Computational and Structural Biotechnology Journal 21 (2023) 3010–3023

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

A novel exosome-derived prognostic signature and risk stratification for breast cancer based on multi-omics and systematic biological heterogeneity

Fei Long [a,1], Haodong Ma [a,1], Youjin Hao [b,1], Luyao Tian [a], Yinghong Li [c], Bo Li [b], Juan Chen [a], Ying Tang [a], Jing Li [a], Lili Deng [a], Guoming Xie [a,*], Mingwei Liu [a,*]

[a] Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing 400016, PR China
[b] College of Life Sciences, Chongqing Normal University, Chongqing 401331, PR China
[c] Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, PR China

ARTICLE INFO

Article history:
Received 18 November 2022
Received in revised form 29 April 2023
Accepted 11 May 2023
Available online 12 May 2023

Keywords:
Breast cancer
Tumor heterogeneity
Risk stratification
Exosome-derived signature
Prognostic panel

ABSTRACT

Tumor heterogeneity remains a major challenge for disease subtyping, risk stratification, and accurate clinical management. Exosome-based liquid biopsy can effectively overcome the limitations of tissue biopsy, achieving minimal invasion, multi-point dynamic monitoring, and good prognosis assessment, and has broad clinical prospects. However, there is still lacking comprehensive analysis of tumor-derived exosome (TDE)-based stratification of risk patients and prognostic assessment for breast cancer with systematic dissection of biological heterogeneity. In this study, the robust corroborative analysis for biomarker discovery (RCABD) strategy was used for the identification of exosome molecules, differential expression verification, risk prediction modeling, heterogenous dissection with multi-ome (6101 molecules), our ExoBCD database (306 molecules), and 53 independent studies (481 molecules). Our results showed that a 10-molecule exosome-derived signature (exoSIG) could successfully fulfill breast cancer risk stratification, making it a novel and accurate exosome prognostic indicator (Cox *P* = 9.9E-04, *HR* = 3.3, 95% *CI* 1.6–6.8). Interestingly, *HLA-DQB2* and *COL17A1*, closely related to tumor metastasis, achieved high performance in prognosis prediction (86.35% contribution) and accuracy (Log-rank *P* = 0.028, *AUC* = 85.42%). With the combined information of patient age and tumor stage, they formed a bimolecular risk signature (Clinmin-exoSIG) and a convenient nomogram as operable tools for clinical applications. In conclusion, as an extension of ExoBCD, this study conducted systematic analyses to identify prognostic multi-molecular panel and risk signature, stratify patients and dissect biological heterogeneity based on breast cancer exosomes from a multi-omics perspective. Our results provide an important reference for in-depth exploration of the "biological heterogeneity - risk stratification - prognosis prediction".

1. Introduction

Breast cancer (BC) is the most common cancer with a high incidence rate and the leading cause of death in women [1]. Its spatio-temporal heterogeneity at the molecular level is an important factor in cancer cell metastasis, drug resistance, recurrence, and poor prognosis after treatment [2–5]. Therefore, an in-depth understanding of the heterogeneity will greatly benefit accurate clinical management and therapeutic decision-making [6].

Currently, many studies have been conducted to identify risk factors and prognosis-related subtype-specific genes in terms of genomic alterations, gene expression, and tumor microenvironment through tissue biopsy, which realizes the tumor grade diagnosis and the risk assessment of patients. Based on the expression profiles (microarray data) of 8102 genes, Perou et al. first classified breast cancer into Luminal, HER2$^+$, Basal-like, and normal-like types [7]. In 2003, Sorlie et al. subdivided the Luminal type into Luminal A and Luminal B subtypes [8]. In 2009, Parker et al. discovered an array of 50 genes for rapid differentiation of five types of breast cancer and

* Corresponding authors.
E-mail addresses: guomingxie@cqmu.edu.cn (G. Xie), liumingwei@cqmu.edu.cn (M. Liu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.csbj.2023.05.013

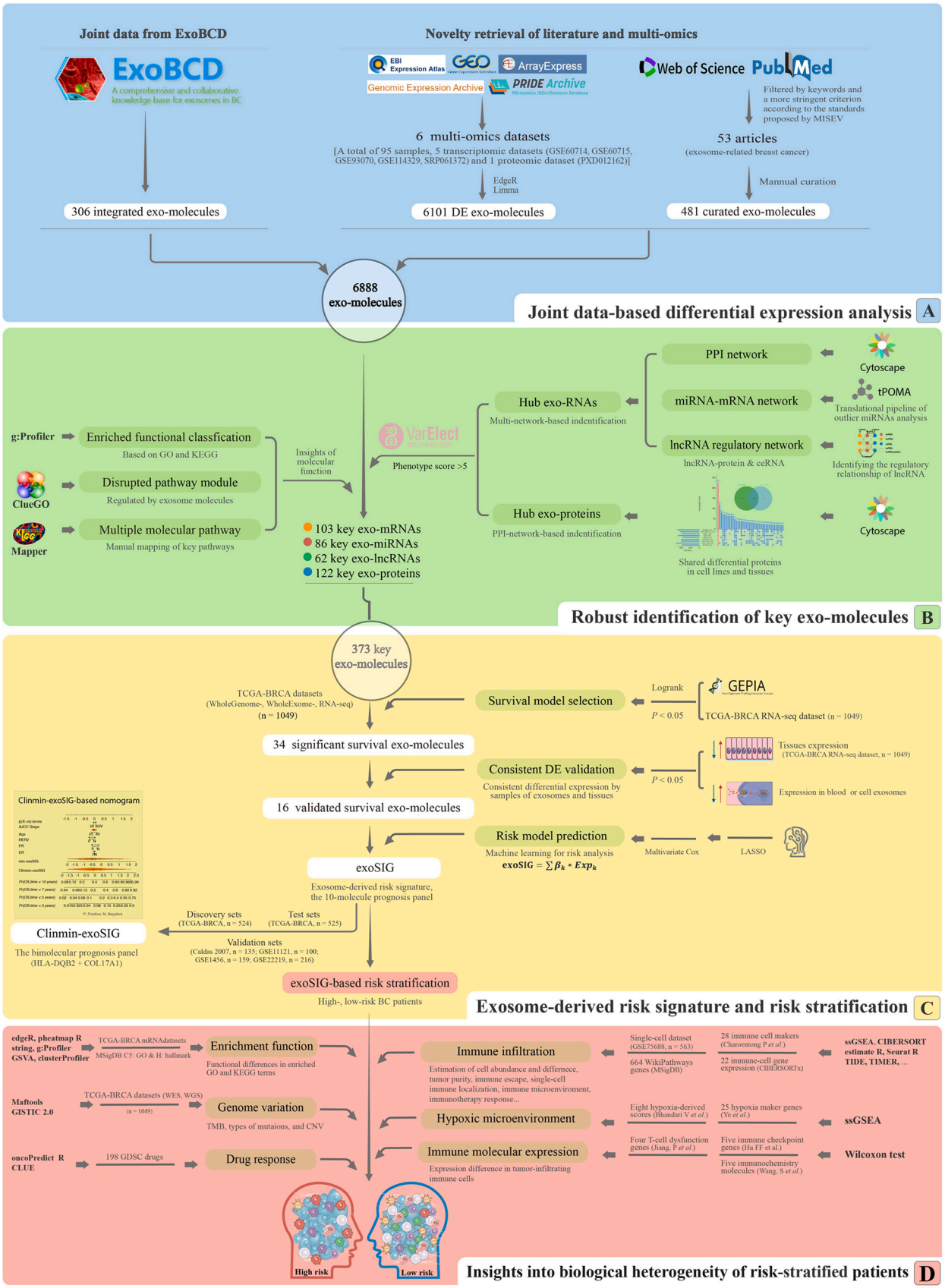

Fig. 1. Schematic diagram of this study. (A) Integration of multi-omics data and differential expression analysis. (i) Integration of data from literatures and databases (ExoBCD, GEO, and PRIDE). (ii)Differential expression analysis of 6010 differential exoMols. (B) Key exoMols screening based on the RCABD strategy. (i) Identification of hub exo-RNAs or exo-proteins was performed based on the PPI network, miRNA-mRNA network, and lncRNA regulation network. (ii) Functional enrichment analysis of 373 key exoMols. (C) Exosome-derived prognostic signature (exoSIG) established by survival analysis, tissue-exosome expression filtration, and LASSO analysis. (D) Multi-dimensional heterogeneity analysis of risk patients.

named it PAM 50 [9]. In 2017, triple-negative breast cancer (TNBC) was divided into basal-like immunosuppressive, immunomodulatory, mesenchymal, and androgen-dependent subtypes [10]. Although the molecular typing of breast cancer has continuously improved with the development of high-throughput sequencing technology, its classification is still controversial.

The deficiencies of tissue biopsy are barriers to accurate molecular stratification and disease typing of heterogeneous tumors. Tissue biopsy cannot achieve multi-point, real-time, and repeated sampling, which inevitably leads to sampling and detection bias, resulting in changes in heterogeneity and the ineffective characterization of tumors [11,12]. Ultimately, it affects the accuracy of the at-risk patient classification and prognostic assessment, leading to inappropriate treatment or wrong treatment decisions. Fortunately, tumor liquid biopsy technologies, such as the detection of circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), and tumor-derived exosomes (TDEs), have been developed as a novel diagnostic method to monitor tumor progression in a non-invasive manner [13–15]. Compared to CTCs and ctDNA, TDEs are more abundant, stable and provide a higher level of information, as they contain proteins, nucleotides, lipids, oligosaccharides and metabolites [16,17]. These biomarkers on exosomes can reflect the physiological state and disease progression of their source cells. Therefore, TDEs can generate molecular profiles of heterogeneous tumors regarding transcriptomic, genetic, and proteomic useful for diagnosis, dynamic monitoring, risk stratification, and prognostic evaluation [18,19]. And a comprehensive understanding of tumor heterogeneity in a biological context and its clinical application value based on TDEs will contribute to better management of BC patients to improve treatment response and prognosis.

Presently, some studies have focused on the identification of molecular biomarkers in BC exosomes for diagnosis and prognosis [20]. Stevic et al. found that the $HER2^+$ and TNBC patients had their own distinct exosomal miRNA expression profiles and specifically dysregulated exosomal miRNA networks [21]. 18 exo-miRNAs with diagnostic or prognostic value were identified, of which *miR-155* and *miR-301* showed excellent prognostic effects on pathological regression. Recent studies have revealed that aberrant miRNA expressions are closely associated with genomic instability (GI), such as genomic amplification or deletion, chromosomal translocation, microsatellite instability, high mutation level, and other GI events. Bao et al. identified three GI-derived signatures (miGISig) (*miR-421*, *miR-128–1*, and *miR-128–2*) through a comprehensive analysis of genome-wide miRNA expression profiles and somatic mutations of plasma exosomes, which enables early detection and prognostic risk stratification with minimal invasion for breast cancer [22]. Moreover, some other combined proteins located on the surface of or within exosomes, such as CD63 and *miR-21* [23], endothelial locus-1 protein (Del-1) and fibronectin [24,25], EpCAM and CD24 [26], as well as tetraspanin CD9, ADAM10, heat-shock protein HSP70 and Annexin-1 [27], could be detected as cancer biomarkers in serum, urine or pleural effusion-derived exosomes.

However, these non-multi-omics studies are limited to the single mapping of molecular profiles of TDEs and the identification of relevant markers for breast cancer [28]. There is still a lack of systematic analysis of stratified patients and their risk prognosis in a biological context of tumor heterogeneity based on exosome-derived factors. In particular, whether exosome-derived markers have higher sensitivity and specificity than traditional markers is a question we need to keep in mind [29]. Therefore, a robust corroborative analysis for biomarker discovery strategy (RCABD) was proposed to identify novel exosome-derived BC risk signatures. A 10-molecule panel, as the exosome-derived risk signature (exoSIG), was obtained based on RNA-seq and proteomic datasets (95 samples), 53 independent studies, and our ExoBCD database [30]. Our results suggest that the two of exosome molecules, *HLA-DQB2* and *COL17A1*, whose downregulation is significantly associated with BC metastasis and can form a clinical minimal exoSIG (Clinmin-exoSIG) combined with patient age and tumor stage for accurate prognostic prediction. In addition, exoSIG was found to have a high performance in stratified risk and prognosis prediction, which was characterized and witnessed by biological heterogeneities of risk patients in genomic alterations, functional enrichment, and immune and hypoxic microenvironment. In conclusion, a multi-dimensional analysis was carried out to identify prognostic exosome signatures, stratify risk patients, and characterize the BC heterogeneity. Our results provide a valuable reference for the in-depth exploration of the "biological heterogeneity - risk stratification - prognosis prediction".

2. Materials and methods

Based on our previous paradigm of the "data-driven and literature-based paradigm with corroborative analysis" [30], a multi-omics-based strategy and pipeline were proposed: the robust corroborative analysis for biomarker discovery (RCABD). It mainly included data collection, selection and analysis, robust molecular identification, machine learning-based risk model building, and systematic insights into the risk heterogeneity of BC patients (Fig. 1). The description was as follows:

2.1. Joint data-based differential expression analysis

Differentially expressed exosomal molecules (exoMols) were identified by integrating published results, high-throughput RNA-seq, and proteomic data (Fig. 1A) : (i) 481 and 306 breast cancer-associated exoMols were obtained from 53 independent studies (Table S2) and the ExoBCD database (https://exobcd.liumwei.org/Download), respectively; (ii) four microarray datasets [GSE60714 (n = 4), GSE60715 (n = 6), GSE93070 (n = 4), and GSE114329 (n = 18)], one RNA-seq dataset [SRP061372 (n = 8)], and one proteomic dataset [PXD012162 (n = 55)] of BC exosomes were retrieved from the GEO (http://www.ncbi.nlm.nih.gov/geo) and the PRIDE (https://www.ebi.ac.uk/pride) database, respectively (Table S3). After the format conversion, quality control, index comparison, and normalization, differential expression analysis was performed to identify key molecules in BC-associated exosomes.

2.2. Robust identification of key exoMols

Differentially expressed molecules were divided into key exo-RNAs or exo-proteins with multi-network-based analysis, including STRING PPI, tPOMA miRNA-mRNA, lncRNA-protein regulatory, and ceRNA (Fig. 1B).

2.3. Exosome-derived risk signature and risk stratification

The determination of exoSIG risk signature and the identification of corresponding molecule panels were realized by LASSO regression and multivariate Cox analysis (Fig. 1C). Risk stratification was performed using the exoSIG score calculated by the following formula:

$$exoSIG = \sum_i Coefficient(Candidate_i)^* Expression(Candidate_i)$$

2.4. Insights into biological heterogeneity of risk patients

Risk grouping of breast cancer patients was performed based on the exoSIG score. Then, biological heterogeneities between risk patients were compared in terms of tumor mutation burden, SNP, CNV, enriched function, immune filtration, and hypoxia state (Fig. 1D).

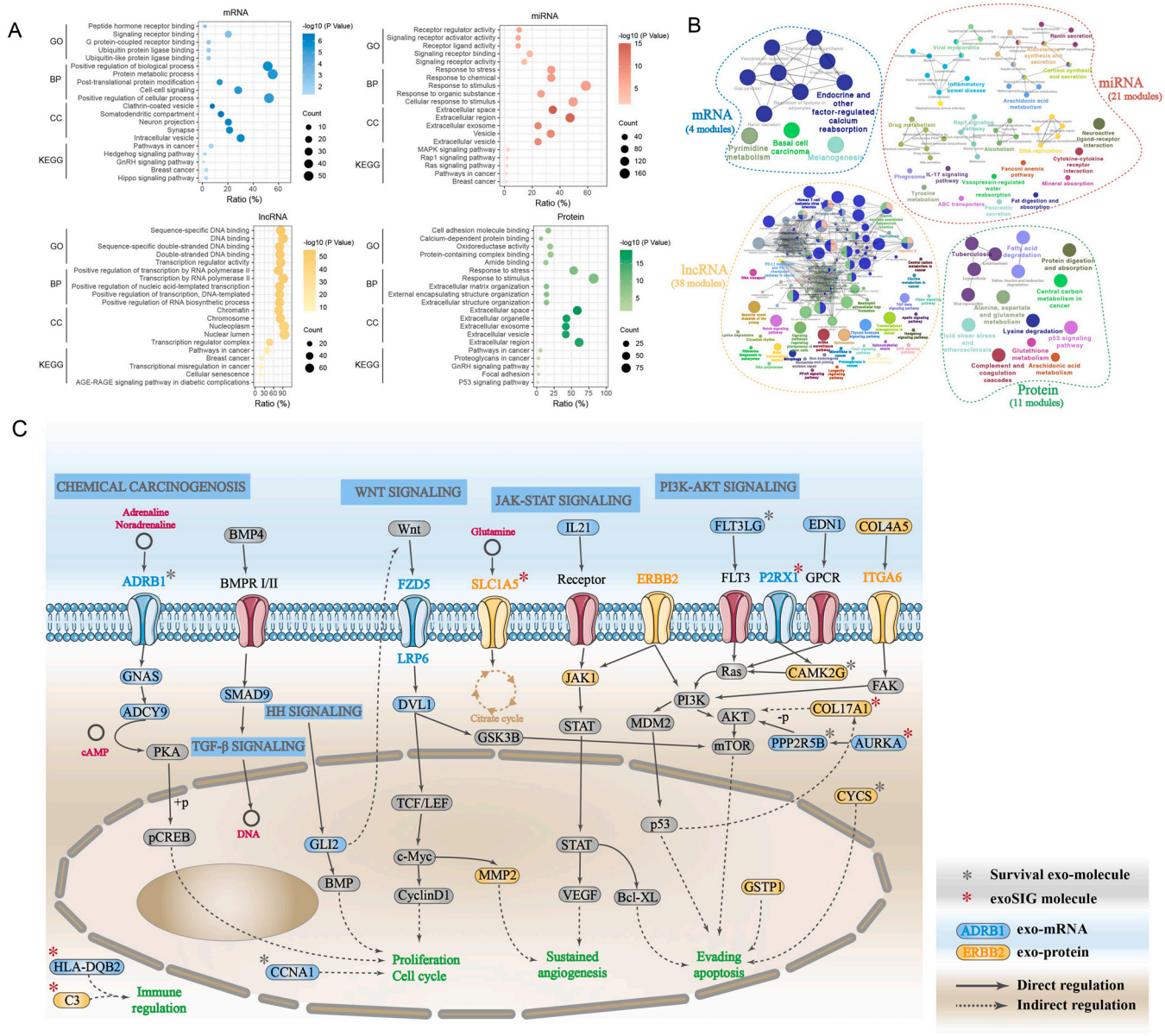

Fig. 2. Functional analysis of 373 key exoMols. (A) GO and KEGG enrichment analysis of key exoMols. (B) Interaction network mediated by key mRNAs, miRNAs, lncRNAs, and proteins in exosomes. (C) Signaling pathway networks of key molecules were generated by KEGG Mapper, including 16 exo-mRNAs (blue) and 11 exo-proteins (yellow).

Detailed descriptions of the above processes were presented in Supplementary Methods. And the involved tools and sources were included shown in Table S1.

3. Results

3.1. Functional overview on key exoMols

A total of 373 molecules were identified as key exoMols (Table S4, see the Supplementary Methods for details). GO & KEGG enrichment analysis revealed that they were associated with receptor recognition, stimulus response, transcriptional activation, metabolism, extracellular vesicles, and cancer-related pathways (Fig. 2A). Further modular analysis revealed that the regulatory networks downstream of key exoMols were significantly diverse (Fig. 2B, Table S5). Compared with exo-mRNAs (4 modules) and exo-proteins (11 modules), the pathways regulated by exo-miRNAs (21 modules) and exo-lncRNAs (38 modules) were more diverse. Further analysis revealed that exo-miRNAs are associated with mismatch repair and DNA replication. While, exo-lncRNAs are involved in the diabetes process, transcriptional disorder, and cancer pathways (such as the Notch and TGF-β signalling pathways). Exo-mRNAs are associated with endocrine pathways (such as insulin, renin and salivary secretion, and regulation of lipolysis in adipocytes). Exo-proteins are closely associated with the regulation of "p53 signalling pathway", "fatty acid degradation", "valine, leucine and isoleucine degradation", and "alanine, aspartate and glutamate metabolism" (Fig. S1).

A pathway map consisting of 27 key exoMols was obtained by KEGG Mapper analysis (Fig. 2C). ADRB, GNAS, and ADCY9 are important in chemical carcinogenesis. Norepinephrine (NE)-induced activation of the β2 adrenergic receptor 1 (ADRB1) has been shown to accelerate tumor growth via activating GNAS and ADCY9. The binding of BMP4 can activate the TGF-β signaling pathway to regulate DNA replication. In the canonical Wnt pathway, the binding of Wnt protein to the Frizzled receptor (FZD5) activates the Dishvelled (DVL1) and regulates cell proliferation. The activated JAK-STAT pathway by ERBB2 can trigger angiogenesis and cell apoptosis. In

addition, survival molecules, such as *FLT3LG*, *P2RX1*, *COL17A1*, and *AURKA*, were enriched in the PI3K-Akt signaling pathway.

4. Exosome-derived risk signature (exoSIG)

For the indicative effect of the above 373 key exoMols on patient survival, survival analyses were performed, and 34 key exoMols were identified, including four mRNAs, ten miRNAs, three lncRNAs, and 17 proteins (Table S6). Validated expression of 16 of 34 key exoMols was further analyzed to identify those highly correlated with patient survival based on TCGA. Subsequently, a risk signature panel consisting of ten molecules was obtained by LASSO regression analysis [31] (Table S7). Finally, the exoSIG multivariate Cox survival model was constructed based on the obtained panel:

$$\begin{aligned} exoSIG &= 0.022{*}EXP_{miR\_324\_5p} + 0.001{*}EXP_{miR\_99a\_5p} - 0.062 \\ &{*}EXP_{HLA\_DQB2} - 0.025{*}EXP_{P2RX1} - 0.012{*}EXP_{AURKA} + 0.136 \\ &{*}EXP_{RTCA} - 0.031{*}EXP_{LINC01055} - 0.011{*}EXP_{SLC1A5} - 0.063 \\ &{*}EXP_{COL17A1} - 0.017{*}EXP_{C3} \end{aligned}$$

The Kaplan-Meyer analysis [32] showed that the above 10 molecules were all strongly associated with the overall survival (OS) in BC patients (Fig. S2). Among them, *HLA-DQB2* was strongly associated with the survival (Log-rank $P = 0.0029$, $HR = 0.91$, 95% *CI* 0.84–0.99), then followed by *COL17A1* (Log-rank $P = 0.015$, $HR = 0.92$, 95% *CI* 0.87–0.97). Survival analysis based on the exoSIG score showed that high-risk patients had a worse survival prognosis than low-risk patients (discovery dataset: Log-rank $P = 0.0079$, $HR = 2.9$, 95% *CI* 1.5–5.8; test set: Log-rank $P = 0.011$, $HR = 3.1$, 95% *CI* 1.5–6.2) (Fig. S2).

Univariate Cox analysis showed that exoSIG, AJCC stage, patient age, metastasis coded, and node coded were significantly associated with patient survival, among which metastasis coded had the highest hazard ratio (Cox $P = 1.6E\text{-}07$, $HR = 5.9$, 95% *CI* 3–12) (Fig. 3A). Multivariate Cox analysis showed that, except for node, the remaining four factors were significantly associated with the patient survival. Among them, exoSIG had the highest hazard ratio (Cox $P = 9.9E\text{-}04$, $HR = 3.3$, 95% *CI* 1.6–6.8) (Fig. 3B).

Clinical survival features of high-risk patients were worse than those of low-risk patients, such as estrogen-receptor (ER) status (Chi-square test, $P < 0.001$), progesterone-receptor (PR) status (Chi-square test, $P < 0.001$), node coded (Chi-square test, $P < 0.05$), AJCC stage (Chi-square test, $P < 0.001$), 50-gene signature (PAM50) (Chi-square test, $P < 0.001$) and age (Student T test, $P < 0.01$) (Fig. 3C). High-risk patients with ER^-, PR^-, high stage, PAM50-LumB, or PAM50-Basal subtypes had a higher ratio than low-risk patients (Fig. 3D). Concurrently, patients with ER^-, PR^-, high stage, PAM50-LumB, or PAM50-Basal subtype had higher exoSIG score (Fig. 3E).

5. Tumor risk heterogeneity in patients

5.1. Genomic changes

Higher non-synonymous tumor mutation burden (TMB) was found in protein-coding regions of the genome of high-risk patients (Fig. 4A). Seventeen of the top 25 genes with the highest mutation frequency (3–41%) were presented in both risk patients (Fig. 4B). Interestingly, the opposite frequency was observed for *PIK3CA* (high / low risk, 27% / 40%) and *TP53* (high / low risk, 41% / 26%). And eight mutation types are shared, of which the nonsense mutation is the most dominant in patients, while nonstop mutation occurs in the *GATA3* gene of the high-risk ones. SNP scanning of gene coding regions revealed that C > T mutation was common for BC patients, but the frequency of C > G mutation was higher in the high-risk patients (Fig. 4B). Mutually exclusive mutations were detected in *TP53-GATA3*, *TP53-MAP3K1*, and *TP53-CDH1* in both risk patients. However, the co-mutation of *PIK3CA-MAP3K1* and *PIK3CA-KMT2C* was unique for the low- and high-risk patients, respectively (Fig. 4C). Mutations enriched in the DNA binding domain (Fig. 4D) of the corresponding protein may be an important reason for the deterioration of tumor suppression efficiency and the decrease in patient survival (Log-rank $P = 0.042$) (Fig. 4E).

CNV analysis showed that high-risk patients had more pronounced CNV events (Fig. 4F), and copy-number amplifications were significantly enriched in oncogenes (*MDM2*, *MYC*, *MYBL1*, and *ERBB2*) and immune-related genes (*CD24 and CD79B*) (Fig. 4G). Deletions were found in tumor suppression genes (*NFRSF10A/B/C/D*, *RB1*, and *TNFSF11*) and other immune-related genes (*PDCD1* and *LCP1*). Moreover, genes with significant copy number variation were also closely associated with many cancer-related biological processes/pathways, such as mitochondrial electron transport chain (*SDHC*), histone modification (*CBX2/4/8*), WNT signalling pathway (*FZD3*, *APC2*), cell adhesion (*NECTIN4*, *PECAM1*) and lipid transport (*APOA2*) (Fig. 4G).

5.2. Differences in immune capacity and cell proliferation of risk patients

Gene expression analysis showed that most of the genes (top 50 differentially expressed genes between two risk groups) were downregulated in the high-risk patients (Fig. 5A), including *COL17A1*, *COL14A1*, and *MFAP4* related to extracellular matrix and *SOX17*, *GLI1*, *TNN* involved in the negative regulation of WNT signalling pathway (Fig. 5B). GSVA analysis showed that several immune-related terms were highly enriched in low-risk patients, such as dendritic cell apoptotic process, regulation of natural killer cell differentiation, MHC class II protein complex, T cell receptor complex, and immune receptor activity (Fig. 5C). However, cell proliferation-associated terms were enriched in the high-risk patients, such as mismatch repair and DNA replication (Fig. 5C). The result of GSEA analysis revealed that 14 signaling pathways were generally inhibited in high-risk patients, including inflammatory response, interferon-gamma response, and TNFA signalling via NFKB (Fig. 5D). In conclusion, compared with the low-risk patients, immunosuppression and increased cell proliferation may be important contributing factors to the poor survival outcomes.

5.3. Heterogeneity of immune infiltration and changes in the hypoxic microenvironment

Immune infiltration analysis showed that the proportion of 28 types of immune cells significantly decreased in high-risk patients (Wilcoxon test, $P < 0.05$) (Fig. 6A), suggesting that they had lower levels of immune cell infiltration. Notably, the immune infiltration level of M1 macrophages was significantly reduced in high-risk patients, while that of M0 and M2 macrophages increased (Wilcoxon test, $P < 0.05$) (Fig. 6B & C). In addition, immune subtype analysis found that more low-risk patients were included in the C3 inflammatory subtype characterized by increased Th1/Th17 expression [33]. However, more high-risk patients were included in C4 lymphocyte-depleted subtype with low Th1 expression and high M2 infiltration. According to tumor microenvironment subtyping standard described by Bagaev et al. [34], correspondence analysis showed that more high-risk patients were the immune desert subtypes, while low-risk patients were IE/F (immune-enriched, fibrotic) subtypes (Fig. 6D). The ESTIMATE score also supported that the low-risk patients were the IE/F subtype with a higher stromal score and stromal cell infiltration (Fig. 6E). Hypoxia-responsive gene expression analysis revealed that high-risk patients had a higher hypoxia score (Fig. 6F & G, Table S8). These results strongly demonstrated the

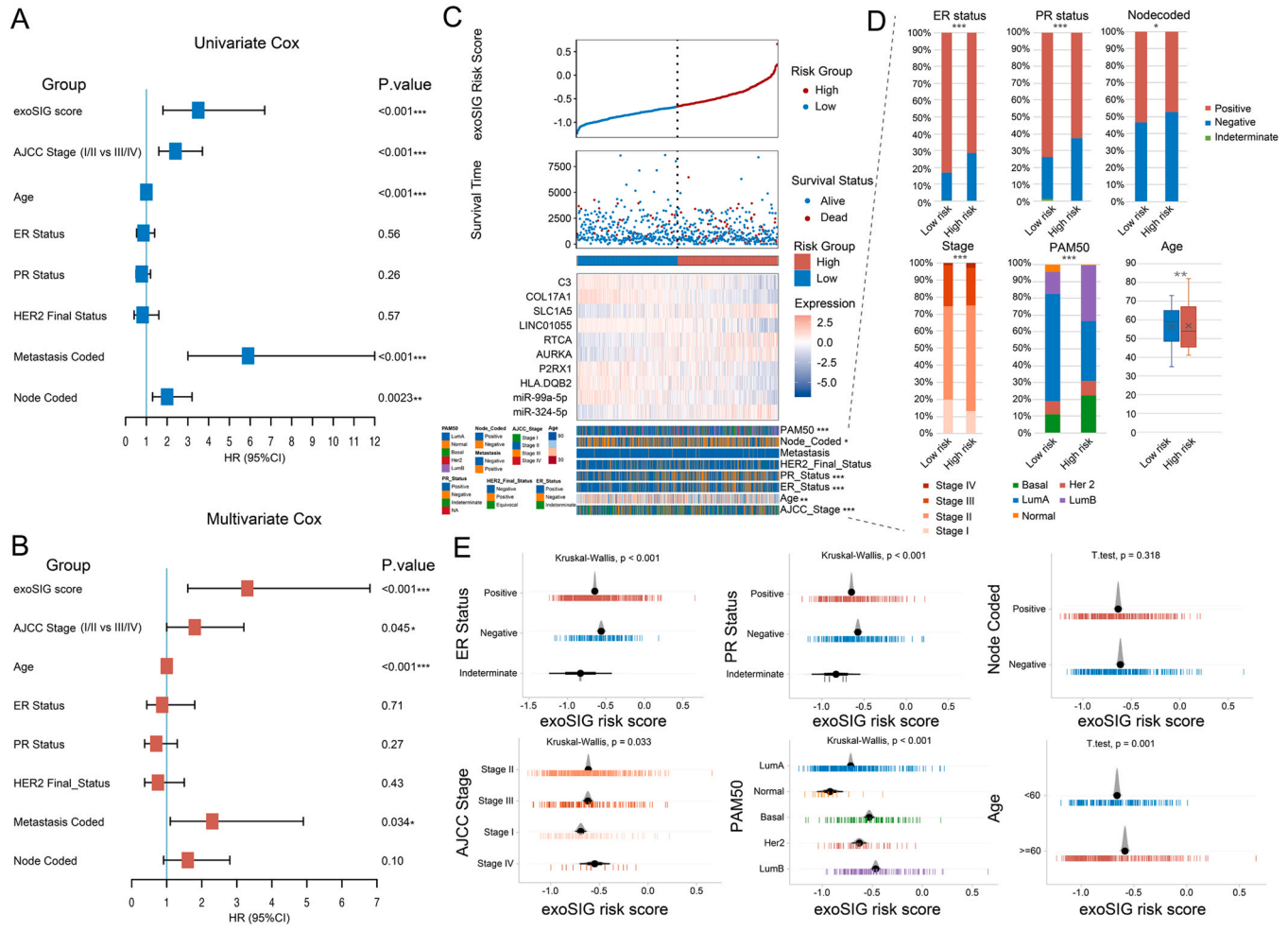

Fig. 3. Association of exoSIG risk score with characteristics of breast cancer patients. (A & B) Univariate and multivariate cox hazard ratio (HR) distributions of exoSIG risk scores versus seven clinical characteristics. (C) Risk grouping based on exoSIG risk scores. Up-panel: correlation between risk grouping and risk score; Middle-panel: correlation between survival status and survival time; Low-panel: Heatmap of expression patterns of 10 exoMols in risk patients. (D) Distribution of different clinical characteristics in the low-risk and high-risk patients. (E) Correlation of clinical characteristics of breast cancer patients with exoSIG risk scores. Breast cancer patients with ER-, PR-, high stage, high age, PAM50-LumB, and PAM50-Basal subtype had higher exoSIG risk scores.

negative correlation of the hypoxic microenvironment with immune cell infiltration.

5.4. Differential expression of immune-related molecules

Several immune chemotactic factors (*CCL4*, *CCL5*, *CXCR3*, *CXCL9*, and *CXCL10*) [35] have been studied to understand the differences in immune infiltration-related molecules among risk patients. The results showed that their expression significantly decreased in the high-risk patients compared with the low-risk ones (Fig. 6H). There were, however, significant increases of expression of five immune checkpoint genes (*CTLA4*, *LAG3*, *PDCD1*, *CD274*, and *TIGIT*) [36] in low-risk patients (Fig. 6I). And a significant up-regulation of T cell dysfunction-associated genes (*TGFB1* and *SERPINB9*) and a significant down-regulation of *CCDC43* and *TRIT1* [37] were also observed (Fig. 6J). In addition, tumor immune dysfunction and exclusion (TIDE) score [37] in low-risk patients was significantly higher than that of high-risk patients (T-test, $P < 0.0001$), suggesting a poor efficacy of immunotherapy (Fig. 6K). The ESTIMATE analysis revealed that low-risk patients had a higher stromal score (Fig. 6E), consistent with the TIDE prediction.

In conclusion, there were significant differences in the expression profiles of immunochemokines and immune checkpoint-related genes among risk patients. These heterogeneities may contribute to more precise tumor immunotherapy. However, immune escape and stromal infiltration in low-risk patients may be the main obstacles to immunotherapy.

5.5. Response to targeted drugs varies by risk group

Exploring chemotherapy/targeted drugs for patients in different exoSIG risk groups may help to prolong patient survival. We further performed a drug response profile analysis based on 198 GDSC chemotherapy/targeted therapeutic drugs for exoSIG risk groups. It was found that most of the drugs were more sensitive to patients in the low exoSIG risk group (Fig. S3A), and patients in the low-risk group were sensitive to gemcitabine, mitoxantrone, teniposide, and camptothecin (Fig. S3B), while patients in the high-risk group were sensitive to axitinib, tozasertib, lapatinib and sapitinib (Fig. S3C). We further investigate the expression of exoSIG molecules between lapatinib drug-tolerant persisters (DTP) and parental (P) BC cell lines from GSE155341. *COL17A1* demonstrated high expression in DTP cells while *RTCA*, *AURKA* and *SLC1A5* decreased (Fig. S3D), which echoed their expression patterns of two risk groups. Personalized drug treatment strategies for patients at different exoSIG risk groups will further improve their survival. Concurrently, we predicted 9 small molecule drugs, including YC-1, EMD-1214063, naphazoline, lisinopril, and BRD-series (K22478004, K17893805, K05096562,

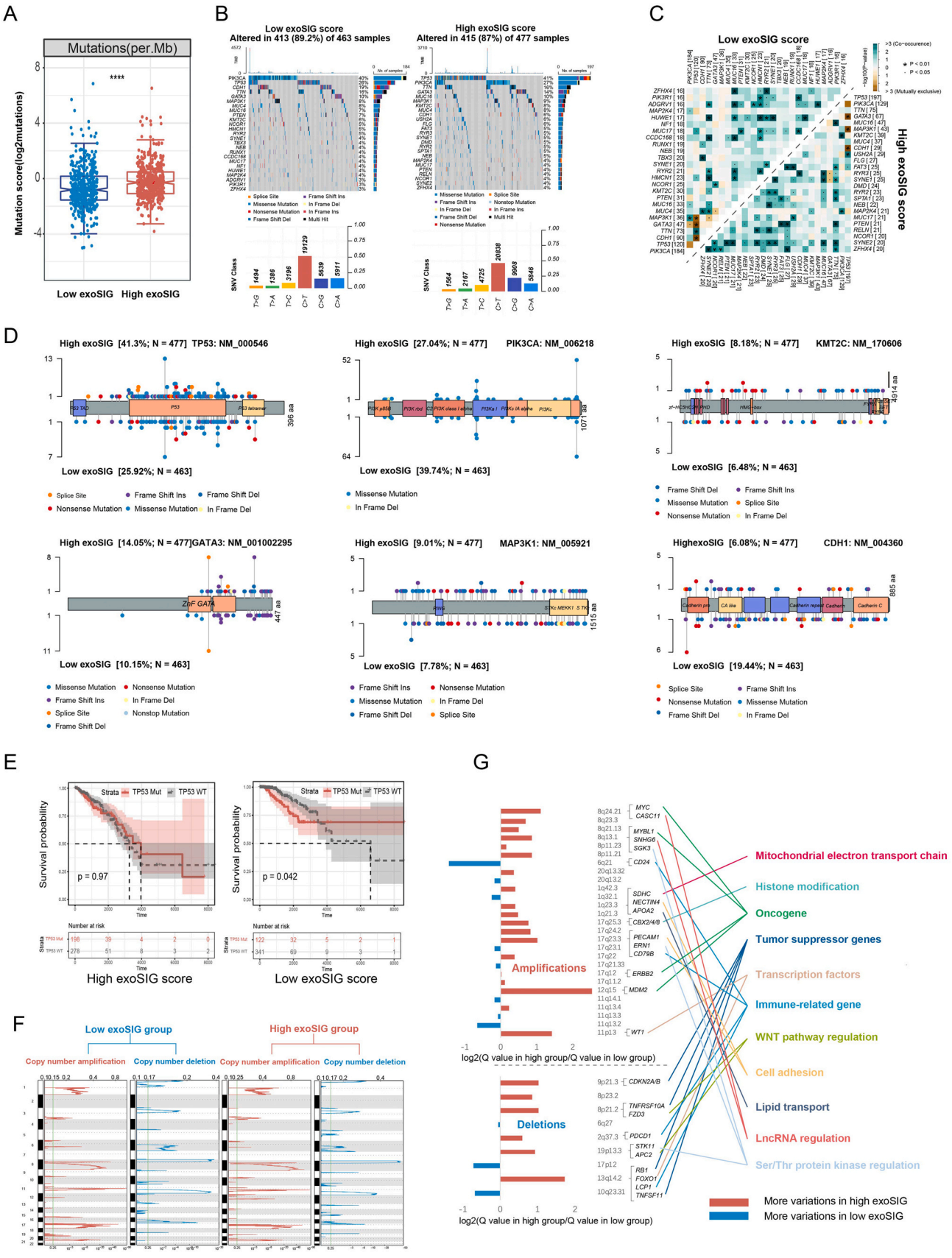

Fig. 4. Genomic alterations in risk patients with low exoSIG or high-exoSIG score. (A) Comparison of tumor mutation burden (TMB). (B) Oncoplot of mutation, deletion, insertion, and frameshift. (C) Gene co-occurrence and mutually exclusive mutation of genes associated with breast cancers. (D) Comparison of different mutation sites of *TP53*, *PIK3CA*, *KMT2C*, *GATA3*, *MAP3K1*, and *CDH1*. (E) Relationship between mutations in *TP53* and patient survival. (F) Copy number variation (CNV) patterns in different risk cohorts. (G) Association of CNV with signaling pathways. Genes with high-level CNV were highly correlated with cancers.

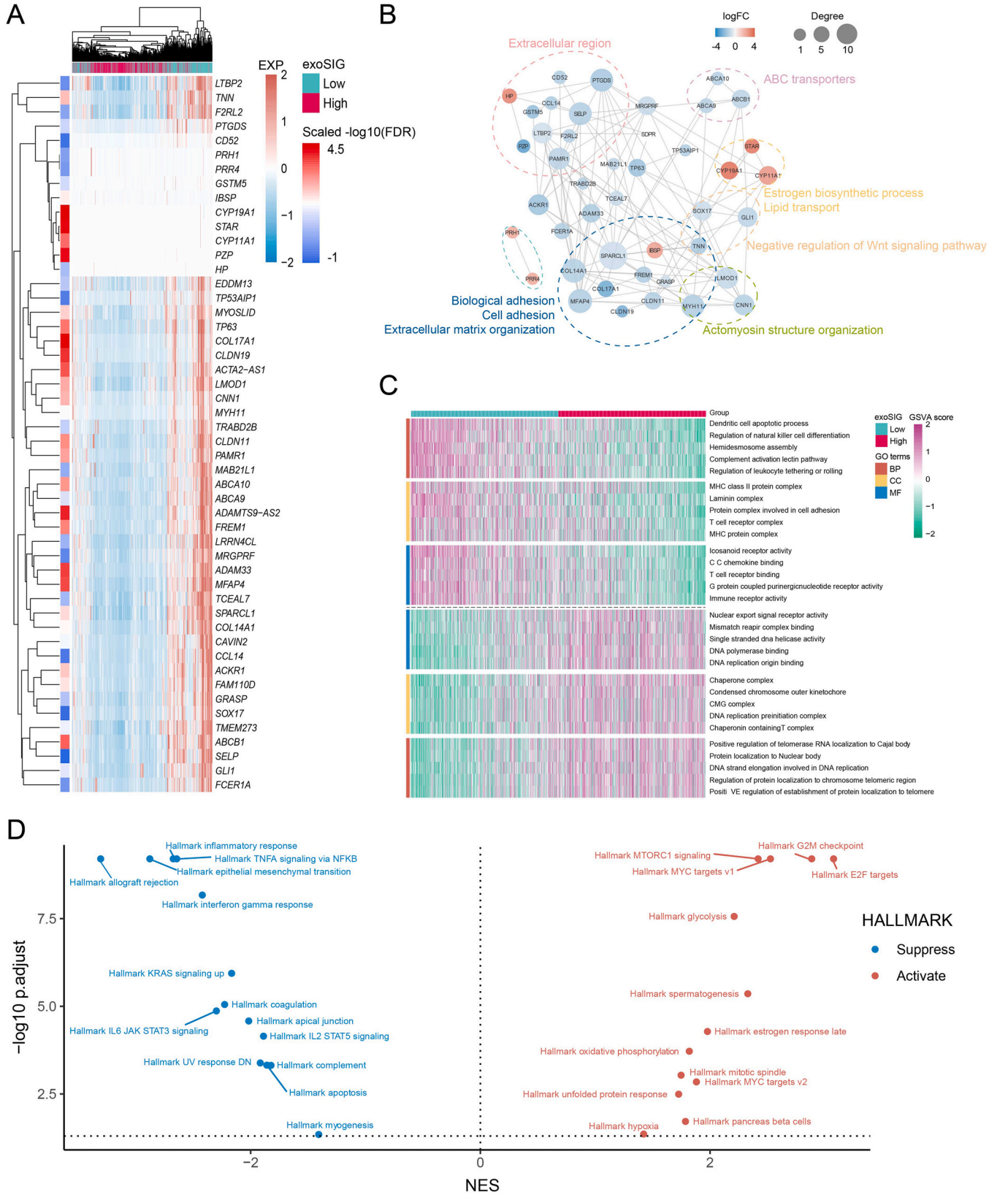

Fig. 5. Functional analysis of differentially expressed genes (DEGs) in low and high-risk patient groups. (A) Heatmap of the top 50 DEGs with the highest fold change. (B) PPI network of top 50 DEGs based on the STRING website. (C) Heatmap of enriched GO terms by GSVA. Immune-related terms were highly enriched in low-score patients, but not tumor proliferation-associated terms. (D) Volcano plot of enriched cancer hallmarks by GSEA. 14 hallmark pathways were significantly suppressed in the high-score patients, while 13 pathways were activated ($P < 0.01$, $FDR < 0.05$).

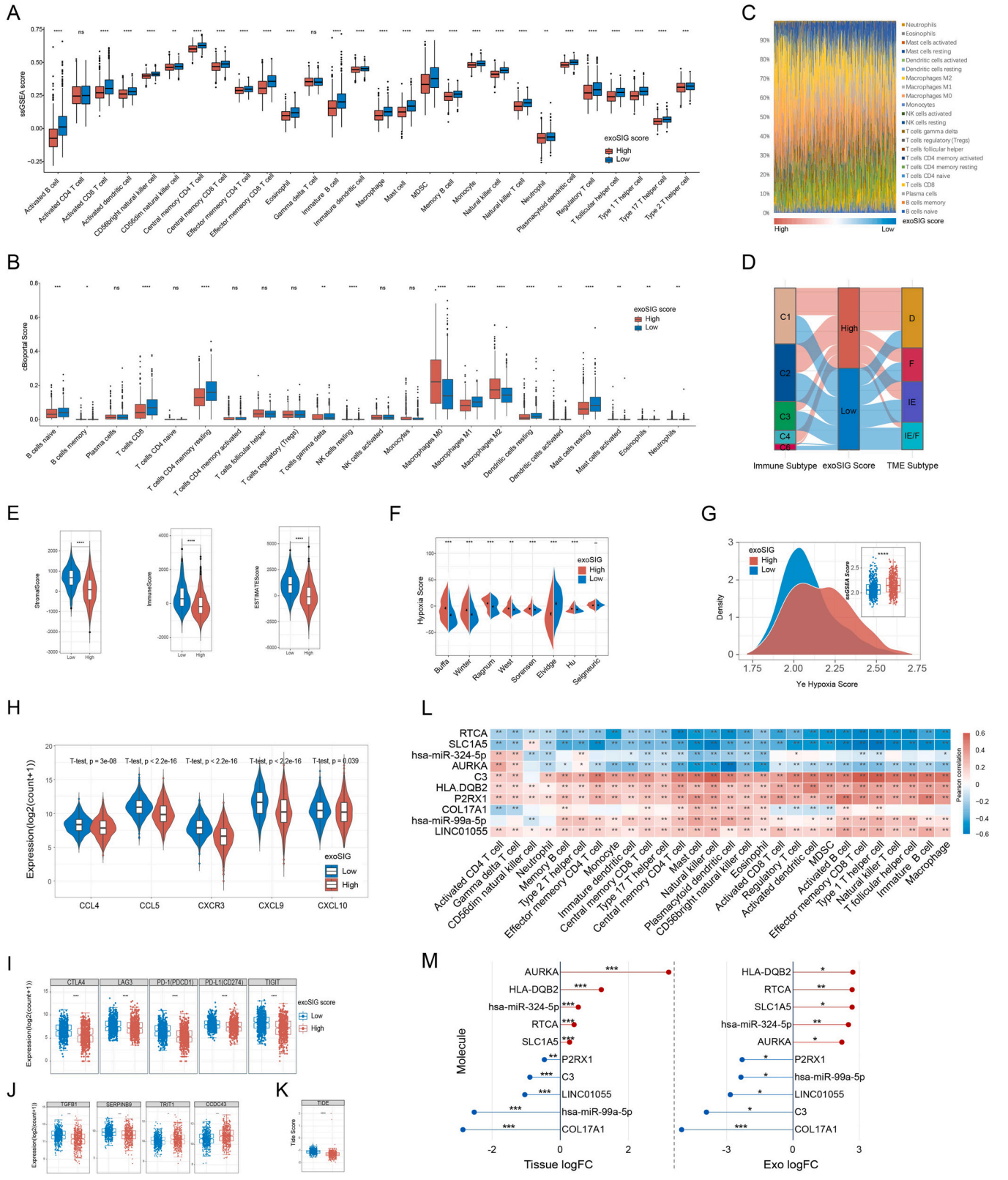

Fig. 6. Immune infiltration and tumor microenvironment (TME) analysis of low and high-risk patients. (A) The boxplot of 28 infiltrated immune cell types was calculated by ssGSEA. (B) The boxplot of 22 infiltrated immune cell types was calculated by CIBERSORT. (C) Distribution of 22 infiltrated immune cell types. (D) The Sankey diagram visualized the association between risk patients and immune subtypes, and TME subtypes. (E) The violin plot shows lower immune infiltration, stromal, and ESTIMATE score in high-risk patients (T-test; ****, $P < 0.0001$). (F & G) Violin and density plot of significantly increased hypoxic score in high-risk patients (Wilcoxon test; $^{*}P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$). (H) Violin plot of expression levels of antitumor-related immune genes in high-risk patients. (I) Box plot of expression levels of immune checkpoint-associated genes. (J & K) Box plot of expression levels of immune escape-related genes and TIDE scores. (L) Heat map of associations of 28 immune cell types with 10 exoSIG molecules (*, $P < 0.05$; **, $P < 0.01$). (M) Differential expression levels of 10 exoSIG molecules in tissues and exosomes in health and tumor patients of breast cancer.

K68988758, K52340217) to reverse the gene expression profile of patients at high exoSIG risk group based on the CLUE database (Fig. S3E). Moreover, HIF1A, MET and DRD2 were identified as potential targets for the small molecule drugs YC-1, EMD-1214063 and naphazoline via molecular docking, respectively ((Fig. S3F-H)).

6. Relationship between the immune heterogeneity and prognosis based on exoSIG molecules

Given the association between exoSIG scores and immune infiltration, the potential role of ten exoSIG molecules in the immune microenvironment was further explored. With increasing exoSIG score, six molecules (*HLA-DQB2*, *COL17A1*, *LINC01055*, *C3*, *P2RX1*, and *miR-99a-5p*) were down-regulated, while the other four molecules (*AURKA*, *SLC1A5*, *RTCA*, and *miR-324–5p*) were up-regulated (Fig. 3C).

Subsequently, the correlation of gene expression profiles and immune cell abundance was assessed and revealed that four exoSIG molecules (*RTCA*, SLC1A5, *miR-324–5p*, *and AURKA*) were highly negatively correlated with the abundance of most immune infiltrating cells. However, the other six molecules (*HLA-DQB2*, *P2RX1*, *COL17A1*, *miR-99a-5p, LINC01055*, and *C3*) showed the opposite relationship (Fig. 6L & M). Notably, the UALCAN analysis [38] confirmed that only *HLA-DQB2,* a hypomethylated gene, was positively correlated with immune cell infiltration (Fig. S4A). The tumor purity of BC patients with high *HLA-DQB2* expression (Wilcoxon test, $P < 0.0001$) was lower, and its expression level decreased with cancer stage (Fig. S4B-F). In addition, high expression of *HLA-DQB2* had a positive association with the abundance of monocytes and macrophages in cancer patients (Fig. S4G-I). Further analysis showed that several immune-related signalling pathways or entries were enriched, such as the T-cell antigen receptor (TCR) signalling pathway (*NES* = 2.84, *FDR* = 5.76E-09), chemokine signalling pathway (*NES* = 2.75, *FDR* = 5.76E-09), and immune cells-microRNAs interactions in tumor microenvironment (*NES* = 2.45, *FDR* = 8.17E-07) (Fig. S4J). These results were well supported by the survival analysis (Fig. S2) and provided evidence that *HLA-DQB2* might be a player in tumor suppression by activating monocytes and, or macrophages.

Taken together, the above ten exoSIG molecules have their characteristics in tumor promotion or tumor suppression, which fully reflects a closer relationship between the heterogeneity in tumor immune and patient prognosis. Among them, *HLA-DQB2* (Log-rank test, $P = 0.0029$) and *COL17A1* (Log-rank test, $P = 0.015$) were the most prominent candidates with a high clinical prognostic value (Table S7, Fig. S2).

7. Bimolecular clinical prognostic signature - Clinmin-exoSIG

Further analysis showed that the prognostic result of minimal exoSIG (min-exoSIG) (*COL17A1 + HLA-DQB2*) was in good agreement with that of 10-molecule exoSIG (86.35%) (Fig. S5). The fitted formula for min-exoSIG was:

$$\min\_exoSIG = -0.078^{*}EXP_{COL17A1} - 0.075^{*}EXP_{HLA\_DQB2}$$

As shown in Fig. 7A-C, min-exoSIG showed a good prediction performance in the overall survival (OS) evaluation (discovery data: Log-rank $P = 0.0079$; test dataset: Log-rank $P = 0.022$; validation dataset: Log-rank $P = 0.021$). In addition, min-exoSIG also has a good performance in the analyses of RFS, DRFS, DMFS based on dataset GSE1456 (Log-rank $P = 0.022$), GSE22219 (Log-rank $P = 0.020$) and GSE11121 (Log-rank $P = 0.018$), indicating that it has the value of predicting the recurrence and metastasis of breast cancer (Fig. 7D-F).

To further increase the accuracy of prognosis, a combined prognostic signature (Clinmin-exoSIG) was constructed by integrating min-exoSIG, AJCC stage, and age of BC patients. The results revealed that the Clinmin-exoSIG had a good prognosis in the discovery dataset (Log-rank $P = 0.0027$), test dataset (Log-rank $P = 0.0036$) and validation dataset (Log-rank $P = 0.028$) (Fig. 7G-I). The fitted formula was:

$$Clinmin\_exoSIG = 1.114^{*}\min\_exoSIG \quad score + 0.935^{*}stage + 0.031^{*}age$$

Time-dependent ROC (tdROC) analysis found that the Clinmin-exoSIG had the best prognostic performance, compared with min-exoSIG, AJCC stage, age, and the commonly used prognostic indicators (CEA, CA153, and HER2). The average AUC value for tdROC analysis of the Clinmin-exoSIG was 72.89% (discovery dataset), 72.62% (test dataset), and 85.42% (validation dataset), respectively (Fig. 7J-L). Moreover, four external signatures that were assessed as having good prognostic performance were compared with Clinmin-exoSIG, which showed that Clinmin-exoSIG had a high AUC, although Oncotype DX demonstrated the best predictive performance (Fig. S6A-D). Concurrently, we found that Clinmin-exoSIG also has prognostic predictive potential in pan-cancer (Fig. S7A-C). For the facilitation of clinical application, a nomogram was generated to predict 3-, 5-, 7-, and 10-year overall survival by the comprehensive analysis of Clinmin-exoSIG, age, AJCC stage, ER status, PR status, and HER2 status (Fig. S8A). The C-index of this model was 0.756, and the calibration curve confirmed that the predicted results were consistent with the actual survival (Fig. S8B).

8. Discussion

Tumor heterogeneity remains a major challenge for accurate subtyping, risk stratification, and effective treatment [39], which urgently requires a systematic dissection of tumor heterogeneity in a biological context [40,41]. Tissue biopsy is currently the gold standard for molecular analysis and characterization of breast cancer, but its inherent defects, such as single-point invasive surgery and discontinuous detection, seriously affect the diagnostic accuracy of tumor heterogeneity. Exosome-based liquid biopsy enables non-invasive, reproducible, and spatiotemporal dynamic molecular monitoring, and has become a new technology for early diagnosis, prognostic stratification, and precise treatment [42]. In this study, the identification of prognostic molecules in exosomes and the construction of a risk model were performed based on multi-omics data on breast cancer, and the risk stratification and risk heterogeneity of stratified patients were performed.

In-depth molecular characterization of tumor heterogeneity is an effective means of patient risk stratification. Our findings suggested that high-risk patients belong to a type of PAM50-LumB/Basal [HR^{+}(ER^{+}/PR^{+})/HER^{+} or TNBC] subtype, lymphocyte depleted (C4) immune subtype (Th1 suppressed and high M2 response), or immune desert subtype, comparing with the low-risk ones (Table S9). High risk ones are also characterized by high TNM stage and enriched in cancer-promoting factors: i) A higher mutation frequency was found in tumor suppressor genes (e.g. *TP53*), but lower in tumor-promoting genes (e.g. *PIK3CA*). ii) Expression levels of genes involved in immunity, immune chemotaxis, cell epidermal adhesion, and immune checkpoint were significantly downregulated, while expression levels of genes associated with hypoxia were significantly up-regulated. iii) Infiltration of immune cells decreased remarkably, whereas that of macrophages M0 and M2 macrophages were increased. These biological carcinogenic factors may be the root cause of high-risk patients' relative sensitivity to pathway-related enzyme inhibitors (such as axitinib, tozasertib, lapatinib, and sapitinib etc.) and their poor prognosis. On the contrary, patients with low exoSIG scores showed prominent tumor suppressor factors, including ER+, PR+, low stage and PAM50-Luminal/Normal subtypes, and were sensitive to DNA replication, genomic mutations, and cell cycle inhibitors (such as gemcitabine, mitoxantrone, teniposide, and camptothecin, etc.). Moreover, functional insights of the 10-molecule

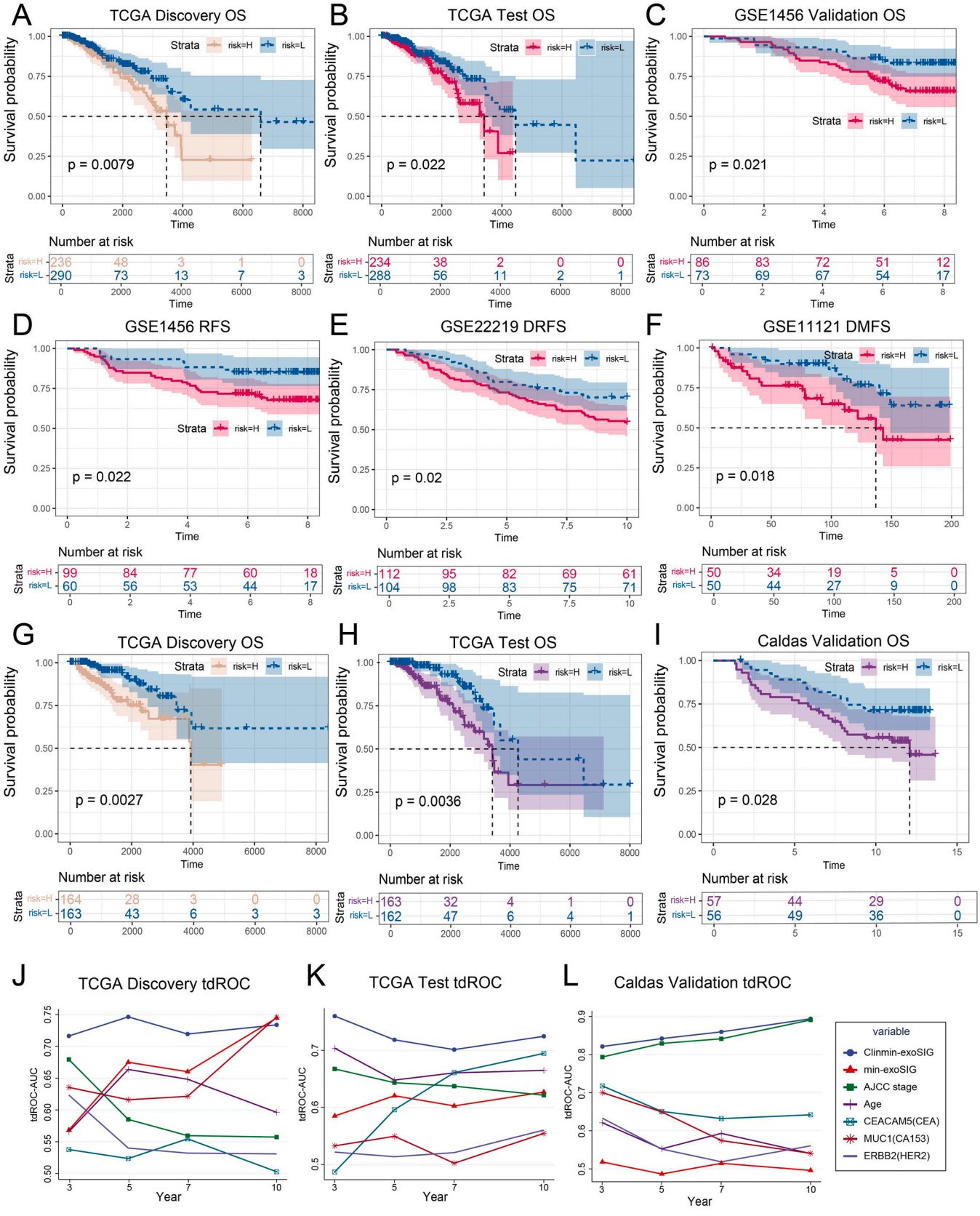

Fig. 7. Survival analysis based on min-exoSIG and Clinmin-exoSIG signatures. (A-C) Overall survival analysis based on min-exoSIG signatures from discovery-, test- and validation-dataset. (D-F) Survival analysis based on RFS, DRFS, and DMFS calculated by min-exoSIG signatures. (G-I) Overall survival analysis based on Clinmin-exoSIG signatures. (J-L) Time-dependent receiver operating characteristic (tdROC) curve of 3-, 5-, 7-, 10-year overall survival.

prognostic panel and risk prediction modeling proved that exoSIG can be used as a novel TDEs-based indicator for BC patient stratification. Especially, this prognostic panel contains two functional molecular groups: tumor-promotion factors (TPF) (*miR-324–5p*, *AURKA*, *RTCA*, and *SLC1A5*), and immune-related molecules (IRM) (*miR-99a-5p*, *LINC01055*, *P2RX1*, *HLA-DQB2*, *C3*, and *COL17A1*). Among TPF, The *AURKA* gene encodes a serine/threonine kinase, which belongs to the Aurora kinase family. AURKA is involved in centrosome replication, segregation, and maturation, and plays an important role in the mitotic cell cycle of eukaryotes [43]. The abnormally expressed *AURKA* often leads to the instability of the genome, increasing the frequency of gene mutations, and can participate in various intracellular signaling pathways directly or indirectly promoting tumorigenesis, which makes it be one of the reasons for the occurrence of malignant tumors [44]. *SLC1A5* (also known as ASCT2), coding as solute carrier family 1 member 5 located in the plasma membrane, is an important amino acid transporter highly expressed in various tumor tissues and cells, such as breast cancer [45] and lung cancer [46], and responsible for transporting glutamine for the nutritional requirements when inflammation, high proliferation and metabolic reprogramming in stem and cancer cells [47]. In addition, studies also found that the high expression of *SLC1A5* is closely associated with the biological characteristics and poor prognosis of tumor malignancy [48,49], so it can be used as an important prognostic indicator for predicting malignant tumors. Among IRM, *miR-99a-5p* has been found to suppress γδ T cell activation and their cytotoxicity to tumor cells, and rapidly respond to tumorigenesis, which has implications for interventional approaches to γδ T cell-mediated cancer therapy [50]. Ma et al. reported that the expression of *P2RX1* (a member of the ATP-gated ion channel receptor family) was closely related to the prognosis of lung adenocarcinoma and the activation of mast and B cells in the microenvironment [51]. In pancreatic cancer liver metastases TME, the infiltration of *P2RX1*-negative neutrophil subsets increases and mediates the metastasis of cancer cells by expressing PD-L1 immunosuppressive molecules [52].

As the functions on cancer reviewed above, these two groups of molecules showed their great value in monitoring at-risk patients. Our findings manifested that the expression of TPF and IRM increased and decreased in high-risk patients with poor prognosis, respectively, while there were the opposite expressions in low-risk patients. Particularly in IRM, *HLA-DQB2* and *COL17A1* are biomolecules closely related to immune regulation and tumor metastasis, whose low expression levels were strongly associated with the poor prognosis of high-risk patients. Being a member of the MHC class II family, HLA-DQB2 is an important molecule involved in the T cell immune response and antigen presentation. It also plays an important role in immunotherapy with anti-PD-1 and anti-CTLA-4 antibodies and the treatment of TNBC patients with progression-free survival (PFS) [53,54]. COL17A1 (collagen type XVII alpha 1 chain) plays a critical role in maintaining the link between the intracellular and the extracellular structural elements involved in epidermal adhesion. And the aberrant expression of *COL17A1* has been reported in epithelial, squamous, and colorectal cancers [55–57]. In breast cancer, it was also found that *COL17A1* is suppressed by DNA methylation in the promoter region and inactivation of p53 [57,58], and the high expression of *COL17A1* inhibits the cell growth and proliferation through deactivation of the AKT/mTOR signalling pathway, which is closely associated with the overall survival of breast cancer patients [59]. Taken together, combining with the clinical information, the bimolecular risk signature Clinmin-exoSIG has a high prognostic accuracy (validation dataset : Log-rank $P = 0.028$, AUC = 85.42%). And the generated nomogram has a good clinical application value.

In this study, we proposed a strategy of robust corroborative analysis for biomarker discovery (RCABD), which includes i) joint-data-based differential expression analysis and identification of risk signatures. Coinciding with this strategy, Qiu et al. identified three prognostic exosome molecules (*IVL*, *CXCL13*, and *AP2S1*) from our ExoBCD database to successfully perform TNBC risk signature discovery and prognostic evaluation of tumor immune microenvironment [60]. ii) comprehensive multi-approach-based analysis, including multi-network-based identification, consistent DE validation, survival & risk model prediction, and multiple large cohort verification strongly supports the robust identification of prognostic panel and exoSIG, as well as comprehensive insights into risk heterogeneity at the level of the genome, transcriptome, immunity, and hypoxic microenvironment providing the strong evidence for the stratification of risk patients. Personalized dosing strategies for patients at different risk will further improve their survival.

As an extension of ExoBCD, this study attempts to address the issues related to tumor heterogeneity and exosome-based tumor liquid biopsy. We also achieved stratification prediction and risk assessment related to breast cancer exosomes, which lays a foundation for the ExoBCD update. In addition, our results also pointed out the directions for follow-up studies on the molecular mechanisms of *HLA-DQB2* and *COL17A1* in tumor immune regulation and metastasis, as well as functional exploration of key exoSIG molecules (*P2RX1*, *AURKA*, *RTCA*, *miR-99a-5p*, *miR-324–5p*, *LINC01055*, and *SLC1A5*) in the immune microenvironment or in promoting tumor progression. Of course, the limitations of this study are the limited number of publicly available breast cancer exosome datasets. In addition, although Clinmin-exoSIG has the potential advantage of achieving convenient clinical prognostic prediction with "minimal molecular detection", it still needs to be validated in a larger clinical cohort. In future work, exosome-based molecular identification, localization and functional analysis will help to further confirm the important functions of these key exoSIG molecules and their mechanisms of action.

In summary, we obtained an exosome-derived risk signature (exoSIG) on breast cancer and multi-molecular panels through a multi-omics-based RCABD strategy. Based on a comprehensive dissection of biological heterogeneity and the existing typing factors, we found that exoSIG can be used as a novel prognostic indicator for the stratified patients of breast cancer. Combining clinical information, the bimolecular risk signature, Clinmin- exoSIG (*HLA-DQB2* + *COL17A1*), enables accurate prognostic evaluation and stratification of risk patients. In conclusion, our findings provide a valuable reference for in-depth exploration of the "biological heterogeneity - risk stratification - prognosis prediction" and further advance the implementation of precision therapy.

Institutional review board

Not applicable.

Funding

This work was funded by the Science Innovation Program of College of Laboratory Medicine, Chongqing Medical University (CX202104); Natural Science Foundation of Chongqing (cstc2019jcyj-msxmX0271, cstc2021jcyj-msxm0317); Science and Technology Research Plan Project of Chongqing Education Commission (KJQN202100418).

Author declaration

We confirm no known and existing conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the

criteria for authorship but are not listed. Further, the order of authors listed in the manuscript has been approved, with certainty, by all of us. We understand that the Corresponding Author is the sole contact for the Editorial process, including communications with Editorial Manager and the office. He is responsible for the progress, including submissions, revisions and final approval of proofs communicating with the other authors. We confirm that we have provided a current, correct email address accessible by the Corresponding Author.

Informed consent

Not applicable.

CRediT authorship contribution statement

Fei Long, Haodong Ma, Youjin Hao: Conceptualization. **Fei Long**: Methodology, Formal analysis, Visualization. **Haodong Ma**: Investigation. **Haodong Ma, Youjin Hao**: Data curation. **Fei Long, Mingwei Liu**: Writing − original draft preparation. All authors: Writing − review & editing. **Mingwei Liu, Guoming Xie**: Supervision, Project administration.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgments

We would like to thank the peer reviewers input in improving this manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.05.013.

References

[1] Sung H., Ferlay J., Siegel R.L., et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians, 2021. 71(3):209–249.
[2] Bianchini G, Balko JM, Mayer IA, et al. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol 2016;13(11):674–90.
[3] Turashvili G, Brogi E. Tumor heterogeneity in breast cancer. Front Med 2017;4:227.
[4] Vessoni AT, Filippi-Chiela EC, Lenz G, et al. Tumor propagating cells: drivers of tumor plasticity, heterogeneity, and recurrence. Oncogene 2020;39(10):2055–68.
[5] Prat A, Pineda E, Adamo B, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast 2015;24:S26–35.
[6] Rivenbark AG, O'Connor SM, Coleman WB. Molecular and cellular heterogeneity in breast cancer: challenges for personalized medicine. Am J Pathol 2013;183(4):1113–24.
[7] Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797). 747-52.
[8] Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;100(14). 8418-23.
[9] Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27(8):1160.
[10] Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin Cancer Res 2015;21(7):1688–98.
[11] Russano M, Napolitano A, Ribelli G, et al. Liquid biopsy and tumor heterogeneity in metastatic solid tumors: the potentiality of blood samples. J Exp Clin Cancer Res 2020;39:1–13.
[12] Natrajan R, Sailem H, Mardakheh FK, et al. Microenvironmental heterogeneity parallels breast cancer progression: a histology–genomic integration analysis. PLoS Med 2016;13(2):e1001961.
[13] Palmirotta R, Lovero D, Cafforio P, et al. Liquid biopsy of cancer: a multimodal diagnostic tool in clinical oncology. Ther Adv Med Oncol 2018;10. 1758835918794630.
[14] Ignatiadis M, Dawson S-J. Circulating tumor cells and circulating tumor DNA for precision medicine: dream or reality? Ann Oncol 2014;25(12):2304–13.
[15] De Rubis G, Krishnan SR, Bebawy M. Liquid biopsies in cancer diagnosis, monitoring, and prognosis. Trends Pharmacol Sci 2019;40(3):172–86.
[16] Mashouri L, Yousefi H, Aref AR, et al. Exosomes: composition, biogenesis, and mechanisms in cancer metastasis and drug resistance. Mol Cancer 2019;18(1):1–14.
[17] Yu D, Li Y, Wang M, et al. Exosomes as a new frontier of cancer liquid biopsy. Mol Cancer 2022;21(1):56.
[18] Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. Science 2020;367:6478.
[19] Tian X, Shen H, Li Z, et al. Tumor-derived exosomes, myeloid-derived suppressor cells, and tumor microenvironment. J Hematol Oncol 2019;12(1):1–18.
[20] Huang T, Deng C-X. Current progresses of exosomes as cancer diagnostic and prognostic biomarkers. Int J Biol Sci 2019;15(1):1.
[21] Stevic I, Müller V, Weber K, et al. Specific microRNA signatures in exosomes of triple-negative and HER2-positive breast cancer patients undergoing neoadjuvant therapy within the GeparSixto trial. BMC Med 2018;16(1):1–16.
[22] Bao S, Hu T, Liu J, et al. Genomic instability-derived plasma extracellular vesicle-microRNA signature as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer. J Nanobiotechnol 2021;19(1):1–14.
[23] Ando W, Kikuchi K, Uematsu T, et al. Novel breast cancer screening: combined expression of miR-21 and MMP-1 in urinary exosomes detects 95% of breast cancer without metastasis. Sci Rep 2019;9(1):1–10.
[24] Moon P-G, Lee J-E, Cho Y-E, et al. Identification of developmental endothelial locus-1 on circulating extracellular vesicles as a novel biomarker for early breast cancer detection. Clin Cancer Res 2016;22(7):1757–66.
[25] Moon P-G, Lee J-E, Cho Y-E, et al. Fibronectin on circulating extracellular vesicles as a liquid biopsy to detect breast cancer. Oncotarget 2016;7(26):40189.
[26] Rupp A-K, Rupp C, Keller S, et al. Loss of EpCAM expression in breast cancer derived serum exosomes: role of proteolytic cleavage. Gynecol Oncol 2011;122(2):437–46.
[27] Galindo-Hernandez O, Villegas-Comonfort S, Candanedo F, et al. Elevated concentration of microvesicles isolated from peripheral blood in breast cancer patients. Arch Med Res 2013;44(3):208–14.
[28] Halvaei S, Daryani S, Eslami-S Z, et al. Exosomes in cancer liquid biopsy: a focus on breast cancer. Mol Ther-Nucleic Acids 2018;10:131–41.
[29] Rezaie J, Feghhi M, Etemadi T. A review on exosomes application in clinical trials: Perspective, questions, and challenges. Cell Commun Signal 2022;20(1):1–13.
[30] Wang X, Chai Z, Pan G, et al. ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer. Brief Bioinform 2021;22:3.
[31] Friedman J., Hastie T. and Tibshirani RJRpv glmnet: Lasso and elastic-net regularized generalized linear models, 2009. 1(4).
[32] Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 2010;1(4):274.
[33] Thorsson V, Gibbs DL, Brown SD, et al. The immune landscape of cancer. Immunity 2018;48(4):812–30. e14.
[34] Bagaev A, Kotlov N, Nomie K, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell 2021;39(6):845–65. e7.
[35] Wang S., Xiong Y., Zhang Q., et al. Clinical significance and immunogenomic landscape analyses of the immune cell signature based prognostic model for patients with breast cancer. Briefings in Bioinformatics, 2021. 22(4):bbaa311.
[36] Hu F.-F., Liu C.-J., Liu L.-L., et al. Expression profile of immune checkpoint genes and their roles in predicting immunotherapy response. Briefings in Bioinformatics, 2021. 22(3):bbaa176.
[37] Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med 2018;24(10):1550–8.
[38] Chandrashekar DS, Bashel B, Balasubramanya SAH, et al. UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses. Neoplasia 2017;19(8):649–58.
[39] Fisher R, Pusztai L, Swanton C. Cancer heterogeneity: implications for targeted therapeutics. Br J Cancer 2013;108(3):479–85.
[40] Bu L, Baba H, Yoshida N, et al. Biological heterogeneity and versatility of cancer-associated fibroblasts in the tumor microenvironment. Oncogene 2019;38(25):4887–901.
[41] Dent A, Diamandis P. Integrating computational pathology and proteomics to address tumor heterogeneity. J Pathol 2022.
[42] Li S, Yi M, Dong B, et al. The role of exosomes in liquid biopsy for cancer diagnosis and prognosis prediction. Int J Cancer 2021;148(11):2640–51.
[43] Du R, Huang C, Liu K, et al. Targeting AURKA in Cancer: Molecular mechanisms and opportunities for Cancer therapy. Mol Cancer 2021;20(1):1–27.
[44] Jia L, Lee HS, Wu CF, et al. SMAD4 suppresses AURKA-induced metastatic phenotypes via degradation of AURKA in a TGFβ-independent manner. Mol Cancer Res 2014;12(12):1779–95.
[45] Van Geldermalsen M, Wang Q, Nagarajah R, et al. ASCT2/SLC1A5 controls glutamine uptake and tumour growth in triple-negative basal-like breast cancer. Oncogene 2016;35(24):3201–8.
[46] Hassanein M, Qian J, Hoeksema MD, et al. Targeting SLC1a5–mediated glutamine dependence in non-small cell lung cancer. Int J Cancer 2015;137(7):1587–97.
[47] Yoo HC, Park SJ, Nam M, et al. A variant of SLC1A5 is a mitochondrial glutamine transporter for metabolic reprogramming in cancer cells. Cell Metab 2020;31(2):267–83. e12.

[48] Liu Y, Yang L, An H, et al. High expression of Solute Carrier Family 1, member 5 (SLC1A5) is associated with poor prognosis in clear-cell renal cell carcinoma. Sci Rep 2015;5(1):1–10.
[49] El-Ansari R, Craze ML, Alfarsi L, et al. The combined expression of solute carriers is associated with a poor prognosis in highly proliferative ER+ breast cancer. Breast Cancer Res Treat 2019;175(1):27–38.
[50] Zhu Y, Zhang S, Li Z, et al. miR-125b-5p and miR-99a-5p downregulate human γδ T-cell activation and cytotoxicity. Cell Mol Immunol 2019;16(2):112–25.
[51] Ma C, Luo H, Cao J, et al. Identification of a novel tumor microenvironment–associated eight-gene signature for prognosis prediction in lung adenocarcinoma. Front Mol Biosci 2020:242.
[52] Wang X, Hu L-P, Qin W-T, et al. Identification of a subset of immunosuppressive P2RX1-negative neutrophils in pancreatic cancer liver metastasis. Nat Commun 2021;12(1):1–17.
[53] Park I, Hwang S, Song I, et al. Expression of the MHC class II in triple-negative breast cancer is associated with tumor-infiltrating lymphocytes and interferon signaling. PloS One 2017;12(8):e0182786.
[54] Rodig S, Gusenleitner D. Jackson D, et al. MHC proteins confer differential sensitivity to CTLA-4 and PD-1 blockade in untreated metastatic melanoma. Sci Transl Med 2018;10:450.
[55] Moilanen JM, Kokkonen N, Löffek S, et al. Collagen XVII expression correlates with the invasion and metastasis of colorectal cancer. Hum Pathol 2015;46(3):434–42.
[56] Moilanen JM, Löffek S, Kokkonen N, et al. Significant role of collagen XVII and integrin β4 in migration and invasion of the less aggressive squamous cell carcinoma cells. Sci Rep 2017;7(1):1–11.
[57] Thangavelu PU, Krenács T, Dray E, et al. In epithelial cancers, aberrant COL17A1 promoter methylation predicts its misexpression and increased invasion. Clin Epigenet 2016;8(1):1–13.
[58] Yodsurang V, Tanikawa C, Miyamoto T, et al. Identification of a novel p53 target, COL17A1, that inhibits breast cancer cell migration and invasion. Oncotarget 2017;8(34):55790–803.
[59] Lothong M, Sakares W, Rojsitthisak P, et al. Collagen XVII inhibits breast cancer cell proliferation and growth through deactivation of the AKT/mTOR signaling pathway. PloS One 2021;16(7):e0255179.
[60] Qiu P, Guo Q, Yao Q, et al. Characterization of exosome-related gene risk model to evaluate the tumor immune microenvironment and predict prognosis in triple-negative breast cancer. Front Immunol 2021;12:736030.