Bioinformatics analysis of Human Kallikrein 5 (KLK5) Expression in Metaplastic breast carcinoma: An Independent Cancer Biomarker

Yue Song (✉️ 445604106@qq.com)

Research article

Keywords: bioinformatics analysis, metaplastic breast carcinoma, differentially expressed genes, biomarkers, prognosis

Posted Date: May 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1263138/v1

License: ☒️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

Metaplastic breast carcinoma (MBC) has been proven to consistently harbour a basal-like immunophenotype and expression profile (lack of estrogen receptor (ER), progesterone receptor (PR), and HER2 expression), but has the worse prognosis and the faster tumorigenesis, lacking of standard effective treatment. We attempt to identify differentially expressed genes (DEGs) by searching for public datasets to help elucidate the underlying mechanisms of MBC development and provide new biomarkers and essential prognostic factors for MBC.

Methods

DEGs in TNBC (triple-negative breast cancer) and non-TNBC were screened out by analyzing the GSE76275 and The Cancer Genome Atlas, while GSE165407 identified the DEGs between metaplastic tumors and non-metaplastic tumors from TNBC patients. There were 2 co-expression DEGs identified but only KLK5 has significant differential expression in MBC and TNBC. Metascape and David online biological information databases were performed to reveal the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis and Gene Ontology (GO) analysis of DEGs. To detect the KLK5 expression in different subtypes of breast cancer, both online databases, including UALCAN, GEPIA, HPA and Breast cancer gene-expression miner, and qPCR were used. KLK5-associated gene analysis was performed by TCGA and LinkedOmics database was used to detect the co-expression genes of KLK5. Finally, Kaplan–Meier plotter showed the survival datas.

Results

KLK5 was screened by comparing DEGs in TNBC and MBC, which was over expression in TNBC but downregulated in MBC. The KEGG enrichment analysis and GO analysis revealed that estrogen signaling pathway was the most TNBC related pathway. Epithelial-to-mesenchymal transition (EMT) was the vital pathogenic mechanism in MBC by using KEGG and GO analysis of DEGs in MBC and non-MBC, as well as overlapping genes of the significantly different genes and co-expression genes of KLK5. What’s more, the prognosis analysis showed that KLK5 was statistically correlated with RFS mainly in LN positive TNBC subtypes, which were highly similar to MBC.

Conclusion

Our research identified the KLK5 by bioinformatic analysis to help us explore the underlying mechanism of MBC differing from TNBC and provide potential targets for diagnosis and prognosis of MBC.
Introduction

Breast cancer is one of the most common cancer with high incidence rate in women worldwide[1], which was divided into four subtypes: luminal, HER2 positive, basal, and normal-like tumors firstly[2]. Afterwards, six subgroups were put forward by Kast. et al: normal-like, luminal A and B, HER2-positive, basal-like, and claudin-low [3]. MBC represent a morphologically heterogenous group of invasive breast cancers characterized by the presence of malignant epithelial cells showing features of myoepithelial differentiation[4]. Most MBC display a triple-negative phenotype (lack of expression of ER, PR, and HER2), and are classified as basal-like or claudin-low molecular subtype, but have a more aggressive clinical behavior and fewer effective therapies[5].

Although MBCs have been proven to consistently harbour a basal-like (TNBCs) immunophenotype and expression profile, it exhibits lower response rates to the neo(adjuvant) systemic treatment, as well as has higher rates of disease progression, recurrence, and mortality than TNBCs[6, 7]. For our limited understanding of its pathogenesis, MBCs still receive similar treatments to TNBCs but are typically chemoresistant. Compared with other breast cancer subtypes, MBCs, especially with spindle cell metaplasia, frequently display stem cell–like and EMT characteristics[8, 9]. However, MBCs also express some other mutation genes involved in Phosphatidylinositol 3-kinase signaling(PI3K) pathway and DNA damage repair mechanisms, but were found not to be driven by a highly recurrent expressed fusion gene or a highly recurrent expressed mutation affecting previously described cancer genes[10–13]. Further studies need to deeply explore the specific mechanisms of MBCs and find some key genes to provide an opportunity for targeted therapy.

Due to the character of early relapse, highly invasion and metastasis, lack of standard effective treatment and poor prognosis, it is necessary to explore the key DEGs in MBC differing from other histologic types of TNBC, which will help us to determine the prognosis of MBC and find a breakthrough in the onset mechanism of MBC. In this study, we attempted to identify DEGs by searching for public datasets and performing bioinformatics analysis to find some key genes in tumorigenesis and progression of MBC compared with TNBC.

Methods

Microarray data

GSE76275 and GSE165407 were obtained from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), which is a free public repository for data storage, including microarray data and next-generation sequencing. The GSE76275 datasets was obtained from the Affymetrix Human Genome U133 Plus 2.0 Array, consisting of 198 TNBC tissue samples and 67 non-TNBC breast cancer tissue samples. The GSE165407 datasets was obtained from Illumina HiSeq 2000 (Homo sapiens), consisting of 8 treatment-naïve metaplastic tumors and 20 treatment-naïve non-metaplastic tumors from
triple-negative breast cancer patients. Whole-transcriptome sequencing (WTS) data of 122 TNBC and 655 non-TNBC samples were downloaded from TCGA database (http://cancergenome.nih.gov).

Screening for DEGs

The adjusted P-values (adj. P) and Benjamini and Hochberg false discovery rate were applied to provide a balance between discovery of statistically significant genes and limitations of false-positives. P-value < 0.001 and |fold change (FC)| ≥ 1.5 were set as the thresholds for identifying DEGs in GEO databases. RNA-sequencing data downloaded from the TCGA database. Differential mRNA expression was analyzed in TNBC samples and non-TNBC samples with the R software with the Limma package. The screening criteria utilized in TCGA: P-value < 0.001 and |fold change (FC)| ≥ 2.

The groups (KLK5 low and KLK5 high) were divided based on the median values of the KLK5 transcript by using the RNAseq data from TCGA. The Limma package in R 3.3.3 was used to screen DEGs between the KLK5 high and KLK5 low groups in breast cancer. Adjusted P < 0.001 and |log1.2 FC| ≥ 1.5 were used as cut-off values for identifying DEGs.

Analysis of co-target genes by venn diagram

FunRich is a stand-alone software tool used mainly for functional enrichment and interaction network analysis of genes and proteins. Besides, the results of the analysis can be depicted graphically in the form of Venn, which showed the co-expressed genes of different groups.

GO functional enrichment and KEGG analysis

The database for annotation, visualization and integrated discover, Metascape[14] (http://metascape.org/gp/index.html#/main/step1) is an online biological information database that integrates biological data and analysis tools, was introduced to perform functional enrichment analysis for the DEGs. David is another online tool (http://david.abcc.ncifcrf.gov)[15, 16], which provides a comprehensive set of high-throughput functional gene analyses to understand the enrichment of KEGG for DEGs and GO enrichment analysis for co-expression DEGs. KEGG is a database resource for understanding high-level functions and biological systems from large-scale molecular datasets generated by high-throughput experimental technologies. GO is a major bioinformatics tool to annotate genes and analyze biological process of these genes. P-value < 0.05 was considered as statistically significant.

Survival data from Kaplan–Meier plotter

The prognostic values of KLK5 in TNBC were analyzed using Kaplan–Meier plotter (KM plotter) database[17], a software available online that specializes survival analysis. In brief, the selected gene was entered into the database, after which Kaplan–Meier survival plots were generated and hazard ratio (HR), 95% confidence intervals (CI), log rank P-value were displayed on the webpage. P-value of < 0.05 was regarded as statistically significant.
Analysis of target gene expression using UALCAN, GEPIA and HPA database

UALCAN (http://ualcan.path.uab.edu/), a portal for facilitating tumor subgroup gene expression and survival analyses, provides easy access to publicly available cancer transcriptome data including TCGA[18]. GEPIA is a new web-based tool which can provide differential expression analysis of tissues by using TCGA[19]. These two databases were utilized to analyze the expression level of KLK5 in breast cancer based on breast cancer subclasses. Statistical analysis of comparison between normal group and TNBC group was performed and log-rank P-value was observed in the database. The online website HPA [20] (https://www.proteinatlas.org/) contains the human transcriptomic and proteomic data in cells, tissues, and organs from human normal or pathological tissues using RNA sequencing (RNA-Seq) analysis and immunohistochemistry (IHC). Thus, KLK5 expression was analyzed in various breast cancer cell lines.

LinkedOmics database analysis

A new and unique tool LinkedOmics (http://www.linkedomics.orglogin.php) [21] is used to determine KLK5 co-expression in breast cancer. KLK5 co-expression genes in breast cancer was determined by analyzing mRNA sequencing datas from the TCGA database, and Pearson's correlation coefficient was calculated to analyze the data. Volcano plots displaying the results were generated.

Breast cancer gene-expression miner

The expression of KLK5 mRNA in different subtypes of breast cancer was analyzed using the Breast Cancer Gene-Expression Miner (bcGenExMinerv4.5, http://bcgenex.centregauducheau.fr/BC-GEM) [22].

RNA extraction and quantitative-PCR (qPCR)

TRIzol reagent (Invitrogen/Thermo Fisher Scientific) was used to extract the RNA from the HS578T and MDA-MB-231 cell lines. Equal amounts of RNA were converted into cDNA using targeted genes primers by using a Prime Script RT Reagent kit (Takara Bio, Inc., Otsu, Japan). Quantitative PCR (qPCR) was performed using a Reverse Transcription System (Promega, Madison, WI, USA) on a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) according to the instructions of the manufacturer. Each sample was analyzed in triplicate. Relative expression was calculated using the 2-ΔΔCq method (25). The PCR conditions were as follows: 35 cycles of 95˚C for 30 sec, 56˚C for 30 sec, 72˚C for 90 sec, and a final extension at 72˚C for 5 min.

Statistical analysis

The results were shown as mean ± SD. Differences between two groups were estimated using unpaired Student’s t-test. A two-tailed value of P < 0.05 was considered as statistically significant.
Results

Identification of DEGs between TNBC and non-TNBC samples.

To identify DEGs of array GSE76275 downloaded from GEO, we analyzed a total of 198 TNBC tissue samples and 67 non-TNBC samples by screening genes with $|\log_2$ fold change (FC)| larger than 1.5 and adj P-value < 0.001. Based on this analysis, a total of 217 DEGs were found to be significantly differentially expressed in TNBC and non-TNBC samples, including 62 upregulated and 155 downregulated genes. Differential expression analyses of protein-coding mRNAs in 122 TNBC samples and 655 non-TNBC samples from the TCGA data sets were performed using R with Limma package (Log2FC $\geq$ 2 and P<0.001). A total of 1235 DEGs were identified. Sangerbox software was used to provide a volcano plot of both GSE76275 and TCGA DEGs (Fig.1A and B). We next built Venn to identify 111 co-expression DEGs(Table S1) in total both in GSE76275 and TCGA datasets (Fig.1C).

KEGG pathway enrichment and GO functional enrichment analysis of DEGs between TNBC and non-TNBC samples

To further analyze the enriched pathways of DEGs between TNBC and non-TNBC samples, KEGG pathway analysis of co-expression DEGs by using DAVID revealed that estrogen signaling pathway is the most enriched pathway in the development of breast cancer(Fig.2A), which is the common pathway in both GSE76275 and TCGA DEGs. GO analysis in David online tool was performed to further investigate the biological functions of the 111 screened DEGs, including biological process (BP), cellular component (CC) and molecular function (MF). The enriched GO functions of BP category included positive regulation of transcription from RNA polymerase II promoter, response to estrogen, negative regulation of gene expression, utero embryonic development and regulation of ion transmembrane transport; In the cellular component group, DEGs were mainly enriched in extracellular exosome, extracellular space, anchored component of membrane and cell surface; In the molecular function, greater percentages of DEGs were involved in protein homodimerization activity, calcium ion binding, enzyme binding and transcription factor binding(Fig.2B).

Identification of DEGs between metaplastic TNBC and non-metaplastic TNBC samples.

MBC is considered to be one of the most malignant types of TNBC. As the result, we next aimed to find the difference of DEGs in MBC and other subtypes of TNBC. First, we performed the different expression analysis between metaplastic TNBC and non-metaplastic TNBC by download of the GEO database GSE165407, 108DEGs(Table S2), including 69 up-regulated genes and 39 down-regulated genes, were identified by screening genes with $|\log_2$ fold change (FC)| $\geq$ 1.5 and adj P-value < 0.001. Next, volcano plot and heatmap of genes screened by GSE165407 were provided by using R with pheatmap package (Figure3 A and B). Finally, there are 2 DEGs both in GSE76275, TCGA and GSE165407, of which KLK5 probably played a different role in TNBC and MBC. As shown in Table1, the expression of KLK5 was over expression in TNBC, but down expression in MBC, which suggested that KLK5 might be one of the piovotal molecules that play a key role in the mechanism of tumorigenesis in MBC.
KEGG pathway enrichment analysis and GO functional enrichment analysis of DEGs between metaplastic TNBC and non-metaplastic TNBC samples.

The distinction of molecular alterations and the pathway enrichment in MBC and other forms of TNBC is poor understood. To this purpose, we used R to perform KEGG pathway enrichment analysis and GO functional enrichment analysis (Figure 4). MBC subtypes exhibit significant pathways, including IL-17 signaling pathway, Cell adhesion molecules and transcriptional misregulation in cancer. The enriched GO terms were mainly associated with extracellular matrix organization, extracellular structure organization, external encapsulating structure organization, collagen-containing extracellular matrix and glycosaminoglycan binding. Compared with non-metaplastic TNBC, MBC-specific increases related to epithelial-to-mesenchymal transition and extracellular matrix. Next, we detected the expression of KLK5 in breast cancer and tried to explore if KLK5 mainly mediates the EMT process of MBC and thus promoted MBC progression.

Expression of Kallikrein-related peptidase 5 (KLK5) in breast cancer

Because of the different expression levels exhibited by KLK5 in TNBC and MBC, we next analyzed the expression of KLK5 in breast cancer using different databases. Firstly, we found that KLK5 was significantly expressed in breast cancers of all types of human cancers by using the GEPIA database (Fig 5 A); In the HPA database, the KLK5 expression levels were mainly distributed in breast tissue cell lines, including breast glandular cells, breast myoepithelial cells and breast glandular cells (Fig 5 B). Using the GEPIA and UALCAN databases, we next tested the mRNA expression and protein levels of KLK5 in breast cancer. The transcriptional levels of KLK5 were significantly downregulated in breast cancer compared with that in normal tissues in GEPIA, and the protein levels of KLK5 were consistent with the RNA levels in breast cancer analyzed by UALCAN (Figure 5 C and D). Finally, we performed PCR to detect the expression of KLK5 in TNBC cell line MDA-MB-231 and metaplastic-like breast cancer cell line HS578T respectively, and the results were consistent with that of our bioinformatic analysis (Fig 5E). It seemed that KLK5 mainly acted as a suppressor gene in breast cancer. However, the results above showed that KLK5 became to be an oncogene and was over expression in TNBC with statistical significance. While, in MBC, KLK5 regained its property of suppressor gene with significantly downregulated expression levels in MBC. Thus, it is necessary to elucidate the expression of KLK5 in different subtypes of breast cancer.

The relationship between KLK5 expression and clinical indicators in breast cancer patients

According to different clinical indicators, we next compared KLK5 expression among groups of patients by using the bc-GenExMiner online tool. As shown in Figure 6, the expression of KLK5 was significantly higher in ≤51-year group compared with >51-year group. ER, PR status was negatively associated with KLK5 expression, while HER-2 status was positively associated with KLK5 expression. Breast cancer patients with basal-like and TNBC subtypes expressed higher KLK5 levels, which is consistent with our bioinformatic analysis. From the results above, we speculate that the status of ER and PR has a more important influence on the expression of KLK5 in TNBC. However, it is possible that the status of ER and...
PR has no similar effect on the expression of KLK5 in MBC, which need deeper exploration of its function in MBC. Besides, breast cancer patients with positive nodal status (N) had no statistical significant compared with negative nodal status.

**KLK5 prognosis in breast cancer**

Kaplan–Meier curve analyses represented the prognostic data of KLK5 by using www.kmplot.com website. Relapse-free survival (RFS) was chosen to predict the prognosis of patients with lymph nodes (LN) positive or negative in different subtypes of breast cancer. As shown in Figure 7, high expression of KLK5 was significantly correlated with worse RFS in breast cancer with LN positive but had no statistically significant association with LN negative breast cancer, which implied that KLK5 was an oncogene in breast cancer. However, our results above has proven that KLK5 was significantly downregulated in breast cancer via different methods, we speculated that low expression of KLK5 might have no clinical significant in breast cancer without accurate subtypes. As our bioinformatic analysis above showed that KLK5 was oncogene in TNBC but suppressor gene in MBC, we further analyzed the Kaplan–Meier curve in TNBC. To our surprise, there was no correlation between KLK5 expression and RFS in either LN positive or negative TNBC. As MBC were highly similar to mesenchymal and mesenchymal stem-like subtype for all of them are related with EMT, we conducted the Kaplan–Meier curve analysis in mesenchymal and mesenchymal stem-like subtype with LN positive or negative respectively. Consequently, the KLK5 was significantly associated with RFS in both subtypes with LN positive, and was consistent with that of our bioinformatic analysis in MBC being a suppressor gene. Taken together, KLK5 might serve as a biomarker in MBC, especially with LN positive.

**KLK5-associated gene analysis between KLK5 high and KLK5 low in breast cancer patients**

To further explore the role of KLK5 in breast cancer, we firstly compared the transcriptomes of the KLK5 high and KLK5 low groups base on the TCGA database. A total of 726 DEGs between the KLK5 low and KLK5 high groups (p ≤ 0.001, |log1.2 FC| ≥ 1.5, Table S3) were screened; Among them, 515 genes were significantly upregulated and 211 genes were downregulated in the KLK5 high group, respectively (Fig 8A). Next, we analyzed the co-expressed genes of KLK5 by using the LinkedOmics database. As shown in Fig 8B, a total of 1443 co-expression genes were significant correlated with KLK5 in breast cancer (False discovery rate, FDR ≤ 0.001, p ≤ 0.001, and |cor.| ≥ 0.3; Table S4). 1246 genes were positively correlated with KLK5 expression, and 197 genes were negatively correlated with KLK5 in breast cancer. There are 231 overlapping genes determined by comparison of the DEGs between the KLK5 low and KLK5 high groups and the co-expressed genes (Table S5). The 231 overlapping genes consisted of 229 up-regulated genes and 2 down-regulated genes, which were screened out for further study (Fig 8C and D).

**Functional analysis of the overlapping genes**
In order to investigate the possible biological function of KLK5 in breast cancer, KEGG and GO annotation of the 231 overlapping genes were performed by using Metascape free tool online. Among the 20 top clusters of enriched sets, extracellular matrix organization (GO:0030198) and cell-cell adhesion via plasma-membrane adhesion molecules (GO:0098742) were directly correlated with EMT, which proved that KLK5 was involved in the EMT process of breast cancer (Fig 9A). Furthermore, KLK5 had a high level of enrichment in both breast cancer and breast tissue, further suggesting KLK5 regulated the development of EMT in breast cancer.

Discussion

Breast cancer could be divided into diverse group of diseases according to morphology, presentation, molecular profile, and response to therapy [23]. TNBCs, are immunohistochemically characterized by ER, PR and HER2 negative, get more attractive attention for lacking of effective treatment and associating with poor prognosis [24]. MBC is a rare breast cancer subtype belonging to TNBC but poorly responsive to contemporary systemic therapy. Many previous studies have explored the molecular characteristics of MBC[25, 26]. It has been reported that MBCs might not be driven by a highly-recurrent/ pathognomonic expressed fusion gene or expressed mutation affecting known cancer genes between the distinct MBC subtypes[13]; The heterogeneity of MBC has been investigated at the genetic level, as well as the proteomic profiles, and provide a unique opportunity to impact precision therapies to this aggressive form of breast cancer. However, few potential diagnostic and prognostic markers between MBCs and TNBCs are discovered.

KLK5 might be one of the pivotal molecules that play a key role in the mechanism of tumorigenesis in MBC, because of the completely opposite expression levels of KLK5 in TNBC and MBC. KLK5 is a secreted serine protease encoded by KLK5 gene, located on chromosome 19q13.4[27]. It is mainly present in human epidermis by involving in skin desquamation, but also found in other organs like breast, ovarian, testis, vagina, esophagus, etc [28-30]. Interesting, KLK5 plays a different role in different cancers. The overexpression of KLK5 predicts the worse prognosis of various types of cancer, including cutaneous squamous cell carcinoma, colorectal adenoma-carcinoma, bladder tumor, ovarian cancer acting as an oncogene [31-4]; while the character of cancer suppressor gene is mainly found in hormone-dependent tumors, like, prostate cancer, vaginal cancer and breast cancer [30],[35-6]. In our study, KLK5 mainly acts as a suppressor gene in breast cancer without precise molecular subtypes. While, it is surprised that KLK5 has different expression levels in TNBC and MBC. The results analyzed by bc-GenExMiner online tool suggest that KLK5 is over expressed in breast cancer patients with basal-like and TNBC subtypes, as well as either ER or PR negative ones. A possible explanation for the abnormal expression of KLK5 in TNBC may be because of its correlation with the estrogen signaling pathway, which is proven to be one of the vital pathways in TNBC. However, the over expression of KLK5 in TNBC has no statistical significance in the prognostic data by analyzing RFS in either LN positive or negative TNBC. In the other sides, the expression of KLK5 in MBC is significantly down-regulated compared with that in TNBC. The differentially expression of KLK5 in MBC could not be simply explained by the status of ER and PR, which need further exploration.
In order to get deeper insight of the function of KLK5 in breast cancer, we screened out 231 overlapping genes determined by comparison of the DEGs between the KLK5 low and KLK5 high groups and the co-expressed genes, and KEGG and GO analysis suggest that EMT but estrogen signaling pathway is the vital pathogenic mechanism by which KLK5 plays an important role in breast cancer, which is also the core pathway by using GO and KEGG analysis in MBC. It has been reported that KLK5 could interrupt ECM physical barriers and cells’ interaction by cleaving extracellular matrix components directly or indirectly via the KLK5-induced activation of other extracellular proteases, thus facilitating the invasiveness, metastasis and angiogenesis of cancer cells [37]; Therefore, the downregulated of KLK5 might have connection with EMT in MBC to promote the development of breast cancer. As MBC was most similar to the mesenchymal and mesenchymal stem-like TNBC subtypes which were also both related with EMT, we conducted the Kaplan–Meier curve analysis in mesenchymal and mesenchymal stem-like subtype with LN positive or negative respectively, and the results showed that low expression of KLK5 had statistical correlation with poor prognosis in LN positive ones. The KM analysis showed that KLK5 was statistically related with worse prognosis in LN positive subtypes, but has no statistical significance in LN negative subtypes, which further demonstrated that the key mechanism of downregulated KLK5 promoted MBC progression might be related to accelerate tumor metastasis by EMT.

Our analysis shows that KLK5 is significantly down regulated in breast cancer without obvious prognostic value. We therefore speculate that the mechanism by which KLK5 regulates EMT is not the core part that promotes the progression of breast cancer. On the other sides, the mechanism of KLK5 overexpression in TNBC is not yet clear. As RFS analysis shows that KLK5 has no statistical significant in TNBC, but KLK5 is negatively correlated with metastatic RFS in MBC by Kaplan–Meier plotter, the expression alteration of KLK5 that distinguish MBC from TNBC is therefore worth for deeper research. Still a lot of questions remain open on why only the low expression of KLK5 in MBC has clinical significance, and if downregulated KLK5 promotes the EMT progression of MBC through a different mechanism from breast cancer. Taken together, the role of KLK5 in MBC is clearly more important than it in breast cancer as well as TNBC, KLK5 may inhibit EMT in different mechanisms from the known studies during tumor progression in MBC, which need further exploration.

In conclusion, in addition to being a diagnosis biomarker in breast cancer, KLK5 is proved to be also a prognosis biomarker in MBC by our study. Significantly decreased KLK5 expression was detected in MBC but over expression in TNBC by bioinformatics analysis. The different expression of KLK5 in MBC and TNBC may be considered as a novel and independent biomarker for the differential diagnosis between MBC and TNBC. What’s more, KLK5 low expression is associated with the aggressiveness of MBC via promoting EMT development, so KLK5 could be a reliable prognostic factor in MBC. Further studies are needed to clarify the involvement of KLK5 in EMT progression of MBC.

Declarations
Acknowledgements

Not applicable.

Authors' contributions

Yehui Shi and Guiying Bai designed the experiments; Yue Song analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Funding

No funding was received.

Availability of data and materials

The datasets used and/or analyzed during the current study are available.

Please contact the author to get the datasets.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

References


### Table

**Table 1: The LogFC and P value of KLK5 in different datasets**

<table>
<thead>
<tr>
<th>Datasets</th>
<th>LogFC</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE76275</td>
<td>1.57</td>
<td>1.28E-13</td>
</tr>
<tr>
<td>TCGA</td>
<td>3.45</td>
<td>7.18E-56</td>
</tr>
<tr>
<td>GSE165407</td>
<td>-3.62</td>
<td>2.28E-06</td>
</tr>
</tbody>
</table>

### Figures
Figure 1

Volcano plot of the DEGs in GSE76275 and TCGA.

A and B: The black dots represent genes that are not differentially expressed between TNBC and non-TNBC tissues, and the red dots and green dots represent the upregulated and downregulated genes in TNBC tissues, respectively. A shows the DEGs screened from GSE76275; B shows the DEGs screened from TCGA. C: Venn shows the co-expression genes in GSE76275 and TCGA.
Figure 2

KEGG pathway enrichment analysis and GO functions of DEGs in GSE76275 and TCGA.

A. Enriched KEGG pathway analysis of the co-expression genes in GSE76275 and TCGA; B. GO functions for the co-expression genes in GSE76275 and TCGA. The GO analysis divided co-expression DEGs into three functional groups: biological process, cellular component, and molecular function. The blue column represents the P-value; the orange column represents the number of DEGs.
Figure 3

Volcano plot and heatmap of the GSE165407

A shows the DEGs screened by the condition (LogFC≥2, P≤0.001). The black dots represent genes that are not differentially expressed between TNBC and non-TNBC tissues, and the red dots and green dots represent the upregulated and downregulated genes in TNBC tissues, respectively. B Heatmap of DEGs in GSE165407.
Figure 4

KEGG pathway enrichment analysis and GO functions of DEGs in GSE165407

A. Enriched KEGG pathway analysis of the DEGs in GSE165407; B. The GO analysis divided DEGs into three functional groups: biological process, cellular component, and molecular function.
Figure 5

The expression of KLK5 in breast cancer.

A. KLK5 expression in various cancer tissues and normal tissues analyzed by GEPIA; B. KLK5 expression in breast cancer cell lines, analyzed by HPA; C. KLK5
expression in breast cancer tissues and normal tissues analyzed by GEPIA; D. The protein expression of KLK5 in primary breast cancer and normal tissues analyzed by UALCAN cancer database. E. The mRNA expression of KLK5 detected by PCR in metaplastic-like breast cancer cell line HS578T and TNBC cell line MDA-MB-231 compared with that in databases of our researches.

Figure 6

Box plot revealing the relationship between KLK5 expression and different clinical indicators
Data shown for age (A), ER (B), PR (C), ER&PR (D), HER-2 (E), basal-like status (F), triple-negative status (G), basal-like & triple-negative status(H), nodal status (I)

Figure 7

Kaplan–Meier survival curve of KLK5 in different subtypes of TNBC.
A and B. Kaplan–Meier survival curve of KLK5 in lymph nodes negative and positive breast cancer; C and D. Kaplan–Meier survival curve of KLK5 in lymph nodes negative and positive TNBC; E and F. Kaplan–Meier survival curve of KLK5 in lymph nodes negative and positive in mesenchymal subtype of TNBC; G and H. Kaplan–Meier survival curve of KLK5 in lymph nodes negative and positive in mesenchymal stem-like subtype of TNBC.

Figure 8

Genome-wide genes associated with KLK5 expression.

A. Volcano plot of different gene-expression profiles between the KLK5 low and KLK5 high groups. B. Volcano plots for the analysis of the co-expression genes associated with KLK5 expression using the LinkedOmics. C. Overlapping genes between positively correlated genes and significantly increased genes. D. Overlapping genes between negatively correlated genes and significantly reduced genes.
Figure 9

Enrichment of functions and signaling pathways of the overlapping genes in breast cancer.

A. Enrichment of differentially expressed genes in tissues and cells. B. Enrichment of overlapping genes in diseases. C. Analysis of GO and KEGG pathway associated with
KLK5 expression.

Supplementary Files

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

- Additionalfile1.xlsx