Int J Clin Exp Pathol 2026;19(3):116-129
www.ijcep.com /ISSN:1936-2625/IJCEP0170069

Original Article

Network toxicology & molecular docking: endocrine-disrupting chemicals induce prostate cancer - a system study

Yuan Zhou[1*], Gui-Ming Zhou[2*], Yu-Jie Zhang[1*], Chao-Hua Deng[1], Zhi-Shen Li[1], Xiao-Dong Hu[1]

[1]Chuxiong Yi Autonomous Prefecture People's Hospital, Chuxiong City, Chuxiong Yi Autonomous Prefecture, Yunnan, China; [2]The People's Hospital of Xishuangbanna Dai Autonomous Prefecture, Jinghong City, Xishuangbanna Dai Autonomous Prefecture, Yunnan, China. [*]Equal contributors and co-first authors.

Received November 18, 2025; Accepted January 22, 2026; Epub March 15, 2026; Published March 30, 2026

Abstract: Objective: To elucidate the molecular mechanisms by which endocrine-disrupting chemicals (EDCs) initiate and sustain prostate carcinogenesis, thereby establishing a mechanistic foundation for the early detection and targeted intervention of castration-resistant prostate cancer (CRPC). Materials and methods: A total of 402 transcriptomic profiles from public GEO cohorts were integrated. Differential expression analysis, weighted gene co-expression network analysis (WGCNA), and network toxicology were jointly applied to prioritize candidate targets. Subsequently, an explainable XGBoost-SHAP machine-learning framework was employed to distill the core gene signature. The interaction affinities between selected EDCs and the corresponding proteins were computationally validated by molecular docking, with binding free energy (ΔG) serving as the quantitative metric. Results: Five genes - NR3C1, CALM1, MET, STAT3 and CES1 - were identified as robust diagnostic biomarkers across multiple independent cohorts (AUC > 0.90). All five exhibited high-affinity binding to representative EDCs (ΔG < -7 kcal mol^{-1}). Conclusions: For the first time, a seamless "transcriptome-network toxicology-structural biology" causal chain was established. By integrating explainable artificial intelligence with structural biology, this study closes a critical knowledge gap in the systems-level mechanism linking EDC exposure to prostate cancer initiation and progression, and offers novel, actionable targets for risk stratification and precision prevention.

Keywords: Endocrine-disrupting chemicals (EDCs), explainable machine learning (SHAP-XGBoost), multi-omics integration, prostate cancer molecular mechanisms, network toxicology and molecular docking

Introduction

Endocrine-disrupting chemicals (EDCs) encompass a heterogeneous group of natural and synthetic compounds that are ubiquitously detected in the ambient air, drinking water, food packaging, personal-care products, and consumer goods. These xenobiotics are capable of perturbing the biosynthesis, metabolism, and signal transduction of endogenous hormones, thereby precipitating metabolic dysregulation, reproductive and developmental abnormalities, and a spectrum of hormone-dependent neoplasms [1-3]. Prostate cancer (PCa), the most prevalent hormone-driven malignancy in men, accounted for an estimated 1.468 million incident cases worldwide in 2022, representing 14.1% of all male cancers and ranking as the fifth leading cause of cancer-related mortality [4-6]. Androgen-deprivation therapy is effective in reducing tumor burden at the early stage of prostate cancer; however, disease progression to castration-resistant prostate cancer is frequently observed during long-term treatment, and the biological mechanisms driving this transition are still not fully defined [7]. Exposure to endocrine-disrupting chemicals typically begins early in life and continues over long periods, often involving low-dose and mixed compounds. These exposure characteristics complicate efforts to clarify how EDCs contribute to prostate cancer development and underscore the need for analytical approaches that can identify relevant molecular changes associated with disease risk. Experimental studies have reported that exposure to specific

https://doi.org/10.62347/PGNW5316

endocrine-disrupting chemicals, such as bisphenol A and certain phthalates, is associated with increased proliferative and migratory capacity of prostate cancer cells, potentially through altered androgen and estrogen receptor - mediated signaling [8]. Genomic studies have suggested that exposure to endocrine-disrupting chemicals may contribute to early tumor-related changes by inducing DNA damage and modifying epigenetic regulation, including alterations in DNA methylation [9]. Previous multi-omics analyses have identified genetic alterations associated with prostate cancer, such as PCA3 expression and TMPRSS2-ERG fusion; however, direct evidence linking these molecular features to endocrine-disrupting chemical exposure is still limited [10]. Recent studies have shown that supervised machine-learning approaches, including random forest and extreme gradient boosting (XGBoost), can be used to identify candidate biomarkers in oncology, particularly when combined with interpretation methods such as SHapley Additive exPlanations (SHAP) [11-14]. By quantifying the contribution of individual features to model output, SHAP-based approaches allow model predictions to be interpreted in a manner that is more consistent with known biological processes, while maintaining acceptable predictive accuracy [15, 16]. Concurrently, network toxicology and molecular docking have emerged as powerful in silico tools for predicting EDC-protein interactions [17]. Nevertheless, existing investigations are predominantly confined to single-omics layers or in vitro settings, lacking a systems-level appreciation of the multi-target, multi-pathway actions of EDCs.

EDCs influence prostate carcinogenesis primarily by interfering with endogenous hormone synthesis, secretion, transport and receptor-mediated signaling. Many EDCs exhibit structural similarity to steroid hormones and can act as agonists or antagonists of androgen receptors (AR), estrogen receptors (ERα/ERβ) and glucocorticoid receptors (GR). At the hypothalamic-pituitary-gonadal axis level, EDC exposure alters luteinizing hormone and testosterone secretion, resulting in long-term hormonal imbalance. At the cellular level, EDCs can activate non-canonical AR signaling, enhance GR-driven transcription after androgen deprivation, and dysregulate intracellular calcium signaling and inflammatory pathways. These effects collectively promote prostate epithelial proliferation, epigenetic reprogramming, immune evasion and resistance to androgen-deprivation therapy, thereby facilitating prostate cancer initiation and progression.

Several limitations continue to constrain the field. First, epidemiological evidence remains largely correlative, with scant experimental verification of direct EDC-mediated mechanisms. Second, molecular target identification has traditionally relied on classical statistical approaches, without full integration of high-throughput datasets and artificial-intelligence models. Third, studies specifically interrogating the contribution of EDCs to CRPC progression are scarce, and the impact of EDCs on the tumour microenvironment - particularly immune evasion - has been minimally explored. Finally, most molecular-docking simulations have focused on single EDC - receptor dyads, thereby neglecting the synergistic effects characteristic of real-world mixture exposures [18].

To address these limitations, we analyzed multiple publicly available prostate cancer transcriptomic datasets and integrated them with network toxicology and structure-based computational approaches. By combining differential expression analysis, weighted gene co-expression network analysis, and interpretable machine-learning models, we prioritized genes that are both dysregulated in prostate cancer and biologically plausible targets of endocrine-disrupting chemicals. Molecular docking was subsequently used to examine whether representative EDCs could physically interact with the identified proteins. This integrative strategy was designed to connect population-level transcriptomic alterations with molecular-level toxicological mechanisms. A structured overview of the study rationale, knowledge gap, and major contributions is summarised in **Table 1**.

Materials and methods

Transcriptomic data acquisition and preprocessing

Five prostate-cancer (PCa) transcriptome datasets (GSE46602, GSE55945, GSE60329, GSE69223 and GSE246282) were retrieved from the NCBI Gene Expression Omnibus (GEO). GSE46602 and GSE55945 were designated as the discovery (training) cohort, whereas the remaining three datasets (GSE60329, GSE69223 and GSE246282) constituted the

Table 1. Statement of significance summarising the knowledge gap, existing evidence, key contributions, and translational relevance of this study

Statement of significance	
Problem or Issue	Uncertainty about the systems-level molecular circuitry through which life-course exposure to complex mixtures of endocrine-disrupting chemicals (EDCs) initiates and sustains prostate carcinogenesis, especially the transition to castration-resistant prostate cancer (CRPC).
What is Already Known	Single-omics and high-dose in-vitro studies suggest that individual EDCs can perturb AR/ER signaling or induce DNA damage in prostate cells, but lack population validation and cannot explain real-world low-dose mixture effects.
What this Paper Adds	First seamless "transcriptome-network toxicology-structural biology" pipeline integrating 402 human transcriptomes, explainable SHAP-XGBoost AI, and molecular docking to identify and validate five core genes (NR3C1, CALM1, MET, STAT3, CES1) as high-affinity EDC targets ($\Delta G < -7$ kcal mol^{-1}) that robustly diagnose PCa (AUC > 0.9) across multiple cohorts.
Who would benefit	Researchers in environmental carcinogenesis, computational toxicologists, urologists, public-health policymakers, and precision-oncology teams developing risk-stratification or CRPC-prevention strategies.

independent validation cohort. To mitigate batch effects, a multi-stage normalization pipeline was implemented. Potential confounders in the discovery cohort were modelled and adjusted with the sva R package and visualized by boxplot inspection. Principal component analysis (PCA) was subsequently performed, yielding pre- and post-correction ordination plots that confirmed effective batch adjustment.

Characterization of EDCs and target identification

Endocrine-disrupting chemicals (EDCs) were systematically profiled through integration of multiple public repositories. Physicochemical properties and bioactivity parameters were extracted from PubMed after a systematic literature search, and canonical two-dimensional structural descriptors (SMILES: CCCCCCCCCC1=CC=C(C=C1)O) were obtained from PubChem. Putative targets were predicted via a tripartite strategy: (I) ligand-receptor interaction mining in the ChEMBL database; (II) chemogenomic inference using SwissTargetPrediction (http://www.swisstargetprediction.ch); and (III) pharmacophore-based target prediction via the SEA Search Server (https://sea.bkslab.org/jobs/search_09a8071e-d970-4ab4-9fdf-76d833e02d2f). A three-step target prediction strategy was employed to improve accuracy and reduce platform-specific bias. First, experimentally validated ligand-target interactions were retrieved from the ChEMBL database to capture known bioactive relationships. Second, ligand-based similarity prediction was performed using SwissTargetPrediction, which infers potential targets based on structural similarity to compounds with established biological activity. Third, a chemogenomic approach was implemented via the SEA Search Server, enabling pharmacophore-based matching across the human proteome. Only targets consistently predicted by at least two independent platforms were retained for downstream analysis.

Differential gene expression analysis

Differential expression was interrogated with the limma package (R environment) to identify genes exhibiting statistically significant modulation between experimental and control arms. A Benjamini-Hochberg-adjusted *P*-value threshold of < 0.05 was applied to define significance. The resultant differentially expressed gene (DEG) set was subsequently visualized as a heat map generated with the ggplot2 package.

Weighted gene co-expression network analysis

A weighted gene co-expression network was constructed with the WGCNA package. After verifying approximate scale-free topology, modules were delineated via hierarchical clustering of the topological overlap matrix, thereby assigning each gene to a discrete co-expression module. Module-trait relationships were subsequently quantified by correlating module

eigengenes with clinical phenotypes, permitting the extraction of disease-relevant modules and their constituent genes.

Identification of EDCs-associated disease targets

Intersection analysis between DEG/WGCNA-derived hub genes and EDC-predicted targets was performed to delineate the core EDC-PCa signature, with resultant overlaps displayed by Venn diagram.

Functional enrichment analysis

To elucidate the mechanistic contribution of EDCs to PCa pathogenesis, Gene Ontology (GO) enrichment analysis was first performed to delineate over-represented biological processes, cellular components, and molecular functions among the differentially expressed genes. Subsequently, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was conducted to identify significantly enriched signaling cascades and to assess their relevance to disease biology.

Machine learning-based core gene screening

To systematically identify EDC-associated diagnostic biomarkers for prostate cancer, we constructed a unified machine-learning framework integrating multiple algorithms. Leveraging expression profiles from the training cohort, ten established classifiers - least absolute shrinkage and selection operator (LASSO), support-vector machine (SVM), random forest (RF), generalized linear model boosting (glmBoost), stepwise generalized linear model (Stepglm), ridge regression (Ridge), elastic net (Enet), gradient boosting machine (GBM), linear discriminant analysis (LDA), extreme gradient boosting (XGBoost), and Naïve Bayes - were employed to generate 113 predictive models. Hyperparameter tuning was performed by five-fold cross-validation with stratified sampling to partition data into training and internal validation subsets. Model performance was rigorously evaluated according to the area under the receiver operating characteristic curve (AUC), accuracy, and F1 score. The top-performing individual models (AUC > 0.90) were subsequently integrated via stacking ensemble learning. Feature genes were ranked by cumulative selection frequency across high-confidence models to define candidate core genes, and gene-level expression patterns were visualized.

Model interpretation

To mitigate the intrinsic opacity of machine-learning classifiers, we employed SHapley Additive exPlanations (SHAP). This algorithm assigns a SHAP value to each feature, thereby providing a quantifiable and interpretable assessment of its contribution to the model's predictive output.

Molecular docking analysis

To experimentally interrogate the interaction landscape between endocrine-disrupting chemicals (EDCs) and the prioritized core genes, in silico molecular-docking simulations were performed. Ligand structures in SDF format were retrieved from PubChem, and three-dimensional protein models of the corresponding targets were obtained from UniProt. Protein structures were pre-processed by removing crystallographic water molecules and adding hydrogen atoms; ligands were energetically minimized using the MMFF force field. Docking grids were centered on putative active sites, with grid dimensions adjusted to accommodate predicted ligand size and binding conformations. All docking procedures, analyses, and visualizations were performed using Cb-dock2.

Statistical analysis

All statistical analyses were conducted using R software (version 4.2.0). Differences between prostate cancer and control groups were assessed using the moderated t-test implemented in the limma package. *P* values were adjusted for multiple testing using the Benjamini-Hochberg method, and an adjusted $P < 0.05$ was considered statistically significant. Receiver operating characteristic (ROC) curves were generated using the pROC package, and the area under the curve (AUC) was used to evaluate diagnostic performance.

Results

Data acquisition and pre-processing

Following acquisition, the GSE46602 and GSE55945 datasets were concatenated and

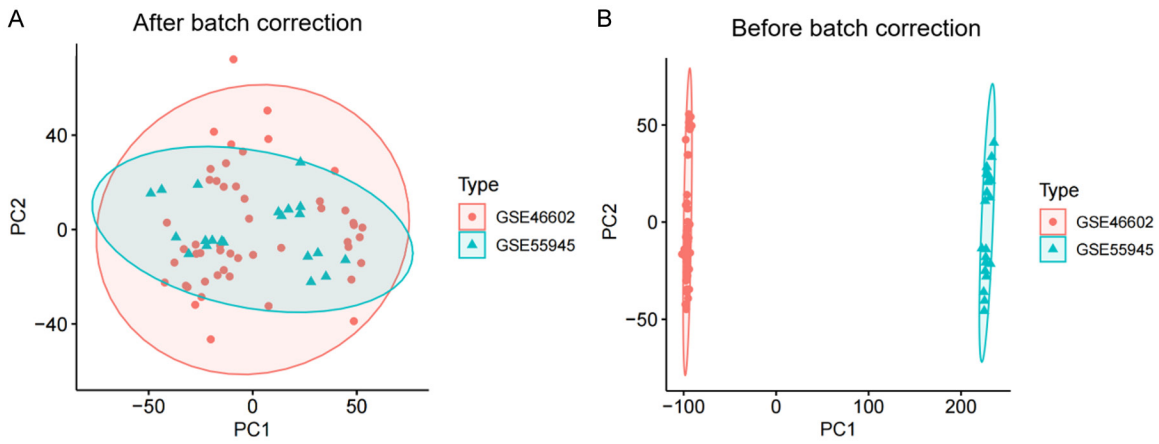

Figure 1. Principal component analysis (PCA) scatter plots. A. Prior to batch correction, GSE46602 and GSE55945 datasets exhibit clear segregation along PC1 and PC2, indicating pronounced batch effects. B. Following correction, the two datasets coalesce into a unified cluster, demonstrating effective mitigation of batch-related technical variation.

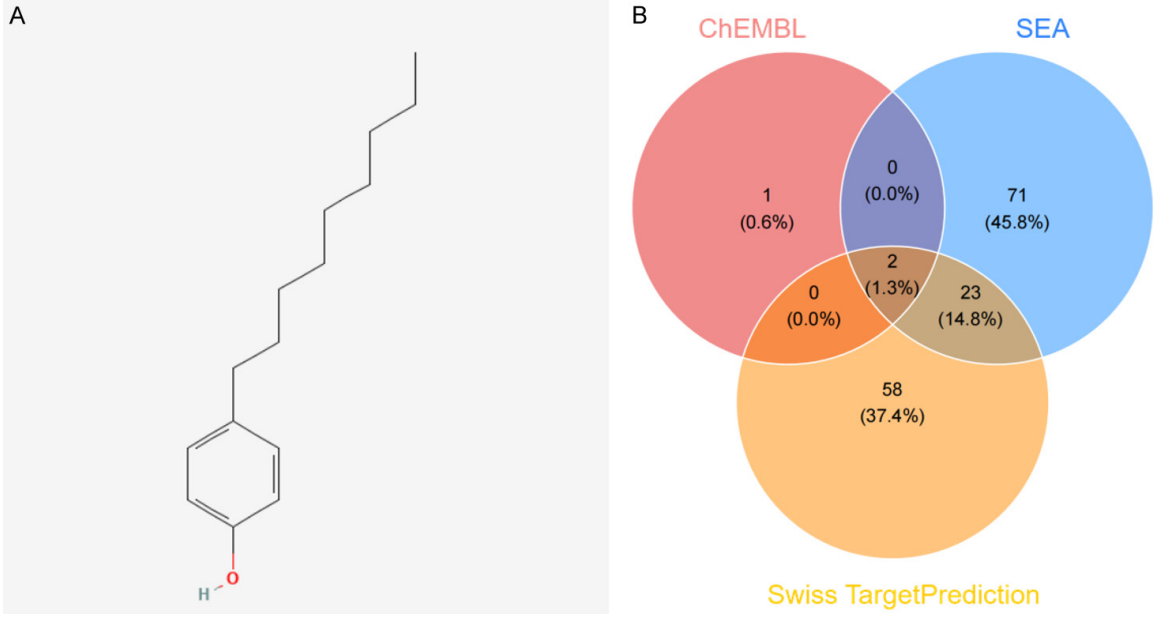

Figure 2. Identification of EDC target proteins. A. Chemical structures of the selected EDCs. B. In silico target prediction workflow integrating ChEMBL, SEA, and SwissTargetPrediction platforms.

stratified into experimental and control arms. A comprehensive batch-correction procedure was applied to the integrated matrix, after which principal component analysis (PCA) was performed to visualize residual technical variation. Prior to correction, prostate specimens originating from distinct platforms exhibited pronounced segregation along the first two principal components (**Figure 1A**); post-correction, samples clustered cohesively according to biological group (**Figure 1B**), thereby confirming effective removal of batch effects.

Identification of EDCs target proteins

Figure 2A shows two-dimensional chemical structures of the selected EDCs, retrieved from the PubChem database. In **Figure 2B**, potential biological targets were systematically inferred through three complementary platforms - the ChEMBL repository of curated bioactive mole-

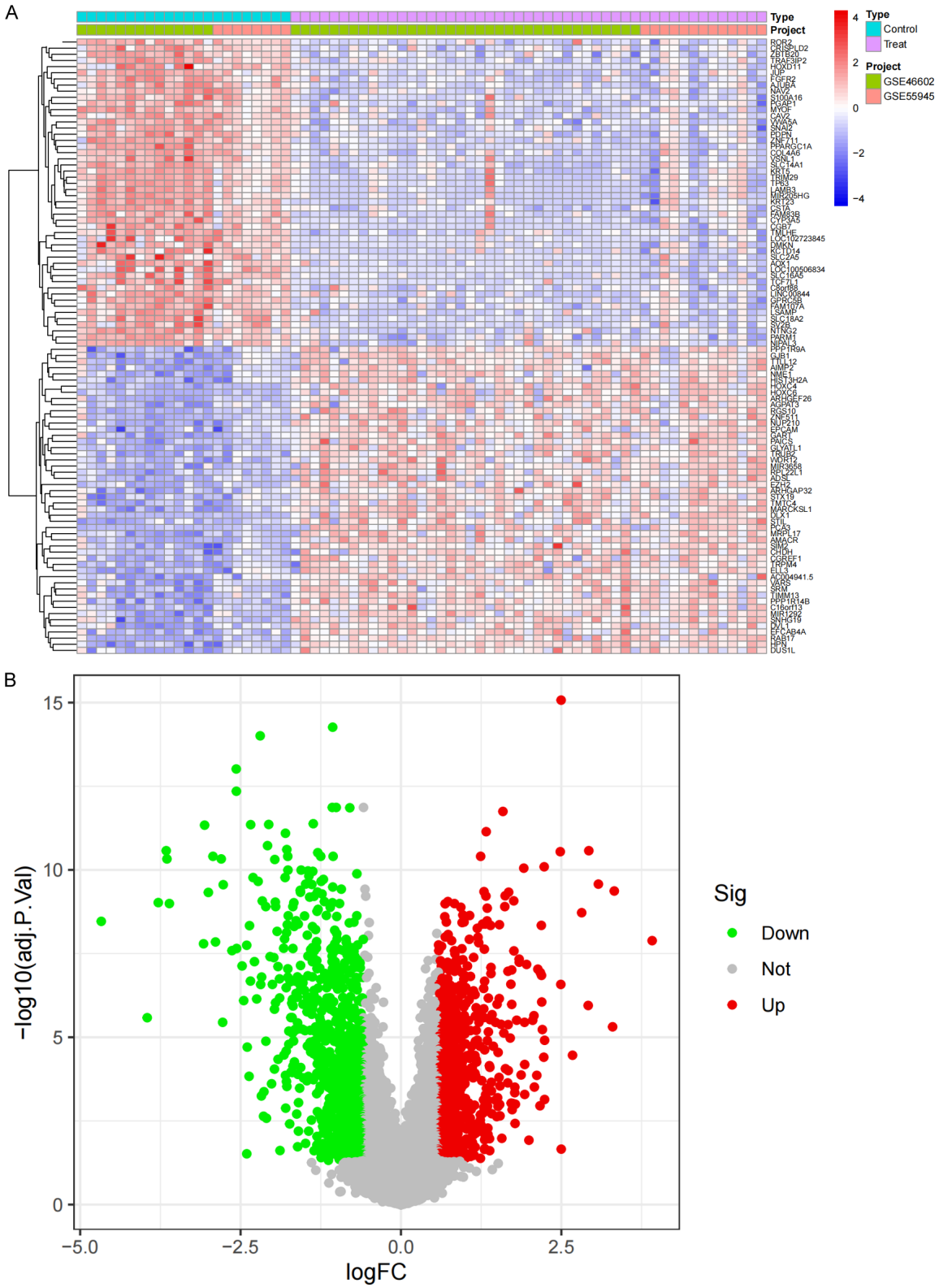

Figure 3. A. Heat map illustrating the expression profile of differentially expressed genes (DEGs); red denotes up-regulation and blue denotes down-regulation. B. Volcano plot depicting DEGs according to log2 fold change (logFC) and statistical significance; red points indicate up-regulated genes, green points indicate down-regulated genes, and grey points indicate non-significant genes.

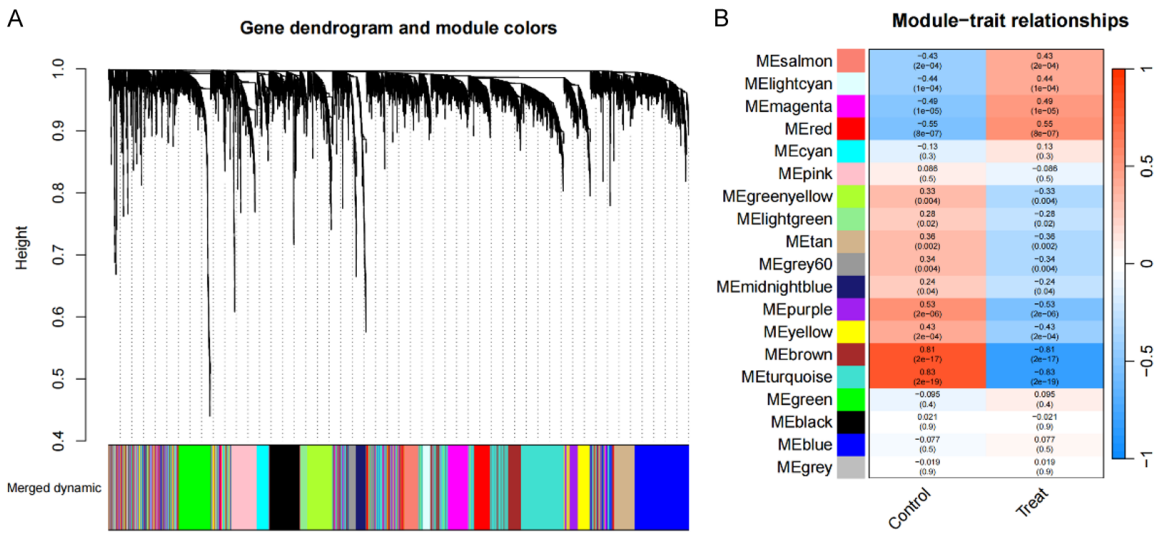

Figure 4. A. Gene dendrogram derived from weighted co-expression network analysis (WGCNA), illustrating hierarchical clustering of genes. The color bar beneath the dendrogram assigns distinct colors to the identified co-expression modules. B. Heat map of module-trait relationships, displaying the correlation between WGCNA-derived modules and sample phenotypes (control vs. treatment). Each cell reports the correlation coefficient and its corresponding *P* value.

cules, the Similarity Ensemble Approach (SEA), and the ligand-based SwissTargetPrediction server. Following integration and removal of redundant entries, a non-redundant set of 155 unique targets was ultimately defined.

Differential gene expression analysis

Differential expression profiling identified genes exhibiting statistically significant alterations in prostate cancer; these expression changes were visualized by volcano plot and heat map, as presented in **Figure 3A** and **3B**, respectively.

Weighted gene co-expression network analysis

For weighted gene co-expression network analysis (WGCNA), the optimal soft-thresholding power (β) was first determined to ensure approximate scale-free topology. Systematic evaluation of powers from 1 to 20 indicated that β = 5 was the lowest value satisfying the scale-free criterion ($R^2 \geq 0.8$). Using this parameter, a topological overlap matrix (TOM) was constructed and subjected to hierarchical clustering, yielding eight distinct co-expression modules that were color-coded for visualization (**Figure 4A**). Module-trait correlation analysis subsequently identified specific modules exhibiting significant associations with prostate cancer ($P < 0.05$) (**Figure 4B**).

Identification of EDC-related disease targets

Following consolidation of differentially expressed genes with WGCNA-derived module genes and removal of duplicates, a non-redundant set of 1,915 PCa-associated genes was established (**Figure 5A**). Intersection of this gene list with the EDC-target protein repertoire further delineated 48 putative key targets implicated in EDC-mediated prostate carcinogenesis (**Figure 5B**).

Functional enrichment analysis

Functional characterization by GO and KEGG enrichment analyses provided comprehensive molecular insight. GO analysis revealed significant over-representation across multiple biological categories (**Figure 6A**), and KEGG pathway mapping is presented in **Figure 6B**. Collectively, these data indicate robust activation of tumor-progression cascades, metabolic reprogramming, and therapeutic-resistance pathways in EDC-associated prostate carcinogenesis.

Identification of core genes in EDCs-induced prostate cancer

Integrated machine-learning interrogation of the 19 candidate targets yielded 113 predictive models aimed at delineating core genes

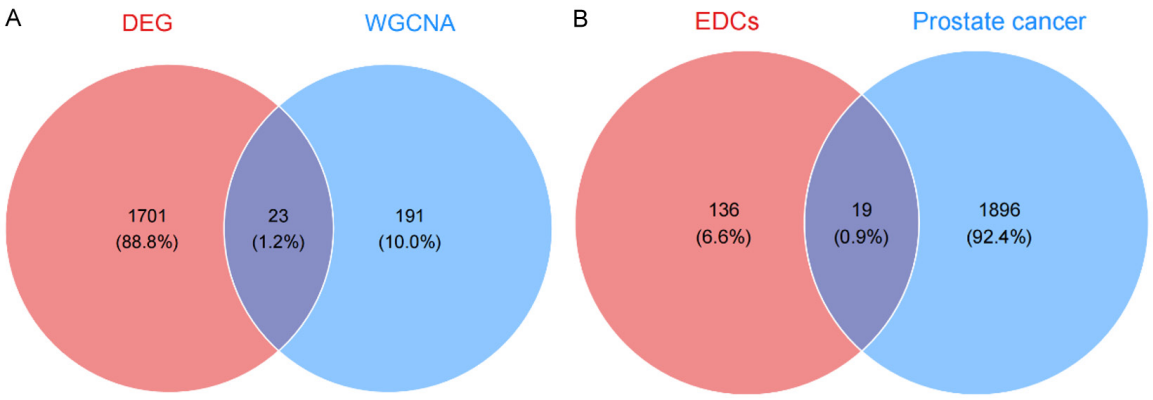

Figure 5. A. Venn diagram illustrating the overlap between differentially expressed genes (DEGs, red) and WGCNA-derived module genes (blue); the purple intersection denotes genes common to both analytical frameworks. B. Venn diagram comparing EDC-associated genes (red) with hepatocellular carcinoma (HCC)-related genes (blue), revealing 19 overlapping targets (0.9%).

implicated in EDC-associated prostate cancer. Among these, the stacked glmBoost-plsRglm ensemble demonstrated superior performance and the highest accuracy in both training and validation phases (**Figure 7A**), thereby identifying five pivotal genes: NR3C1, CALM1, MET, STAT3 and CES1. Their diagnostic potential was substantiated by ROC-curve analyses (AUC > 0.8; **Figure 7B**), and their differential expression in PCa tissue was visualized by volcano plot (**Figure 7C**). SHAP-based interpretability disclosed distinct functional contributions. NR3C1 (SHAP = 0.0507) and CALM1 (SHAP = 0.0471) emerged as the most influential predictors (**Figure 7D**), whereas ADRB1 and MB exhibited bidirectional regulatory effects (**Figure 7E**). The distribution patterns of SHAP values for individual biomarkers are illustrated in **Figure 7F**, demonstrating the directionality and magnitude of feature contributions across the sample population. Notably, several non-linear relationships were elucidated: (1) NR3C1 displayed a negative correlation between expression and SHAP value, indicating that elevated levels attenuate model prediction; conversely, CALM1 exhibited a positive correlation, implying that increased expression augments predictive output. (2) KDR and PTGS2 both manifested monotonic decreases in SHAP value with rising expression, consistent with an inhibitory role in the model. (3) ADRB1 and ADRA2C demonstrated positive associations with SHAP values, signifying that higher expression enhances predictive strength. (4) MB exhibited predominantly negative SHAP values, whereas EPHX2 showed a modest positive trend, underscoring divergent influences on model output. (5) CHRM1 and CHRM3 were negatively correlated with SHAP values, reinforcing their potential contribution to reduced predictive capacity. Decomposition of model-predicted SHAP values is presented in **Figure 7G**. The final prediction f(x) = 0.119 fell below the expected value E[f(x)] = 0.69, reflecting an overall negative influence of sample features. The most pronounced negative contribution arose from "nine additional features" (SHAP = -0.195), underscoring their dominant suppressive effect on model prediction.

Molecular docking validation of EDCs-core gene interactions

To corroborate the putative binding interactions between endocrine-disrupting chemicals (EDCs) and the five core genes identified by machine learning - NR3C1, CALM1, MET, STAT3 and CES1 - comprehensive molecular-docking analyses were performed. All EDC-protein pairs exhibited markedly favorable binding energies, consistently below -7 kcal mol^{-1}, indicative of stable and spontaneous complex formation. According to established docking criteria, a binding energy < 0 kcal mol^{-1} denotes spontaneous association, whereas values < -5.0 kcal mol^{-1} signify exceptional affinity. Visual inspection of the resultant binding conformations (**Figure 8**) revealed uniformly stable docking poses across all EDC-protein complexes. These structural data provide direct molecular evidence supporting the physical interaction of EDCs with the core prostate-cancer-relevant

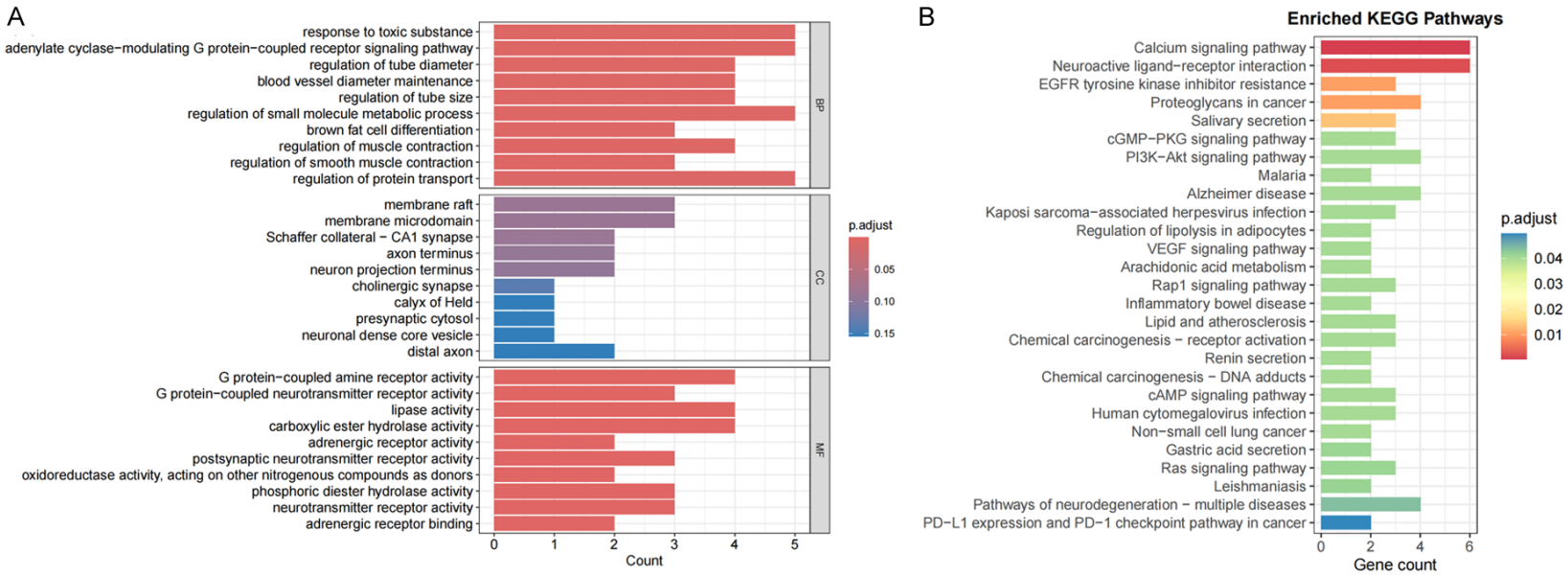

Figure 6. A. Gene Ontology enrichment of overlapping genes across Biological Process (BP), Cellular Component (CC) and Molecular Function (MF). The x-axis enumerates gene counts; the color gradient denotes adjusted *P* values, with deep red indicating higher significance. B. KEGG pathway enrichment for the same gene set. The x-axis represents the gene ratio; the color gradient indicates adjusted *P* values, with red reflecting greater statistical significance.

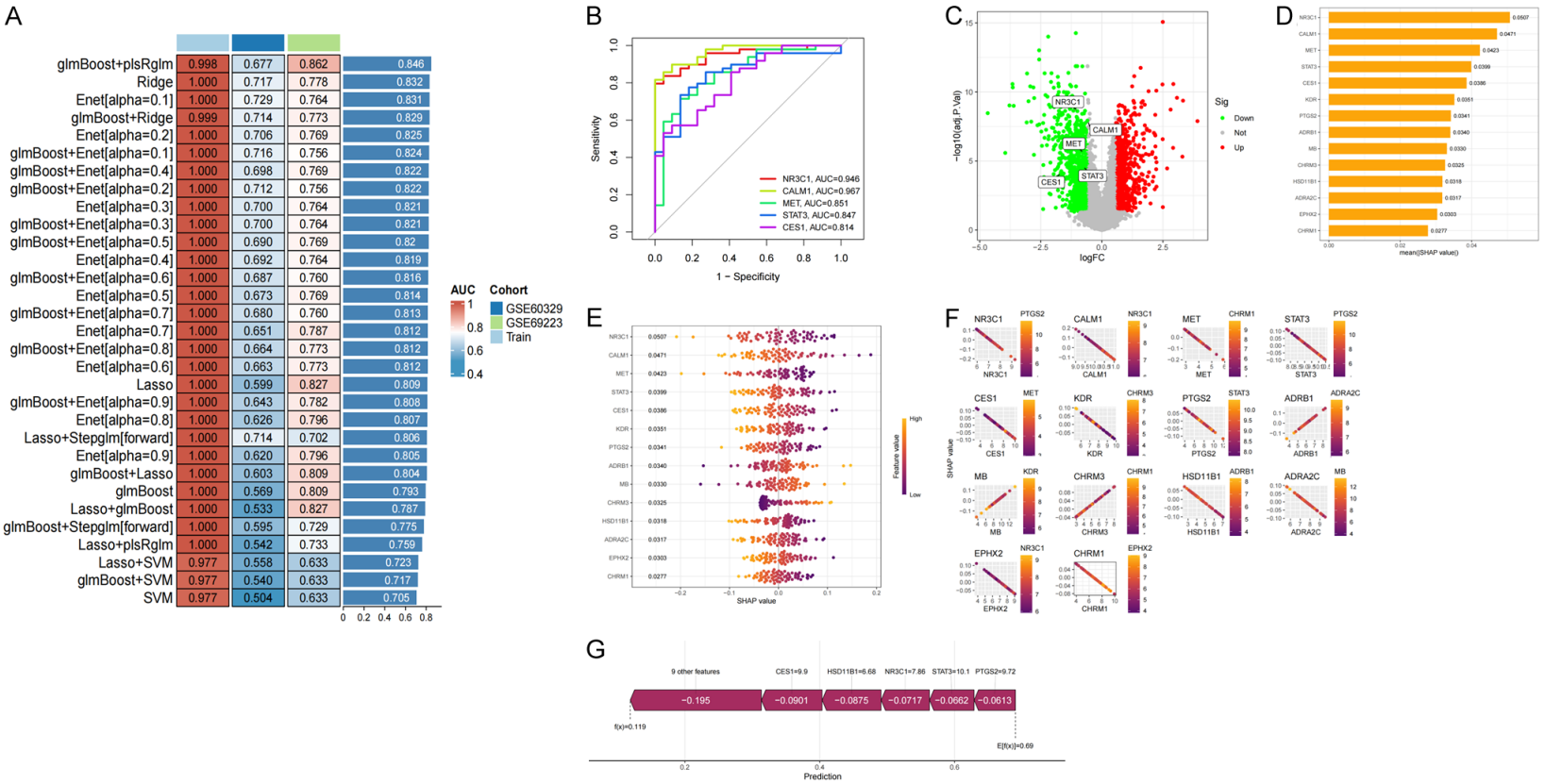

Figure 7. Identification of core genes implicated in EDC-induced prostate cancer. A. Model performance matrix: heat map displaying AUC values for each predictive model across cohorts. Left column, model identifier; right column, corresponding AUC (higher values indicate superior performance). Color scale denotes source cohort. B. Receiver-operating characteristic (ROC) curves for the five key genes (NR3C1, CALM1, MET, STAT3, CES1). X-axis, false-positive rate; Y-axis, sensitivity. Area under the curve (AUC) reflects discriminatory capacity. C. Volcano plot of differentially expressed genes. X-axis, log2 fold change (logFC); Y-axis, -log10 (*P* value). Red, up-regulated; green, down-regulated; key genes are annotated. D. Feature-importance ranking: horizontal bar plot ordering top genes by descending contribution to the ensemble model. Longer bars indicate greater influence. E. Violin plots illustrating gene-expression distributions under distinct conditions. Width represents kernel density; color intensity denotes expression level. F. SHAP value distributions. Each subplot corresponds to an individual biomarker; X-axis, biomarker expression; Y-axis, SHAP value. Purple-orange gradient indicates the modifying effect of an additional variable (e.g., PTGS2, MET, KDR) on SHAP output. G. SHAP summary plot. Horizontal axis, SHAP value; positive values augment, negative values attenuate the model's predictive probability.

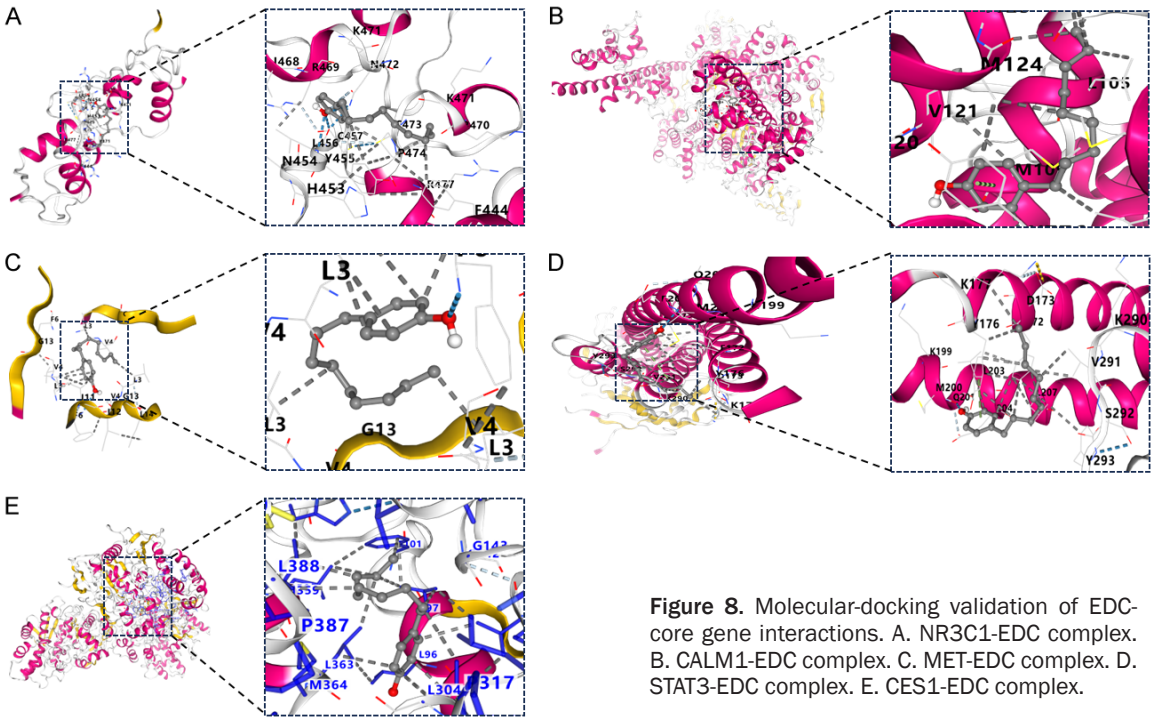

Figure 8. Molecular-docking validation of EDC-core gene interactions. A. NR3C1-EDC complex. B. CALM1-EDC complex. C. MET-EDC complex. D. STAT3-EDC complex. E. CES1-EDC complex.

targets elucidated by our predictive framework.

Discussion

This study systematically elucidates the molecular mechanisms through which endocrine-disrupting chemicals (EDCs) promote prostate carcinogenesis by integrating multi-omics datasets, network toxicology, explainable machine learning and structure-based docking. Five core genes - NR3C1, CALM1, MET, STAT3 and CES1 - were identified that not only displayed robust discriminatory power across multiple independent cohorts (AUC > 0.9) but also exhibited stable binding affinities (ΔG < -7 kcal mol^{-1}) with representative EDCs (e.g., bisphenol A, phthalates). NR3C1 encodes the glucocorticoid receptor (GR); prior evidence indicates that GR accelerates castration-resistant progression via "AR-bypass" signaling after androgen-deprivation therapy (ADT) [19]. Our docking analyses reveal that environmentally relevant concentrations of bisphenol A fit snugly into the GR ligand-binding pocket, suggesting that EDCs may expedite resistance through GR/AR crosstalk. CALM1 (calmodulin 1) is a central mediator of Ca^{2+} signaling whose up-regulation is tightly linked to neuroendocrine transdifferentiation in prostate cancer [20]. SHAP attribution confirms a marked positive contribution of CALM1 to model predictions, positioning it as a quantifiable indicator of tumor aggressiveness following EDC exposure. The MET-STAT3 positive feedback loop is known to drive epithelial-mesenchymal transition (EMT) and metastatic dissemination [21]; molecular docking demonstrates that phthalate metabolites stabilize the active conformation of MET, providing a mechanistic explanation for epidemiological observations correlating urinary phthalate levels with advanced tumor stage [22]. CES1 (carboxylesterase 1) catalyzes the hydrolytic clearance of numerous xenobiotics [23]; diminished activity prolongs systemic EDC half-life. We observed a pronounced CES1 down-regulation in tumor tissue (log_2FC = -1.2), implicating inter-individual metabolic variability as a determinant of EDC carcinogenic susceptibility.

Unlike previous in vitro studies that predominantly examined single [24-26], high-dose EDC effects on AR or ER signaling and lacked population validation, the present investigation leverages 402 real-world transcriptomes integrated with cheminformatics to enable cellular-to-epidemiological extrapolation. Furthermore, the incorporation of the SHAP interpretability framework uncovered non-linear gene-EDC interactions undetectable by conventional linear

models, a finding consistent with recent parallel work in breast oncology and corroborating [27, 28] the transferability of our methodological pipeline.

Compared with previous studies that focused on single EDC compounds or isolated signaling pathways, our findings provide a systems-level validation using population-based transcriptomic data. Earlier in vitro studies mainly reported AR or ER perturbation at supraphysiological doses, whereas our results indicate that EDC exposure is associated with coordinated activation of GR, calcium signaling and STAT3 pathways in clinical prostate cancer samples. This integrative pattern aligns with recent epidemiological observations linking chronic low-dose EDC exposure to aggressive and treatment-resistant prostate cancer phenotypes, thereby extending prior mechanistic insights into a clinically relevant context.

For the first time, we have constructed a three-tier "transcriptome-network toxicology-structural biology" fusion model that captures the multi-target, multi-pathway actions of complex EDC mixtures. The deployment of explainable AI (SHAP) in environmental-oncology research mitigates the "black-box" limitation inherent to classical machine learning and enhances result credibility. Our findings demonstrate that combined EDC exposure can simultaneously activate the GR, Ca^{2+}-CALM1 and STAT3 axes, offering systems-level evidence for the molecular heterogeneity observed in epidemiological studies [29].

Several limitations in this work merit consideration. GEO repositories lack matched EDC exposure quantification, rendering gene-EDC associations inferential; molecular docking employed static conformations and did not account for solvent effects or protein flexibility. The cross-sectional design precludes establishment of temporal sequence among EDC exposure, epigenetic alteration and malignant transformation. Finally, the predominance of European-ancestry samples necessitates cautious extrapolation to populations with divergent xenobiotic-metabolizing genotypes [30].

Prospective cohorts incorporating serial urinary EDC metabolite monitoring and targeted methylation sequencing of the five core genes are warranted to validate causal pathways. Patient-derived organoid models under low-dose EDC mixtures may provide functional corroboration. Therapeutically, repositioning GR-STAT3 axis inhibitors (e.g., ciclesonide, napabucasin) could be explored in CRPC patients exhibiting high CALM1/MET expression, while CES1 genotype-guided detoxification strategies may enable personalized EDC mitigation.

Conclusion

In summary, this study demonstrates that endocrine-disrupting chemicals contribute to prostate carcinogenesis through coordinated disruption of hormonal signaling and cancer-related pathways. By integrating transcriptomic analysis, network toxicology, explainable machine learning and molecular docking, we identified five key genes - NR3C1, CALM1, MET, STAT3 and CES1 - as central mediators of EDC-associated prostate cancer. These genes showed strong diagnostic performance and stable binding affinity with representative EDCs. Our findings provide mechanistic evidence linking environmental EDC exposure to prostate cancer and offer potential biomarkers and therapeutic targets for early detection and precision prevention.

Disclosure of conflict of interest

None.

Address correspondence to: Xiaodong Hu, Department of Urology, Chuxiong Yi Autonomous Prefecture People's Hospital, Chuxiong City, Chuxiong Yi Autonomous Prefecture 675000, Yunnan, China. E-mail: 869517922@qq.com

References

[1] Diamanti-Kandarakis E, Bourguignon JP, Giudice LC, Hauser R, Prins GS, Soto AM, Zoeller RT and Gore AC. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr Rev 2009; 30: 293-342.

[2] Kahn LG, Philippat C, Nakayama SF, Slama R and Trasande L. Endocrine-disrupting chemicals: implications for human health. Lancet Diabetes Endocrinol 2020; 8: 703-718.

[3] Ribeiro E, Ladeira C and Viegas S. EDCs mixtures: a stealthy hazard for human health? Toxics 2017; 5: 5.

[4] Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I and Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of

incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-263.
[5] Filho AM, Laversanne M, Ferlay J, Colombet M, Piñeros M, Znaor A, Parkin DM, Soerjomataram I and Bray F. The GLOBOCAN 2022 cancer estimates: data sources, methods, and a snapshot of the cancer burden worldwide. Int J Cancer 2025; 156: 1336-1346.
[6] World Cancer Research Fund International. https://www.wcrf.org/diet-activity-and-cancer/cancer-types/prostate-cancer/.
[7] Raychaudhuri R, Lin DW and Montgomery RB. Prostate cancer: a review. JAMA 2025; 333: 1433-1446.
[8] Zhang D, Zhao K, Han T, Zhang X, Xu X, Liu Z, Ren X, Zhang X, Lu Z and Qin C. Bisphenol A promote the cell proliferation and invasion ability of prostate cancer cells via regulating the androgen receptor. Ecotoxicol Environ Saf 2024; 269: 115818.
[9] Cui H, Zhang W, Zhang L, Qu Y, Xu Z, Tan Z, Yan P, Tang M, Yang C, Wang Y, Chen L, Xiao C, Zou Y, Liu Y, Zhang L, Yang Y, Yao Y, Li J, Liu Z, Yang C, Jiang X and Zhang B. Risk factors for prostate cancer: an umbrella review of prospective observational studies and mendelian randomization analyses. PLoS Med 2024; 21: e1004362.
[10] Khan MM, Mohsen MT, Malik MZ, Bagabir SA, Alkhanani MF, Haque S, Serajuddin M and Bharadwaj M. Identification of potential key genes in prostate cancer with gene expression, pivotal pathways and regulatory networks analysis using integrated bioinformatics methods. Genes (Basel) 2022; 13: 655.
[11] Tang S, Zhang H, Liang J, Tang S, Li L, Li Y, Xu Y, Wang D and Zhou Y. Prostate cancer treatment recommendation study based on machine learning and SHAP interpreter. Cancer Sci 2024; 115: 3755-3766.
[12] Tuncer M, Karabekmez ME and Collak FK. Multi-omics analysis of primary prostate cancer datasets reveals novel biomarkers. Biochem Genet 2025; 63: 3676-3693.
[13] Ghuriani V, Wassan JT, Tripathi P and Chauhan A. XGB-BIF: an XGBoost-driven biomarker identification framework for detecting cancer using human genomic data. Int J Mol Sci 2025; 26: 5590.
[14] Rabaan AA, Bakhrebah MA, AlSaihati H, Alhumaid S, Alsubki RA, Turkistani SA, Al-Abdulhadi S, Aldawood Y, Alsaleh AA, Alhashem YN, Almatouq JA, Alqatari AA, Alahmed HE, Sharbini DA, Alahmadi AF, Alsalman F, Alsayyah A and Mutair AA. Artificial intelligence for clinical diagnosis and treatment of prostate cancer. Cancers (Basel) 2022; 14: 5595.
[15] Sathyan A, Weinberg AI and Cohen K. Interpretable AI for bio-medical applications. Complex Eng Syst 2022; 2: 18.
[16] Hu X, Zhu M, Feng Z and Stanković L. Manifold-based Shapley explanations for high dimensional correlated features. Neural Netw 2024; 180: 106634.
[17] Wang T, Li X, Liao G, Wang Z, Han X, Gu J, Mu X, Qiu J and Qian Y. AFB1 triggers lipid metabolism disorders through the PI3K/Akt pathway and mediates apoptosis leading to hepatotoxicity. Foods 2024; 13: 163.
[18] Moloi TP, Ziqubu K, Mazibuko-Mbeje SE, Mabaso NH and Ndlovu Z. Aflatoxin B(1)-induced hepatotoxicity through mitochondrial dysfunction, oxidative stress, and inflammation as central pathological mechanisms: a review of experimental evidence. Toxicology 2024; 509: 153983.
[19] Sakellakis M and Flores LJ. Is the glucocorticoid receptor a key player in prostate cancer?: A literature review. Medicine (Baltimore) 2022; 101: e29716.
[20] De Falco M and Laforgia V. Combined effects of different endocrine-disrupting chemicals (EDCs) on prostate gland. Int J Environ Res Public Health 2021; 18: 9772.
[21] Trasande L and Sargis RM. Endocrine-disrupting chemicals: mainstream recognition of health effects and implications for the practicing internist. J Intern Med 2024; 295: 259-274.
[22] Bleak TC and Calaf GM. Breast and prostate glands affected by environmental substances (Review). Oncol Rep 2021; 45: 20.
[23] Her L and Zhu HJ. Carboxylesterase 1 and precision pharmacotherapy: pharmacogenetics and nongenetic regulators. Drug Metab Dispos 2020; 48: 230-244.
[24] Greener JG, Kandathil SM, Moffat L and Jones DT. A guide to machine learning for biologists. Nat Rev Mol Cell Biol 2022; 23: 40-55.
[25] Jiang T, Gradus JL and Rosellini AJ. Supervised machine learning: a brief primer. Behav Ther 2020; 51: 675-687.
[26] Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C and Wen H. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J Environ Manage 2023; 332: 117357.
[27] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L, Wolf DM, Shriver CD and Thorsson V; Cancer Genome Atlas Research Network, Hu H. An integrated TCGA Pan-Cancer clinical data resource to drive high-quality survival outcome analytics. Cell 2018; 173: 400-416, e411.

[28] Chaoying L, Chao M, Xiangrui Y, Yingjian H, Gang Z, Yunhan R and Yu G. Risk factors of bone metastasis in patients with newly diagnosed prostate cancer. Eur Rev Med Pharmacol Sci 2022; 26: 391-398.
[29] Enwald M, Lehtimäki T, Mishra PP, Mononen N, Murtola TJ and Raitoharju E. Human prostate tissue micrornas and their predicted target pathways linked to prostate cancer risk factors. Cancers (Basel) 2021; 13: 3537.
[30] Rivera-Núñez Z, Ashrap P, Barrett ES, Llanos AAM, Watkins DJ, Cathey AL, Vélez-Vega CM, Rosario Z, Cordero JF, Alshawabkeh A and Meeker JD. Personal care products: demographic characteristics and maternal hormones in pregnant women from Puerto Rico. Environ Res 2022; 206: 112376.