Air pollution and hepatocellular carcinoma: Integrated network toxicology, machine learning, molecular docking and multiomics analysis

XIAORONG WANG[1*], HUIHAO QIN[2*], JING JING[2], YUN LIU[1], LEI GUO[1], HONGQUAN LIU[3] and WEI ZHANG[4]

[1]The Third Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, P.R. China; [2]Department of Radiology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210028, P.R. China; [3]Department of Neurology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210028, P.R. China; [4]Department of Interventional Radiology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210028, P.R. China

Received March 19, 2026; Accepted May 19, 2026

DOI: 10.3892/ol.2026.15737

Abstract. Air pollution is closely related to the incidence and prognosis of hepatocellular carcinoma (HCC); however, the mechanisms remain elusive. Therefore, using network toxicology and molecular docking analysis, the present study aimed to identify candidate molecular pathways potentially linking 10 air pollutants (APs) and HCC through machine learning and multi-omics analyses. Using weighted gene co-expression network analysis, differential gene expression analysis and multiple online databases, AP-related and HCC-related genes (AP-HCC-Gs) were obtained for functional and pathway enrichment analyses. A total of 101 combinations of 10 machine learning algorithms were applied to construct an AP-related and HCC-related prognostic signature (AP-HCC-PS), and three machine learning algorithms were used to screen AP-related and HCC-related diagnostic markers (AP-HCC-DMs). Based on the AP-HCC-PS and AP-HCC-DMs, molecular docking and nomogram analyses were conducted. The AP-HCC-PS and AP-HCC-DMs were intersected, and single-cell RNA, reverse transcription-quantitative PCR, small interfering RNA, wound healing and immune cell infiltration analyses were performed. A total of 43 AP-HCC-Gs were identified, which were mainly involved in cell proliferation, oxidative stress and the immune response. The AP-HCC-PS (seven AP-HCC-Gs) and two AP-HCC-DMs showed good prognostic and diagnostic value. Molecular docking analysis suggested the potential binding affinity between these targets and APs. The hub gene proteasome 20S subunit β5 (PSMB5) was highly expressed in regulatory T cells (Tregs) and positively associated with Treg infiltration. AP exposure may be associated with HCC incidence through two AP-HCC-DMs and with prognosis through seven AP-HCC-Gs, with PSMB5 potentially representing an important candidate gene in HCC progression. The present study provided a multi-dimensional analytical framework for identifying candidate genes and pathways potentially linking APs to HCC, thereby advancing hypotheses regarding environmental carcinogenesis.

Correspondence to: Professor Wei Zhang, Department of Interventional Radiology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, 100 Shizijie, Hongshan Road, Nanjing, Jiangsu 210028, P.R. China
E-mail: zw2307@163.com

Professor Hongquan Liu, Department of Neurology, Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, 100 Shizijie, Hongshan Road, Nanjing, Jiangsu 210028, P.R. China
E-mail: njliuhq@126.com

*Contributed equally

Key words: air pollutants, hepatocellular carcinoma, network toxicology, machine learning, molecular docking

Introduction

With the acceleration of global industrialization, air pollution has become one of the major drivers of cancer and cancer-related mortality. Environmental pollutants and other extrinsic factors contribute to 70-90% of cancers, and in particular, exposure to air pollutants (APs), such as fine particulate matter ($PM_{2.5}$), nitrogen dioxide (NO_2) and ozone (O_3), is associated with the morbidity and mortality of hepatocellular cancer (HCC) (1-3).

The incidence, mortality and prognosis of HCC are affected by air pollution. In six European adult cohorts recruited between 1985 and 2005, a positive linear association was observed between NO_2, $PM_{2.5}$ and HCC incidence (4). A study of 940 patients with HCC revealed that, after adjusting for confounders and co-pollutants, for every 1-$\mu g/m^3$ increase

in $PM_{2.5}$ concentration or every 1-ppb increase in NO_2 concentration, the HCC-related mortality increased by 11 and 8%, respectively, and the survival time was markedly decreased (5). It is widely recognized that air pollution affects the onset and progression of HCC mainly through pathways such as oxidative stress, inflammation and epigenetic processes; however, to the best of our knowledge, markers have remained elusive (6,7).

Network toxicology and molecular docking offer novel approaches to explore the link between APs and diseases, overcoming epidemiological limitations and revealing mechanistic associations with HCC (8). In the present study, 10 prevalent APs [benzene, sulfur dioxide (SO_2), sulfur trioxide (SO_3), nitric oxide (NO), carbon monoxide (CO), NO_2, O_3, toluene, vinyl chloride and 1,3-butadiene] were selected based on their ubiquity in urban-industrial airsheds and relevance to real-world exposure risks. These compounds are representative of major sources: Traffic exhaust [CO, nitrogen oxides (NO_x), benzene, toluene and 1,3-butadiene], fossil-fuel combustion and power generation (SO_2 with secondary SO_3 formation, NO_x and CO) and petrochemical/industrial activities (benzene, toluene, vinyl chloride and 1,3-butadiene). O_3 forms secondarily through photochemical reactions between NO_x and volatile organic compounds. These gaseous toxicants are predominantly absorbed by organisms via inhalation, with dermal and gastrointestinal absorption occurring under high-exposure or occupational conditions. To date, only limited studies have reported pollutant-specific mechanisms for these 10 APs in HCC, with existing evidence primarily demonstrating that benzene produces toxic intermediates (such as benzene oxide) that disrupt DNA repair, and that 1,3-butadiene induces carcinogenesis by forming DNA adducts and causing chromosomal aberrations (9,10). To elucidate their potential mechanistic relevance in HCC, an integrative approach combining network toxicology, machine learning, molecular docking and multi-omics analyses was employed.

Materials and methods

Evaluation of AP toxicity. Air pollution-related genes were obtained from Pu *et al* (11), who investigated seven air pollutants in HCC. To expand the pollutant list, three additional pollutants, namely 1,3-butadiene, sulfur trioxide and vinyl chloride, were included in the present study. These three pollutants were searched in PubChem (https://pubchem.ncbi.nlm.nih.gov/) using their exact chemical names as search terms, namely '1,3-butadiene', 'sulfur trioxide' and vinyl chloride. The corresponding pollutant-related genes were then retrieved and integrated with the genes associated with the seven pollutants reported by Pu *et al* (11), resulting in a ten-pollutant air pollution-related gene set. Pu *et al* (11) mainly focused on air pollutant-related immune genes in HCC based on seven pollutants. By contrast, the present study was based on a broader set of genes associated with ten air pollutants and was not limited to air pollution-related immune genes. This approach may provide a more comprehensive view of the potential molecular association between air pollution exposure and HCC. The ADMETLAB 3.0 platform (https://admetlab3.scbdd.com) and the ProTox3 database (https://tox.charite.de/protox3) were used to assess carcinogenicity with their default prediction settings. For each AP, the canonical chemical structure was submitted directly to both platforms without additional parameter modification. The carcinogenicity-related probability output generated by each platform was recorded and a value >0.5 was considered indicative of carcinogenic potential. Pollutants were defined as carcinogenic when either ADMETLAB 3.0 or ProTox3 returned a carcinogenicity probability >0.5.

Acquisition of target genes of APs. The human target genes for the 10 APs were obtained from the SuperPred database (https://prediction.charite.de/), Swiss Target Prediction database (http://www.swisstargetprediction.ch/), STITCH database (http://stitch.embl.de/) and Similarity Ensemble Approach database (https://sea.bkslab.org/). The genes from all databases were merged and duplicates were removed.

Acquisition of HCC-related target genes. RNA-sequencing data for HCC were derived from The Cancer Genome Atlas (TCGA; 50 normal samples; 374 HCC samples; https://www.cancer.gov/ccg/research/genome-sequencing/tcga), the International Cancer Genome Consortium (ICGC; 232 HCC samples; https://dcc.icgc.org/) and GSE14520 (221 HCC samples; https://www.ncbi.nlm.nih.gov/geo/). Within each dataset, expression matrices were preprocessed and normalized according to the original data format, and cross-dataset batch effects were corrected using the ComBat function in the 'sva' R package (version 4.3.2; https://www.r-project.org/) before downstream integrative analyses. In the weighted gene co-expression network analysis (WGCNA), genes with a standard deviation of expression level <0.5 were removed. The optimal soft threshold was determined to be 18. Dynamic tree cutting and hierarchical clustering were then used to identify modules. The cutting height was 0.3 and the minimum module size was 100 genes.

Differentially expressed genes (DEGs) between HCC and normal liver samples with $|\log_2$ fold change$|>1$ and false discovery rate (FDR) $q<0.05$ were subsequently identified.

HCC-related genes were retrieved from the GeneCards database (https://www.genecards.org/) with a threshold of >5 score.

Functional and pathway enrichment analyses. The 'Org.Hs.eg.db', 'enrichplot' and 'clusterProfiler' packages in R software were used to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of AP-related and HCC-related genes (AP-HCC-Gs). GeneMANIA (https://genemania.org/) was used to evaluate the biological functions.

mRNA-microRNA (miRNA) regulatory network. TargetScan Human (https://www.targetscan.org), miRDB (https://mirdb.org/), miRanda (https://www.microrna.org/) and miRWalk (https://mirwalk.umm.uni-heidelberg.de/) were utilized to predict the binding of miRNAs and mRNAs. Cytoscape (v3.8.2; https://cytoscape.org/) was used for visualization.

Construction and validation of the AP-HCC-prognostic signature (PS) using machine learning. Batch effects between the TCGA cohort (training set) and ICGC cohort (external validation set) were corrected using the 'sva' R package. After univariate Cox regression analysis was performed in the

training set to identify AP-HCC-Gs associated with overall survival, 10 machine-learning algorithms, including least absolute shrinkage and selection operator (LASSO), Ridge, StepCox, CoxBoost, random survival forest, generalized boosted regression modeling, survival support vector machine, supervised principal components (SuperPC), elastic net and partial least squares regression for Cox, were used to develop 101 prognostic models with 10-fold cross-validation. These models included both single-algorithm models and sequential combinations, in which the first algorithm was applied for variable selection and the second algorithm was used for model fitting. For single-algorithm models, all candidate variables were entered directly into model construction. Only genes shared by the TCGA and ICGC datasets were retained, and the expression matrices were standardized before analysis. Model combinations yielding ≤3 variables at the variable-selection step or retaining ≤3 variables in the final fitted model were excluded (seed, 1,234). Model performance was evaluated using the concordance index (C-index) in both the training and external validation cohorts, and models were ranked according to the average C-index across cohorts. The top-ranked model was retained for downstream analyses.

Principal component analysis (PCA), the area under the curve (AUC), Kaplan-Meier (K-M) curves, risk curves and univariate and multivariate Cox analyses were used to further validate the robustness.

Molecular docking. Protein structures were obtained from the Alphafold (https://alphafold.ebi.ac.uk/) and Protein Data Bank (https://www.rcsb.org/) databases. CB-Dock2 (https://cadd.labshare.cn/) was used for molecular docking analysis.

Identification of AP-HCC-diagnostic markers (DMs) and hub genes. For diagnostic marker screening, LASSO, support vector machine recursive feature elimination (SVM-RFE) and random forest algorithms were applied in the discovery dataset using 10-fold cross-validation or the corresponding internal optimization procedure as implemented in each algorithm. Genes consistently retained across the three methods were defined as candidate diagnostic markers and were subsequently evaluated by receiver operating characteristic (ROC) analysis ('timeROC' R package). Because these markers were derived from the discovery pipeline, their diagnostic performance should be interpreted as preliminary until further independent validation is available. Subsequently, AP-HCC-DMs were further intersected with AP-HCC-PS genes to obtain hub genes. The AUC was used to evaluate the diagnostic value.

Construction of nomograms. Using the 'rms' R package, prognostic and diagnostic nomograms were constructed. The C-index was used to evaluate the consistency of the predicted values and the observed values.

Single-cell analysis and immune-cell infiltration analysis. Single-cell analysis was used to localize the hub gene within immune cell subpopulations and, through cell-cell communication analysis, to predict the communication frequency between the hub gene-enriched immune subset and other immune cells, thereby indicating potential AP-gene-immune cell pathways through which AP-induced gene expression may promote an immunosuppressive tumor microenvironment in HCC.

Reverse transcription-quantitative PCR (RT-qPCR), small interfering RNA (siRNA) transfection and wound-healing assay. THLE-2 and Huh-7 cells were obtained from the Institute of Cell Research, Chinese Academy of Sciences. Huh-7 cells were cultured in DMEM (Gibco; Thermo Fisher Scientific, Inc.), whereas THLE-2 cells were cultured in bronchial epithelial growth medium (BEGM; Lonza, Switzerland) with the recommended growth supplements added. Both media were supplemented with 10% fetal bovine serum (Gibco; Thermo Fisher Scientific, Inc.) and 1% penicillin-streptomycin (Gibco; Thermo Fisher Scientific, Inc.). All cells were maintained at 37˚C in a humidified incubator with 5% CO_2. Total RNA was isolated using TRIzol reagent (cat. no. 15596-026; Invitrogen; Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions. cDNA was synthesized using a reverse transcription kit (K1622; Fermentas; Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions. RT-qPCR was performed using a SYBR Green PCR kit (K0223; Thermo Fisher Scientific, Inc.) on an ABI 7300 Real-Time PCR System (Applied Biosystems; Thermo Fisher Scientific, Inc.). The PCR reaction mixture (25 μl) contained 10 μl SYBR Green Mix, 1 μl forward primer, 1 μl reverse primer, 2 μl cDNA template and 11 μl nuclease-free water. The thermocycling conditions were as follows: Initial denaturation at 95˚C for 10 min, followed by 40 cycles of denaturation at 95˚C for 15 sec and annealing/extension at 55˚C for 45 sec. Melting curve analysis was subsequently performed at 95˚C for 15 sec, 60˚C for 1 min, 95˚C for 15 sec and 60˚C for 15 sec. Relative mRNA expression levels were quantified using the $2^{-\Delta\Delta Cq}$ method with GAPDH as the internal control (12). The primer sequences used for RT-qPCR were as follows: Proteasome 20S subunit β5 (PSMB5) forward, 5'-GCTACAGCGGGTGCTTACAT-3' and reverse, 5'-TTCCCAGAAGCTGCAATCCG-3'; GAPDH forward, 5'-CACCCACTCCTCCACCTTTG-3' and reverse, 5'-CCACCACCCTGTTGCTGTAG-3'.

Cells were transfected with PSMB5-targeting siRNA or the corresponding negative control (NC) siRNA using Lipofectamine 2000 (cat. no. 11668-019; Invitrogen; Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions. At 24 h post-transfection, cells were harvested for total RNA extraction, and RT-qPCR was performed to determine the knockdown efficiency of PSMB5 at the transcript level. The siRNA sequences used for transfection were as follows: siNC, 5'-UUCUCCGAACGUGUCACGUTT-3'; and siPSMB5, 5'-CGAAAUGCUUCAUGGAACA-3'.

For the wound-healing assay, cells in the control and experimental groups were seeded in 35-mm dishes at a density of $8x10^5$ cells/dish and cultured until reaching a confluent monolayer. A linear scratch was created perpendicular to the reference line marked on the bottom of the dish using a 10-μl pipette tip. Detached cells were removed by PBS washing and cells were then incubated in serum-free medium. Images were captured at 0, 24 and 48 h at the same marked position under a microscope to evaluate cell migration. Wound closure was quantified using ImageJ software (version 1.53; National Institutes of Health).

Single-cell analysis and immune-cell infiltration analysis. Single-cell analysis was used to localize the hub gene within immune cell subpopulations and, through cell-cell communication analysis, to predict the communication frequency between the hub gene-enriched immune subset and other immune cells, thereby indicating potential AP-gene-immune cell pathways through which AP-induced gene expression may promote an immunosuppressive tumor microenvironment in HCC.

In the present study, scRNA-seq experiments were not performed on newly collected samples. Instead, publicly available scRNA-seq data from 12 HCC samples, consisting of treatment-naïve tumor tissues from patients with pathologically confirmed primary HCC, were obtained from the China National Gene Bank Nucleotide Sequence Archive (CNP0000650; https://db.cngb.org/cnsa).

After excluding genes expressed in <3 cells and retaining cells with an nFeature of 6,000-10,000, nCount of 4,000-15,000 and percentage of mitochondrial gene counts (percent.mt) <5, 19,516 genes and 8,680 cells were obtained. For scRNA-seq data, LogNormalize was applied for data normalization, the 'Harmony' package was applied for correction of batch effects and PCA was applied for dimensionality reduction. The 'FindNeighbors' and 'FindClusters' functions were used for cell clustering and resolution determination. Uniform manifold approximation and projection were used for visualization. Cell annotation was performed using marker genes from previous studies (13-15). Cell-cell communication was analyzed using the CellChat R package.

The infiltration abundance of 22 immune cell types was calculated using CIBERSORT (https://cibersortx.stanford.edu/). Pearson analysis was further utilized to evaluate the correlation between the hub genes and immune cell types in HCC.

Statistical analysis. Statistical analyses were performed using R software (version 4.3.2; https://www.r-project.org/). Unless otherwise specified, two-group comparisons of continuous variables were performed using Student's t-test or the Wilcoxon rank-sum test, depending on data distribution. Categorical variables were compared using the χ^2 test. Survival analyses were conducted using K-M curves ('survival' R package) and Cox regression models. P-values were two-sided, and $P<0.05$ was considered to indicate a statistically significant difference. Where appropriate, multiple testing was adjusted for using the FDR method.

Results

Toxicity of APs and AP-related target genes. Fig. S1 shows the research flowchart of the present study. All 10 APs exhibited carcinogenicity based on both toxicity prediction platforms (Table SI; Fig. 1A). Target genes for the 10 APs were obtained: 133 for 1,3-butadiene, 82 for benzene, 70 for CO, 92 for NO, 87 for NO_2, 88 for O_3, 92 for SO_2, 83 for SO_3, 102 for toluene and 173 for vinyl chloride. After removing duplicates, a total of 354 AP-related target genes were identified (Table SII).

Identification of HCC-related target genes. WGCNA identified six co-expression modules (Fig. 1B). Among them, the blue module (3,344 genes) had the highest correlation with HCC ($r=0.7$; $P=3\times10^{-12}$; Fig. 1D and E). Module membership and gene significance showed a significant correlation ($r=0.64$; $P<1\times10^{-200}$; Fig. 1C).

Differential gene expression analysis of mRNAs in TCGA-liver HCC dataset identified 6,479 DEGs (Fig. 1F).

A total of 4,423 HCC-related genes were obtained from GeneCards. After taking intersection, 655 HCC-related genes were obtained (Fig. 1G; Table SIII).

mRNA-miRNA regulatory network and GO/KEGG enrichment analyses of AP-HCC-Gs. The intersection of 354 AP-related genes and 655 HCC-related genes was determined, revealing 43 AP-HCC-Gs (Fig. 1H; Table SIII).

GO analysis revealed that the AP-HCC-Gs were primarily involved in 'mitotic cell cycle phase transition' (GO:0044772), 'regulation of protein serine/threonine kinase activity' (GO:0071900), 'protein autophosphorylation' (GO:0046777), 'positive regulation of transferase activity' (GO:0051347), 'nuclear division' (GO:0000280) and 'G_2/M transition of mitotic cell cycle' (GO:0000086) (Fig. S2A; Table SIV).

KEGG pathway analysis suggested that the top three most enriched pathways were 'microRNAs in cancer', 'cell cycle' and 'MAPK signaling pathway' (Fig. S2B; Table SV). Furthermore, a number of immune and metabolic pathways were enriched, including the 'Fc epsilon RI signaling pathway', 'neutrophil extracellular trap formation', 'growth hormone synthesis, secretion and action', 'chemical carcinogenesis-reactive oxygen species' and 'thyroid hormone signaling pathway'.

In addition, due to the highly enriched 'miRNAs in cancer' pathway, a miRNA-mRNA network was constructed, providing a reference for relevant studies to understand core regulatory axes driven by miRNAs (Fig. S3; Table SVI).

Construction and validation of the AP-HCC-PS. The univariate Cox analysis identified 36 AP-HCC-Gs significantly associated with HCC prognosis (Fig. 2G; Table SVII). Among the 101 models, the StepCox (both) + SuperPC model and the StepCox (backward) + SuperPC model achieved the highest average C-index (Fig. 2C). As both models showed equivalent predictive performance and retained the same seven genes, namely FKBP prolyl isomerase 1A (FKBP1A), NAD(P)H quinone dehydrogenase 1 (NQO1), aurora kinase B (AURKB), hydroxymethylbilane synthase (HMBS), enhancer of zeste 2 polycomb repressive complex 2 subunit (EZH2), PSMB5 and X-ray repair cross complementing 1 (XRCC1), the StepCox (both) + SuperPC model was retained as the final model because it preserved variables selected in both directions and was therefore considered to be more comprehensive. In this signature, lower expression of these genes was associated with improved prognosis (Fig. S4A-F).

PCA and K-M curve analysis indicated that the AP-HCC-PS could distinguish between different prognoses and had good predictive performance (Fig. 2E and F). The risk curve and heat map indicated that the higher the AP-HCC-PS gene expression level, the worse the prognosis (Fig. 2D and H). The univariate Cox analysis [hazard ratio (HR), 2.224; 95% CI, 1.701-2.908; $P<0.001$; Fig. S4G] and

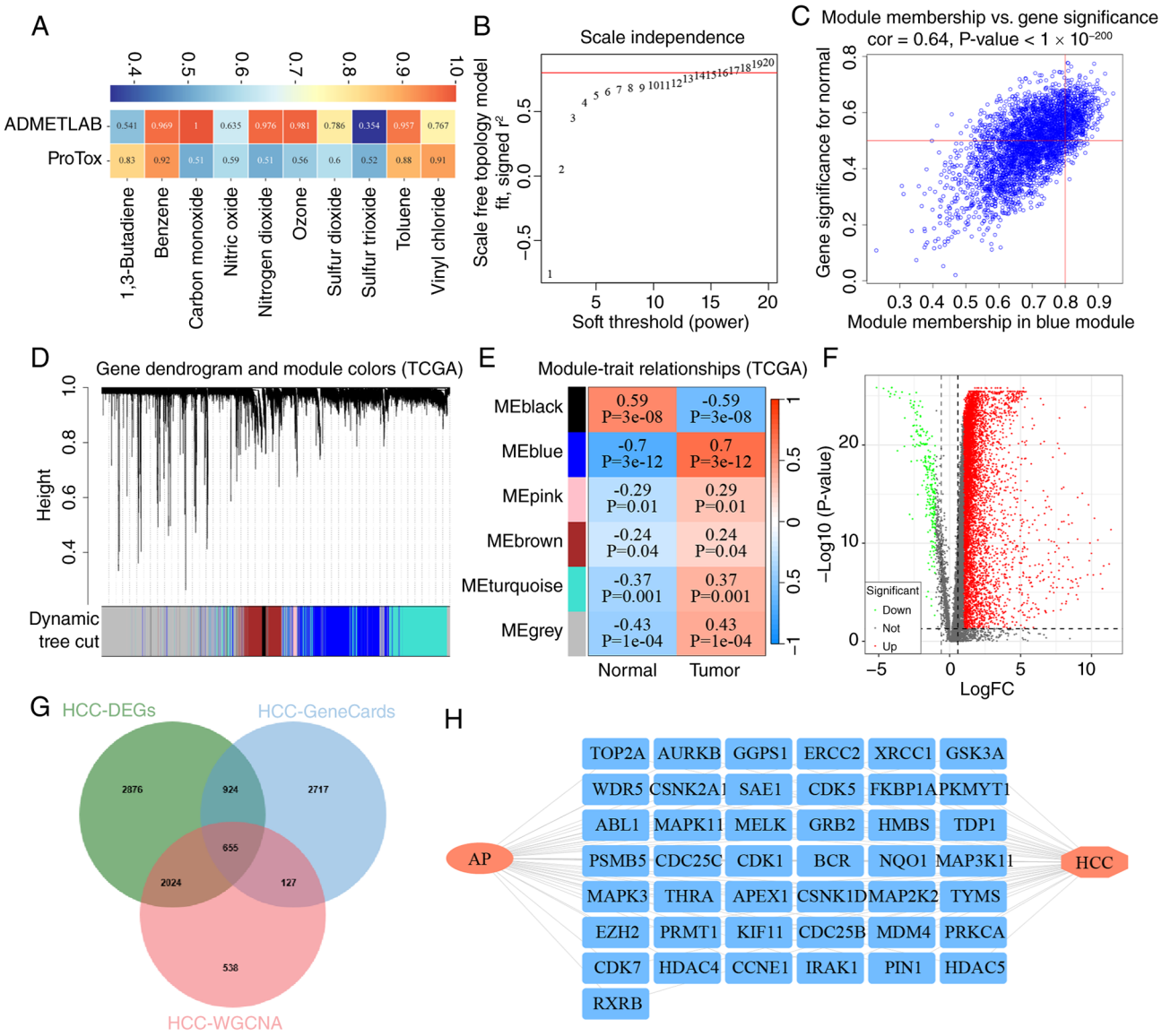

Figure 1. Identification of AP-HCC-G. (A) Heatmap of toxicity evaluation. (B) Scale independence. (C) Correlation between the gene significance and module membership. (D) Cluster dendrogram of weighted gene co-expression network analysis. (E) Module-trait correlation heatmap. (F) Volcano plot of DEGs. (G) Venn diagram of DEGs, module genes, and genes from GeneCards. (H) Correlation plot of AP-HCC-G. AP-HCC-G, AP-related and HCC-related genes; AP, air pollutant; HCC, hepatocellular carcinoma; cor, correlation coefficient; TCGA, The Cancer Genome Atlas; DEGs, differentially expressed genes; FC, fold change; ME, module eigengene.

multivariate Cox analysis (HR, 2.076; 95% CI, 1.548-2.783; P<0.001; Fig. S4H) showed that the AP-HCC-PS exhibited predictive performance independent of clinical features. Decision curve analysis (DCA) revealed that the AP-HCC-PS yielded a higher net benefit than other clinical features across a range of thresholds (Fig. S4I). The ROC curve suggested that the AP-HCC-PS had better predictive performance (AUC, 0.768; Fig. 2B) than clinical features, and had good predictive performance in predicting 1-year (AUC, 0.768), 3-year (AUC, 0.679) and 5-year (AUC, 0.657) survival (Fig. 2A). In the training set, the model sensitivity decreased with longer prediction horizons, which may be partly attributable to the limited sample size and the introduction of additional noise over time. The ICGC cohort and the additional validation cohort GSE14520 further validated the generalizability and robustness of the model (Figs. S5A-G and S6A-G).

Identification of AP-HCC-DMs. LASSO (Fig. 3A and B), Random-forest algorithm (Fig. 3C and D) and SVM-RFE (Fig. 3E and F) methods identified 16, 7 and 15 genes, respectively. After taking the intersection, PKMYT1 and PSMB5 were identified as AP-HCC-DMs (Fig. 3G; Table SVIII). The AUCs indicated that PKMYT1 (AUC, 0.971) and PSMB5 (AUC, 0.965) showed high diagnostic value (Fig. 3H).

Construction of prognostic and diagnostic nomograms. The prognostic nomogram (C-index, 0.722; 95% CI, 0.674-0.769; Fig. S7B and D) and the diagnostic nomogram (C-index, 0.984; 95% CI, 0.972-0.995; Fig. S7A and C) showed good accuracy in predicting HCC incidence and prognosis. DCA revealed that both prognostic and diagnostic nomograms yielded a superior net benefit across a range of thresholds, indicating favorable clinical accuracy (Fig. S7E and F). The

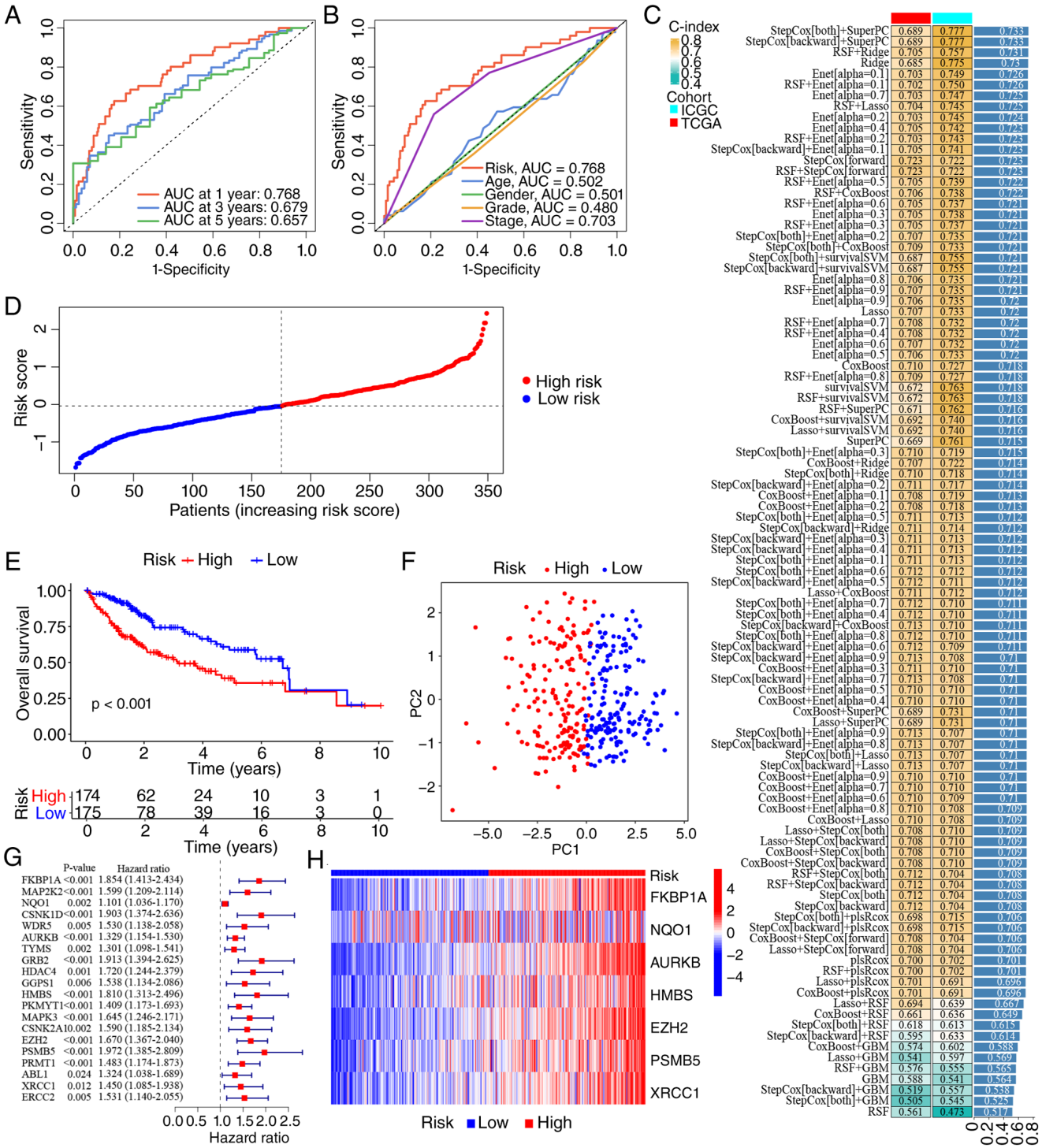

Figure 2. Construction and validation of AP-HCC-PS. (A) ROC curves predicting 1-, 3-, and 5-year survival. (B) ROC curve of clinical features. (C) C-index values of 101 machine learning models. (D) Risk curves based on AP-HCC-PS. (E) Kaplan-Meier survival curve based on AP-HCC-PS (chi-square test). (F) PC analysis based on AP-HCC-PS. (G) Univariate Cox regression analysis. (H) Heatmap of signature gene expression. TCGA, The Cancer Genome Atlas; ICGC, International Cancer Genome Consortium; ROC, receiver operating characteristic; AUC, area under ROC curve; PC, principal component; AP, air pollutant; HCC, hepatocellular carcinoma; AP-HCC-PS, AP-related and HCC-related prognostic signature; Lasso, least absolute shrinkage and selection operator; StepCox, stepwise Cox regression; survival-SVM, survival support vector machine; CoxBoost, Cox proportional hazards model with componentwise likelihood-based boosting; Ridge, ridge regression; SuperPC, supervised principal components; Enet, elastic net; plsRcox, partial least squares regression for Cox models; RSF, random survival forest; GBM, gradient boosting machine.

DCA demonstrated that both the AP-HCC-PS risk score and the integrated clinical nomogram achieved a higher net benefit than individual clinical features over a wide spectrum of threshold probabilities, with the nomogram outperforming the risk score alone (Fig. S7G).

Molecular docking analysis. The molecular docking analysis showed that the predicted docking scores between the AP-HCC-PS/AP-HCC-DM targets and the 10 APs were generally low, with average binding energies ranging from -2 to -3 kcal/mol. These findings provided only exploratory support

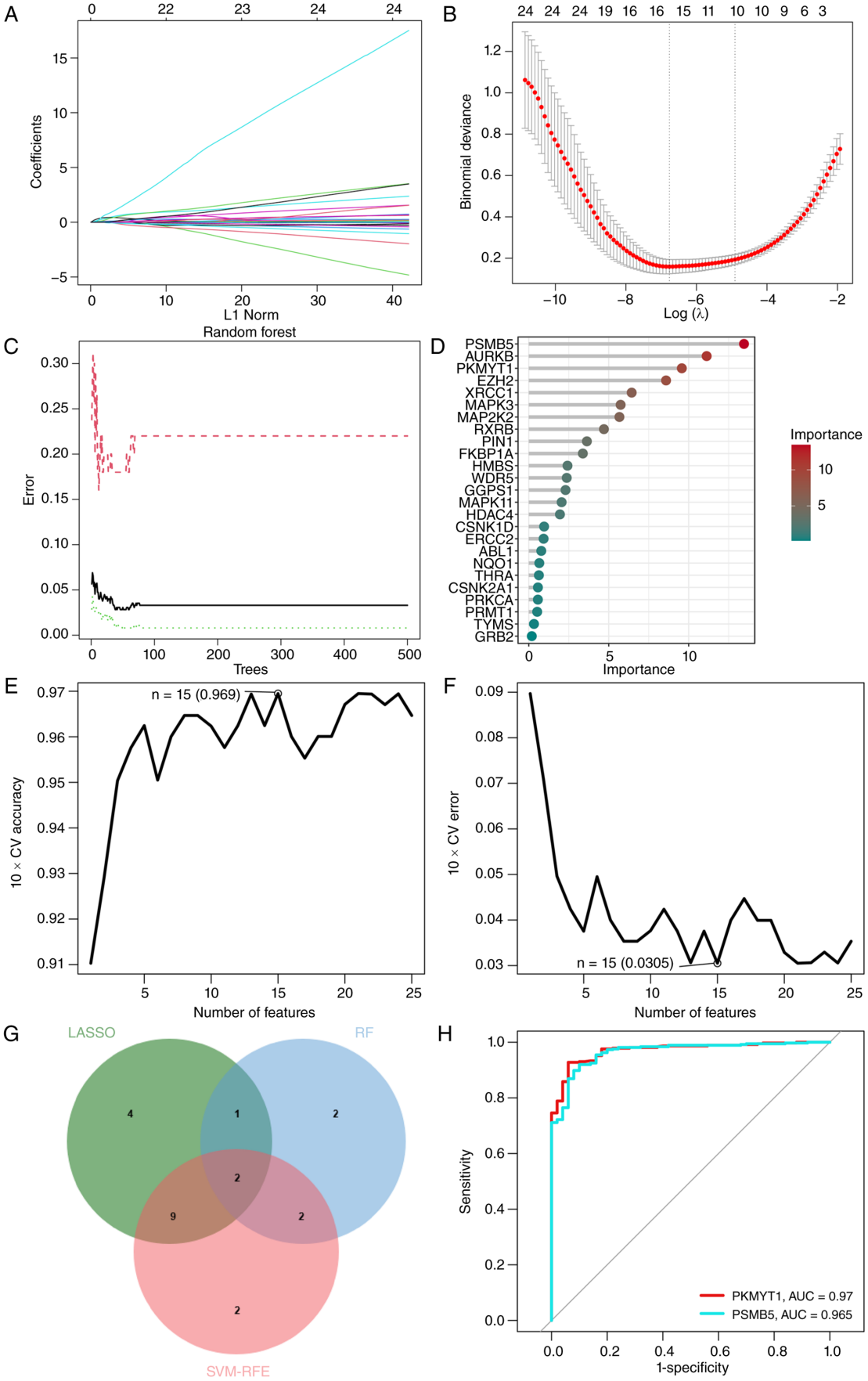

Figure 3. Identification of AP-HCC-DM using machine learning. (A) LASSO coefficient profiles. (B) LASSO lambda. (C) RF error rate. (D) RF importance ranking of AP-HCC-G. (E) SVM-RFE accuracy of AP-HCC-G (n=number of features). (F) SVM-RFE error of AP-HCC-G (n=number of features). (G) Venn diagram of genes identified by 3 machine learning methods. (H) ROC curve of 2 AP-HCC-DM. AP, air pollutant; HCC, hepatocellular carcinoma; AP-HCC-DM, AP-related and HCC-related diagnostic marker; AP-HCC-G, AP-related and HCC-related genes; ROC, receiver operating characteristic; AUC, area under ROC curve; SVM-RFE, support vector machine recursive feature elimination; CV, cross-validation.

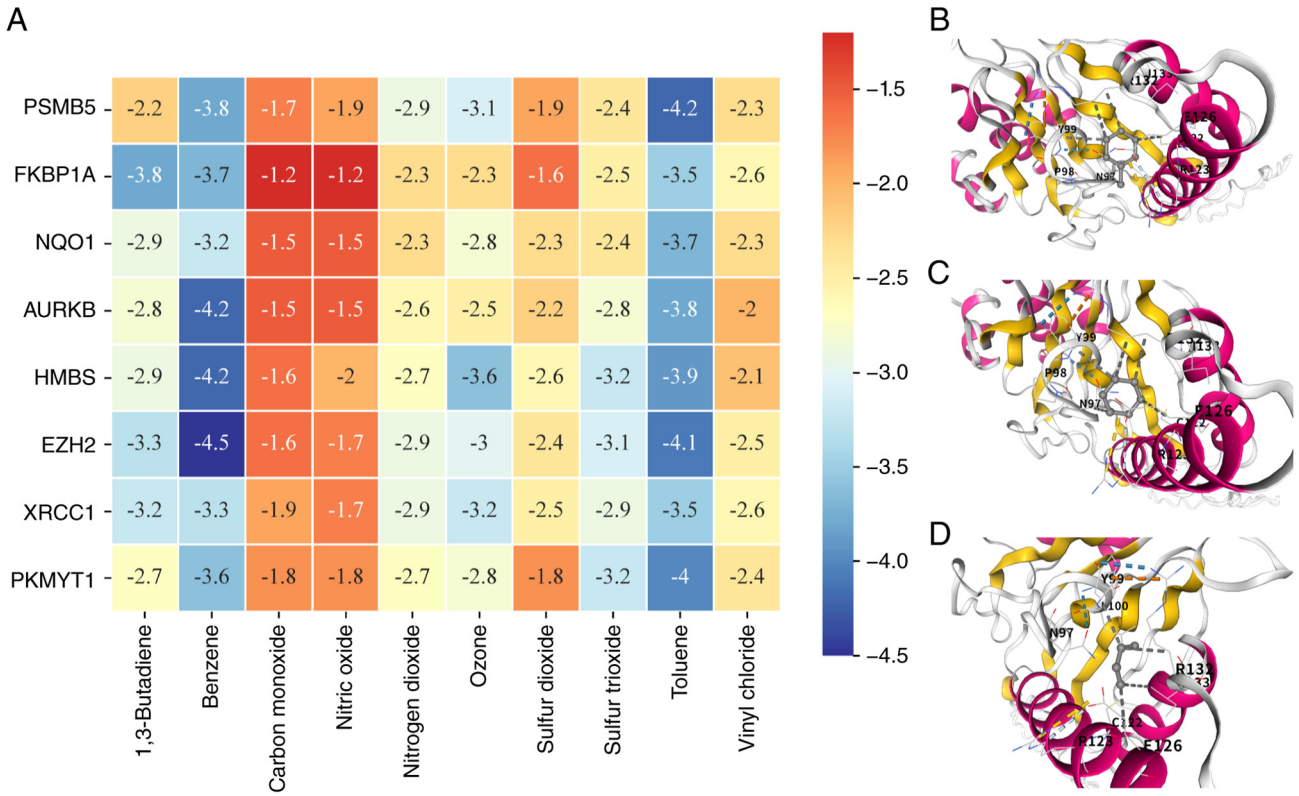

Figure 4. Molecular docking analysis (lower energy indicates stronger binding affinity). (A) Heatmap of binding energies (kcal/mol) of molecular docking analysis. Molecular docking images of (B) PSMB5 and toluene, (C) PSMB5 and benzene and (D) PSMB5 and 1,3-butadiene.

for potential target-pollutant interactions (Fig. 4A; Table SIX). Among these exploratory results, PSMB5 showed relatively lower docking scores with toluene (-4.2; Fig. 4B), benzene (-3.8; Fig. 4C) and 1,3-butadiene (-2.2; Fig. 4D), suggesting that it may warrant further investigation. Furthermore, molecular docking of known inhibitors for the seven targets was performed to aid the interpretation of the AP-gene relationships. For all seven targets, the inhibitor-target binding energies were ≤-6 kcal/mol, whereas the pollutant-target docking scores were substantially weaker (Table SX). This result confirmed the robustness of the docking procedure and suggested that AP-gene interactions differ from drug-target inhibition: While drugs achieve rapid inhibition through strong binding, APs may modulate target traits through long-term and cumulative effects on gene expression.

Knockdown of PSMB5 inhibits the migration of HCC cells. After intersecting the AP-HCC-PS with the two AP-HCC-DMs (Fig. S8A), PSMB5 was identified as the hub gene, and its high expression was found to be associated with poor prognosis in HCC (Fig. S8B).

RT-qPCR results showed that PSMB5 was highly expressed in HCC cells (Fig. 5A; Table SXI). PSMB5 expression was significantly decreased following siRNA-mediated knockdown (Fig. 5B; Table SXII). The wound-healing assay showed no significant difference among the three groups at 0 h (Fig. 5C). At 24 h, the PSMB5 siRNA group differed significantly from both the blank group (P=0.0082) and the NC group (P=0.0003), while no significant difference was found between the blank and NC groups. At 48 h, the wound-healing rate in the PSMB5 siRNA group was significantly lower than that in the blank group (P<0.0001) and the NC group (P=0.0002) (Fig. 5C-I; Table SXIII).

Single-cell analysis and immune-cell infiltration analysis of the hub gene. PSMB5 mainly interacts with its family members and is involved in biological functions such as immune responses, oxidative stress and classical cancer pathways (Fig. S8C). scRNA-seq analysis was strategically employed to resolve the cellular context of PSMB5. scRNA-seq results showed that nCount RNA was negatively correlated with percent.mt (r=-0.72) and positively correlated with nFeature RNA (r=0.54) (Fig. 6A). Based on the elbow plot, the top 10 principal components were selected for subsequent analysis (Fig. S9C). Six clusters were identified (resolution, 0.2; Figs. 6B, and S9A and B), which were subsequently annotated into five cell types (Fig. 6C and D): B cells, stromal cells, regulatory T cells (Tregs), hepatocytes and cytotoxic T cells (CTLs). Among the immune cell populations, PSMB5 was mainly expressed in Tregs (Figs. 6E and S9E). CIBERSORT revealed a positive correlation between PSMB5 and Tregs (Fig. 6F; Table SXIV). Tregs (sender) and CTLs (receiver) exhibited the highest interaction number, which was mainly mediated by the MIF-(CD74+CXCR4) axis (Figs. 6G and H, and S9D). These findings suggested that PSMB5 may be related to Treg activity and may participate in Treg-CTL communication via the MIF-(CD74+CXCR4) axis, potentially contributing to an immunosuppressive tumor microenvironment in HCC.

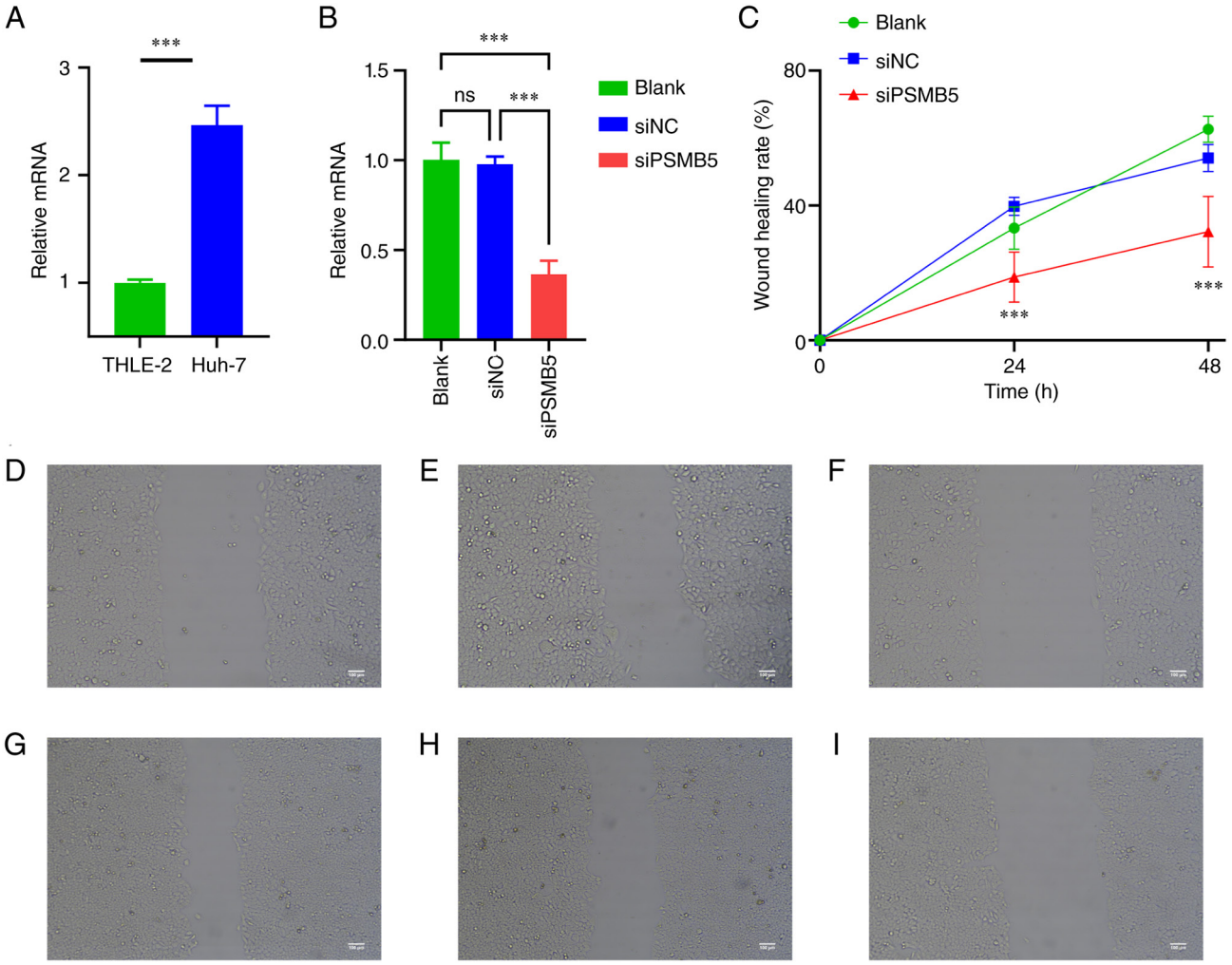

Figure 5. RT-qPCR, siRNA and wound-healing experiments for PSMB5. (A) RT-qPCR of PSMB5. (B) RT-qPCR of PSMB5 after siRNA transfection. (C) Wound healing rate. Representative images of wound-healing assay for groups (D) Blank-24 h, (E) siNC-24 h, (F) siPSMB5-24 h, (G) Blank-48 h, (H) siNC-48 h and (I) siPSMB5-48 h (scale bars, 100 μm). ***P<0.001; ns, no significance. siRNA, small interfering RNA; RT-qPCR, reverse transcription quantitative PCR; NC, negative control.

Discussion

Air pollutants are important environmental risk factors for the development of HCC and poor prognosis. Epidemiological studies and meta-analyses have shown that long-term exposure to multiple atmospheric pollutants is significantly associated with increased liver cancer incidence and mortality. For example, Odat *et al* (16) reported in a meta-analysis that nitrogen dioxide exposure was significantly associated with an increased risk of liver cancer. Ecological and occupational studies have also reported positive associations between exposure to vinyl chloride, benzene and sulfur oxides and increased HCC risk (2,9,17). In addition, the longitudinal study by Chin *et al* (5) showed that long-term post-diagnosis exposure to $PM_{2.5}$ and NO_2 was significantly associated with shorter survival in patients with HCC. Furthermore, limited experimental evidence further suggests that exposure to 1,3-butadiene may induce abnormalities in DNA-related processes and increase genomic instability in mouse liver tissue, whereas vinyl chloride exposure may contribute to hepatocarcinogenesis by inducing genomic mutations and chronic liver injury (18,19). Furthermore, nitric oxide and ozone have been reported to promote oxidative stress, inflammatory responses, angiogenesis and activation of oncogenic pathways, thereby facilitating HCC progression and reducing survival outcomes (20-22). Collectively, these findings suggest that APs are important factors involved in the initiation and progression of HCC and warrant further investigation.

In the present study, 43 AP-HCC-Gs were obtained by intersecting AP-related genes with HCC-related genes. GO and KEGG enrichment analyses suggested that AP-HCC-Gs were associated with biological functions and pathways potentially relevant to HCC progression, including cell growth and proliferation, metabolic reprogramming, oxidative stress and the immune response. The constructed AP-HCC-PS incorporated seven AP-HCC-Gs and showed good prognostic predictive ability. The AP-HCC-PS was further intersected with two AP-HCC-DMs to obtain PSMB5 as the hub gene. PSMB5 was mainly expressed in Tregs and was associated with the survival of HCC. A prognostic nomogram was also constructed, which demonstrated favorable predictive performance in multiple datasets. However, given that the model was developed after feature selection and internal optimization, this result should be interpreted cautiously until confirmed in

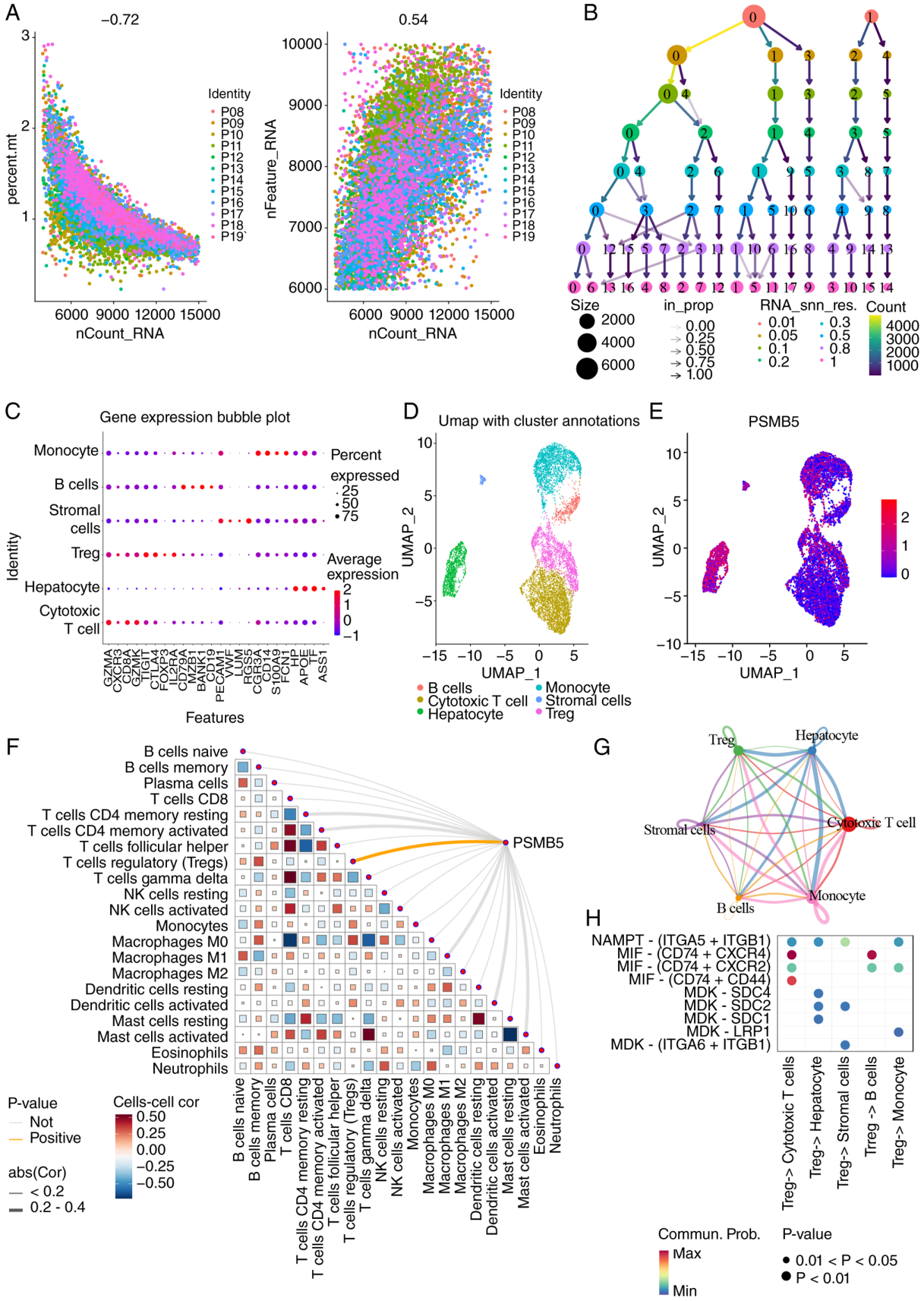

Figure 6. Single-cell RNA sequencing analysis and immune-cell infiltration analysis. (A) Correlation between percentage of mitochondrial genes and total number of unique molecular identifiers. (B) Selection of clustering resolutions. (C) Marker genes of cell types. (D) Annotated cell types. (E) PSMB5 expression in cells. (F) Correlation between PSMB5 and immune-cell infiltration abundance. (G) Receptor-ligand for cell communication. (H) Circle plot of interaction number in cell-cell communication. UMIs, unique molecular identifiers; nCount RNA, total RNA counts; nFeature RNA, number of detected RNA features; percent.mt, percentage of mitochondrial genes; RNA snn res., RNA shared nearest neighbor resolution; UMAP, uniform manifold approximation and projection; PSMB5, proteasome 20S subunit beta 5; Treg, regulatory T cell; NK, natural killer; cor, correlation; abs(Cor), absolute correlation coefficient; Commun. Prob., communication probability.

independent external cohorts. Molecular docking suggested that these targets (AP-HCC-PS and AP-HCC-DMs) may have potential binding affinity for the 10 APs.

As a member of the PSMB family, PSMB5 is the core subunit of the 20S proteasome (23). High PSMB5 expression can promote proteasome overactivation, which leads to the degradation of tumor suppressors and cell cycle checkpoint inhibitors, thereby promoting HCC (24,25). Additionally, PSMB5 and its family members may contribute to the formation of an immunosuppressive niche in HCC by affecting antigen presentation by major histocompatibility complex class I as well as the oxidative stress pathway (26,27). Notably, the present study revealed that PSMB5 was expressed in T cells and positively correlated with Tregs. Upregulation of PSMB5 expression may induce the release of IFNγ and activate the immunoproteasome (PSMB8/9/10) to produce specific antigenic peptides, which promotes sustained crosstalk between antigen-presenting cells and Tregs, and thus, enhances immunosuppressive effects (28,29). Meanwhile, increased abundance of Tregs in HCC has been shown to be associated with CD8 T-cell impairment and poor prognosis in HCC (30). The present study also demonstrated that Tregs communicated with CTLs via MIF-(CD74+CXCR4). MIF signaling has been shown to inhibit T-cell antitumor activity (31,32). These findings raise the possibility that AP-related molecular signatures may be linked to PSMB5 expression in Tregs and to an immunosuppressive tumor microenvironment in HCC.

PKMYT1 is a cell cycle-regulated kinase. Overexpression of PKMYT1 attenuates the interaction between GSK3β and β-catenin, and activates β-catenin/T-cell factor signaling, which could promote the migration and metastasis of HCC cells (33). PKMYT1 knockdown has been found to mediate autophagy and induce apoptosis via MAPK and PI3K/Akt/mTOR signaling pathways to inhibit the proliferation and migration of HCC cells (34). In addition, PKMYT1 levels in breast cancer are regulated by estrogen receptor α (ERα) and E2F family members, and are associated with CDK4/6 inhibitor resistance, while ERα and E2F are also important targets for HCC (35).

The other six AP-HCC-Gs that made up the AP-HCC-PS were closely associated with the prognosis of HCC. AURKB is a protein serine/threonine kinase and can directly activate the expression of cyclin D1, thereby promoting the proliferation of HCC cells (36). AURKB could also promote HCC progression via the PI3K/AKT/mTOR pathway, while its downregulation markedly inhibits HCC cell proliferation, migration and invasion (37). Additionally, upregulation of AURKB can induce the hyperpolyploidization of hepatocytes and their transformation into HCC cells (38). EZH2, a histone methyltransferase, primarily promotes tumorigenesis through epigenetic modifications such as catalyzing the trimethylation of histone H3 lysine 27, leading to the suppression of tumor suppressors such as p21 and chromodomain helicase DNA binding protein 5 (39,40). EZH2 also promotes HCC progression by inhibiting pyruvate kinase L/R expression to inactivate natural killer cells residing in the liver (41). FKBP1A is a cis-trans prolyl isomerase with elevated expression in HCC. Knockdown of FKBP1A reduces cell proliferation, migration and invasion, and induces autophagy via the PI3K/AKT/mTOR signaling pathway (42). HMBS encodes hydroxymethylbilane synthase and serves an indirect role in heme biosynthesis (43). The high expression of HMBS may stem from the adaptive response of tumor cells to metabolic stress, including the compensatory upregulation of 5'-aminolevulinate synthase 1 or activation of Wnt/β-catenin signaling, but the loss of its core function due to bi-allelic HMBS inactivation could lead to massive accumulation of toxic substrates such as α-lipoic acid and metabolic disorders, worsening HCC progression (43,44). NQO1 is a flavoenzyme that catalyzes the two-electron reduction of quinone to hydroquinone (45). NQO1 is highly expressed in HCC, and promotes HCC cell proliferation and migration by activating PI3K/AKT and MAPK/ERK signaling (46). Furthermore, NQO1 has been used for the development of targeted anticancer drugs because of its ability to generate compounds capable of DNA-alkylated or reactive oxygen species when hydroquinone is unstable (47,48). XRCC1 primarily functions as a scaffold protein in base excision repair and single-strand break repair to maintain genomic stability (49). XRCC1 overexpression can lead to anticancer drug resistance (50). Targeted inhibition of XRCC1 in HCC cells could increase the cytotoxicity of cisplatin (51). The classic oncogene STAT3 has been demonstrated to upregulate the expression levels of XRCC1, thereby promoting cancer cell proliferation (52,53). Low XRCC1 expression is associated with good prognosis in cancer types such as HCC, glioblastoma and lung adenocarcinoma, and is also associated with immune checkpoint gene expression (54).

By identifying molecular markers for HCC prognosis and pathogenesis, the present study provided novel biomarkers for early diagnosis and prognostic prediction, filling a critical knowledge gap in this research field. However, there are still certain limitations. First, the present study was mainly based on bioinformatics analyses of publicly available datasets. Although preliminary experimental validation of PSMB5 was performed using RT-qPCR, siRNA transfection and wound-healing assays, the evidence remains limited and protein-level validation is lacking. Additional *in vitro* and *in vivo* experiments, such as western blot analysis, immunohistochemistry and functional perturbation assays, are needed to further strengthen the conclusions. Second, the selected APs were not quantified in blood or liver tissue, and multiple cohorts lacked individualized exposure data. Therefore, dose-response relationships and causal inference could not be established. Future studies integrating exposure assessment, toxicokinetic analysis and longitudinal designs are needed to better define the relationship between pollutant exposure and gene expression. Third, although cross-validation and external validation were performed, the machine learning workflow involved multiple feature selection and model selection steps, and feature selection was not evaluated within a fully nested resampling framework. Thus, the performance of the diagnostic markers and nomogram may still be optimistic and requires further validation in larger independent cohorts. Finally, the molecular docking analysis should be interpreted cautiously. Because most predicted pollutant-target binding energies were weak and because the biological relevance of conventional docking for small gaseous, volatile or reactive pollutants remains uncertain, these results should be regarded as exploratory support for target prioritization rather than mechanistic evidence of direct binding.

In conclusion, the present study identified seven AP-HCC-Gs associated with HCC prognosis and two AP-HCC-Gs associated with HCC incidence. Among them, PSMB5 may be a key target linking APs and HCC. In future studies, it is recommended to extract the pollutants for *in vitro* experiments, including flow cytometry, Cell Counting Kit-8 and western blot assays, to further validate their effects on immune responses and the underlying molecular mechanisms.

Acknowledgements

Not applicable.

Funding

This work was supported by a grant from the Medical Scientific Research Project of Jiangsu Commission of Health (grant no. M2024056).

Availability of data and materials

The data generated in the present study are available from the corresponding author on reasonable request.

Authors' contributions

XW and HQ conceptualized and designed the study, developed the analytical strategy, performed the statistical and bioinformatics analyses, used the relevant software for data analysis and visualization, wrote the original draft, and reviewed and edited the manuscript. WZ and HL contributed to the study design, supervised the study, administered the project, reviewed the analytical results, and contributed to writing the original draft. JJ, YL and LG contributed to data processing, statistical analysis, interpretation of the analytical results, and reviewing and editing the manuscript. XW and WZ checked and confirmed the authenticity of all the raw data. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Wu S, Powers S, Zhu W and Hannun YA: Substantial contribution of extrinsic risk factors to cancer development. Nature 529: 43-47, 2016.
2. Cicalese L, Raun L, Shirafkan A, Campos L, Zorzi D, Montalbano M, Rhoads C, Gazis V, Ensor K and Rastellini C: An ecological study of the association between air pollution and hepatocellular carcinoma incidence in texas. Liver Cancer 6: 287-296, 2017.
3. De Guzman R and Schiller J: Air pollution and its impact on cancer incidence, cancer care and cancer outcomes. BMJ Oncol 4: e000535, 2025.
4. So R, Chen J, Mehta AJ, Liu S, Strak M, Wolf K, Hvidtfeldt UA, Rodopoulou S, Stafoggia M, Klompmaker JO, *et al*: Long-term exposure to air pollution and liver cancer incidence in six European cohorts. Int J Cancer 149: 1887-1897, 2021.
5. Chin WS, Pan SC, Huang CC, Chen PJ and Guo YL: Exposure to air pollution and survival in follow-up after hepatocellular carcinoma. Liver Cancer 11: 474-482, 2022.
6. Barouki R, Samson M, Blanc EB, Colombo M, Zucman-Rossi J, Lazaridis KN, Miller GW and Coumoul X: The exposome and liver disease-how environmental factors affect liver health. J Hepatol 79: 492-505, 2023.
7. VoPham T and Jones RR: State of the science on outdoor air pollution exposure and liver cancer risk. Environ Adv 11: 100354, 2023.
8. Kinaret PAS, Ndika J, Ilves M, Wolff H, Vales G, Norppa H, Savolainen K, Skoog T, Kere J, Moya S, *et al*: Toxicogenomic profiling of 28 nanomaterials in mouse airways. Adv Sci (Weinh) 8: 2004588, 2021.
9. Xie Q, Liu S, Li X, Wu Q, Luo Z, Fu X, Cao W, Lan G, Li D, Zheng W and Chen T: Dinuclear zinc(II) complexes containing (benzimidazol-2-yl)benzene that overcome drug resistance in hepatocellular carcinoma cells through induction of mitochondria fragmentation. Dalton Trans 43: 6973-6976, 2014.
10. Zhou G and Zhao X: Carcinogens that induce the A:T > T:A nucleotide substitutions in the genome. Front Med 12: 236-238, 2018.
11. Pu L, Zhang X, Pu C and Sun P: Air pollution-related immune gene prognostic signature for hepatocellular carcinoma: Network toxicology, machine learning and multi-omics analysis. Front Immunol 16: 1638445, 2025.
12. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods 25: 402-408, 2001.
13. Rocque B, Barbetta A, Singh P, Goldbeck C, Helou DG, Loh YE, Ung N, Lee J, Akbari O and Emamaullee J: Creation of a single cell RNASeq meta-atlas to define human liver immune homeostasis. Front Immunol 12: 679521, 2021.
14. Li H, Qu L, Yang Y, Zhang H, Li X and Zhang X: Single-cell transcriptomic architecture unraveling the complexity of tumor heterogeneity in distal cholangiocarcinoma. Cell Mol Gastroenterol Hepatol 13: 1592-1609.e9, 2022.
15. Wang J, Xu Y, Chen Z, Liang J, Lin Z, Liang H, Xu Y, Wu Q, Guo X, Nie J, *et al*: Liver immune profiling reveals pathogenesis and therapeutics for biliary atresia. Cell 183: 1867-1883.e26, 2020.
16. Odat RM, Hussain HAH, Alshwayyat S, Qasem HM, Marzouk HA, Alghraibah MS, Al-Hasaseen DM, Bani Salameh MM, Itayem A, Shtayat AA, *et al*: Global air pollution exposure and hepatocellular carcinoma: A systematic review and meta-analysis of epidemiological studies. BMC Public Health 25: 2972, 2025.
17. Lotti M: Do occupational exposures to vinyl chloride cause hepatocellular carcinoma and cirrhosis? Liver Int 37: 630-633, 2017.
18. National Toxicology Program: NTP toxicology and carcinogenesis studies of 1,3-butadiene (CAS No. 106-99-0) in B6C3F1 mice (inhalation studies). Natl Toxicol Program Tech Rep Ser 434: 1-389, 1993.
19. Weihrauch M, Lehnert G, Kockerling F, Wittekind C and Tannapfel A: p53 mutation pattern in hepatocellular carcinoma in workers exposed to vinyl chloride. Cancer 88: 1030-1036, 2000.
20. Park SJ, Lee SK, Lim CR, Park HW, Liu F, Kim S and Kim BC: Heme oxygenase-1/carbon monoxide axis suppresses transforming growth factor-β1-induced growth inhibition by increasing ERK1/2-mediated phosphorylation of Smad3 at Thr-179 in human hepatocellular carcinoma cell lines. Biochem Biophys Res Commun 498: 609-615, 2018.
21. Peng JP, Zheng S, Xiao ZX and Zhang SZ: Inducible nitric oxide synthase expression is related to angiogenesis, bcl-2 and cell proliferation in hepatocellular carcinoma. J Zhejiang Univ Sci 4: 221-227, 2003.
22. Peng J, Wang S, Wang Y, Yu W, Zha Y and Gao S: Effects of ozone exposure on lipid metabolism in Huh-7 human hepatoma cells. Front Public Health 11: 1222762, 2023.
23. Tanaka K: The proteasome: Overview of structure and functions. Proc Jpn Acad Ser B Phys Biol Sci 85: 12-36, 2009.

24. Liu L, Fu Y, Zheng Y, Ma M and Wang C: Curcumin inhibits proteasome activity in triple-negative breast cancer cells through regulating p300/miR-142-3p/PSMB5 axis. Phytomedicine 78: 153312, 2020.
25. Wang CY, Li CY, Hsu HP, Cho CY, Yen MC, Weng TY, Chen WC, Hung YH, Lee KT, Hung JH, *et al*: PSMB5 plays a dual role in cancer development and immunosuppression. Am J Cancer Res 7: 2103-2120, 2017.
26. Rousseau A and Bertolotti A: Regulation of proteasome assembly and activity in health and disease. Nat Rev Mol Cell Biol 19: 697-712, 2018.
27. McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, Yoder JA, Andrade J and de Jong JL: Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc Natl Acad Sci USA 113: E5014-E5023, 2016.
28. Ghannam K, Martinez-Gamboa L, Spengler L, Krause S, Smiljanovic B, Bonin M, Bhattarai S, Grützkau A, Burmester GR, Häupl T and Feist E: Upregulation of immunoproteasome subunits in myositis indicates active inflammation with involvement of antigen presenting cells, CD8 T-cells and IFNΓ. PLoS One 9: e104048, 2014.
29. Solé C, Domingo S, Penzo E, Moliné T, Porres L, Aparicio G, Ferrer B and Cortés-Hernández J: Downregulation of miR-885-5p promotes NF-κB pathway activation and immune recruitment in cutaneous lupus erythematosus. J Invest Dermatol 143: 209-219. e13, 2023.
30. Fu J, Xu D, Liu Z, Shi M, Zhao P, Fu B, Zhang Z, Yang H, Zhang H, Zhou C, *et al*: Increased regulatory T cells correlate with CD8 T-cell impairment and poor survival in hepatocellular carcinoma patients. Gastroenterology 132: 2328-2339, 2007.
31. Vilbois S, Xu Y and Ho PC: Metabolic interplay: Tumor macrophages and regulatory T cells. Trends Cancer 10: 242-255, 2024.
32. Waibl PJ, Hoyt-Miggelbrink A, Tomaszewski WH, Wachsmuth LP, Lorrey SJ, Wilkinson DS, Lerner E, Woroniecka K, Finlay JB, Ayasoufi K and Fecci PE: Antigen presentation by tumor-associated macrophages drives T cells from a progenitor exhaustion state to terminal exhaustion. Immunity 58: 232-246.e6, 2025.
33. Liu L, Wu J, Wang S, Luo X, Du Y, Huang D, Gu D and Zhang F: PKMYT1 promoted the growth and motility of hepatocellular carcinoma cells by activating beta-catenin/TCF signaling. Exp Cell Res 358: 209-216, 2017.
34. Wu F, Tu C, Zhang K, Che H, Lin Q, Li Z, Zhou Q, Tang B, Yang Y, Chen M and Shao C: Knockdown of PKMYT1 is associated with autophagy inhibition and apoptosis induction and suppresses tumor progression in hepatocellular carcinoma. Biochem Biophys Res Commun 640: 173-182, 2023.
35. Chen A, Kim BJ, Mitra A, Vollert CT, Lei JT, Fandino D, Anurag M, Holt MV, Gou X, Pilcher JB, *et al*: PKMYT1 is a marker of treatment response and a therapeutic target for CDK4/6 inhibitor-resistance in ER+ breast cancer. Mol Cancer Ther 23: 1494-1510, 2024.
36. Andrade AAR, Pauli F, Pressete CG, Zavan B, Hanemann JAC, Miyazawa M, Fonseca R, Caixeta ES, Nacif JLM, Aissa AF, *et al*: Antiproliferative activity of N-acylhydrazone derivative on hepatocellular carcinoma cells involves transcriptional regulation of genes required for G2/M transition. Biomedicines 12: 892, 2024.
37. Zhu G, Luo L, He Y, Xiao Y, Cai Z, Tong W, Deng W, Xie J, Zhong Y, Hu Z and Shan R: AURKB targets DHX9 to promote hepatocellular carcinoma progression via PI3K/AKT/mTOR pathway. Mol Carcinog 63: 1814-1826, 2024.
38. Lin H, Huang YS, Fustin JM, Doi M, Chen H, Lai HH, Lin SH, Lee YL, King PC, Hou HS, *et al*: Hyperpolyploidization of hepatocyte initiates preneoplastic lesion formation in the liver. Nat Commun 12: 645, 2021.
39. Porazzi P, Nason S, Yang Z, Carturan A, Ghilardi G, Guruprasad P, Patel RP, Tan M, Padmanabhan AA, Lemoine J, *et al*: EZH1/EZH2 inhibition enhances adoptive T cell immunotherapy against multiple cancer models. Cancer Cell 43: 537-551.e7, 2025.
40. Ma J, Zhang Y, Li J, Dang Y and Hu D: Regulation of histone H3K27 methylation in inflammation and cancer. Mol Biomed 6: 14, 2025.
41. Wang B, Liu Y, Liao Z, Wu H, Zhang B and Zhang L: EZH2 in hepatocellular carcinoma: Progression, immunity, and potential targeting therapies. Exp Hematol Oncol 12: 52, 2023.
42. Ye W, Shi Z, Zhou Y, Zhang Z, Zhou Y, Chen B and Zhang Q: Autophagy-related signatures as prognostic indicators for hepatocellular carcinoma. Front Oncol 12: 654449, 2022.
43. Molina L, Zhu J, Trépo E, Bayard Q, Amaddeo G; GENTHEP Consortium; Blanc JF, Calderaro J, Ma X, Zucman-Rossi J and Letouzé E: Bi-allelic hydroxymethylbilane synthase inactivation defines a homogenous clinico-molecular subtype of hepatocellular carcinoma. J Hepatol 77: 1038-1046, 2022.
44. Fontanellas A and Avila MA: Hydroxymethylbilane synthase (aka porphobilinogen deaminase): A novel metabolic tumor suppressor gene in hepatocellular carcinoma. J Hepatol 77: 912-914, 2022.
45. Ross D and Siegel D: The diverse functionality of NQO1 and its roles in redox control. Redox Biol 41: 101950, 2021.
46. Dimri M, Humphries A, Laknaur A, Elattar S, Lee TJ, Sharma A, Kolhe R and Satyanarayana A: NAD(P)H quinone dehydrogenase 1 ablation inhibits activation of the phosphoinositide 3-kinase/Akt serine/threonine kinase and mitogen-activated protein kinase/extracellular signal-regulated kinase pathways and blocks metabolic adaptation in hepatocellular carcinoma. Hepatology 71: 549-568, 2020.
47. Jaiswal AK: Regulation of genes encoding NAD(P)H:quinone oxidoreductases. Free Radical Bio Med 29: 254-262, 2000.
48. Yang X, Duan J and Wu L: Research advances in NQO1-responsive prodrugs and nanocarriers for cancer treatment. Future Med Chem 14: 363-383, 2022.
49. Wang J, Sadeghi CA, Le LV, Le Bouteiller M and Frock RL: ATM and 53BP1 regulate alternative end joining-mediated V(D)J recombination. Sci Adv 10: eadn4682, 2024.
50. Li G, Wang D, Zhai Y, Pan C, Zhang J, Wang C, Huang R, Yu M, Li Y, Liu X, *et al*: Glycometabolic reprogramming-induced XRCC1 lactylation confers therapeutic resistance in ALDH1A3-overexpressing glioblastoma. Cell Metab 36: 1696-1710.e10, 2024.
51. Zhang R, Niu Y and Zhou Y: Increase the cisplatin cytotoxicity and cisplatin-induced DNA damage in HepG2 cells by XRCC1 abrogation related mechanisms. Toxicol Lett 192: 108-114, 2010.
52. Wright G, Sonavane M and Gassman NR: Activated STAT3 is a novel regulator of the XRCC1 promoter and selectively increases XRCC1 protein levels in triple negative breast cancer. Int J Mol Sci 22: 5475, 2021.
53. Wright GM and Gassman NR: Glucose increases STAT3 activation, promoting sustained XRCC1 expression and increasing DNA repair. Int J Mol Sci 23: 4314, 2022.
54. Wang G, Li Y, Pan R, Yin X, Jia C, She Y, Huang L, Yang G, Chi H and Tian G: XRCC1: A potential prognostic and immunological biomarker in LGG based on systematic pan-cancer analysis. Aging (Albany NY) 16: 872-910, 2024.