Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

VOC-Associated Prognostic Targets MAOB, CPA3, and CCNA2 in Lung Adenocarcinoma: An Integrated Transcriptomic, Single-Cell, and Experimental Validation Study

Shilan Liu
Chengdu Fifth People's Hospital
Heng Li
Chengdu Fifth People's Hospital
Guilin Zeng
Chengdu Fifth People's Hospital
Yingxue Hu
Chengdu Fifth People's Hospital
Haitao Huang
Chengdu Fifth People's Hospital
Qin Yan
Chengdu Fifth People's Hospital
SHUAI LI
Chengdu Fifth People's Hospital
Yamao Li
Chengdu Fifth People's Hospital
Xiao Liu
Lxzl111@yeah.net

Chengdu Fifth People's Hospital

Research Article

Keywords: Volatile organic compounds, lung adenocarcinoma, single-cell transcriptomics, network toxicology, tumor immune microenvironment, molecular docking

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11440160/v1

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

Additional Declarations: No competing interests reported.

Abstract

Objective

To integrate network toxicology and single-cell transcriptomics approaches to investigate the effects of volatile organic compound (VOC) exposure on lung adenocarcinoma (LUAD), and to identify key genes and signaling pathways.

Methods

Targets of eight common VOCs (benzene, bromodichloromethane, chloroform, dibromochloromethane, 1,4-dichlorobenzene (1,4-DCB), ethylbenzene, methyl tert-butyl ether, and o-xylene) were predicted using multiple databases. Key genes were identified via differential expression, functional enrichment, Cox and LASSO regression. Single-cell sequencing was leveraged to analyze tumor heterogeneity and cell communication. We then used molecular docking and molecular dynamics simulations to confirm the binding of VOCs to their targets. Finally, the exposure concentrations of 1,4-DCB and ethylbenzene were screened, and the CCK-8 experiment was carried out. Optimal exposure concentrations of 1,4-DCB and ethylbenzene were determined by CCK-8 assay and applied to A549 and LAD2 cells for mRNA and protein detection.

Results

Twenty-two VOC-associated LUAD DEGs were identified and enriched in toxic response, drug metabolism, and amino acid metabolism pathways. Three prognostic genes were further pinpointed: CCNA2, MAOB, and CPA3. Single-cell analysis revealed that CCNA2 is highly expressed in immunosuppressive proliferating cells, while CPA3 and MAOB are highly expressed in mast cells. Cell communication analysis indicated that tumor tissues formed an immunosuppressive microenvironment. Molecular modeling confirmed that 1,4-DCB forms a stable association with MAOB and CCNA2. At 10 mg/L 1,4-DCB and 100 μM ethylbenzene, expression changes of CCNA2, MAOB, and CPA3 in A549 and LAD2 cells were consistent with the bioinformatic predictions.

Conclusion

CCNA2, CPA3, and MAOB were identified as potential core VOC-associated genes in LUAD and preliminarily validated in vitro. VOC exposure may reshape the tumor immune microenvironment through these three genes, thereby promoting LUAD initiation and progression.

1 Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide [1, 2], there are approximately 2.5 million new cases of lung cancer and 1.8 million deaths globally each year [3]. Lung adenocarcinoma (LUAD), the most common histological subtype of non-small cell lung cancer (NSCLC), is also the most prevalent type of lung cancer among never-smokers [4–6]. Although molecularly targeted therapies (such as EGFR tyrosine kinase inhibitors and ALK inhibitors) and immune checkpoint inhibitors (such as anti-PD-1/PD-L1 monoclonal antibodies) have significantly improved clinical outcomes for LUAD patients, the 5-year overall survival rate for patients with advanced disease remains low [7, 8]. The fundamental reasons for this are that the disease is often diagnosed at an advanced stage and that tumors exhibit high molecular and spatial heterogeneity [9]. In addition to smoking, environmental air pollution—particularly fine particulate matter ($PM_{2.5}$)—has been identified as a major environmental driver of lung cancer in nonsmokers. Recent studies have revealed that $PM_{2.5}$ does not primarily induce new gene mutations but rather selectively promotes the abnormal

proliferation of lung epithelial progenitor cells harboring oncogenic driver mutations such as EGFR or KRAS, thereby initiating the tumorigenic process [10]. This finding shifts the research paradigm of environmental carcinogenesis from "genetic damage-driven" to "microenvironment and cellular state reprogramming-driven," offering a new perspective on how environmental exposures reshape lung tissue homeostasis.

Volatile organic compounds (VOCs) are a class of organic chemicals that possess high saturated vapor pressure at standard temperature and pressure and readily volatilize into the gas phase. Based on their chemical structure, VOCs primarily include aromatic hydrocarbons (e.g., benzene, toluene, ethylbenzene, xylene), halogenated hydrocarbons (e.g., dichloromethane, trichloroethylene), and oxygen-containing organic compounds (e.g., formaldehyde, acetone) [11]. Their sources are widespread, encompassing industrial emissions, vehicle exhaust, building and decorative materials, furniture, and everyday consumer products; human exposure primarily occurs through inhalation, though it can also result from skin contact or dietary intake [12]. Notably, indoor VOC concentrations (such as those released during renovation, from furniture, and from cooking fumes) are often significantly higher than outdoor levels, posing a public health risk that cannot be ignored [13, 14]. Epidemiological evidence indicates that prolonged exposure to high-concentration VOC environments can significantly increase the risk of lung cancer [15]. From a molecular toxicological perspective, VOCs can promote malignant transformation in the lungs through multiple mechanisms, including inducing oxidative stress, inhibiting DNA damage repair, activating the CYP450 enzyme system to generate reactive metabolic intermediates, and disrupting the dynamic balance between cell proliferation and apoptosis. However, the systematic molecular mechanisms by which VOC exposure drives LUAD development remain unclear; in particular, there is a lack of in-depth analysis regarding how VOCs specifically influence different immune and stromal cell subsets within the tumor microenvironment, thereby reshaping intercellular communication and immune surveillance functions. To address these critical knowledge gaps, this study integrates network toxicology, multi-omics analysis, and computational modeling techniques to establish a multi-level, cross-scale research framework for systematically elucidating the biological mechanisms by which VOC exposure promotes LUAD progression. Specifically, we used network toxicology strategies based on multi-source databases, such as TCMSP, SwissTargetPrediction, and STITCH, to predict potential target genes for eight representative VOCs. By integrating clinical transcriptomic and survival data from the TCGA-LUAD cohort, we identified core target genes with independent prognostic value. Additionally, we used high-quality single-cell RNA sequencing datasets, such as GSE131907, to elucidate the specific expression patterns of key target genes across various cellular subpopulations in the tumor microenvironment. We also examined changes in ligand-receptor-mediated cellular interaction networks. Finally, we validated the binding stability, conformational dynamics, and key interacting residues between VOC small molecules and key target proteins through molecular docking and 100 ns all-atom molecular dynamics simulations. Preliminary experiments were conducted to verify the expression level of core genes and test the feasibility of the aforementioned bioinformatics analysis. This study aims to provide a theoretical basis and data support for early warning, risk stratification, and the discovery of intervention targets for VOC-associated lung cancer.

2 Methods

2.1 Data Acquisition and Preprocessing

LUAD transcriptomic datasets were retrieved from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and served as the training cohort, encompassing RNA sequencing (RNA-seq) data from 528 tumor tissue samples and 59 paired adjacent normal tissue samples, together with corresponding clinical and survival information. In parallel, the GSE118370 dataset was accessed from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geoprofiles/) as an independent validation cohort, containing microarray expression profiles derived from 7 matched pairs of LUAD and normal tissues. Prior to analysis, all datasets were preprocessed through background correction, gene identifier conversion, and expression normalization. To characterize

intratumoral heterogeneity, single-cell RNA sequencing data (GSE131907) from 11 paired LUAD tumor and adjacent normal lung tissue samples were also obtained.

2.2 Identification of VOC Targets

Considering that co-exposure to multiple VOCs is commonplace in real-world settings, eight frequently encountered VOCs were selected for analysis: benzene, bromodichloromethane, chloroform, dibromochloromethane, 1,4-dichlorobenzene (1,4-DCB), ethylbenzene, methyl tert-butyl ether (MTBE), and o-xylene. Potential human protein targets for these VOCs were predicted by cross-referencing four databases: SwissTargetPrediction (https://www.expasy.org/resources/swisstargetprediction), PharmMapper (http://www.lilab-ecust.cn/pharmmapper/), SEA (https://sea.bkslab.org/), and STITCH (http://stitch.embl.de/). Targets retrieved across databases were merged and deduplicated to generate a non-redundant comprehensive VOC target list.

2.3 Differential Expression Analysis of Transcriptomic Data

Differential expression analysis was performed on the TCGA-LUAD RNA-seq dataset using the limma package, with cut-off criteria of $|\log_2(\text{fold change})| > 1$ and Benjamini-Hochberg-adjusted p-value < 0.05 to define differentially expressed genes (DEGs) in LUAD versus normal control samples. The resulting DEGs were then overlapped with the compiled VOC target gene set, and the intersecting genes were retained as candidate genes for downstream functional characterization and validation.

2.4 Functional Enrichment Analysis of Candidate Genes

To explore the biological significance of candidate genes implicated in VOC-driven LUAD, functional enrichment analysis was carried out using the clusterProfiler package. Both Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment and Gene Ontology (GO) term enrichment analyses were conducted. Enrichment significance for each KEGG pathway and GO term was evaluated via hypergeometric testing, followed by Benjamini-Hochberg correction for multiple comparisons; terms with adjusted p-value < 0.05 were considered significantly enriched.

2.5 Selection of Prognostic Genes

To screen for candidate genes with independent prognostic relevance, univariate Cox proportional hazards regression was first applied to all candidate genes using the "survival" R package, based on LUAD patients with available survival data in the training cohort. Genes satisfying $P < 0.05$ and hazard ratio (HR) ≠ 1 were retained as having potential prognostic significance. These candidates were subsequently subjected to Least Absolute Shrinkage and Selection Operator (LASSO) penalized regression via the "glmnet" R package, with 10-fold cross-validation employed to determine the optimal regularization parameter and finalize the prognostic gene signature.

Expression levels of the selected key genes in LUAD versus normal tissues were further compared in the independent validation cohort GSE118370 using a two-sample t-test, to assess the reproducibility and robustness of their differential expression patterns.

2.6 Single-Cell RNA-seq Data Analysis

To delineate the cell-type-specific expression profiles of the identified key genes within the tumor microenvironment (TME), in-depth analysis was conducted using single-cell transcriptomic data. The gene-by-cell expression matrix was loaded into the Seurat (v4.4.0) package. High-quality cells were retained based on per-cell gene detection count (nFeature_RNA > 200) and mitochondrial gene proportion (percent.mt < 20); doublets were identified and excluded using scDblFinder (v1.22.0).

Expression data were normalized per cell using the "LogNormalize" method, and highly variable genes were identified via the "FindVariableFeatures" function to serve as feature genes for downstream analysis. The feature gene expression

matrix was subsequently scaled and centered with the "ScaleData" function, followed by dimensionality reduction via Principal Component Analysis (PCA). Batch effects across samples were corrected using harmony (v1.2.3).

Following dimensionality reduction, cell clustering was performed using the "FindNeighbors" and "FindClusters" functions, which construct a Shared Nearest Neighbor (SNN) graph and apply the Louvain algorithm to partition cells into distinct subpopulations. Each cluster was annotated by matching its top marker genes against canonical cell-type signatures. Non-linear dimensionality reduction was performed using the "RunUMAP" function to generate Uniform Manifold Approximation and Projection (UMAP) embeddings for visual representation of cell subgroup distributions and relationships. Immune checkpoint-related functional scores for individual cells were computed using the AddModuleScore function in Seurat, based on curated gene sets.

2.7 Cell-Cell Communication Analysis

Intercellular communication patterns within the TME were examined using CellChat v2 (http://cellchat.org/), which quantifies signaling interactions between tumor cells and surrounding cell populations. Communication networks were reconstructed by computing ligand-receptor interaction pairs, with interaction strength inferred from the expression levels of cognate ligands and receptors. In parallel, scDiffCom (v1.0.0), a dedicated single-cell RNA-seq tool for differential intercellular communication analysis, was applied to reconstruct ligand-receptor interaction networks. Statistically significant differences in ligand-receptor interactions between LUAD and normal tissue were identified, with multiple testing correction applied throughout to ensure analytical reliability. GO and KEGG functional enrichment analyses were subsequently performed on the significantly altered communication pathways to interpret their biological relevance. This approach enabled characterization of the regulatory mechanisms governing intercellular communication in LUAD versus normal tissue contexts.

2.8 Molecular Docking

Molecular docking was performed to evaluate the direct binding capacity of VOCs to key LUAD target proteins. Three-dimensional structures of VOCs were retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/) and preprocessed by removing solvent molecules, adding hydrogen atoms, and assigning Gasteiger charges. Target protein structures were downloaded from the AlphaFold database (https://alphafold.ebi.ac.uk/) and prepared by stripping co-crystallized ligands and water molecules and adding hydrogen atoms. Docking search spaces were defined around the predicted active sites of each target protein, with grid box dimensions and coordinates configured accordingly. Input files for both ligands and receptors were prepared using AutoDock Tools, and docking calculations were executed in AutoDock Vina to estimate binding free energies between VOCs and the target proteins.

2.9 MD Simulation

To assess the stability of the VOC–protein complexes identified through molecular docking, all-atom MD simulations were performed. Simulations were run using GROMACS 2021.1, employing the Amber14SB force field for protein parameterization and the Generalized Amber Force Field (GAFF) for ligand molecules. Each complex was embedded in a 10 × 10 × 10 nm cubic periodic box, solvated with TIP3P explicit water, and charge-neutralized by adding Na^+ or Cl^- counter-ions. After energy minimization to relieve steric clashes, sequential 100 ps equilibration runs were conducted under NVT and NPT ensembles, with temperature controlled at 298 K via the V-rescale thermostat and pressure maintained at 1 atm using the Parrinello-Rahman barostat. Production MD simulations of 100 ns were then carried out with a 2 fs integration time step, with coordinate snapshots saved every 20 ps.

Trajectory analyses included: root mean square deviation (RMSD) to monitor conformational stability, radius of gyration (Rg) to assess binding pocket compactness, and hydrogen bond counting to track dynamic ligand–protein interactions. Binding free energies were computed using the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) approach to quantitatively evaluate ligand–protein binding affinity. Together, these analyses characterized the dynamic

behavior of VOC ligands within their respective binding pockets and the contribution of key interacting residues, offering mechanistic insights into how VOC exposure may drive LUAD development.

2.10 Cell Culture and Processing

The human non-small cell lung cancer cell line A549 was obtained from the American Type Culture Collection (ATCC, Manassas, VA, USA). The LAD2 human mast cell line was obtained from Shanghai YaJi Biological (Shanghai, China). We thaw the cryopreserved A549 and LAD2 cells initially using a water bath at 37°C. A549 cells were seeded in DMEM complete medium, which contained 10% FBS, 1% streptomycin, and gentamicin. LAD2 cells were seeded in a complete DMEM medium. The medium contained 10% FBS and 1% streptomycin and gentamicin. It also contained 100 ng/mL recombinant human SCF. Cultivate both cell lines in a 37°C and 5% CO_2 incubator. When the A549 cells reached 80–90% confluence, the medium was discarded, the cells were washed twice with PBS and digested with 0.25% trypsin for two to three minutes. Digestion was then terminated by adding complete medium. The cells were then passaged at a 1:3 ratio for expansion. When the LAD2 cell density reached 1.0-1.5 × 10^6/mL, the cell suspension was harvested, centrifuged at 1,000 rpm for 5 min, the supernatant was discarded, and the cells were resuspended in fresh medium and passaged at a 1:2 ratio.

2.11 VOC Exposure Concentration Screening

In order to explore the exposure conditions of 1,4-DCB and ethylbenzene, a preliminary CCK-8 experiment was carried out in our study. Log-phase cells were harvested from the A549 and LAD2 groups, after which they were seeded at a density of 5 × 10^3 cells per well in a 96-well plate. There were three replicates in each group, and a blank control was established. According to the reported experimental in vivo, the experimental concentration of 1,4-DCB was divided into 0 (G0), 10 mg/L (G1), 50 mg/L (G2), and 150 mg/L (G3) from low to high [16]. A549 cells were added with complete medium containing the above concentration of 1,4-DCB, 200 μL/well. For LAD2 group, the concentration gradient of ethylbenzene was set to 0 (G0), 100 μM (G1), 500 μM (G2) and 1 mM (G3), which was based on previous studies [17]. The threshold setting is mainly based on the actual exposure level and the maximum saturation concentration in the human body. At the same time, 100 ng/mL SCF was added to the medium. G0 is the DMSO solvent control group, with a final concentration of ≤ 0.1%. After addition, the plates were sealed and placed in a closed culture system for incubation at 37°C and 5% CO_2. Following 24 hours of incubation, 10 μL of CCK-8 solution was added to each well. Then, the plates were incubated at 37°C in the dark for four hours. The OD values were measured at a wavelength of 450 nm using a microplate reader.

The screening criterion is the highest effective concentration at which cell viability is at least 80%. This ensures that any changes in gene expression are not caused by toxic-induced. Following the concentration screening experiments described above, the optimal concentrations of 1,4-DCB and ethylbenzene were added to the A549 and LAD2 cell lines to detect mRNA and protein expression levels.

2.12 Real-Time Quantitative Reverse Transcription PCR (RT-qPCR) Test

The A1 group of A549 cells was added the same amount of DMSO as above, while the experimental group (A2) was added a concentration of 1,4-DCB complete medium as determined by the preliminary experiment. The solvent control group (M1) of LAD2 cells was added the DMSO complete medium containing SCF, and the M2 group was added the same concentration of ethylbenzene complete medium containing SCF, as determined by the pre-experiment. Three replicates were set up in each group. After adding the sample, the plate was sealed and cultured in an independent closed culture system at 37°C and 5% CO_2 for 24 hours.

Collect cells in each group. In the A549 group, the medium was discarded and washed twice with PBS, and 1 mL RNAiso Easy was added to each well. After full lysis, it was transferred to a centrifuge tube. In LAD2 group, the cell suspension

was collected and centrifuged at 1000 rpm for 5 min. After discarding the supernatant, the cells were washed twice with PBS, and 1 mL RNAiso Easy was added to each tube. Next, Solution H was added, standing at room temperature for 5 min, and centrifuged at 12,000 × g at room temperature for 15 min. The supernatant was added with an equal volume of isopropanol to precipitate RNA, washed twice with 75% ethanol, dissolved in RNase-free water after drying, and RNA concentration and purity were determined by NanoDrop 2000. RNA samples from each group were reverse transcribed into cDNA using PrimeScript RT kit, and qPCR amplification was performed using TB Green Premix Ex Taq with GAPDH as internal reference. CCNA2 was detected in group A1 and group A2. Reaction conditions: 95°C for 30 s; 95°C denaturation for 5 s, 60°C annealing/extension for 30 s and 40 cycles; the melting curve is 65°C → 95°C. Three technical repetitions were set for each sample, and the relative expression level was calculated by $2^{-\Delta\Delta Ct}$ method.

2.13 Western Blotting (WB) Analysis

Collect cells from each group, wash them, add RIPA lysis buffer, and lyse on ice for 30 min. Centrifuge at 12,000×g at 4°C for 15 min. Then collect the supernatant, determine the protein concentration using the BCA method, add 5× loading buffer, denature on a boiling water bath for 10 min, and store at − 80°C for later use. Take denatured protein samples from each group (30 μg/well), perform SDS-PAGE electrophoresis, wet-transfer to a PVDF membrane, block with 5% skim milk for 1 h, incubate separately with CCNA2, MAOB, CPA3, and GAPDH primary antibodies (overnight at 4°C), wash three times with TBST, incubate with HRP-labeled secondary antibody at room temperature for 1 h, wash three times with TBST, perform ECL chemiluminescence detection, and analyze the grayscale images using ImageJ.

2.14 Statistical Analysis

All statistical analyses were conducted in R (version 4.3.0). Differential expression was assessed with the limma package, applying cut-offs of $|\log_2(\text{fold change})| > 1$ and Benjamini-Hochberg (BH)-adjusted $P < 0.05$. GO and KEGG enrichment analyses were implemented via the clusterProfiler package, with BH-adjusted $P < 0.05$ as the significance threshold. Associations between gene expression and overall survival were evaluated by univariate Cox proportional hazards regression ($P < 0.05$, HR ≠ 1); LASSO penalized regression with 10-fold cross-validation, implemented via the glmnet package, was used for final prognostic gene selection. Between-group expression differences in the validation cohort were assessed using a two-sample t-test. Single-cell-level comparisons were performed using the Wilcoxon rank-sum test with BH correction. A p-value threshold of 0.05 was applied throughout unless explicitly stated otherwise. Data were analyzed using GraphPad Prism software. For the preliminary experiment, a one-way ANOVA combined with Tukey's post hoc test was used; for the main experiment, a two-sided independent-samples t-test was used to compare the two groups. A $P < 0.05$ was considered statistically significant.

3 Results

3.1 Identification of Differentially Expressed Genes in LUAD and VOC Target Genes

To investigate differentially expressed genes in LUAD and their association with VOC exposure, we conducted differential expression analysis based on the TCGA-LUAD dataset, identifying a total of 1,952 DEGs in LUAD, with 900 upregulated genes and 1,052 downregulated genes (Fig. 1A).

Next, we predicted target genes for eight environmentally common VOCs using the SwissTargetPrediction, PharmMapper, SEA, and STITCH databases, yielding a total of 144 predicted target genes (Supplementary Fig. 1). Through set analysis, we observed a certain degree of overlap among these VOC targets (Fig. 1B).

By intersecting the DEGs with the VOC target genes (Fig. 1C), we identified 22 potential genes that met both criteria of differential expression in LUAD and being potential VOC targets. Among these, 9 genes were upregulated and 13 genes

were downregulated in LUAD tissues (Fig. 1E). These genes may play crucial roles in the process of VOC-induced LUAD development.

A: Volcano plot of DEGs between LUAD and normal tissues; yellow, blue, and gray dots indicate upregulated, downregulated, and non-significant genes; B: Set analysis of VOC target genes across eight compounds; C: Venn diagram of LUAD DEGs and VOC target genes; D: GO enrichment analysis of candidate genes (BP, biological process; MF, molecular function); E: Expression heatmap of 22 common target genes across samples (blue to red: low to high); F: KEGG pathway enrichment analysis of candidate genes.

3.2 Functional Enrichment Analysis of Potential Genes

To explore the functional associations of these 22 potential genes, we conducted GO and KEGG functional enrichment analyses. GO analysis (Fig. 1D) revealed that these genes were primarily enriched in biological processes (BP) such as response to toxic substances, organic hydroxy compound transport, positive regulation of secretion, and amine metabolism; in terms of molecular functions (MF), they were mainly enriched in oxidoreductase activity, carboxylesterase activity, and retinol binding. KEGG pathway analysis (Fig. 1F) further indicated that these genes were significantly enriched in drug metabolism-cytochrome P450 and multiple amino acid metabolism pathways, suggesting that VOCs may induce cancer by affecting various metabolic pathways in lung tissue. MAOA and MAOB participated in multiple metabolic pathways, with the tyrosine metabolism pathway showing the highest gene proportion and statistical significance.

3.3 Evaluation of Prognostic Value of Key Genes

We performed univariate Cox regression analysis on the 22 potential genes, identifying 7 genes significantly associated with prognosis (Fig. 2A). Among these, high expression of CCNA2, KIF11, and PFN2 was associated with poor prognosis (HR > 1), while high expression of CAT, MAOB, MAOA, and CPA3 was associated with better prognosis (HR < 1).

To further select the key gene combination with the most prognostic value, we employed LASSO regression analysis (Fig. 2B). Through 10-fold cross-validation to determine the optimal lambda value (Fig. 2C), we further identified four key genes, including CCNA2, CAT, MAOB, and CPA3, which demonstrated robust prognostic predictive ability during tumor progression (Fig. 2D).

We further validated in the validation set, under the threshold condition of $P < 0.1$, the expression patterns of CCNA2, MAOB, and CPA3 were consistent with the training set (Fig. 2E). The difference in CAT between LUAD and normal tissues was not significant ($P = 0.25$).

A: Forest plot of univariate Cox regression for 22 candidate genes; red and green squares indicate genes associated with poor (HR > 1) or favorable (HR < 1) prognosis; B: LASSO regression coefficient trajectories across lambda values; C: 10-fold cross-validation curve for optimal lambda selection (red dashed line); D: Kaplan-Meier survival curves of four prognostic genes stratified by expression level (red, high; blue, low); E: Box plots and heatmap of key gene expression in LUAD versus normal tissues in the validation cohort; two-sample t-test.

3.4 LUAD Heterogeneity and Expression Characteristics of Key Genes at the Single-Cell Level

We classified cells into eight major cell types based on marker gene expression patterns (Fig. 3B): T cells, NK cells, B cells/plasma cells, myeloid cells, mast cells, epithelial cells, endothelial cells, and fibroblasts (Fig. 3A). Comparing the distribution of cell types between tumor and normal tissue samples (Figs. 3C, 3D), we found a significant increase in the proportion of epithelial cells in tumor tissues, which may be a result of abnormal proliferation of lung epithelium in

tumors; the proportions of T cells and B cells/plasma cells were significantly increased, while the proportions of NK cells and myeloid cells were significantly reduced, suggesting remodeling of the immune microenvironment within the tumor.

Analyzing the expression characteristics of our previously identified key genes in different cell types, we found that CPA3 and MAOB were highly expressed in mast cells (Fig. 3E), while CCNA2 was mainly expressed in T cells, B cells/plasma cells, myeloid cells, and epithelial cells, with the highest expression in T cells (Fig. 3F). These results indicate that VOC target genes exhibit cell type-specific expression patterns, potentially reflecting their diverse roles in the tumor microenvironment.

3.5 T/NK Cell Subgroup Analysis and Association with VOC Target Genes

We further divided T/NK cells into eight subgroups: Naive CD8T/CD4T, IFN+ CD4T, Cytotoxic CD8T_1, Cytotoxic CD8T_2, Exhausted CD8T, Proliferating, Treg, and NK (Fig. 3G). Further analysis comparing the subgroup distribution in normal and tumor tissues (Figs. 3H, 3J) revealed that the proportions of Naive CD8T/CD4T, Exhausted CD8T, and Treg cells were significantly increased in tumor tissues, while the proportions of NK and Cytotoxic CD8T_2 cells were significantly reduced. This result indicates insufficient functional cytotoxic T cells and active immunosuppressive cells in tumor tissues, suggesting that LUAD tumors establish an immunosuppressive microenvironment.

Examining the marker gene expression patterns of these subgroups (Fig. 3I), we found that the Cytotoxic CD8T_2 subgroup highly expressed NKG7, KLRD1, and GNLY, indicating stronger killing capacity than Cytotoxic CD8T_1. The Proliferating subgroup, as a highly proliferative subgroup, specifically expressed genes such as MKI67 and TOP2A, while also highly expressing immune suppression-related genes such as PDCD1 and LAG3. We hypothesized that proliferating cells within the tumor are primarily immunosuppressive cells. To validate this hypothesis, we further scored cells based on immune checkpoint expression, and the results showed that Tregs had the highest immunosuppression scores, followed by Exhausted CD8T and Proliferating cells (Fig. 3K), suggesting that the Proliferating subgroup consists of immunosuppressive proliferating cells.

A: UMAP of all cells colored by annotated cell type; B: Dot plot of canonical marker genes per cell type; C: UMAP colored by tissue origin (tumor, yellow; normal, blue); D: Bar chart of cell-type proportions in tumor versus normal tissues; E: UMAP feature plots of key gene expression across cell types; F: Dot plot of CCNA2 expression across cell types; G: UMAP of T/NK cell subpopulations; H: T/NK cell subgroup composition by tissue origin; I: Marker gene expression across T/NK subpopulations; J: Proportional distribution of T/NK subgroups in tumor versus normal tissues; K: Violin plots of immune checkpoint scores across T/NK subgroups; Wilcoxon rank-sum test; ns, not significant; ****, $P < 0.0001$.

3.6 Cell Communication Networks in the LUAD Microenvironment

To decipher cellular interactions in the LUAD microenvironment, we analyzed cell communication networks in normal (Fig. 4A) and tumor tissues (Fig. 4B) using CellChat and scDiffCom tools. Comparing overall intercellular communication intensities between the two states, we observed that in the tumor state, communication intensities between multiple cell types and Cytotoxic CD8T_2 and Exhausted CD8T were generally enhanced, interactions between Cytotoxic CD8T_1 and other cells were significantly reduced, and cell communication between fibroblasts and myeloid cells was significantly weakened. This result further confirms the formation of an immunosuppressive microenvironment within the tumor.

From the comparison of overall signal intensity patterns in normal and tumor states (Fig. 4C), in the tumor state, the communication signals between fibroblasts, epithelial cells, proliferating cells, and mast cells with Cytotoxic CD8T_2 showed the most significant increase. The signaling axes involved in this process include CCL19:CXCR3, CXCL13:CXCR3, COL10A1:ITGB1_ITGA1, and SEMA7A:ITGB1_ITGA1 (Fig. 4E), indicating that the body attempts to recruit functional cytotoxic immune cells to kill tumors through various signaling pathways. Simultaneously, interactions between

proliferating cells and mast cells with Exhausted CD8T also increased significantly (Fig. 4D), indicating that tumors can promote the strengthening of immunosuppressive networks.

KEGG enrichment analysis based on receptor-ligand pairs showed that in the tumor state, cell adhesion and chemokine-related signaling pathways were significantly enhanced (Fig. 4F). We compared changes in signaling axis interaction intensities between normal and tumor states, with results showing that FAM3C:PDCD1, SIRPG:CD47, and CXCL13:CXCR3 signals were significantly increased in the tumor state (Fig. 4G). Specifically, proliferating cells can attract the accumulation of immune cells by secreting the inflammatory chemokine CXCL13. Additionally, various cells (including NK cells, Tregs, endothelial cells, B cells/plasma B cells, fibroblasts, Cytotoxic CD8T-1, and epithelial cells) bind to inhibitory receptor PDCD1 on the surface of Exhausted CD8T, Proliferating, and Cytotoxic CD8T_2 cells through the FAM3C ligand, suppressing T cell inflammatory activity and causing tumor immune escape. Exhausted CD8T can bind to the CD47 receptor on other cells through SIRPG, inhibiting the phagocytosis of exhausted CD8T cells by phagocytes, further promoting the immunosuppressive environment within the tumor. Within the CXCL13:CXCR3 network, proliferating cells act as the primary senders, releasing CXCL13 signals to a wide range of T-cell subsets (including IFN^+ CD4 T cells, cytotoxic CD8 T-1 cells, NK cells, Tregs, cytotoxic CD8 T-2 cells, exhausted CD8 T cells, and naive CD8 T/CD4 T cells), suggesting that proliferating cells may actively recruit and shape the composition of immune cells in the tumor microenvironment via the CXCL13-CXCR3 signaling pathway (Fig. 4H).

A/B: Heatmaps of intercellular communication strength in normal and tumor tissues; C: Top 20 upregulated sender–receiver cell-type pairs in LUAD by ORA score; D: Differential communication network between LUAD and normal tissues (red, increased; blue, decreased); E: Heatmap of significantly upregulated ligand-receptor pairs in LUAD; F: Top 20 upregulated ligand-receptor interactions in LUAD; G: Top 20 upregulated ligand-receptor pairs ranked by interaction strength; H: Chord diagram of cell communication patterns for three key ligand-receptor pairs.

3.7 Molecular Docking and MD Simulation of VOCs with Key Proteins

Our research systematically revealed the molecular network of VOCs inducing LUAD development. Through integrated transcriptomic analysis, we identified three key genes playing central roles in VOC carcinogenesis: CCNA2, CPA3, and MAOB. These genes cover multiple functional domains, including cell cycle regulation, various metabolic pathways, and immune regulation, constituting the molecular basis of VOC carcinogenic effects.

Through molecular docking techniques, we simulated the binding modes of different VOCs with CCNA2, CPA3, and MAOB protein structures (Table 1). 1,4-DCB exhibited comparable high affinities for MAOB and CCNA2, with binding energies of -5.8 kcal/mol and − 5.1 kcal/mol, respectively. Ethylbenzene had binding energies of -5.6 kcal/mol and − 5.3 kcal/mol with CPA3 and MAOB, respectively, indicating its ability to stably bind with target proteins.

1,4-DCB primarily binds in the MAOB pocket through hydrophobic interactions, with key residues including TYR393, PRO265, and HIS273 (Fig. 5A). Similarly, 1,4-DCB forms hydrogen bonds with PRO195 and THR310 residues of CCNA2 and multiple hydrophobic interactions in the binding pocket, with key residues including PRO195, PRO352, and VAL311 (Fig. 5C). Ethylbenzene primarily binds in the CPA3 pocket through hydrophobic interactions, with key residues including HIS304, ALA376, and ILE351 (Fig. 5B).

Table 1
Molecular docking binding energies of candidate compounds and their targets.

Drug	Target	Binding energy (kcal/mol)
MAOB	1,4-Dichlorobenzene	-5.8
MAOB	Ethylbenzene	-5.3
CPA3	Ethylbenzene	-5.6
CCNA2	1,4-Dichlorobenzene	-5.1

A: MAOB–1,4-DCB complex; panel (a), 2D interaction diagram; panel (b), 3D binding conformation with key residues. B: CPA3–ethylbenzene complex. C: CCNA2–1,4-DCB complex.

Subsequently, we conducted MD simulations, analyzing the RMSD (Fig. 6A), Rg (Fig. 6B), Solvent Accessible Surface Area (SASA) (Fig. 6C), and Root Mean Square Fluctuation (RMSF) (Fig. 6D) during the simulation process. The results showed that the CCNA2-1,4-DCB complex maintained relatively stable RMSD (approximately 0.15 nm), radius of gyration (approximately 1.8 nm), and solvent accessible surface area (SASA, approximately 130 nm^2) during the 100 ns simulation, with overall low RMSF values for the complex, with most residues having RMSF values below 0.2 nm, except for a few high fluctuation regions at the N-terminus, indicating good stability of this binding. Similarly, the MAOB − 1,4-DCB complex also exhibited relative stability throughout the simulation. The CPA3 - Ethylbenzene complex underwent conformational adjustment during the initial 30 ns before stabilizing, with the final RMSD maintained at approximately 0.5 nm, radius of gyration stable at approximately 2.2 nm, SASA gradually decreasing to approximately 180 nm^2, and most regions having RMSF values less than 0.3 nm, indicating that the protein conformation became more compact during the binding process.

Finally, we further calculated and validated the binding affinities of these complexes using the binding free energy calculation (MM/PBSA) method (Table 2). The CCNA2–1,4-DCB complex and MAOB − 1,4-DCB complex exhibited the most stable binding, with binding free energies of -49.86 kJ/mol and − 43.94 kJ/mol, respectively. These molecular simulation results provide a structural basis for direct interactions between 1,4-DCB and Ethylbenzene with the key proteins CCNA2, CPA3, and MAOB, supporting their mechanistic roles in VOC-induced LUAD development.

Table 2
Binding free energy analysis of candidate compound-target complexes.

Ligand	Item	ΔG_{vdw}(kJ/mol)	ΔG_{ele}(kJ/mol)	ΔG_{PB}(kJ/mol)	ΔG_{np}(kJ/mol)	ΔGbind(kJ/mol)
MAOB − 1,4-Dichlorobenzene	Energy	-66.48	-5.33	37.00	-9.13	-43.94
	Delta	5.89	2.64	11.24	0.59	11.47
CPA3 - Ethylbenzene	Energy	-76.86	-9.16	70.08	-9.10	-25.05
	Delta	4.35	2.57	9.84	0.54	9.42
CCNA2–1,4-Dichlorobenzene	Energy	-85.05	-3.94	48.18	-9.04	-49.86
	Delta	3.04	1.52	7.23	0.48	7.96

A: RMSD curves of three complexes; B: Rg curves reflecting conformational compactness; C: SASA curves monitoring solvent exposure; D: Per-residue RMSF profiles reflecting local flexibility.

3.8 Effects of VOC Exposure on the Expression of Hub Genes

To further validate the biological relevance of the identified hub genes under VOC exposure, we examined cell viability and the expression profiles of CCNA2, MAOB, and CPA3 in A549 (human LUAD cells) and LAD2 (human mast) cell lines following VOC treatment.

As shown in Fig. 7A, cell viability of both A549 and LAD2 cells decreased after exposure to increasing concentrations of VOCs (G0-G3), with the highest dose (G3) causing the most pronounced reduction compared with the untreated control (G0) ($P < 0.05$, $P < 0.001$). Therefore, the highest concentration at which cell viability was greater than 80% was selected. The concentrations of 1,4-DCB (10 mg/L) and ethylbenzene (100 μM), which were used in the G1 group, were chosen for subsequent expression experiments.

WB analysis revealed that the protein level of CCNA2 was markedly upregulated in the VOC-exposed group (A2) compared with the control group (A1) (Fig. 7B), and quantitative analysis confirmed a significant increase in both protein expression and mRNA expression of CCNA2 (Fig. 7C; $P < 0.001$).

Conversely, the protein levels of MAOB and CPA3 were substantially decreased in the VOC-treated group (M2) relative to the control (M1) (Fig. 7D). The relative protein levels of both MAOB and CPA3 demonstrated a significant reduction (Fig. 7E; $P < 0.001$). Consistently, RT-qPCR analysis showed that the mRNA expression levels of MAOB ($P < 0.001$) and CPA3 ($P < 0.01$) were significantly downregulated following VOC exposure (Fig. 7F).

A: Cell viability of A549 (left) and LAD2 (right) cells following exposure to graded concentrations of VOCs (G0, control; G1-G3, increasing VOC doses), assessed by CCK-8 assay; B: Representative WB bands showing CCNA2 protein levels in the control (A1) and VOC-treated (A2) groups, GAPDH was used as the loading control; C: Quantitative analysis of the relative CCNA2 protein levels (left) and mRNA expression levels (right) between A1 and A2 groups; D: Representative WB bands of MAOB and CPA3 proteins in the control (M1) and VOC-treated (M2) groups, GAPDH served as the internal reference; E: Quantitative analysis of the relative protein levels of MAOB (left) and CPA3 (right) in M1 and M2 groups; F: Relative mRNA expression levels of MAOB (left) and CPA3 (right) determined by RT-qPCR in M1 versus M2 groups. Data are presented as mean ± SD from at least three independent experiments. Statistical significance was determined by Student's t-test (for two-group comparisons) or one-way ANOVA followed by Tukey's post-hoc test (for multi-group comparisons). *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ versus the control group.

4 Discussion

This study employed a combined approach of network toxicology, large-scale transcriptomic analysis, single-cell RNA sequencing, and molecular modeling to elucidate the molecular effects of VOC exposure on LUAD at the computational level. The study identified 22 candidate target genes, whose biological functions were enriched in the “response to toxic substances,” “drug metabolism-cytochrome P450,” and “amino acid metabolism” pathways, strongly aligning with the classical carcinogenic logic of small-molecule environmental pollutants. Based on these findings, three prognosis-related genes—CCNA2, MAOB, and CPA3—were further identified. Single-cell resolution analysis revealed that these three genes belong to two distinct cell subpopulations in the LUAD microenvironment: an immunosuppressive proliferative cell cluster and mast cells.

The carcinogenic potential of inhaled VOCs has been extensively documented in the literature. Epidemiological studies confirm that VOC exposure is a significant risk factor for respiratory tract tumors in non-smokers [11]. VOCs can promote lung tumorigenesis through multiple pathways, including oxidative stress, cell growth and apoptosis, chronic inflammation, and disruption of cellular antioxidant defense systems [15]. The network toxicology strategy employed in this study advances the aforementioned population-level associations to a mechanistic investigation at the genetic level. Previous metabolomics studies have confirmed that VOC exposure can disrupt multiple metabolic pathways, including amino acid metabolism and fatty acid β-oxidation, and suggest that mitochondrial dysfunction may be one of the

underlying carcinogenic mechanisms [12]. Consistent with this, the candidate genes identified in this study are enriched in the cytochrome P450 drug metabolism and amino acid metabolism pathways. Aromatic and halogenated hydrocarbons require metabolic activation to generate reactive intermediates, and the lung epithelium is the primary site for such metabolism [18]. This finding provides toxicological rationale for the functional target genes identified in this study.

Among the prognostic genes associated with the effects of VOCs on LUAD, CCNA2 is the only gene whose high expression is associated with a poorer prognosis, which is consistent with reports in the cancer literature. CCNA2 regulates the two key cell cycle phases, G1/S and G2/M, by binding to CDK2 and CDK1, respectively. It is highly expressed in various cancers, including breast cancer, prostate cancer, and LUAD, and is negatively correlated with LUAD prognosis [19, 20]. This is consistent with the results of expression verification in this study. From the perspective of mechanism, up-regulation of CCNA2 may be involved in the regulation of cell plasticity in lung cancer cell lines. Smoking-induced CCNA2 expression promotes the differentiation of alveolar type II (AT2) cells into AT2-like cells with tumor stem cell characteristics by interfering with WNT/β-catenin signaling and triggering ubiquitin-dependent β-catenin degradation [21]. Single-cell analysis in this study revealed that CCNA2 is selectively enriched in proliferating cell subpopulations, which simultaneously overexpress immune checkpoint and exhaustion marker genes, indicating that CCNA2-positive proliferating cells serve as a critical juncture where cell cycle drive and local immune suppression coexist. The results further confirmed the correlation between CCNA2 overexpression and poor prognosis of LUAD. This finding has a reported biological basis, as cell populations with the highest proliferation indices typically activate both proliferation and immune crosstalk regulatory programs simultaneously [22]. Furthermore, cell communication analysis suggests that these CCNA2-positive proliferating cells are a major source of intratumoral CXCL13. These findings suggest that VOC exposure may promote tumor progression by inducing CCNA2 overexpression to expand the population of immunosuppressive proliferating cells.

MAOB is a mitochondrial outer membrane flavoprotein that degrades biogenic amines through oxidative deamination and produces hydrogen peroxide [23]. In this study cohort, low MAOB expression in LUAD tissues was associated with a poorer prognosis. In stromal cells, MAOB promotes tumor proliferation and invasion by driving reactive oxygen species (ROS)-mediated stromal formation and activating the CXCL12/CXCR4 paracrine signaling axis [24]. In contrast, within tumor cells, MAOB primarily functions as a tumor suppressor; its low expression in various malignancies, including glioblastoma, lung cancer, and colorectal cancer, is closely associated with poor prognosis [25]. This mechanism is consistent with the experimental results of expression level in this study. MAOB has been reported to inhibit tumor migration, proliferation, and EMT processes, while enhancing the sensitivity of LUAD to cisplatin, so its low expression is closely related to the poor prognosis of LUAD [26]. Furthermore, MAOB has been shown to contribute to radiation resistance in non-small cell lung cancer; its pharmacological inhibition (e.g., tanshinone, selegiline) can restore tumor cell sensitivity to ionizing radiation by antagonizing NF-κB [27]. This study found that 1,4-DCB forms a stable complex with the MAOB hydrophobic pocket, and the key contact residues highly align with the hydrophobic substrate-binding cavity structure mediated by Tyr398 and Tyr435 as reported in the literature [28]. It is therefore inferred that the long-term occupation of this pocket by halogenated aromatic VOCs may interfere with MAOB-dependent monoamine metabolism, thereby regulating the local redox balance and immunometabolic state of lung tissue, and ultimately promoting the progression of LUAD.

CPA3 is a mast cell protease with variable expression patterns; together with chymotrypsin and trypsin, it constitutes one of the three core protease components of the mast cell secretome [29]. In this study, CPA3 was found to be expressed exclusively in specific mast cell subsets, and its overall high expression level in tumor tissues was associated with a better patient prognosis. This is also consistent with our results of mRNA and protein expression. The role of mast cells in LUAD exhibits a significant duality. On the one hand, pro-inflammatory mast cell subsets can secrete IL-18 via the NLRP3-dependent pyroptosis pathway, thereby activating mucosa-associated invariant T (MAIT) cells and enhancing antitumor cytotoxicity [30]. On the other hand, mast cell density in LUAD is positively correlated with the enrichment of

CCR2⁺ cytotoxic T cells and improved overall survival, a finding particularly pronounced in ground-glass opacities [31]. Mechanistically, granule proteases released by mast cells (including CPA3) exert protective anti-metastatic effects via the CXCL16/CD1d/NKT cell axis in models of lung cancer metastasis [32]. According to the results of this study, CPA3 was down-regulated by VOC exposure in LUAD, which may affect the presentation of cancer cells from CD1d to NKT cells and weaken the anti-metastasis ability of mast cells [33]. It is therefore hypothesized that chronic exposure to VOCs may not completely inhibit CPA3 enzymatic activity, but rather may weaken the protective inflammatory program of mast cells in early-stage pulmonary lesions by altering mast cell granule components and their downstream immunoregulatory functions.

Analysis at single-cell resolution has further elucidated the LUAD tumor microenvironment and the expression patterns of key genes at the single-cell level. The cytotoxic CD8T_2 subset exhibits higher expression levels of core cytotoxic molecules such as NKG7, KLRD1, and GNLY, suggesting that it possesses stronger granule exocytosis and targeted killing capabilities compared to the cytotoxic CD8T_1 subset. NKG7 has been demonstrated to be a key regulatory molecule essential for CD8 T and NK cells to execute granulocyte exocytosis and antitumor cytotoxicity; its absence significantly impairs cytotoxic effects and tumor control in LUAD models [34]. GNLY (granulysin), stored in cytotoxic granules alongside perforin and cathepsin, is a crucial effector protein mediating target cell death by lymphocytes [35]. KLRD1 (CD94) forms heterodimers with NKG2 family receptors and participates in the recognition and activation regulation of HLA-E by CD8 T and NK cells [36]. The synergistic overexpression of these three components constitutes a true effector program with antitumor efficacy. Notably, while the Proliferating subpopulation highly expresses proliferation markers such as MKI67 and TOP2A, it also significantly upregulates immune checkpoint molecules such as PDCD1 and LAG3. This suggests that this subpopulation does not represent the expansion of effective effector T cells but rather exists in a compensatory state characterized by the coexistence of proliferation and exhaustion [37]. Furthermore, Tregs, exhausted CD8 T cells, and proliferating cells exhibit high immune suppression scores. Tumor-infiltrating Tregs have been demonstrated in the literature to exert potent inhibitory functions through the co-expression of multiple immune checkpoint molecules, including CTLA-4, PD-1, TIGIT, LAG3, CCR8, LAGN, and IL1R2; their accumulation in tumor tissues is closely associated with poor prognosis in LUAD [38–40]. The accumulation of exhausted CD8 T cells is a core feature of T-cell dysfunction in LUAD, accompanied by persistent high expression of PD-1, TIM-3, LAG3, and TIGIT, as well as the progressive loss of effector molecules [37].

In this study, Cell Communications analysis identified three significantly upregulated signaling axes in tumor tissue: FAM3C:PDCD1, SIRPG:CD47, and CXCL13:CXCR3. PDCD1 (PD-1), a key target in non-small cell lung cancer, is typically activated on exhausted T cells, leading to a progressive loss of cytotoxicity and cytokine secretion capacity [41]. SIRPG is primarily expressed on T cells, and its interaction with CD47 is involved in T cell endothelial migration and homeostasis maintenance. CD47-SIRP has been reported as a potential innate immune checkpoint, and targeted antibodies and fusion proteins are currently in clinical development [42]. In LUAD, enhanced CD47-SIRPα signaling on tumor-associated macrophages is associated with reduced tumor-infiltrating lymphocytes and poorer overall survival [43]. CXCL13 is a key chemokine involved in recruiting immune cells and forming tertiary lymphoid structures (TLS); its high expression is associated with improved efficacy of immune checkpoint inhibitors [44]. However, CD4⁺CXCL13⁺ exhaustion T cells have recently been identified as a characteristic immunological marker of aggressive LUAD lesions [45]. This study further reveals that plasma cells and mast cells are both major sources of intratumoral CXCL13 secretion. Concurrently, depletion of CD8⁺ T cells and Tregs leads to impaired cytotoxic T cell function, ultimately resulting in a tumor immune escape microenvironment.

Based on the above evidence, this study proposes, through computational analysis, a hypothesis that VOC exposure promotes the development of LUAD via two pathways. First, aromatic and halogenated VOCs (particularly 1,4-DCB and ethylbenzene) can specifically bind to hydrophobic pockets of proteins associated with the cell cycle and mast cell function (CCNA2, MAOB, CPA3); stable RMSD and Rg trajectories from MD simulations provide structural support for

this. Second, by disrupting MAOB and CPA3 to affect mast cell functional states, while simultaneously amplifying immunosuppressive proliferative cell populations via CCNA2, VOC exposure can drive the LUAD microenvironment toward a state of immune evasion mediated by FAM3C:PDCD1, SIRPG:CD47, and CXCL13-related axes.

However, this study has limitations. Although the differential expression of CCNA2, MAOB, and CPA3 was preliminarily validated in VOC-exposed A549 and LAD2 cells, their functional roles and impact on the tumor immune microenvironment still require further verification in established in vivo VOC exposure models and clinical cohorts. Also, molecular docking and MD simulations do not constitute direct evidence of binding at the cellular level; subsequent studies should incorporate experimental validation. In summary, by integrating network toxicology, transcriptomic prognostic analysis, single-cell heterogenity analysis, cell communication, and molecular simulations, this study identified the key genes CCNA2, MAOB, and CPA3 as potential targets through which VOC exposure influences the progression of LUAD. Furthermore, the effects of environmental small-molecule pollutants on tumors should not be limited to analysis at the whole-tissue level but should be examined at single-cell resolution, as these effects are often concentrated in specific, identifiable cell subpopulations within the tumor microenvironment.

Declarations

Conflicts of Interest

The authors declare no conflicts of interest related to this study.

Consent to Publish Declaration

Not applicable.

Ethics and Consent to Participate Declaration

Not applicable. This study was based on publicly available datasets and established human cell lines (A549 and LAD2), and did not involve any human participants, identifiable patient information, human-derived primary samples, or animal experiments. Therefore, ethical approval and informed consent were not required.

Funding

This work was supported by grants from the following sources: the 2025 Annual Joint Entrepreneurship Fund for Universities and Committees (Grant No. WXLH202501313, WXLH202501284, and WXLH202406001), and the 2025 Special Research Project on Traditional Chinese Medicine.

Author Contribution

S-LL and G-LZ conceived, designed and performed the key bioinformatics analyses, interpreted the main findings, and drafted the manuscript. H-TH assisted with single-cell RNA sequencing analysis and data interpretation. QY contributed to data validation and quality control. Y-XH performed the molecular docking experiments. SL performed the molecular dynamics simulations. Y-ML assisted with data collection and visualization. HL and XL supervised the study, provided critical feedback, revised the manuscript, and acquired the funding. All authors discussed the results and approved the final manuscript.

Acknowledgments

We would like to express our sincere gratitude to all the participants for their contributions to this study. We also acknowledge the publicly available databases including TCGA, GEO, and CTD for providing the data resources used in this research.

Data Availability

All data used in this study are publicly available. The LUAD transcriptomic sequencing data and clinical information were obtained from The Cancer Genome Atlas (TCGA) database (https://www.cancer.gov/ccg/research/genome-sequencing/tcga). The single-cell RNA sequencing data were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). VOC-related target information was retrieved from the Comparative Toxicogenomics Database (CTD, https://ctdbase.org/) and Swiss Target Prediction (http://www.swisstargetprediction.ch/). LUAD-associated gene sets were obtained from the GeneCards database (https://www.genecards.org/) and DisGeNET database (https://www.disgenet.org/). Protein structural data for molecular docking and molecular dynamics simulations were sourced from the RCSB Protein Data Bank (https://www.rcsb.org/) and AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/).

References

1. Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. 2023;20:624–39. https://doi.org/10.1038/s41571-023-00798-3.
2. Guo L, Zhu C, Cai L, Zhang X, Fang Y, Chen H, et al. Global burden of lung cancer in 2022 and projected burden in 2050. Chin Med J (Engl). 2024;137:2577–82. https://doi.org/10.1097/CM9.0000000000003268.
3. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/caac.21834.
4. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet. 2021;398:535–54. https://doi.org/10.1016/S0140-6736(21)00312-3.
5. Hendriks LEL, Remon J, Faivre-Finn C, Garassino MC, Heymach JV, Kerr KM, et al. Non-small-cell lung cancer. Nat Rev Dis Primers. 2024;10:71. https://doi.org/10.1038/s41572-024-00551-9.
6. Luo W, Zeng Z, Jin Y, Yang L, Fan T, Wang Z, et al. Distinct immune microenvironment of lung adenocarcinoma in never-smokers from smokers. Cell Rep Med. 2023;4:101078. https://doi.org/10.1016/j.xcrm.2023.101078.
7. Guo Q, Liu L, Chen Z, Fan Y, Zhou Y, Yuan Z, et al. Current treatments for non-small cell lung cancer. Front Oncol. 2022;12:945102. https://doi.org/10.3389/fonc.2022.945102.
8. Li Y, Yan B, He S. Advances and challenges in the treatment of lung cancer. Biomed Pharmacother. 2023;169:115891. https://doi.org/10.1016/j.biopha.2023.115891.
9. Siegel RL, Giaquinto AN, Jemal A, Cancer statistics. 2024. CA Cancer J Clin 2024;74:12–49. https://doi.org/10.3322/caac.21820
10. Hill W, Lim EL, Weeden CE, Lee C, Augustine M, Chen K, et al. Lung adenocarcinoma promotion by air pollutants. Nature. 2023;616:159–67. https://doi.org/10.1038/s41586-023-05874-3.
11. Maung TZ, Bishop JE, Holt E, Turner AM, Pfrang C. Indoor Air Pollution and the Health of Vulnerable Groups: A Systematic Review Focused on Particulate Matter (PM), Volatile Organic Compounds (VOCs) and Their Effects on Children and People with Pre-Existing Lung Disease. Int J Environ Res Public Health. 2022;19. https://doi.org/10.3390/ijerph19148752.

12. Yan M, Yang J, Zhu H, Zou Q, Zhao H, Sun H. Volatile organic compound exposure in relation to lung cancer: Insights into mechanisms of action through metabolomics. J Hazard Mater. 2024;480:135856. https://doi.org/10.1016/j.jhazmat.2024.135856.
13. Zhou L, Jiao X, Yang B, Yuan W, Zhao W, Zhang L, et al. The Impact of Indoor Environments on the Abundance of Urban Outdoor VOCs. Environ Sci Technol. 2025;59:9654–64. https://doi.org/10.1021/acs.est.4c13133.
14. Pál L, Lovas S, McKee M, Diószegi J, Kovács N, Szűcs S. Exposure to volatile organic compounds in offices and in residential and educational buildings in the European Union between 2010 and 2023: A systematic review and health risk assessment. Sci Total Environ. 2024;945:173965. https://doi.org/10.1016/j.scitotenv.2024.173965.
15. Hussain MS, Gupta G, Mishra R, Patel N, Gupta S, Alzarea SI, et al. Unlocking the secrets: Volatile Organic Compounds (VOCs) and their devastating effects on lung cancer. Pathol Res Pract. 2024;255:155157. https://doi.org/10.1016/j.prp.2024.155157.
16. Auerbach SS, Cora MC, Liu YF, Luh J, Prince LM, Roberts GK et al. NIEHS Report on the In Vivo Repeat Dose Biological Potency Studies of 1,4-Dichlorobenzene (CASRN 106-46-7) in Female Sprague Dawley (Hsd:Sprague Dawley(®) SD(®)) Rats and B6D2F1/Crl Mice (Whole-body Inhalation Studies): NIEHS Report 13. Research Triangle Park (NC): 2026. https://doi.org/10.22427/NIEHS-13
17. Croute F, Poinsot J, Gaubin Y, Beau B, Simon V, Murat JC, et al. Volatile organic compounds cytotoxicity and expression of HSP72, HSP90 and GRP78 stress proteins in cultured human cells. Biochimica et Biophysica Acta (BBA) -. Mol Cell Res. 2002;1591:147–55. https://doi.org/10.1016/S0167-4889(02)00271-9.
18. Castell JV, Donato MT, Gómez-Lechón MJ. Metabolism and bioactivation of toxicants in the lung. The in vitro cellular approach. Exp Toxicol Pathol. 2005;57(Suppl 1):189–204. https://doi.org/10.1016/j.etp.2005.05.008.
19. Zhang L, Wang S, Wang L. Comprehensive analysis pinpoints CCNA2 as a prognostic and immunological biomarker in non-small cell lung cancer. BMC Pulm Med. 2025;25:14. https://doi.org/10.1186/s12890-025-03490-7.
20. Gao T, Han Y, Yu L, Ao S, Li Z, Ji J. CCNA2 is a prognostic biomarker for ER+ breast cancer and tamoxifen resistance. PLoS ONE. 2014;9:e91771. https://doi.org/10.1371/journal.pone.0091771.
21. He Q, Qu M, Xu C, Wu L, Xu Y, Su J, et al. Smoking-induced CCNA2 expression promotes lung adenocarcinoma tumorigenesis by boosting AT2/AT2-like cell differentiation. Cancer Lett. 2024;592:216922. https://doi.org/10.1016/j.canlet.2024.216922.
22. Atitey K, Anchang B. Mathematical Modeling of Proliferative Immune Response Initiated by Interactions Between Classical Antigen-Presenting Cells Under Joint Antagonistic IL-2 and IL-4 Signaling. Front Mol Biosci. 2022;9:777390. https://doi.org/10.3389/fmolb.2022.777390.
23. Jaisa-Aad M, Muñoz-Castro C, Healey MA, Hyman BT, Serrano-Pozo A. Characterization of monoamine oxidase-B (MAO-B) as a biomarker of reactive astrogliosis in Alzheimer's disease and related dementias. Acta Neuropathol. 2024;147:66. https://doi.org/10.1007/s00401-024-02712-2.
24. Pu T, Wang J, Wei J, Zeng A, Zhang J, Chen J, et al. Stromal-derived MAOB promotes prostate cancer growth and progression. Sci Adv. 2024;10:eadi4935. https://doi.org/10.1126/sciadv.adi4935.
25. Alsaad I, Abdel Rahman DMA, Al-Tamimi O, Alhaj SA, Sabbah DA, Hajjo R et al. Targeting MAO-B with Small-Molecule Inhibitors: A Decade of Advances in Anticancer Research (2012–2024). Molecules 2024;30. https://doi.org/10.3390/molecules30010126
26. Liang Z, Zhang H, Jiang J, Li C, Yuan F. USP44 Stabilizes MAOB via Deubiquitination to Inhibit Cisplatin Resistance in Lung Adenocarcinoma. Int J Genomics 2026;2026:7433804. https://doi.org/10.1155/ijog/7433804
27. Son B, Jun SY, Seo H, Youn H, Yang HJ, Kim W, et al. Inhibitory effect of traditional oriental medicine-derived monoamine oxidase B inhibitor on radioresistance of non-small cell lung cancer. Sci Rep. 2016;6:21986. https://doi.org/10.1038/srep21986.

28. Huang H-C, Hsieh Y-H, Hsiao C-H, Lin C-Y, Wang S-S, Ho K-H, et al. MAOB expression correlates with a favourable prognosis in prostate cancer, and its genetic variants are associated with the metastasis of the disease. J Cell Mol Med. 2024;28:e18229. https://doi.org/10.1111/jcmm.18229.
29. Atiakshin D, Kostin A, Trotsenko I, Samoilova V, Buchwalow I, Tiemann M. Carboxypeptidase A3-A Key Component of the Protease Phenotype of Mast Cells. Cells 2022;11. https://doi.org/10.3390/cells11030570
30. Fan F, Wang J, Liu K, Zhang S, Gao J, Li X, et al. Mast cells boost anti-tumor potency of MAIT cells via inflammasome-dependent secretion of IL-18. Nat Commun. 2025;16:6074. https://doi.org/10.1038/s41467-025-61324-w.
31. Fan F, Gao J, Zhao Y, Wang J, Meng L, Ma J, et al. Elevated Mast Cell Abundance Is Associated with Enrichment of CCR2 + Cytotoxic T Cells and Favorable Prognosis in Lung Adenocarcinoma. Cancer Res. 2023;83:2690–703. https://doi.org/10.1158/0008-5472.CAN-22-3140.
32. Grujic M, Paivandy A, Gustafson A-M, Thomsen AR, Öhrvik H, Pejler G. The combined action of mast cell chymase, tryptase and carboxypeptidase A3 protects against melanoma colonization of the lung. Oncotarget. 2017;8:25066–79. https://doi.org/10.18632/oncotarget.15339.
33. Grujic M, Paivandy A, Gustafson A-M, Thomsen AR, Öhrvik H, Pejler G. The combined action of mast cell chymase, tryptase and carboxypeptidase A3 protects against melanoma colonization of the lung. Oncotarget. 2017;8:25066–79. https://doi.org/10.18632/oncotarget.15339.
34. Ng SS, De Labastida Rivera F, Yan J, Corvino D, Das I, Zhang P, et al. The NK cell granule protein NKG7 regulates cytotoxic granule exocytosis and inflammation. Nat Immunol. 2020;21:1205–18. https://doi.org/10.1038/s41590-020-0758-6.
35. Dotiwala F, Lieberman J. Granulysin: killer lymphocyte safeguard against microbes. Curr Opin Immunol. 2019;60:19–29. https://doi.org/10.1016/j.coi.2019.04.013.
36. Voogd L, van den Broek RL, van Wolfswinkel M, Franken KLMC, Ruibal P, Jespers W, et al. HLA-E/peptide complexes differentially interact with NKG2A/CD94 and T cell receptors. J Immunol. 2025;214:595–605. https://doi.org/10.1093/jimmun/vkae068.
37. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;24:978–85. https://doi.org/10.1038/s41591-018-0045-3.
38. De Simone M, Arrigoni A, Rossetti G, Gruarin P, Ranzani V, Politano C, et al. Transcriptional Landscape of Human Tissue Lymphocytes Unveils Uniqueness of Tumor-Infiltrating T Regulatory Cells. Immunity. 2016;45:1135–47. https://doi.org/10.1016/j.immuni.2016.10.021.
39. Kadota K, Nitadori J-I, Adusumilli PS. Prognostic value of the immune microenvironment in lung adenocarcinoma. Oncoimmunology. 2013;2:e24036. https://doi.org/10.4161/onci.24036.
40. Zhang P, Zhang X, Cui Y, Gong Z, Wang W, Lin S. Revealing the role of regulatory T cells in the tumor microenvironment of lung adenocarcinoma: a novel prognostic and immunotherapeutic signature. Front Immunol. 2023;14:1244144. https://doi.org/10.3389/fimmu.2023.1244144.
41. Li Z, Wang T, Liu J, Qi W, Lv Q, Xu Y, et al. The mechanism and research progress of PD-1/PD-L1 on immune escape of lung cancer. Transl Cancer Res. 2025;14:6041–51. https://doi.org/10.21037/tcr-2025-230.
42. Zhang W, Huang Q, Xiao W, Zhao Y, Pi J, Xu H, et al. Advances in Anti-Tumor Treatments Targeting the CD47/SIRPα Axis. Front Immunol. 2020;11:18. https://doi.org/10.3389/fimmu.2020.00018.
43. Giatromanolaki A, Mitrakas A, Anestopoulos I, Kontosis A, Koukourakis IM, Pappa A, et al. Expression of CD47 and SIRPα Macrophage Immune-Checkpoint Pathway in Non-Small-Cell Lung Cancer. Cancers (Basel). 2022;14. https://doi.org/10.3390/cancers14071801.
44. Park S, Cha H, Kim HS, Lee B, Kim S, Kim TM, et al. Transcriptional upregulation of CXCL13 is correlated with a favorable response to immune checkpoint inhibitors in lung adenocarcinoma. Cancer Med. 2023;12:7639–50.

https://doi.org/10.1002/cam4.5460.

45. Chen X, Zhang J, Lu T, Du M, Li F, Dong D, et al. CD4(+)CXCL13(+) exhausted T cells drive immune microenvironment divergence in synchronous double primary lung adenocarcinoma with different degrees of invasiveness. Transl Lung Cancer Res. 2026;15:17. https://doi.org/10.21037/tlcr-2025-1-1489.

Figures

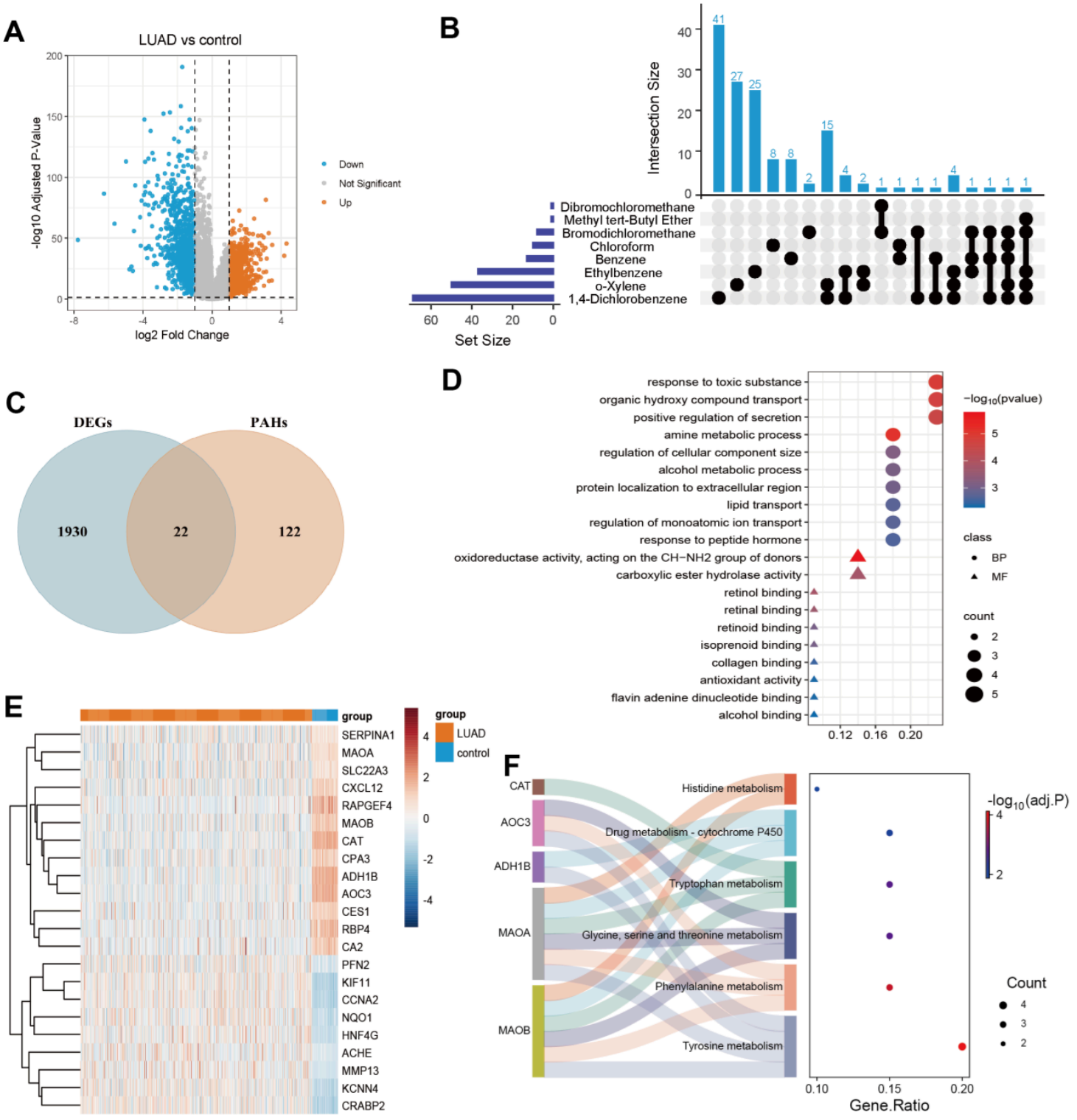


Figure 1

Analysis of differentially expressed genes in LUAD and their association with VOC target genes.

A: Volcano plot of DEGs between LUAD and normal tissues; yellow, blue, and gray dots indicate upregulated, downregulated, and non-significant genes; B: Set analysis of VOC target genes across eight compounds; C: Venn diagram of LUAD DEGs and VOC target genes; D: GO enrichment analysis of candidate genes (BP, biological process; MF, molecular function); E: Expression heatmap of 22 common target genes across samples (blue to red: low to high); F: KEGG pathway enrichment analysis of candidate genes.

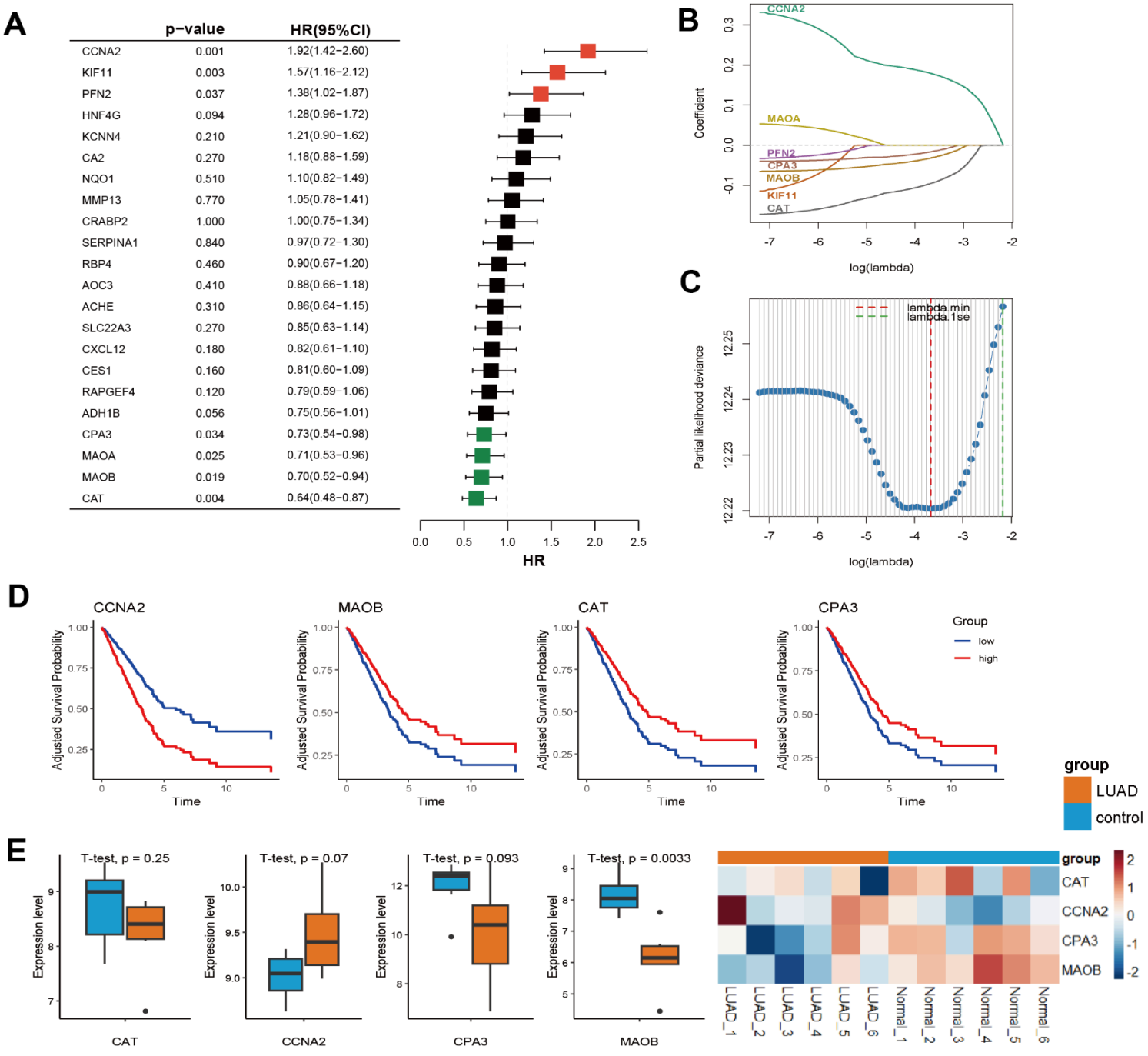


Figure 2

Expression characteristics and prognostic value analysis of key genes.

A: Forest plot of univariate Cox regression for 22 candidate genes; red and green squares indicate genes associated with poor (HR > 1) or favorable (HR < 1) prognosis; B: LASSO regression coefficient trajectories across lambda values; C: 10-fold cross-validation curve for optimal lambda selection (red dashed line); D: Kaplan-Meier survival curves of four prognostic genes stratified by expression level (red, high; blue, low); E: Box plots and heatmap of key gene expression in LUAD versus normal tissues in the validation cohort; two-sample t-test.

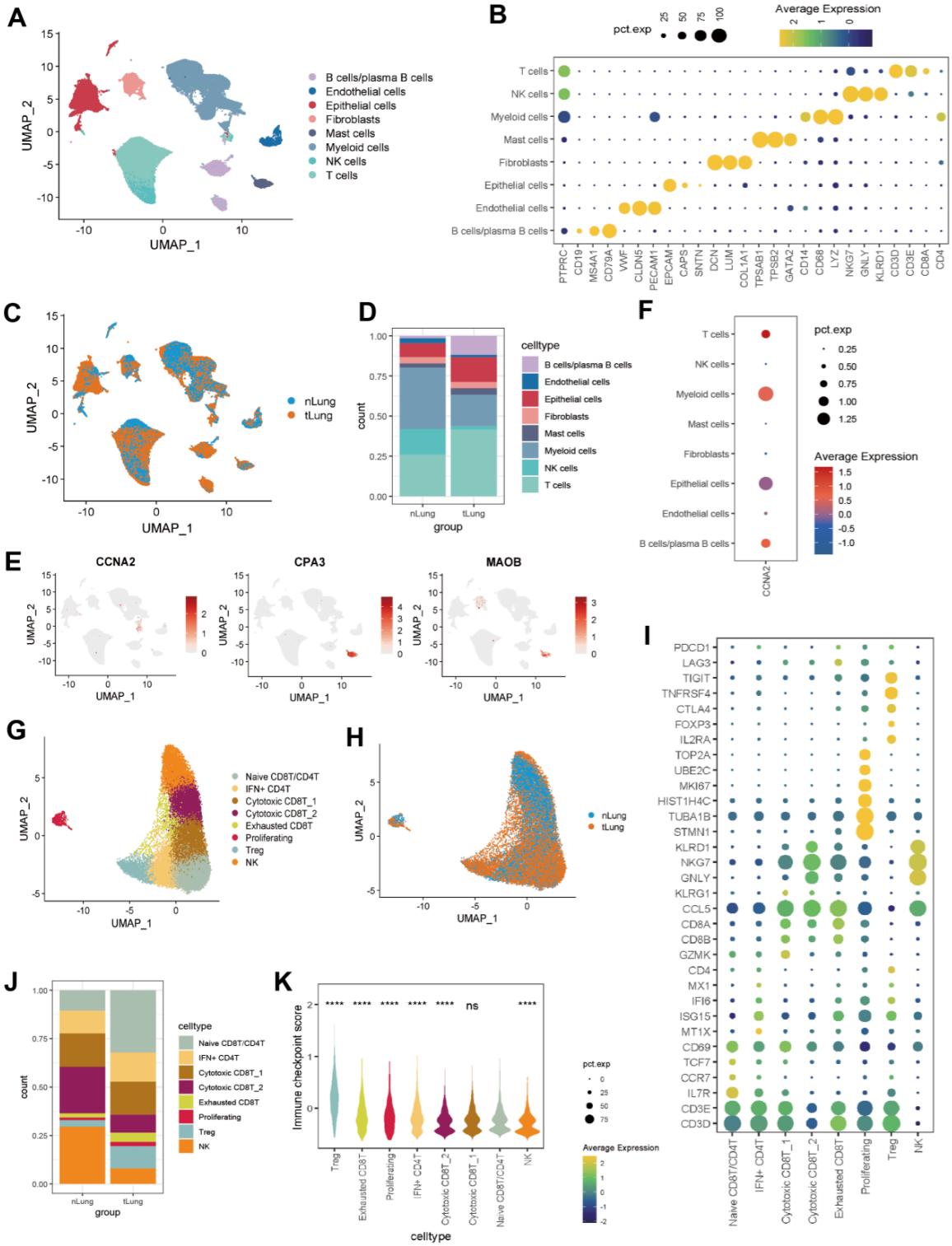


Figure 3

Analysis of LUAD tumor heterogeneity and expression characteristics of key genes at the single-cell level.

A: UMAP of all cells colored by annotated cell type; B: Dot plot of canonical marker genes per cell type; C: UMAP colored by tissue origin (tumor, yellow; normal, blue); D: Bar chart of cell-type proportions in tumor versus normal tissues; E: UMAP feature plots of key gene expression across cell types; F: Dot plot of CCNA2 expression across cell types; G: UMAP of T/NK cell subpopulations; H: T/NK cell subgroup composition by tissue origin; I: Marker gene expression across T/NK subpopulations; J: Proportional distribution of T/NK subgroups in tumor versus normal tissues; K: Violin plots of immune checkpoint scores across T/NK subgroups; Wilcoxon rank-sum test; ns, not significant; ****, $P < 0.0001$.

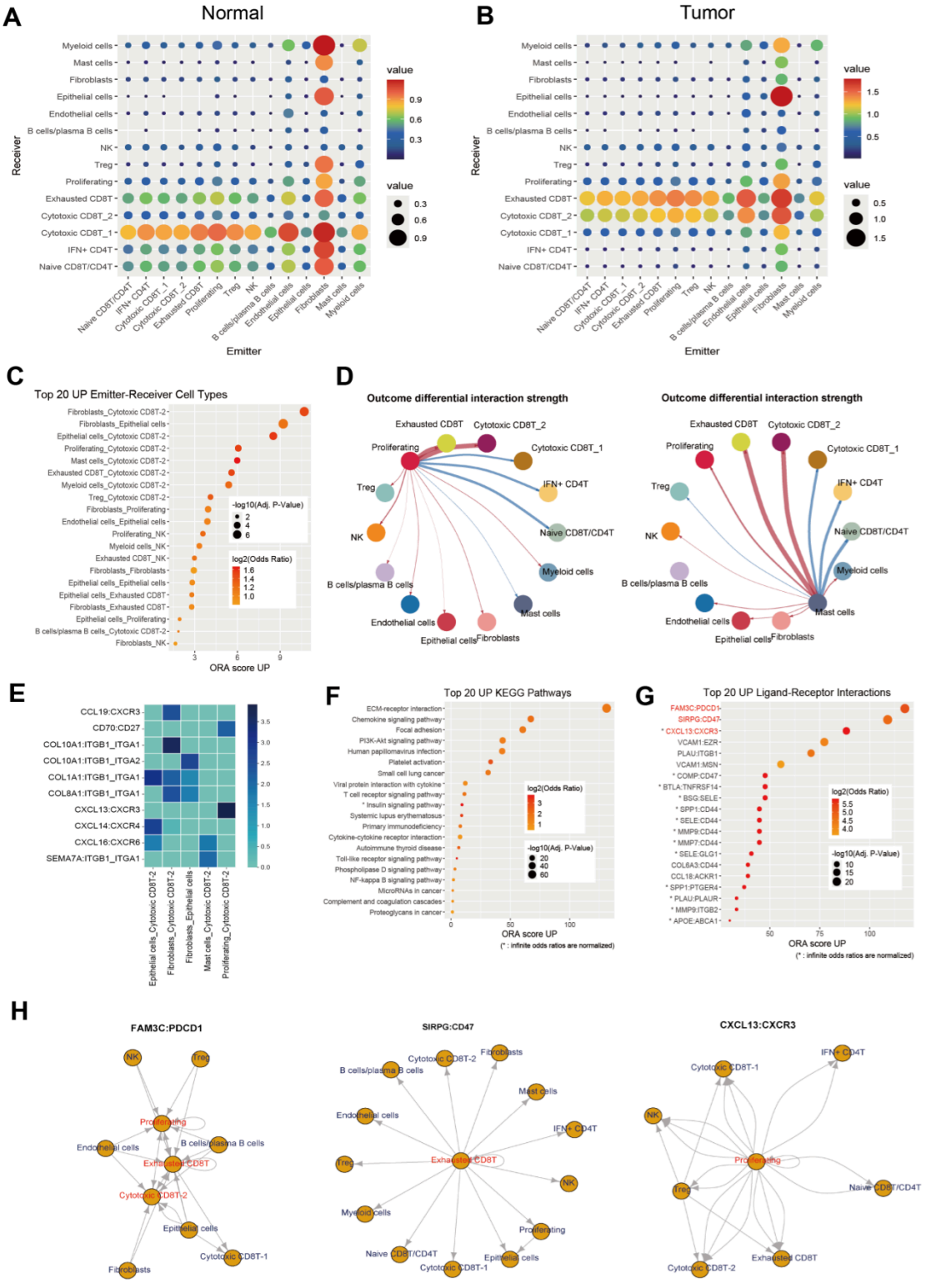
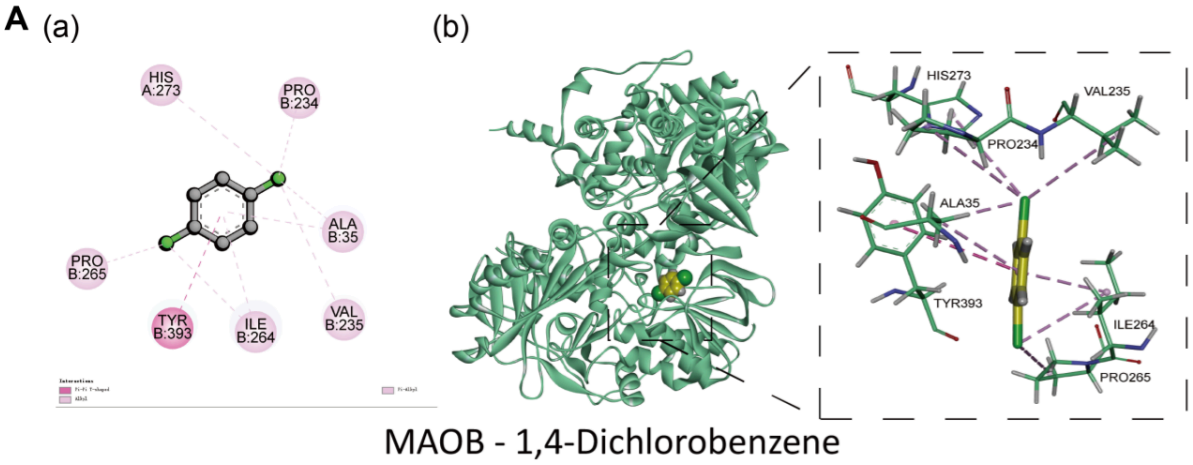
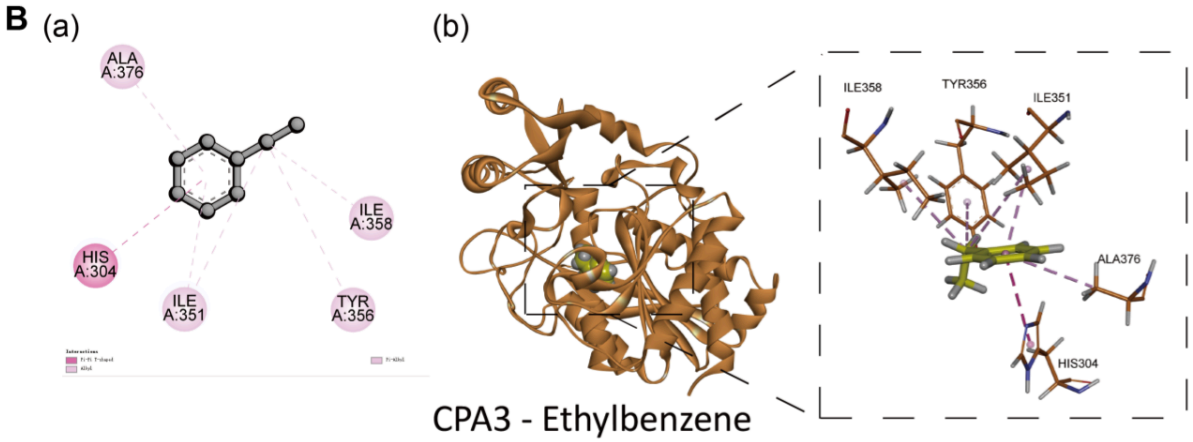


Figure 4

Cell communication networks in the LUAD microenvironment.

A/B: Heatmaps of intercellular communication strength in normal and tumor tissues; C: Top 20 upregulated sender–receiver cell-type pairs in LUAD by ORA score; D: Differential communication network between LUAD and normal tissues (red, increased; blue, decreased); E: Heatmap of significantly upregulated ligand-receptor pairs in LUAD; F: Top 20 upregulated ligand-receptor interactions in LUAD; G: Top 20 upregulated ligand-receptor pairs ranked by interaction strength; H: Chord diagram of cell communication patterns for three key ligand-receptor pairs.





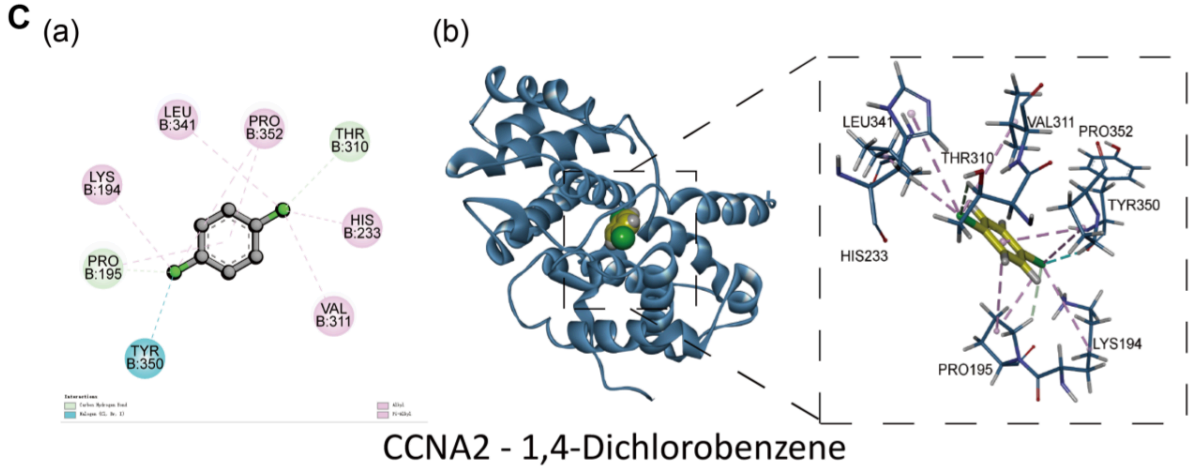


Figure 5

Molecular docking analysis of active components with key targets.

A: MAOB–1,4-DCB complex; panel (a), 2D interaction diagram; panel (b), 3D binding conformation with key residues. B: CPA3–ethylbenzene complex. C: CCNA2–1,4-DCB complex.

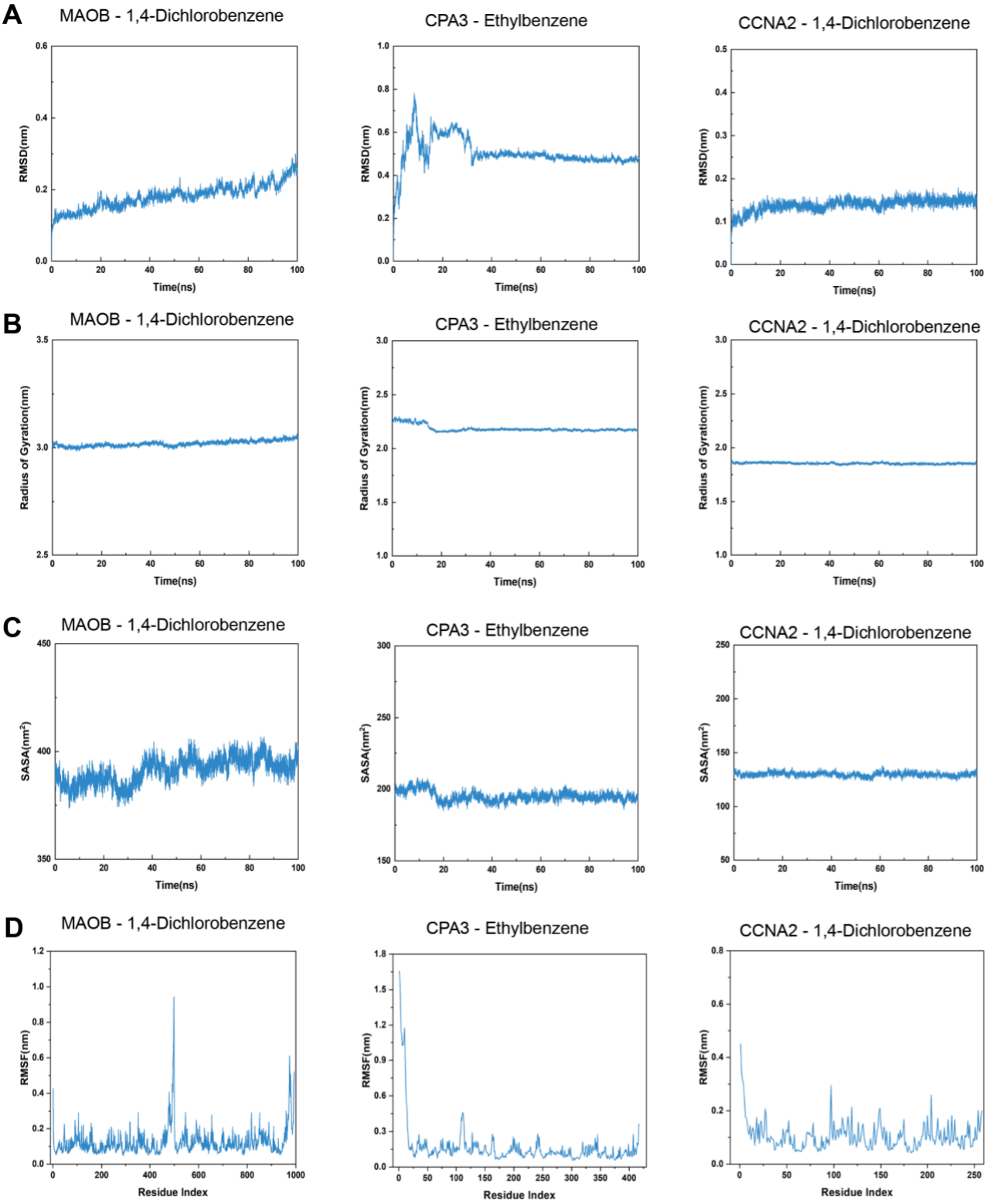


Figure 6

MD simulation analysis.

A: RMSD curves of three complexes; B: Rg curves reflecting conformational compactness; C: SASA curves monitoring solvent exposure; D: Per-residue RMSF profiles reflecting local flexibility.

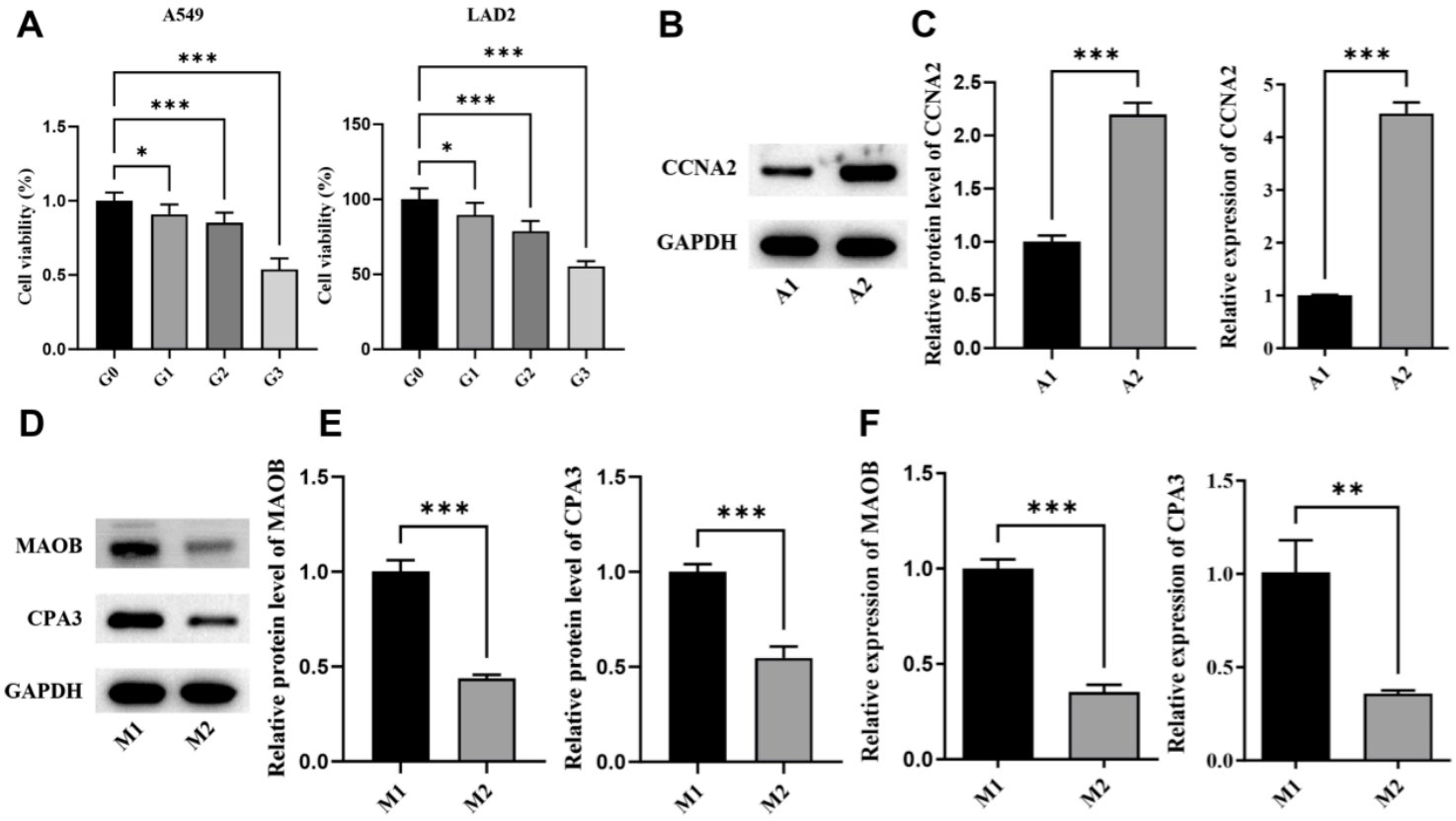


Figure 7

Effects of VOC exposure on cell viability and the expression of CCNA2, MAOB, and CPA3.

A: Cell viability of A549 (left) and LAD2 (right) cells following exposure to graded concentrations of VOCs (G0, control; G1-G3, increasing VOC doses), assessed by CCK-8 assay; B: Representative WB bands showing CCNA2 protein levels in the control (A1) and VOC-treated (A2) groups, GAPDH was used as the loading control; C: Quantitative analysis of the relative CCNA2 protein levels (left) and mRNA expression levels (right) between A1 and A2 groups; D: Representative WB bands of MAOB and CPA3 proteins in the control (M1) and VOC-treated (M2) groups, GAPDH served as the internal reference; E: Quantitative analysis of the relative protein levels of MAOB (left) and CPA3 (right) in M1 and M2 groups; F: Relative mRNA expression levels of MAOB (left) and CPA3 (right) determined by RT-qPCR in M1 versus M2 groups. Data are presented as mean ± SD from at least three independent experiments. Statistical significance was determined by Student's t-test (for two-group comparisons) or one-way ANOVA followed by Tukey's post-hoc test (for multi-group comparisons). *P< 0.05, **P< 0.01, ***P< 0.001 versus the control group.

Supplementary Files

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

- AnnotatedWesternMAOBCPA3andGAPDHBlotImages.pdf
- UncroppedWesternBlotCCNA2.tif

- UncroppedWesternBlotCPA3.tif
- UncroppedWesternBlotMAOBCPA3GAPDH.tif
- UncroppedWesternBlotCCNA2GAPDH.tif
- AnnotatedWesternCCNA2GAPDHBlotImages.pdf
- UncroppedWesternBlotMAOB.tif