanwhile, low-risk group had a higher score of CCR, HLA, check-point, cytolytic activity, APC co-inhibition, inflammation promotion, para inflammation, T cell co-stimulation, T cell co-inhibition, and type II INF response (adjusted p value < 0.05, Fig. 5D).

Expression levels of key genes in the clinical samples

The levels of the 16 genes in the BC and paired adjacent normal tissues of TCGA were compared to explore the clinical significance of the signature.The result showed that the mRNA expression levels were presented in Fig. 6A-P.The highly expressed genes included AGPAT1,PCYT2, DGKZ,PLPP2,CEL,LYPLA1,GK,PLA2G2D,and low expression genes have ENPP6,ADPRM,PAFAH1B1,SGMS2,CHPT1,SGPP1,CHKB,PCYT1A.

Discussion

Previous studies have developed a number of tumour molecular signatures for breast cancer.[6] The assessment of estrogen receptor, progesterone receptor, and human epidermal growth factor receptor 2 is mandatory for all invasive breast tumors, and have been widely used in clinical practice.[6] But outcomes of studying patients based on a single gene lack of accuracy. Lack of effective and reliable prognostic biomarkers still remains a major problem to improve the clinical outcomes of breast cancer patients. The relationship between lipid metabolism and cancer is multifaceted. Recent studies have shown that reprogramming of cellular lipid metabolism directly leads to malignant transformation and progression.[8, 9] It broadened our understanding of how lipid metabolism was relevant to cancer biology. To our knowledge, this study is the first to use LMGs to predict the prognosis of BC. In this study, we comprehensively analyzed 193 LMGs in BC samples and their relationships with prognosis. In functional analyses, these genes were significantly enriched in several immune cell types and immune-related pathways.

We identified 116 DEGs that were significantly differentially expressed between BC samples

and adjacent non-cancer samples, supporting the general importance of lipid metabolism in the pathogenesis and progression of BC. Moreover, we found 18 of these genes were related to OS in a univariate Cox regression analysis. These results demonstrated the importance of studying LMGs in breast tumors. The combination of univariate analysis and LASSO Cox regression was conducted to screen genes to indicate either poor or good prognosis and to construct a robust gene signature, which has been used widely in studies.[10–12] In the result, we established a new lipid-related 16-gene signature. According to the results of the ROC curve, the AUC of the risk score established by 16 LMGs was more than 0.7. This result showed that the risk score established by these 16 LMGs had a high value for predicting prognosis.

In recent years, the tumor microenvironment (TME) has been a research hotspot. During BC progression, tumor immune microenvironment remodeling with the changing of the ratio of immune cells and releasing of multiple immune inhibitory and reactive cytokines is a critical feature.[13, 14] The role of the tumor microenvironment (TME) in breast cancer immunomodulation is vitally important,[15] a better understanding of the immune cell infiltrate in the breast cancer microenvironment is crucial for the development of more effective therapeutic approaches.[16] By analyzing the pathways and functions of genes enriched in differences between the high and low risk groups, we observed a strong association with immune function. Given the above situation, we intended to unearth more information about the

immunological characteristics of the LMGs signature. As excepted, patients in the low-risk groups had higher fractions of CD8 + T cells and NK cells. CD8 + T cells are the main kind of cytolytic lymphocytes in tumors, while NK cells, a type of cytotoxic lymphocytes, are crucial constituents of the innate immune system and play a key role in immune surveillance.[15] This phenomenon may indicate that CD8 + T cells and NK cells are correlated with a favorable prognosis in tumor due to their ability to target and kill tumor cells, consistent with previous studies.[17–21] PD-1 and PD-L1 constitute an essential inhibitory mechanism which causes T cell exhaustion. That's the main reason why PD-L1 has drawn increasing attention of researchers concerned.[22–24] There have been many studies of lipid metabolism effects on tumor immunity. Recently, a potent sphingolipid metabolite regulated tumorigenesis, and responsed to chemotherapy and immunotherapy by affecting the trafficking, differentiation or effector function of tumor-infiltrating immune cells (TIICs), and implicated in many processes that are important for BC.[25] But the potential mechanism of LMGs and tumor immune microenvironment in BC prognosis remains unknown, and requires further investigation.

Conclusions

In summary, we established a robust 16-gene prognostic signature for breast cancer by comprehensive mining of the transcriptional profiles. Our study provides a new understanding of the prognostic value of tumor progress in BC and will benefit the prognosis assessment of BC patients. Targeting lipid metabolism in cancer cells could have therapeutic benefits. However, there are several limitations to our study. First, the present research is a retrospective study based on TCGA databases. Therefore, prospective clinical research should be performed to validate the application of this model. Secondly, the underlying mechanisms of lipid metabolism-associated genes on BC, and its relation to tumor immune status remain relatively enigmatic and warrant further investigation.

Material And Methods

Data collection

The RNA sequencing (RNA-seq) expression data and clinicopathological information of female breast cancer patients from 1053 breast cancer tissue samples and 111 nontumor tissue samples were downloaded from the TCGA BRCA dataset (https://portal.gdc.cancer.gov/). Probes were transformed to corresponding Entrez gene names referring to the annotation files. In addition, 193 genes in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway associated with metabolism were also extracted from the “c2.cp.kegg.v7.0.symbols” gene sets in GSEA platform (https://www.gsea-msigdb.org/gsea/downloads.jsp). Pathways including glycerolipid metabolism, glycerophospholipid metabolism, ether lipid metabolism and sphingolipid metabolism, genes in datasets were extracted for further analysis. The detailed flow-process diagram of this study is shown in Fig. 1. The study was conducted in accordance with the Declaration of Helsinki(as revised in 2013).

Construction of a lipid-metabolism model

First, the expression level of lipid metabolism genes was extracted from the total gene expression list. If a gene appeared more than once in the same sample, the limma of Bioconductor R package was utilized for averaging operations.[26] Second the limma was utilized to identify differentially expressed genes (DEGs) between breast cancer tissue samples and adjacent noncancerous breast tissue samples. The false discovery rate (FDR) threshold was set at $p < 0.05$ for DEG calling. To establish the lipid metabolism-related risk model, univariate Cox regression analysis was performed on the differentially expressed lipid metabolism-related genes. A total of 18 prognostic related differential genes were obtained by the intersection of DEGs and prognostic genes,The mutation rates of prognostic LMGs were analyzed by the cBioPortal for Cancer Genomics online website (https://www.cbioportal.org).[27] The OncoPrint schematic was constructed in cBioPortal (TCGA provisional) to directly reflect the mutation of 817 BRCA patients.

A protein–protein interaction (PPI) network of proteins encoded by all overlapping DEGs with prognostic value was visualized using String (http://string-db.org).[28]

To avoid overfitting, the least absolute shrinkage and selection operator (LASSO) was utilized to select variables with high prognostic value.[29] Next, 1000 LASSO iterations were performed for prognostic model construction using the 'glmnet' package in R.and their regression coefficients were obtained. Finally, the formula of the risk score was composed as follows, and risk scores were computed: Risk score = $\sum ni = \sum Coefi \times xi$, where xi represents the normalized expression level of target gene i and Coefi represents the regression coefficient. According to the median risk score in TCGA dataset, 1014 patients in the data set were divided into high-risk and low-risk groups after samples with a survival time of zero were removed. The Kaplan-Meier plot were used to evaluate survival differences between the high- and low-risk groups. Receiver operator characteristic curves (ROC) and area under the curves (AUC) were used to evaluate the availability of the prognosis model via the 'survivalROC' R package. To analyze the distribution differences between different groups, PCA was performed using the 'prcomp' function in the STATS package in R. A t-SNE analysis was implemented using the R package Rtsne (https://github.com/jkrijthe/Rtsne).

Univariate and Multivariate Cox Regression Analysis

Univariate cox regression analysis was presented for assessment of the prognostic values of the risk score and clinical features (age, stage, T classification, N classification, M classification). Then, multivariate cox regression analysis was used to determine which prognostic factors could independently predict the survival of patients.

Functional enrichment and pathway analysis

BC patients were divided into the high- and low-risk groups based on the median risk score,Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses for all selected DEGs between the two risk,[30, 31] cohorts were performed with the 'clusterProfiler' package in BioConductor using $|log2FC| \geq 1$ and $FDR < 0.05$ as thresholds. Finally, we determined the scores of 16 tumor-infiltrating immune cells and 13 immune-related functions for samples by ssGSEA.

Statistical analysis

We used R software (version 3.6.2) to perform all statistical analyses. The Student's t-test was used to compare gene expression levels between BC samples and non-cancer samples. Heatmaps of the LASSO analysis genes were plotted using the 'heatmap' R package. The OS for the two risk groups was evaluated by Kaplan-Meier (KM) survival curves and log-rank test. The 'survivalROC' package in R software was used to plot the ROC curve and calculate the area under the ROC curve (AUC). Bioconductor R package 'GSVA' was used to compare ssGSEA enrichment scores for immune cells and immune-related pathways between the two groups (i.e. high- or low-risk groups).[32] Unless otherwise stated, $p < 0.05$ was considered statistically significant.

Abbreviations

BC breast cancer;

TCGA The Cancer Genome Atlas

LASSO least absolute shrinkage and selection operator

ssGSEA single-sample gene set enrichment analysis

LMGs lipid metabolism-associated genes

FDR false discovery rate

Declarations

Acknowledgments

The authors gratefully acknowledge cBioPortal, GSEA platform and TCGA database, which made the data available.

Authors' contributions

This research was conducted in collaboration with all authors. MG and HW performed the data curation and analysis. WY, XL, HS and XZ analyzed and interpreted the results. XA and SW drafted and reviewed the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81572616 and 81772845) and the Natural Science Foundation of Shandong Province (No. ZR2017MH016).

Availability of data and materials

The raw data of this study are derived from cBioPortal (https://www.cbioportal.org), GSEA platform (https://www.gsea-msigdb.org/gsea/downloads.jsp) and TCGA database (https://portal.gdc.cancer.gov/), which are publicly available databases.

Declarations

Ethics approval and consent to participate

Not necessary.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Footnotes

Publisher's Note

References

1. Sung H, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209–49.
2. Baenke F, et al. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. Dis Model Mech. 2013;6(6):1353–63.
3. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74.
4. Pavlova NN, Thompson CB. *The Emerging Hallmarks of Cancer Metabolism* Cell Metab. 2016;23(1):27–47.
5. Guerra B, Issinger OG. *Role of Protein Kinase CK2 in Aberrant Lipid Metabolism in Cancer*. Pharmaceuticals (Basel), 2020. 13(10).
6. Giridhar KV, Liu MC. Available and emerging molecular markers in the clinical management of breast cancer. Expert Rev Mol Diagn. 2019;19(10):919–28.
7. Wang G, et al. The prediction of survival in Gastric Cancer based on a Robust 13-Gene Signature. J Cancer. 2021;12(11):3344–53.
8. Beloribi-Djefaflia S, Vasseur S, Guillaumond F. Lipid metabolic reprogramming in cancer cells. Oncogenesis. 2016;5(1):e189.

9. DeBerardinis RJ, Chandel NS. *Fundamentals of cancer metabolism* Sci Adv. 2016;2(5):e1600200.
10. Jiang P, et al. The construction and analysis of a ferroptosis-related gene prognostic signature for pancreatic cancer. Aging. 2021;13(7):10396–414.
11. Wang S, et al. Identification of a ferroptosis-related gene signature (FRGS) for predicting clinical outcome in lung adenocarcinoma. PeerJ. 2021;9:e11233.
12. Ge M, et al. A Ferroptosis-Related Signature Robustly Predicts Clinical Outcomes and Associates With Immune Microenvironment for Thyroid Cancer. Front Med (Lausanne). 2021;8:637743.
13. Deng W, et al. Recombinant Listeria promotes tumor rejection by CD8(+) T cell-dependent remodeling of the tumor microenvironment. Proc Natl Acad Sci U S A. 2018;115(32):8179–84.
14. Marra A, Viale G, Curigliano G. Recent advances in triple negative breast cancer: the immunotherapy era. BMC Med. 2019;17(1):90.
15. Zheng H, et al., *Tumor Microenvironment: Key Players in Triple Negative Breast Cancer Immunomodulation*. Cancers (Basel), 2021. 13(13).
16. Nascimento C, Ferreira F. Tumor microenvironment of human breast cancer, and feline mammary carcinoma as a potential study model. Biochim Biophys Acta Rev Cancer. 2021;1876(1):188587.
17. Ruth JH, et al., *CD6 is a target for cancer immunotherapy*. JCI Insight, 2021. 6(5).
18. Shan CK, et al. Pingyangmycin enhances the antitumor efficacy of anti-PD-1 therapy associated with tumor-infiltrating CD8(+) T cell augmentation. Cancer Chemother Pharmacol. 2021;87(3):425–36.
19. Buque A, et al. NK cells beat T cells at early breast cancer control. Oncoimmunology. 2020;9(1):1806010.
20. Levy EM, Roberti MP, Mordoh J. Natural killer cells in human cancer: from biological functions to clinical applications. J Biomed Biotechnol. 2011;2011:676198.
21. Ali HR, et al. Association between CD8 + T-cell infiltration and breast cancer survival in 12,439 patients. Ann Oncol. 2014;25(8):1536–43.
22. McDermott DF, Atkins MB. PD-1 as a potential target in cancer therapy. Cancer Med. 2013;2(5):662–73.
23. Duraiswamy J, et al. Dual blockade of PD-1 and CTLA-4 combined with tumor vaccine effectively restores T-cell rejection function in tumors. Cancer Res. 2013;73(12):3591–603.
24. Kamphorst AO, et al. Proliferation of PD-1 + CD8 T cells in peripheral blood after PD-1-targeted therapy in lung cancer patients. Proc Natl Acad Sci U S A. 2017;114(19):4993–8.
25. Nema R, Kumar A. Sphingosine-1-Phosphate Catabolizing Enzymes Predict Better Prognosis in Triple-Negative Breast Cancer Patients and Correlates With Tumor-Infiltrating Immune Cells. Front Mol Biosci. 2021;8:697922.
26. Ritchie ME, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
27. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.

28. Szklarczyk D, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43(Database issue):D447-52.
29. Duan J, et al. Generalized LASSO with under-determined regularization matrices. Sig Process. 2016;127:239–46.
30. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25–9.
31. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
32. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

Figures

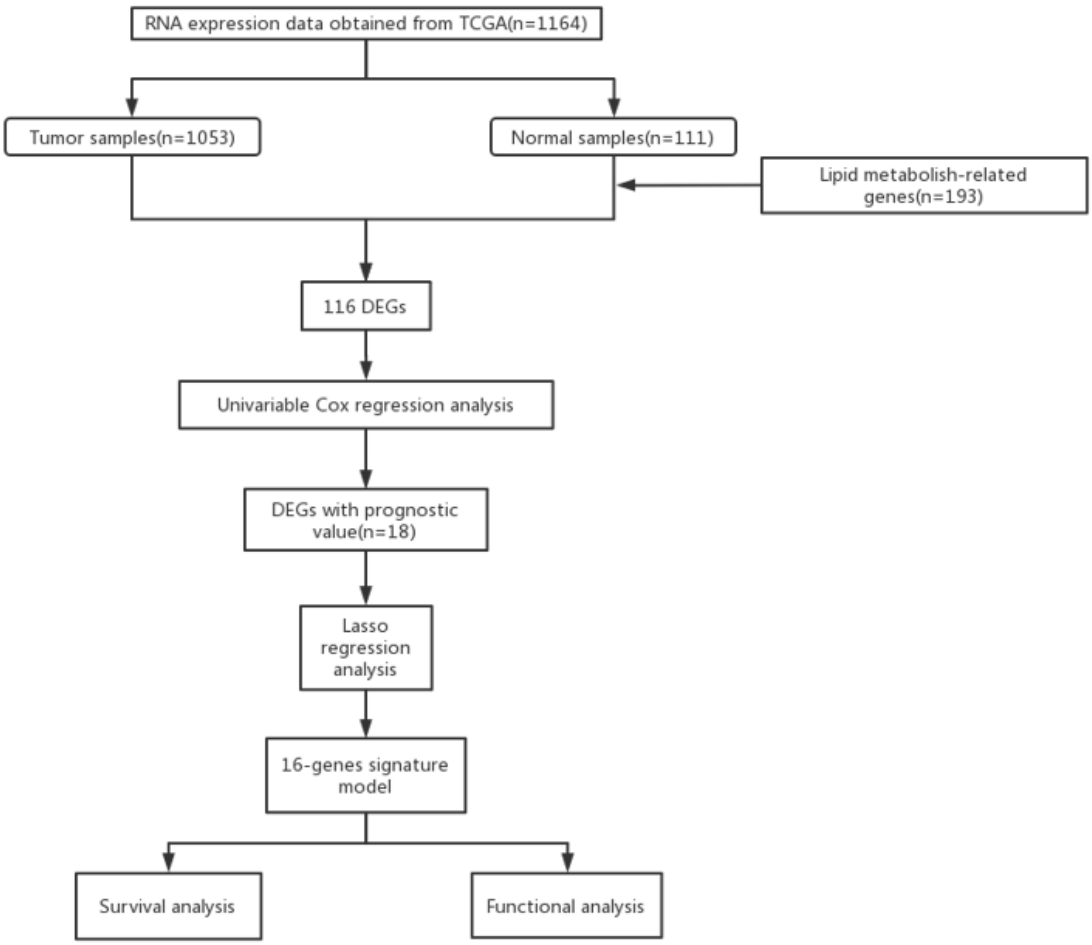

Figure 1

Flowchat of 16-genes signature model construction

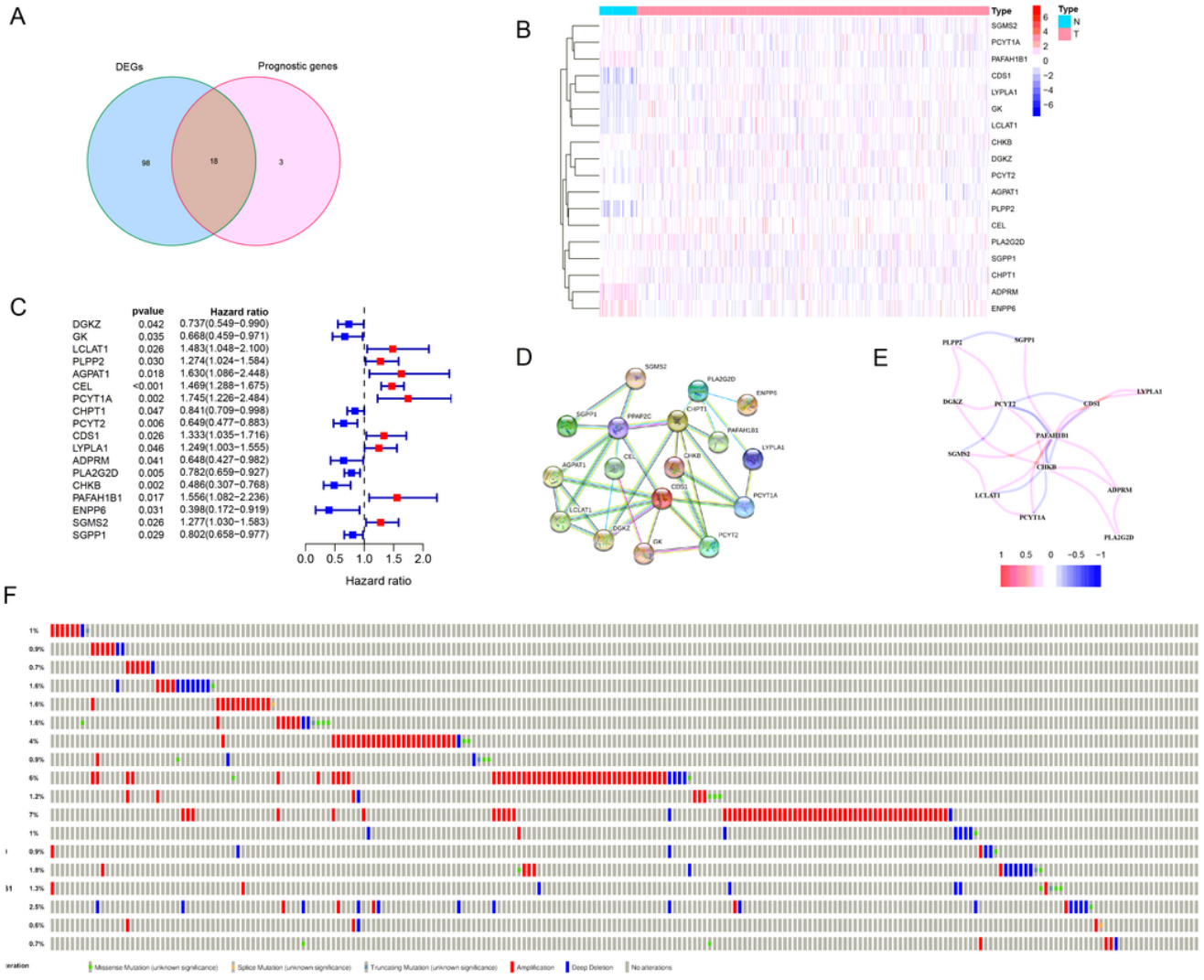

Figure 2

Identification of candidate genes related to lipid metabolism in breast cancer. (A) Venn diagram illustrating prognostic DEGs between tumor and adjacent nontumor samples. (B) Heatmap analysis of 18 prognostic DEGs. (C) Forest plot with hazard ratios from the survival analysis based on the univariate Cox regression model using gene expression levels as variables. (D) Construction and visualization of a protein-protein interaction (PPI) network of 18 genes generated using the STRING database. Yellow lines represent text-mining evidence in the PPI network, and black lines represent co-expressed proteins. (E)Network analysis of internal correlations among 18 candidate genes. Correlation coefficients are indicated by different colors (F)Alteration of the 18 candidate genes in clinical samples.

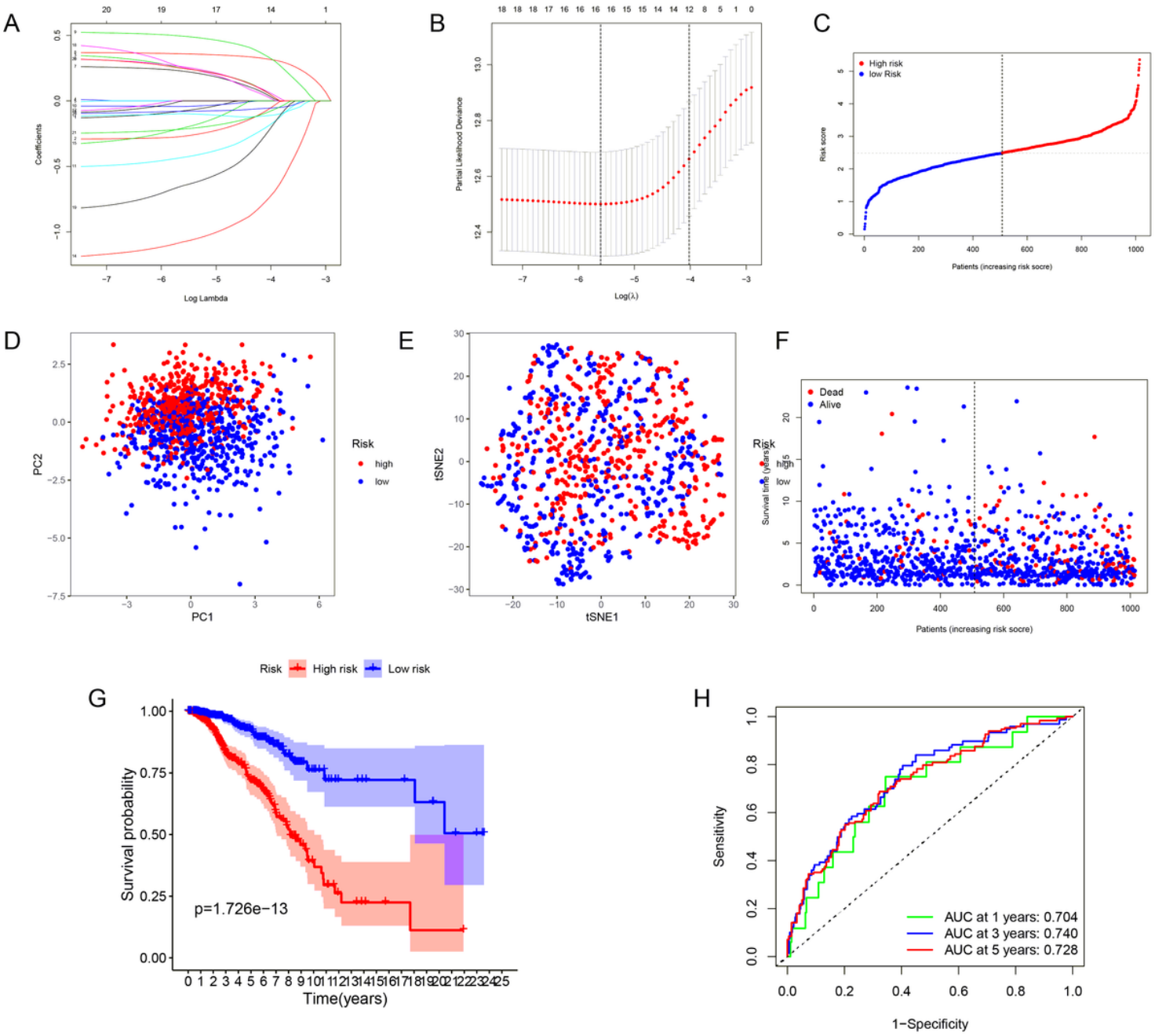

Figure 3

Synthetic analysis of a prognostic gene signature A. LASSO coefficient profiles of the 18 genes in BC samples. B. Selection of the optimal parameter (lambda) in the LASSO model for BC. C. The distribution and median value of the risk scores among BC samples D. Score plot for the principal component analysis (PCA). E. Two-dimensional projection of BC-seq data from TCGA by a t-SNE analysis. F. OS status, OS, and risk score in the TCGA cohort G. OS by Kaplan-Meier curves for patients in the two risk groups H. AUC of time-dependent ROC curves verified the prognostic performance of the risk score

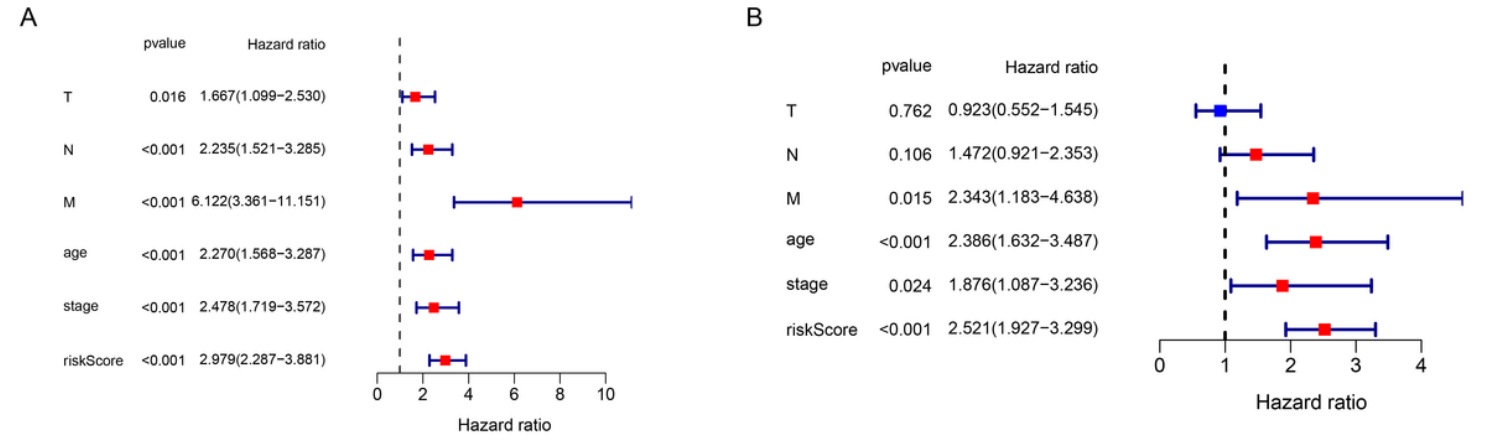

Figure 4

Independent prognostic analysis of risk scores and clinical parameters. A.The univariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients B.The multivariate Cox regression analysis of the associations between the risk scores and clinical parameters and the OS of patients.

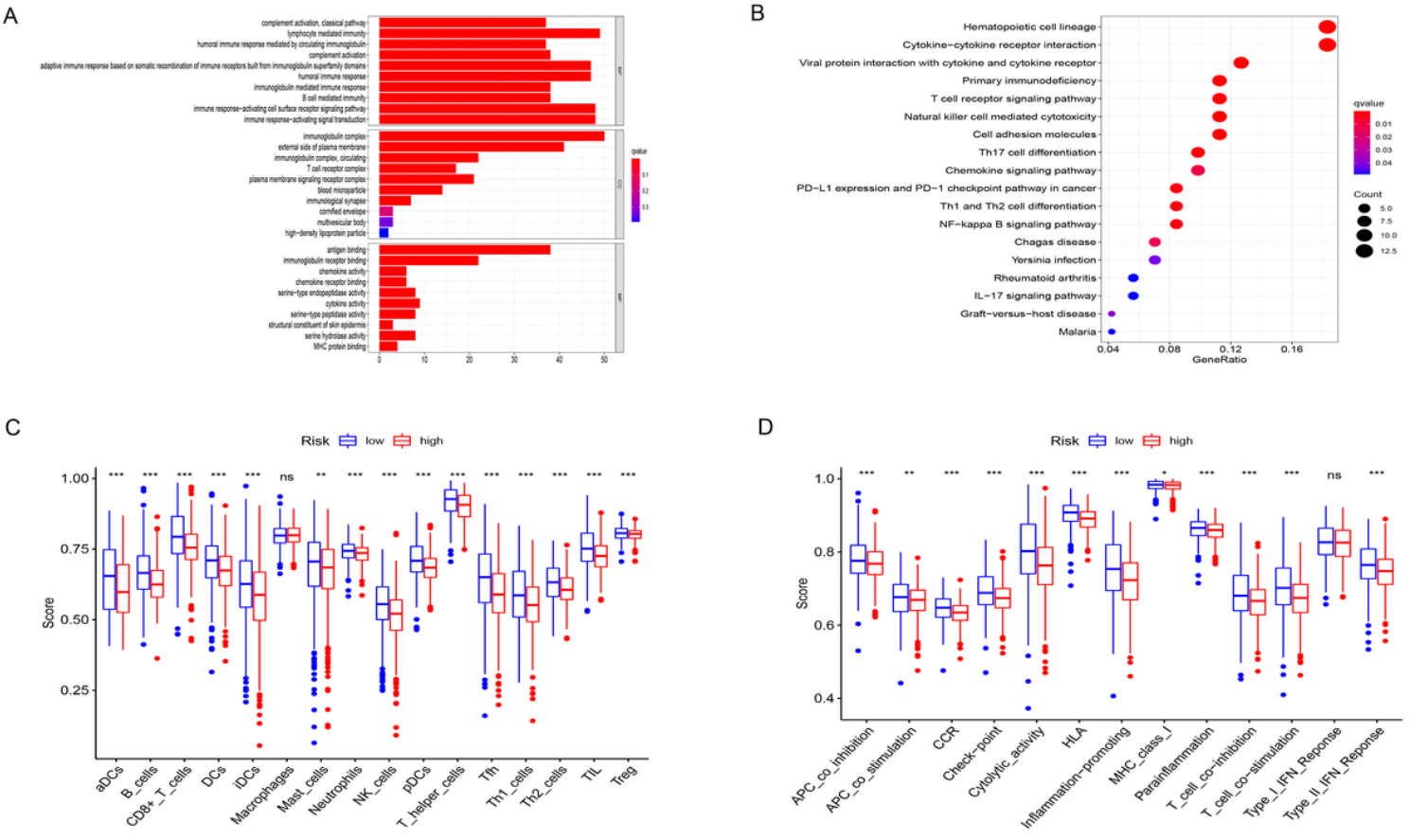

Figure 5

Functional enrichment analyses and ssGSEA enrichment scores in the TCGA cohort A.GO analysis showing the biological processes, cellular components and molecular functions enrichment of DEGs in

two groups B.KEGG analysis of DEGs in two groups C.the distribution of ssGSEA enrichment scores of 16 immune cells between high-risk and low-risk groups in BC samples D.the distribution of ssGSEA enrichment scores of 13 immune-related biological processes between high-risk and low-risk groups in BC samples

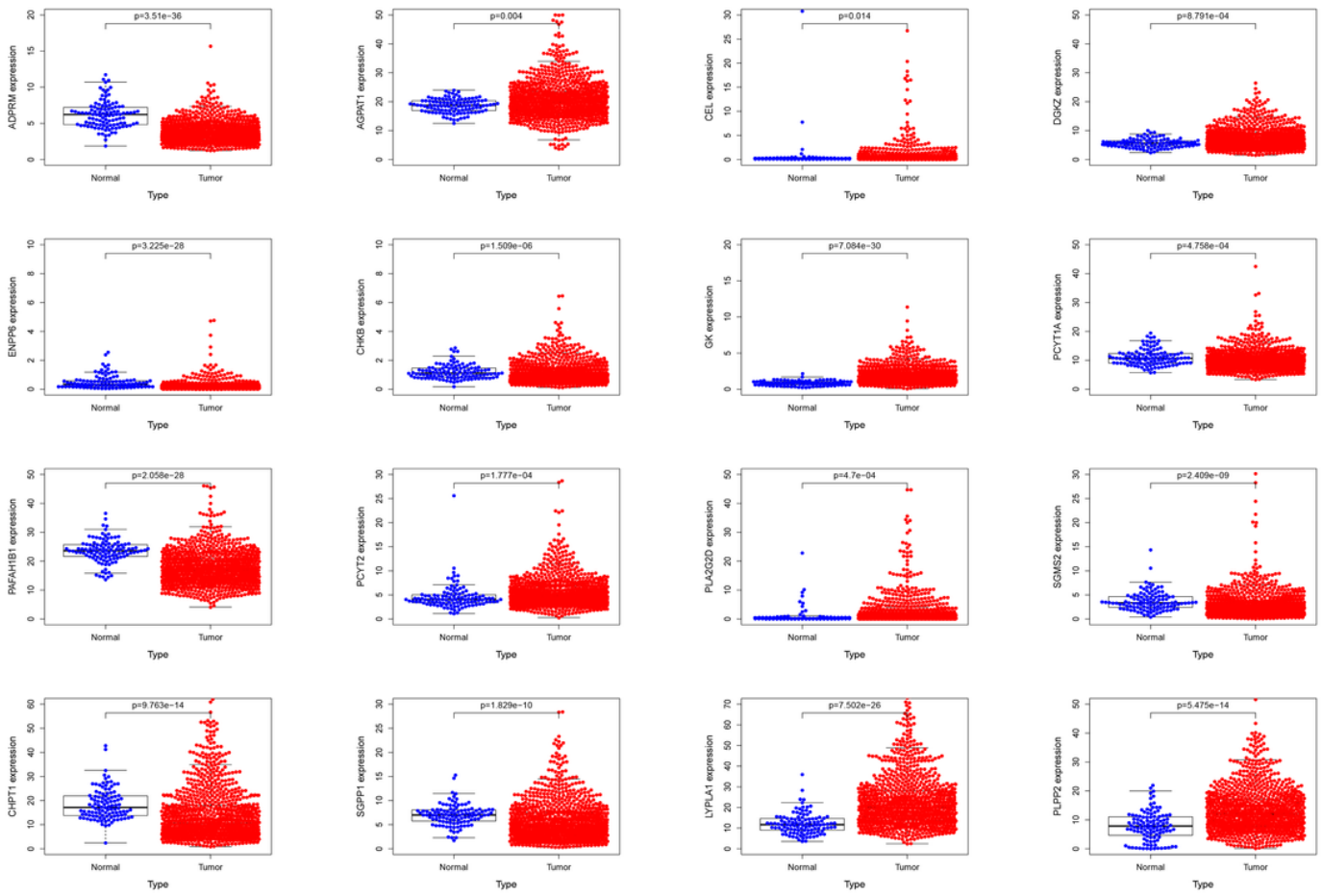

Figure 6

The mRNA expression level of 16 LMGs in TCGA