Development of a novel three-lncRNA immune-related signature as a prognostic indicator for hepatocellular carcinoma

Bo Hu
Department of Liver Surgery
Xiao-Bo Yang
Department of Liver Surgery
Xinting Sang (sangxt@pumch.cn)
Peking Union Medical College Hospital https://orcid.org/0000-0003-1952-0527

Research

Keywords: Hepatocellular carcinoma, Immune, Signature, Long non-coding RNA, The Cancer Genome Atlas

Posted Date: February 18th, 2020

DOI: https://doi.org/10.21203/rs.2.23831/v1

Abstract

Background: Hepatocellular carcinoma (HCC) is one of the deadliest malignancies. Currently, there is still a lack of effective treatment. Our purpose was to develop an immune-related prognosis lncRNA signature with regard to HCC.

Methods: A total of 14,142 lncRNAs and 331 immune genes were obtained from The Cancer Genome Atlas (TCGA) and the Molecular Signatures Database to construct the immune-related lncRNAs co-expression networks. Moreover, the tumor samples collected from TCGA were randomized as training set and testing set, among which, the testing set and the entire set were used for verification. Subsequently, gene set enrichment analysis (GSEA) and principal component analysis (PCA) were employed for functional annotation.

Results: An immune-related signature consisting of *AC015908.3*, AC068987.4 and AL365203.2 was identified among HCC patients. Under different conditions, patients in low-risk group exhibited longer overall survival (OS) than those in high-risk group ($P < 0.001$). Moreover, the as-constructed signature was an independent factor, which showed marked association with patient OS ($P < 0.001$, hazard ratio (HR) = 1.407). These findings were further validated in testing set and the entire set. Additionally, GSEA results revealed the different immune states between low-risk and high-risk groups. On the other hand, lncRNA-related mRNAs were also extracted to depict the networks.

Conclusion: Our findings indicate that the three-lncRNA immune-related signature shows prognostic value for HCC.

Background

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related deaths worldwide, which affects about 800,000 cases annually[1]. Generally, HCC is treated with surgical resection and chemotherapy, but the mortality rate remains high[2], and the prognosis for HCC patients is affected by various factors. By contrast, the therapeutic regimens that counteract the immunosuppressive mechanisms potentially change the clinical outcomes of HCC, which prompts us to further explore the relationship of the abnormal local immune status with HCC development and prognosis[3]. Moreover, long non-coding RNAs (lncRNAs) possess a wide range of functional activities[4, 5], and their dysregulation, such as aberrant methylation modification, may contribute to epithelial-mesenchymal transition (EMT), proliferation, invasion and migration of HCC cells through activating various pathways, like the JAK2-STAT3 signal transduction pathway and the PTEN-PI3K/Akt pathway [6-8]. Meanwhile, the immune-related lncRNAs have also become the research hotspots. For instance, Jiang *et al.* demonstrated an immune-related lncRNA, namely, the lnc-epidermal growth factor receptor (EGFR), which stimulated the differentiation of T-regulatory (Treg) cells, thus promoting the immune escape of HCC [9]. Additionally, another gene, lncRNA cox-2, is found to prevent the immune escape and metastasis of HCC through altering the M1/M2 macrophage polarization[10]. Moreover, Li *et al.* described an HCC-derived

exosomal lncRNA TUC339, which was also implicated in regulating macrophage activation and polarization; besides, the authors further identified that the cytokine-cytokine receptor interaction, Toll-like receptor signaling, regulation of the actin cytoskeleton and cell proliferation were related to the TUC339 function in macrophages[11]. Therefore, it is speculated in this study that aberrant expression of immune-related lncRNAs may have prognostic value for HCC patients, which can thereby serve as the potential therapeutic targets.

This study aimed to gain insight into the potential clinical value of the immune-related lncRNA signature in prognosis stratification, as well as their potential as biomarkers in targeted HCC therapy. Typically, the RNA sequencing data were used in combination with the clinical information to select the HCC-specific lncRNAs, and the computational methods were also applied in assessing the overall survival (OS) for HCC patients. Further, the expression status and prognostic landscape of the as-constructed prognosis model were systematically analyzed. The molecular characteristics related to the immune status and multidimensional characteristics of the signature-lncRNAs were explored through bioinformatic analyses, including functional enrichment and co-expression analysis. Notably, results obtained in this study contribute to providing certain foundation for subsequent in-depth immune-related studies, which show great promise in individualized treatment of HCC.

Materials And Methods

Data collection

The level 3 mRNA expression profiles and clinical data were collected from 374 HCC and 50 normal control samples within TCGA database (https://cancergenome.nih.gov). Patients with no available survival data or those with the survival of $\leq$ 30 days were excluded, since they might die of lethal complications (such as heart failure, intracranial infection and hemorrhage) rather than HCC. Relevant data were downloaded from the publicly available database, so no additional ethical approval was required.

Mining of lncRNA profiles and grouping of patients

The lncRNA profiles were determined according to an established mining method[12]. First of all, genes were identified as protein-coding genes or non-coding genes based on their Refseq IDs or Ensembl IDs, and only the long non-coding genes in the NetAffx Annotation files were retained. Thereafter, the obtained lncRNAs expression values were subjected to log10 transformation for subsequent analysis. For the duplicated samples, the average lncRNAs expression levels of all duplicated samples were regarded as the eventual expression levels. Typically, a total of 14,142 lncRNAs were obtained from TCGA.

In addition, altogether 331 immune-related genes were extracted from the Molecular Signatures Database v7.0. [13] (Immune system process M13664, Immune response M19817; http://www.broadinstitute.org/gsea/msigdb/index.jsp). Ultimately, 542 immune-related lncRNAs were identified through constructing the immune-lncRNAs co-expression networks at the thresholds of correction

coefficient ≥ 0.4 and P ≤ 0.001. Furthermore, the missing follow-up information was filtered out, and 343 unique HCC samples, which were randomized as the training group (n=172) and the testing group (n=171) using the R package “caret”[14], were incorporated into subsequent analysis. Of them, the training set was adopted to establish a prognostic multi-lncRNA signature, whereas the testing set and the entire set were employed to validate the predictive power of the as-established signature.

Construction of the immune-related prognostic lncRNA signature

The univariate Cox regression analysis was performed to determine the association of immune-related lncRNA expression with patient overall survival (OS). Specifically, lncRNAs with the *P*-value of <0.001 were considered as the prognostic lncRNAs, which were further analyzed by the least absolute shrinkage and selection operator (LASSO) Cox regression for identifying the optimal candidates using the “glmnet” package in R[15]. Notably, LASSO is the penalized regression, which employs an L1 penalty to shrink the regression coefficients toward zero, thereby eliminating numerous variables based on the principle that fewer predictors are selected in the presence of a larger penalty. In this study, the nominated lncRNAs with the LASSO regression coefficients of <0.1 were ignored. Thereafter, a multi-lncRNA prognostic signature was built to predict patient OS. The risk score was assigned using this model according to the linear combination of the immune-related lncRNAs expression levels weighted by the regression coefficient (β). Specifically, β was calculated by the log-transformed hazard ratio (HR) obtained from univariate Cox regression analysis. Particularly, the risk score for every patient was calculated as follows: Risk score $=\beta_{gene1}$ × lncRNA1 expression level + β_{gene2} × lncRNA2 expression level + ⋯ + β_{genen} × lncRNAn expression level. Besides, the median risk score obtained from this model was set as the cutoff value to assign all patients into high- or low-risk group. Subsequently, the prognostic genes were screened from univariate Cox regression analysis based on the 331 immune-related lncRNAs to construct a three-lncRNA signature as the prognostic indicator.

Verification of the immune-related prognostic lncRNA signature

To confirm the prognostic significance of the candidate lncRNA signature, the Kaplan-Meier (K-M) curves were plotted for the training set, testing set and the entire set, respectively. Moreover, the time-dependent receiver operating characteristic (ROC) curve was employed to assess the predictive significance of the signature for survival prediction among the three groups through calculating the area under the curve (AUC) [16]. Afterwards, both univariate and multivariate Cox proportional hazard regression analyses were performed to ascertain whether the three-lncRNA signature predicted prognosis independently from diverse clinical factors, including age, gender, histologic grade, clinical stage, T stage, albumin level, platelet level, AFP level and vascular invasion). Also, principal components analysis (PCA) was also conducted as the dimension-reducing procedure to identify a small set of synthetic variables, so as to explore the model grouping capacity. Notably, PCA is a statistical technique to determine the key variables in a multidimensional data set, which explains the observational differences and is utilized to simplify the analysis and visualization of the multidimensional data sets[17].

Gene set enrichment analysis (GSEA)

GSEA is a computational approach to determine whether a priori defined gene set shows statistically significant and concordant differences between two biological states[18]. Therefore, GESA was performed in this study using the JAVA program (https://www.broadinstitute.org/gsea), so as to identify the immune lncRNA-related gene sets, among which, the MSigDB H: hallmark gene sets were used as the functional gene sets, whereas each lncRNA expression was utilized to be the phenotype. Following 1000 permutations, the top 20 gene sets with the FDR q < 0.25 and P < 0.05 were considered to be significantly enriched. Furthermore, the top 6 gene sets in terms of the normalized enrichment score (NES) were employed in multiple GSEA, so as to demonstrate a whole picture of the signaling pathway involved in the immune-related lncRNA signature in HCC.

Relationships of immune-related lncRNA prognostic index with immune cell infiltration

The TIMER online database, which a web resource to systemically evaluate the clinical impact of various immune cells on diverse cancer types, analyzes and visualizes the abundances of tumor-infiltrating immune cells [19]. It covers 10,009 samples across 23 cancer types from TCGA to estimate the abundance of six tumor-infiltrating immune cell subtypes, including B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells. Therefore, it can be easily used to determine the relationship of immune cells infiltration with other parameters. In this study, the immune infiltrate levels of TCGA HCC patients were downloaded, and the associations of the prognostic signature with immune cells infiltration were calculated.

Extraction of lncRNA-related mRNA and functional enrichment analysis

To predict the molecular function of each candidate lncRNA, the immune lncRNA-related mRNAs were screened by Pearson correlation analysis using the TCGA dataset (P-value and FDR q-value of < 0.001), which were then applied in further Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analyses using the Bioconductor “clusterProfiler” package, and the statistical significance thresholds were set at P-value and FDR q-value of < 0.001[20]. Typically, mRNAs with the absolute Pearson correlation coefficient values of >0.4 were screened and ranked based on the correlation degree. 50 mRNAs at both ends of the three lists were selected to construct the lncRNA-mRNA co-expression networks, and gene-gene co-expression networks connected with signature lncRNAs. In addition, the Cytoscape software (version 3.7.2), which is applicable for large network, which is greatly flexible in terms of network analysis, import, and visualization of additional data[21]. Besides, GeneMANI is a real-time multiple association network integration algorithm, which is used to predict gene function[22]. Both of them were employed in the current study to depict the comprehensive networks. Afterwards, three co-expressed mRNAs with the greatest positive Pearson correlation coefficients were introduced into the UALCAN database, so as to analyze the expression level and the survival outcomes. On the other hand, UALCAN, which adopts TCGA level 3 RNA-seq and clinical data from 31 cancer types, is an interactive web-portal for the in-depth analyses of TCGA gene expression data, and it is also utilized to assist in analysis[23]. Additionally, UALCAN is characterized by its user-friendly features that, it allows to analyze

the relative expression of a query gene(s) across tumor and normal samples, as well as among various tumor sub-groups based on individual cancer stage, tumor grade or other clinicopathological features. UALCAN is publicly available at http://ualcan.path.uab.edu.

Statistical methods

The R (v.3.6.1) software was employed for all statistical analyses. The K-M curve was plotted by the R package “survival” [24] and “survminer”. In addition, the area under the survival ROC curve (AUC) was calculated using the R software package, so as to validate the performance of the as-constructed prognostic signature[25]. Additionally, the univariate and multivariate Cox regression analyses were also performed by the R package “survival”, while PCA was implemented through the limma[26] and scatterplot3d[27] packages. Of them, the univariate Cox regression analysis was conducted based on the survival status. Besides, GSEA (http://www.broadinstitute.org/gsea/index.jsp) was performed for functional annotation, and the limma package was adopted to predict the target mRNAs of immune-related lncRNAs through co-expression . A two-sided P-value of < 0.05 was considered as statistically significant.

Results

Identification of the immune-related lncRNAs

After analyzing the transcriptome RNA-sequencing data of HCC samples downloaded from TCGA data portal, 14,142 genes were identified as lncRNAs, among which, 542 were deemed as immune-related lncRNAs, as suggested by the co-expression networks constructed with 331 immune-related genes extracted from the Molecular Signatures Database v7.0 (Immune system process M13664, Immune response M19817). Table 1 presents the clinical information of all samples in the training set, the testing set and the entire set.

Construction of the three-lncRNA immune-related signature for HCC patients

After analyzing and optimizing the training set data, a model composed of three novel immune-related lncRNAs, namely, *AC015908.3*, *AC068987.4* and *AL365203.2*, was constructed, which represented a relatively simple and accurate way to predict HCC prognosis (Table 2). To obtain the results, LASSO Cox regression analysis was carried out to screen the most putatively informative lncRNAs, and the “leave-one-out-cross-validation” approach was employed to facilitate parameter selection (Fig. 1A and 1B). Further, the expression patterns of three lncRNAs were investigated in all TCGA HCC samples, and the results suggested that, compared with the corresponding non-tumorous liver tissues, *AC015908.3* expression was remarkably down-regulated in tumor tissues, while *AC068987.4* and *AL365203.2* expression was markedly up-regulated (Fig. 2A-2C, respectively). Besides, there were significant differences in the expression levels of these three lncRNAs between high- and low-risk groups in the training set, the testing set and the entire set (P<0.0001; Fig. 2D-2L, respectively). Thereafter, all HCC patients in the training set were classified as two groups based on the median risk score, namely, the low-

and high-risk groups, and the difference in OS between these two groups was statistically significant ($P < 0.001$). Typically, the calculating formula was shown below: [*AC015908.3* expression level* (-0.4054)] + [*AC068987.4* expression level* (0.1099)] + [*AL365203.2* expression level* (0.0875)].

Survival analysis based on the three lncRNAs in the training, testing and entire sets

Fig.3A shows the comparison of survival differences between the two groups in training set. Moreover, such findings were further verified in the testing set and the entire set (Fig. 3B and 3C). Additionally, the AUC values were 0.788, 0.739 and 0.796 in the training set, the testing set and the entire set, separately, which indicated moderate and stable potentials for the immune-lncRNA-based prognostic signature to monitor survival (Fig. 3D - 3F). The presently constructed signature achieved the optimal AUC value compared with those of other conventional clinicopathological features in TCGA patients, simultaneously reflecting its excellent predicting ability. Besides, the impacts of high and low risks on the prognosis within subgroups with identical clinical features were also evaluated. In the entire set, the low-risk group showed markedly superior prognosis to the high-risk group among all age, gender, clinical stage, tumor grade and T stage subgroups (P<0.01; Fig. 4A-4J, respectively). To further explore the independent prediction of the model, univariate and multivariate analyses were performed, indicating that risk score could be identified as an independent prognostic indicator (Table 3). Fig.5A-5C depict the risk score distribution, survival status and expression of three signature lncRNAs in the training set, the testing set and the entire set. Clearly, their distributions were similar, thus supporting the robust predictive capacity of our three-lncRNA-based risk score assessment model.

PCA proved the model grouping capacity

PCA was further conducted to examine the difference between low- and high-risk groups based on the immune-related lncRNA signature, immune-lncRNAs, immune genes and the entire gene expression profiles. The results obtained based on our model showed that low- and high-risk groups were generally distributed at different directions (Fig. 6A). Nonetheless, the distributions of high- and low-risk groups displayed in Fig. 6B -6D were relatively scattered, which confirmed that our prognosis signature was capable of distinguishing the high-risk group from the low-risk group.

Functional enrichment analysis revealed different immune states between high- and low-risk groups

Functional annotation was further carried out by means of GSEA, and the results suggested that the differentially expressed genes (DEGs) between these two groups were enriched in the immunological signature gene sets (c7. All. V7.0. symbol). In addition, the significant enrichment gene sets in the top six NES were selected, as displayed in Fig. 7A – 7F. According to our results, GSE20727, GSE39110, GSE14308, GSE17974, GSE1460 and GSE24634 were differentially enriched in the high-risk phenotype. Afterwards, multiple GSEA was conducted to further illustrate the whole picture of the six gene sets involved in the immune-lncRNA prognosis index for HCC (Fig. 7G). Furthermore, the relationship of immune-related lncRNA signature with immunocyte infiltration was also explored, so as to examine whether the immuno-genome accurately reflected the tumor immune microenvironment status (Figure 8A-

8F). Our results suggested that, for high risk patients, the levels of B cells (Cor=0.143; p=0.008), $CD4^+$T cells (Cor=0.112; p=0.039), $CD8^+$T cells (Cor=0.249; p=2.978e-06), macrophages (Cor=0.470; p=2.802e−20), neutrophils (Cor=0.312; p=3.702e−09), and dendritic cells (Cor=0.274; p=2.507e−07), markedly increased in tumor microenvironment (TME) (P<0.05), indicating the difference in immune status between high- and low-risk groups.

Co-expression network and functional enrichment analyses

mRNAs with the Pearson correlation with lncRNA of >0.4 were selected and ranked according to the association with lncRNA, and altogether 300 mRNAs were plotted into three network maps (Fig. 10A – 10C). In addition, the interactive regulatory networks of these mRNAs were also constructed (Fig. 10D – 10F). Functional enrichment analyses demonstrated that 'cell adhesion molecule binding', 'cadherin binding', 'helicase activity', 'ATPase activity', 'catalytic activity (acting on RNA)', and 'transcription coactivator activity' were the most markedly enriched functions; whereas 'spliceosome', 'cell cycle', 'RNA transport', 'fatty acid degradation' and 'DNA replication' were the most remarkably enriched pathways. The results obtained from UALCAN showed that over-expression of the solute carrier family 38 member 1 (*SLC38A1*), which was positively correlated with *AL365203.2*, was observed in TCGA HCC tissues (P < 0.001), and it was associated with the shorter OS (P < 0.001) (Supplementary Fig.1A and 1B). Besides, the AT-rich interaction domain 3A (*ARID3A*) related to *AC068987.4* was evidently over-expressed within tumor tissues (P < 0.001), but it showed no ability in predicting prognosis (P = 0.28) (Supplementary Fig.1C and 1D). Meanwhile, the transmembrane protein 220 (TMEM220) expression level was down-regulated in HCC tumor samples (P < 0.001), which also served as a favorable prognostic factor (P = 0.039) (Supplementary Fig.1E and 1F).

Discussion

In this study, a series of complex bioinformatics analyses were carried out, which identified three lncRNAs associated with HCC prognosis, including *AC015908.3*, *AC068987.4* and *AL365203.2*. The immune-related risk model constructed based on these three lncRNAs was able to distinguish low-risk from high-risk samples with a relatively high prognosis accuracy. For the training set, HCC patients in low-risk group had longer OS than those in high-risk group, and similar results were also obtained in the testing set and the entire set. Moreover, when patients were stratified based on age (≤60/>60), gender (female/male), clinical stage (I and II/III and IV), grade (G1 and G2/G3 and G4) and T stage (T1 and T2/T3 and T4), the as-constructed prognosis model also displayed the potentials to predict the differential prognosis between high and low risk groups. Further PCA confirmed that our prognosis signature had grouping capacity. Therefore, the identified immune-lncRNA signature might be implicated in HCC initiation and progression, which rendered them with the potentials as the valuable clinical biomarkers. In addition, the gene expression profiles of high risk HCC cases selected by our signature were enriched in various immune-related gene sets, and subsequent analysis further confirmed the increased infiltration of B cells, $CD4^+$T cells, $CD8^+$T cells, macrophages, neutrophils, and dendritic cells in TME of high risk HCC patients, especially for the latter three immune cell types.

To the best of our knowledge, no existing study has reported the direct correlation between *AC015908.3, AC068987.4* or *AL365203.2* expression and HCC prognosis, and the functions of these three lncRNAs in HCC remain unclear so far. Moreover, the detailed studies with regard to these three novel lncRNAs are lacking at the moment. *AC068987.4* and *AL365203.2* were found to be the negative prognostic genes, while *AC015908.3* was the positive prognostic gene. Accordingly, *AC068987.4* and *AL365203.2* expression in high-risk groups among the three sets was remarkably higher than that in low-risk groups. Compared with normal tissues, the expression levels of the above-mentioned two adverse prognostic factors in TCGA HCC samples were also remarkably up-regulated. There are few reports on the immune-related functions of these three genes, and the identified lncRNAs may regulate immune function either directly or indirectly. Among the above-mentioned immune-related gene sets, GSE39110 displayed an immunization regimen, which delivered the T cell receptor (TCR) signals through a defined antigenic peptide, transmitted the inflammatory signals through lipopolysaccharide (LPS), and propagated the growth and differentiation signals through IL-2R that initially favored Ag-specific CD8+ T cells, for the sake of rapid and substantial development of TCR transgenic OVA-specific OT-I CD8+T cells into T effector-memory cells [28]. Compared with low-risk samples, the high-risk samples were enriched with genes that were down-regulated in $CD8^+$T cells at 3 days after immunization, which might suggest that high-risk patients had poor ability to transform Ag-specific CD8+ T cells into T effector-memory cells. Of interest, previous studies have demonstrated that, the tumor-infiltrating and interleukin-33–producing effector-memory CD8+ T cells in the resected HCC prolong patient survival[29], thus reflecting a protective effect. In another gene set, GSE14308, which contained genes that were up-regulated in Th2 cells compared with in Th17 cells, was also closely relate to the high-risk groups. The markedly increased levels of anti-inflammatory cytokines, such as IL-4, IL-5, IL-8, and IL-10, which were secreted by Th2 cells, were also reported to be unique to liver tissues in metastatic HCC patients[30], and such results might reflect a more malignant state in high risk patients. On the other hand, genes that were up-regulated in CD25+ T cells compared with CD25- T cells were also suggested to be associated with the high-risk groups. For instance, Lee *et al* illustrated that the increased CD4+ CD25+ T cells, which appeared to suppress the dendritic cells-activated immune response, was correlated with HCC tumor size in TME[31].

Moreover, the relationships of the signature with immune cell infiltration were deliberated to reflect the immune microenvironment status of HCC. Interestingly, our analysis indicated that the immune-related signature constructed in this study was positively correlated with the infiltration of six immune cell types, especially for macrophages, neutrophils and dendritic cells (DCs), suggesting that the higher infiltration levels of these immune cells might be observed in high-risk patients. Consistently, recent studies have expounded that the high levels of tumor-infiltrating macrophages and neutrophils predict dismal prognosis for primary HCC patients. Typically, the intramural neutrophil infiltration is elaborated to be promoted by the CXCL5 and CXCR2-CXCL1 axis; besides, it displays marked correlation with shorter OS and HCC recurrence, and serves as an independent prognostic factor[32-34]. Moreover, the increased infiltration of tumor-associated macrophages dominated by the M2 macrophages, which may be ascribed to the deletion of the Hippo signaling, has been reported to produce the Wnt/β-catenin signaling and trigger the elevated intramural FoxP3+ Treg population, while this in turn accelerates HCC

progression [35-37]. In addition, accumulating evidence reveals a role of DCs as an adverse prognostic factor for HCC. For instance, Zhou *et al.* set forth that, the intratumoral infiltration by plasmacytoid DCs was a novel indicator of the poor prognosis for HCC patients, which might be achieved through inducing the immune tolerogenic and inflammatory TME comprising regulatory T and IL-17-producing cells[38]. Such results have underscored the importance of tumor-associated DC cells in predicting the prognosis for HCC patients. Taken together, the as-constructed model not only predicted the prognosis for HCC patients, but also suggested the different immune conditions of patients in high- and low-risk groups. The immune-related signature was correlated with survival, which might serve as the biomarkers to assess the feasibility of various immunotherapies. Nevertheless, to investigate the detailed relationship between immune-related lncRNA expression and the actual immune phenotype, future studies should make efforts to development more novel immune-related signatures in RNA sequencing datasets and further validate them experimentally. The three lncRNAs identified in this study provided clues for discovering the particular immunotherapies for HCC patients.

It is known that lncRNAs participate in various biological processes, such as transcription, translation, cell differentiation, chromatin modification, as well as regulation of gene expression, cell cycle and nuclear-cytoplasmic trafficking[39, 40]. To further explore the broader functions of these three lncRNAs, co-expression and functional enrichment analyses were performed subsequently. After co-expression analysis on those three ncRNAs, mRNAs with the Pearson correlation of >0.4 were selected for further enrichment analysis. The results demonstrated that these three lncRNA-related protein-coding genes were mainly involved in cadherin binding, helicase activity, and ATPase activity, which thus indirectly revealed the potential functions of the signature-lncRNAs from another perspective. Numerous works have elaborated that, the reduced expression of E-cadherin as well as catenin complex, such as α-, β-, γ-catenin, occurs frequently in HCC, which contributes to tumor progression and development [41, 42]. On the other hand, several genes that are involved in ATPase activity or helicase activity are also illustrated to have a bearing on HCC progression. RuvB-like 2 (*RUVBL2*), an ATPase and putative DNA helicase known to interact with β-catenin and cellular v-myc myelocytomatosis viral oncogene homolog (c-myc), is confirmed to be over-expressed in most HCCs, which is associated with the enhanced tumorigenicity [43]. Further, Grigoletto *et al*. reported that the ATPase activity of Reptin, which was linked to the Walker A and B domains, was required for its effects on tumor cell growth and viability in HCC[44]. Based on the KEGG pathway enrichment analysis, the spliceosome pathway was revealed to be the most enriched one, and most of its genes were demonstrated to be up-regulated in HCC[45]. The spliceosome is a ribonucleoprotein complex to control the (alternative) splicing, which participates in cell cycle, invasion, metastasis and angiogenesis [46]. Previous studies indicate that, targeting spliceosome reduces cancer cell proliferation [14], and triggers the mTOR blockade and autophagy[47]. Our founding suggested that, the signature-lncRNAs might also promote HCC cell metastasis and proliferation through the spliceosome pathway. However, further study is warranted to characterize the tumor-specific role of lncRNAs. Moreover, mRNAs showing the highest Pearson correlation coefficient with signature-lncRNAs were also examined. Specifically, TMEM220, which showed the highest positive correlation coefficient with *AC015908.3* according to our co-expression analysis (Cor: 0.662, $P < 0.001$), was proved to be remarkably

down-regulated in TCGA HCC tissues compared with normal liver tissues, and it was deemed as the favorable prognostic factor. *ARID3A*, a nuclear matrix-associated transcription factor that stimulates the expression of immunoglobulin heavy chain (IgH) expression as well as the progression of the Cyclin E1/E2F-dependent cell cycle, displays marked correlation with *AC068987.4* (Cor: 0.636, $P < 0.001$), as mentioned in a transcriptional regulatory network in the context of HCC[48, 49]. In our study, *ARID3A* expression in HCC samples was markedly up-regulated, but its effect on prognosis should be further investigated. Additionally, *SLC38A1*, a crucial glutamine transporter, plays an essential role in nutrient uptake, energy production, chemical metabolism, detoxification, and neurotransmitter cycling[50]. As the mRNA that had the highest positive correlation with *AL365203.2* in our study (Cor: 0.65, $P < 0.001$), *SLC38A1* was demonstrated to be activated in HCC tissues[51], which was linked with the worse OS for HCC patients. Taken together, the potential functions mentioned in this study have enriched the scope of action of the novel signature-lncRNAs and assisted in understanding these three genes at a deeper level.

Nonetheless, some limitations should be noted in this study, which should be taken into consideration when interpreting our results. Firstly, the transcriptomic analysis only reflected some aspects of immune status, rather than the global map. Secondly, the immunocyte-specific gene sets applied in this study were limited to 6 major immune cell types, as a result, differences in the more specialized immune cell subtypes (like the differently polarized macrophages or myeloid-derived suppressor cells) might not be recognized in this study, even though they were known to be mechanistically linked to HCC progression and stage[16, 52]. Finally, our results were not validated via another independent cohort, which was also a limitation of this study, and the reliability of our molecular results was still challenged by the lack of experiments in vitro or in vivo.

Conclusions

To sum up, the three lncRNAs screened in this study (including *AC015908.3, AC068987.4* and *AL365203.2*) may serve as the novel lncRNA biomarkers to predict HCC prognosis. Moreover, there is difference in immune state between the low- and high-risk groups divided according to the median riskscore value. Further studies on these lncRNAs and the associated genes may shed more lights on the underlying mechanism of HCC development.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this article and its supplementary information files.

Funding

Not applicable.

Author contribution

Xin-Ting Sang and Bo Hu created the idea for the review. Bo Hu performed the selection of literature, drafted the manuscript, and prepared the figures. Xiao-Bo Yang and Xin-Ting Sang revised the manuscript. All authors read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest.

Acknowledgements

The authors would like to thank the HCCDB, UALCAN, Kaplan-Meier plotter, TIMER, LinkedOmics and TCGA databases for the availability of the data.

References

1. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* International journal of cancer, 2015. **136**(5): p. E359-E386.
2. Llovet, J.M., R. Montal, and A. Villanueva, *Randomized trials and endpoints in advanced HCC: Role of PFS as a surrogate of survival.* Journal of hepatology, 2019.
3. Khemlina, G., S. Ikeda, and R. Kurzrock, *The biology of Hepatocellular carcinoma: implications for genomic and immune therapies.* Molecular cancer, 2017. **16**(1): p. 149.
4. Mercer, T.R., M.E. Dinger, and J.S. Mattick, *Long non-coding RNAs: insights into functions.* Nature reviews genetics, 2009. **10**(3): p. 155.
5. L, L., et al., *Long noncoding RNA expression profiling in cancer: Challenges and opportunities.* Genes, chromosomes & cancer, 2019. **58**(4): p. 191-199.
6. SX, Y., et al., *Long noncoding RNA, the methylation of genomic elements and their emerging crosstalk in hepatocellular carcinoma.* Cancer letters, 2016. **379**(2): p. 239-44.
7. Y, W., et al., *Long Noncoding RNA HOST2 Promotes Epithelial-Mesenchymal Transition, Proliferation, Invasion and Migration of Hepatocellular Carcinoma Cells by Activating the JAK2-STAT3 Signaling Pathway.* Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology, 2018. **51**(1): p. 301-314.
8. Y, L., et al., *Long Noncoding RNA lncARSR Promotes Doxorubicin Resistance in Hepatocellular Carcinoma via Modulating PTEN-PI3K/Akt Pathway.* Journal of cellular biochemistry, 2017. **118**(12):

p. 4498-4507.

9. R, J., et al., *The long noncoding RNA lnc-EGFR stimulates T-regulatory cells differentiation thus promoting hepatocellular carcinoma immune evasion.* Nature communications, 2017. **8**(undefined): p. 15129.
10. Y, Y., et al., *Long non-coding RNA cox-2 prevents immune evasion and metastasis of hepatocellular carcinoma by altering M1/M2 macrophage polarization.* Journal of cellular biochemistry, 2018. **119**(3): p. 2951-2963.
11. X, L., et al., *Regulation of Macrophage Activation and Polarization by HCC-Derived Exosomal lncRNA TUC339.* International journal of molecular sciences, 2018. **19**(10): p. undefined.
12. Zhang, X., et al., *Long non-coding RNA expression profiles predict clinical phenotypes in glioma.* Neurobiology of disease, 2012. **48**(1): p. 1-8.
13. Subramanian, A., et al., *Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
14. Kuhn, M., *Building predictive models in R using the caret package.* Journal of statistical software, 2008. **28**(5): p. 1-26.
15. Friedman, J.H., T. Hastie, and R. Tibshirani, *Regularization Paths for Generalized Linear Models via Coordinate Descent.* Journal of Statistical Software, 2010. **033**.
16. Blanche, P., J.r. Dartigues, and H. Jacqmin□adda, *Estimating and comparing time□dependent areas under receiver operating characteristic curves for censored event times with competing risks.* Statistics in Medicine, 2013. **32**(30): p. 5381-5397.
17. Raychaudhuri, S., J. Stuart, and R. Altman, *Principal components analysis to summarize microarray experiments: application to sporulation time series.* 2000. **5**: p. 455.
18. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences, 2005. **102**(43): p. 15545-15550.
19. Li, T., et al., *TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells.* Cancer Research. **77**(21): p. e108-e110.
20. Yu, G., et al., *clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters.* OMICS-A JOURNAL OF INTEGRATIVE BIOLOGY. **16**(5): p. 284-287.
21. NT, D., et al., *Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data.* Journal of proteome research, 2019. **18**(2): p. 623-632.
22. Mostafavi, S., et al., *GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function.* **9**(1 Supplement).
23. Chandrashekar, D.S., et al., *UALCAN: a portal for facilitating tumor subgroup gene expression and survival analyses.* Neoplasia, 2017. **19**(8): p. 649-658.

24. Diboun, I., et al., *Microarray analysis after RNA amplification can detect pronounced differences in gene expression using limma.* **7**(1): p. 252-0.
25. Heagerty, P.J., et al., *Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker.* Biometrics.
26. ME, R., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic acids research, 2015. **43**(7): p. e47.
27. Koutecký, P., *MorphoTools: a set of R functions for morphometric analysis.* Plant Systematics & Evolution, 2015. **301**(4): p. 1115-1121.
28. Castro, I., M.J. Dee, and T.R. Malek, *Transient Enhanced IL-2R Signaling Early during Priming Rapidly Amplifies Development of Functional CD8+ T Effector-Memory Cells.* Journal of Immunology. **189**(9): p. 4321-4330.
29. Brunner, S.M., et al., *Tumor-infiltrating, interleukin-33–producing effector-memory CD8+ T cells in resected hepatocellular carcinoma prolong patient survival.* Hepatology, 2015. **61**.
30. Qin, L.X., *Inflammatory Immune Responses in Tumor Microenvironment and Metastasis of Hepatocellular Carcinoma.* **5**(3): p. 203-209.
31. WC, L., et al., *The impact of CD4+ CD25+ T cells in the tumor microenvironment of hepatocellular carcinoma.* Surgery, 2012. **151**(2): p. 213-22.
32. Li, L., et al., *CXCR2–CXCL1 axis is correlated with neutrophil infiltration and predicts a poor prognosis in hepatocellular carcinoma.* Journal of Experimental & Clinical Cancer Research, 2015. **34**(1): p. 129.
33. Zhou, S.L., et al., *Overexpression of CXCL5 mediates neutrophil infiltration and indicates poor prognosis for hepatocellular carcinoma.* Hepatology, 2012. **56**(6): p. 2242-2254.
34. Li, Y.-W., et al., *Intratumoral neutrophils: a poor prognostic factor for hepatocellular carcinoma following resection.* Journal of hepatology, 2011. **54**(3): p. 497-505.
35. Ding, T., et al., *High tumor-infiltrating macrophage density predicts poor prognosis in patients with primary hepatocellular carcinoma after resection.* Human pathology, 2009. **40**(3): p. 381-389.
36. Shirabe, K., et al., *Role of tumor-associated macrophages in the progression of hepatocellular carcinoma.* Surgery today, 2012. **42**(1): p. 1-7.
37. Tian, Z., et al., *Macrophages and hepatocellular carcinoma.* Cell & bioscience, 2019. **9**(1): p. 79.
38. Zhou, Z.-J., et al., *Intratumoral plasmacytoid dendritic cells as a poor prognostic factor for hepatocellular carcinoma following curative resection.* Cancer Immunology, Immunotherapy, 2019: p. 1-11.
39. Kim, E.-D. and S. Sung, *Long noncoding RNA: unveiling hidden layer of gene regulatory networks.* Trends in Plant Science. **17**(1): p. 0-21.
40. Yoon, J.-H., K. Abdelmohsen, and M. Gorospe, *Posttranscriptional Gene Regulation by Long Noncoding RNA.* Journal of Molecular Biology. **425**(19): p. 3723-3730.

41. Zhai, B., et al., *Reduced expression of E-cadherin/catenin complex in hepatocellular carcinomas.* World journal of gastroenterology: WJG, 2008. **14**(37): p. 5665.
42. Fransvea, E., et al., *Blocking transforming growth factor–beta up-regulates E-cadherin and reduces migration and invasion of hepatocellular carcinoma cells.* Hepatology (Baltimore, Md.). **47**(5): p. 1557-1566.
43. *Overexpression and role of the ATPase and putative DNA helicase RuvB-like 2 in human hepatocellular carcinoma.* Hepatology, 2007. **46**(4).
44. Grigoletto, A., et al., *The ATPase Activity of Reptin Is Required for Its Effects on Tumor Cell Growth and Viability in Hepatocellular Carcinoma.* Molecular Cancer Research. **11**(2): p. 133-139.
45. W, X., et al., *Meta-analysis of gene expression profiles indicates genes in spliceosome pathway are up-regulated in hepatocellular carcinoma (HCC).* Medical oncology (Northwood, London, England), 2015. **32**(4): p. 96.
46. RI, S. and N. M, *Alternative splicing in cancer: noise, functional, or systematic?* The international journal of biochemistry & cell biology, 2007. **39**(null): p. 1432-49.
47. RJ, v.A., et al., *The spliceosome as target for anticancer treatment.* British journal of cancer, 2009. **100**(2): p. 228-32.
48. Chen, J., et al., *Integrative Analysis of Microarray Data to Reveal Regulation Patterns in the Pathogenesis of Hepatocellular Carcinoma.* Gut and liver, 2016. **11**(1): p. 112-120.
49. Lin, D., et al., *Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer.* **6**(1): p. 23.
50. Gu, S., et al., *Characterization of an N-system Amino Acid Transporter Expressed in Retina and Its Involvement in Glutamine Transport *.* Journal of Biological Chemistry, 2001. **276**.
51. Wang, K., et al., *Activation of SNAT1/SLC38A1 in human breast cancer: correlation with p-Akt overexpression.* Bmc Cancer, 2013. **13**.
52. Shen, P., et al., *Increased circulating Lin−/lowCD33+HLA-DR− myeloid-derived suppressor cells in hepatocellular carcinoma patients.* 2014. **44**(6): p. 639-650.

Tables

Table 1. Summary of training set, testing set and entire set

Clinical factor	Training set	Testing set	Entire set
	SEX		
Male	118	117	235
Female	54	54	108
	AGE		
≤60	86	78	164
>60	86	93	179
	GEADE		
G1-G2	105	109	214
G3-G4	65	59	124
NA	2	3	5
	STAGE		
I-II	124	113	237
III-IV	38	46	84
NA	10	12	22
	SURVIVAL SATATUS		
Alive	112	112	224
Dead	60	59	119

NA: non-available

Table 2. lncRNAs in the risk assessment model

lncRNA	Coefficient	HR	Lower. 95	Upper. 95	*P* - value
AC015908.3	-0.4054	0.67	0.5	0.89	0.006
AC068987.4	0.1099	1.12	1.0	1.22	0.019
AL365203.2	0.0875	1.09	1.0	1.18	0.035

lncRNA, long noncoding RNA; HR, hazard ratio

Table 3. Univariate and multivariate Cox regression analyses of clinicopathologic characteristics associated with overall survival in The Cancer Genome Atlas samples

Variable	The training set				The testing set				The entire set			
	Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis		Univariate analysis		Multivariate analysis	
	HR (95% CI)	*P* - value	HR (95% CI)	*P* - value	HR (95% CI)	*P* - value	HR (95% CI)	*P* - value	HR (95% CI)	*P* - value	HR (95% CI)	*P* - value
Age (>60/≤60)	1.032 (0.997-1.068)	0.072	-	-	1.004 (0.975-1.035)	0.774	-	-	1.018 (0.995-1.041)	0.125	-	-
Gender (male/female)	0.500 (0.231-1.081)	0.078	-	-	0.812 (0.357-1.646)	0.619	-	-	0.601 (0.345-1.048)	0.073	-	-
Grade(G4/G3/G2/G1)	2.188 (1.244-3.850)	0.007[a]	2.059 (1.025-4.137)	0.042[a]	0.971 (0.990-1.731)	0.971	-	-	1.458 (0.997-2.133)	0.052	-	-
Stage (IV/III/II/I)	1.135 (0.711-1.812)	0.596	-	-	2.121 (1.368-3.287)	<0.001[a]	2.629 (0.876-7.888)	0.085	1.595 (1.184-2.148)	0.002[a]	1.593 (0.613-4.143)	0.339
T stage (T4/T3/T2/T1)	1.167 (0.750-1.226)	0.494	-	-	1.919 (1.234-2.985)	0.004[a]	0.710 (0.238-2.121)	0.540	1.524 (1.134-2.047)	0.005[a]	0.983 (0.385-2.508)	0.971
Albumin (>3.5/≤3.5g/dL)	0.885 (0.639-1.226)	0.462	-	-	1.120 (0.716-1.752)	0.618	-	-	0.964 (0.740-1.255)	0.784	-	-
Platelet (>250/≤250 × 109/L)	0.999 (0.995-1.003)	0.835	-	-	1.000 (0.999-1.001)	0.398	-	-	0.999 (0.998-1.001)	0.288	-	-
AFP (> 20/≤ 20 ng/mL)	1 (0.999-1.002)	0.666	-	-	0.999 (0.997-1.003)	0.458	-	-	0.999 (0.997-1.001)	0.435	-	-
Vascular invasion (macro/micro/none)	1.465 (0.764-2.809)	0.250	-	-	2.448 (1.376-4.357)	0.002[a]	2.715 (1.326-5.559)	0.006[a]	1.920 (1.263-2.920)	0.002[a]	1.962 (1.193-3.022)	0.006[a]
Risk score	1.686 (1.304-2.179)	<0.001[a]	1.435 (1.021-2.016)	0.037[a]	1.317 (1.041-1.666)	0.021[a]	1.367 (1.008-1.864)	0.044[a]	1.455 (1.233-1.716)	<0.001[a]	1.371 (1.120-1.677)	0.002[a]

[a], Statistically significant. AFP, Alpha-fetoprotein; HR, Hazard ratio; CI, confidence interval.

Figures

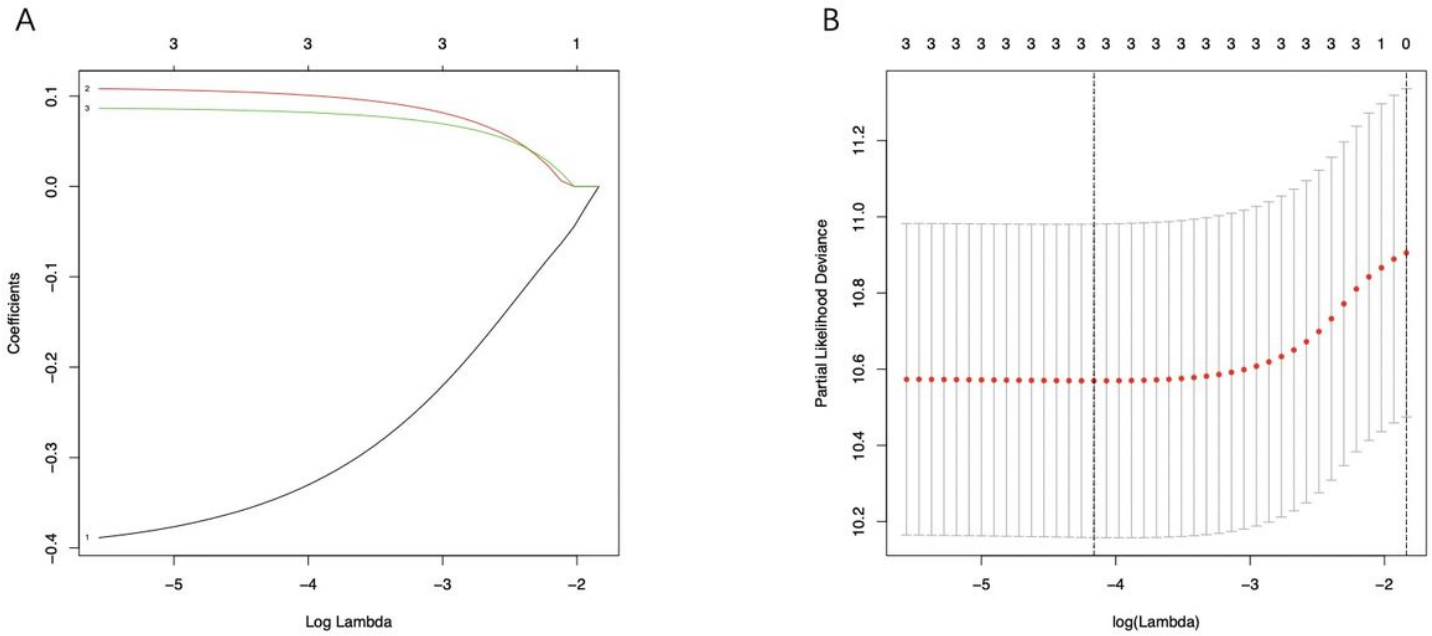

Figure 2

LASSO coefficient profiles of the three lncRNAs are depicted in (A). (B) show the selection of the tuning parameter (lambda) in the LASSO model by tenfold cross-validation based on minimum criteria for OS; the lower X axis shows log (lambda), and the upper X axis shows the average number of OS-genes. The Y axis indicates partial likelihood deviance error. Red dots represent average partial likelihood deviances for

every model with a given lambda, and vertical bars indicate the upper and lower values of the partial likelihood deviance errors. The vertical black dotted lines define the optimal values of lambda, which provides the best fit.

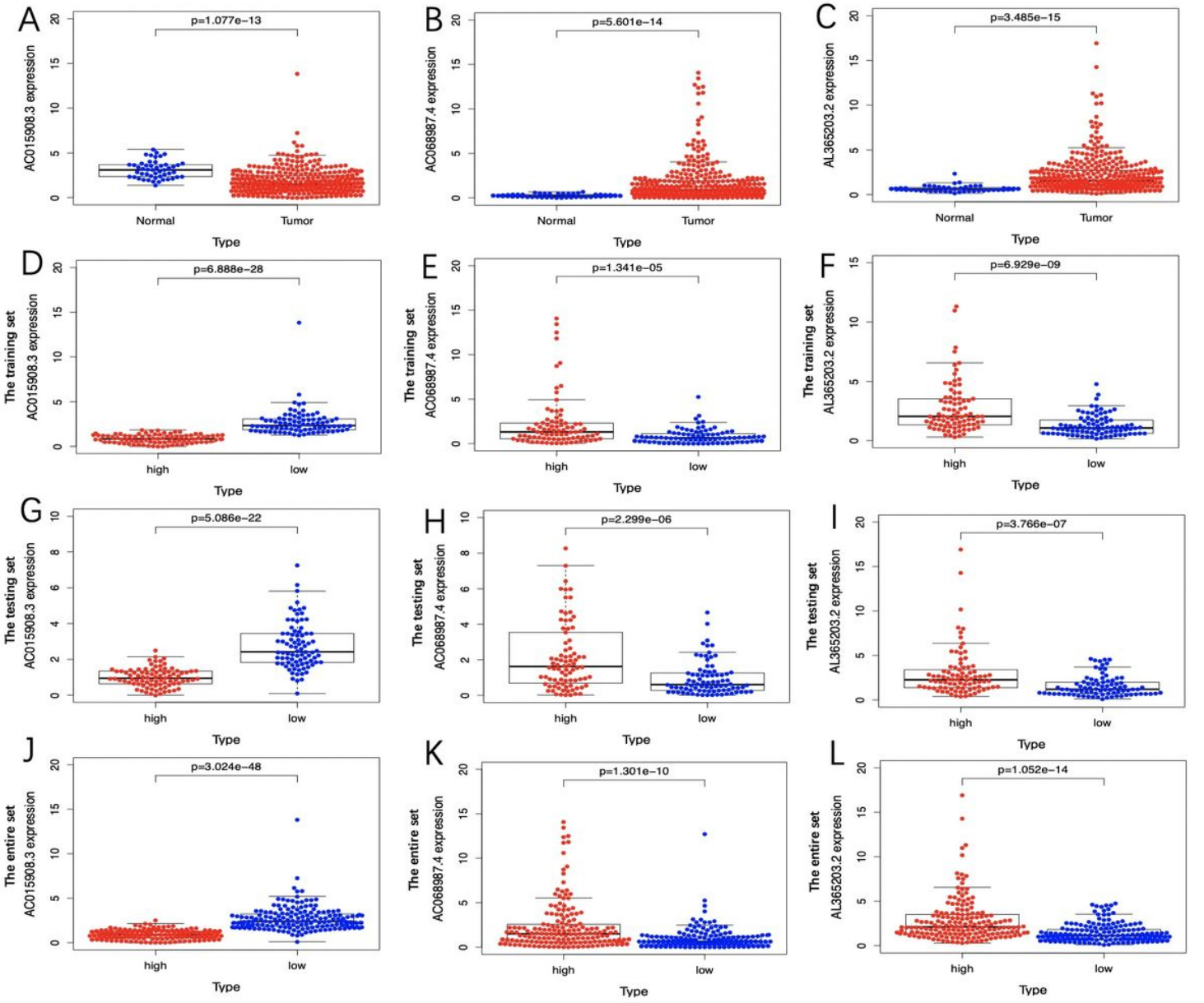

Figure 4

Expression of three signature lncRNAs in samples of The Cancer Genome Atlas (TCGA) –hepatocellular carcinoma (HCC) cohort. (A) – (C) show the expression of three lncRNAs in hepatocellular carcinoma samples and normal tissues. (D) – (F) show the expression of three genes between high- and low - risk groups in the training group. Similar contents of testing set are exhibited in (G) - (I) while entire set in (J) – (L).

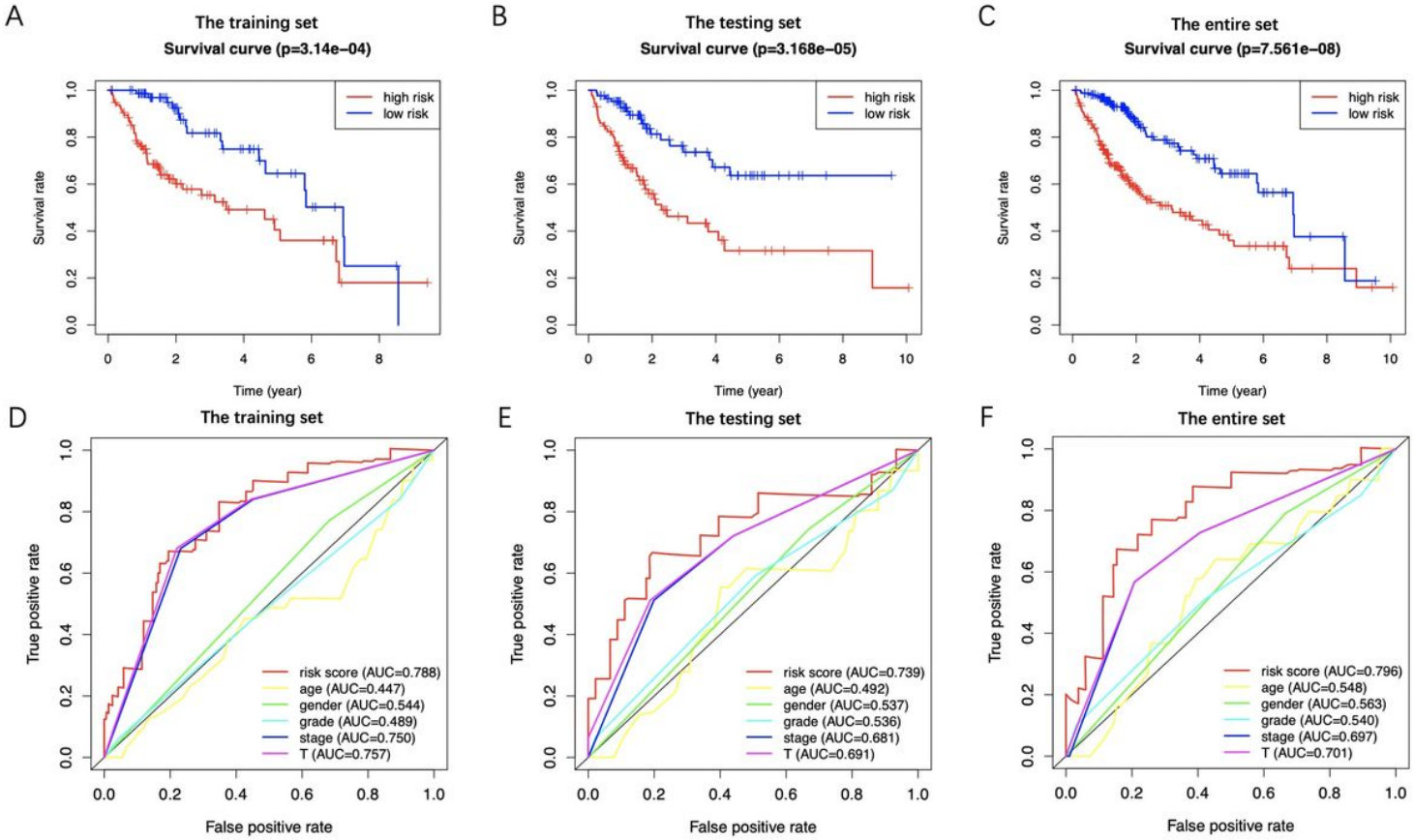

Figure 6

Survival curves of patients in high risk group and low risk group of training set (A), testing set (B) and entire set (C). Patients in high-risk group suffered shorter overall survival. (D) – (F) show survival-dependent receiver operating characteristic (ROC) curves validation at 1 – year of prognostic value of the prognostic index in the three sets (training set, testing set and and entire set, respectively).

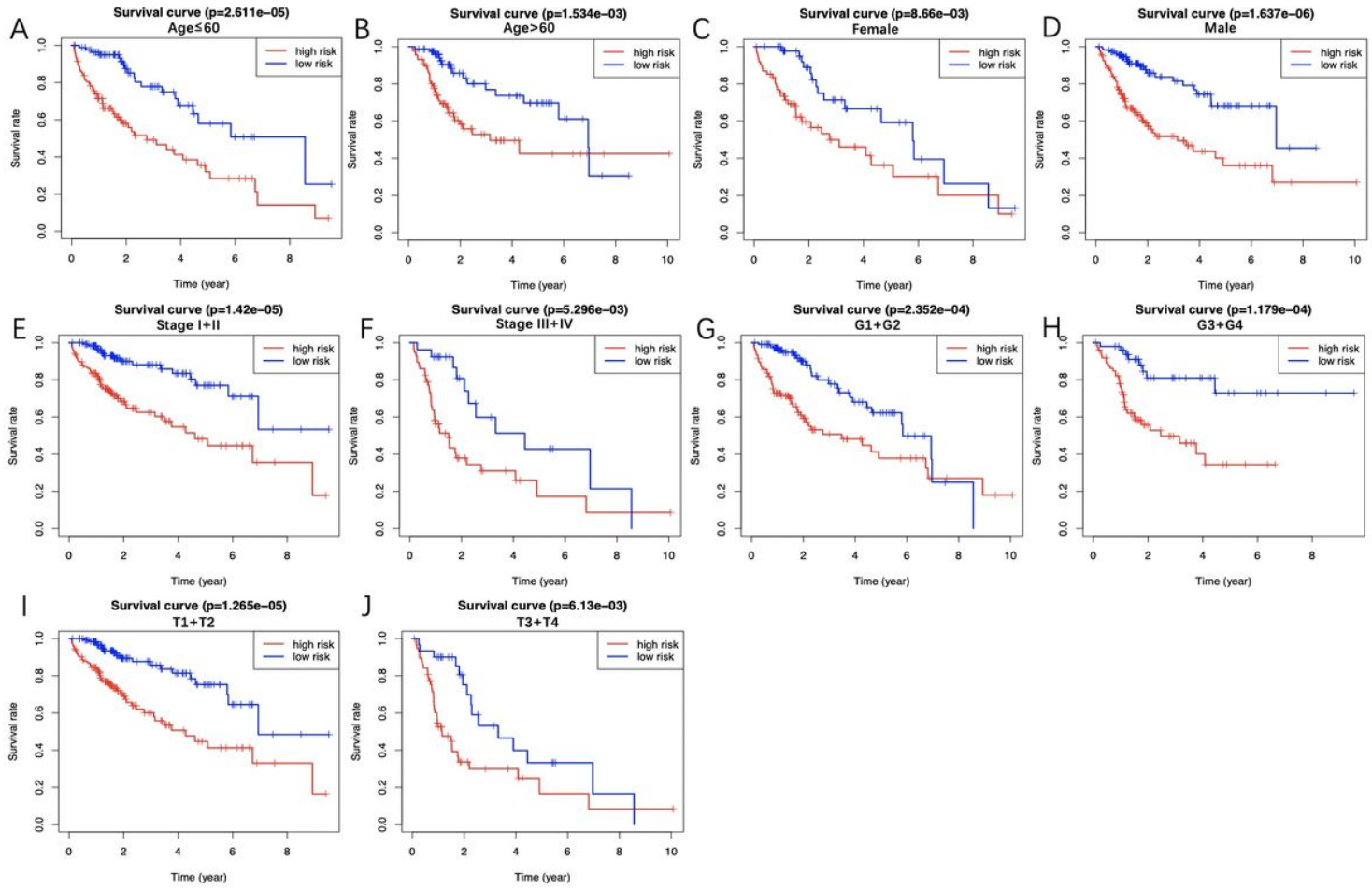

Figure 8

The overall survival differences between the high-risk group and the low-risk group were shown under the conditions of classifying patients by age (A) - (B), sex (C) – (D), stage (E) – (F), grade (G) – (H) and T stage (I) – (J). Detailed notes are described in the main text.

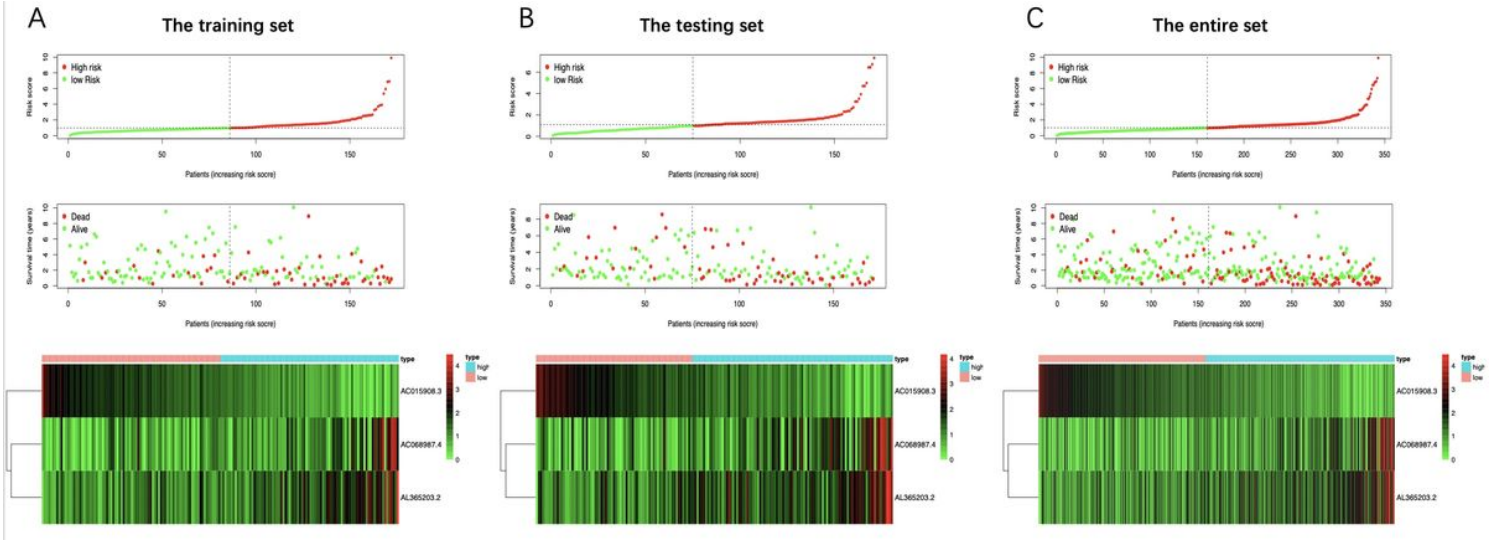

Figure 10

Distribution of risk score, overall survival (OS), gene expression in (A) training set, (B) testing set and (C) entire set. Distribution of risk score, OS and heat map of the expression of three signature lncRNAs in low-risk and high-risk groups are listed in the picture from top to bottom.

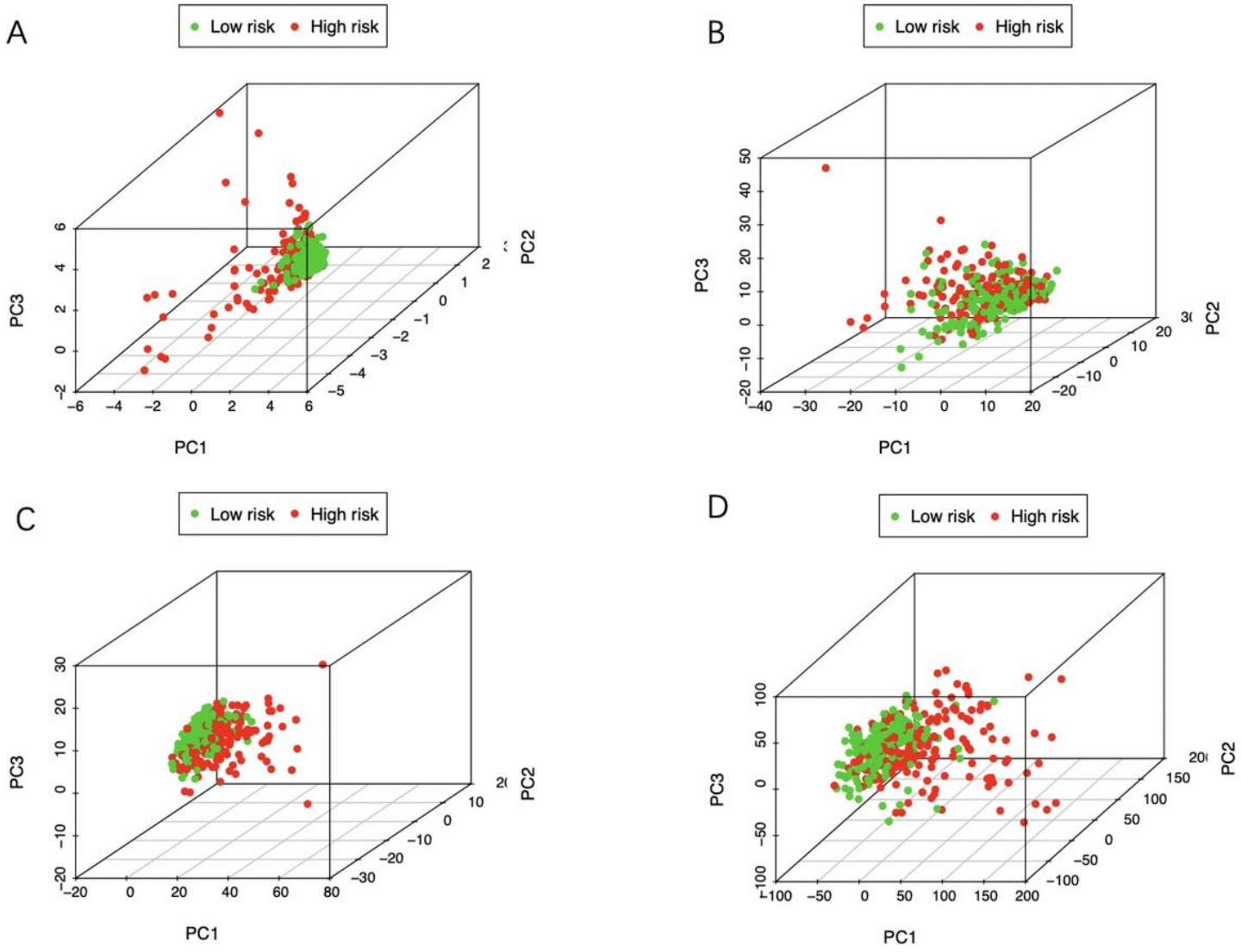

Figure 12

Principal components analysis between low- and high-risk groups based on (A) immune-related lncRNA signature, (B) immune-lncRNAs, (C) immune genes and (D) the entire gene expression profiles.

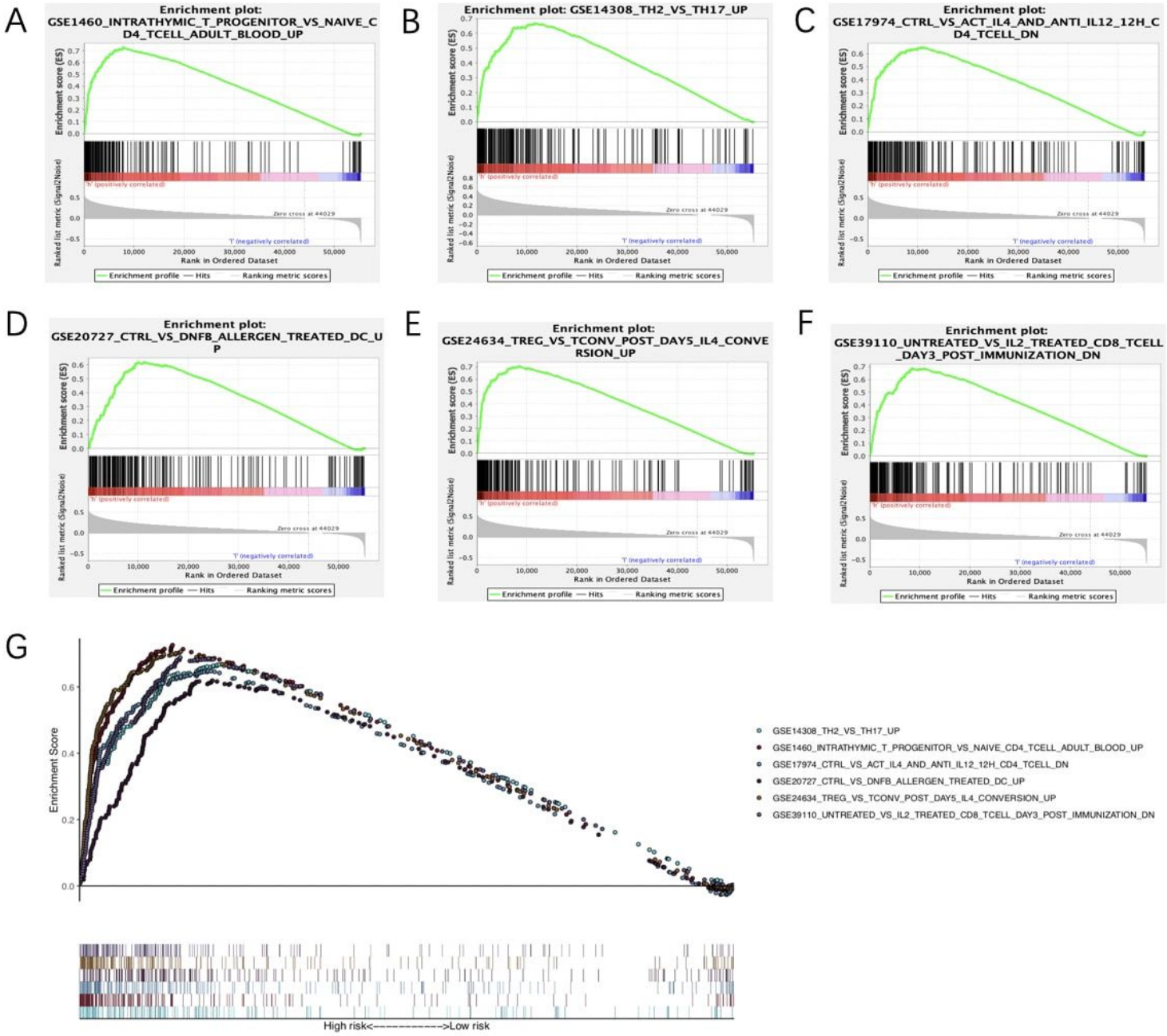

Figure 14

Enrichment plots from gene set enrichment analysis (GSEA). GSEA results showing (A) GSE1460_intrathymic T progenitor vs naïve CD4+Tcell adult blood_UP, (B) GSE14308_Th2 vs Th17 UP, (C) GSE17974_CTRL vs act IL4 and anti IL12 12H CD4+Tcell_DN, (D) GSE20727_CTRL vs DNFB allergen treated DC_UP, (E) GSE24634_Treg vs Tconv post day5 IL4 conversion_UP, (F) GSE39110_untreated vs IL2 treated CD8+Tcell day3 post immunization_DN are positively differentially enriched in high risk group that screened our by the lncRNA-related signature. (G) summarizes the above six gene sets.

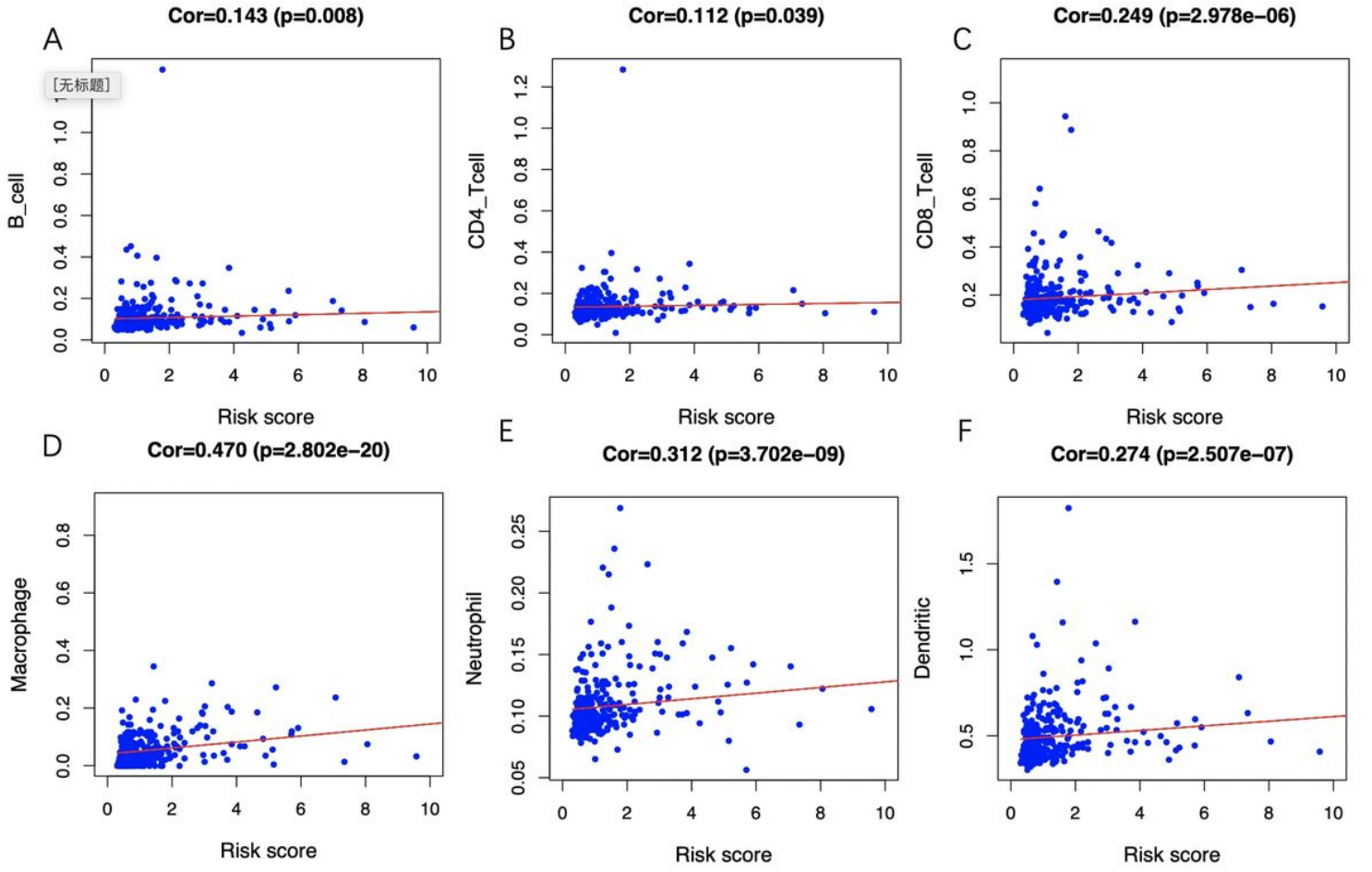

Figure 16

Relationships between the immune-related prognostic index and infiltration abundances of six types of immune cells. The correlation was performed by using Pearson correlation analysis. (A) B cells; (B) CD4+Tcells; (C) CD8+T cells; (D) neutrophils; (E) macrophages; and (F) dendritic cells.

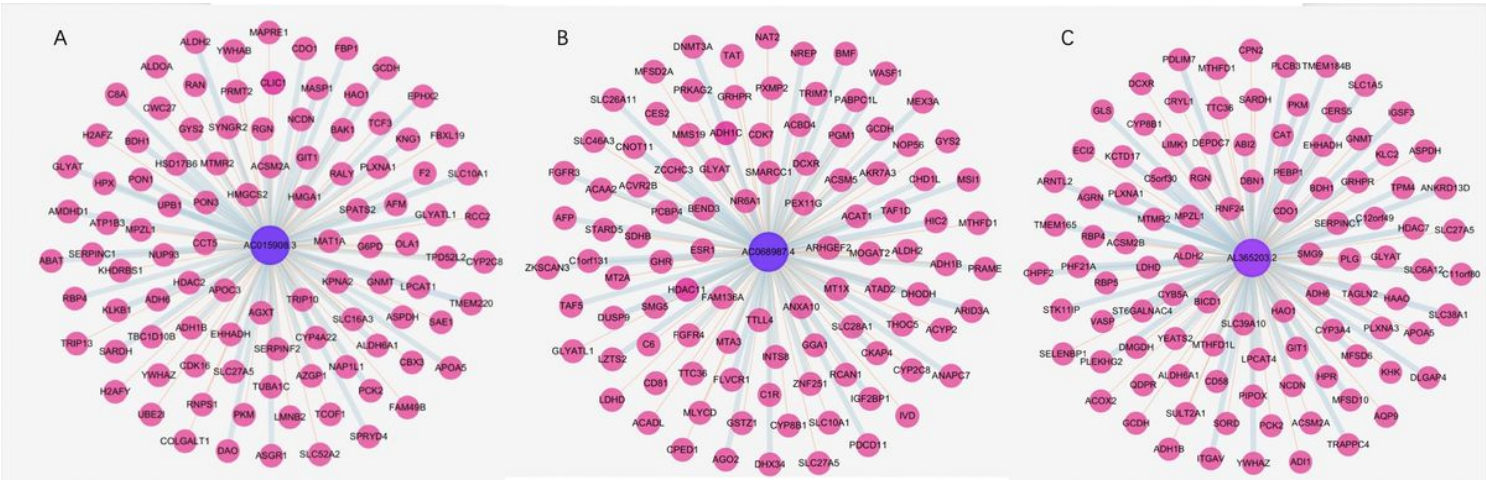

Figure 18

. Co-expression networks of signature lncRNA-gene interactions. Top mRNAs co-expressed with (A) AC015908.3, (B) AC068987.4, (C) AL365203.2. Each lncRNA is associated with 100 co-expressed mRNAs.

Blue lines represent positive correlation, while yellow lines represent negative correlation.