scientific reports

Check for updates

OPEN

Integrative network toxicology and experimental evidence reveal mechanisms underlying diethyl phthalate-induced initiation and progression of endometrial cancer

Xi Chen[1,7], Zijing Wang[2,7], Fengfeng Wang[3], Yuexiao Wu[1], Dan Hu[4], Yuemei Cheng[4], Yijuan Xing[4], Junhong Du[4], Tao Jiang[5]✉, Yongxiu Yang[4,6]✉, Xiaolei Liang[4,6]✉ & Xuehan Bi[4,6]✉

Diethyl phthalate (DEP) is a ubiquitous environmental endocrine-disrupting chemical (EDC). Epidemiological studies have suggested a potential association between DEP exposure and an increased risk of endometrial cancer (EC); however, its underlying molecular mechanisms remain largely unclear. Four GEO datasets were integrated, and differential expression analysis combined with weighted gene co-expression network analysis (WGCNA) was performed to identify candidate genes potentially linking DEP exposure to EC. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted to explore relevant signaling pathways. Machine learning models, coupled with Shapley Additive Explanations (SHAP), were employed to prioritize key genes. Molecular docking and molecular dynamics (MD) simulations were used to assess the binding affinity between DEP and the identified targets. A series of in vitro experiments in EC cell lines were subsequently conducted to validate the biological effects of DEP. Nineteen overlapping DEP–EC genes were identified, predominantly enriched in the MAPK, cAMP, and cGMP–PKG signaling pathways. Among them, FOS, NR4A1, ADRA2C, JUN, and SLC6A2 were prioritized as core genes through machine learning and SHAP analysis. Molecular simulations confirmed stable binding between DEP and these targets. In vitro assays demonstrated that DEP exposure induces oxidative stress, significantly enhances ERK1/2 and AKT phosphorylation, upregulates Cyclin D1/CDK4 expression, promotes G1/S phase transition, and facilitates EC cell proliferation. These findings suggest that DEP may promote endometrial carcinogenesis by triggering oxidative stress-mediated signaling crosstalk and accelerating cell cycle progression. This study establishes a multi-layered methodological framework-from computational screening and machine learning to experimental validation-offering novel mechanistic insights into the carcinogenic potential of environmental endocrine disruptors such as DEP.

Keywords Endometrial cancer, Diethyl phthalate, Network toxicology, Machine learning, Molecular dynamics simulation

Endometrial cancer (EC) is one of the most prevalent malignancies of the female reproductive system and ranks sixth in global cancer incidence among women[1]. According to GLOBOCAN 2020 data, approximately 417,000 new EC cases and 97,000 related deaths occur annually worldwide[2]. The incidence is particularly high

[1]The First Clinical Medical College of Lanzhou University, Lanzhou 730000, Gansu, China. [2]The First Clinical Medical College, General Hospital of Ningxia Medical University, Yinchuan 750004, China. [3]Wushan Hospital of Traditional Chinese Medicine, Tianshui 741306, Gansu, China. [4]Department of Obstetrics and Gynecology, the First Hospital of Lanzhou University, Lanzhou 730000, Gansu, China. [5]Department of Anal-Colorectal Surgery, General Hospital of Ningxia Medical University, 804 Shengli Road, Yinchuan 750004, China. [6]Department of Gynecology, the First Hospital of Lanzhou University, Chengguan District, No. 1, Donggangxi Road, Lanzhou 730000, Gansu, China. [7]These authors contributed equally: Xi Chen and Zijing Wang. ✉email: jtyy1209@163.com; yxyanglzu@163.com; liangxl07@lzu.edu.cn; bxh2022@163.com

in Western countries, where EC accounts for over 20% of all female reproductive tract malignancies[3]. In China, the incidence of EC has increased markedly, reaching 6.1 per 100,000 population in 2015-representing a more than 100% rise over the past decade-primarily driven by population aging and changes in lifestyle patterns[4]. Epidemiological studies have consistently linked EC development to prolonged estrogen exposure, obesity, diabetes, and metabolic syndrome[5]. However, these established risk factors alone do not fully explain the rapid and widespread increase in EC incidence, indicating that, in addition to endogenous hormonal imbalances, exogenous environmental exposures, particularly endocrine-disrupting chemicals (EDCs), may also contribute to EC initiation and progression.

Among various exogenous risk factors, EDCs have attracted considerable attention for their ability to interfere with endogenous hormone synthesis and signal transduction[6]. Phthalates (PAEs), one of the most pervasive classes of EDCs, are widely used in plastic products, personal care items, food packaging, and medical devices[7]. In recent years, accumulating evidence from experimental and animal studies has demonstrated the tumor-promoting potential of PAEs across multiple cancer types[8]. For instance, dibutyl phthalate (DBP) enhances the proliferation of MCF-7 breast cancer cells and exhibits estrogenic activity[9], whereas butyl benzyl phthalate (BBP) promotes breast cancer cell migration via activation of the ERK1/2 signaling pathway[10]. In ovarian cancer, mono(2-ethylhexyl) phthalate (MEHP), a metabolite of di(2-ethylhexyl) phthalate (DEHP), induces epithelial–mesenchymal transition (EMT) and increases the invasiveness of SKOV3 cells through the PPARα–PI3K/AKT/NF-κB signaling axis[11]. Moreover, chronic DEHP exposure in animal models has been linked to sustained proliferation and endocrine disruption in prostate tissue[12], as well as the progression of hepatocellular carcinoma via oxidative stress and MAPK pathway activation[13]. Together, these studies suggest that phthalates can modulate proliferation-associated signaling (e.g., MAPK and PI3K/AKT) and oxidative-stress pathways, which are highly relevant to hormone-responsive cancers. These findings suggest that PAEs may contribute to tumorigenesis through diverse molecular mechanisms and play a pivotal role in malignancies of the female reproductive tract. Among these compounds, diethyl phthalate (DEP) is of particular concern due to its widespread environmental prevalence[14]. Because EC is an estrogen-responsive malignancy, DEP may plausibly affect EC through endocrine-related and redox-related mechanisms. However, the molecular basis of DEP-associated endometrial carcinogenesis remains largely unexplored.

Although DEP has been implicated in female reproductive toxicity, mechanistic evidence specifically linking DEP exposure to endometrial carcinogenesis remains limited, and integrated multi-level investigations in EC are still lacking[15]. Comprehensive, multi-level investigations exploring the mechanistic link between DEP exposure and endometrial carcinogenesis are notably lacking. To address this knowledge gap, the present study adopted an integrative strategy combining network toxicology and bioinformatics to identify candidate genes and signaling pathways potentially mediating DEP-induced endometrial tumorigenesis. Molecular docking and molecular dynamics (MD) simulations were employed to evaluate putative interactions between DEP and key molecular targets. Furthermore, in vitro experiments were conducted to investigate the effects of DEP on cell proliferation, cell cycle progression, and the activation of critical signaling pathways in EC cells. Through this multidimensional and integrative approach, the study aims to systematically elucidate the molecular underpinnings of DEP's involvement in EC. These findings not only provide a theoretical foundation for understanding tumor risks associated with environmental chemical exposure but also offer new insights to foster interdisciplinary research at the intersection of environmental toxicology and oncology.

Methods

Data acquisition and preprocessing

This study analyzed four publicly available GEO transcriptomic cohorts (GSE17025, GSE36389, GSE63678, and GSE115810). These datasets profiled primary human EC tissues and included non-tumor endometrium controls with clear sample annotations (https://www.ncbi.nlm.nih.gov/geo/). Key characteristics of each cohort are summarized in Supplementary Tables S1-2. Raw expression matrices from these datasets were subjected to quantile normalization using the *limma* package in R. To evaluate the effectiveness of normalization, boxplots were generated with *ggplot2*, following data reshaping using the *reshape2* package.

Differential expression analysis

Differential expression analysis was performed using the *limma* package in R. A design matrix was constructed based on sample groupings, followed by linear model fitting and empirical Bayes moderation to estimate gene-wise differential expression statistics. Multiple testing correction was applied using the Benjamini–Hochberg method to control the false discovery rate (FDR). Genes were considered differentially expressed if they met the criteria of $|\log_2 FC| > 0.585$ (equivalent to an absolute fold change > 1.5 on the linear scale, since $\log2(1.5) \approx 0.585$) and FDR < 0.05. Hierarchical clustering and heatmap visualization of the identified differentially expressed genes (DEGs) were conducted using the *pheatmap* package. The overall distribution of gene expression changes was further illustrated using volcano plots, generated with the *ggplot2* package.

Weighted gene co-expression network analysis (WGCNA)

A weighted gene co-expression network was constructed using the *WGCNA* package in R. Initially, low-variance genes were filtered out, and outlier samples were identified and removed based on hierarchical clustering of sample distances. The optimal soft-thresholding power was determined to ensure scale-free topology, after which the adjacency matrix and the corresponding topological overlap matrix (TOM) were computed. Subsequently, hierarchical clustering was applied to the TOM to identify gene modules using a dynamic tree cutting algorithm. Modules with similar expression profiles were merged based on the correlation of their module eigengenes. The association between each module and relevant clinical traits was then assessed. Finally, key modules and hub genes were identified using the metrics of gene significance (GS) and module membership (MM).

Identification of DEP-associated targets

DEP-associated targets were collected from three online resources: ChEMBL (https://www.ebi.ac.uk/chembl/), STITCH (http://stitch.embl.de/), and SwissTargetPrediction (http://www.swisstargetprediction.ch/). In ChEMBL, DEP was searched as a compound and the list of reported compound–target associations was downloaded from the compound page; when target species information was available, only Homo sapiens targets were retained. In STITCH, DEP-related protein targets were retrieved from the chemical–protein interaction output, retaining only interactions with a confidence score > 0.4. In SwissTargetPrediction, predicted targets were obtained from the server output, retaining only targets with probability ≥ 0.1. After harmonizing target identifiers to gene symbols and merging the three lists, duplicate targets were removed to generate a non-redundant DEP target set. Using these criteria, 255 targets were obtained from ChEMBL, 7 from STITCH, and 25 from SwissTargetPrediction, resulting in 287 unique DEP-associated genes. The detailed target lists are provided in Supplementary Table S3.

Candidate key gene selection and network construction

An overlap analysis was performed among the DEGs, genes from WGCNA-identified modules, and predicted targets of DEP. A Venn diagram illustrating the intersecting gene sets was generated using the *ggVennDiagram* package. The intersecting genes between DEGs and WGCNA modules were considered EC-related candidates supported by both differential expression and co-expression evidence. These EC-related genes were further intersected with DEP-predicted targets to obtain DEP-EC overlapping genes. To visualize the relationship between DEP and EC, a DEP-gene-EC interaction network (also referred to as a gene-disease network in this study) was constructed in Cytoscape. The network included three node types: DEP (compound), EC (disease), and the DEP-EC overlapping genes. Edges represented compound–gene links (genes predicted as DEP targets) and gene-disease links (genes identified as EC-related). The resulting network was used for visualization and descriptive network analysis.

Functional enrichment analysis

The shared genes identified from the intersection analysis were first annotated using the *org.Hs.eg.db* package. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were subsequently conducted using the *clusterProfiler* package to investigate associated biological processes and signaling pathways[16]. Multiple testing correction was applied using the Benjamini–Hochberg method, with statistical significance defined as $P < 0.05$ and FDR < 0.05. Enrichment results were visualized using the *enrichplot* and *ggplot2* packages, with bar charts and dot plots employed to display significantly enriched functional categories and pathways.

Machine learning modeling and validation

To evaluate the predictive value of genes associated with DEP, normalized expression matrices from all included GEO cohorts were integrated. Shared genes were extracted, and batch effects were corrected using the ComBat method from the *sva* package. Samples were partitioned into training and testing sets based on their dataset of origin. A total of 113 machine learning algorithm combinations were systematically assessed. These included regularized regression models (LASSO, Ridge, Elastic Net), support vector machine (SVM), random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), partial least squares–generalized linear model (PLS-GLM), stepwise regression (Stepwise GLM), and additional boosting methods such as *glmBoost*. Each algorithm performed feature selection followed by classifier construction based on the selected gene set. External validation was carried out using an independent test set to evaluate the generalizability of the models. Model performance was quantified by the area under the receiver operating characteristic curve (AUC), calculated with the *pROC* package. Comparative performance across models and cohorts was visualized using the *ComplexHeatmap* package. Full details of the machine-learning modeling workflow (including dataset sizes, endpoint definition, feature descriptors, algorithm implementations, and parameter settings) are provided in the Supplementary Method S1.

Model interpretability analysis

To interpret the optimal classifier, Shapley Additive exPlanations (SHAP) were used to quantify the contribution of each gene to the model output (predicted probability). For an individual sample, positive SHAP values indicate contributions that increase the predicted probability toward the EC class, whereas negative values decrease it toward the control class. Global feature importance was summarized using the mean absolute SHAP value (mean|SHAP|). SHAP values were computed using permutation-based SHAP (permshap) and visualized with shapviz. Specifically, the importance bar plot summarizes global feature importance (mean|SHAP|); the beeswarm plot depicts the distribution of SHAP contributions across samples together with feature values; dependence plots display how changes in a feature relate to its SHAP contribution and enable inspection of potential interactions; and force/decision plots illustrate, at the individual-sample level, how multiple features cumulatively shift predictions from the baseline (expected value) to the final output. In addition, single-gene ROC analysis was performed to provide an intuitive reference for the standalone discriminative ability of selected genes.

Molecular Docking

Molecular docking of the selected ligand with target proteins was conducted using the CB-Dock2 online platform (http://cao.labshare.cn/cb-dock2/). CB-Dock2 employs the CurPocket algorithm to automatically detect potential binding pockets within protein structures and utilizes the AutoDock Vina scoring engine to perform docking calculations. Target proteins included the crystal structures of ADRA2C (PDB ID: 6KUW),

NR4A1 (PDB ID: 2QW4), and SLC6A2 (PDB ID: 8HFE), obtained from the Protein Data Bank (PDB, https://www.rcsb.org/), as well as predicted structures of FOS and JUN downloaded from the AlphaFold Protein Structure Database (https://www.alphafold.ebi.ac.uk/). Upon uploading the protein structures to CB-Dock2, automated preprocessing was performed, including the removal of water molecules and hetero-ligands and the addition of hydrogen atoms, followed by identification of the top five binding pockets ranked by docking score. The three-dimensional structure of the ligand DEP was retrieved from the PubChem database (https://pubchem.ncbi.nlm.nih.gov/), converted into mol2 format, and uploaded to the platform to enhance docking precision. Multiple docking simulations were executed for each binding pocket, and the corresponding binding energies (kcal/mol) and three-dimensional conformations were recorded. The conformation exhibiting the lowest binding energy was selected for subsequent analysis. Docking results and protein–ligand interactions were visualized using PyMOL.

MD simulation

To further assess the stability and dynamic behavior of the protein–ligand complexes, all-atom MD simulations were performed for 100 ns using GROMACS. The AMBER99SB-ILDN force field was employed for proteins, and ligand parameters were generated using the GAFF force field, with restrained electrostatic potential (RESP) charges calculated via Gaussian 16W. Each system was solvated in a TIP3P water model, and Na^+ ions were added to neutralize the net charge. Energy minimization was conducted using the steepest descent algorithm. Long-range electrostatic interactions were treated with the particle mesh Ewald (PME) method, and bond constraints involving hydrogen atoms were managed using the LINCS algorithm. Simulations were carried out in the NPT ensemble at 300 K and 1 bar, with a 2 fs time step. Trajectories were recorded every 10 ps for subsequent analysis. Post-simulation evaluations included root mean square deviation (RMSD), radius of gyration (Rg), solvent-accessible surface area (SASA), root mean square fluctuation (RMSF), hydrogen bond analysis, and Gibbs free energy landscape (FEL) profiling, to comprehensively evaluate the conformational stability and dynamic properties of the complexes.

Cell culture

The human EC cell lines HEC-1-A and Ishikawa were obtained from GeneChem Gene Technology Co., Ltd. (Shanghai, China). The cell line authentication certificates provided by the supplier are included in Supplementary Methods S2. Ishikawa cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Hyclone, Logan, UT, USA) supplemented with 10% fetal bovine serum (FBS) (Siji Green, China) and 1% penicillin–streptomycin (Servicebio, Wuhan, China). HEC-1-A cells were cultured in McCoy's 5A medium (Procell, China) supplemented with 10% FBS and 1% penicillin–streptomycin. All cells were incubated at 37 °C in a humidified atmosphere containing 5% CO_2.

Drug treatment

DEP (MCE, HY-Y0284) was dissolved in dimethyl sulfoxide (DMSO) to prepare a 1 mM stock solution, which was stored at -20 °C in the dark. Prior to use, a 100 × intermediate working solution containing 10% DMSO was prepared and subsequently diluted into the culture medium at a 1:100 ratio, yielding a final DMSO concentration of 0.1% across all treatment groups. We first performed a range-finding dose–response experiment to identify biologically active, non-cytotoxic concentrations. Cells were treated with DEP (0.1 nM–100 μM) for 24 h to generate dose–response curves. Subsequent mechanistic assays were conducted only at the selected concentrations identified in this screening. For mechanistic assays, three concentrations-10 μM, 1 μM, and 10 nM-were selected based on preliminary findings. All experiments included a 0.1% DMSO vehicle control and were independently performed in triplicate or more.

Cell Viability Assay

The effect of DEP on EC cell viability was evaluated using the Cell Counting Kit-8 (CCK-8) assay. Cells were seeded at a density of 5×10^3 cells/well in 96-well plates and allowed to adhere for 24 h. Subsequently, they were treated with various concentrations of DEP for an additional 24 h. At 0, 24, and 48 h post-treatment, CCK-8 reagent (BioScience, China) was added to each well, followed by incubation at 37 °C in the dark for 2 h. Cell viability was assessed by measuring absorbance at 450 nm using a microplate reader (Thermo Scientific, USA).

EdU Assay

Cell proliferation was assessed using a 5-ethynyl-2′-deoxyuridine (EdU) incorporation assay. Cells were seeded into 96-well plates and incubated for 24 h, followed by treatment with various concentrations of DEP for an additional 24 h. The culture medium was replaced with EdU working solution (10 μM, Abbkine, KTA2031). Cells were incubated at 37 °C for 2 h protected from light. Following incubation, cells were fixed with 4% paraformaldehyde and permeabilized with 0.5% Triton X-100. The EdU reaction solution was applied for 30 min, after which cell nuclei were counterstained with Hoechst 33342. Imaging was performed using a fluorescence microscope equipped with a 20 × objective lens, and the percentage of EdU-positive cells was quantified using ImageJ software.

Cell cycle analysis

Following treatment, cells were processed in accordance with the manufacturer's protocol. A PI (propidium iodide)-based DNA staining buffer containing RNase A (Lianke Biotech, CCS012) was added, and cells were incubated at room temperature in the dark for 30 min. Cell cycle distribution was analyzed using a NovoCyte flow cytometer (Agilent Technologies, USA), and the proportions of cells in the G0/G1, S, and G2/M phases were quantified using *NovoExpress* analytical software.

Western Blotting (WB)

Total cellular protein was extracted using RIPA lysis buffer (Solarbio, Beijing, China) supplemented with phenylmethylsulfonyl fluoride (PMSF) and a protease inhibitor cocktail. Protein concentrations were quantified using the bicinchoninic acid (BCA) assay. A total of 15 μg protein was loaded per lane for SDS-PAGE and transferred onto PVDF membranes (Millipore, USA). Membranes were blocked with 3% bovine serum albumin (BSA) for 2 h at room temperature, followed by overnight incubation at 4 °C with primary antibodies targeting ERK1/2, phospho-ERK1/2 (p-ERK1/2), PI3K, p-PI3K, AKT, p-AKT, GSK3β, p-GSK3β, Rb, p-Rb, CDK4, Cyclin D1, β-Actin, and GAPDH. After washing, membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 1 h at room temperature. Protein bands were visualized using enhanced chemiluminescence (ECL) and subjected to densitometric analysis using ImageJ software. Detailed information on all antibodies used is provided in Supplementary Table S4.

Reactive Oxygen Species (ROS) Detection

Intracellular ROS levels were measured using a dihydroethidium (DHE)-based ROS detection kit (Pulilai, China). After 24-h treatment with DEP at concentrations of 10 nM, 1 μM, and 100 μM, cells were incubated with 5 μM DHE working solution at 37 °C in the dark for 30 min. DHE fluorescence was assessed by fluorescence microscopy and flow cytometry. Mean fluorescence intensity (MFI) was quantified using ImageJ to evaluate intracellular superoxide-related oxidative signals.'[

Superoxide Dismutase (SOD) Detection

SOD activity was assessed using a WST-1-based SOD assay kit (Jiancheng, Nanjing, China). Following DEP exposure, cells were washed with PBS, lysed, and centrifuged at 12,000×*g* for 10 min at 4 °C. The resulting supernatant was collected for analysis. Protein concentrations were determined using the bicinchoninic acid (BCA) assay. SOD activity was measured according to the manufacturer's protocol, and absorbance was recorded at 450 nm. Results were expressed as enzyme activity (U) per milligram of protein. All experiments were performed independently in triplicate, with technical duplicates included for each condition.

Statistical analysis

Bioinformatics analyses were performed using R software (version 4.2.0). Experimental data were analyzed with GraphPad Prism (versions 8.0 and 9.0). Two-tailed Student's t-tests were used for comparisons between two groups, and one-way ANOVA followed by Tukey's post hoc test was applied for multiple-group comparisons. Unless otherwise specified, $P < 0.05$ was considered statistically significant. Statistical significance is indicated as ns (not significant), $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

Ethical approval and compliance with guidelines

This study comprised bioinformatics analyses of publicly available datasets and in vitro experiments using established human cell lines only (no human tissues were used). The bioinformatics analyses utilized open-access datasets (e.g., TCGA, GEO), which do not require additional ethical approval. All experimental procedures involving human cell lines were reviewed and approved by the Ethics Committee of The First Hospital of Lanzhou University (Approval No. LDYYLL2024-735). All methods were carried out in accordance with the relevant institutional and national guidelines and regulations.

Results

Differential expression characteristics of EC-related genes

An overview of the study design is presented in Fig. 1. Following normalization of the four GEO datasets, gene expression profiles were found to be generally consistent across samples (Figures S1A–B). Prior to batch effect correction, PCA revealed clear separation among the different cohorts (Figure S1C). However, after applying the ComBat method, the samples exhibited substantial overlap in the PCA space (Figure S1D), indicating successful elimination of batch effects. In the integrated transcriptomic dataset, a total of 1583 DEGs were identified, including 787 downregulated and 796 upregulated genes. The heatmap demonstrated that these DEGs effectively distinguished tumor samples from controls, while the volcano plot visually depicted the global distribution of upregulated and downregulated genes (Fig. 2A,B).

Co-expression modules associated with EC

A weighted gene co-expression network was constructed based on the integrated expression matrix. Hierarchical clustering of samples revealed no significant outliers (Figure S2A). Soft-thresholding analysis determined that a power value of $\beta = 6$ best satisfied the scale-free topology criterion (Figure S2B). A total of 11 stable co-expression modules were subsequently identified (Fig. 2C). Correlation analysis between module eigengenes and clinical traits showed that the blue ($r = 0.46$, $P = 8 \times 10^{-10}$), green ($r = 0.50$, $P = 8 \times 10^{-12}$), and purple ($r = 0.38$, $P = 5 \times 10^{-7}$) modules were significantly associated with the ECphenotype (Fig. 2D, Figure S2C), suggesting that genes within these modules may play important roles in the initiation and progression of EC.

Identification of DEP-associated candidate genes

A total of 645 EC-related candidate genes were identified by intersecting the DEGs (n = 1583) with genes from the WGCNA co-expression modules (n = 2341), yielding 645 shared genes (Fig. 3A). Subsequently, predicted DEP targets (n = 194) were integrated with the EC-related candidate gene set (n = 645), resulting in 19 overlapping DEP-EC genes (Fig. 3B). These overlapping genes were considered potential mediators linking DEP exposure to EC. In the DEP-gene-EC interaction network, JUN, STAT3, FOS, and KLF5 showed higher connectivity and occupied central positions (Fig. 3C), highlighting them as key nodes for downstream prioritization.

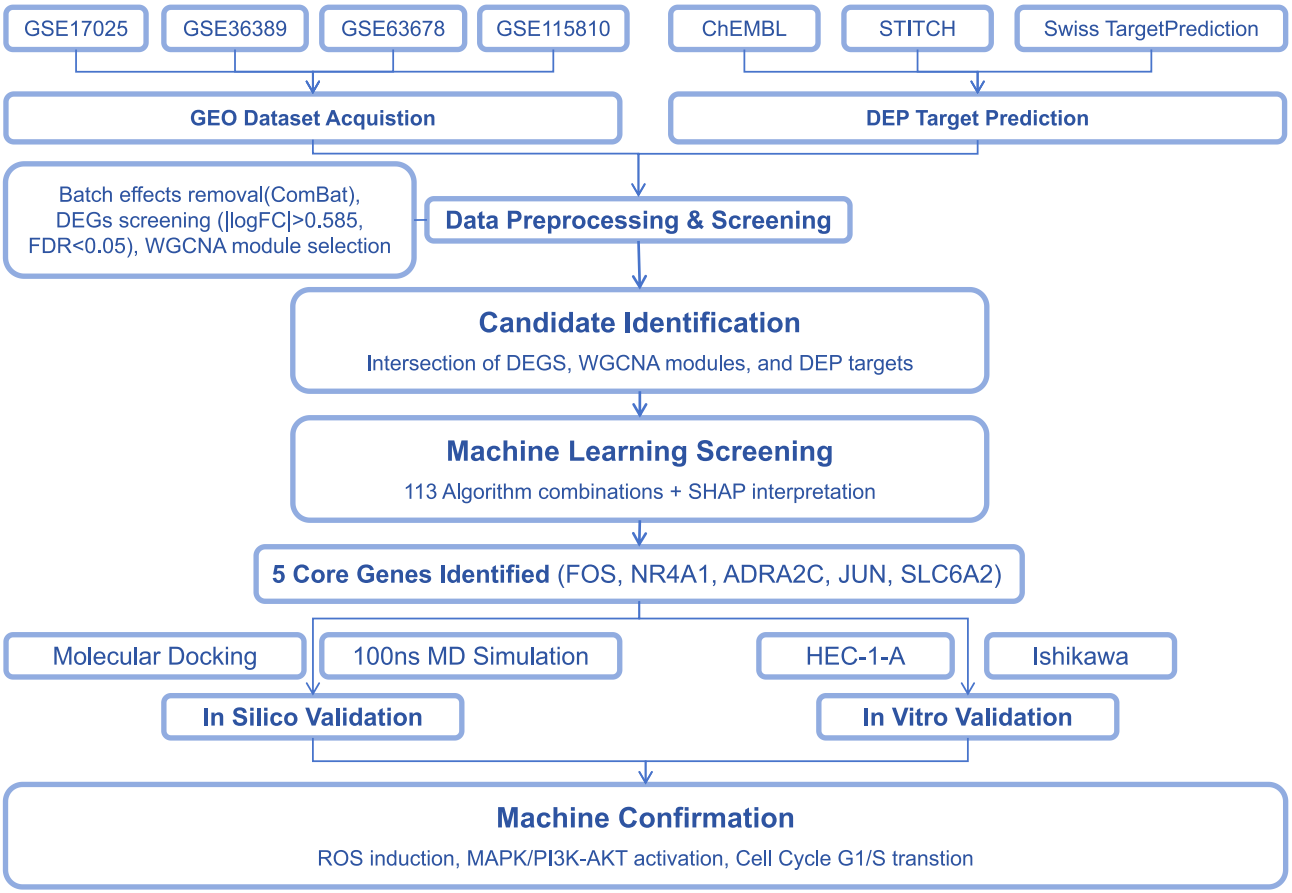

Fig. 1. Overall workflow and major findings of this study.

Functional enrichment analysis of DEP-EC candidate genes

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted on the intersecting DEP–EC gene set. GO analysis revealed significant enrichment in biological processes (BP) such as miRNA transcriptional regulation, skeletal muscle tissue and organ development, cell differentiation, and cell cycle progression. At the cellular component (CC) level, the genes were associated with the RNA polymerase II transcription regulator complex, endoplasmic reticulum–Golgi intermediate compartment, calcium channel complexes, and the spindle apparatus. For molecular functions (MF), they were enriched in DNA-binding transcription factor activity, nuclear receptor activity, SMAD binding, and phosphatase binding (Fig. 4A,B). KEGG pathway analysis further indicated significant enrichment in the MAPK signaling pathway, cAMP and cGMP–PKG signaling pathways, chemical carcinogenesis-related pathways, and the apoptosis pathway (Fig. 4C,D). Collectively, these findings suggest that DEP may contribute to the initiation and progression of EC by modulating multiple signaling cascades and transcription factor regulatory networks.

Machine learning-based identification of DEP targets and key genes

To systematically evaluate the predictive value of DEP-associated candidate genes in EC, a total of 113 machine-learning algorithm combinations were assessed. Across these models, RF/GBM-based approaches (including ensemble strategies) achieved the highest overall performance, with AUCs approaching or exceeding 0.90 in both the training set and independent validation datasets (Fig. 5A), where models were ranked by their mean AUC across datasets (blue bar). SHAP analysis of the optimal model identified FOS, NR4A1, ADRA2C, JUN, and SLC6A2 as the top five contributors, with FOS showing the highest mean absolute SHAP value (mean|SHAP|) (Fig. 5B,C). SHAP importance was calculated for the complete set of 15 genes retained in the optimal model (Fig. 5B). Dependence analysis suggested potential feature interactions, including FOS–ADRA2C and STAT3–KIF11 (Fig. 5D). At the individual-sample level, representative SHAP plots for an example correctly classified sample (chosen as the correctly classified sample with predicted probability closest to the decision threshold) further supported the contribution patterns of the top features (Fig. 5E,F). In addition, single-gene ROC analyses were performed for the top five SHAP-ranked genes (FOS, NR4A1, ADRA2C, JUN, and SLC6A2), each achieving an AUC > 0.79 (Fig. 5G). Collectively, these findings highlight FOS, NR4A1, ADRA2C, JUN, and SLC6A2 as key DEP-associated signatures contributing to EC classification and prioritize them for downstream mechanistic investigation.

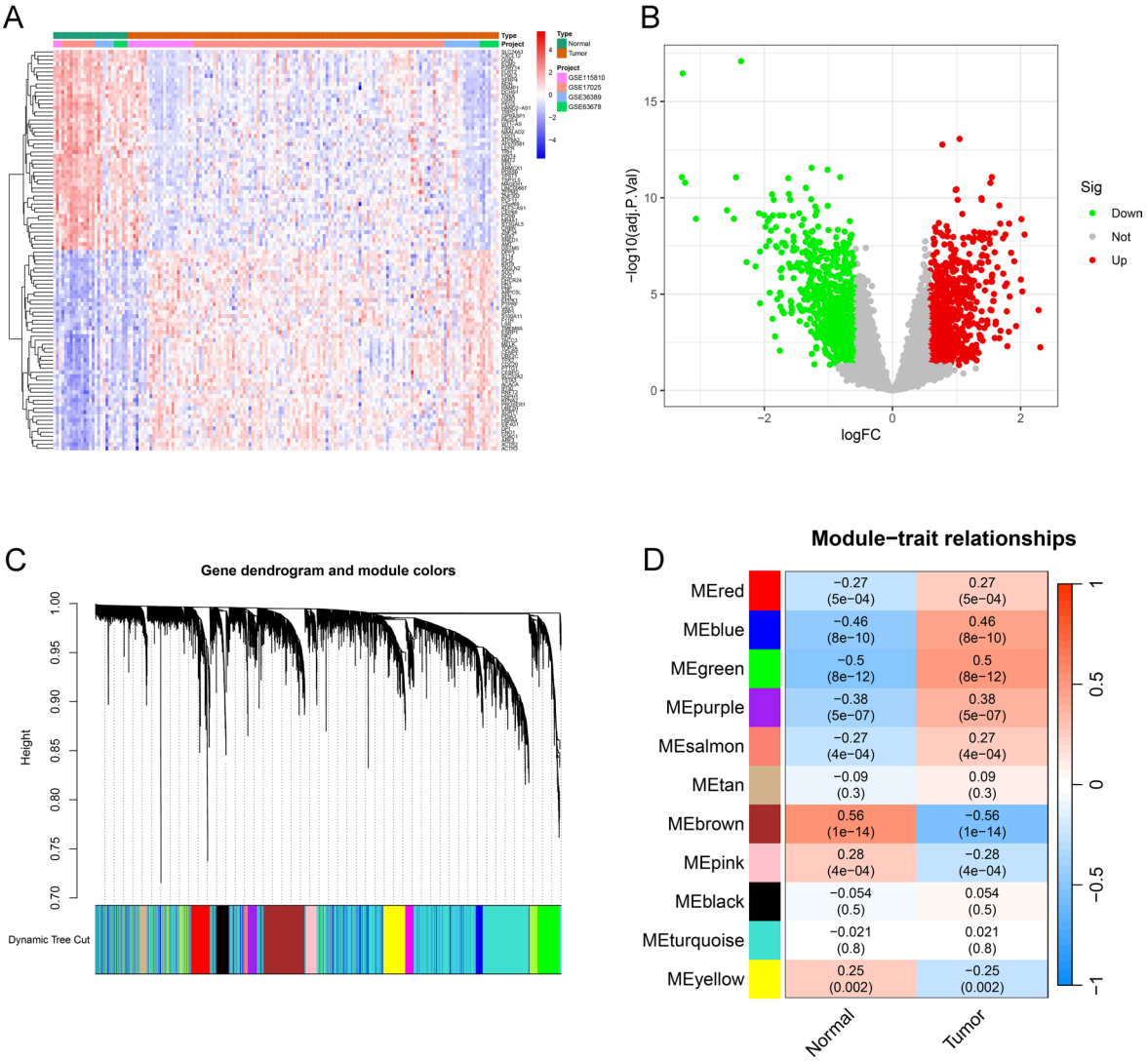

Fig. 2. Identification of candidate genes associated with EC. (**A**) Heatmap of DEGs between endometrial EC and normal controls. (**B**) Volcano plot of DEGs showing significantly upregulated (red) and downregulated (green) genes. (**C**) WGCNA identified 11 distinct gene modules, each represented by a different color. (**D**) Module–trait correlation heatmap demonstrating that the blue (r = 0.46, $P = 8 \times 10^{-10}$), green (r = 0.50, $P = 8 \times 10^{-12}$), and purple (r = 0.38, $P = 5 \times 10^{-7}$) modules were positively associated with EC.

Binding modes and stability of DEP with key target proteins

Molecular docking suggested that DEP could be accommodated by all five targets, with more favorable docking scores observed for ADRA2C (− 6.0 kcal/mol), NR4A1 (− 5.8 kcal/mol), and SLC6A2 (− 6.8 kcal/mol) than for FOS (− 4.0 kcal/mol) and JUN (− 4.6 kcal/mol) (Fig. 6A–E). Hydrogen bond analysis indicated polar contacts between DEP and ADRA2C (TYR327, SER108), NR4A1 (HIS41), FOS (CYS50, ASP52, THR51), and JUN (GLN329, GLN327, GLN326). In contrast, no hydrogen bonds were detected in the SLC6A2 complex, suggesting that this interaction may be dominated by hydrophobic and van der Waals forces (Fig. 6A–E).

MD simulations provided additional support for these observations. The SLC6A2, ADRA2C and NR4A1 complexes exhibited rapid conformational adaptation during the initial simulation phase, after which RMSD values stabilized at relatively low levels (~ 0.25–0.6 nm). Both the Rg and SASA decreased and reached steady states (SLC6A2: Rg ≈ 2.50 nm, SASA: 250 → 230 nm^2; ADRA2C: Rg ≈ 3.10 nm, SASA: 250 → 220 nm^2; NR4A1: Rg ≈ 1.81 nm, SASA: 128 → 115–120 nm^2), indicating enhanced structural compactness and complex stability. RMSF remained below 0.3 nm for the majority of residues, except for slight increases at terminal or loop regions. The number of hydrogen bonds was maintained at 1–2, increasing to 2–3 in ADRA2C during later simulation stages. FEL analysis revealed a single deep energy basin for each complex, supporting relatively stable conformational states. By contrast, the FOS and JUN complexes underwent marked conformational rearrangements during the early simulation period. RMSD values converged at ~ 4 nm for FOS and ~ 3 nm for JUN, accompanied by a rapid Rg decline from > 4 to ~ 3 nm. SASA decreased significantly (FOS: ~ 600 → 220–250 nm^2; JUN: ~ 400 → 200–220 nm^2), while RMSF values increased notably (0.5–2.0 nm), indicating elevated

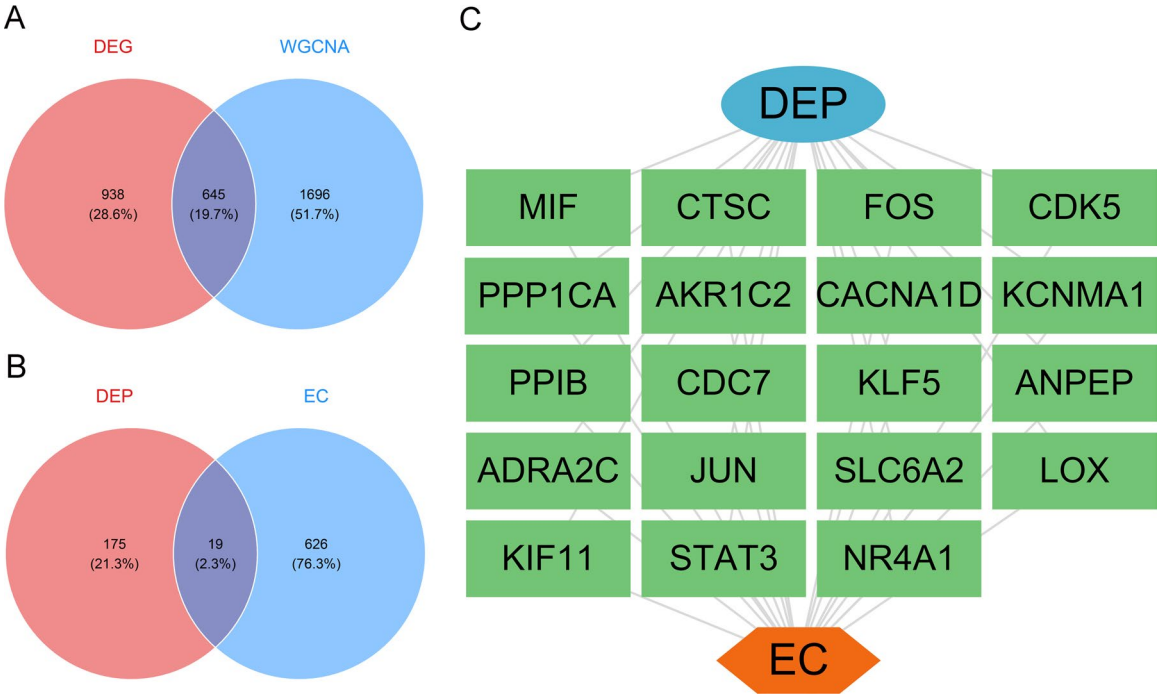

Fig. 3. Identification of DEP-related candidate targets in EC. (**A**) Venn diagram showing the overlap between DEGs (n = 1583) and WGCNA modules (n = 2341), yielding 645 intersecting genes. (**B**) Venn diagram of potential targets of DEP (n = 194) and EC-related candidate genes (n = 645), identifying 19 overlapping genes. (**C**) Network diagram illustrating the shared candidate genes linking DEP exposure to EC, including ADRA2C, JUN, NR4A1, SLC6A2, STAT3, and others.

residue flexibility. Hydrogen bond numbers fluctuated between 1 and 3 in both complexes, with JUN occasionally reaching up to 4. FEL analysis showed multiple shallow basins in FOS, and a narrow deep basin with surrounding shallow regions in JUN, collectively indicating lower conformational stability compared to ADRA2C, NR4A1, and SLC6A2 (Fig. 7A–F). Together, these results suggest that DEP forms more stable and compact complexes with ADRA2C, NR4A1, and SLC6A2, while its interactions with FOS and JUN are comparatively weaker and characterized by greater conformational flexibility and dynamic fluctuations.

DEP enhances EC cell proliferation and induces oxidative stress

Guided by our in silico analyses, DEP–EC overlapping genes were enriched in proliferation-associated signaling pathways, and the SHAP-prioritized core genes are functionally linked to stress-responsive transcription and oxidative redox regulation. Therefore, we first evaluated whether DEP exposure promotes proliferative phenotypes in EC cells and whether oxidative stress is induced as a plausible upstream driver. CCK-8 assays demonstrated that DEP significantly enhanced the proliferative capacity of both HEC-1-A and Ishikawa EC cell lines (Fig. 8A,B). We initially conducted a broad range-finding dose–response experiment (0.1 nM–100 μM) to define a biologically active and non-cytotoxic window. Based on the dose–response profiles and supported by prior in vitro studies, the most commonly used effective concentrations of DEP in tumor and reproductive cell models are within 10 nM–10 μM[17,18]. Therefore, 10 μM, 1 μM, and 10 nM were selected as representative doses for subsequent mechanistic investigations. EdU incorporation assays further confirmed a significant increase in nuclear fluorescence intensity following DEP exposure, indicating enhanced DNA synthesis activity (Fig. 8C–E). Concurrently, DEP treatment resulted in a marked increase in intracellular ROS levels (Fig. 8F) and a reduction in SOD activity in both cell lines (Figure S3), suggesting that oxidative stress contributes to the proliferative effects induced by DEP.

DEP activates the MAPK and PI3K/AKT signaling pathway and promotes G1/S phase progression in EC cells

Flow cytometric analysis revealed that DEP exposure significantly reduced the proportion of HEC-1-A and Ishikawa cells in the G0/G1 phase, while increasing the proportion of cells in the S phase, indicating that DEP promotes G1/S phase cell cycle progression in EC cells (Fig. 9A,B). At the molecular level, WB analysis demonstrated that DEP treatment induced phosphorylation of ERK, which was accompanied by elevated phosphorylation of PI3K and AKT, signifying activation of the ERK/PI3K–AKT signaling axis. Additionally, phosphorylation of GSK3β was markedly increased, along with upregulated expression of Cyclin D1 and CDK4. These changes suggest that DEP-mediated inactivation of GSK3β may stabilize and promote the accumulation of Cyclin D1 and CDK4, thereby facilitating the G1 to S phase transition (Fig. 9C,D).

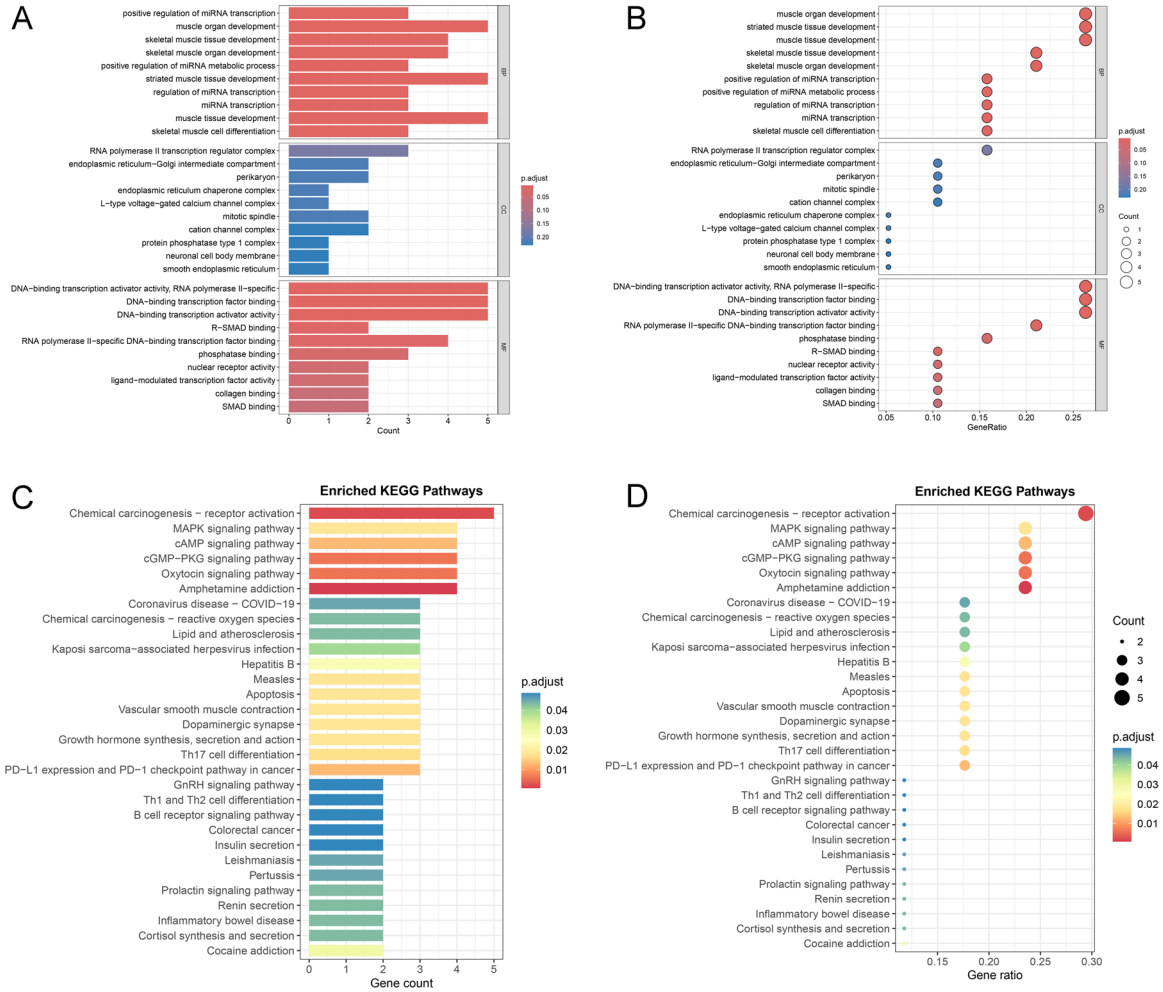

Fig. 4. Functional enrichment analysis of DEP-related candidate genes. (**A**, **B**) GO enrichment analysis showing significant BP, CC, and MF of candidate genes. (**C**, **D**) KEGG pathway analysis highlighting signaling pathways potentially associated with DEP-induced EC. KEGG materials are used under license from Kanehisa Laboratories for Open Access publication in Scientific Reports.© Kanehisa Laboratories.

Discussion

DEP is one of the most extensively used PAE plasticizers. Its metabolite, monoethyl phthalate (MEP), is detected in the urine of more than 95% of individuals[19], indicating that DEP is a pervasive environmental contaminant. Although its estrogenic potency is relatively weak, mounting evidence suggests that DEP can alter cellular behavior by disrupting multiple receptors and signaling pathways. For instance, Fiocchetti et al. reported that DEP activates estrogen receptors (ER), triggering downstream gene transcription and thereby promoting cell proliferation and oxidative stress[20]. Similarly, Gopalakrishnan et al. demonstrated that low-dose DEP exposure induces aberrant expression of multiple genes in mammary tissue, with substantial overlap observed between dysregulated genes in preneoplastic and tumor tissues[21]. These findings imply that, while DEP is not generally considered a classical genotoxic carcinogen, it may facilitate the progression of hormone-responsive cancers. In EC, emerging epidemiological data have associated DEP exposure with disease advancement. A case–control study by Knez et al. observed that increased urinary levels of DEP and other low molecular weight PAEs were significantly correlated with enhanced EC invasiveness[22]. However, these studies primarily establish correlations and offer limited insights into the underlying molecular mechanisms. Moreover, existing experimental research has predominantly focused on PAEs such as DEHP and DBP, with limited attention given to DEP. To bridge this knowledge gap, the present study integrated network toxicology and bioinformatics to systematically identify DEP-related candidate genes and signaling pathways involved in EC. Molecular docking and MD simulations further confirmed stable interactions between DEP and key factors, including FOS, NR4A1, ADRA2C, JUN, and SLC6A2. These findings were substantiated by in vitro experiments, which demonstrated that DEP promotes cell proliferation and modulates critical signaling pathways, thereby offering novel mechanistic insights into its potential role in EC pathogenesis.

Our functional enrichment analysis showed that the 19 DEP-EC overlapping genes were enriched in multiple signaling pathways, including MAPK, cAMP, and cGMP-PKG, as well as tumor-associated pathways such as chemical carcinogenesis. These pathways have been implicated in cell proliferation and cell-cycle regulation

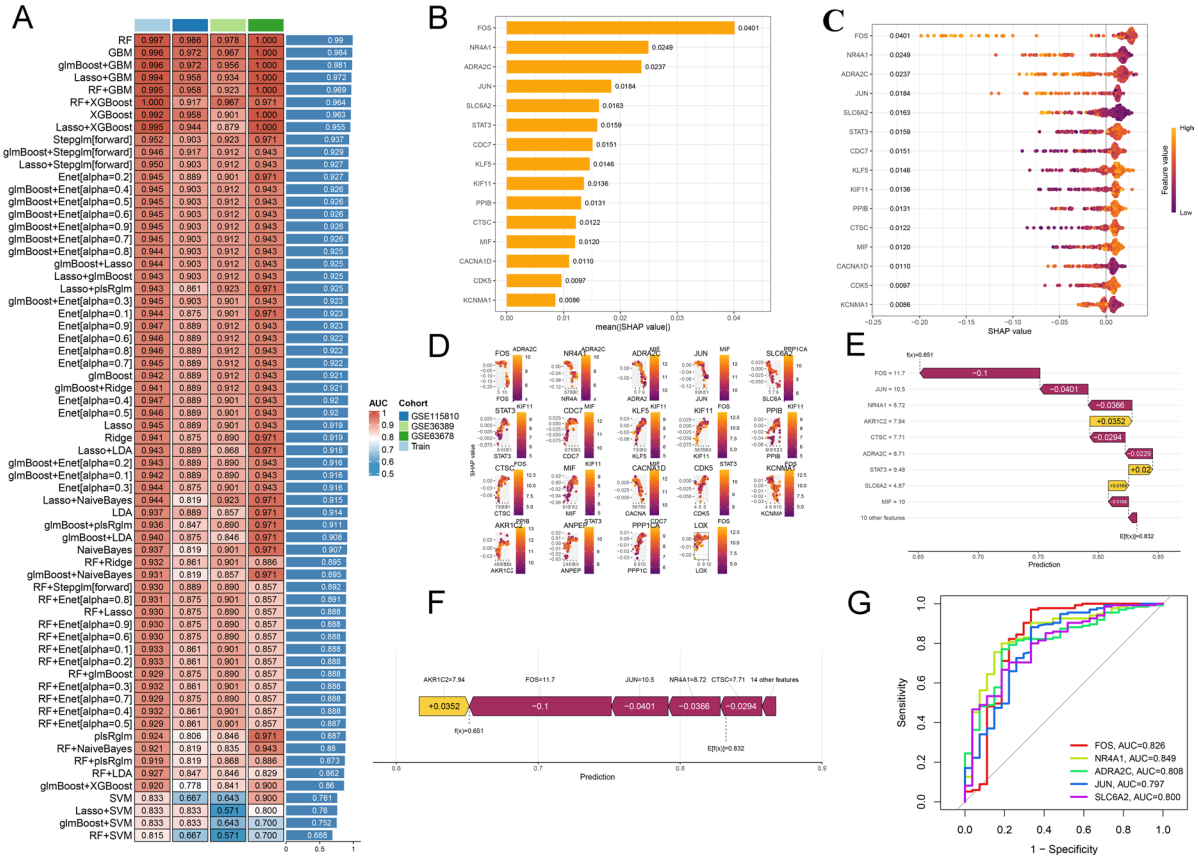

Fig. 5. Core gene screening and model validation of DEP-associated targets in EC. (**A**) Heatmap of AUC values for different machine learning models across multiple cohorts. (**B**) Bar plot of mean SHAP values, ranking candidate genes by their average contribution to the predictive model. (**C**) SHAP summary (beeswarm) plot showing feature importance distributions and the relationship between gene expression levels and SHAP values. (**D**) SHAP dependency plots illustrating potential gene–gene interactions and cooperative effects. (**E**, **F**) SHAP force plots (decision plots) displaying the impact of core genes on model prediction at the individual sample level. (**G**) ROC curves of representative core genes (FOS, NR4A1, ADRA2C, JUN, SLC6A2), demonstrating their diagnostic performance.

in EC and other hormone-responsive malignancies. The MAPK/ERK pathway is a well-established regulator of cell-cycle progression, and its downstream transcriptional programs can increase Cyclin D1 expression and facilitate the G1/S transition[23,24]. In EC tissues, AP-1 family members (c-Fos, Fra-2, and JunB) have been reported to correlate with cell-cycle regulators such as Cyclin D1 and CDK4[25], supporting a potential link between MAPK activity and cell-cycle control. In our study, DEP exposure was accompanied by increased ERK1/2 phosphorylation and upregulation of Cyclin D1 and CDK4, which is consistent with activation of a pro-proliferative signaling state. Notably, although KEGG enrichment did not directly prioritize the PI3K-AKT pathway, we observed increased AKT phosphorylation, suggesting that PI3K/AKT signaling may also be engaged under DEP exposure. PI3K-AKT signaling is frequently dysregulated in type I EC and has been associated with tumor grade and prognosis[26]. Estrogen has also been shown to activate PI3K/AKT via GPR30, leading to Cyclin D1 upregulation and sustained EC cell proliferation[27]. Together, these findings raise the possibility that DEP may modulate MAPK and PI3K/AKT signaling in parallel, thereby favoring Cyclin D1/CDK4 complex formation and cell-cycle progression. cAMP/PKA and cGMP/PKG pathways are known to exhibit crosstalk with MAPK signaling, and their effects can be context dependent[28–30]. Given that DEP increased intracellular ROS and reduced SOD activity in our experiments, oxidative stress may act as an upstream modulator that amplifies these signaling networks. However, we acknowledge that the current data primarily support associations between DEP exposure and pathway activation. The functional necessity of ERK/AKT signaling and ROS in mediating DEP-induced proliferation and G1/S progression remains to be determined. Future studies using pathway inhibition (e.g., MEK/ERK or PI3K/AKT inhibitors) and ROS-scavenging approaches (e.g., NAC) will be required to establish causality.

For candidate gene prioritization, this study employed machine learning combined with SHAP-based interpretability analysis to identify five core genes-FOS, JUN, SLC6A2, NR4A1, and ADRA2C-from the pool of DEP-EC-associated genes. These genes showed the highest contribution scores in predictive modeling and exhibited good discriminatory performance in single-gene ROC analyses. Biologically, FOS and JUN are core components of the AP-1 transcription factor complex and are involved in cell-cycle-related transcriptional

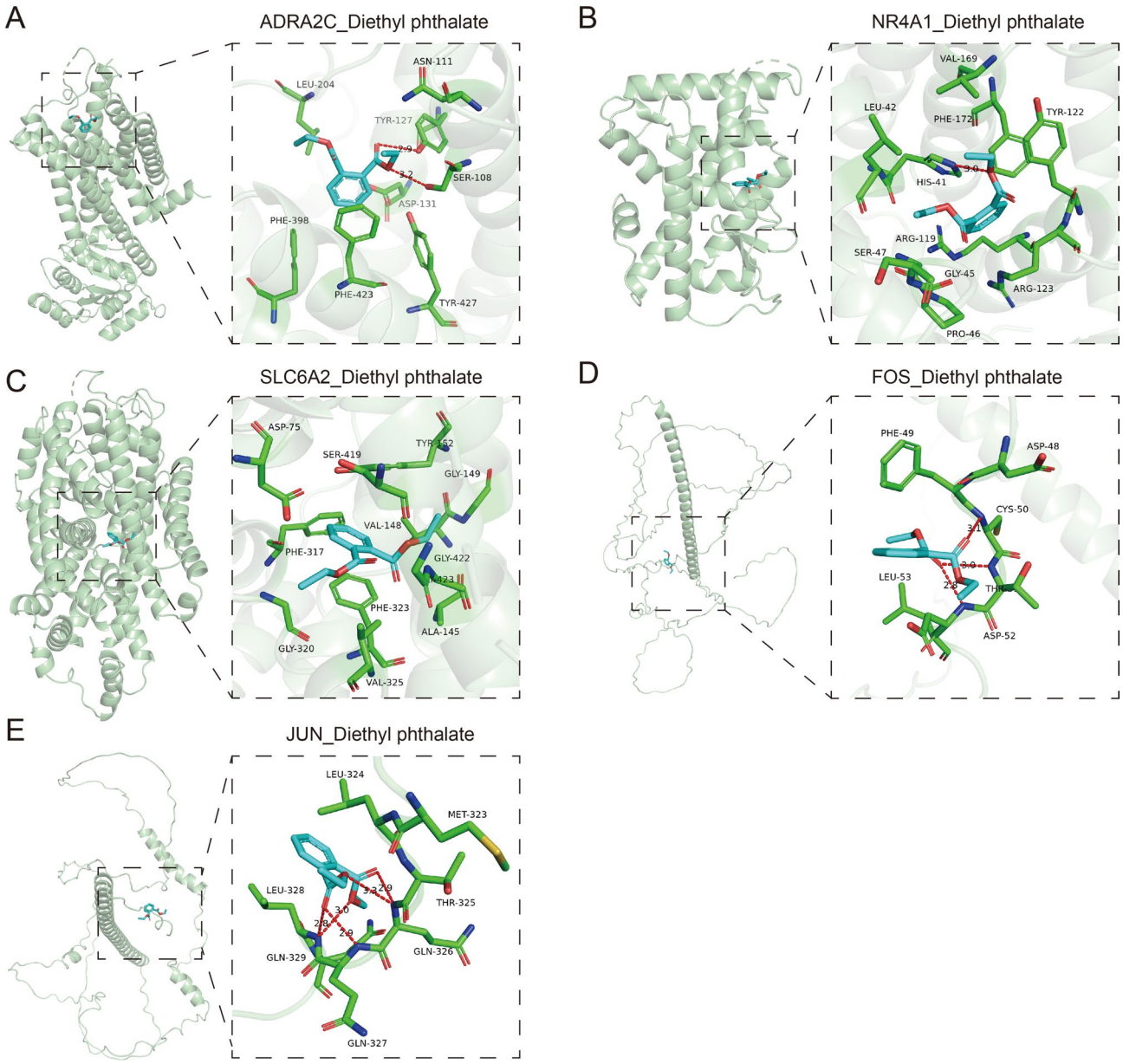

Fig. 6. Molecular docking results of DEP with core target proteins. The docking modes of DEP with (**A**) ADRA2C, (**B**) NR4A1, (**C**) SLC6A2, (**D**) FOS, and (**E**) JUN are shown. The binding pockets and major hydrogen bond or hydrophobic interactions are indicated.

programs[31]. SLC6A2 encodes the norepinephrine transporter and is linked to adrenergic signaling, which has been associated with stress-related tumor progression, although its role in EC remains underexplored[32]. NR4A1, an orphan nuclear receptor, has been implicated in oxidative stress responses and metabolic adaptation in tumor cells[33]. ADRA2C, an α2-adrenergic receptor subtype connected to cAMP/PKA signaling, has rarely been investigated in EC, suggesting potential relevance in DEP-related signaling regulation. To further explore molecular plausibility at the structural level, molecular docking and 100 ns MD simulations suggested putative binding between DEP and these targets, with relatively stable interaction patterns observed for several complexes. To improve biological interpretability, we further mapped the predicted binding poses to known functional regions of each target. For ADRA2C, DEP was predicted to occupy a small-molecule–accessible cavity within the receptor structure, consistent with the canonical ligand-binding region of adrenergic GPCRs. For NR4A1, docking was performed on the ligand-binding region, and the predicted contacts clustered within/adjacent to the putative pocket. For SLC6A2, DEP was predicted to bind within a transporter-accessible cavity implicated in substrate/inhibitor recognition. In contrast, FOS/JUN are AP-1 transcription factors whose DNA-binding and dimerization functions are mediated by the bZIP domain; therefore, whether DEP binding overlaps with this functional region remains uncertain, and the FOS/JUN docking results should be interpreted cautiously.

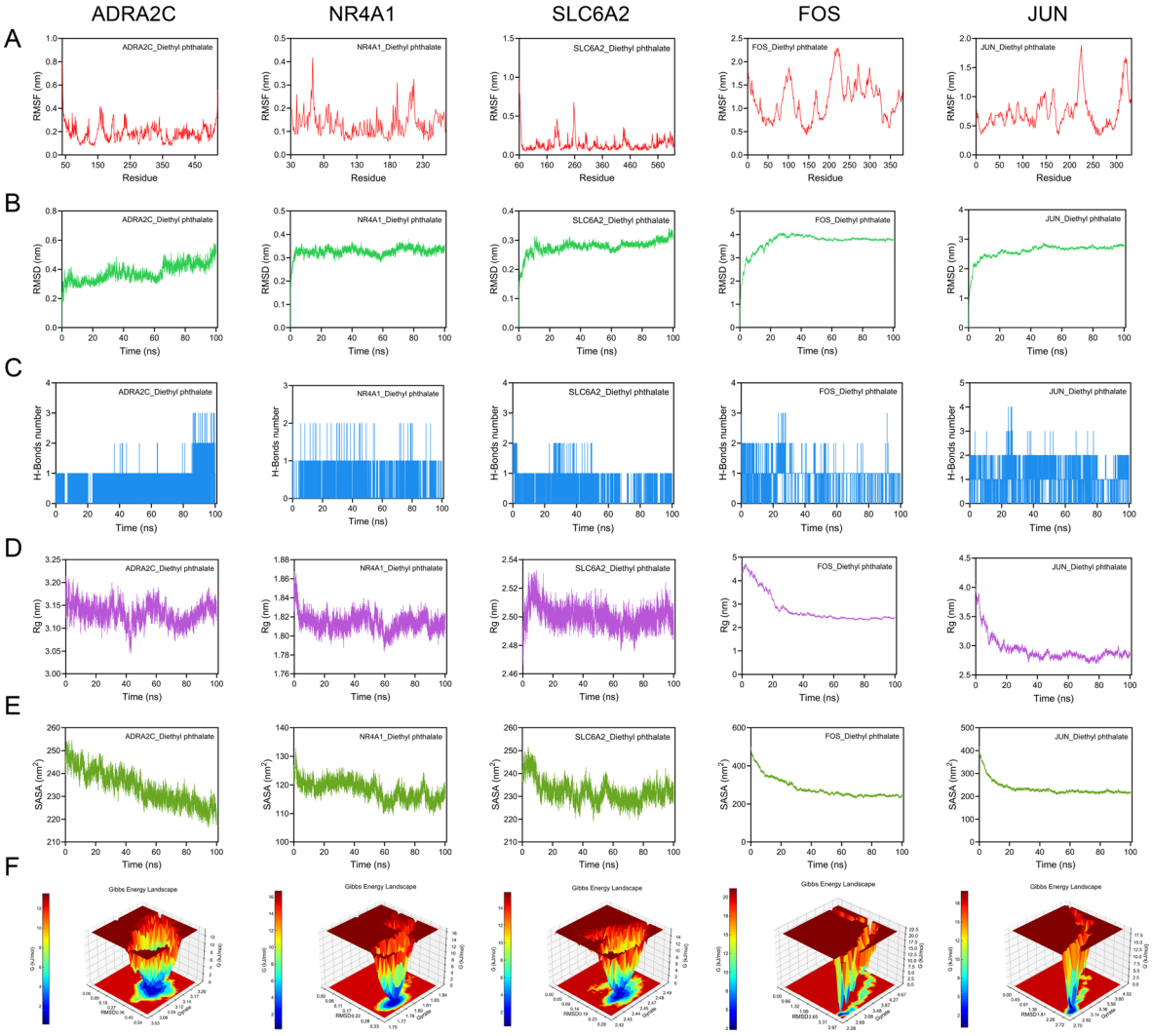

Fig. 7. Molecular dynamics simulation of DEP with five core target proteins. The 100-ns MD simulations of DEP binding to ADRA2C, NR4A1, SLC6A2, FOS, and JUN were evaluated. (**A**) RMSF, (**B**) RMSD, (**C**) hydrogen bond number, (**D**) Rg, (**E**) SASA, and (**F**) FEL are shown for each complex.

Notably, the integration of machine learning and SHAP has been utilized in previous environmental toxicology studies. For instance, Xia et al. identified key toxicological biomarkers in cadmium exposure[34], while Li et al. applied this approach to explore heavy metal–related cardiovascular diseases[35]. However, such studies typically remain at the "exposure–phenotype" level, without mechanistic molecular validation. The innovation of the present study lies in the integration of machine learning-based gene prioritization with molecular simulations and cellular experiments, enabling cross-level validation. This approach establishes a comprehensive investigative framework that spans exogenous chemical exposure, mechanistic elucidation at the molecular level, and phenotypic validation in cell models.

In conclusion, this study comprehensively elucidates the potential molecular associations between DEP exposure and EC. Our findings suggest that DEP promotes EC cell proliferation by inducing ROS accumulation, enhancing the phosphorylation of ERK1/2 and AKT, upregulating Cyclin D1/CDK4 expression, and accelerating G1/S phase transition. These results indicate that DEP may contribute to tumorigenesis through oxidative stress–mediated crosstalk between key signaling pathways, including MAPK and PI3K/AKT. Compared with previous research, this study not only provides novel molecular evidence supporting the tumor-promoting effects of DEP but also establishes a multilayered investigative framework that integrates network-based screening, machine learning, and experimental validation. This strategy addresses a longstanding gap in the mechanistic understanding of DEP's toxicological impact. However, several limitations should be acknowledged. This study lacks validation in animal models and clinical specimens and does not account for the complexity of environmental co-exposures. Future research should incorporate epidemiological cohorts, in vivo models, and multi-omics analyses to further validate and expand upon these mechanistic insights. Moreover, investigating the combined effects of DEP with other EDCs may provide deeper insights into their potential synergistic roles in EC pathogenesis, ultimately contributing to more effective strategies for disease prevention and control.

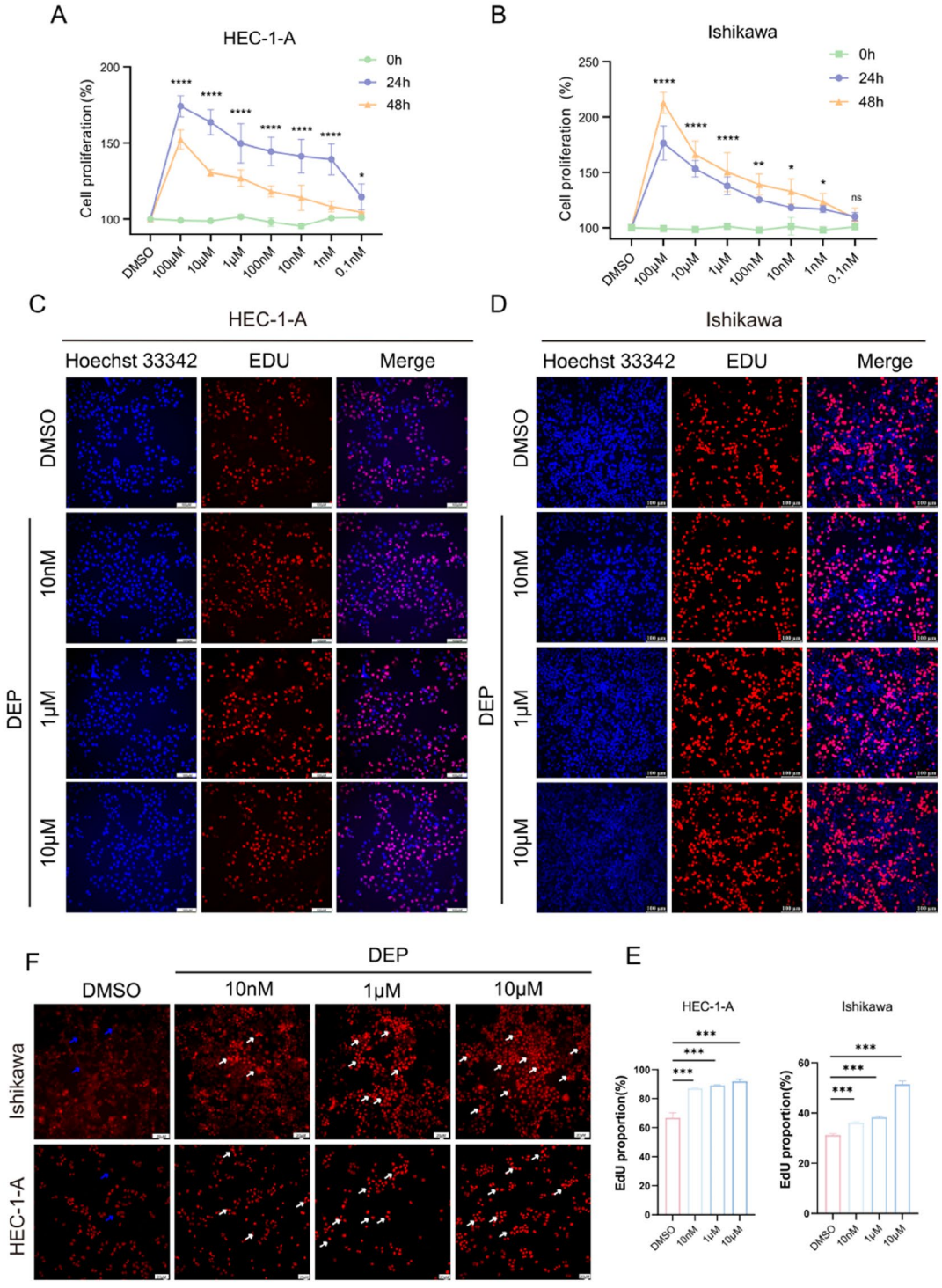

Fig. 8. DEP inhibited proliferation and induced ROS accumulation in EC cells. (**A**, **B**) CCK-8 assay showing the effect of different concentrations of DEP on proliferation of HEC-1-A (**A**) and Ishikawa (**B**) cells at 0, 24, and 48 h (n = 3). (**C**, **D**) Representative EdU staining images of HEC-1-A (**C**) and Ishikawa (**D**) cells treated with DEP at the indicated concentrations (DMSO, DEP at 10 nM, 1 μM, and 10 μM). Nuclei were counterstained with Hoechst 33,342 (blue), and EdU-positive cells are shown in red. Scale bar = 100 μm. (**E**) Quantification of EdU-positive cells in HEC-1-A and Ishikawa cell lines (n = 3). (**F**) Representative fluorescence images of ROS staining in HEC-1-A and Ishikawa cells following DEP treatment. White arrows indicate cells with increased ROS levels, and blue arrows indicate cells with lower ROS levels (n = 3). Scale bar = 20 μm. Data are presented as mean ± SD; statistical significance was determined by one-way ANOVA followed by Tukey's post hoc test. (*$P < 0.05$, ***$P < 0.001$, ****$P < 0.0001$).

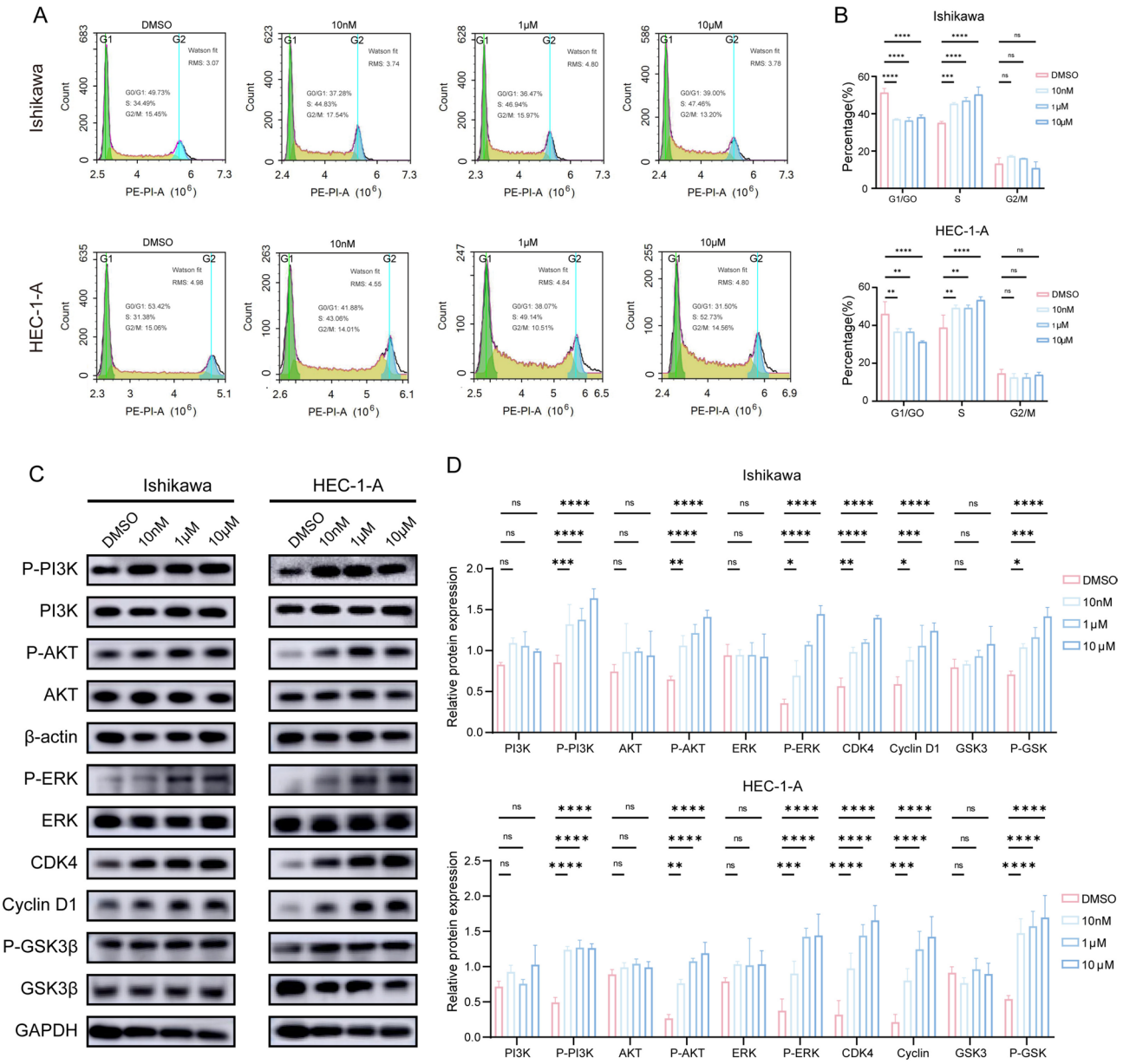

Fig. 9. DEP induced cell cycle arrest and modulated the PI3K/AKT signaling pathway in EC cells. (**A**) Flow cytometry analysis of cell cycle distribution in Ishikawa and HEC-1-A cells treated with DEP at the indicated concentrations (DMSO, DEP at 10 nM, 1 μM, and 10 μM). (**B**) Quantification of the proportion of cells in G0/G1, S, and G2/M phases (n = 3). (**C**) WB analysis of PI3K, P-PI3K, AKT, P-AKT, ERK, P-ERK, CDK4, Cyclin D1, GSK3β, and P-GSK3β expression in Ishikawa and HEC-1-A cells after DEP treatment. (**D**) Densitometric quantification of protein expression levels normalized to GAPDH or β-actin (n = 3). Data are presented as mean ± SD; statistical significance was determined by one-way ANOVA followed by Tukey's post hoc test. (ns = not significant, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$).

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 3 October 2025; Accepted: 4 February 2026
Published online: 10 February 2026

References

1. Gu, B. et al. Variations in incidence and mortality rates of endometrial cancer at the global, regional, and national levels, 1990–2019. *Gynecol. Oncol.* **161**(2), 573–580 (2021).

2. Feng, J., Lin, R., Li, H., Wang, J. & He, H. Global and regional trends in the incidence and mortality burden of en dometrial cancer, 1990–2019: Updated results from the Global Burden of Disease Study, 2019. *Chin. Med. J. (Engl.)* **137**(3), 294–302 (2024).
3. Lortet-Tieulent, J., Ferlay, J., Bray, F. & Jemal, A. International patterns and trends in endometrial cancer incidence, 197 8–2013. *J. Natl. Cancer Inst.* **110**(4), 354–361 (2018).
4. Qiu, S., Jiang, S., Ye, Q., Yang, Y. & Li, X. Global trends and geographical disparities in the incidence of uterine cancer from 1990 to 2021. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **311**, 114066 (2025).
5. Tang, W.-Z. et al. The global burden of polycystic ovary syndrome, endometriosis, uterine fibroids, cervical cancer, uterine cancer, and ovarian cancer from 1990 to 2021. *BMC Public Health* **25**(1), 1774 (2025).
6. Pesonen, M. & Vähäkangas, K. Contribution of common plastic-related endocrine disruptors to epithel ial-mesenchymal transition (EMT) and tumor progression. *Chemosphere* **309**(Pt 1), 136560 (2022).
7. Akbariani, M., Omidi, M., Shahabi, Z., Haghi-Aminjan, H. & Shadboorestan, A. The AhR pathway regulation in phthalates-induced cancer promotion, pro gression and metastasis: A scoping review. *Cancer Cell Int.* **25**(1), 27 (2025).
8. Meng, M. et al. Association between urinary phthalates and phthalate metabolites and cancer risk: A systematic review and meta-analysis. *Heliyon* **10**(8), e29684 (2024).
9. Ahern, T. P. et al. Phthalate exposure and breast cancer incidence: A Danish Nationwide Co hort Study. *J. Clin. Oncol.* **37**(21), 1800–1809 (2019).
10. Wang, Y.-C. et al. Benzyl butyl phthalate promotes breast cancer stem cell expansion via SPHK1/S1P/S1PR3 signaling. *Oncotarget* **7**(20), 29563–29576 (2016).
11. Leng, J. et al. Low-dose mono(2-ethylhexyl) phthalate promotes ovarian cancer developm ent through PPARα-dependent PI3K/Akt/NF-κB pathway. *Sci. Total Environ.* **790**, 147990 (2021).
12. Tsai, Y.-S. et al. Phthalate exposure induces microRNA-5010/Nrf2-EGR1/GDF15 signaling exp ression in prostate cancer. *Ecotoxicol. Environ. Saf.* **290**, 117759 (2025).
13. Ghosh, J., Das, J., Manna, P. & Sil, P. C. Hepatotoxicity of di-(2-ethylhexyl)phthalate is attributed to calcium aggravation, ROS-mediated mitochondrial depolarization, and ERK/NF-κB pathway activation. *Free Radic. Biol. Med.* **49**(11), 1779–1791 (2010).
14. Zhu, Q. et al. Occurrence, spatial distribution and ecological risk assessment of phthalate esters in water, soil and sediment from Yangtze River Delta, China. *Sci. Total Environ.* **806**(Pt 4), 150966 (2022).
15. Chen, J. et al. Exposure to diethyl phthalates impairs ovarian reserve and oocyte maturation in female mice. *Ecotoxicol. Environ. Saf.* **303**, 118956 (2025).
16. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**(D1), D672–D677 (2025).
17. Chen, F.-P. & Chien, M.-H. Chern IY-Y: Impact of low concentrations of phthalates on the effects of 17β-estra diol in MCF-7 breast cancer cells. *Taiwan J. Obstet. Gynecol.* **55**(6), 826–834 (2016).
18. Rashiya N, Sangavi J, Padmini N, Langeswaran K, Alagarsamy A, Selvakumar G, Saravanan M. In silico and in vitro analysis of diethyl phthalate as a quorum sensing inhibitor and its antitumor evaluation against MDA-MB-231 cell lines. *Mol. Divers*. (2025). https://doi.org/10.1007/s11030-11025-11202-w.
19. Long, S. E., Kahn, L. G., Trasande, L. & Jacobson, M. H. Urinary phthalate metabolites and alternatives and serum sex steroid hormones among pre- and postmenopausal women from NHANES, 2013–16. *Sci. Total Environ.* **769**, 144560 (2021).
20. Fiocchetti, M. et al. The peculiar estrogenicity of diethyl phthalate: Modulation of estrogen receptor α activities in the proliferation of breast cancer cells. *Toxics* **9**(10), 237 (2021).
21. Gopalakrishnan, K. et al. Gene expression profiles for low-dose exposure to diethyl phthalate in rodents and humans: A translational study with implications for breast carcinogenesis. *Sci. Rep.* **10**(1), 7067 (2020).
22. Knez, J., Zganec, J., Kotnik, P., Knez, M. & Sobočan, M. 2022-RA-1323-ESGO Endometrial cancer aggressiveness may be associated with exposure to phthalates. *Int. J. Gynecol. Cancer* **32**, A144–A145 (2022).
23. Vasjari, L., Bresan, S., Biskup, C., Pai, G. & Rubio, I. Ras signals principally via Erk in G1 but cooperates with PI3K/Akt for Cyclin D induction and S-phase entry. *Cell Cycle* **18**(2), 204–225 (2019).
24. Cai, W. et al. Targeting cyclin D1 as a therapeutic approach for papillary thyroid carcinoma. *Front. Oncol.* **13**, 1145082 (2023).
25. Bamberger, A. M., Milde-Langosch, K., Rössing, E., Goemann, C. & Löning, T. Expression pattern of the AP-1 family in endometrial cancer: Correlations with cell cycle regulators. *J. Cancer Res. Clin. Oncol.* **127**(9), 545–550 (2001).
26. Hernández-Silva, C. D., Villegas-Pineda, J. C. & Pereira-Suárez, A. L. Expression and role of the G protein-coupled estrogen receptor (GPR30/GPER) in the development and immune response in female reproductive cancers. *Front. Endocrinol. (Lausanne)* **11**, 544 (2020).
27. Chen, C. et al. LSD1 sustains estrogen-driven endometrial carcinoma cell proliferation through the PI3K/AKT pathway via di-demethylating H3K9 of cyclin D1. *Int. J. Oncol.* **50**(3), 942–952 (2017).
28. Zhang, H. et al. cAMP-PKA/EPAC signaling and cancer: the interplay in tumor microenvironment. *J. Hematol. Oncol.* **17**(1), 5 (2024).
29. Lan, T. et al. Increased endogenous PKG I activity attenuates EGF-induced proliferati on and migration of epithelial ovarian cancer via the MAPK/ERK pathway. *Cell Death Dis.* **14**(1), 39 (2023).
30. Lv, Y. et al. Nucleotide de novo synthesis increases breast cancer stemness and meta stasis via cGMP-PKG-MAPK signaling pathway. *PLoS Biol.* **18**(11), e3000872 (2020).
31. Chen, J. et al. Overexpression of FOS enhances the malignant potential of eutopic endometrial stromal cells in patients with endometriosis-associated ovaria n cancer. *Oncol. Rep.* **53**(4), 45 (2025).
32. Zhang, H. et al. Venlafaxine antagonizes the noradrenaline-promoted colon cancer progression by inhibiting the norepinephrine transporter. *Cell Death Discov.* **9**(1), 152 (2023).
33. Jiang, C. et al. NR4A1 suppresses breast cancer growth by repressing c-Fos-mediated lip id and redox dyshomeostasis. *Exp. Mol. Med.* **57**(4), 804–819 (2025).
34. Xia, T. & Han, K. Machine learning prediction model with shap interpretation for chronic bronchitis risk assessment based on heavy metal exposure: A nationally representative study. *BMC Pulm. Med.* **25**(1), 252 (2025).
35. Li, X. et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US ad ults via SHAP: Findings of the US NHANES from 2003 to 2018. *Chemosphere* **311**(Pt 1), 137039 (2023).

Author contributions

Xi Chen: Conceptualization, Methodology, Data curation, Formal analysis, Visualization, Writing – original draft. Zijing Wang: Methodology, Software, Data curation, Validation, Visualization, Writing – original draft. Fengfeng Wang: Investigation, Validation, Resources. Yuexiao Wu: Data curation, Validation. Dan Hu: Investigation, Validation. Yuemei Cheng: Resources, Data curation. Yijuan Xing: Investigation, Formal analysis. Junhong Du: Resources, Validation. Tao Jiang: Methodology, Supervision, Writing – review & editing. Xuehan Bi: Conceptualization, Project administration, Supervision, Funding acquisition, Writing – review & editing. Xiaolei Liang: Conceptualization, Supervision, Funding acquisition, Writing – review & editing. Yongxiu Yang: Conceptualization, Project administration, Supervision, Writing – review & editing.

Funding
This work was supported by the National Natural Science Foundation of China (Nos. 81960278, 82360303, and 82260543), the Joint Fund of Gansu Province (No. 25JRRA1258), the Gansu Provincial Key Research and Development Project (No. 24YFFA038), and the Science and Technology Innovation Major Project of the Health Commission of Gansu Province (No. GSWSZD2025-03).

Declarations

Competing interests
The authors declare no competing interests.

Ethics statement
The study protocol was approved by the Ethics Committee of The First Hospital of Lanzhou University (Approval No. LDYLL2024-735). The study involved only publicly available datasets and established human cell lines, and thus did not require informed consent.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-39325-6.

Correspondence and requests for materials should be addressed to T.J., Y.Y., X.L. or X.B.