Discover Oncology

ANALYSIS Open Access

Deciphering the potential pathogenic mechanisms of 3-BHA in ovarian cancer through integrated bioinformatics and machine learning strategies

Yifei Shi[1], Dong Niu[1] and Chunhui Jin[2*]

*Correspondence:
Chunhui Jin
wxzy013@njucm.edu.cn
[1]Graduate School of Nanjing, University of Traditional Chinese Medicine, Nanjing 210023, Jiangsu, China
[2]Wuxi Hospital, Nanjing University of Chinese Medicine, Wuxi 214071, Jiangsu, China

Abstract

Ovarian cancer (OC) remains a malignancy characterized by obscure risk factors and unfavorable prognosis. While 3-tert-butyl-4-hydroxyanisole (3-BHA) is suspected of exerting toxic effects on ovarian health, the precise molecular mechanisms underlying its impact remain elucidated. This study aims to systematically investigate the potential pathogenic mechanisms of 3-BHA in the progression of OC.Integrated transcriptomic data from the GEO database (GSE18520 and GSE40595) were analyzed. A synergistic computational framework was employed, incorporating Differentially Expressed Genes (DEGs) identification, Weighted Gene Co-expression Network Analysis (WGCNA), multiple machine learning algorithms, and SHapley Additive exPlanations (SHAP) analysis to achieve high-interpretability feature selection.Five hub genes—CXCR4, CCL7, CXCL8, CXCR2, and CX3CL1—were identified, all demonstrating robust diagnostic efficacy with AUC values of 0.911, 0.882, 0.823, 0.772, and 0.837, respectively. Prognostic profiling via GEPIA3 highlighted CXCR2 overexpression as a potential critical biomarker driving poor clinical outcomes in OC. Furthermore, molecular docking validated the strong binding affinity of 3-BHA with CX3CL1 and CXCR2. Subsequent 100 ns molecular dynamics simulations and thermodynamic stability assessments confirmed the structural stability of the 3-BHA-CXCR2 complex. By integrating bioinformatics and computational toxicology, this study deciphers the potential mechanistic landscape through which 3-BHA influences OC. These findings not only refine the toxicological understanding of 3-BHA but also provide novel candidates for early diagnosis and prognostic risk stratification in OC.

Keywords 3-tert-Butyl-4-hydroxyanisole(3-BHA), Network toxicology, Machine learning, Molecular docking, Ovarian cancer

1 Introduction

Ovarian cancer (OC) remains one of the most lethal gynecological malignancies worldwide. Due to the insidious nature of its early-onset clinical symptoms, the high propensity for chemoresistance, and the ongoing deficiency in effective screening modalities,

OC poses a severe threat to women's health and survival [1].According to projections by the American Cancer Society, approximately 20,890 new cases of OC are anticipated in the United States in 2025 alone, accompanied by an estimated 12,730 related deaths. While the average five-year relative survival rate for all cancers has improved significantly—from 49% in the mid-1970s to 69% during 2014–2020—the corresponding rate for OC remained stagnant at only 51% for the same period. This discrepancy underscores the protracted progress in enhancing OC prognosis and highlights the formidable challenges persisting in clinical management [2]. The precise etiological landscape of OC has yet to be fully elucidated. Conventional paradigms suggest that its onset is multifactorial, involving age, genetic predisposition, and socioeconomic factors such as employment status. Furthermore, individual physiological factors, including endometriosis and hormone replacement therapy, have been validated as contributors to increased risk [3]. However, emerging evidence in recent years reveals that the pathogenesis and progression of OC involve far more intricate biological processes, including aerobic glycolysis (the Warburg effect) [4], mitochondrial metabolic reprogramming [5], and dysregulation of the cellular microenvironment [6].

Synthetic Phenolic Antioxidants (SPAs) represent a critical class of industrial chemical additives primarily utilized to impede oxidative degradation. These compounds are ubiquitously incorporated into food products—particularly to extend the shelf life of edible oils [7]—as well as cosmetics and polymers. The pervasive application of SPAs in consumer goods has catalyzed significant academic scrutiny regarding their potential biosafety [8]. In light of their prevalence in the food supply chain, international regulatory bodies, such as the European Union, have established rigorous benchmarks, mandating a maximum permissible limit of 0.01 g/kg for individual SPAs and a cumulative threshold of 0.2 g/kg [9].

This study focuses on 3-tert-Butyl-4-hydroxyanisole (3-BHA), a prominent representative of the SPA family. Steroidogenesis assays have demonstrated that 3-BHA can elicit potential estrogenic effects, thereby disrupting steroid hormone homeostasis both in vivo and in vitro [10]. Furthermore, an array of toxicological investigations has corroborated the cytotoxicity and genotoxicity of 3-BHA [11], suggesting detrimental impacts on organismal development and reproductive functions. Prior research has indicated that 3-BHA suppresses hepatocyte differentiation and exhibits latent toxicity during embryonic development and organogenesis [12]. Given the bioaccumulative nature of 3-BHA, its established role as an endocrine disruptor, and the inherent correlation between OC pathogenesis and hormonal imbalances, exploring the environmental toxicological impact and molecular underpinnings of 3-BHA in OC is of profound scientific and practical significance. Recent studies have successfully integrated machine learning with multi-omics to explore programmed cell death in renal carcinoma [13] and prognostic markers in bladder cancer [14], setting a precedent for our systematic investigation into 3-BHA-induced OC progression.

Consequently, this study leverages advanced bioinformatic methodologies to elucidate the potential mechanisms by which 3-BHA influences OC. Utilizing OC transcriptomic datasets from the GEO database (GSE18520 and GSE40595) alongside 3-BHA-related gene sets, we executed a systematic analysis involving Differentially Expressed Genes (DEGs) and Weighted Gene Co-expression Network Analysis (WGCNA) to identify critical modules and interaction networks. Functional enrichment analyses (GO and

KEGG) were subsequently performed to unveil core biological processes. Furthermore, a predictive model was constructed using ensemble machine learning algorithms, with gene contributions interpreted via SHapley Additive exPlanations (SHAP) to isolate definitive hub genes. The clinical relevance of these genes was further validated through prognostic assessment on the GEPIA3 platform. Finally, molecular docking and 100 ns molecular dynamics (MD) simulations, complemented by thermodynamic stability analysis, were conducted to verify the binding affinity and structural stability of 3-BHA with core targets, specifically CX3CL1 and CXCR2. These findings collectively deepen the toxicological understanding of 3-BHA and offer novel perspectives for the early diagnosis and prognostic stratification of OC.

2 Materials and methods

2.1 Preprocessing of OC datasets

The GSE18520 and GSE40595 datasets were retrieved from the NCBI GEO database (Platform: GPL570). GSE18520 comprises 10 normal ovarian surface epithelium (OSE) brushings and 53 OC tissue samples; GSE40595 includes 8 normal ovarian stroma samples and 31 OC stroma samples. All utilized data were microarray-based tissue profiles. To mitigate batch effects across datasets, the "ComBat" function from the R package sva was employed. The efficacy of batch correction was validated via Principal Component Analysis (PCA) to ensure data consistency, and the integrated dataset served as the training set for subsequent machine learning models. Differentially Expressed Genes (DEGs) were identified using the limma package with the significance criteria set at an adjusted $P < 0.05$ and $|\log FC| > 0.585$. Subsequently, Weighted Gene Co-expression Network Analysis (WGCNA) was performed. A scale-free topology was constructed by selecting a minimum soft-threshold power (β) corresponding to a scale-free fit index (R^2) approaching 0.8. A Topological Overlap Matrix (TOM) was then calculated to define co-expression modules through dynamic tree cutting. Key modules were identified by calculating the correlation between module eigengenes and OC clinical traits for downstream analysis.

2.2 Acquisition of 3-BHA target gene sets

The SMILES structure of 3-BHA was retrieved from PubChem ([15] https://pubchem.ncbi.nlm.nih.gov/). Toxicity profiles were predicted using ADMETlab 3.0 [16] (https://admetmesh.scbdd.com/service/evaluation/cal) and ProTox 3.0 (https://tox.charite.de/protox3/). Target genes associated with 3-BHA were aggregated from ChEMBL (https://www.ebi.ac.uk/chembl/), SEA (https://sea.bkslab.org/), STITCH (http://stitch.embl.de/), and SwissTargetPrediction (http://www.swisstargetprediction.ch/) databases, filtered for Homo sapiens, and standardized using the UniProt database [17] (https://www.uniprot.org/). A non-redundant 3-BHA-related gene set was established by merging the results from these platforms.

2.3 Screening of key genes and functional enrichment

Candidate genes were identified through Venn analysis of the OC-related (DEGs/WGCNA) and 3-BHA-related gene sets using the ggvenn package. A Protein-Protein Interaction (PPI) network was constructed via the STRING database [18] (https://cn.string-db.org/) with a minimum confidence threshold of 0.4. Within Cytoscape [19],

node importance was ranked using four algorithms via the CytoHubba and MCODE plugins: Edge Percolated Component (EPC), Maximal Clique Centrality (MCC), Molecular Complex Detection (MCODE), and Maximum Neighborhood Component (MNC). Specifically, the top 10 nodes from EPC, MCC, and MNC, along with the top 6 nodes from MCODE, were extracted. The intersection of these four algorithms defined the Key Genes.Functional characterization was performed using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses via R packages including clusterProfiler and org.Hs.eg.db. Significance was defined by an adjusted $P < 0.05$ and a count ≥ 1.

2.4 Machine learning model selection and interpretation

Three external datasets (GSE52037, GSE54388, and GSE105437) from the GPL570 platform were utilized. GSE52037 comprises 10 normal ovarian samples (Normal) and 10 OC tissue samples; GSE54388 includes 6 human ovarian surface epithelium (HOSE) samples (Normal) and 16 OC tissue samples; GSE105437 comprises 5 normal ovarian samples (Normal) and 10 OC tissue samples as the test set (Test). We systematically compared the performance of multiple machine learning algorithms, including Gradient Boosting Machine (GBM), Lasso Regression, Random Forest (RF), and Support Vector Machine (SVM). Data were normalized (mean = 0, SD = 1), and Gaussian noise (SD = 0.01) was introduced to improve robustness. Hyperparameters were optimized via 10-fold cross-validation to prevent overfitting. The model with the highest Area Under the Curve (AUC) was selected as the optimal predictor.

SHAP (SHapley Additive exPlanations) analysis was subsequently applied to quantify the contribution of each feature to OC, identifying the top five contributors as Hub Genes. The expression levels of these hub genes were validated between Normal and OC groups, and their diagnostic performance was evaluated using Receiver Operating Characteristic (ROC) curves.

2.5 GEPIA3 prognostic analysis

The correlation between hub gene expression and OC prognosis was evaluated using GEPIA3(https://gepia3.bioinfoliu.com/). Survival analysis was performed on a cohort from TCGA and GTEx (212 OC and 212 normal samples). The "Survival" module was utilized with a median cut-off (50% high vs. 50% low) to assess the impact of gene expression on Overall Survival (OS).

2.6 Molecular docking

PDB files for hub proteins (receptors) were obtained from RCSB PDB [20] (https://www.rcsb.org/), and the 3D structure of 3-BHA (ligand) was sourced from PubChem. Blind molecular docking was performed in triplicate on the CB-Dock2 platform [21](https://cadd.labshare.cn/cb-dock2/index.php) to ensure reproducibility. The platform automat ically preprocessed structures by removing water molecules and heteroatoms. Binding affinity was quantified using the Vina score, where lower values indicate higher structural stability. The optimal binding pose was selected based on the lowest Vina score and visualized in 3D.

2.7 Molecular dynamics (MD) simulation

Gromacs 2025 was employed for a 100 ns MD simulation to evaluate the dynamic stability of the 3-BHA-target complexes under physiological conditions [22]. The protein and ligand were parameterized using the AMBER14SB and GAFF force fields, respectively. The system was solvated in a TIP3P explicit water box and neutralized with Na^+. Following energy minimization via the steepest descent method, the system underwent 500 ps of NVT and NPT equilibration (coupling constant: 0.1 ps) at 300 K and 1 Bar.The production MD run lasted 100 ns (50,000,000 steps, 2 fs/step). Structural stability, flexibility, and solvent accessibility were assessed using Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (Rg), Solvent Accessible Surface Area (SASA), and hydrogen bonding profiles [23]. Non-covalent interactions were quantified using the Protein-Ligand Interaction Profiler (PLIP) [24]. Finally, the thermodynamic landscape was evaluated via Gibbs Free Energy Landscape (FEL) analysis.

3 Results

3.1 Integration of OC datasets and identification of DEGs

To expand the study cohort, we integrated samples from the GSE18520 and GSE40595 datasets. To eliminate cross-platform batch effects (Fig. 1A), data correction was performed. Principal Component Analysis (PCA) demonstrated a distinct shift from clustered batch-specific distributions to a cohesive unified distribution, validating the effectiveness of the normalization (Fig. 1B). Subsequent DEG analysis between the Normal and OC groups yielded 2,865 significantly altered genes, comprising 1,340 upregulated and 1,525 downregulated transcripts (Fig. 1C).

Weighted Gene Co-expression Network Analysis (WGCNA) was executed on the integrated matrix. By evaluating the scale-free topology fit index (R^2), a soft-threshold power (β) of 8 was selected, achieving an R^2 approach to 0.8 (indicated by the red line in Fig. 2A). This ensured the biological relevance of the scale-free network. Through dynamic tree cutting (Fig. 2C) and module merging (Fig. 2D), 12 co-expression modules were identified. Sample clustering confirmed the absence of significant outliers (Fig. 2B). Correlation analysis between module eigengenes (MEs) and clinical phenotypes revealed that the MEred module exhibited the strongest positive correlation with OC (Cor = 0.58, P = 1e-10). Although the grey module (MEgrey) showed high correlation, it was excluded as it represents unassigned genes (Fig. 2E). The MEred module, containing 256 genes with high gene significance (Fig. 2F), was prioritized for further investigation.

3.2 Toxicological profiling and target gene set of 3-BHA

Pharmacokinetic and toxicological assessments via ADMETlab 3.0 and ProTox 3.0 flagged 3-BHA for potential carcinogenicity (Table 1), providing a safety-based rationale for this study. By aggregating data from ChEMBL, SEA, STITCH, and SwissTargetPrediction and removing redundant entries, a finalized 3-BHA-related gene set comprising 495 targets was established.

3.3 Candidate gene identification and functional enrichment analysis

Venn analysis intersecting the 3-BHA gene set (Fig. 3A) and the OC-related DEGs/WGCNA modules (Fig. 3B) yielded 71 overlapping candidate genes (Fig. 3C). A PPI network constructed via STRING displayed 1,305 edges across 121 nodes with an average

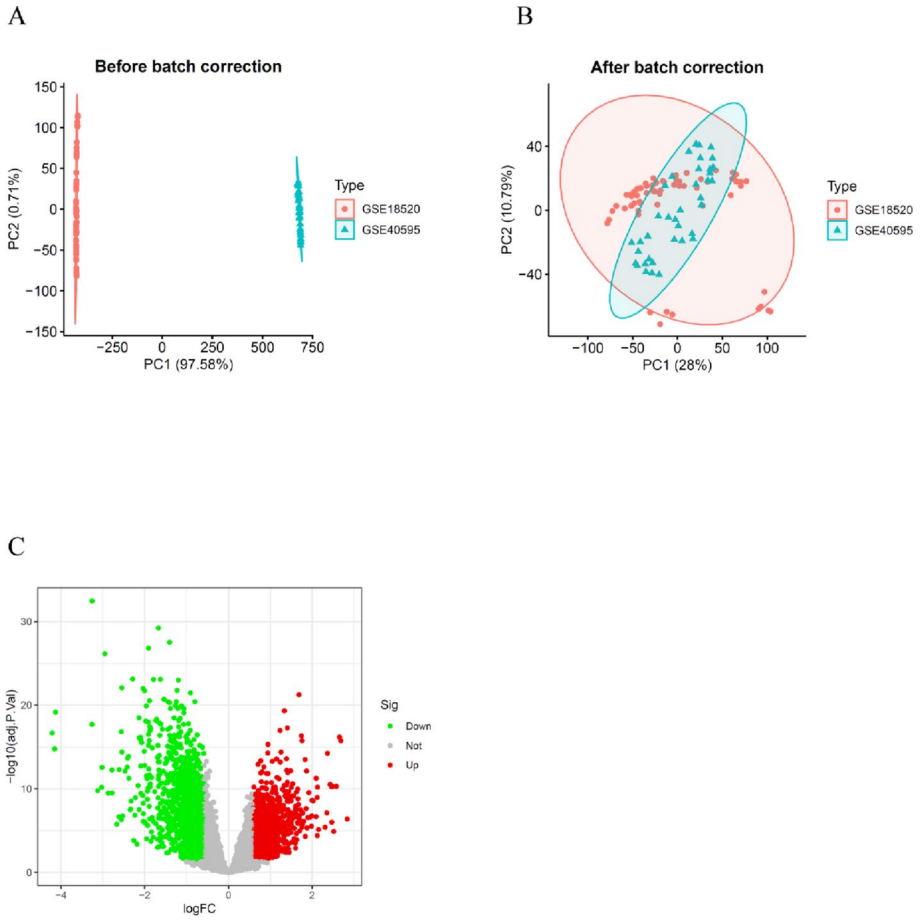

Fig. 1 Identification of DEGs. PCA scatter plot illustrating the presence of inherent batch effects between the GSE18520 and GSE40595 datasets prior to batch correction (A). PCA scatter plot demonstrating the successful mitigation of batch effects following correction, reflecting enhanced data consistency and comparability across integrated datasets (B). Volcano plot highlighting the distribution of Differentially Expressed Genes (DEGs) based on |logFC| and statistical significance. Red dots signify upregulated genes, green dots signify downregulated genes, and grey dots represent non-significant genes (C).

node degree of 21.6, reflecting a highly interconnected interactome. Using the CytoHubba and MCODE algorithms, seven high-ranking genes—CCL7, CCL11, CXCL8, CXCR2, CXCR4, CX3CL1, and CX3CR1—were shortlisted as pivotal nodes (Fig. 3D).

Functional enrichment analysis elucidated the biological impact of these candidates. GO analysis (Fig. 4A) identified 272 Biological Processes (BP), emphasizing cell chemotaxis and myeloid leukocyte migration. Cellular Component (CC) enrichment highlighted transmembrane transporter complexes and mitochondrial respiratory chain components (e.g., NADH dehydrogenase complex), aligning with Molecular Function (MF) results involving oxidoreductase and electron transfer activities.

KEGG pathway analysis (Fig. 4B) mapped the candidates to 45 pathways (adjusted $P < 0.05$), notably Chemical carcinogenesis-reactive oxygen species, MAPK signaling, and various neurodegenerative pathways. A Sankey diagram (Fig. 4C) further visualized the distribution, confirming that five core genes (CXCR4, CCL7, CXCL8, CXCR2, CX3CL1) are integral to the Chemokine signaling pathway and cytokine-receptor interactions.

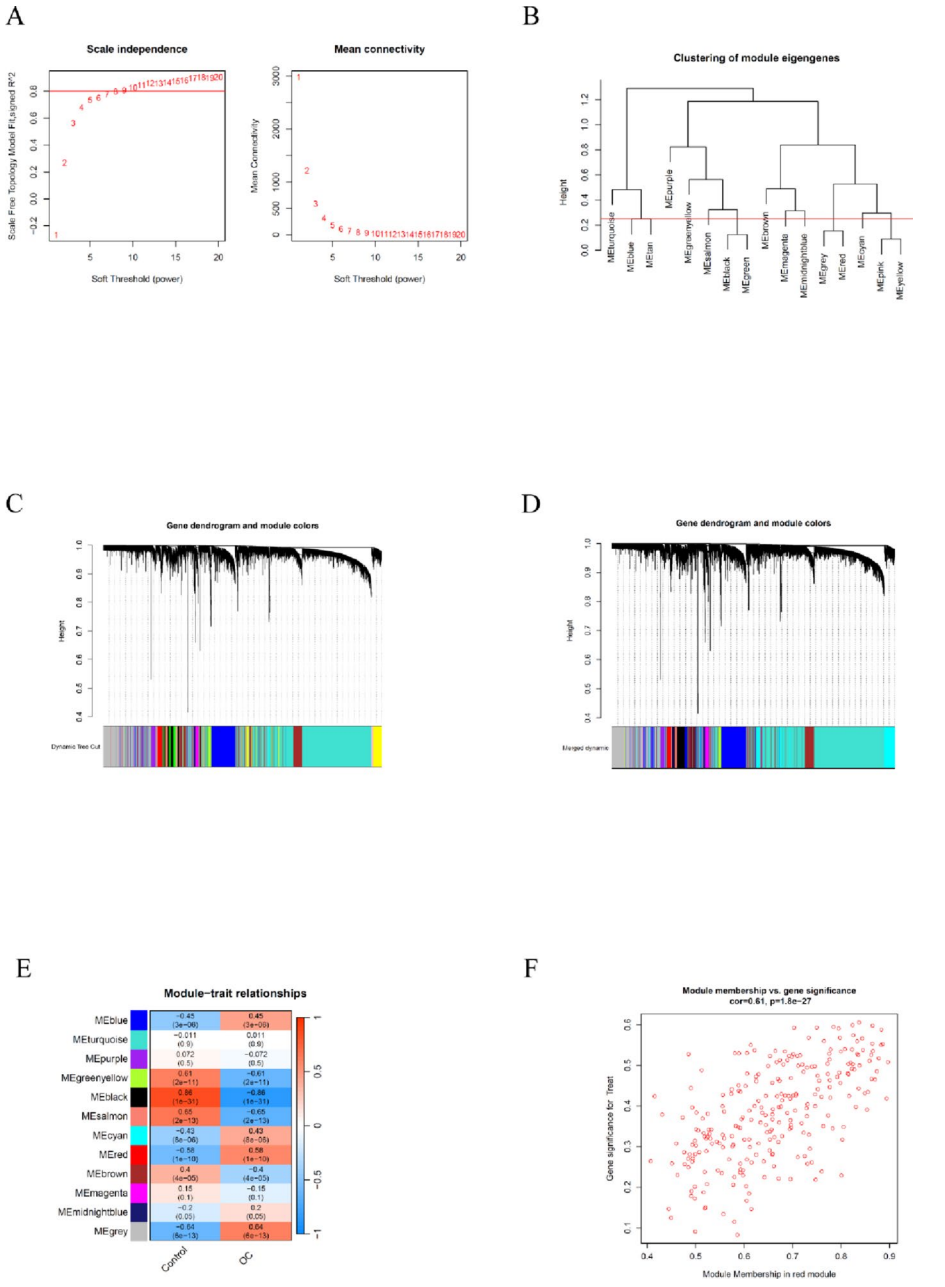

Fig. 2 Identification of WGCNA. Analysis of the scale-free topology fit index to determine the optimal soft-threshold power(β). At β = 8, the fit index reached approximately 0.8, ensuring the construction of a biologically relevant scale-free network (A). Module eigengene clustering dendrogram illustrating the partitioning of co-expression modules and the subsequent merging of closely related clusters (B). Hierarchical clustering dendrogram of genes based on the dynamic tree-cutting method; the colored bars in the lower panel represent distinct co-expression modules assigned to gene clusters (C). Hierarchical clustering dendrogram showing gene organization based on co-expression patterns, with color-coded assignments indicating the finalized module structure (D). Module-trait relationship heatmap displaying the correlation between WGCNA-identified modules and clinical status (Normal vs. Tumor). Each cell contains the correlation coefficient and the corresponding *P*-value (E). Scatter plot highlighting the Gene Significance (GS) vs. Module Membership (MM) for the MEred module, showing a robust linear correlation (Cor = 0.61, *P* = 1.8e-27) (F).

Table 1 The toxicology prediction results of 3-BHA

Abbreviation	Name	Molecular formula	SMILES	ADMET lab3.0 (Carcinogeni City)	ProTox 3.0 (Carcinogeni City)
3-BHA	3-tert-butyl-4-hydroxyanisole	$C_{11}H_{16}O_2$	CC(C)(C)C1 = C(C = CC(= C1)OC)O	0.452(The output value is the probability of being toxic.)	Active

3.4 Machine learning optimization and model interpretation

Before modeling, all features were standardized to enhance robustness. We systematically evaluated GBM, Lasso, RF, and SVM architectures. The ensemble "Stepglm[both]+Lasso" model emerged as the superior predictor, achieving an AUC of 0.989 in the training set and a mean AUC of 0.921 (Fig. 5A). This optimal model incorporated six genes: CCL7, CXCL8, CXCR2, CXCR4, CX3CL1, and CX3CR1.

Although the sample size in this study was inherently constrained by the availability of public datasets, we mitigated the potential risks of overfitting and ensured the robustness and generalizability of our machine learning models by integrating multiple complementary algorithms and employing SHAP interpretability analysis. To decrypt the model's logic, SHAP analysis was utilized. The SHAP Bee Swarm Plot (Fig. 5B) illustrated that elevated expression of these genes consistently increased the predicted risk of OC. The global importance ranking was: CXCR4 (0.0798) > CCL7 (0.0731) > CXCL8 (0.0412) > CXCR2 (0.0391) > CX3CL1 (0.0375). SHAP dependence plots (Fig. 5C) further elucidated how individual expression levels drive risk contributions, often modulated by CXCR4 interactions. Based on these insights, five Hub Genes (CXCR4, CCL7, CXCL8, CXCR2, CX3CL1) were finalized.

ROC analysis confirmed the diagnostic utility of these five genes, with individual AUC values ranging from 0.772 to 0.911. Moreover, the combined AUC reached 0.984, demonstrating significantly higher diagnostic accuracy than any individual gene (Fig. 5D). Volcano plots (Fig. 5E) verified their expression polarity: CCL7 and CXCR2 were downregulated, while CXCR4, CXCL8, and CX3CL1 were significantly upregulated in OC.

3.5 Prognostic validation via GEPIA3

Clinical relevance was assessed through the GEPIA3 platform. Survival analysis (Fig. 6) revealed that CXCR4 ($P = 0.0112$) and CXCR2 ($P = 0.0332$) were significantly correlated with Overall Survival (OS). CXCR4 high expression acted as a protective factor (HR = 0.73), consistent with its upregulation in our DEG analysis. Conversely, CXCR2 overexpression was identified as a critical risk factor (HR = 1.31), marking it as a key biomarker driving poor clinical outcomes in OC.

3.6 Molecular docking analysis

Docking simulations explored the binding landscape between 3-BHA and the hub targets: CXCR4(PDBID: 6HSR)、CCL7(PDBID: 7SCU)、CXCL8(PDBID: 4XDX)、CXCR2(PDBID: 5TYT)、CX3CL1(PDBID: 4XT1). Binding energies below -5.0 kcal/mol signify robust affinity. Results showed favorable interactions across all targets, with the lowest binding energies for CX3CL1 (-6.0 kcal/mol) and CXCR2 (-5.6 kcal/mol), suggesting their central role in 3-BHA-mediated OC progression. The 3D binding configurations were visualized using CB-Dock2 (Fig. 7).

3.7 Molecular dynamics (MD) simulation and thermodynamic landscape

A 100 ns MD simulation was conducted to evaluate the dynamic stability of the 3-BHA-CXCR2 complex (Fig. 8A). PLIP-based micro-analysis revealed a sophisticated interaction network (Fig. 8B): 3-BHA's hydroxyl oxygen established high-strength hydrogen bonds with HIS-102 (2.29 Å), MET-96 (2.50 Å), and LYS-84 (2.95 Å). This "multi-chain

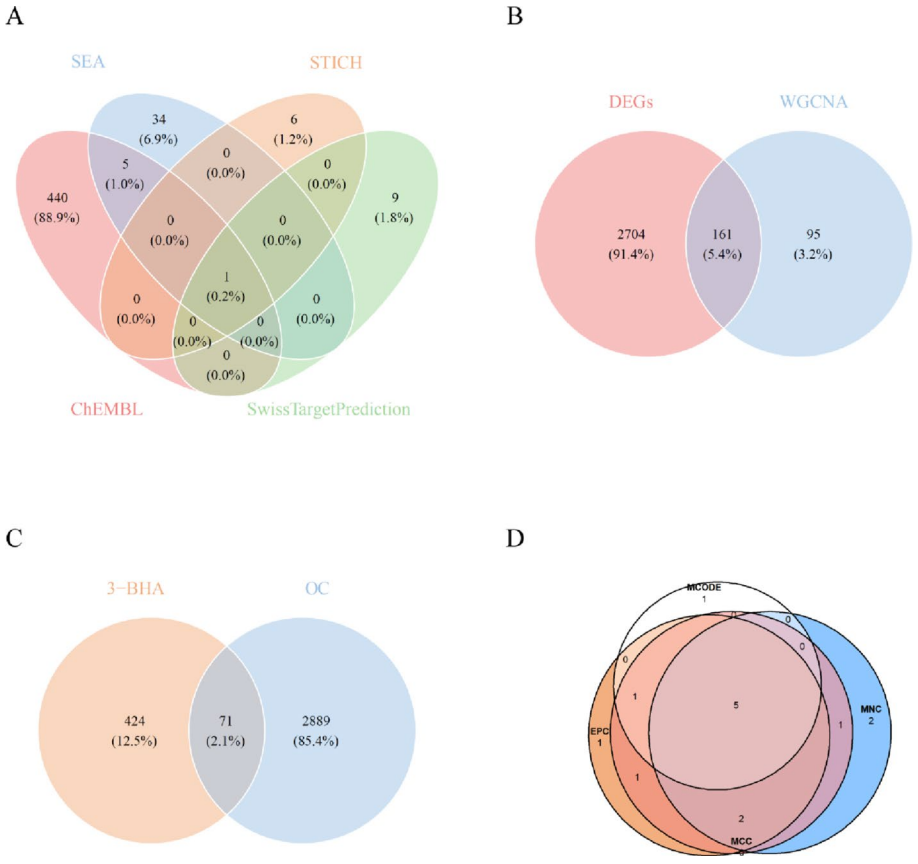

Fig. 3 Identification of important genes. Venn diagram illustrating the aggregation of 3-BHA target genes retrieved from multiple pharmacological platforms (ChEMBL, SEA, STITCH, and SwissTargetPrediction), showing the establishment of a comprehensive non-redundant gene library (A). Venn diagram intersecting DEGs (red) and WGCNA-identified modules (blue) to define robust OC-associated gene clusters (B). Venn diagram identifying candidate genes by intersecting the 3-BHA-related gene set (yellow) with the OC-associated gene set (blue); the overlapping region represents potential targets for 3-BHA-mediated impact on OC(C). Identification of seven pivotal genes based on the consensus of four topological algorithms within the Protein-Protein Interaction (PPI) network. The intersection of Edge Percolated Component (EPC), Maximal Clique Centrality (MCC), Molecular Complex Detection (MCODE), and Maximum Neighborhood Component (MNC) rankings was utilized to ensure the reliability of the hub gene selection (D).

anchoring" mechanism ensures exceptional ligand retention. Additionally, the tert-butyl group maintained hydrophobic contacts with THR-95 and LYS-84.

Trajectory analysis confirmed the complex's integrity. Hydrogen bonds remained stable (1–3 bonds, peaking at 6; Fig. 9A). The RMSD stabilized at ~1.5 nm after a 40 ns induced-fit period, with the ligand's own RMSD fluctuating within a negligible 0.15 nm threshold, precluding dissociation (Fig. 9B). Constant Radius of Gyration (Rg ≈ 2.3 nm) and SASA profiles (Fig. 9D and E), alongside minimal RMSF fluctuations (0.15–0.38 nm) in key residues, validated the formation of a rigid, biologically significant complex.

Finally, Free Energy Landscape (FEL) analysis exhibited a singular, deep "funnel-shaped" energy well. The spontaneous convergence to this thermodynamic minimum (0 kJ/mol vs. >14 kJ/mol for surrounding states) confirms that 3-BHA embedding into the CXCR2 pocket is a self-driven, highly stable energetic process(Fig. 8C-D).

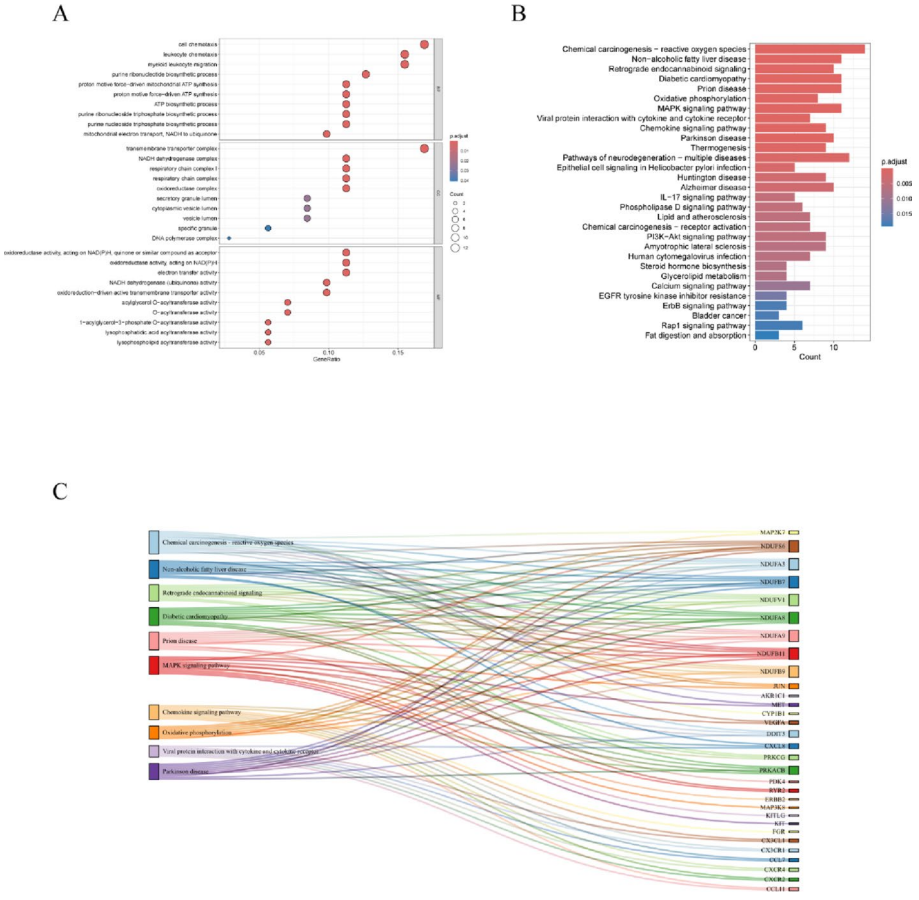

Fig. 4 GO and KEGG enrichment results of the candidate genes. GO functional annotation illustrating the enrichment of candidate genes across three categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF), highlighting the predominant involvement in chemotaxis and mitochondrial metabolic activities (A). KEGG pathway analysis revealing the primary signaling cascades and metabolic pathways associated with the candidate genes, including chemical carcinogenesis and MAPK signaling (B). Sankey diagram visualizing the connectivity and distribution between candidate genes and the top 10 enriched KEGG pathways, illustrating the multi-target regulatory landscape of 3-BHA in OC (C).

4 Discussion

In this study, an integrative framework combining network toxicology, machine learning [25], and 100 ns molecular dynamics (MD) simulations was employed to elucidate the potential pathogenic mechanisms of 3-BHA in ovarian cancer (OC). Consistent with the methodologies used to decipher mitotic catastrophe and ADME gene signatures in bladder cancer [26], our study underscores the power of machine learning in identifying the critical chemokine network driving 3-BHA-induced malignancy. By leveraging GEO transcriptomic data and toxicological databases, we identified a core gene signature comprising CXCR4, CCL7, CXCL8, CXCR2, and CX3CL1, all of which exhibited robust diagnostic performance. Notably, prognostic analysis via GEPIA3 highlighted CXCR2 as a critical biomarker, with its overexpression significantly correlating with diminished overall survival (OS). Molecular docking and MD simulations further confirmed that 3-BHA achieves high-affinity, specific binding within the CXCR2 active pocket through a stable hydrogen-bonding network. These findings not only refine the toxicological profile of 3-BHA but also provide novel evidence for early OC diagnosis and prognostic stratification.

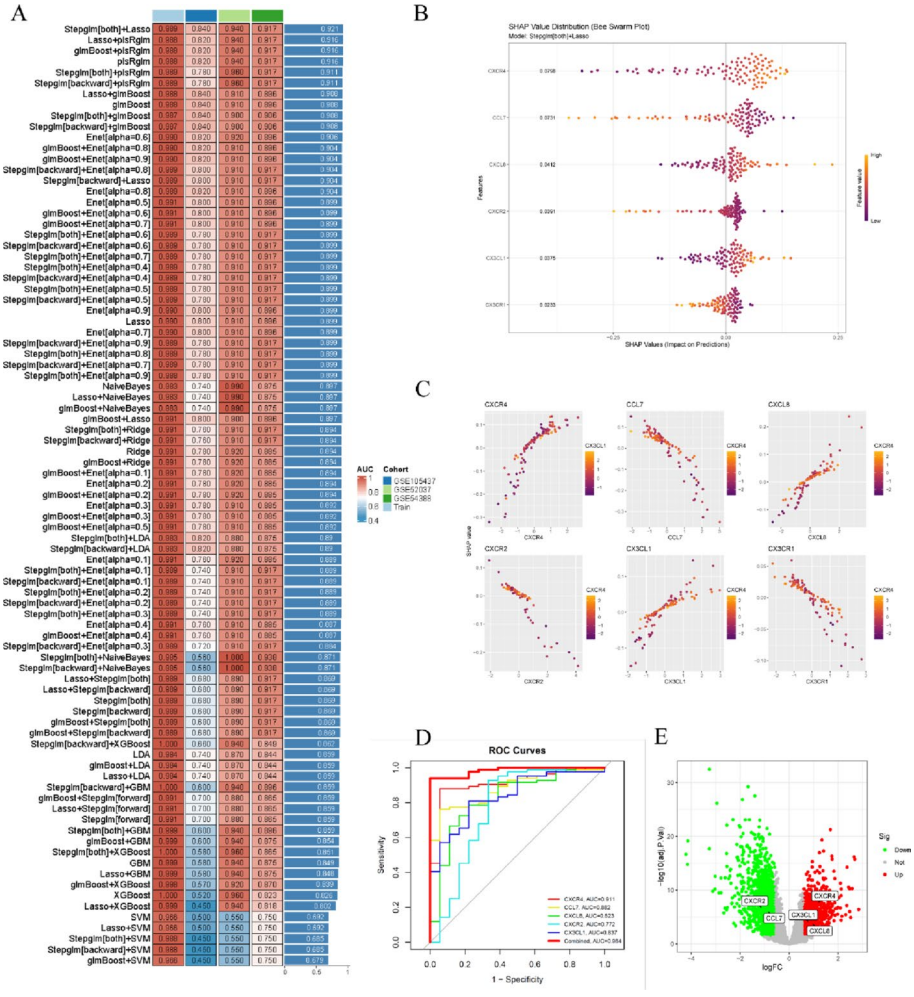

Fig. 5 The progress of machine learning and validation of SHAP. Performance Heatmap illustrating the comparative analysis of various machine learning models based on Area Under the Curve (AUC) values, identifying the optimal predictive architecture for OC (A). SHAP Bee Swarm Plot displaying the distribution of gene contributions within the ensemble "Stepglm[both]+Lasso" model. The horizontal width reflects data density, while the color gradient signifies the expression levels of individual genes (B). SHAP Dependence Plots elucidating how the expression levels of individual genes (X-axis) influence their marginal contribution to the model (Y-axis); the color-coding represents interaction effects with other pivotal features, predominantly CXCR4 (C). Receiver Operating Characteristic (ROC) curves evaluating the diagnostic efficacy of the five identified hub genes (CXCR4, CCL7, CXCL8, CXCR2, and CX3CL1). The X-axis denotes the False Positive Rate, the Y-axis denotes the True Positive Rate, and the AUC values quantify the predictive robustness (D). Volcano plot visualizing the expression polarity of the core genes. Red and green dots indicate significantly upregulated and downregulated genes, respectively, while grey dots represent non-significant features. Critical hub genes are explicitly labeled for clarity (E).

Network toxicology has emerged as a transformative approach that integrates proteomics, genomics, and bioinformatics to map the intricate relationships between environmental compounds, target genes, and diseases [27]. By coupling this with machine learning [28] and interpretability tools like SHAP [29], we achieved a precise identification of risk-associated targets. The subsequent validation through MD simulations [30] transitioned the study from static predictions to a dynamic biochemical understanding, effectively bridging the gap between computational screening and structural biology.

The etiological landscape of OC is complex, and the rising concern over long-term, low-dose exposure to synthetic phenolic antioxidants (SPAs) like 3-BHA is well-founded. 3-BHA is ubiquitous in daily consumer goods, and emerging evidence suggests it may

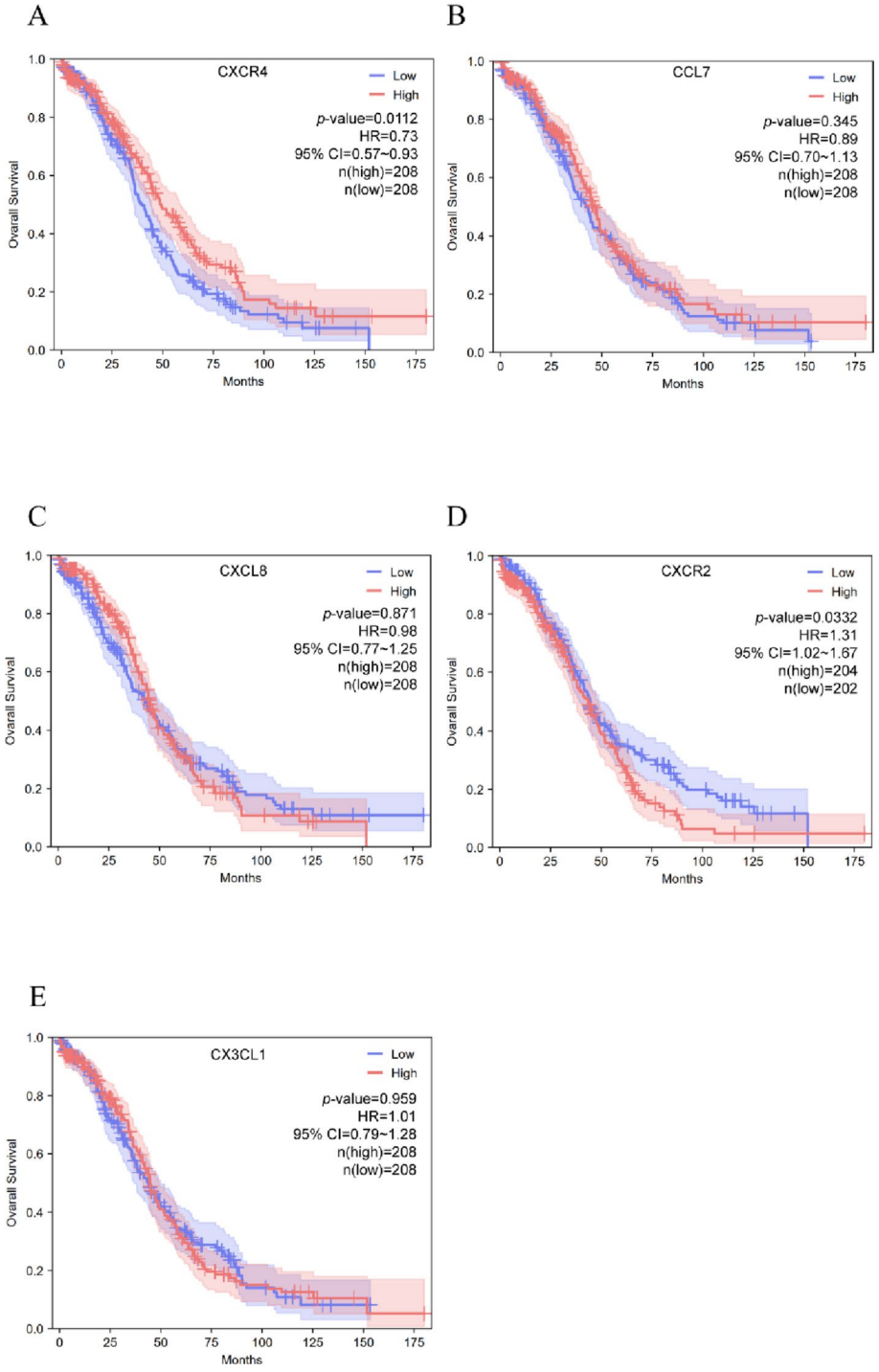

Fig. 6 Results of GEPIA(A-E)

act synergistically with other pollutants. For instance, co-exposure to nanoplastics and 3-BHA has been shown to amplify nephrotoxicity in murine models [31]. Given the unfavorable prognosis of OC and the bioaccumulative nature of SPAs, investigating the impact of 3-BHA on ovarian health is of profound clinical and public health significance.

Our GO enrichment analysis localized the functional impact of 3-BHA-related OC genes to immune cell recruitment, nucleic acid metabolism, and mitochondrial redox

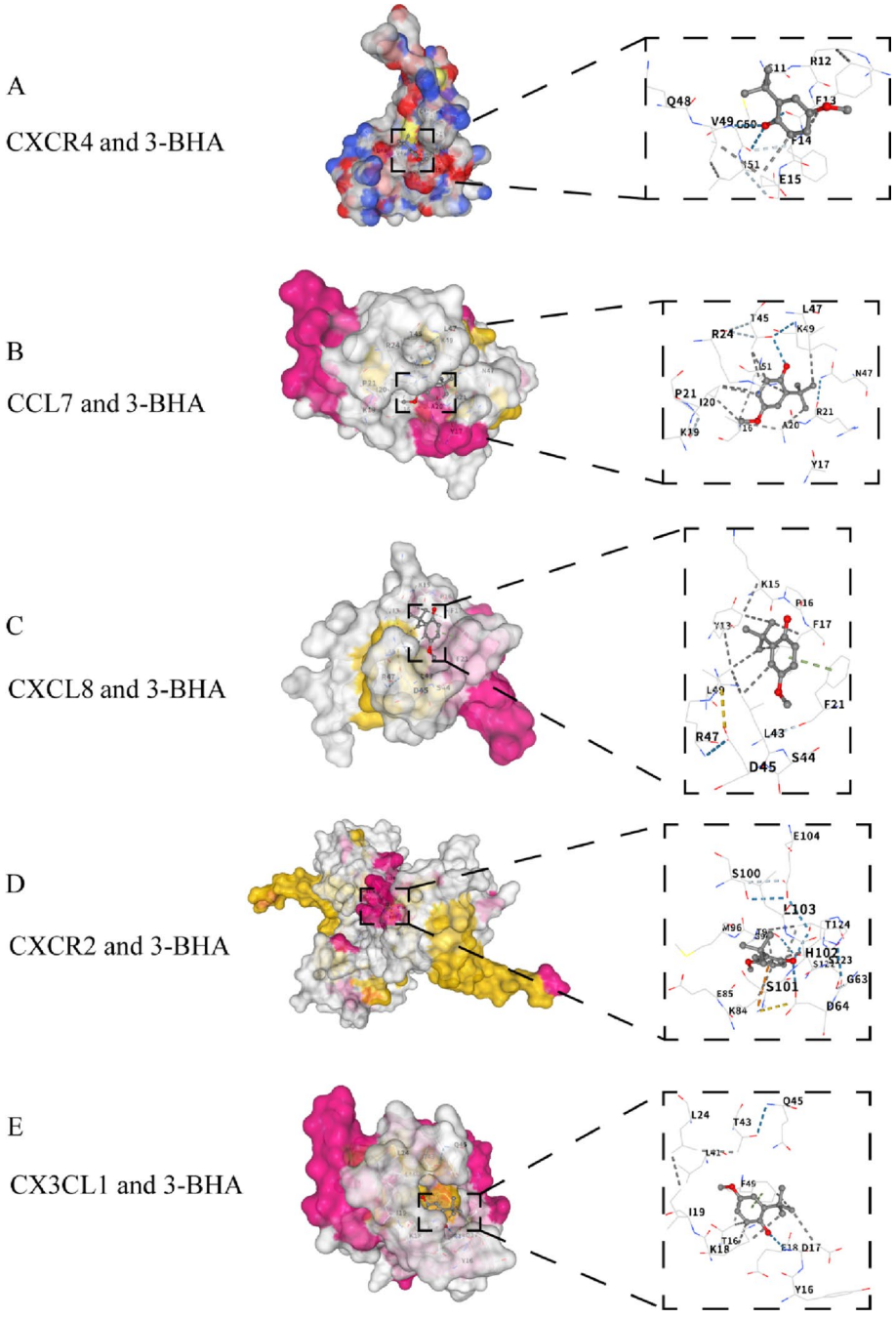

Fig. 7 Molecular docking results of the core genes with 3-BHA(A-E)

homeostasis. The enrichment in cell chemotaxis and myeloid leukocyte migration underscores 3-BHA's potential role in modulating the inflammatory landscape of the tumor microenvironment. Furthermore, the involvement in purine ribonucleotide biosynthesis suggests a mechanism supporting the heightened metabolic demands of proliferating malignant cells. In the cellular component (CC) category, the focus on transmembrane transporters and mitochondrial respiratory chain complexes—coupled with the molecular function (MF) of oxidoreductase activity—indicates that these target

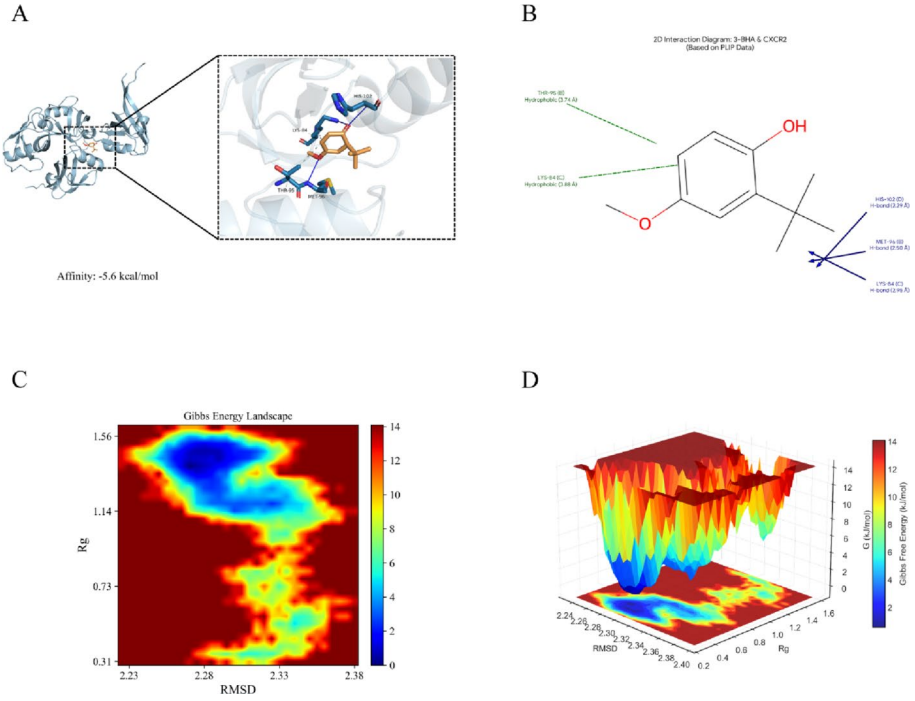

Fig. 8 Binding Mode and Intermolecular Interactions between 3-BHA and CXCR2. Three-dimensional (3D) binding architecture of the 3-BHA-CXCR2 complex at the equilibrium state. This panel illustrates the spatial orientation and conformational fit of 3-BHA within the active pocket, highlighting the relative positioning of critical residues involved in the interaction (A). Two-dimensional (2D) interaction diagram generated via the PLIP (Protein-Ligand Interaction Profiler) algorithm. This visualization details the precise non-covalent bonding network, including specific hydrogen bonds and hydrophobic interactions, between 3-BHA and key amino acid residues at the atomic level (B). Gibbs Free Energy Landscape (FEL) illustrating the thermodynamic stability of the complex from a multi-dimensional perspective. Using RMSD and Radius of Gyration (Rg) as reaction coordinates, the energy landscape exhibits a distinct, singular deep energy well, indicating a highly favorable and stable binding state (C). Close-up characterization of the energy landscape, where the dark blue core region reveals an intensely concentrated and deep "funnel-shaped" energy valley. This thermodynamic profile demonstrates that the binding of 3-BHA to CXCR2 not only achieves conformational convergence in spatial orientation but also resides in an energetically optimized state of exceptional stability (D).

genes may drive OC progression by simultaneously enhancing oxidative phosphorylation and facilitating immune cell infiltration.

KEGG pathway analysis revealed that 3-BHA may trigger OC through oxidative stress, genomic damage, and the MAPK signaling pathway [32]. The MAPK cascade, involving subtypes such as ERK, JNK, and p38 [33], is a classic conduit for regulating cell proliferation, differentiation, and apoptosis [34]. Notably, p38 MAPK inhibition has been shown to reverse paclitaxel resistance in OC cells [35], highlighting its clinical importance. The enrichment of the Chemical carcinogenesis-ROS pathway emphasizes the role of reactive oxygen species in driving oncogenic transformation [36]. Additionally, systemic metabolic stressors, such as non-alcoholic fatty liver disease and diabetic cardiomyopathy, may facilitate ovarian tumor development by promoting mitochondrial dysfunction and lipid metabolic reprogramming.

In the present study, the carcinogenic potential of 3-BHA appears to be closely linked to the synergy between 3-BHA-induced oxidative stress (ROS) and chemokine signaling pathways. Our survival analysis revealed an intriguing phenomenon: CXCR4 emerged as a protective factor (HR = 0.73), which seemingly contradicts its conventional role in promoting tumor metastasis. However, considering the complex microenvironment

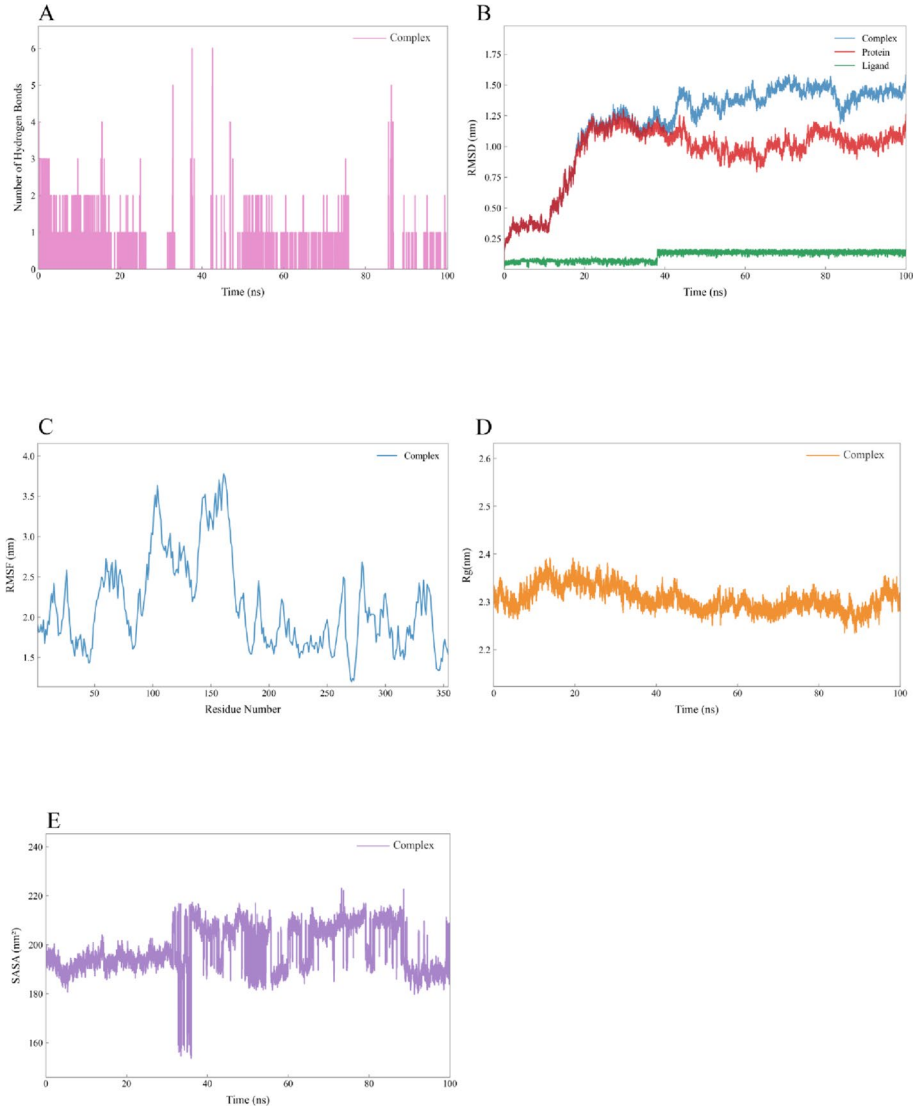

Fig. 9 Dynamic Stability and Structural Evolution during 100 ns MD Simulation between 3-BHA and CXCR2. Temporal evolution of intermolecular hydrogen bonds between 3-BHA and CXCR2, illustrating the sustained stability and frequency of polar interactions throughout the dynamic simulation (A). Root Mean Square Deviation (RMSD) profiles for both the protein backbone and the ligand over the 100 ns trajectory. These curves evaluate the structural convergence of the complex and the precise positional stability of 3-BHA within the binding pocket (B).Root Mean Square Fluctuation (RMSF) profile of CXCR2 residues, highlighting the local flexibility of the binding site and the global protein scaffold; low fluctuation values in the active site residues indicate robust ligand-induced stabilization(C).Radius of Gyration (Rg) and Solvent Accessible Surface Area (SASA) trajectories, utilized to assess the structural compactness and the degree of solvent exposure of the CXCR2 protein upon binding with 3-BHA (D-E).

under 3-BHA exposure, this "protective effect" might reflect a feedback mechanism where CXCR4 recruits specific anti-tumor immune cell populations in response to environmental stress-induced damage. Furthermore, CXCR4 does not function in isolation; rather, it constitutes a dynamic chemokine network alongside CCL7, CXCL8, CXCR2, and CX3CL1. 3-BHA-induced ROS may act as a priming signal to trigger the recruitment of immunosuppressive cells or metabolic reprogramming mediated by this network, collectively shaping a microenvironment conducive to OC progression. Recent evidence emphasizes that tumor microenvironment heterogeneity and immune cell

infiltration are critical in regulating OC progression [37–39]. Our findings align with these studies, suggesting that the 3-BHA-triggered chemokine axis promotes malignancy by shaping an immunosuppressive microenvironment through specific immune cell recruitment.

Molecular docking revealed a binding affinity of −5.6 kcal/mol between 3-BHA and CXCR2. While representing moderate affinity in computational toxicology, the specific interaction pattern with residues S101, H102, and L103 supports stable site occupancy. The subsequent 100 ns MD simulation provided high-fidelity evidence of complex stability in a dynamic aqueous environment. The convergence of the RMSD curve and the minimal ligand displacement (< 0.15 nm) serve as a "gold standard" for identifying stable binding. This was further corroborated by Rg and SASA data, indicating that CXCR2 maintains a compact, stable fold upon 3-BHA binding.

Thermodynamically, the Free Energy Landscape (FEL) demonstrated that the 3-BHA-CXCR2 complex spontaneously converges to a global minimum energy state, effectively precluding non-specific binding or random dissociation. PLIP analysis identified a robust "anchoring network" involving hydrogen bonds with HIS-102, MET-96, and LYS-84. We hypothesize that this binding may induce allosteric effects in the CXCR2 transmembrane domain, triggering downstream cascades such as the PI3K or MAPK pathways [40].

Specifically, 3-BHA may activate the MAPK cascade via CXCR2, leading to the inhibition of apoptosis and the promotion of proliferative escape. Since p38 MAPK is highly sensitive to ROS, the phenolic structure of 3-BHA might act synergistically to create a pro-carcinogenic microenvironment. The low RMSF values in the binding domain, enhanced by the hydrophobic shielding of the tert-butyl group, provide a self-consistent chain of evidence confirming that 3-BHA is a highly reliable binding partner for CXCR2.

5 Limitations

While this study provides a high-efficiency framework for screening environmental toxins, it is not without limitations. Our conclusions are primarily derived from computational analyses of publicly available gene expression datasets from the GEO database, specifically utilizing five datasets (GSE18520, GSE40595, GSE52037, GSE54388, GSE105437). Following data preprocessing and partitioning [41], the machine learning model was trained on a cohort comprising 18 normal samples and 84 tumor samples, and evaluated on an independent test set of 21 normal samples and 36 tumor samples. Our conclusions are primarily based on these public databases and computational simulations, and have not been experimentally validated. Future research must incorporate in vitro and in vivo experiments to validate the specific activation of the MAPK pathway by 3-BHA. Furthermore, while we have identified CXCR2 as a key biomarker, multi-exposure models are needed to assess the cumulative effects of mixed environmental pollutants. Drawing inspiration from concepts like the Glutamine Metabolism Prognostic Index (GMPI) [42], future studies could refine these chemokine genes as definitive features for predicting patient outcomes and optimizing therapeutic strategies in OC.

6 Conclusions

This study systematically evaluated the potential toxicological implications of 3-BHA on ovarian health through an integrated bioinformatics and computational framework. Our findings identified a signature of chemokine-related genes, with CXCR2 emerging as a central hub target linking 3-BHA exposure to ovarian cancer (OC) progression and unfavorable prognosis. These results provide a robust scientific foundation for the safety assessment of 3-BHA and offer novel molecular leads for the early diagnosis and prognostic risk stratification of OC.

While this research preliminarily elucidates the mechanistic underpinnings by which 3-BHA influences OC pathogenesis, these computational insights warrant further validation through in vivo and in vitro experimental models. Overall, our work underscores the necessity of monitoring environmental antioxidant exposure and paves the way for integrating multi-omics strategies in gynecological oncology.

Acknowledgements
We would like to express our sincere gratitude to the data provided by GEO, GEPIA3.0 and other databases, which allowed us to complete this research. At the same time, we are very grateful to the various online tools for bioinformatics analysis developed by scientific researchers, which are convenient for us to learn and use.

Author contributions
Chunhui Jin (corresponding author) designed the project, Yifei Shi analyzed data and wrote the manuscript, Dong Niu revised the manuscript.

Funding
This study is supported by the National Natural Science Foundation of China(No. 82274269) and Medical Key Strategic Project of Wuxi Municipal Health Commission.

Data availability
The RNA-seq datasets analyzed in this study were obtained from the NCBI Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE18520, GSE40595, GSE52037, GSE54388, and GSE105437.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 31 January 2026 / Accepted: 16 April 2026
Published online: 03 May 2026

References

1. Ma X, Qu L, Wu Q, et al. Advances in ovarian cancer: biological insights, therapeutic innovations, and future perspectives. J Ovar Res. 2025. https://doi.org/10.1186/s13048-025-01921-x.
2. Siegel RL, Tyler B, Kratzer AN, Sung H, Jemal A. Cancer statistics, 2025. Cancer J Clin. 2025;75(1):10–45. https://doi.org/10.3322/caac.21871.
3. Zheng G, Baandrup L, Wang J, et al. Ovarian cancer risk factors in relation to family history. JNCI J Natl Cancer Inst. 2024;116(11):1767–74. https://doi.org/10.1093/jnci/djae164.
4. Xu Y, Zhao S, Shen Y, et al. MARCH5 promotes aerobic glycolysis to facilitate ovarian cancer progression via ubiquitinating MPC1. Apoptosis. 2024;29(7–8):1232–45. https://doi.org/10.1007/s10495-024-01962-5.
5. Nan Y, Wu X, Luo Q, et al. OTUB2 silencing promotes ovarian cancer via mitochondrial metabolic reprogramming and can be synthetically targeted by CA9 inhibition. Proc Natl Acad Sci. 2024;121(19): e2315348121. https://doi.org/10.1073/pnas.2315348121.
6. Qi R, Yang J, Shen S, Yue Yu, Qing Y. Role of the tumor microenvironment in chemotherapy resistance in ovarian cancer and targeted therapy. J Ovar Res. 2025. https://doi.org/10.1186/s13048-025-01927-5.
7. Wu Z, Shi X, Wang X, Luo Y, Wang X. Modeling and analysis of synergistic phenolic antioxidants in edible oil: integrating molecular simulation and artificial neural network approach. J Mol Struct. 2024;1315:138959. https://doi.org/10.1016/j.molstruc.2024.138959.

8. Xu X, et al. Synthetic phenolic antioxidants: metabolism, hazards and mechanism of action. Food Chem. 2021;353:129488. https://doi.org/10.1016/j.foodchem.2021.129488.
9. Carocho Márcio, Morales P, Isabel CFR, et al. Antioxidants: reviewing the chemistry, food applications, legislation and role as preservatives. Trends Food Sci Technol. 2018;71(January):107–20. https://doi.org/10.1016/j.tifs.2017.11.008.
10. Yang X, Song W, Liu N, et al. Synthetic phenolic antioxidants cause perturbation in steroidogenesis in vitro and in vivo. Environ Sci Technol. 2018;52(2):850–8. https://doi.org/10.1021/acs.est.7b05057.
11. Vandghanooni S, Forouharmehr A, Eskandani M, et al. Cytotoxicity and DNA fragmentation properties of butylated hydroxyanisole. DNA Cell Biol. 2013;32(3):98–103. https://doi.org/10.1089/dna.2012.1946.
12. Nijiati N, Wang M, Julaiti M, Wubuli D, Huang P, Bowen H. Investigating the hepatic developmental toxicity of synthetic phenolic antioxidant 3-BHA: a comprehensive network-based approach combined with human embryonic stem cell-derived hepatocyte models. Ecotoxicol Environ Saf. 2025;305:119235. https://doi.org/10.1016/j.ecoenv.2025.119235.
13. Dai H, Zhang X, Yin L, et al. Integrating machine learning and multi-omics analysis to explore treg-associated programmed cell death features in clear cell renal cell carcinoma. Cancer Cell Int. 2026;26(1):15. https://doi.org/10.1186/s12935-025-04133-x.
14. Dai H, Yu Z, Zhao Y, et al. Integrating machine learning models with multi-omics analysis to decipher the prognostic significance of mitotic catastrophe heterogeneity in bladder cancer. Biol Direct. 2025;20(1):56. https://doi.org/10.1186/s13062-025-00650-x.
15. Kim S, Chen J, Cheng T, et al. PubChem 2025 update. Nucleic Acids Res. 2025;53(D1):D1516-25. https://doi.org/10.1093/nar/gkae1059.
16. Fu, Li S, Shi J, et al. ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. Nucleic Acids Res. 2024;52(W1):W422-31. https://doi.org/10.1093/nar/gkae236.
17. The UniProt Consortium, Bateman A, Martin M-J, et al. UniProt: The Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523–31. https://doi.org/10.1093/nar/gkac1052.
18. Szklarczyk D, Nastou K, Koutrouli M, et al. The STRING database in 2025: protein networks with directionality of regulation. Nucleic Acids Res. 2025;53(D1):D730-37. https://doi.org/10.1093/nar/gkae1113.
19. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. https://doi.org/10.1101/gr.1239303.
20. Bittrich S, Segura J, Duarte JM, Burley SK, Rose Y. Exploring protein 3D similarities via comprehensive structural alignments. Bioinformatics. 2024;40(6):btae370. https://doi.org/10.1093/bioinformatics/btae370.
21. Liu Y, Yang X, Gan J, Chen S. CB-Dock2: improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. Nucleic Acids Res. 2022;50(W1):W159-64. https://doi.org/10.1093/nar/gkac394.
22. Wang J, Deng Q, Lu Q. Exploring the toxicological effects of DOTP exposure on periodontitis by combining molecular docking and molecular dynamics simulations. Sci Rep. 2025;15(1):20915. https://doi.org/10.1038/s41598-025-05740-4.
23. Dong Y, Zhu J. Network toxicology reveals glyphosate mechanisms in kidney injury and cancer. Sci Rep. 2025;15(1):31132. https://doi.org/10.1038/s41598-025-17067-1.
24. Adasme MF, Katja L, Linnemann SN, Bolz, et al. PLIP 2021: Expanding the Scope of the Protein–Ligand Interaction Profiler to DNA and RNA. Nucleic Acids Res. 2021;49(W1):W530–34. https://doi.org/10.1093/nar/gkab294.
25. Dai H, Zhao K, Zhao Y, et al. Machine learning model in multi-omics perspective demystifies the prognostic significance of crotonylation heterogeneity in clear cell renal cell carcinoma. BMC Urol. 2025;25(1):229. https://doi.org/10.1186/s12894-025-01914-4.
26. Dai H, Zhang X, Zhao Y, et al. ADME gene-driven prognostic model for bladder cancer: a breakthrough in predicting survival and personalized treatment. Hereditas. 2025;162(1):42. https://doi.org/10.1186/s41065-025-00409-4.
27. Xu S, Jiang L, Zhang Z et al. Deciphering Bisphenol A (BPA)-elicited osteoarthritis mechanisms through network toxicology and molecular docking, then de novo generation of novel therapeutic candidates. BMC Pharmacol Toxicol. 2025. https://doi.org/10.1186/s40360-025-01069-w.
28. Wang Z, Wang C, Wan C, et al. Exploring the mechanisms of Benzo[a]Pyrene-induced lung adenocarcinoma based on network toxicology and machine learning. BMC Pharmacol Toxicol. 2025. https://doi.org/10.1186/s40360-025-01064-1.
29. Lin L, Lapi F, Galuzzi BG, Vanoni M, Alberghina L. Mechanistically informed machine learning links non-canonical TCA cycle activity to warburg metabolism and hallmarks of malignancy. PLoS Comput Biol. 2025;21(12):e1013384. https://doi.org/10.1371/journal.pcbi.1013384.
30. Xia M, Liu X, Wang K, Liang B, Xiao P. From sweetener to risk factor: network toxicology, molecular docking and molecular dynamics reveal the mechanism of aspartame in promoting coronary heart disease. Chemico-Biol Interact. 2026;424:111876. https://doi.org/10.1016/j.cbi.2025.111876.
31. Xiong G, Shi H, Jiang Y, et al. Combined effects of nanoplastics and 3-BHA at environmentally relevant concentrations significantly aggravated kidney injury via TGF-β/SMAD signaling pathway in mice. Ecotoxicol Environ Saf. 2025;303:119014. https://doi.org/10.1016/j.ecoenv.2025.119014.
32. Kim EK, Eui-Ju C. Pathological roles of MAPK signaling pathways in human diseases. Biochim Biophys Acta (BBA) Mol Basis Dis. 2010;1802(4):396–405. https://doi.org/10.1016/j.bbadis.2009.12.009.
33. Achkar IW, Abdulrahman N, Al-Sulaiti H, Joseph JM, Uddin S, Mraiche F. Cisplatin based therapy: the role of the mitogen activated protein kinase signaling pathway. J Transl Med. 2018;16(1):96. https://doi.org/10.1186/s12967-018-1471-1.
34. Li J, Meng Y, Wu X, Sun Y, et al. Polyamines and related signaling pathways in cancer. Cancer Cell Int. 2020;20(1):539. https://doi.org/10.1186/s12935-020-01545-9.
35. Fan Li-li, Chen X, Zhang Xiao-yu, Li Ze-min, Shen Y. Octreotide-paclitaxel conjugate reverses paclitaxel resistance by P38 mitogen-activated protein kinase (MAPK) signaling pathway in A2780/taxol human ovarian cancer cells. Med Sci Monit. 2020. https://doi.org/10.12659/MSM.922612.
36. Hayes JD, Dinkova-Kostova AT, Tew KD. Oxidative stress in cancer. Cancer Cell. 2020;38(2):167–97. https://doi.org/10.1016/j.ccell.2020.06.001.
37. Gong Z, Du M, Li Y, et al. Machine learning identifies TIME subtypes linking EGFR mutations and immune states in lung adenocarcinoma. NPJ Digit Med. 2025;8(1):796. https://doi.org/10.1038/s41746-025-02172-2.

38. Xia W, Liu J, Chen R, et al. Molecular subtypes and prognostic signature rooted in disulfidptosis highlight tumor microenvironment in lung adenocarcinoma. Chin J Cancer Res. 2025;37(5):796–820. https://doi.org/10.21147/j.issn.1000-9604.2025.05.11.
39. Zhang P, Liang X, Ye B, et al. Metabolic reprogramming signature predicts immunotherapy efficacy in lung adenocarcinoma: targeting SLC25A1 to overcome immune resistance. Chin J Cancer Res. 2025;37(6):1000–19. https://doi.org/10.21147/j.issn.1000-9604.2025.06.11.
40. Cheng Y, Ma Xue-lei, Wei Yu-quan, et al. Potential roles and targeted therapy of the CXCLs/CXCR2 axis in cancer and inflammatory diseases. Biochim Biophys Acta (BBA) Rev Cancer. 2019;1871(2):289–312. https://doi.org/10.1016/j.bbcan.2019.01.005.
41. Dai H, Zhao K, Zhao Y, et al. Machine learning model in multi-omics perspective demystifies the prognostic significance of crotonylation heterogeneity in clear cell renal cell carcinoma. BMC Urol. 2025;25(1):229. https://doi.org/10.1186/s12894-025-01914-4.
42. Gao L, Wei Z, Ying F, et al. Glutamine metabolism prognostic index predicts tumour microenvironment characteristics and therapeutic efficacy in ovarian cancer. J Cell Mol Med. 2024;28(7):e18198. https://doi.org/10.1111/jcmm.18198.

Publisher's Note