



OPEN

Environmental exposure to perfluorooctane sulfonate and its role in esophageal cancer progression: a comprehensive bioinformatics and experimental study

Fangyi Liu^{1,2}, Yansong Lin^{1,2}, Lixuan Liu^{1,2}, Yinli Zheng^{1,2} & Jingping Yun^{1,2}✉

Esophageal cancer (ESCA) is a significant malignancy with rising global incidence rates and considerable impacts on patient survival and quality of life. Current diagnostic and therapeutic strategies face limitations, necessitating research into its underlying mechanisms and potential biomarkers for early diagnosis. This study aims to investigate the role of perfluorooctane sulfonate (PFOS), an environmental toxicant, in the development of ESCA through a comprehensive bioinformatics approach. Using the TCGA-ESCA dataset, we identified differentially expressed genes (DEGs) and intersected them with PFOS-related toxicity targets predicted via Comparative Toxicogenomics Database (CTD) and SuperPred. Machine learning (Random Forest, XGBoost, LASSO, SVM) were applied to prioritize core targets. Survival analysis, *in vitro* qPCR (ESO-26/FLO-1 cells), and molecular docking were performed. Immune infiltration and pathway activity (GSVA) were assessed. We identified 98 PFOS-related DEGs in ESCA, enriched in hypoxia response, epithelial migration, and cancer-associated pathways (e.g., AGE-RAGE, PI3K-Akt). Machine learning highlighted three core targets: PLAU, TOP2A, and BAX. High expression of these genes correlated with poor survival (PLAU, $p = 0.047$) and was upregulated in ESCA tissues. PFOS exposure significantly elevated their expression in esophageal cancer cells. Molecular docking revealed strong binding affinities between PFOS and core targets. GSVA linked PLAU/TOP2A/BAX to oncogenic pathways (angiogenesis, DNA repair), while immune analysis showed PLAU's association with stromal infiltration and TOP2A's negative correlation with CD8⁺ T cells. PFOS exacerbates ESCA by dysregulating PLAU, TOP2A, and BAX, which drive tumor progression via immune modulation, genomic instability, and oncogenic signaling. These targets may serve as biomarkers and therapeutic vulnerabilities for PFOS-associated ESCA, underscoring the need for environmental regulation and targeted therapies.

Keywords PFOS, Esophageal cancer, Toxicity targets, Machine learning, Molecular docking, Molecular dynamics, Immune infiltration

Esophageal cancer (ESCA) represents a significant global health challenge, characterized by rising incidence rates and a profound impact on patient survival and quality of life¹. The disease not only inflicts a heavy toll on affected individuals but also imposes substantial economic burdens on healthcare systems due to high treatment costs and lost productivity². Current diagnostic and therapeutic modalities, including surgical intervention, chemotherapy, and radiotherapy, are hindered by limitations such as late-stage diagnosis and the development of treatment resistance^{3,4}. These challenges underscore a critical research gap, necessitating the exploration of novel approaches to enhance early detection and treatment efficacy⁵. In this context, the present study aims

¹State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-Sen University Cancer Center, Guangzhou 510060, People's Republic of China. ²Department of Pathology, Sun Yat-Sen University Cancer Center, Guangzhou 510060, People's Republic of China. ✉email: yunjp02412@aliyun.com; yunjp@sysucc.org.cn

to investigate the role of perfluorooctane sulfonate (PFOS), an environmental toxicant, in the pathogenesis of ESCA.

Recent studies have suggested that various environmental factors may play a pivotal and significant role in the complex development of esophageal cancer, a serious and often life-threatening condition^{6,7}. Among these environmental influences, PFOS has garnered considerable attention as a persistent organic pollutant, known for its widespread presence in the environment and its potential carcinogenic properties, raising concerns about its impact on human health and the urgent need for further research into its effects⁸. Prior research has indicated correlations between PFOS exposure and various health risks, including cancer, highlighting the necessity for further exploration of its specific effects on esophageal cancer^{9,10}. Understanding the association between PFOS and the carcinogenic processes involved in ESCA could provide valuable insights into the disease's etiology and pave the way for preventive strategies. Furthermore, a recent investigation has revealed that exposure to environmental factors and the accumulation of PFOS in the body may significantly contribute to the progression of esophageal squamous cell carcinoma¹¹. Although there is an expanding body of research linking environmental pollutants to cancer, a considerable gap persists in our understanding of the precise mechanisms by which PFOS affects the development of esophageal cancer. Most existing research has concentrated on general correlations, neglecting the molecular alterations caused by PFOS exposure^{12–14}. This deficiency in detailed insight highlights the necessity for advanced bioinformatics methods to analyze extensive datasets, which would allow researchers to pinpoint crucial toxicological targets and clarify their interactions with PFOS.

In this study, we adopt a comprehensive bioinformatics methodology that includes differential expression analysis, machine learning algorithms, and molecular docking simulations. These tools allow for the identification of differentially expressed genes (DEGs) associated with PFOS exposure and their potential roles in the progression of esophageal cancer. By leveraging these techniques, we aim to uncover the molecular mechanisms underlying the relationship between PFOS and esophageal cancer, ultimately identifying novel biomarkers that can aid in disease diagnosis and treatment.

The primary objective of this research is to elucidate how PFOS impacts the development of esophageal cancer at a molecular level and to explore the potential of identified biomarkers for clinical application. By integrating bioinformatics approaches, we hope to provide a clearer understanding of the influence of environmental toxins on cancer pathogenesis and contribute to the ongoing efforts in cancer research and public health. The findings from this study could significantly advance our knowledge of esophageal cancer and facilitate the development of targeted therapies aimed at mitigating the effects of environmental toxins like PFOS.

Methods

Collection of ESCA-related targets

We made use of genomic data that is publicly accessible from the TCGA-ESCA dataset, which comprises RNA-sequencing information for 163 ESCA tumor samples and 11 samples of normal tissue. This dataset was obtained from the Cancer Genome Atlas (TCGA) portal (<https://portal.gdc.cancer.gov/>). Differentially expressed genes (DEGs) between ESCA tumor tissues and normal tissues were identified with the limma package in R. The threshold for considering DEGs significant was set at $p\text{-adj} < 0.05$.

Prediction of PFOS toxicity targets

Potential toxicity targets of PFOS were predicted using the Comparative Toxicogenomics Database (CTD) (<https://ctdbase.org/>) and SuperPred (<https://prediction.charite.de/index.php>)^{15,16}. Venn diagram analysis was conducted utilizing an online tool (<https://bioinfo.gp.cnb.csic.es/tools/venny/index.html>) to pinpoint overlapping targets that are differentially expressed in connection with PFOS exposure and are involved in the pathogenesis of esophageal cancer.

Expression and PPI analysis

The expression levels of the identified toxicity targets were visualized in a comprehensive heatmap, which was generated using the sophisticated “ComplexHeatmap” package. In addition, a thorough protein-protein interaction (PPI) analysis was conducted utilizing the extensive STRING database, which is a valuable resource for understanding the functional associations between proteins (<https://cn.string-db.org/>)¹⁷. The resulting interaction network was then elegantly visualized with the advanced Cytoscape software, specifically version 3.8.2. The top 40 hub genes were selected based on degree centrality (the number of connections per node) within the PPI network, as nodes with higher degrees are considered biologically pivotal in network analysis. Nodes with degree values below the top 40 were excluded to focus on the most interconnected and potentially influential targets.

Enrichment analyses

Kyoto Encyclopedia of Genes and Genomes (KEGG) (<https://www.kegg.jp/>) provides an integrated database platform for the systematic representation and computational analysis of biological systems¹⁸. Enrichment analyses were conducted to gain insights into the functional roles of the identified genes, specifically utilizing the Gene Ontology (GO) and KEGG enrichment analyses. These analyses were performed with the aid of the “clusterProfiler” package, version 3.18.0. In the context of the GO analysis, we considered three distinct categories: biological process (BP), cellular component (CC), and molecular function (MF), each providing a unique perspective on gene functionality. To effectively communicate our findings, the top 10 pathways identified through these analyses were visually represented using informative bar plots.

Machine learning algorithms

Four distinct machine learning algorithms were employed to identify and analyze key toxicological targets, each chosen for its unique strengths and capabilities in handling complex datasets. The Random Forest (RF) algorithm was implemented using the “randomForest” package, allowing for robust classification and regression tasks through its ensemble learning approach. Meanwhile, the XGBoost model was developed utilizing the “xgboost” package, which is renowned for its efficiency and performance in gradient boosting, particularly in scenarios involving large datasets and high-dimensional feature spaces. LASSO regression, a technique known for its ability to perform both variable selection and regularization to enhance the prediction accuracy and interpretability of the statistical model, was executed using the “glmnet” package. Lastly, the Support Vector Machine (SVM) algorithm was executed through the “e1071” package, which is widely recognized for its effectiveness in classification tasks, especially in high-dimensional spaces. To visualize the relationships and intersections among the identified toxicological targets, an UpSet Venn diagram was created using the “UpSetR” package. Detailed descriptions of the machine learning methods are provided in the Supplementary Methods section.

Differential expression analysis

The differential expression of the core toxicity targets, which include PLA2, TOP2A, and BAX, was analyzed through the use of box plots and receiver operating characteristic (ROC) curves, providing a comprehensive visual representation of the data. These informative plots were generated utilizing the “ggplot2” package, alongside the “pROC” package, within the R programming environment.

GSVA analysis

The CancerSEA database has categorized the distinct functional states of 14 different tumor cell types. A prior study introduced the z-score algorithm, which integrates the expression of characteristic genes to reflect pathway activity¹⁹. Using the R package GSVA with the zscore parameter, the 14 functional state gene sets were computed, and a combined z-score was obtained. We then applied the scale function to further normalize this score, defining it as the gene set score. Finally, we calculated the Pearson correlation between each gene and the respective gene set scores.

Immune cell infiltration analysis

The ESTIMATE score, a crucial metric in evaluating tumor microenvironments, was computed using the “estimate” package. Following this, the single-sample Gene Set Enrichment Analysis (ssGSEA) was conducted utilizing the “GSVA” package. To visualize the results, heatmaps that represent the correlation matrix were generated using the “pheatmap” package, allowing for a clear depiction of the relationships among the variables. Furthermore, the correlation between the core toxicity targets and the infiltration of various immune cell types was analyzed employing Pearson correlation.

Cell experiment protocol

The ESO-26 and FLO-1 cell lines, obtained from the American Type Culture Collection (ATCC), were cultured in RPMI-1640 medium (Life Technologies, Shanghai, China) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin. Cells were maintained at 37 °C in a 5% CO₂ humidified incubator. For treatment, cells were incubated with either 2 μM PFOS or a DMSO vehicle control for 48 h. Post-treatment, the cells were washed twice with PBS and harvested for downstream analysis. Total RNA was extracted using a Thermo Fisher Scientific RNA isolation kit, followed by cDNA synthesis with the PrimeScript RT Reagent Kit. Gene expression profiling was performed via quantitative PCR (qPCR) on an ABI 7900HT system (Applied Biosystems). The $2^{-\Delta\Delta C_t}$ method was applied to calculate relative gene expression levels.

Molecular docking analysis

To investigate the binding affinities between PFOS and key proteins, we conducted molecular docking simulations. We sourced 3D protein structures from the PDB (<https://www.rcsb.org/>) and refined them by eliminating water molecules, heteroatoms, and organic ligands using PyMOL. The PFOS ligand was sourced from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) and converted to PDBQT format via AutoDockTools, followed by energy minimization. We defined a 3D grid around each protein's active site, carefully choosing dimensions to encapsulate the binding area. Docking was executed in AutoDock Vina with the prepared protein and ligand files. Binding affinities for each pose were computed, and the poses with the lowest binding energies were chosen. Visualization of the docking results was done using PyMOL 2.5.2.

Molecular dynamics (MD) simulations

MD simulations were carried out using the Desmond/Maestro 2022.1 software package. The system was constructed with the TIP3P water model and incorporated a 0.15 M NaCl solution to neutralize the system's charge and replicate physiological conditions. Prior to the production phase, the system underwent energy minimization via the steepest descent method to eliminate any steric clashes or unfavorable interactions in the initial configuration. The system was then equilibrated for 100 ps under isothermal-isobaric (NPT) conditions at 300 K and 1 bar to stabilize the temperature and pressure. Subsequently, a 100 ns production simulation was performed at the same conditions, and trajectory analysis was conducted using Desmond's built-in tools.

Statistical analysis

Data are expressed as mean values ± standard deviation (SD). Comparisons between groups were performed using an unpaired two-tailed Student's t-test, with a p-value below 0.05 considered statistically significant.

Results

Identification and analysis of PFOS toxicity targets in esophageal cancer

In this study, we utilized the TCGA-ESCA dataset to identify a total of 5,757 differentially expressed genes (DEGs) associated with esophageal cancer. Subsequently, we employed the Comparative Toxicogenomics Database and SuperPred to predict 255 potential toxicity targets of perfluorooctane sulfonate (PFOS). Through the application of the Venn diagram tool, we further refined our analysis to identify 98 toxicity targets that are differentially expressed in response to PFOS exposure and are implicated in the pathogenesis of ESCA (Fig. 1A and Table S1). The identification of these 98 targets is crucial for understanding the molecular mechanisms by which PFOS may contribute to the development of esophageal cancer. In Fig. 1B, we present a heatmap that visualizes the expression levels of the 98 identified toxicity targets within the TCGA-ESCA dataset. The heatmap reveals distinct expression patterns, indicating that these targets may play significant roles in the biological processes associated with ESCA. Furthermore, Fig. 1C depicts the protein-protein interaction (PPI) analysis of the 98 toxicity targets. The PPI network elucidates the complex interactions among these proteins, providing insights into the potential signaling pathways and biological processes that may be affected by PFOS exposure. Overall, our findings contribute to the growing body of evidence linking environmental toxicants, such as PFOS, to the risk of esophageal cancer, emphasizing the need for continued research in this area to better understand the underlying mechanisms and to inform public health strategies.

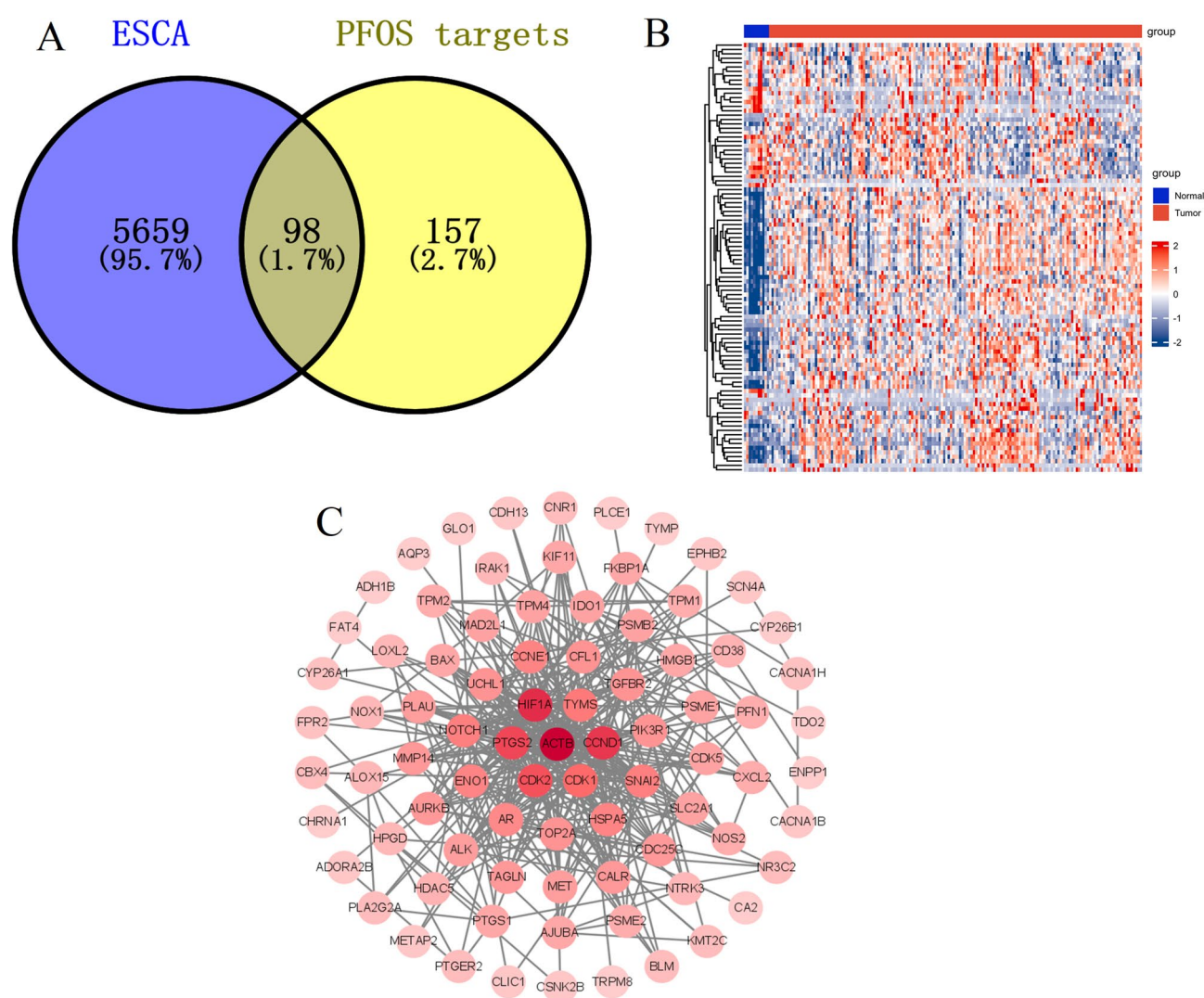


Fig. 1. Analysis of PFOS toxicity targets in esophageal cancer. (A) Venn diagram showing the intersection of DEGs in esophageal cancer from the TCGA-ESCA dataset and 255 predicted PFOS toxicity targets, resulting in 98 overlapping genes. (B) Heatmap displaying the expression levels of the 98 overlapping toxicity targets in the TCGA-ESCA dataset, comparing normal and tumor tissues. The color gradient indicates the level of gene expression, with red representing higher expression and blue representing lower expression. (C) PPI network analysis of the 98 overlapping toxicity targets. Node colours reflect the degree of connectivity, with larger nodes representing more connected proteins.

Enrichment analyses of differentially expressed PFOS-related toxicity targets in esophageal cancer

To elucidate the biological functions and pathways associated with the 98 differentially expressed toxicity targets of PFOS exposure in esophageal cancer, enrichment analyses were performed. As shown in Fig. 2A, the top 10 enriched BP terms include response to oxygen levels, response to hypoxia, response to decreased oxygen levels, epithelium migration, and epithelial cell migration. These results suggest that PFOS exposure may influence cellular responses to hypoxic conditions and migration-related processes, which are critical in tumor progression and metastasis. The top 10 enriched CC terms highlight the involvement of structures and complex assemblies such as the serine/threonine protein kinase complex, cell-substrate junction, focal adhesion, cyclin-dependent protein kinase holoenzyme complex, and protein kinase complex (Fig. 2B). These components are implicated in maintaining cellular architecture and signaling pathways necessary for cell adhesion and migration. The top 10 enriched MF terms include heme binding, tetrapyrrole binding, protein serine/threonine/tyrosine kinase activity, voltage-gated calcium channel activity, and gated channel activity (Fig. 2C). These molecular functions are essential for cellular signal transduction and ion transport, which are often dysregulated in cancer cells. As shown in Fig. 2D, the KEGG pathway enrichment analysis identified several key pathways associated with the 98 toxicity targets. Notable pathways include Small cell lung cancer, Thyroid hormone signaling pathway, MicroRNAs in cancer, Oocyte meiosis, Human T-cell leukemia virus 1 infection, Prostate cancer, the AGE-RAGE signaling pathway in diabetic complications, central carbon metabolism in cancer. The involvement of these pathways underscores the multifaceted roles of PFOS exposure in perturbing various signaling mechanisms and cellular processes that contribute to the development and progression of ESCA. These comprehensive enrichment analyses reveal that PFOS exposure can significantly impact a wide range of biological processes,

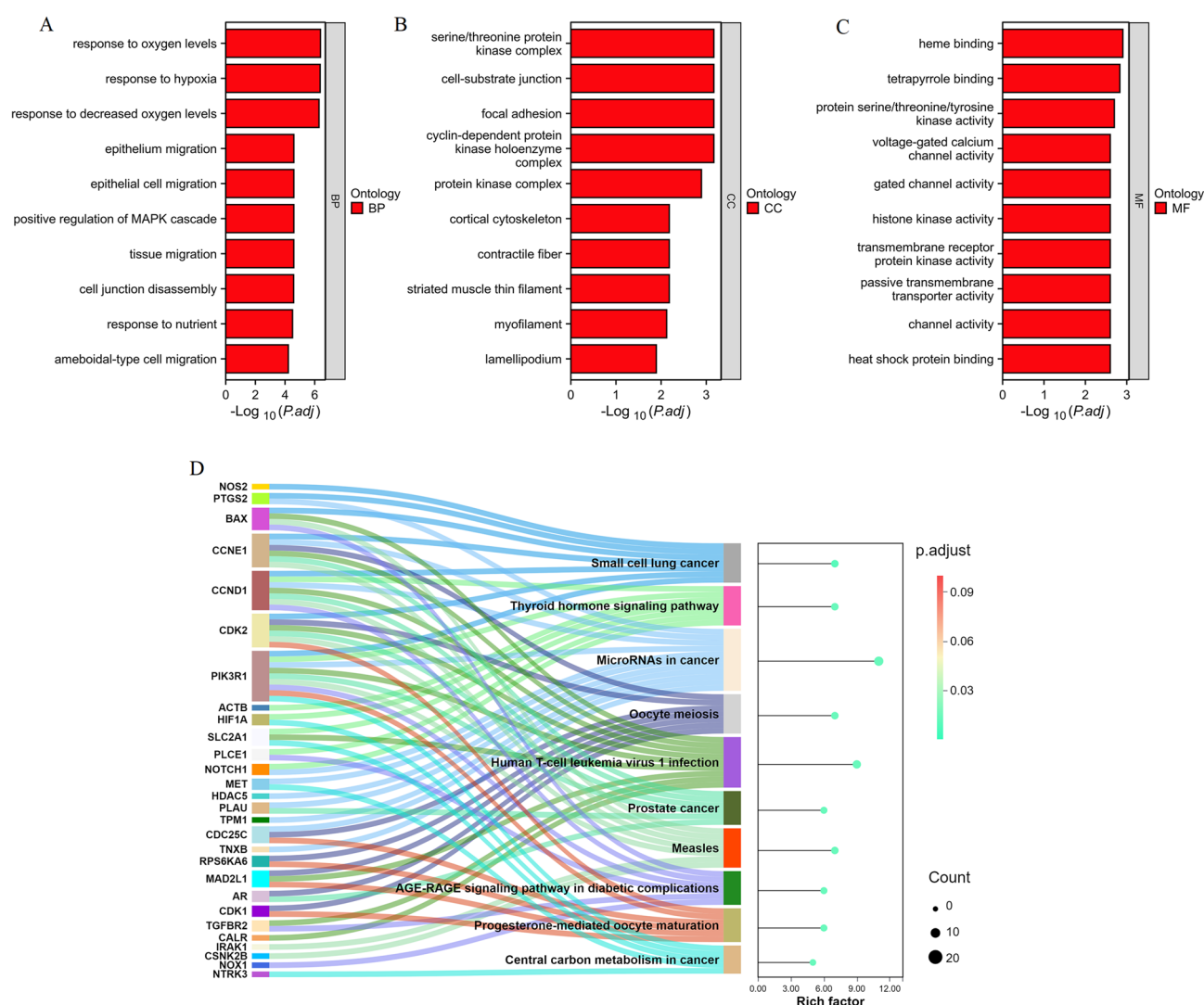


Fig. 2. Enrichment analyses of differentially expressed PFOS-related toxicity targets. GO enrichment analysis of the 98 differentially expressed PFOS toxicity targets categorized by BP (A), CC (B), and MF (C). (D) KEGG pathway enrichment analysis of the 98 targets identifying significant pathways.

cellular components, and molecular functions, providing a deeper understanding of the mechanistic pathways through which PFOS may influence carcinogenesis in the esophagus.

Identification of key toxicological targets using machine learning algorithms

In this study, we employed four distinct machine learning algorithms to identify key toxicological targets associated with esophageal cancer, based on the gene expression profiles of the top 40 toxicity targets derived from the PPI network. Figure 3A illustrates the results obtained using the RF algorithm, which successfully identified 8 significant toxic targets. Figure 3B presents the findings from the XGBoost algorithm, which identified 9 important toxic targets. Figure 3C shows the results from the LASSO algorithm, which pinpointed 11 significant toxic targets. Figure 3D, E depict the outcomes from the SVM algorithm, which identified a total of 24 important toxic targets. Finally, Fig. 3F employs the UpSet Venn diagram to reveal 3 core toxic targets (PLAU, TOP2A, and BAX) that were consistently identified across multiple algorithms. Overall, the integration of machine learning approaches in this study not only enhances our understanding of the toxicological landscape associated with PFOS exposure but also provides a foundation for future research aimed at elucidating the mechanisms underlying esophageal cancer development.

Survival analysis and expression levels of key targets in esophageal cancer.

The survival analysis for key target genes in the TCGA dataset demonstrated distinct survival probabilities based on the expression levels of PLAU, TOP2A, and BAX. Figure 4A shows the Kaplan-Meier survival curve for PLAU, indicating that high expression levels of PLAU are significantly associated with poorer overall survival. The hazard ratio for high expression vs. low expression is 1.92 ($P=0.047$). In Fig. 4B, the survival analysis for TOP2A shows that its high expression is associated with a trend towards worse overall survival, though it did not reach statistical significance ($P=0.061$). Figure 4C illustrates the results for BAX, where high expression also correlates with a trend towards reduced overall survival, with a hazard ratio of 1.89 ($P=0.057$). The gene expression levels of PLAU, TOP2A, and BAX were further analyzed between normal and esophageal cancer tissues using the TNMplot database. Figure 4D shows that the expression levels of PLAU were significantly elevated in esophageal cancer tissues compared to normal tissues ($P<0.001$). Similarly, Fig. 4E indicates a significant overexpression of TOP2A in tumor tissues relative to normal tissues ($P<0.001$). Figure 4F shows that BAX expression was also markedly higher in esophageal cancer tissues compared to normal tissues ($P<0.01$).

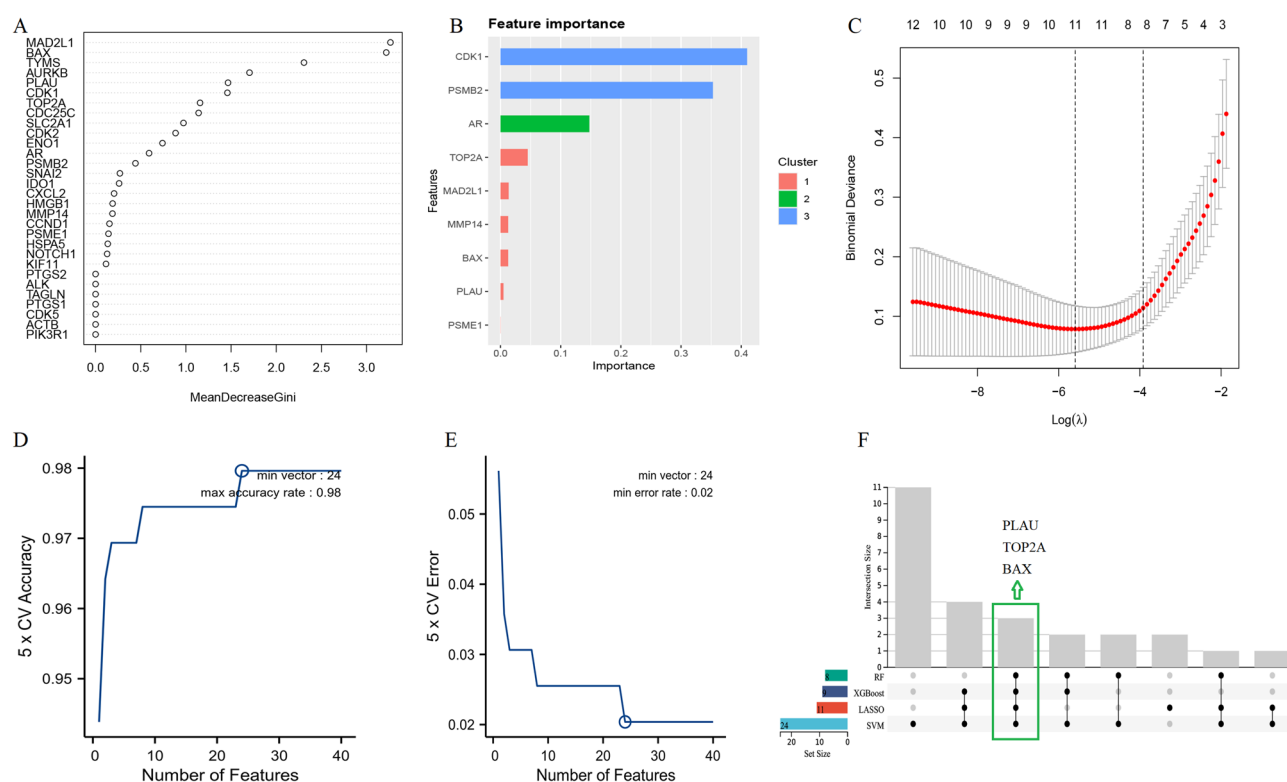


Fig. 3. Identification of key toxicological targets using machine learning algorithms. (A) RF algorithm identified 8 significant toxicity targets based on their mean decrease Gini values. (B) XGBoost algorithm recognized 9 important toxicity targets, illustrating feature importance values. (C) LASSO algorithm identified 11 significant toxicity targets, with the tuning parameter (λ) selection plot showing the binomial deviance. SVM algorithm identified 24 significant toxicity targets. (D) Five-fold cross-validated accuracy plot and (E) cross-validated error plot display the optimal number of features for maximum model performance. (F) UpSet Venn diagram highlighting the 3 core toxicity targets (PLAU, TOP2A, BAX) commonly identified by the RF, XGBoost, LASSO, and SVM algorithms.

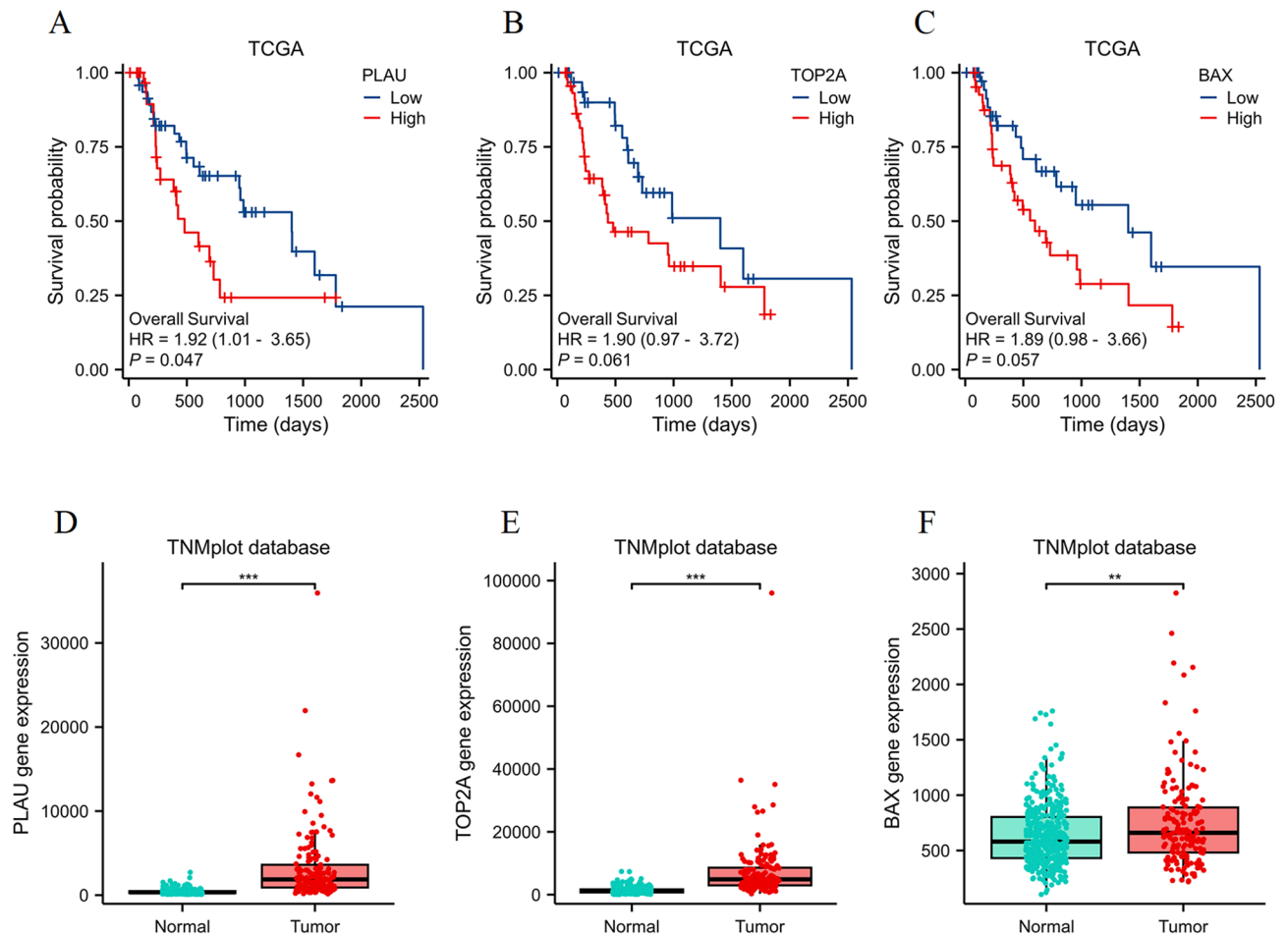


Fig. 4. Survival analysis and expression levels of key targets in esophageal cancer. Kaplan-Meier survival curves for overall survival of esophageal cancer patients from the TCGA dataset based on the expression levels of PLAU (A), TOP2A (B), and BAX (C). Gene expression levels of PLAU (D), TOP2A (E), and BAX (F) in normal and esophageal cancer tissues from the TNMplot database. $**P < 0.01$, $***P < 0.001$.

These findings collectively suggest that high expression levels of PLAU, TOP2A, and BAX are associated with esophageal cancer, and elevated levels of these genes are indicative of poorer prognosis in ESCA patients. The differential expression of these genes in tumor versus normal tissues highlights their potential role as biomarkers and therapeutic targets in esophageal cancer.

Validation of core gene expression in esophageal cancer cells following PFOS exposure

The expression levels of PLAU, TOP2A, and BAX were significantly increased in esophageal cancer cells following PFOS exposure. In ESO-26 cells (Fig. 5A), PFOS treatment led to a notable upregulation of PLAU ($P < 0.01$), TOP2A ($P < 0.05$), and BAX ($P < 0.01$). In FLO-1 cells (Fig. 5B), PFOS exposure resulted in a significant increase in the expression of PLAU ($P < 0.001$), TOP2A ($P < 0.001$), and BAX ($P < 0.01$). These results indicate that PFOS exposure leads to a marked upregulation of PLAU, TOP2A, and BAX gene expressions in esophageal cancer cell lines, suggesting a potential mechanism by which PFOS may influence esophageal carcinogenesis.

GSVA analysis of core toxicity targets

GSVA was performed to analyze the association between the core toxic targets (PLAU, TOP2A, and BAX) and various oncogenic pathways. As illustrated in Fig. S1A, PLAU expression exhibited significant positive correlations with multiple oncogenic pathways, including angiogenesis, apoptosis, differentiation, DNA damage, EMT, hypoxia, inflammation, invasion, metastasis, proliferation, and quiescence. Figure S1B demonstrates that TOP2A expression significantly correlated with cell cycle, DNA damage, DNA repair, and proliferation. In addition, TOP2A negatively correlating with hypoxia and inflammation. In Fig. S1C, BAX was negatively correlated with differentiation and hypoxia. In sum, GSVA enrichment analysis highlights significant associations between PLAU, TOP2A, and BAX expressions with critical oncogenic pathways, underscoring their importance in the molecular pathology of esophageal cancer affected by PFOS exposure. The results provide insights into potential mechanisms through which these core targets contribute to cancer progression.

Correlation analysis of core toxicity targets with immune cell infiltration.

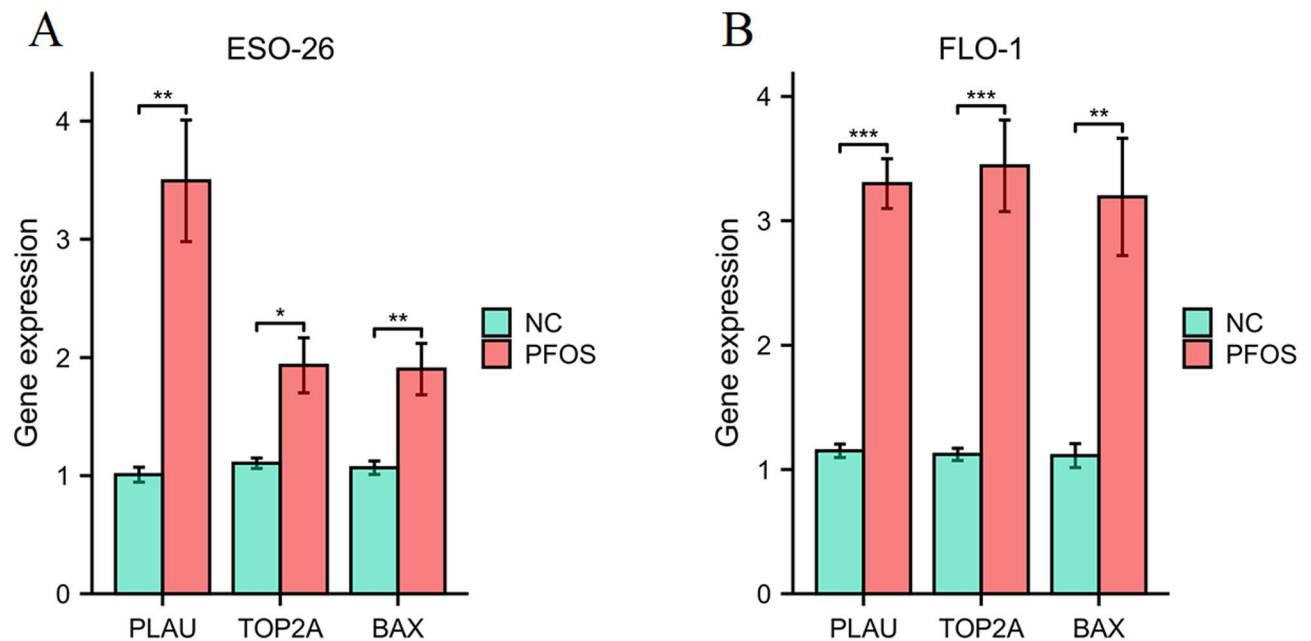


Fig. 5. Upregulation of core genes in esophageal cancer cells following PFOS exposure. Gene expression levels of PLAU, TOP2A, and BAX in ESO-26 cells (A) and FLO-1 cells (B) treated with PFOS (red bars) compared to the control group (NC, green bars). Data are expressed as mean values $\pm$ standard deviation (SD). ** $P < 0.01$, *** $P < 0.001$.

We analyzed the correlation between the core toxicity targets (PLAU, TOP2A, and BAX) and immune cell infiltration using the ESTIMATE and ssGSEA algorithms. Figure S2A presents the correlation matrix of core toxicity targets with StromalScore, ImmuneScore, and ESTIMATEScore. PLAU exhibited a significant positive correlation with both StromalScore ($p < 0.001$) and ESTIMATEScore ($p < 0.01$). TOP2A showed a negative correlation with ImmuneScore ($p < 0.01$) and ESTIMATEScore ($p < 0.05$). BAX did not demonstrate significant correlations with these scores. Figure S2B illustrates the ssGSEA analysis for immune cell infiltration. PLAU expression positively correlated with various immune cells, including iDC, macrophages, NK cells, Tcm, Th1 cells, Th2 cells and several other immune cell types. TOP2A expression was negatively correlated with CD8 T cells, DC, mast cells, neutrophils, and other immune cells. BAX exhibited fewer significant correlations; it was correlated with Tcm. These results suggest that PLAU and TOP2A are involved in immune cell infiltration in esophageal cancer, potentially influencing the tumor microenvironment and immune response mechanisms associated with PFOS exposure.

Molecular docking analysis

Molecular docking analysis was performed to elucidate the interactions between PFOS and the core toxicity targets (BAX, PLAU, and TOP2A). The docking results indicate a significant binding affinity between BAX and PFOS, with a Vina score of -8.5 (Fig. 6A). Similarly, the interaction between PLAU and PFOS demonstrates a moderate binding affinity with a Vina score of -7.1 (Fig. 6B). Notably, TOP2A exhibited the strongest binding affinity to PFOS among the three targets, with a Vina score of -10.2 (Fig. 6C). These findings suggest that PFOS may directly interact with BAX, PLAU, and TOP2A, potentially influencing their functional roles in esophageal cancer pathology. The MD simulation analysis of the interaction between PFOS and PLAU was conducted to elucidate the potential mechanisms in ESCA. As illustrated in Fig. 6D, the root mean square deviation (RMSD) of the PLAU protein remained relatively stable throughout the simulation, ultimately stabilizing between 2.8 Å and 3.2 Å. The RMSD of the ligand PFOS exhibited fluctuations before stabilizing around 6.0 Å. This indicates that, despite initial conformational changes, a more stable binding conformation was achieved through the dynamics simulation.

Discussion

ESCA is a significant global health concern characterized by its high mortality rates and complex treatment challenges. The disease is predominantly classified into two main histological types: squamous cell carcinoma and adenocarcinoma, each associated with distinct risk factors and pathophysiological mechanisms. Early detection remains a critical challenge, as most patients are diagnosed at advanced stages, leading to poor prognoses. The rising incidence of ESCA, particularly in high-risk regions, underscores the urgent need for innovative approaches to improve diagnosis and treatment, as well as to understand the underlying biological mechanisms that contribute to its development and progression.

This study aims to explore the potential relationship between PFOS, an environmental toxicant, and the development of esophageal cancer. By employing a comprehensive bioinformatics approach, including

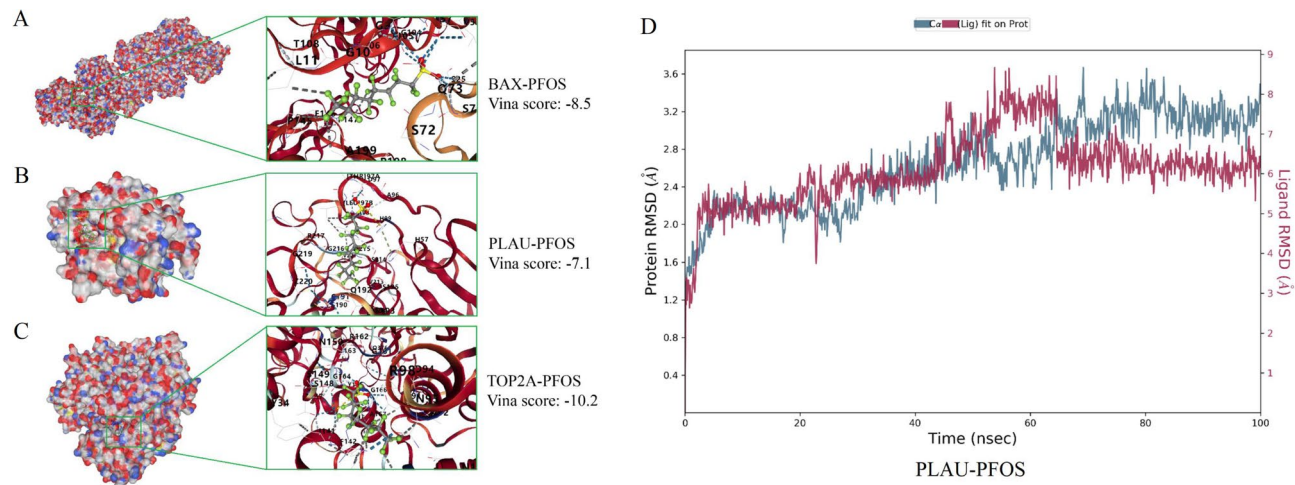


Fig. 6. Molecular docking analysis of core toxicity targets with PFOS. Molecular docking of PFOS with BAX (A), PLAU (B), and TOP2A (C). MD simulation analysis of the interaction between PFOS and PLAU (D).

differential expression analysis, machine learning algorithms, and molecular docking simulations, this research seeks to identify key toxicological targets and elucidate the molecular pathways involved in ESCA related to PFOS exposure. The findings from this study are expected to provide valuable insights into the pathogenesis of ESCA and pave the way for the identification of novel biomarkers for early diagnosis and targeted therapeutic interventions.

The association between PFOS exposure and various cancers has been a subject of increasing interest in recent years. However, the current study did not sufficiently address the direct causality between PFOS exposure and esophageal cancer carcinogenesis. To clarify, we now specifically examine how PFOS may contribute to carcinogenesis in esophageal cancer through the disruption of key biological processes, such as oxidative stress, DNA damage, and immune modulation. Previous studies have indicated that PFOS can disrupt endocrine functions and induce oxidative stress, leading to cellular damage and tumorigenesis^{20–22}. Our enrichment analyses corroborate these findings, as we observed significant involvement of biological processes such as response to hypoxia and epithelial cell migration, which are critical in cancer progression²³. The upregulation of genes related to these processes suggests that PFOS may exacerbate the hypoxic tumor microenvironment, promoting aggressive tumor behavior. Moreover, the identification of key pathways such as the Thyroid Hormone Signaling Pathways and Human T-cell leukemia virus 1 infection align with previous research that has implicated these pathways in various malignancies^{24,25}. The role of the AGE-RAGE signaling pathway in cancer development has been previously explored, and it is known to play a role in inflammation, fibrosis, and tumor progression, which could be particularly relevant in the context of PFOS exposure²⁶. The Thyroid hormone signaling pathway has also been shown to influence the regulation of cell differentiation and proliferation, with implications for tumor growth²⁷. The dysregulation of thyroid hormone signaling is linked to various cancers, including esophageal cancer, as it could contribute to tumor cell proliferation and survival in the presence of environmental toxins such as PFOS. The involvement of microRNAs in cancer, identified in our KEGG pathway enrichment analysis, is particularly noteworthy. MicroRNAs are small non-coding RNAs that regulate gene expression at the post-transcriptional level and have been implicated in cancer progression, including esophageal cancer²⁸. We now hypothesize that PFOS exposure may disrupt the normal expression of microRNAs, leading to the upregulation of oncogenes and downregulation of tumor suppressor genes, thus contributing directly to cancer initiation and progression.

The application of machine learning algorithms in our study has proven effective in identifying core toxicological targets associated with ESCA. The convergence of results from multiple algorithms, including RF, XGBoost, LASSO, and SVM, underscores the robustness of our findings. Notably, the identification of PLAU, TOP2A, and BAX as core targets is particularly significant. These targets are not merely correlated with the presence of esophageal cancer but appear to directly contribute to its carcinogenesis. Our analysis now includes an evaluation of the functional roles of these targets in cancer development, focusing on their contribution to tumorigenesis, cellular survival, and metastasis. Our study found that PLAU expression was significantly upregulated in tumor tissues relative to normal tissues, which is consistent with previous reports linking PLAU to cancer progression²⁹. PLAU has been implicated in multiple oncogenic processes, including angiogenesis, invasion, and metastasis, by regulating the extracellular matrix degradation and promoting cell migration^{30,31}. Our analysis showed that PLAU expression was significantly correlated with critical oncogenic pathways, such as epithelial-mesenchymal transition (EMT), invasion, and metastasis. This aligns with findings in other cancers, where PLAU overexpression facilitates tumor progression by altering the tumor microenvironment³². The strong correlation between PLAU expression and immune cell infiltration (e.g., macrophages, NK cells, and T cells) further supports its role in influencing the tumor immune microenvironment, as previously demonstrated in pancreatic ductal adenocarcinoma³². Thus, PLAU emerges as a promising biomarker for diagnostic purposes and a potential therapeutic target for PFOS-related esophageal cancer. The expression of TOP2A was also significantly

elevated in esophageal cancer tissues, consistent with findings from other studies in various cancers^{33,34}. TOP2A is crucial for DNA replication and repair, and its dysregulation can lead to genomic instability, a hallmark of cancer³⁵. Our study's GSEA analysis indicated that TOP2A expression strongly correlated with DNA damage and proliferation. These findings align with earlier reports that showed elevated TOP2A expression in tumors correlates with higher cell proliferation rates and poor prognosis^{36,37}. Moreover, TOP2A's negative correlation with hypoxia and inflammation pathways in our analysis suggests that TOP2A may play a complex role in modulating the tumor microenvironment, potentially influencing tumor adaptation to stress conditions such as hypoxia and immune infiltration³⁸. In the context of PFOS exposure, the upregulation of TOP2A may be a result of DNA damage induced by environmental toxins. PFOS has been shown to disrupt the DNA repair mechanisms in various cell types, which may enhance the oncogenic potential of molecules like TOP2A^{39,40}. These findings underscore the potential of TOP2A as both a biomarker for esophageal cancer and a target for therapeutic interventions. Interestingly, BAX exhibited a more variable expression pattern in our study, with upregulation observed in cancer tissues compared to normal tissues. The pro-apoptotic protein BAX plays a pivotal role in regulating cell survival and apoptosis by interacting with other Bcl-2 family members to promote mitochondrial outer membrane permeabilization and caspase activation⁴¹. The negative correlations observed between BAX and differentiation and hypoxia pathways suggest that BAX may contribute to the deregulation of apoptotic signaling in esophageal cancer, potentially allowing for tumor cell survival in adverse conditions such as hypoxia. These results corroborate findings in other cancers where reduced BAX expression is linked to poor prognosis and resistance to apoptosis^{42,43}. Thus, BAX might serve as an additional target for therapeutic strategies aimed at restoring apoptotic sensitivity in PFOS-exposed esophageal cancer cells.

The role of PFOS in esophageal cancer development is a growing area of interest, and our study provides new insights into the molecular mechanisms by which this environmental toxin may promote carcinogenesis. PFOS, a perfluorinated compound, is known to induce oxidative stress, DNA damage, and inflammation in various cell types⁴⁴. Our findings indicate that PFOS exposure may dysregulate critical genes involved in DNA repair, apoptosis, cell proliferation, and immune response, contributing to the development and progression of esophageal cancer. Additionally, our analysis of immune cell infiltration revealed that PLAU and TOP2A are involved in modulating the tumor immune microenvironment, suggesting that PFOS exposure may affect the interaction between cancer cells and immune cells. The immunosuppressive effects of PFOS have been documented in other studies, with reports indicating that PFOS exposure can induce immunotoxicity and skew immune responses⁴⁵. These findings point to the importance of investigating the immune-modulatory effects of PFOS in future research, as targeting the tumor microenvironment may offer new therapeutic avenues for PFOS-related cancers.

The limitations of this study warrant careful consideration. Firstly, the absence of wet lab experiments to validate the bioinformatics findings restricts the ability to confirm the biological relevance of the identified differentially expressed genes and toxicological targets. Additionally, without clinical validation, the translational potential of the identified biomarkers remains uncertain, limiting their applicability in real-world settings. These constraints underscore the necessity for subsequent research endeavors that incorporate experimental validation and diverse datasets to substantiate the findings of this investigation. Furthermore, while this study focused on PFOS due to its well-established carcinogenic potential and emerging evidence linking it to esophageal cancer, we acknowledge that other environmentally toxic compounds may also contribute to disease pathogenesis. Future studies should expand this framework to include comparative analyses of additional pollutants, such as PFOA, BPA, or heavy metals, to provide a more comprehensive understanding of environmental toxicants in cancer progression.

Conclusions

In conclusion, this research provides significant insights into the molecular mechanisms by which PFOS exposure may contribute to the development of esophageal cancer. The identification of critical differentially expressed genes and potential biomarkers highlights the importance of environmental factors in cancer pathogenesis. A multidisciplinary approach integrating bioinformatics, experimental validation, and clinical assessments will be pivotal in advancing our understanding of the interplay between environmental toxins and cancer progression.

Data availability

The data used in our study are available from the TCGA (<https://portal.gdc.cancer.gov/>).

Received: 10 April 2025; Accepted: 11 August 2025

Published online: 26 August 2025

References

- Waters, J. K. & Reznik, S. I. Update on management of squamous cell esophageal cancer. *Curr. Oncol. Rep.* **24**, 375–385. <https://doi.org/10.1007/s11912-021-01153-4> (2022).
- Yang, J. et al. Estimating the costs of esophageal cancer screening, early diagnosis and treatment in three high risk areas in China. *Asian Pac. J. Cancer Prevention: APJCP.* **12**, 1245–1250 (2011).
- Musa, I. H., Musa, T. H., Musa, H. H. & Ahmed, M. E. Esophageal cancer epidemiology, diagnosis, and management in Sudan - A review. *Med. J. Malay.* **76**, 691–697 (2021).
- Deboever, N., Jones, C. M., Yamashita, K., Ajani, J. A. & Hofstetter, W. L. Advances in diagnosis and management of cancer of the esophagus. *BMJ (Clinical Res. ed.)*. **385**, e074962. <https://doi.org/10.1136/bmj-2023-074962> (2024).
- Glickman, J. N. Section II: pathology and pathologic staging of esophageal cancer. *Semin. Thorac. Cardiovasc. Surg.* **15**, 167–179 (2003).
- Sheikh, M. et al. Individual and combined effects of environmental risk factors for esophageal cancer based on results from the Golestan cohort study. *Gastroenterology* **156**, 1416–1427. <https://doi.org/10.1053/j.gastro.2018.12.024> (2019).

7. Simba, H., Kuivaniemi, H., Abnet, C. C., Tromp, G. & Sewram, V. Environmental and life-style risk factors for esophageal squamous cell carcinoma in africa: a systematic review and meta-analysis. *BMC Public Health*. **23**, 1782. <https://doi.org/10.1186/s12889-023-16629-0> (2023).
8. Jane, L. E. L. et al. Health-related toxicity of emerging per- and polyfluoroalkyl substances: comparison to legacy PFOS and PFOA. *Environ. Res.* **212**, 113431. <https://doi.org/10.1016/j.envres.2022.113431> (2022).
9. Saikat, S., Kreis, I., Davies, B., Bridgman, S. & Kamanyire, R. The impact of PFOS on health in the general population: a review. *Environ. Sci. Processes Impacts*. **15**, 329–335. <https://doi.org/10.1039/c2em30698k> (2013).
10. Messmer, M. F. et al. Risk of cancer in a community exposed to Per- and Poly-Fluoroalkyl substances. *Environ. Health Insights*. **16**. <https://doi.org/10.1177/11786302221076707> (2022).
11. Liu, Y. et al. Perfluorooctane sulfonate (PFOS) triggers migration and invasion of esophageal squamous cell carcinoma cells via regulation of Zeb1. *Drug Chem. Toxicol.* **45**, 2804–2813. <https://doi.org/10.1080/01480545.2021.1991775> (2022).
12. Sun, D. et al. Long-Term exposure to fine particulate matter and incidence of esophageal cancer: A prospective study of 0.5 million Chinese adults. *Gastroenterology* **165**, 61–70e65. <https://doi.org/10.1053/j.gastro.2023.03.233> (2023).
13. Li, P. et al. The associations of air pollution and socioeconomic factors with esophageal cancer in China based on a Spatiotemporal analysis. *Environ. Res.* **196**, 110415. <https://doi.org/10.1016/j.envres.2020.110415> (2021).
14. Ruano-Ravina, A., Aragónés, N., Pérez-Ríos, M., López-Abente, G. & Barros-Dios, J. M. Residential radon exposure and esophageal cancer. An ecological study from an area with high indoor radon concentration (Galicia, Spain). *Int. J. Radiat. Biol.* **90**, 299–305. <https://doi.org/10.3109/09553002.2014.886792> (2014).
15. Gallo, K., Goede, A., Preissner, R. & Gohlke, B. O. SuperPred 3.0: drug classification and target prediction-a machine learning approach. *Nucleic Acids Res.* **50**, W726–w731. <https://doi.org/10.1093/nar/gkac297> (2022).
16. Davis, A. P. et al. Comparative toxicogenomics database (CTD): update 2023. *Nucleic Acids Res.* **51**, D1257–D1262. <https://doi.org/10.1093/nar/gkac833> (2023).
17. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–d646. <https://doi.org/10.1093/nar/gkac1000> (2023).
18. 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**, D672–d677. <https://doi.org/10.1093/nar/gkac909> (2025).
19. Yuan, H. et al. CancerSEA: a cancer single-cell state atlas. *Nucleic Acids Res.* **47**, D900–d908. <https://doi.org/10.1093/nar/gky939> (2019).
20. Sonthithai, P. et al. Perfluorinated chemicals, PFOS and PFOA, enhance the estrogenic effects of 17 β -estradiol in T47D human breast cancer cells. *J. Appl. Toxicol. JAT*. **36**, 790–801. <https://doi.org/10.1002/jat.3210> (2016).
21. Mao, Z. et al. Perfluorooctane sulfonate induces apoptosis in lung cancer A549 cells through reactive oxygen species-mediated mitochondrion-dependent pathway. *J. Appl. Toxicol. JAT*. **33**, 1268–1276. <https://doi.org/10.1002/jat.2785> (2013).
22. Chen, T. et al. Prenatal PFOS exposure induces oxidative stress and apoptosis in the lung of rat off-spring. *Reprod. Toxicol. (Elsevier N.Y.)*. **33**, 538–545. <https://doi.org/10.1016/j.reprotox.2011.03.003> (2012).
23. Zhao, H., Chen, W., Zhu, Y. & Lou, J. Hypoxia promotes pancreatic cancer cell migration, invasion, and epithelial-mesenchymal transition via modulating the FOXO3a/DUSP6/ERK axis. *J. Gastrointest. Oncol.* **12**, 1691–1703. <https://doi.org/10.21037/jgo-21-359> (2021).
24. Liu, Y. C., Yeh, C. T. & Lin, K. H. Molecular functions of thyroid hormone signaling in regulation of cancer progression and Anti-Apoptosis. *Int. J. Mol. Sci.* **20**. <https://doi.org/10.3390/ijms20204986> (2019).
25. Takeuchi, M., Miyoshi, H. & Ohshima, K. Tumor microenvironment of adult T-cell leukemia/lymphoma. *J. Clin. Exp. Hematopathol. JCEH*. **61**, 202–209. <https://doi.org/10.3960/jslrr.21007> (2021).
26. Kadonosono, T. et al. AGE/RAGE axis regulates reversible transition to quiescent States of ALK-rearranged NSCLC and pancreatic cancer cells in monolayer cultures. *Sci. Rep.* **12**, 9886. <https://doi.org/10.1038/s41598-022-14272-0> (2022).
27. Yang, Y. et al. Thyroid hormone suppresses Medulloblastoma progression through promoting terminal differentiation of tumor cells. *Cancer Cell*. **42**, 1434–1449e1435. <https://doi.org/10.1016/j.ccell.2024.07.008> (2024).
28. Hemmatzadeh, M., Mohammadi, H., Karimi, M., Musavishenas, M. H. & Baradaran, B. Differential role of MicroRNAs in the pathogenesis and treatment of esophageal cancer. *Biomed. pharmacotherapy = Biomedecine Pharmacotherapie*. **82**, 509–519. <https://doi.org/10.1016/j.biopha.2016.05.009> (2016).
29. Ai, C. et al. FOXM1 functions collaboratively with PLAU to promote gastric cancer progression. *J. Cancer*. **11**, 788–794. <https://doi.org/10.7150/jca.37323> (2020).
30. Gao, Y., Ma, X., Lu, H., Xu, P. & Xu, C. PLAU is associated with cell migration and invasion and is regulated by transcription factor YY1 in cervical cancer. *Oncol. Rep.* **49**. <https://doi.org/10.3892/or.2022.8462> (2023).
31. Lin, M. et al. MicroRNA-193a-3p suppresses the colorectal cancer cell proliferation and progression through downregulating the PLAU expression. *Cancer Manage. Res.* **11**, 5353–5363. <https://doi.org/10.2147/cmar.s208233> (2019).
32. Hosen, S. M. Z. et al. Metastatic phenotype and immunosuppressive tumour microenvironment in pancreatic ductal adenocarcinoma: key role of the urokinase plasminogen activator (PLAU). *Front. Immunol.* **13**, 1060957. <https://doi.org/10.3389/fimmu.2022.1060957> (2022).
33. Järvinen, T. A. & Liu, E. T. Topoisomerase α gene (TOP2A) amplification and deletion in cancer—more common than anticipated. *Cytopathology: Official J. Br. Soc. Clin. Cytol.* **14**, 309–313. <https://doi.org/10.1046/j.0956-5507.2003.00105.x> (2003).
34. Uusküla-Reimand, L. & Wilson, M. D. Untangling the roles of TOP2A and TOP2B in transcription and cancer. *Sci. Adv.* **8**, eadd4920. <https://doi.org/10.1126/sciadv.add4920> (2022).
35. Song, L. et al. DNA repair and replication proteins as prognostic markers in melanoma. *Histopathology* **62**, 343–350. <https://doi.org/10.1111/j.1365-2559.2012.04362.x> (2013).
36. Kou, F. et al. TOP2A promotes lung adenocarcinoma cells' malignant progression and predicts poor prognosis in lung adenocarcinoma. *J. Cancer*. **11**, 2496–2508. <https://doi.org/10.7150/jca.41415> (2020).
37. Cai, H., Shao, B., Zhou, Y. & Chen, Z. High expression of TOP2A in hepatocellular carcinoma is associated with disease progression and poor prognosis. *Oncol. Lett.* **20**, 232. <https://doi.org/10.3892/ol.2020.12095> (2020).
38. Wang, X. Q., Li, L. L., Lou, P., Zhang, Q. & Wang, Y. Noval CeRNA axis-mediated high expression of TOP2A correlates with poor prognosis and tumor immune infiltration of hepatocellular carcinoma. *Translational Cancer Res.* **12**, 3486–3502. <https://doi.org/10.21037/tcr-23-755> (2023).
39. Xu, D., Li, C., Wen, Y. & Liu, W. Antioxidant defense system responses and DNA damage of earthworms exposed to perfluorooctane sulfonate (PFOS). *Environ. Pollution (Barking Essex: 1987)*. **174**, 121–127. <https://doi.org/10.1016/j.envpol.2012.10.030> (2013).
40. Xiaopeng, C. & Jin, T. Perfluorooctane sulfonate (PFOS) causes aging damage in the liver through the mt-DNA-mediated NLRP3 signaling pathway. *Ecotoxicol. Environ. Saf.* **262**, 115121. <https://doi.org/10.1016/j.ecoenv.2023.115121> (2023).
41. Rucker, E. B. et al. Bcl-x and Bax regulate mouse primordial germ cell survival and apoptosis during embryogenesis. *Mol. Endocrinol. (Baltimore Md)*. **14**, 1038–1052. <https://doi.org/10.1210/mend.14.7.0465> (2000).
42. Pravdic, Z. et al. The influence of BCL2, BAX, and ABCB1 gene expression on prognosis of adult de Novo acute myeloid leukemia with normal karyotype patients. *Radiol. Oncol.* **57**, 239–248. <https://doi.org/10.2478/raon-2023-0017> (2023).
43. Ito, T. et al. Decreased expression of Bax is correlated with poor prognosis in oral and oropharyngeal carcinoma. *Cancer Lett.* **140**, 81–91. [https://doi.org/10.1016/s0304-3835\(99\)00055-5](https://doi.org/10.1016/s0304-3835(99)00055-5) (1999).

44. Durham, J., Tessmann, J. W., Deng, P., Hennig, B. & Zaytseva, Y. Y. The role of perfluorooctane sulfonic acid (PFOS) exposure in inflammation of intestinal tissues and intestinal carcinogenesis. *Front. Toxicol.* **5**, 1244457. <https://doi.org/10.3389/ftox.2023.1244457> (2023).
45. Qazi, M. R., Nelson, B. D., Depierre, J. W. & Abedi-Valugerdi, M. 28-Day dietary exposure of mice to a low total dose (7 mg/kg) of perfluorooctanesulfonate (PFOS) alters neither the cellular compositions of the thymus and spleen nor humoral immune responses: does the route of administration play a pivotal role in PFOS-induced immunotoxicity? *Toxicology* **267**, 132–139. <https://doi.org/10.1016/j.tox.2009.10.035> (2010).

Author contributions

Fangyi Liu authored the manuscript, Yansong Lin designed the study, Lixuan Liu and Yinli Zheng analyzed the data and created the figures, while Jingping Yun reviewed and edited the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval and consent to participate

This study adhered to Helsinki Declaration guidelines. Bioinformatics analyses of data came from publicly available sources with prior institutional ethical approval.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-15765-4>.

Correspondence and requests for materials should be addressed to J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025