

CAV1 and PECAM1 Mediate the Potential Association between Di(2-ethylhexyl) Phthalate Exposure and Lung Adenocarcinoma Immune Microenvironment Remodeling

Quan Chen¹

The Affiliated Hospital of Southwest Medical University

Yuying Li

lzhlyyhy@126.com

The Affiliated Hospital of Southwest Medical University

Research Article

Keywords: DEHP, LUAD, network toxicology, machine learning, bioinformatics

Posted Date: October 28th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11034260/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

**CAV1 and PECAM1 Mediate the Potential Association between
Di(2-ethylhexyl) Phthalate Exposure and Lung Adenocarcinoma
Immune Microenvironment Remodeling**

Quan Chen¹, Yuying Li^{1*}

¹ Department of Respiratory and Critical Care Medicine, The Affiliated Hospital of Southwest Medical
University, Luzhou 646000, Sichuan Province, P.R. China

Email: aholi2936@126.com

*Corresponding author: Yuying Li. Department of Respiratory and Critical Care Medicine, The
Affiliated Hospital of Southwest Medical University, Luzhou 646000, Sichuan Province, P.R. China.

Email: lzhlyyhy@126.com

Abstract

Di-(2-ethylhexyl) phthalate (DEHP), a widely used plasticizer designated as a priority water pollutant in China, is an established endocrine disruptor with carcinogenic, teratogenic, and mutagenic potential. Although environmental exposure to DEHP through food and water is pervasive, its specific role in lung adenocarcinoma (LUAD) is poorly defined. Here, we systematically identified 485 putative DEHP-associated targets in LUAD by integrating multiple public databases. These targets are significantly enriched in cancer-related pathways such as those regulating epithelial cell proliferation, cell adhesion, extracellular matrix remodeling, and the PI3K-Akt pathway. Three complementary machine-learning algorithms converged on CAV1 and PECAM1 as core targets; their differential expression and diagnostic performance were validated in TCGA and GEO transcriptomic cohorts. CIBERSORT analysis showed that these core targets associate closely with tumor immune infiltration, and single-cell data delineated their cell-type-specific expression. Molecular docking and molecular dynamics simulations demonstrated stable binding of DEHP to CAV1, with a binding free energy of -29.62 ± 3.89 kcal/mol. Drug-response survival analysis indicated that PECAM1 could serve as a predictive marker for pemetrexed chemotherapy. Together, these findings suggest a potential DEHP–CAV1/PECAM1–immune microenvironment regulatory axis in LUAD and provide computational evidence for future experimental validation.

Keywords: DEHP; LUAD; network toxicology; machine learning; bioinformatics

Introduction

LUAD is the predominant subtype of non-small cell lung cancer, comprising about 40% to 50% of all lung cancer cases^[1]. Its incidence and mortality rates are among the highest for malignant tumors^[2].

While targeted therapies and immune checkpoint inhibitors have improved outcomes for some patients, the high tumor heterogeneity and frequent late-stage diagnoses remain major challenges. In addition to genetic factors, environmental exposure significantly influences the onset and progression of lung cancer^[3]. Identifying key environmental risk factors and understanding their carcinogenic mechanisms are crucial for lung cancer prevention and treatment. Di (2-ethylhexyl) phthalate (DEHP), a plasticizer commonly found in plastic products, food packaging, and medical devices, is known to easily leach from these materials, leading to widespread population exposure^[4]. Toxicological studies have shown that DEHP can disrupt endocrine functions, induce oxidative stress, and cause DNA damage and epigenetic changes^[5]. Epidemiological evidence also links DEHP to an increased risk of various tumors^[6]. Thus, investigating the relationship between DEHP and LUAD is significant for both environmental health and tumor etiology. Existing studies have begun to reveal DEHP's cancer-promoting potential. It alters proliferation, apoptosis, and invasion by activating PPAR, modulating the estrogen pathway, and inducing inflammation^[7], and it contributes to tumor microenvironment remodeling and immune regulation^[8]. In the lungs, DEHP exposure has been associated with airway inflammation^[9], impaired lung function^[10], but systematic investigations focused on LUAD remain scarce. The current shortcomings are as follows. First, most studies treat non-small cell lung cancer as a single entity and do not investigate the deep molecular mechanisms specific to LUAD. Second, key-target screening mainly uses conventional differential-expression analysis and does not exploit machine-learning pattern-recognition capabilities, which limits accuracy and robustness. Third, bulk transcriptomes cannot resolve gene-expression features of specific cellular subpopulations. Fourth, there remains a lack of molecular-docking and kinetic-simulation evidence demonstrating whether DEHP can stably bind the proposed core target protein. Collectively, these gaps

impede mechanistic insight and the development of early-warning biomarkers. In addition, whether DEHP-associated core genes can serve as predictive biomarkers for LUAD chemotherapy response also remains unclear. Network toxicology combines target prediction with pathway enrichment to reveal system-level molecular links between toxins and diseases. Machine learning efficiently selects characteristic genes and constructs diagnostic models. Single-cell sequencing supplies cell type-specific expression data and precisely characterizes microenvironmental remodeling. Molecular dynamics simulations assess the binding stability of compounds to target proteins^[11]. Integrating these strategies can overcome the limits of any single method and systematically elucidate the potential molecular network by which DEHP drives LUAD. This study proceeded as follows. We identified intersection targets of DEHP and LUAD using multiple databases. GO and KEGG enrichment analyses revealed core pathways. We applied several machine learning algorithms to select key hub genes and validated their expression differences and diagnostic performance across multiple cohorts. Immune infiltration and single-cell transcriptome analyses identified the cellular sources of these genes and characterized remodeling of the immunosuppressive microenvironment. Molecular docking and molecular dynamics simulations evaluated DEHP binding modes and the stability of its interactions with target proteins. Furthermore, we performed chemotherapeutic drug-response survival analysis to evaluate the potential of core genes in predicting the prognosis of LUAD patients receiving conventional chemotherapy. This study aims to research the carcinogenic mechanism of new environmental pollutants. It also offers references for health risk assessment and the development of individualized clinical chemotherapy regimens.

Methods

All public databases and datasets used in this study, together with their accession numbers and web links, are described in the "Availability of data and materials" section.

Toxicity prediction of DEHP

The standard chemical structure and SMILES representation of DEHP were sourced from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/> ; accessed on 26 April 2026)(Fig. 1A). These SMILES expressions were then entered into ProTox 3.0 (<https://tox.charite.de/protox3/>;accessed on 26 April 2026) to predict DEHP's toxicological characteristics, such as carcinogenicity, cytotoxicity, and organ toxicity. The analysis specifically focused on pulmonary toxicity and carcinogenic potential, offering a theoretical basis for future research.

Identification and integration of potential DEHP targets

The SMILES representations of DEHP were entered into the ChEMBL database(<https://www.ebi.ac.uk/chembl/>;accessed on 27 April 2026) ,specifying the species as human, which resulted in the identification of 90 targets. The English name of the compound was then submitted to the CTD online database(<https://ctdbase.org/>;accessed on 27 April 2026), where gene interactions were selected, yielding a total of 4,600 interacting genes associated with DEHP. Subsequently, the structural formula of DEHP was submitted to the DIGEP-Pred 2.0 database(<https://way2drug.com/digep-pred/index.php/>;accessed on 27 April 2026), with the Dataset set to CTD-mRNA and a cutoff of $P_a > 0.5$, under the condition of Up regulation, resulting in 5,919 entries. Additionally, by directly inputting the SMILES representations into the SwissTargetPrediction database (<https://swisstargetprediction.ch/>;accessed on 27 April 2026),and selecting human as the species, 100 entries were obtained. The target information from these four databases was integrated and

de-duplicated using R language, thereby constructing a comprehensive target gene set for DEHP.

Screening of LUAD-related targets

Using GeneCards(<https://www.genecards.org/>;accessed on 28 April 2026), we searched for the keyword "Lung adenocarcinoma" with a relevance score threshold of 20. In the TTD database(<https://ttd.idrblab.cn/>;accessed on 28 April 2026), we filtered by disease type, specifically selecting 2C25: Lung cancer, and retained only the genes with specific gene names. Additionally, we conducted a search for "Lung adenocarcinoma" in the OMIM database(<https://www.omim.org/>;accessed on 28 April 2026). The genes associated with LUAD identified across these three databases were then de-duplicated and integrated using R language to construct a comprehensive set of related target genes for LUAD.

Functional enrichment analysis

GO enrichment analysis, categorized into biological process (BP), cellular component (CC), and molecular function (MF), along with KEGG pathway enrichment analysis, was performed using the clusterProfiler package. The screening criteria were established at $p < 0.05$ and $q < 0.05$. A bubble chart was generated using the enrichplot package to visualize the results. This analysis aims to elucidate the potential biological processes and molecular pathways associated with LUAD induced by DEHP.

Machine learning-based core target screening

Using 485 potential genes as feature genes and designating "lung adenocarcinoma" as the disease label, we analyzed data from TCGA LUAD. After sorting the data by averaging multiple samples from the same patient, we obtained gene TPM values for 516 tumor samples and 59 normal samples, thereby constructing the gene feature matrix. We employed the glmnet package to construct a LASSO logistic regression model, utilizing 5-fold cross-validation to identify the optimal λ value (lambda.min) and to

filter genes with non-zero coefficients as potential core targets. The Support Vector Machine - Recursive Feature Elimination (SVM-RFE) model was developed using the e1071 and caret packages, gradually eliminating features through recursive feature elimination to identify the gene set that most significantly contributed to the model. Subsequently, we utilized the randomForest package to create a random forest model, compute the importance scores of genes, and select the top 15% of genes based on these scores. The VennDiagram package in R was employed to generate Venn diagrams illustrating the screening results from the three algorithms. The genes identified in the intersection of these three analyses were designated as the core target genes associated with LUAD resulting from DEHP exposure.

Expression validation and diagnostic value analysis of core target genes

Three datasets, specifically GSE19804^[12], GSE32863^[13], and GSE40419^[14], which include LUAD tissues and normal lung tissues, were selected from the GEO database. The expression matrices and their corresponding platform annotation files were downloaded. Using Excel, the probe IDs were converted into gene symbols. In instances where multiple probes corresponded to the same gene, the average value was calculated. In three independent datasets, with $p < 0.05$ and $\log_2\text{FC} \geq 1$ as the screening criteria, we analyzed the expression differences of the core target gene between LUAD tissues and normal tissues using the limma package. The differential expression results were visualized with a violin plot generated by the ggplot2 package. Additionally, we utilized these three datasets, treating the expression level of the core target gene as the variable and "lung adenocarcinoma/normal" as the grouping label, to construct the receiver operating characteristic (ROC) curve. The ROC curve was drawn, and the area under the curve (AUC) value was marked to assess the diagnostic efficacy of individual core target genes and combinations of core genes for LUAD. On the Kaplan-Meier Plotter

website (<https://kmplot.com/analysis/>, accessed on 29 April 2026) , we selected lung cancer, set the median cutoff, and plotted KM survival curves for the core target genes.

Immune microenvironment analysis

Using the CIBERSORT algorithm included in the IOBR package^[15], the LM22 immune cell characteristic gene set was selected. The mean transcriptome data from TCGA LUAD was imported, with permutations set to 1000 and p-values adjusted accordingly. For samples with a significance level of 0.05, the relative infiltration ratios of 22 immune cell types were calculated for each sample. The pheatmap package was employed to generate a heatmap. Subsequently, the correlation between the core target genes (CAV1 and PECAM1) and the infiltration ratios of major immune cells was analyzed, and the results were visualized using a scatter plot created with ggplot2.

single cell RNA analysis

In this study, we analyzed 208,506 single-cell data points from 58 LUAD samples obtained from the GEO database GSE131907^[16]. We selected 10 representative samples, comprising 35,933 cells, which included normal lung tissue, primary tumors, brain metastases, pleural effusions, and lymph node metastases. Objects were constructed using the Seurat package. Following quality control, 35,677 cells were retained for further analysis. Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction clustering were conducted based on hypervariable genes, while the original annotated cell types were preserved. Key analysis of the expression characteristics of CAV1 and PECAM1: In epithelial cells, high and low subgroups were established based on the 75th quantile of CAV1 expression. Differential genes were identified using the Wilcoxon rank sum test (Bonferroni correction, adjusted $p < 0.05$), followed by GO/KEGG enrichment analysis (BH correction) on the upregulated genes. PECAM1 is expressed in endothelial and immune

cells, with a value of 0 defined as positive. The same differential expression and functional enrichment analyses were conducted for PECAM1. For cell-cell communication analysis, ligand-receptor (LR) interaction scores between endothelial cells and immune cells were calculated. Endothelial cells were stratified into CAV1-high/low and PECAM1-high/low subgroups according to the above expression thresholds. Total LR scores were summarized to evaluate the communication strength between endothelial subpopulations and immune cell subsets, and visualized by chord diagrams. To assess endothelial-immune cell communication mediated by CAV1 and PECAM1, ligand-receptor (LR) interaction analysis was performed using the CellChat package. Endothelial cells were grouped into CAV1-high/low and PECAM1-high/low subsets, and LR interaction scores between endothelial and immune cell populations were calculated. Chord diagrams were plotted using the circize package to visualize and compare the communication strength between different groups. All statistical analyses and visualizations were performed in the R environment.

Molecular Docking

The PECAM1 and CAV1 proteins utilized for docking were sourced from the PDB database, identified by PDBID values of 5C14 and 7SC0, respectively. The three-dimensional structures of the small molecules were retrieved from the PUBCHEM database, followed by energy minimization conducted under the MMFF94 force field. In this study, molecular docking was performed using AutoDock Vina version 1.2.3.^[17] Prior to initiating the docking process, the receptor protein was processed with PyMol version 2.5.5, which involved the removal of water molecules, salt ions, and small molecules. Subsequently, a docking box was established to encompass the entire protein structure. Additionally, all processed small molecules and receptor proteins were converted to the PDBQT format required for AutoDock Vina 1.2.3 docking using ADFRsuite 1.0.^[19] During the docking procedure, the

global search detail was set to 32, while all other parameters were maintained at their default settings.

The docking conformation yielding the highest score was designated as the binding conformation.

Finally, visual analysis of the docking results was conducted using PyMol 2.5.5 and Ligplot v2.2.9.

Molecular Dynamics Simulation

The full-atom molecular dynamics simulation was conducted using the small molecule and protein complex derived from docking as the initial structure, employing AMBER 24 software^[20]. Prior to the simulation, the AM1-BCC charge of the small molecule was computed using the antechamber module^[21]. Subsequently, the small molecules and proteins were characterized with the GAFF2 small molecule force field and the ff14SB protein force field^{[22][23]}, respectively. In each system, the LEaP module facilitated the addition of hydrogen atoms. Truncated octahedral TIP3P solvent boxes^[24] were incorporated at a distance of 10 Å from the system, and Na⁺/Cl⁻ ions were introduced to neutralize the system's charge. Finally, the topology and parameter files necessary for the simulation were generated. Energy optimization of the system was first performed using the steepest descent method for 2,500 steps, followed by the conjugate gradient method for an additional 2,500 steps. Upon completion of the energy optimization, the system was heated for 200 ps at constant volume and a controlled heating rate, gradually raising the temperature from 0 K to 298.15 K. Once the system stabilized at 298.15 K, a 500 ps NVT (isothermal isovolume) ensemble simulation was conducted to ensure an even distribution of solvent molecules within the solvent box. Under NPT (constant temperature and constant pressure) conditions, a 500 ps equilibrium simulation was performed for the entire system. Subsequently, the composite system underwent a 100 ns NPT (isothermal-isobaric) simulation under periodic boundary conditions. During this simulation, the truncation distance for non-bonded interactions was set to 10 Å. The Particle Mesh Ewald (PME) method was employed to calculate

long-range electrostatic interactions^[25], while the SHAKE method was utilized to constrain the bond lengths of hydrogen atoms^[26]. Temperature control was achieved using the Langevin algorithm^[27], with a collision frequency (γ) set to 2 ps^{-1} . The system pressure was maintained at 1 atm, the integration step size was 2 fs, and the trajectory was recorded every 10 ps for subsequent analysis.

MMGBSA binding free energy calculation

The binding free energy between proteins and ligands across all systems was calculated using the MM/GBSA method^{[28][29][30]}. In this study, molecular dynamics (MD) trajectories ranging from 90 to 100 ns were employed for the calculations, with the specific formula provided as follows: $\Delta G_{bind} = \Delta G_{complex} - (\Delta G_{receptor} + \Delta G_{ligand}) = \Delta E_{internal} + \Delta E_{VDW} + \Delta E_{elec} + \Delta G_{GB} + \Delta G_{SA}$. In Equation, $\Delta E_{internal}$ represents the internal energy, ΔE_{VDW} represents van der Waals interactions, ΔE_{elec} represents electrostatic interactions. The internal energy includes bond energy (ΔG_{bind}), angle energy (E_{angle}), and torsion energy ($E_{torsion}$). ΔG_{GB} and ΔG_{SA} are collectively referred to as solvation free energy, where ΔG_{GB} denotes the polar solvation free energy and ΔG_{SA} denotes the nonpolar solvation free energy. For ΔG_{GB} , the GB model developed by Nguyen et al.^[31] was adopted in this study (igb = 2). The nonpolar solvation free energy (ΔG_{SA}) was calculated as the product of the surface tension (γ), and the solvent-accessible surface area (SASA), using the formula:

$\Delta G_{SA} = 0.0072 \times \Delta SASA$ ^[32]. Entropy contributions were neglected in this study due to their high computational cost and relatively low prediction accuracy.

Drug-response survival analysis

To explore the clinical implications of core genes in guiding chemotherapy selection, we performed drug-response survival analysis using the GEPIA3 database

(http://gepia3.bioinfo.liu.com/drug_response/; accessed on 10 May 2026). We evaluated hazard ratios

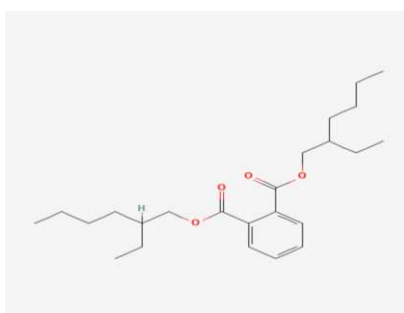
(HR) of overall survival for multiple first-line lung adenocarcinoma chemotherapeutic agents (including pemetrexed, carboplatin, erlotinib, paclitaxel, docetaxel, cisplatin, gemcitabine, vinorelbine, taxol, and alimta) in patients stratified by high- and low-expression levels of CAV1 and PECAM1. Adjusted P-values < 0.05 were considered statistically significant. Forest plots were used to visualize HR values and corresponding 95% confidence intervals (CIs).

Results

Toxicity prediction results of DEHP using ProTox

The chemical structure of di-(2-ethylhexyl) phthalate (DEHP) is shown in Fig. 1A. ProTox predictions indicate that DEHP markedly interferes with metabolic enzymes, notably CYP2E1 and CYP3A4. The compound also binds estrogen and androgen receptors, producing strong endocrine-disrupting effects, and shows a tendency toward hepatotoxicity, nephrotoxicity, and respiratory toxicity (Fig. 1B). DEHP may promote lung tissue injury by disrupting exogenous metabolism and altering hormonal signaling.

A



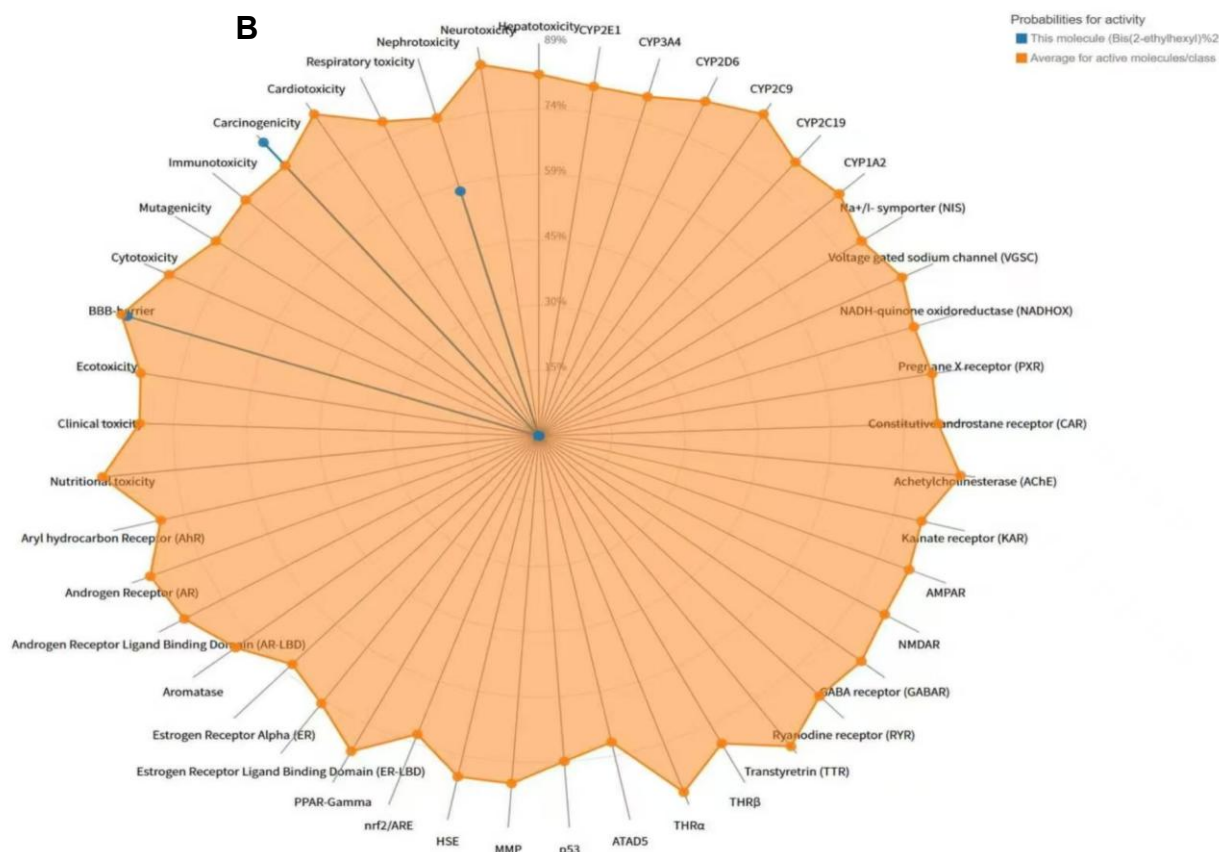


Fig. 1 Chemical structure and toxicity endpoint prediction of DEHP. (A) Chemical structure of DEHP. (B) Radar chart of DEHP-related toxicity target activity probabilities predicted by the ProTox-3.0 web server. The blue line represents the prediction probability of DEHP binding to different toxicity targets, while the orange shaded area indicates the average binding probability of active molecules in each toxicity category. CYP2E1, Cytochrome P450 2E1; CYP3A4, Cytochrome P450 3A4; ER, Estrogen Receptor; PPAR- γ , Peroxisome Proliferator-Activated Receptor Gamma.

Potential targets of DEHP in LUAD

In this study, we identified 7,618 potential DEHP targets from ChEMBL, CTD, SwissTargetPrediction, and DIGEP (Fig. 2A). From GeneCards, OMIM, and TTD, we retrieved 975 genes associated with LUAD (Fig. 2B). The intersection of these sets yielded 485 potential target genes linking DEHP to LUAD (Fig. 2C).

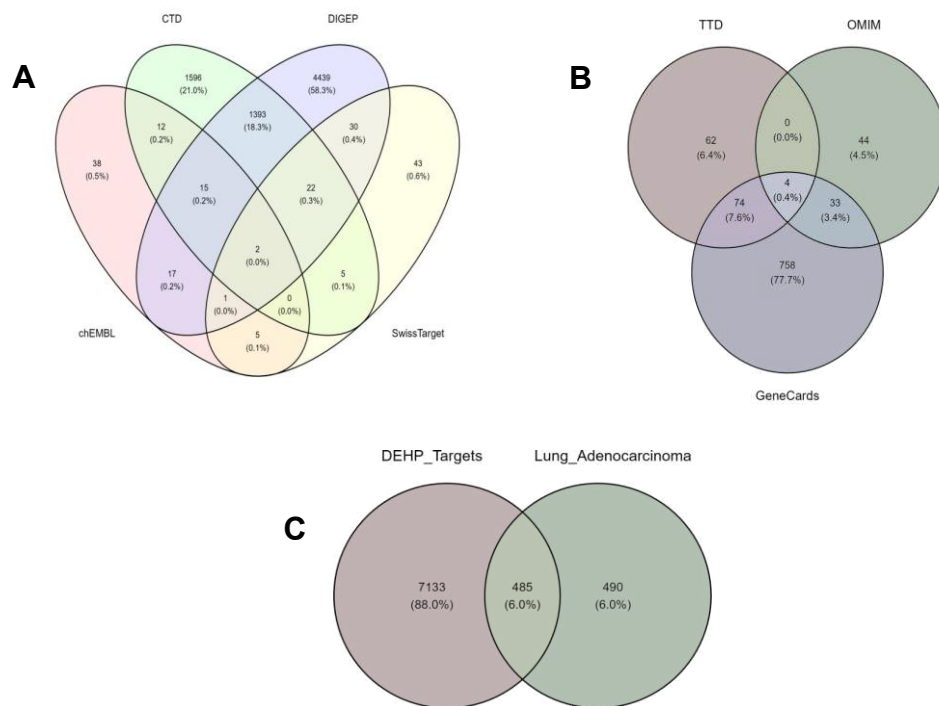


Fig. 2 Venn diagrams for screening DEHP-related potential target genes. (A) Venn diagram of the intersection of DEHP target genes predicted from four databases (CTD, DIGEP-Pred 2.0, ChEMBL, and SwissTargetPrediction). (B) Venn diagram of DEHP-related gene intersections across the TTD, OMIM, and GeneCards databases. (C) Venn diagram of the intersection between DEHP-related target genes and LUAD-related genes. Numbers and percentages in each region represent the count and proportion of target genes in each subset.

Functional and pathway enrichment analysis of potential target genes

Gene Ontology (GO) analysis of the intersection genes was performed in R. At the biological process (BP) level, the genes were primarily associated with dysregulated cell proliferation and death, aberrant signal transduction, angiogenesis and metastasis, hypoxia and metabolic adaptation, and inflammation and immune regulation, providing a multidimensional view of DEHP's cancer-promoting features(Fig. 3A). GO cellular component (GO-CC) terms clustered around membrane rafts, the extracellular matrix, and adhesion sites(Fig. 3B). GO molecular function (GO-MF) terms were

dominated by kinase regulation and transcription factor binding(Fig. 3C). KEGG pathway analysis revealed significant enrichment of cancer-related miRNAs, the PI3K-Akt signaling pathway, proteoglycans in cancer, and cellular senescence pathways, indicating that DEHP contributes to LUAD development by activating multiple oncogenic pathways and by modulating processes related to proliferation, apoptosis, and metastasis(Fig. 3D).

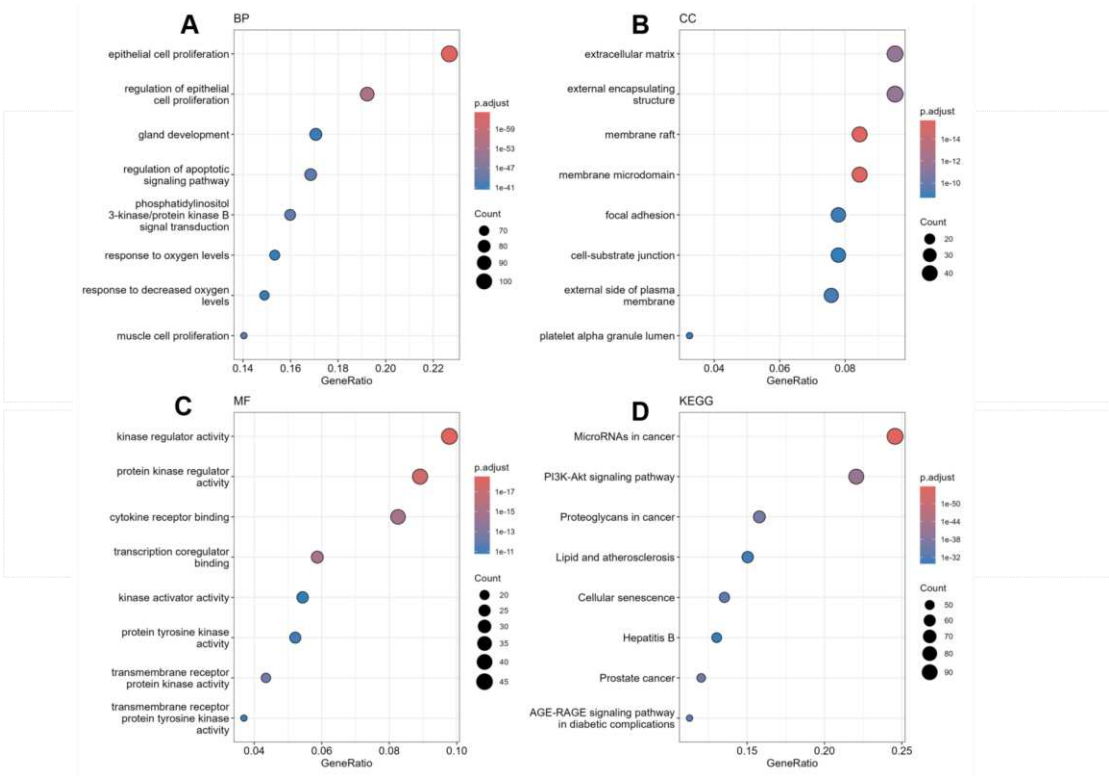


Fig. 3 GO and KEGG enrichment analyses of DEHP-induced lung adenocarcinoma-related target genes. Bubble plots display the top enriched Gene Ontology (GO) terms for Biological Process (BP, A), Cellular Component (CC, B), and Molecular Function (MF, C), together with Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (D). The x-axis indicates the gene ratio of enriched target genes, and the y-axis lists enriched functional terms and pathways. Bubble color corresponds to the adjusted P value (p.adjust), with red indicating greater enrichment significance and blue indicating lower significance; bubble size represents the number of enriched target genes. All enrichment analyses were conducted with the clusterProfiler package in R using screening thresholds of $P < 0.05$ and $q <$

0.05.

Use machine learning to identify core target genes

Using the TCGA LUAD cohort (516 tumors and 59 normal samples), we built expression matrices using intersecting genes as features and identified core targets with three machine-learning methods: LASSO logistic regression, SVM-RFE, and random forest (Fig. 4A). LASSO selected 7 genes, SVM-RFE identified 8 optimal features, and random forest retained the top 15% most important genes(Fig. 4B-F). The intersection of all three methods comprised CAV1 and PECAM1. Both genes were consistently detected across algorithms and are thus designated as core target genes of DEHP-induced LUAD.

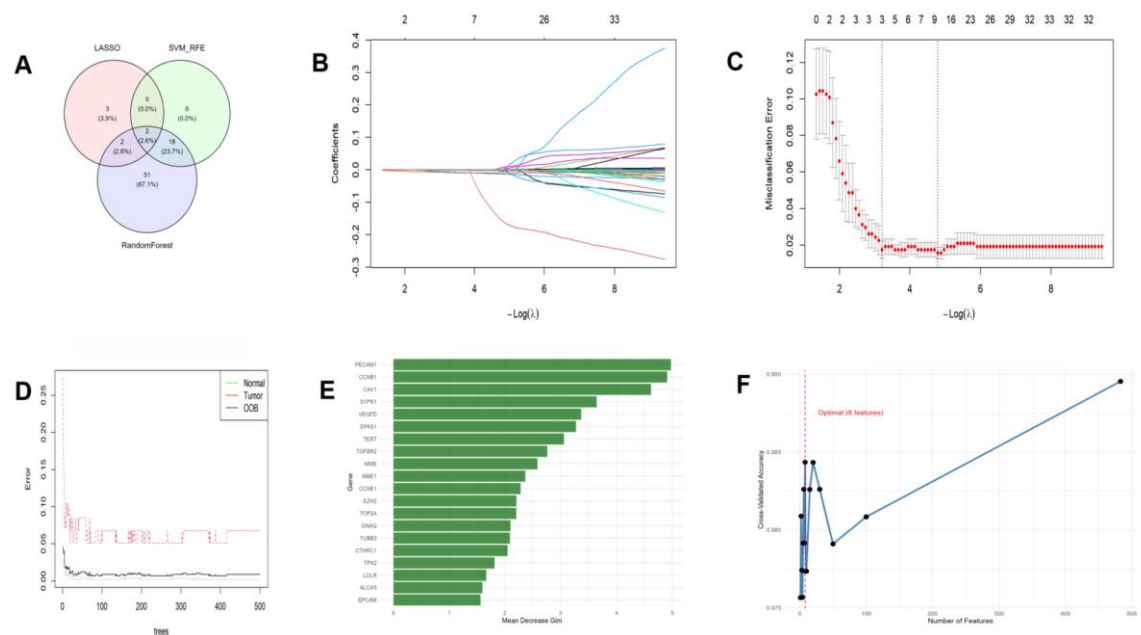


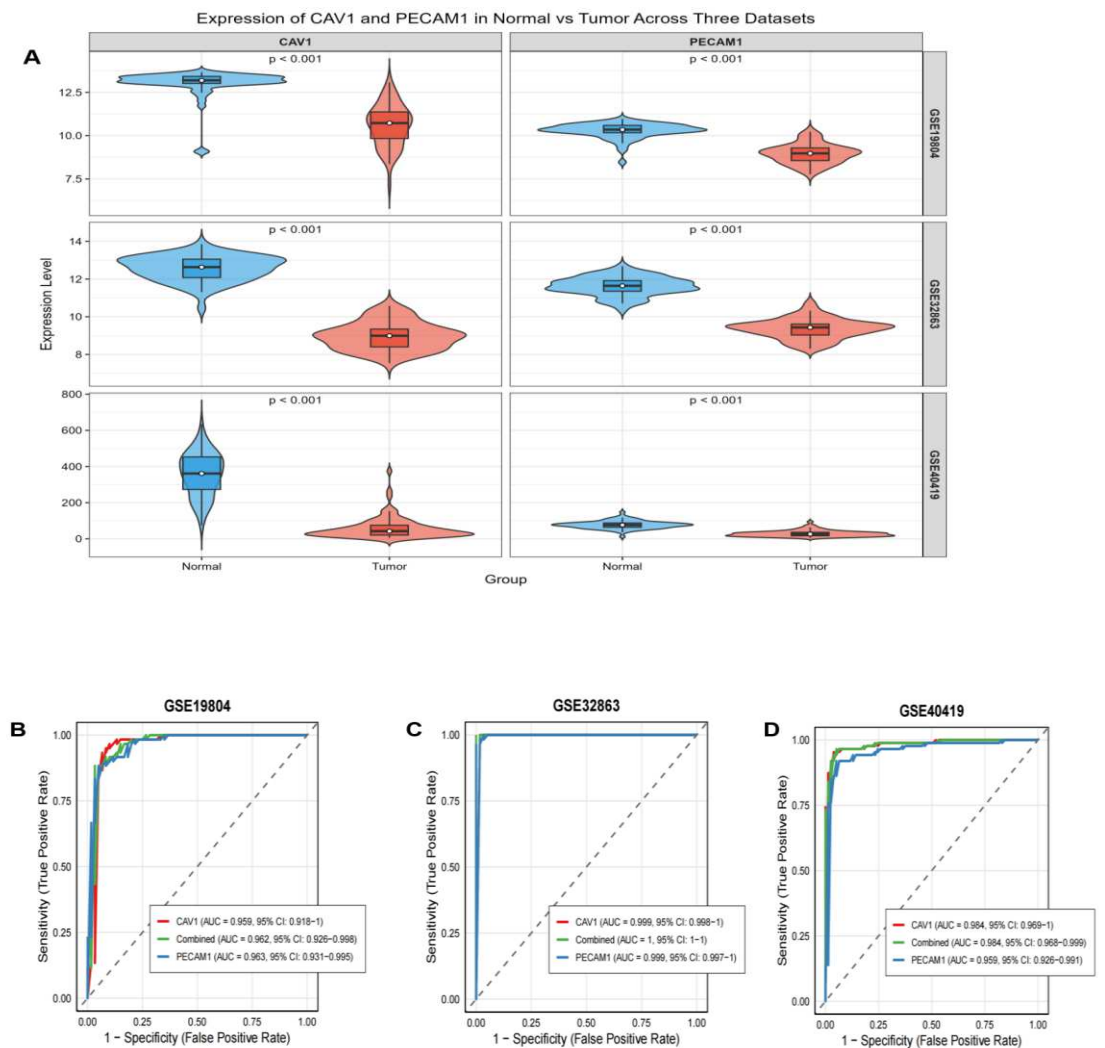
Fig. 4 Machine learning-based screening of core genes for DEHP-induced lung adenocarcinoma.(A) Venn diagram showing the overlap of feature genes selected by three machine learning algorithms (LASSO logistic regression, SVM-RFE, and random forest). The intersection of genes identified by all three methods was taken as the core target genes.(B) LASSO coefficient profile plot for the 485 candidate genes. Each colored line represents the coefficient trajectory of a gene as the

tuning parameter λ varies. The x-axis represents $-\log(\lambda)$, and the y-axis represents the gene coefficients. (C) Five-fold cross-validation curve for selecting the optimal λ in the LASSO model. The red dots indicate binomial deviance values at different λ , with error bars showing standard error. The dashed lines mark the optimal λ ($\lambda_{\min}$) and 1-SE λ (λ_{1se}). (D) Error rate of the random forest model as a function of the number of trees. The green line represents the error rate for normal samples, the red line for tumor samples, and the black line indicates the out-of-bag (OOB) error rate. The model converged with a stable OOB error after approximately 100 trees. (E) Top 20 genes ranked by importance scores (mean decrease in Gini index) from the random forest model. The top 15% of genes with the highest importance were selected as key features. (F) Recursive feature elimination curve for the SVM-RFE model, showing cross-validated accuracy as a function of the number of features. The red dashed line marks the optimal number of features ($n=8$) that maximizes model performance.

Expression Verification, Diagnostic and Prognostic Value Analysis of Core Target Genes

We obtained RNA expression datasets (GSE19804, GSE32863, GSE40419) associated with LUAD from the TCGA and GEO databases to assess core gene expression levels. In LUAD tissues, CAV1 and PECAM1 expression levels are significantly reduced compared to normal tissues (Fig. 5A). Using these datasets, we generated ROC curves and calculated AUC to assess the diagnostic efficacy of these two core genes. The results show that both single-gene and two-gene combinations yield relatively high diagnostic performance. The two-gene combination, in particular, discriminates normal and diseased tissues more accurately (Fig. 5B-D). These findings suggest that CAV1 and PECAM1 are strong candidates for novel biological markers. We further evaluated the prognostic significance of CAV1 and PECAM1 expression in LUAD using Kaplan-Meier survival analysis. Patients with high

CAV1 expression exhibited significantly longer overall survival compared to those with low expression (HR = 0.81, 95% CI: 0.68-0.96, log-rank P = 0.016; Fig. 5E). Similarly, high PECAM1 expression was associated with improved overall survival (HR = 0.69, 95% CI: 0.54-0.88, log-rank P = 0.0027; Fig. 5F). Collectively, these results indicate that reduced expression of CAV1 and PECAM1 correlates with poor prognosis in LUAD patients, further supporting their potential as prognostic biomarkers.



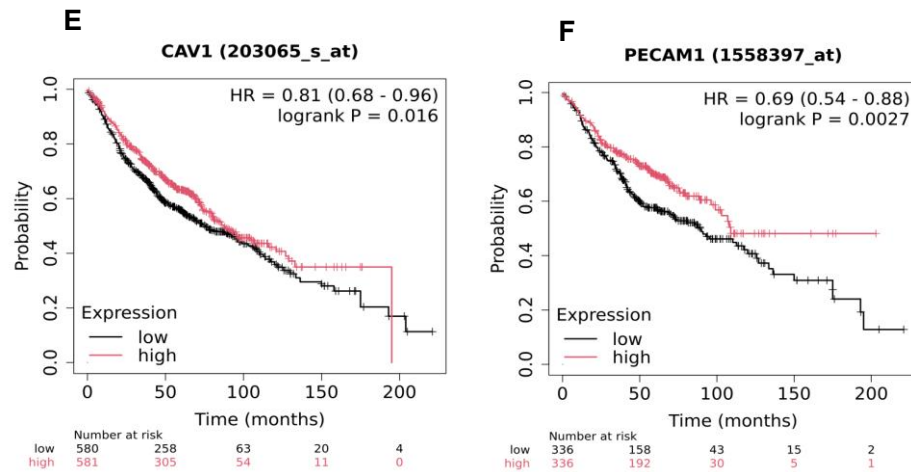


Fig 5. Expression verification, diagnostic efficacy, and prognostic analysis of core genes CAV1

and PECAM1 in LUAD.(A) Violin plots showing the expression levels of CAV1 and PECAM1 in normal and LUAD tissues across three independent datasets (GSE19804, GSE32863, GSE40419).

(B-D) Receiver operating characteristic (ROC) curves for single-gene (CAV1, PECAM1) and two-gene combination models in the three datasets. (E, F) Kaplan-Meier survival curves comparing overall survival between patients with high and low expression of CAV1 and PECAM1. HR, hazard ratio; 95% CI, 95% confidence interval.

Immune Infiltration Analysis

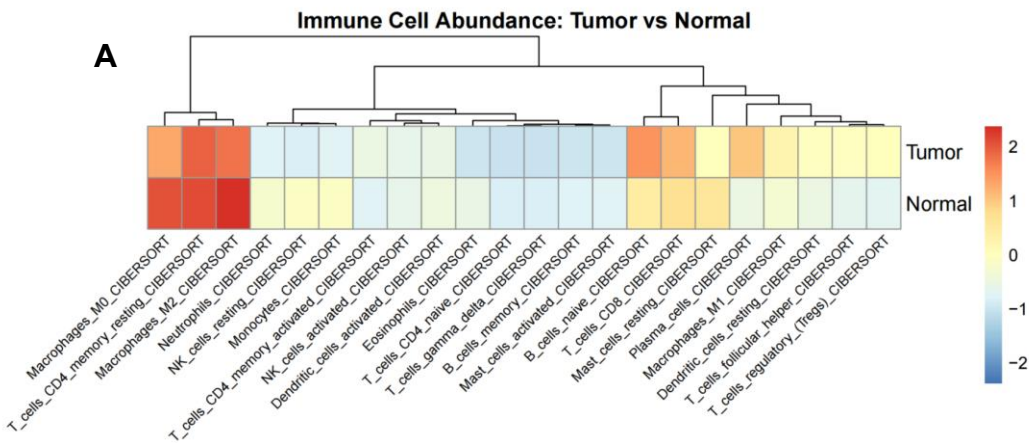
We applied the CIBERSORT algorithm to quantify the proportions of 22 tumor-infiltrating immune cell subsets in 516 LUAD and 59 normal lung tissue samples from the TCGA database. A heatmap was generated to visualize the differences in immune cell abundance between tumor and normal tissues (Fig. 6A), revealing distinct immune landscapes in LUAD. Compared with normal lung tissue, LUAD samples were significantly enriched in regulatory T cells (Tregs), plasma cells, and follicular helper T cells, whereas M2 macrophages, resting mast cells, resting NK cells, and monocytes were more abundant in normal tissues. We further analyzed the correlations among the 22 immune cell subsets in the tumor microenvironment (Fig. 6B), showing a complex network of positive and negative interactions between different immune cell types.

To explore the regulatory roles of CAV1 and PECAM1 in shaping the immune microenvironment, we performed correlation analyses between their expression levels and the infiltration of key immune cell subsets (Fig. 6C-P).

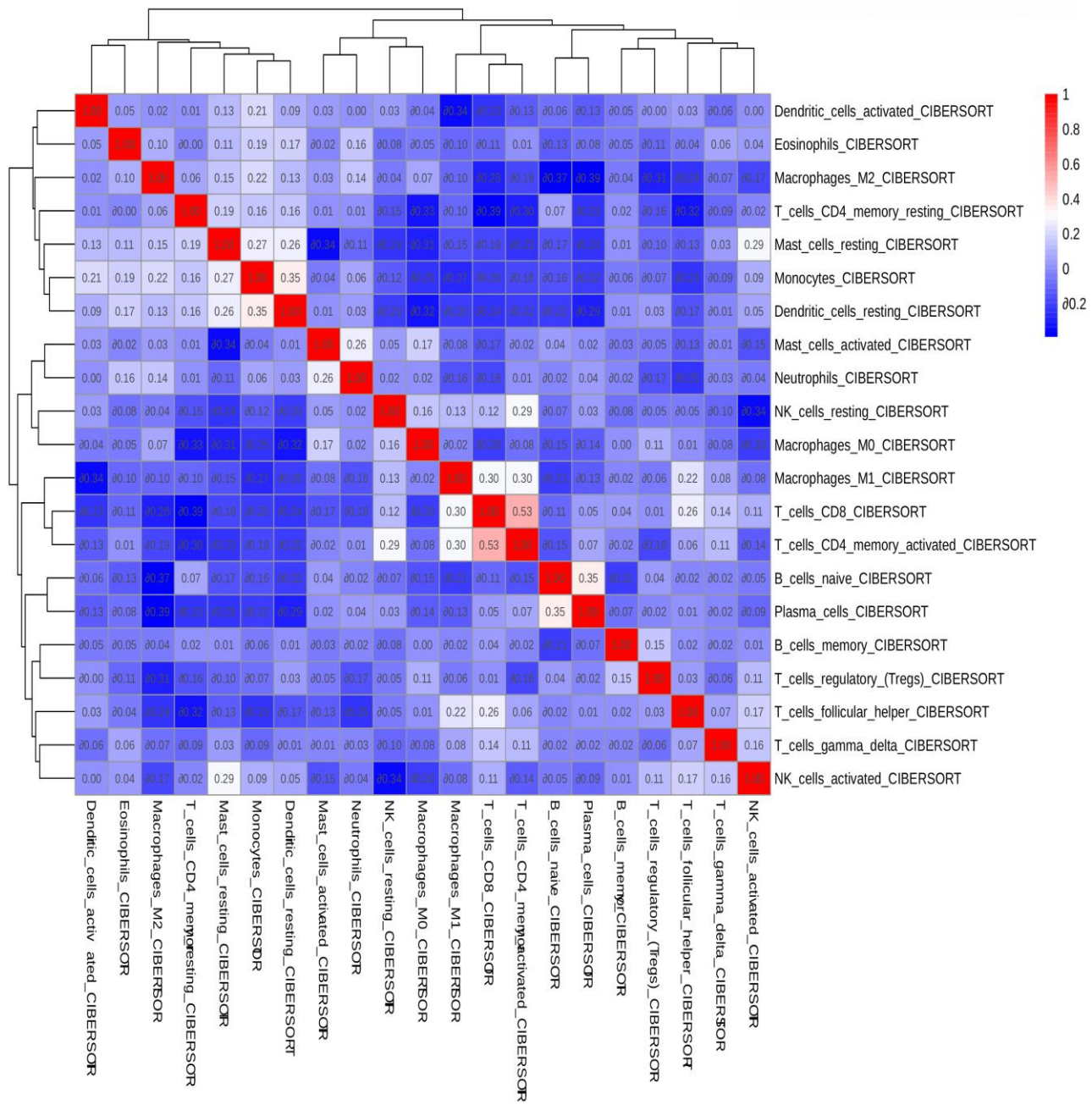
For CAV1: expression was significantly positively correlated with M2 macrophages, resting mast cells, and resting CD4+ memory T cells, and negatively correlated with plasma cells, follicular helper T cells, and regulatory T cells (Tregs).

For PECAM1: expression was significantly positively correlated with M2 macrophages, resting mast cells, neutrophils, and resting CD4+ memory T cells, and negatively correlated with naive B cells, plasma cells, follicular helper T cells, and regulatory T cells (Tregs).

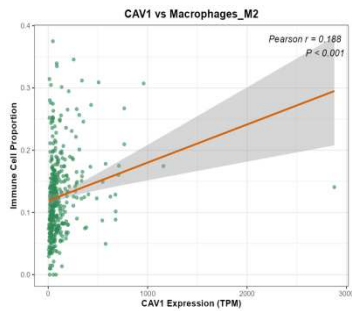
These results suggest that both CAV1 and PECAM1 play critical roles in modulating the recruitment and polarization of immune cells in the LUAD microenvironment, thereby contributing to the formation of an immunosuppressive tumor microenvironment.



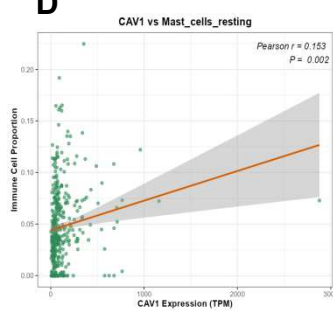
B



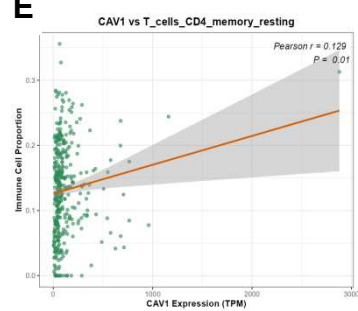
C



D



E



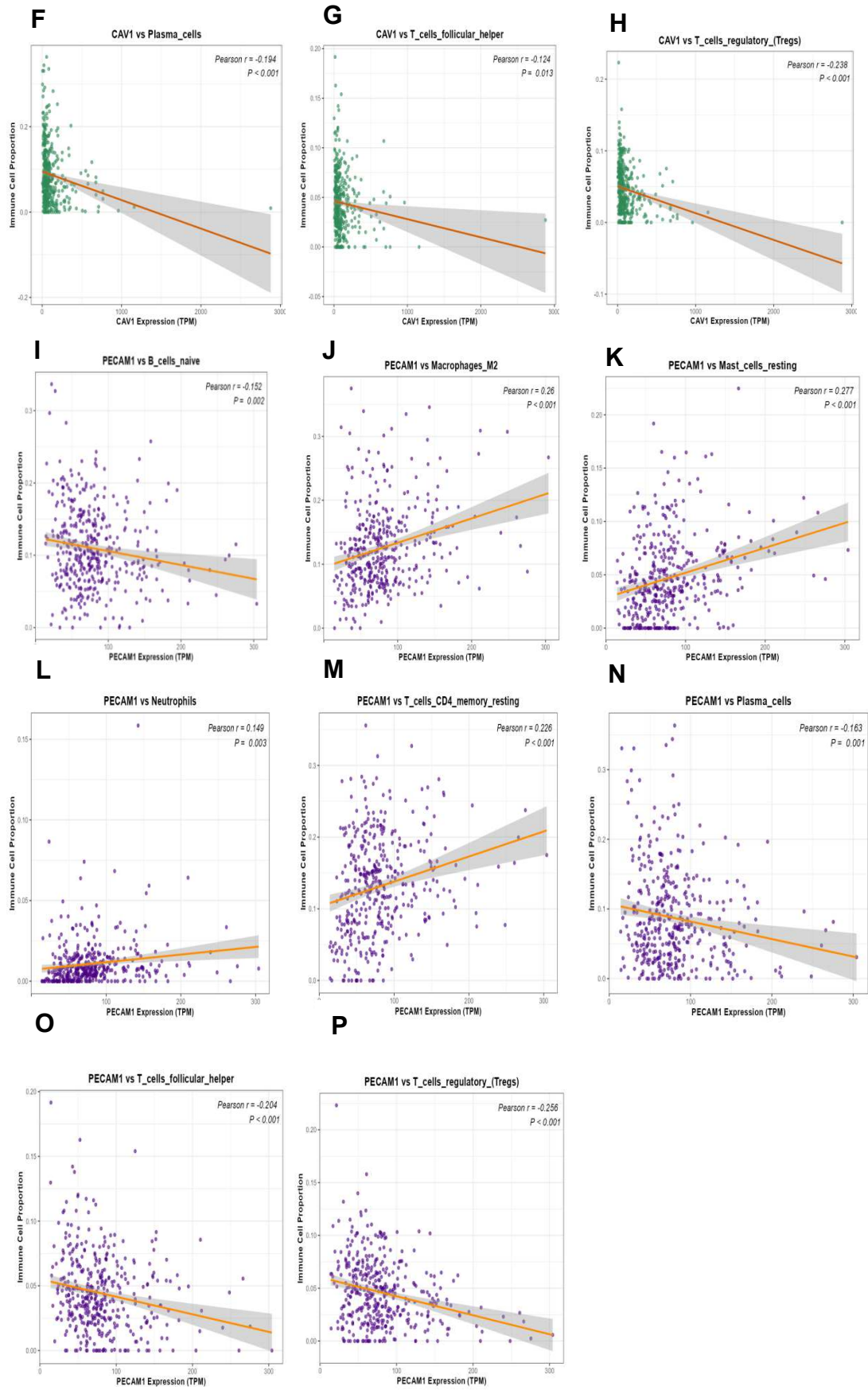
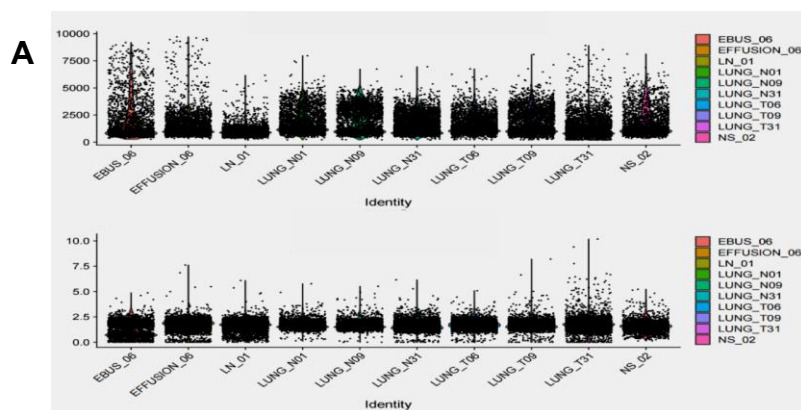


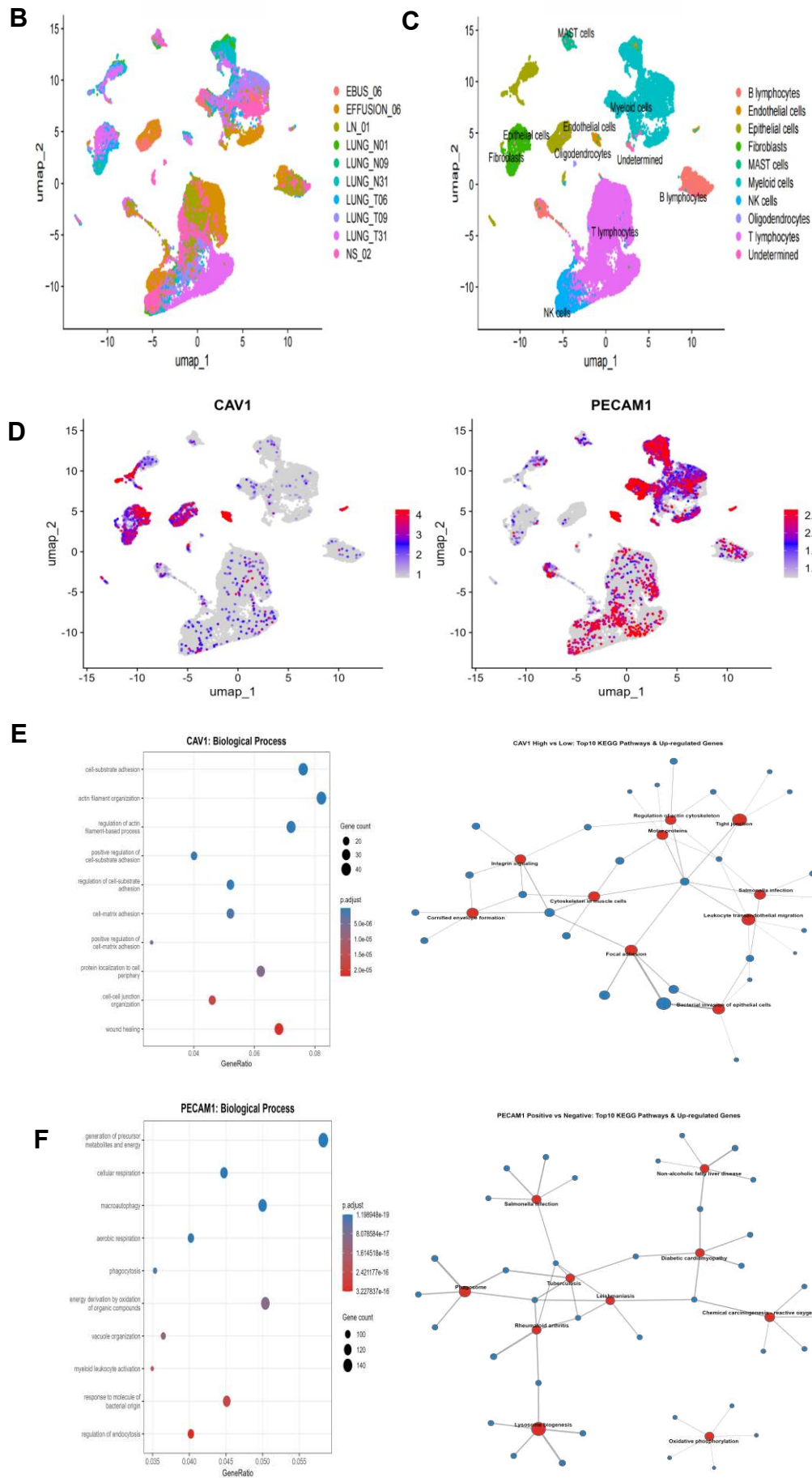
Fig. 6 Correlation between core target genes (CAV1 and PECAM1) and tumor-infiltrating immune cells in LUAD. (A) Heatmap showing the abundance of 22 tumor-infiltrating immune cell subsets in LUAD and normal lung tissues. Red indicates higher abundance, and blue indicates lower abundance. (B) Correlation heatmap of 22 tumor-infiltrating immune cell subsets in the LUAD microenvironment. Red indicates positive correlation, and blue indicates negative correlation. (C-H) Scatter plots showing correlations between CAV1 expression and infiltration levels of key immune cell subsets, including M2-type macrophages, resting mast cells, resting CD4⁺ memory T cells, plasma cells, follicular helper T cells, and regulatory T cells (Tregs). (I-P) Scatter plots showing correlations between PECAM1 expression and infiltration levels of key immune cell subsets, including naive B cells, M2-type macrophages, resting mast cells, neutrophils, resting CD4⁺ memory T cells, plasma cells, follicular helper T cells, and regulatory T cells (Tregs).

Single-cell analysis

To investigate cell-specific expression and potential functions of CAV1 and PECAM1 in the LUAD microenvironment, we performed single-cell transcriptome analysis on the GSE131907 dataset. Quality control indicated that gene counts and mitochondrial gene proportions in each sample met analysis criteria (Fig. 7A), and 35,677 cells were retained for downstream analysis. UMAP dimensionality reduction partitioned cells into major subpopulations, including T lymphocytes, myeloid cells, epithelial cells, and endothelial cells (Fig. 7B-C), with each subpopulation present across samples from different disease states and no evident batch effect. Gene expression patterns revealed that CAV1 was predominantly enriched in epithelial cell subsets, whereas PECAM1 expression was specific to endothelial cells and to several immune cell subsets (Fig. 7D). We therefore performed functional enrichment analyses separately for epithelial cells with high CAV1 expression and for

PECAM1-positive cells. In epithelial cells, genes upregulated in the CAV1 high-expression group were significantly enriched in biological processes including cell-matrix adhesion, wound healing, actin filament organization, and unsaturated fatty acid metabolism. KEGG analysis indicated primary involvement in signaling pathways related to focal adhesion and cell adhesion (Fig. 7E). In endothelial/immune cells, the upregulated genes in the PECAM1-positive group were significantly enriched in processes such as energy metabolism, phagocytosis, and immune cell activation, suggesting that it may be involved in regulating the infiltration and migration of immune cells in the tumor microenvironment (Fig. 7F). To further investigate the regulatory roles of CAV1 and PECAM1 in the LUAD microenvironment, we performed ligand-receptor interaction analysis. Global communication strength analysis confirmed that PECAM1-high endothelial cells possessed enhanced signaling crosstalk with immune compartments (Fig. 7G). Consistently, epithelial cells with high CAV1 expression also displayed significantly increased communication with immune and stromal cells, particularly myeloid cells, T lymphocytes, and fibroblasts (Fig. 7H). Pearson's correlation coefficients (r) and corresponding P-values are presented in each scatter plot.





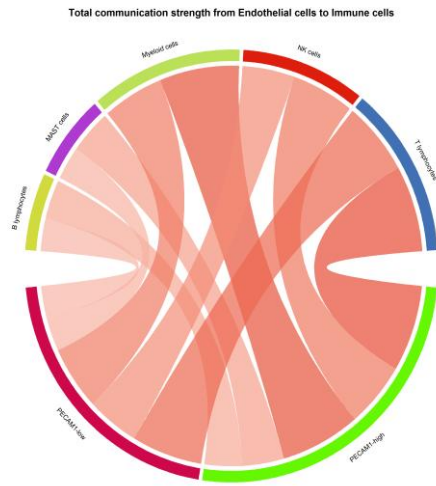
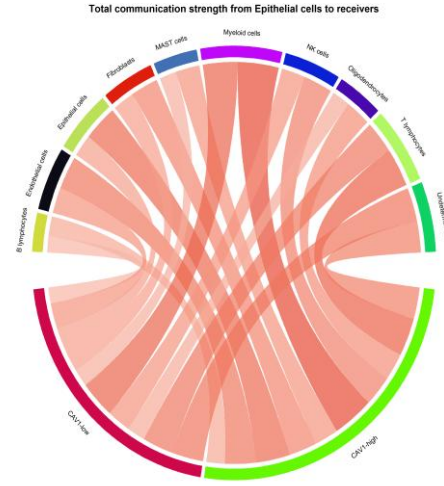
G**H**

Fig 7. Single-cell transcriptomic profiling of LUAD, functional characterization, and epithelial- and endothelial-immune cell communication analysis of CAV1 and PECAM1. (A) Quality control plots showing the number of detected genes per cell (upper) and mitochondrial gene ratio (lower). (B) UMAP visualization of cells colored by sample and cell type. (C) UMAP projection of all cells colored by cell type. (D) Violin plots showing CAV1 and PECAM1 expression across cell types. (E) GO biological process enrichment and KEGG pathway network analysis of upregulated genes in CAV1-high LUAD cells. (F) GO biological process enrichment and KEGG pathway network analysis of upregulated genes in PECAM1-positive LUAD cells. (G) Total communication strength from endothelial cells to immune cells stratified by PECAM1 expression levels. (H) Total communication strength from epithelial cells to immune and stromal cells stratified by CAV1 expression levels. Note: For CAV1, a score ≥ 1.266 is defined as High expression, and the rest are defined as Low expression. For PECAM1, an expression value > 0 is defined as positive (High), and the rest are defined as negative (Low).

Molecular docking

Docking simulation is a practical and efficient approach for probing interactions between small molecules and target sites. We performed docking using Vina 1.2.3 to evaluate compound–protein interactions. In the 7SC0-CAV1_DEHP complex, DEHP establishes extensive hydrophobic contacts with ALA-73, PHE-89, GLU-74, PHE-68, TRP-85, and LYS-86, indicating that the ligand is well accommodated within CAV1's hydrophobic binding cavity. Contact with aromatic and nonpolar residues provides a strong hydrophobic stabilizing effect. Notably, TRP-85 engages in π – π stacking with the ligands, indicating that electronic interactions between aromatic rings contribute substantially to the complex. This interaction both enhances ligand recognition by the protein's aromatic pocket and can increase the stability of the bound conformation. Additionally, a salt bridge between HIS-79 and the ligand indicates that electrostatic interactions also contribute to the complex's stability. The binding of DEHP to CAV1 is driven primarily by hydrophobic interactions and aromatic stacking, with accompanying salt bridges that collectively stabilize the complex, indicating the ligand's spatial adaptability and synergistic interactions at the CAV1 binding site (Fig. 8A). Docking of DEHP with PECAM1(5C14) identified a binding pocket dominated by hydrophobic residues PHE-55, LYS-50, VAL-53, TYR-67, PRO-70, ILE-69, and TYR-80. The hydrophobic effect therefore constitutes the principal driving force maintaining the complex's stability. Multiple aromatic and aliphatic residues jointly encapsulate the ligand, enhancing hydrophobic complementarity with DEHP in the binding pocket and restricting its conformational freedom. A hydrogen bond between TYR-96 and the ligand further improves ligand positioning and stabilizes binding in the pocket. LYS-131 participates in salt-bridge formation, providing an electrostatic anchor that helps consolidate the DEHP–PECAM1 binding mode. Overall, the complex is dominated by hydrophobic interactions and is stabilized further by hydrogen bonds and salt bridges. This network of interactions likely underlies DEHP's persistent

occupancy of the PECAM1 binding pocket(Fig. 8B).

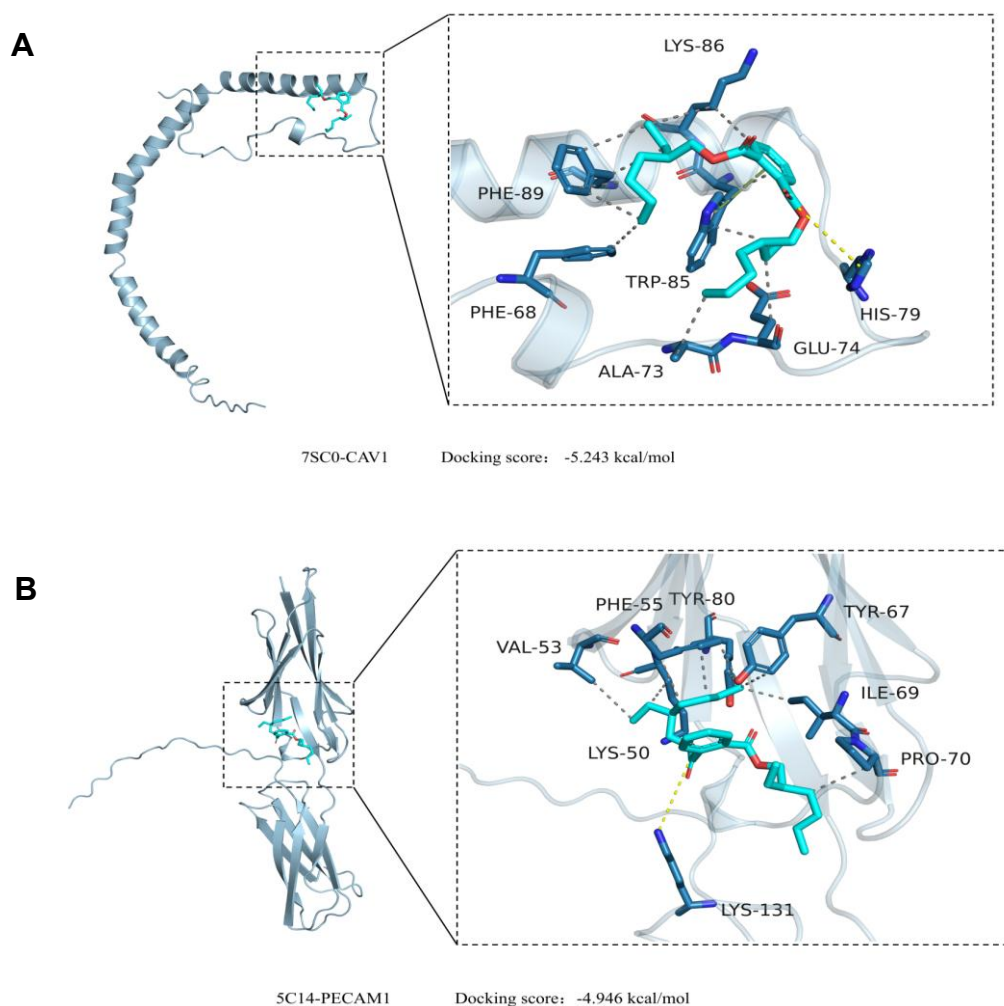


Fig. 8 Molecular docking analysis of DEHP with CAV1 and PECAM1. (A) Three-dimensional binding model of DEHP with CAV1 (PDB: 7SC0). DEHP (yellow stick model) occupies hydrophobic pockets of CAV1 and forms hydrophobic contacts, π - π stacking, and salt bridges with key residues. (B) Three-dimensional binding model of DEHP with PECAM1 (PDB: 5C14). DEHP (yellow stick model) binds the PECAM1 binding site and establishes hydrophobic contacts, hydrogen bonds, and salt bridges with key residues. Legend: Blue cartoon, protein; yellow dotted line, salt bridge; green dotted line, π - π /aromatic interaction; grey dotted line, hydrophobic interaction. Docking score: CAV1, -5.243 kcal/mol; PECAM1, -4.946 kcal/mol.

Molecular dynamics simulation and binding free energy assessment

To investigate the binding mode and dynamic stability of DEHP to the core target protein CAV1, a 100 ns all-atom molecular dynamics simulation was performed using AMBER 24, based on the conformation of the docking complex. The results indicate that the RMSD of the DEHP ligand increases rapidly during the initial phase of the simulation, subsequently stabilizing at a plateau, with no significant dissociation or conformational drift; thus, the binding conformation remains stable (Fig. 9A). After the initial conformational adjustment, the overall RMSD of the complex exhibited stability within a reasonable fluctuation range, and the structure did not experience unfolding or severe rearrangement (Fig. 9B). The RMSF analysis of protein residues reveals that the main structure maintains strong rigidity, while only the local ring regions and terminal residues exhibit moderate flexibility, facilitating ligand recognition and dynamic adaptation at the binding interface (Fig. 9C). The radius of gyration (R_g) of the complex fluctuated steadily throughout the simulation, indicating good compactness of the system, with no occurrences of structural loosening or collapse (Fig. 9D). The number of hydrogen bonds between ligands and proteins exhibits consistent fluctuations, thereby establishing a reliable interaction network (Fig. 9E). The solvent-accessible surface area (SASA) of the compound initially decreases slightly before stabilizing, maintaining a compact conformation without significant expansion (Fig. 9F). The MM-GBSA calculations reveal that the binding free energy of DEHP to CAV1 is -29.62 ± 3.89 kcal/mol, with van der Waals interactions primarily contributing to the binding force (Table 1). Decomposition of the binding energy associated with key amino acid residues indicates that hydrophobic and aromatic residues predominantly facilitate the binding process through hydrophobic stacking and π - π interactions (Fig. 9G). The free energy landscape (FEL) display system is characterized by a concentration in a single low-energy conformation basin, exhibiting excellent

conformational convergence, a high transition energy barrier, and remarkable thermodynamic stability(Fig. 9H,I). These findings confirm that DEHP can form a high-affinity and stable binding complex with CAV1, thereby providing a structural basis for elucidating the molecular mechanism through which DEHP promotes the progression of LUAD.

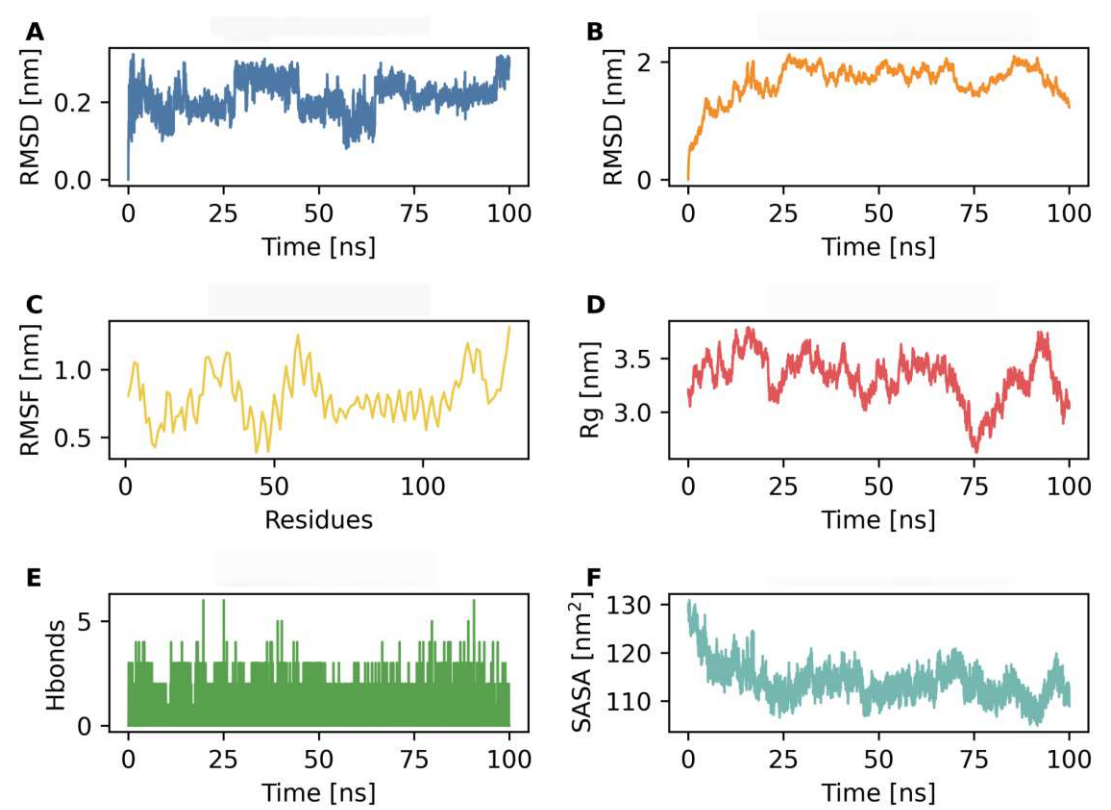


Table 1: Binding free energies and energy components for CAV1-DEHP complex

ΔE_{vdW} : van der Waals energy. ΔE_{elec} : electrostatic energy. ΔG_{GB} : electrostatic contribution to solvation. ΔG_{SA} : non-polar contribution to solvation. G_{bind} : binding free energy.

System	CAV1/ DEHP
name	

ΔE_{vdw}	-35.42 ± 4.33
ΔE_{elec}	-3.75 ± 4.45
ΔG_{GB}	14.36 ± 3.82
ΔG_{SA}	-4.80 ± 0.31
ΔG_{bind}	-29.62 ± 3.89

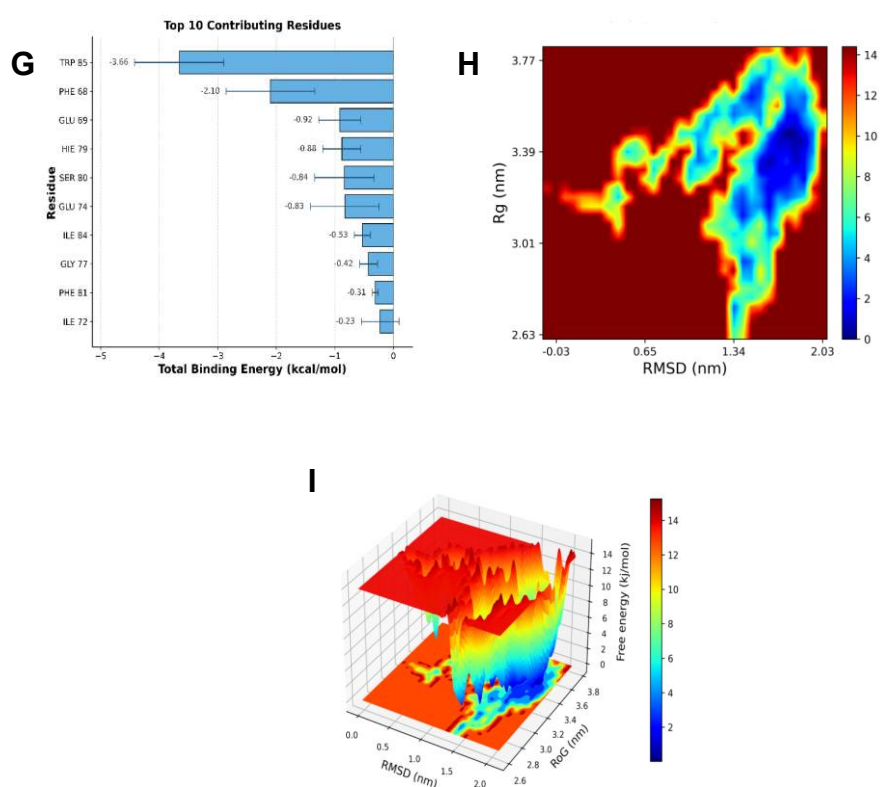


Fig. 9 Molecular dynamics simulation and binding stability analysis of the CAV1–DEHP complex.

(A–F) Stability profiles of the CAV1–DEHP complex during 100 ns simulation, including ligand RMSD (A), complex RMSD (B), protein RMSF (C), radius of gyration (D), hydrogen bond number (E), and solvent-accessible surface area (F). (G) Per-residue binding free energy decomposition. (H,I) Free energy landscape of the system.

Chemotherapeutic drug-response survival analysis

We further investigated whether CAV1 and PECAM1 expression could predict clinical responses to conventional chemotherapeutic agents in LUAD patients using GEPIA3-derived survival data. Forest plot analysis revealed that, in patients with high CAV1 expression, none of the tested chemotherapeutic agents exhibited statistically significant survival benefits (all adjusted $P > 0.05$, Fig. 10 left). In contrast, among patients with high PECAM1 expression, pemetrexed treatment significantly reduced the risk of death (HR = 0.58, 95% CI: 0.34–0.99, adjusted $P < 0.05$) (Fig. 10 right). No other drugs showed significant prognostic effects in the PECAM1-high subgroup. These findings indicate that PECAM1 expression status may serve as a potential predictive biomarker for pemetrexed-based chemotherapy in LUAD patients, providing clinical evidence for individualized therapeutic strategy selection.

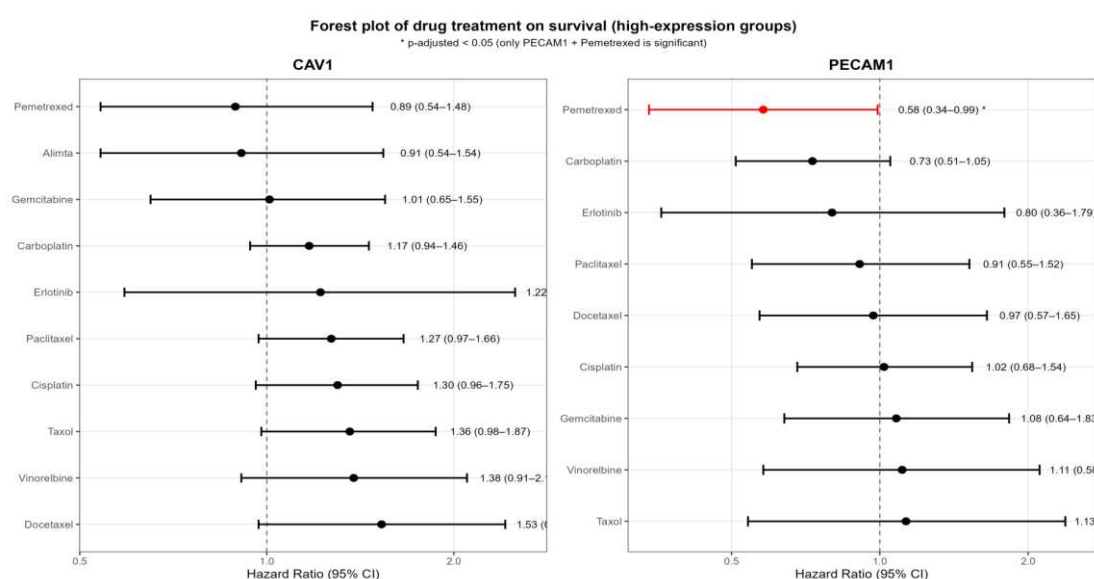


Fig. 10. Forest plot of survival hazard ratios for different chemotherapeutic agents in CAV1-high and PECAM1-high LUAD patients. Hazard ratios (HR) and 95% confidence intervals (CIs) for overall survival under different drug treatments were calculated in patients with high CAV1 (left panel) and high PECAM1 (right panel) expression using the GEPIA3 database. Only pemetrexed showed a statistically significant survival-protective effect in the PECAM1-high subgroup (*adjusted $P < 0.05$).

Discussion

DEHP is a widely produced plasticizer used in polymer materials. It is found in numerous plastic products, including food packaging, medical devices, children's toys, and personal care items^[33]. Because DEHP does not form covalent bonds with the polymer matrix, it can leach continuously into the surrounding environment and cause chronic human exposure via ingestion, inhalation, and dermal contact^[34]. Epidemiological and toxicological studies have shown that DEHP, a representative endocrine disruptor, is linked to adverse outcomes such as abnormal reproductive development^[35] and metabolic disorders^[36]. Recently, concern has grown over its potential carcinogenicity. The International Agency for Research on Cancer of the World Health Organization has classified DEHP as a possible human carcinogen (Group 2B)^[37]. Current research has not yet elucidated the molecular mechanisms linking DEHP to specific cancers, particularly LUAD. As the most common subtype of non-small cell lung cancer, LUAD has been rising in incidence worldwide. A substantial proportion of patients lack a clear history of smoking, which implicates environmental factors in its pathogenesis. Consequently, investigating the relationship between ubiquitous environmental pollutants such as DEHP and the initiation and progression of LUAD has important public health implications. Mono (2-ethylhexyl) phthalate (MEHP), a metabolite of DEHP, has been linked to LUAD^[38]. Nonetheless, the direct role and primary targets of the DEHP prototype remain ambiguous. This study investigates the DEHP prototype to elucidate the relationship between prevalent environmental pollutants, such as DEHP, and the onset and progression of LUAD, a matter of considerable public health importance. This study is the first to integrate network toxicology, machine learning, immune deconvolution, single-cell transcriptomics, molecular dynamics simulation and chemotherapy response prediction to build a systematic association network linking DEHP exposure with LUAD onset and progression. Rather than seeking novel carcinogenic drivers, the work aligns DEHP's toxicological profile with established

LUAD molecular pathways and explicitly connects the environmental pollutant DEHP, core target genes (CAV1, PECAM1), and tumor immune microenvironment remodeling within a unified regulatory framework. This approach contributes an environmental toxicology perspective to current LUAD biology. From a toxicological perspective, ProTox predicts that DEHP exerts pronounced endocrine-disrupting effects and a propensity for respiratory toxicity. The compound may mediate chronic lung injury via multiple mechanisms, including interference with the CYP450 enzyme system, activation of oxidative stress pathways (SRARE), and binding to sex hormone receptors. These findings align with epidemiological reports linking DEHP exposure to increased risks of respiratory disease and lung cancer, implying that persistent DEHP exposure may not create novel oncogenic circuits but rather promote the initiation and progression of LUAD by reshaping established pro-tumor signaling pathways. Network toxicology analysis identified 485 DEHP-LUAD intersection targets. GO and KEGG enrichment showed these targets were strongly associated with disrupted balance between proliferation and apoptosis, PI3K-Akt signaling, MAPK cascade activation, HIF-1-mediated hypoxia responses, and tumor-associated inflammatory processes. Notably, PI3K-Akt and MAPK represent canonical pro-survival and pro-proliferation axes in LUAD, while HIF-1 governs tumor angiogenesis and metabolic reprogramming^[39]. These enrichment results indicate that DEHP-driven LUAD likely involves concurrent dysregulation of several classical pathways. To identify key mediators from a vast number of targets, we then employed three machine learning algorithms, namely LASSO, SVM - RFE, and random forest, for feature selection. All three methods consistently screened out CAV1 and PECAM1 as the core targets of DEHP - induced LUAD. Cross - cohort expression validation indicated that both CAV1 and PECAM1 were significantly down - regulated in LUAD tissues. Moreover, the combined diagnostic model of the two genes demonstrated excellent discriminatory performance in

distinguishing tumor from normal samples (AUC values were stable at around 0.98), suggesting their potential as highly specific clinical diagnostic markers. Kaplan–Meier survival analysis further validated the clinical prognostic value of CAV1 and PECAM1. Patients with high CAV1 expression exhibited significantly longer overall survival (HR = 0.81, 95% CI: 0.68–0.96, log-rank P = 0.016), and high PECAM1 expression was similarly associated with favorable prognosis (HR = 0.69, 95% CI: 0.54–0.88, log-rank P = 0.0027). These results confirm that reduced expression of CAV1 and PECAM1 correlates with poor clinical outcomes in LUAD, supporting their roles as tumor suppressors and independent prognostic indicators. Downregulation of these two genes may thus serve as a warning sign for disease progression and poor survival. CAV1 (Caveolin - 1) is the major structural protein of caveolae in the cell membrane and participates in various physiological processes such as signal transduction, substance transport, cell cycle regulation, mitochondrial function, and bone metabolism^[40]. Recent studies have shown that CAV1 plays a role of tumor suppression . Its low expression in lung cancer is closely associated with poor prognosis and enhanced invasion and metastasis ability^[41]. PECAM1 (platelet endothelial cell adhesion molecule 1, also known as CD31) is an iconic molecule of vascular endothelial cells, which dominates the adhesion connection between endothelial cells, vascular homeostasis, and inflammation regulation^[42]. The down - regulation of CAV1 and PECAM1 may disrupt the structural integrity of alveolar epithelium and vascular endothelium, weakening their tumor - suppressing barrier function and thus synergistically promoting DEHP - induced malignant transformation. At the level of the tumor immune microenvironment, CIBERSORT deconvolution analysis showed that the infiltration levels of regulatory T cells (Tregs), plasma cells, and follicular helper T cells (Tfh) were significantly increased in LUAD tissues, while resting M2 macrophages, mast cells, and NK cells were more enriched in normal lung tissues. This

suggests that tumors actively reshape the immune landscape to construct an inhibitory microenvironment conducive to their own growth. Further correlation analysis revealed that high expression of CAV1 was significantly negatively correlated with Tregs, plasma cells, and Tfh, and positively correlated with M2 macrophages and resting mast cells. As a vascular endothelial marker, PECAM1 was positively correlated with the infiltration of resting mast cells, M2 macrophages, and neutrophils, and negatively correlated with Tregs, Tfh, plasma cells, and naive B cells. These findings indicate that CAV1 and PECAM1 may participate in the formation of the tumor immunosuppressive microenvironment. Specifically, the down-regulation of CAV1 may relieve its inhibitory effect on Tregs recruitment, promoting immune escape. The activation of vascular endothelium mediated by PECAM1 may indirectly regulate the infiltration of effector immune cells into tumor foci by affecting vascular permeability. Single-cell transcriptome sequencing (scRNA-seq) analysis further validated the above findings at the cellular precision level. Analysis based on the GSE131907 dataset showed that CAV1 was highly expressed mainly in myeloid cells, endothelial cells, and fibroblasts, while PECAM1 was specifically enriched in the endothelial cell population (positive rate over 75%). By comparing the epithelial cell subsets with high and low CAV1 expression, we identified 5,176 differentially expressed genes. Functional enrichment analysis suggested that these differentially expressed genes were mainly involved in cell-matrix adhesion, antigen processing and presentation, and the PI3K-Akt signaling pathway, implying that CAV1 may maintain the integrity of the alveolar barrier by supporting the migration and damage-repair ability of epithelial cells. Meanwhile, its leading lipid metabolism pathway ensures normal material transport and energy supply of cells. This function of maintaining tissue homeostasis and repair potential is the molecular basis for CAV1 to play a natural tumor-suppressing role in the normal physiological environment. In the immune and

endothelial cell populations, PECAM1 - positive cells were mainly endothelial cells and myeloid cells, which were enriched in phagocytosis, lysosome function, and immune surveillance - related pathways, suggesting that PECAM1 plays a tumor - suppressing role by maintaining vascular endothelial integrity and innate immune homeostasis. According to cell communication analyses, ligand-receptor crosstalk between immune cells and CAV1/PECAM1-high endothelial or epithelial cells was markedly enhanced relative to that in the low-expression groups. Endothelial cells predominantly interacted strongly with myeloid cells, T lymphocytes and NK cells. PECAM1 showed a more extensive capacity to activate immune cell communication. Collectively, these results indicate that CAV1 and PECAM1 remodel the endothelial-immune communication network. This process alters the recruitment, migration and activation of immune cells, and facilitates the formation of an immunosuppressive microenvironment. Our data offer a new mechanistic explanation for DEHP-mediated progression of LUAD. At the level of structural biology, molecular docking and 100 ns explicit solvent molecular dynamics simulations provided dynamic support for the direct interaction between DEHP and the core targets under near - physiological conditions. The AutoDock Vina docking results showed that DEHP had considerable binding affinity with both CAV1 and PECAM1. The binding free energy of the CAV1 - DEHP complex calculated by MM/GBSA was -29.62 ± 3.89 kcal/mol, indicating a strong affinity between them. The MD trajectory showed that after equilibrium, the RMSD of the receptor - ligand complex fluctuated little and was stable, the radius of gyration and solvent - accessible surface area were basically constant, and there were persistent hydrogen - bond interactions at the binding interface, indicating that the docking conformation corresponded to a thermodynamically favorable binding mode rather than a docking artifact. The MD - based research results greatly strengthened the mechanistic rationality of the DEHP - CAV1/PECAM1 axis in LUAD. In conclusion, these observations suggest that DEHP

exposure can directly bind to and down - regulate CAV1 and PECAM1. On the one hand, it weakens the epithelial barrier and vascular homeostasis, activating classical proliferation and invasion pathways such as PI3K - Akt and MAPK. On the other hand, it promotes the enrichment of immunosuppressive cells such as Tregs by reshaping the immune infiltration landscape, ultimately driving the occurrence and development of LUAD.

Limitations

However, it must be acknowledged that this study has some limitations. First, all results are derived from computational analyses, including database-driven target mining, machine-learning-based feature selection, immune infiltration estimation, single-cell transcriptomic analysis, molecular docking, and molecular dynamics simulations; therefore, these findings should be interpreted as hypothesis generation rather than conclusive mechanistic evidence. Second, the study lacked experimental validation (e.g., DEHP exposure experiments in LUAD cell lines, functional perturbation of CAV1/PECAM1, or in vivo tumor models), which limits the strength of causal inference, and no validating wet-laboratory study has yet been performed. Third, although multiple independent datasets were used, the dataset sizes were moderate (516 tumor and 59 normal samples in the TCGA-LUAD cohort; three GEO validation cohorts of limited sample size; 35,677 cells from a subset of the GSE131907 single-cell dataset); larger patient cohorts and prospective epidemiological data (including biomonitoring of DEHP exposure levels) will further enhance the generalizability of our findings. Therefore, future research should prioritize: (1) validating the expression changes of CAV1/PECAM1 and the activation of downstream pathways in DEHP-exposed LUAD cells and organoid models; (2) clarifying the causal role of CAV1/PECAM1 in DEHP-induced malignant transformation using gene-editing techniques; (3) evaluating the dose-response relationship between the internal DEHP

exposure dose, the risk of LUAD, and the expression levels of CAV1/PECAM1 in tumors by combining biomonitoring data from population cohorts.

Conclusion

In summary, we combined network toxicology, machine learning, multi-omics analyses, immune infiltration assessment, single-cell transcriptomics, and molecular simulations to explore how DEHP contributes to LUAD progression. We identified CAV1 and PECAM1 as core functional genes mediating the oncogenic effects of DEHP, without discovering new cancer-driving factors. These bioinformatic results construct a rational framework to elucidate DEHP-related toxicity in lung cancer. Given that all conclusions are derived from computational analyses, further experimental verification is urgently needed. This study offers valuable hypotheses and references for subsequent mechanistic exploration and clinical translational research on DEHP-associated LUAD.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Quan Chen: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – Original Draft, Visualization.

Yuying Li: Review ,Editing, Supervision, Project administration.

Availability of data and materials

All datasets analysed in the current study are publicly available, and no new datasets were generated.

The TCGA-LUAD RNA-sequencing cohort (516 tumour and 59 normal samples) is available from The Cancer Genome Atlas (TCGA) at the NCI Genomic Data Commons project record

(<https://portal.gdc.cancer.gov/projects/TCGA-LUAD>); expression data were downloaded via the GDC data download tool (https://portal.gdc.cancer.gov/analysis_page?app=Downloads). The four gene expression datasets used for validation are available from the Gene Expression Omnibus (GEO): GSE19804 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19804>), GSE32863 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE32863>), GSE40419 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE40419>), and GSE131907 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131907>). The chemical structure and canonical SMILES of di-(2-ethylhexyl) phthalate (DEHP) were retrieved from PubChem (compound CID 8343, <https://pubchem.ncbi.nlm.nih.gov/compound/8343>). The three-dimensional molecular structures used for docking were retrieved from the Protein Data Bank (PDB): CAV1 (PDB ID 7SC0, <https://www.rcsb.org/structure/7SC0>) and PECAM1 (PDB ID 5C14, <https://www.rcsb.org/structure/5C14>). DEHP-related target genes were predicted from ChEMBL (DEHP compound record CHEMBL1242017, <https://www.ebi.ac.uk/chembl/explore/compound/CHEMBL1242017>), the Comparative Toxicogenomics Database (CTD) chemical-gene interaction query page (<https://ctdbase.org/query.go?type=ixn>, queried with "di-(2-ethylhexyl) phthalate"), DIGEP-Pred 2.0 (<https://way2drug.com/digep-pred/>), and SwissTargetPrediction (<https://swisstargetprediction.ch/predict.php>). LUAD-related genes were retrieved from GeneCards (search results for "Lung adenocarcinoma", <https://www.genecards.org/search/results?q=Lung+adenocarcinoma>, relevance score ≥ 20), the Therapeutic Target Database (TTD) disease-class search results for "2C25: Lung cancer" (<https://ttd.idrblab.cn/search-result/target-drug?engine=diseaseClass&keyword=2C25:+Lung+cancer>),

and Online Mendelian Inheritance in Man (OMIM) search results for "Lung adenocarcinoma" (https://www.omim.org/search?index=entry&limit=10&sort=score+desc%2C+prefix_sort+desc&search=Lung+adenocarcinoma). Toxicity prediction was performed on the ProTox 3.0 compound search page (https://tox.charite.de/protox3/index.php?site=compound_search_similarity). Prognostic analysis was performed on the Kaplan-Meier Plotter lung cancer analysis page (<https://kmplot.com/analysis/index.php?p=service&cancer=lung>), and chemotherapy drug-response survival analysis was performed on the GEPIA3 drug-response analysis page (https://gepia3.bioinfliu.com/drug_response/#drug_resp_result).

Declarations

Ethical Approval and Accordance: Not applicable. This study did not involve human participants, human tissues, or animals. All analyses were conducted entirely using publicly available, de-identified datasets from public databases; therefore, formal ethics committee approval was not required.

Consent to participate: Not applicable.

Consent to publish: Not applicable.

Competing interests: The authors declare that they have no competing interests.

Clinical trial number: Not applicable.

References

- [1].Qin Z, Li L, Wang C. Molecular mechanism of resveratrol in suppressing lung adenocarcinoma based on network pharmacology and molecular docking. *Medicine (Baltimore)*. 2025;104(36):e44489. doi:10.1097/MD.00000000000044489
- [2]. Li N, Li R, Song B, Ge H, Qu Y. Downregulation of hsa-miR-30b-3p Inhibits the Oncogenicity of Lung Adenocarcinoma by Targeting the METTL7B Gene. *J Oncol*. 2022;2022:6883140. Published 2022 May 20. doi:10.1155/2022/6883140

- [3]. Bhaliya K, Anwer M, Wei MQ. Emerging Therapeutic Strategies for Lung Cancer: The Role of Immunotherapy and HPV-Targeted Cancer Vaccines. *Vaccines (Basel)*. 2025;13(9):957. Published 2025 Sep 8. doi:10.3390/vaccines13090957
- [4]. Zhu Z, Qin J, He C, et al. Computational analysis of DEHP's oncogenic role in colorectal cancer. *Discov Oncol*. 2025;16(1):848. Published 2025 May 21. doi:10.1007/s12672-025-02616-x
- [5]. Yu J, Huang G, Gong Q, et al. MicroRNA-375 Mediated Regulation on Pre-mRNA Processing Factor 3 in Zebrafish Embryos Exposed to Di-(2-ethylhexyl)phthalate at Low Concentrations. *Chem Res Toxicol*. 2023;36(1):32-42. doi:10.1021/acs.chemrestox.2c00257
- [6]. Su WC, Tsai YC, Chang TK, et al. Correlations between Urinary Monoethylhexyl Phthalate Concentration in Healthy Individuals, Individuals with Colorectal Adenomas, and Individuals with Colorectal Cancer. *J Agric Food Chem*. 2021;69(25):7127-7136. doi:10.1021/acs.jafc.1c00953
- [7]. Zhou X, Ye W, Xu J, et al. The role of di-(2-ethylhexyl) phthalate in cancer initiation and progression: Mechanisms and health implications. *Sci Total Environ*. 2025;959:178285. doi:10.1016/j.scitotenv.2024.178285
- [8]. Zuo X, Hou X, Zhang Y, Zhang C, Zhang J. DEHP promotes psoriasis via immune modulation and direct molecular interactions: Evidence from epidemiology, multi-omics, and structural simulation. *Sci Total Environ*. 2025;1003:180723. doi:10.1016/j.scitotenv.2025.180723
- [9]. Alfardan AS, Nadeem A, Ahmad SF, Al-Harbi NO, Al-Harbi MM, AlSharari SD. Plasticizer, di(2-ethylhexyl)phthalate (DEHP) enhances cockroach allergen extract-driven airway inflammation by enhancing pulmonary Th2 as well as Th17 immune responses in mice. *Environ Res*. 2018;164:327-339. doi:10.1016/j.envres.2018.02.039
- [10]. Wang CW, Chen SC, Wu DW, et al. Effect of dermal phthalate levels on lung function tests in residential area near a petrochemical complex. *Environ Sci Pollut Res Int*. 2021;28(21):27333-27344. doi:10.1007/s11356-020-12322-6
- [11]. Hu HY, Zhang ZZ, Jiang XY, Duan TH, Feng W, Wang XG. Hesperidin Anti-Osteoporosis by Regulating Estrogen Signaling Pathways. *Molecules*. 2023;28(19):6987. Published 2023 Oct 9. doi:10.3390/molecules28196987

- [12]. Lu TP, Tsai MH, Lee JM, et al. Identification of a novel biomarker, SEMA5A, for non-small cell lung carcinoma in nonsmoking women. *Cancer Epidemiol Biomarkers Prev.* 2010;19(10):2590-2597. doi:10.1158/1055-9965.EPI-10-0332
- [13]. Selamat SA, Chung BS, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res.* 2012;22(7):1197-1211. doi:10.1101/gr.132662.111
- [14]. Seo JS, Ju YS, Lee WC, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* 2012;22(11):2109-2119. doi:10.1101/gr.145144.112
- [15]. Zeng D, Fang Y, Qiu W, et al. Enhancing immuno-oncology investigations through multidimensional decoding of tumor microenvironment with IOBR 2.0. *Cell Rep Methods.* 2024;4(12):100910. doi:10.1016/j.crmeth.2024.100910
- [16]. Kim N, Kim HK, Lee K, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun.* 2020;11(1):2285. Published 2020 May 8. doi:10.1038/s41467-020-16164-1
- [17]. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model.* 2021;61(8):3891-3898. doi:10.1021/acs.jcim.1c00203
- [18]. DeLano, W. L., Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography* 2002, 40 (1), 82-92.
- [19]. Ravindranath, P. A.; Forli, S.; Goodsell, D. S.; Olson, A. J.; Sanner, M. F., AutoDockFR: Advances in Protein-Ligand Docking with Explicitly Specified Binding Site Flexibility. *PLoS Comput. Biol.* 2015, 11 (12), e1004586.
- [20]. Salomon - Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2013, 3 (2), 198-210. DOI: 10.1002/wcms.1121.
- [21]. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Antechamber: an accessory software package for molecular mechanical calculations. *J. Am. Chem. Soc* 2001, 222, U403.
- [22]. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* 2004, 25 (9), 1157-1174.

- [23]. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 2015, 11 (8), 3696-3713. DOI: 10.1021/acs.jctc.5b00255.
- [24]. Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* 2001, 105 (43), 9954-9960. DOI: 10.1021/jp003020w.
- [25]. Sagui, C.; Darden, T. A. Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annu. Rev. Biophys. Biomol. Struct.* 1999, 28 (1), 155-179. DOI: 10.1146/annurev.biophys.28.1.155.
- [26]. Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 2001, 22 (5), 501-508. DOI: 10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.
- [27]. Larini, L.; Mannella, R.; Leporini, D. Langevin stabilization of molecular-dynamics simulations of polymers by means of quasisymplectic algorithms. *J. Chem. Phys.* 2007, 126 (10), 104101. DOI: 10.1063/1.2464095 From NLM.
- [28]. Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* 2011, 51 (1), 69-82. DOI: 10.1021/ci100275a From NLM.
- [29]. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert opinion on drug discovery* 2015, 10 (5), 449-461. DOI: 10.1517/17460441.2015.1032936.
- [30]. Rastelli, G.; Rio, A. D.; Degliesposti, G.; Sgobba, M. Fast and accurate predictions of binding free energies using MM - PBSA and MM - GBSA. *J. Comput. Chem.* 2010, 31 (4), 797-810.
- [31]. Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* 2013, 9 (4), 2020-2034. DOI: 10.1021/ct3010485.
- [32]. Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO).

J. Comput. Chem. 1999, 20 (2), 217-230.

[33]. Ding M, Ma H, Du H, Yang Y, Yu M, Zhang C. Omega-3 fatty acid normalizes postsynaptic density related miRNAs and proteins in hippocampus and prevents DEHP-induced impairment of learning and memory in mice. PLoS One. 2025;20(7):e0313233. Published 2025 Jul 3. doi:10.1371/journal.pone.0313233

[34]. Rowdhwal SSS, Chen J. Toxic Effects of Di-2-ethylhexyl Phthalate: An Overview. Biomed Res Int. 2018;2018:1750368. Published 2018 Feb 22. doi:10.1155/2018/1750368

[35]. Zhang K, Lin Y, Sun X, et al. Differential life cycle toxic effects and molecular mechanisms of Di(2-ethylhexyl) phthalate (DEHP) exposure on the female reproductive system. Toxicology. 2025;518:154275. doi:10.1016/j.tox.2025.154275

[36]. Lin Y, Wei J, Li Y, et al. Developmental exposure to di(2-ethylhexyl) phthalate impairs endocrine pancreas and leads to long-term adverse effects on glucose homeostasis in the rat. Am J Physiol Endocrinol Metab. 2011;301(3):E527-E538. doi:10.1152/ajpendo.00233.2011

[37]. Rusyn I, Corton JC. Mechanistic considerations for human relevance of cancer hazard of di(2-ethylhexyl) phthalate. Mutat Res. 2012;750(2):141-158. doi:10.1016/j.mrrev.2011.12.004

[38]. Wang W, Li J, Qie X. Comprehensive utilization of in silico approach and in vitro experiment to unveil the molecular mechanisms of mono (2-ethylhexyl) phthalate-induced lung adenocarcinoma. Bioorg Chem. 2024;153:107947. doi:10.1016/j.bioorg.2024.107947

[39]. Yuan M, Huang LL, Chen JH, Wu J, Xu Q. The emerging treatment landscape of targeted therapy in non-small-cell lung cancer. Signal Transduct Target Ther. 2019;4:61. Published 2019 Dec 17. doi:10.1038/s41392-019-0099-9

[40]. Williams TM, Lisanti MP. The caveolin proteins. Genome Biol. 2004;5(3):214. doi:10.1186/gb-2004-5-3-214

[41]. Dong J, Du R, Zhang Y, Hong B, Sun G. Caveolin-1 inhibits the proliferation and invasion of lung adenocarcinoma via EGFR degradation. Sci Rep. 2025;15(1):21654. Published 2025 Jul 1. doi:10.1038/s41598-025-05259-8

[42]. Liu Y, Zhang Z, Li W, Tian S. PECAM1 Combines With CXCR4 to Trigger Inflammatory Cell Infiltration and Pulpitis

Progression Through Activating the NF- κ B Signaling Pathway. *Front Cell Dev Biol.* 2020;8:593653. Published 2020 Dec 23.

doi:10.3389/fcell.2020.593653