Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 21 No. 11 (2025) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v21i11.56915

PAPER

Big Data Analytics Reveals Pyrethrins' Breast Cancer Risks: A Deep Learning-Enhanced Study Combining Mendelian Randomization and Molecular Dynamics

Zikang Jiang[1,2],
Jinghui Sung[1,2],
Weijie Li[1,2], **Yixin Zhuang**[3],
Yuanpeng Huang[1,2](✉),
Ting-Yu Chen[4]

[1]Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, China

[2]Xiamen Hospital, Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China

[3]Xiamen University, Xiamen, Fujian, China

[4]National Chin-Yi University of Technology, Taizhong, Taiwan, China

huangyp998@xmu.edu.cn

ABSTRACT

Pyrethrins (PYs), widely used insecticides, raise concerns about chronic carcinogenicity. To address limited mechanistic evidence for PY-associated breast cancer (BC) risk, we integrated multi-omics data via a computational framework combining network toxicology, Mendelian randomization (MR), deep learning-enhanced screening, molecular docking, and dynamics simulations. Computational screening identified 16 high-affinity targets with binding energies < -7.5 kcal/mol, indicative of stable interactions. Molecular dynamics simulations further validated the structural stability of PY-target complexes. MM/PBSA analyses confirmed that both PY I and II exhibit thermodynamically spontaneous interactions with diverse targets, demonstrating binding free energies ranging from -9.53 to -27.37 kcal/mol. Complex interactions among targets were constructed. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analyses revealed significant associations between these targets and BC pathways. Our findings indicate that PYs may promote breast carcinogenesis by disrupting oncogenic signaling networks, justifying prioritized regulatory reassessment and biomonitoring. Nevertheless, further validation through *in vivo* and *in vitro* experiments remains imperative.

KEYWORDS

big data, deep learning, pyrethrin (PY), breast cancer (BC), network toxicology, Mendelian randomization (MR), molecular docking, molecular dynamics simulations

1 INTRODUCTION

Pyrethrins (PYs), derived from *Tanacetum cinerariifolium*, are primarily composed of Pyrethrin I (PY I) and II (PY II). These compounds have been utilized for over a century in pest management due to their potent bioactivity and rapid degradation kinetics [1]. In agricultural practices, these compounds are extensively employed to

Jiang, Z., Sung, J., Li, W., Zhuang, Y., Huang, Y., Chen, T.-Y. (2025). Big Data Analytics Reveals Pyrethrins' Breast Cancer Risks: A Deep Learning-Enhanced Study Combining Mendelian Randomization and Molecular Dynamics. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(11), pp. 81–96. https://doi.org/10.3991/ijoe.v21i11.56915

Article submitted 2025-05-29. Revision uploaded 2025-07-21. Final acceptance 2025-07-26.

control arthropod pests in crop production systems, formulate pet hygiene products, and mitigate the transmission of vector-borne diseases.

Research has explored the mechanisms through which pyrethroids, such as permethrin and cypermethrin, increase the risk of breast cancer (BC) by interacting with hormone receptors and cellular pathways [2]. Exposure to pyrethroids can lead to changes in thyroid hormone levels, which are dose-dependent and can affect growth and development, including breast development [3]. In contrast to organophosphate pesticides, such as chlorpyrifos, which are associated with an increased risk of BC [4], evidence for PYs remains inconclusive. Notably, mechanistic toxicology studies identify PYs as endocrine-disrupting chemicals that may affect reproductive and metabolic functions by interfering with hormonal regulation [5]. However, a comprehensive understanding of the causal link between PY exposure and BC risk, the identification of novel molecular targets beyond canonical endocrine pathways, and the biophysical validation of PY interactions with BC-relevant proteins remain significant knowledge gaps. These gaps stem from limitations inherent in single-method approaches: epidemiological studies struggle with confounding and establishing causality for specific mechanisms; isolated *in vitro* or computational studies lack population-level genetic support and rigorous dynamic validation. Therefore, to address these interconnected gaps and systematically evaluate the potential impact of PYs on BC risk, this study employs an integrative analytical framework combining network toxicology, Mendelian randomization (MR) for causal inference, deep learning for broad target prediction, molecular docking for affinity screening, and molecular dynamics (MD) simulations with binding free energy calculations for structural and energetic validation. This multi-omics big data integration overcomes the limitations of individual methods, providing cross-scale evidence to predict PYs' breast carcinogenicity while addressing mechanistic uncertainties in risk assessments, supporting evidence-based policies with scientific rigor.

2 MATERIALS AND METHODS

2.1 Targets collection

The targets of PY I and II were primarily retrieved from the ChEMBL database. Additionally, the canonical chemical structures of PY I and II were submitted to SwissTargetPrediction for computational target prediction to augment the dataset. Potential BC-related targets were systematically collated from the GeneCards, OMIM, PharmGKB, and Therapeutic Target databases. Concurrently, we downloaded the GSE42568 and GSE15852 datasets from the GEO database.

2.2 Identification of overlapping targets

Intersection analysis was performed between the targets of PY I and II and BC-associated targets curated from public databases, with common targets identified and visualized via venn diagrams generated in R Studio. The resultant overlapping targets were postulated as potential targets through which PY I and II, respectively, may contribute to BC pathogenesis.

2.3 Deep learning predicts PYs binding to shared BC targets

DeepPurpose [6], a deep learning-based framework, employs an encoder-decoder architecture to predict drug-target interactions (DTI). We employed its MPNN-CNN architecture, where MPNN processes compound SMILES strings to capture graph-based molecular features, while 1D-CNN extracts local residue motifs from protein sequences. The concatenated embeddings were fed into a 3-layer MLP (512→256→1 neurons) to predict binding affinities, which quantify the interaction strength between the drug and target. The model dynamically adapts to regression tasks or classification tasks (binary binding prediction) based on the input data characteristics. The SMILES notations of PY I and II were retrieved from PubChem, and their corresponding amino acid sequences of overlapping BC target proteins were obtained from the PDB database. Using the pre-trained DeepPurpose framework (model_MPNN_CNN_DAVIS) without modification, as documented in the library (https://github.com/kexinhuang12345/DeepPurpose), binding affinity scores between PY I and II and their respective overlapping targets were calculated and ranked in descending order based on prediction confidence. The DeepPurpose framework was selected due to its rigorously benchmarked performance against state-of-the-art DTI methods (e.g., KronRLS, DeepDTA) in its original publication. As our study utilizes its pre-trained model for inference—not proposing a new algorithm—benchmarking was deemed redundant given the exhaustive validation in its source work. Instead, we prioritized orthogonal validation via molecular docking, MD simulations, and MR to ensure prediction reliability.

2.4 MR-based causal inference for BC target identification

This study applied two-sample MR under core assumptions, using standardized protocols. Instrumental variables (IVs) were selected from GWAS summary data with rigorous quality controls. Blood pQTL (plasma protein quantitative trait loci) data from FinnGen [7] included Olink (619 samples, 2,925 proteins) and SomaScan platforms (828 samples, 7,596 proteins). BC datasets (FinnGen code: DF12) were also retrieved, comprising 5 distinct outcome definitions: carcinoma in situ of breast, intraductal carcinoma in situ, lobular carcinoma in situ, ER-negative BC, and ER-positive BC. TSMR analyses paired BC outcomes with Olink and SomaScan pQTL data. IVs ($F > 10$, $r^2 < 0.001$, $p < 5 \times 10^{-6}$) were selected from exposure GWAS. Exposure-outcome SNP data (β, SE, alleles) were harmonized, resolving strand/palindromic conflicts. Causal effects were estimated via *IVW* (primary method), supplemented by MR-Egger and weighted median models. The *IVW* method achieves unbiased estimation by integrating the effect ratios of individual SNPs, expressed as formula (1).

$$\beta_{IVW} = \frac{\Sigma\left(\beta_{y_j}\beta_{x_j} \cdot \sigma_{y_j}^{-2}\right)}{\Sigma\left(\beta_{x_j}^{2} \cdot \sigma_{y_j}^{-2}\right)} \tag{1}$$

Where β_{y_j} and β_{x_j} represent the SNP-outcome and SNP-exposure effect estimates, respectively, and $\sigma_{y_j}^{-2}$ denotes the inverse variance of the SNP-outcome association.

Furthermore, we assessed the heterogeneity of IVs using Cochran's Q test, examined horizontal pleiotropy through the intercept term of MR-Egger regression, detected and removed outlier SNPs via the MR-PRESSO method, and validated the influence of individual SNPs on overall effects by leave-one-out analysis.

2.5 Molecular docking screening for high-affinity PY targets in BC

BC-associated protein targets (DeepPurpose score >7 or MR-linked) underwent molecular docking with PYs I/II. Ligand structures were sourced from PubChem;

proteins were obtained via RCSB PDB or AlphaFold predictions. Dockey [8] facilitated batch docking: ligands/receptors were preprocessed (removing solvents/cofactors), grids were generated via AutoGrid, and binding energies were calculated using AutoDock4. Each docking produced 100 conformations, with the lowest-energy results selected.

2.6 MD simulations

To further validate the binding stability of PYs with BC targets and elucidate their interaction mechanisms, MD simulations were performed on protein targets corresponding to genes with docking binding energies < –7.5 kcal/mol. The initial structures of the complexes were derived from the optimal docking poses obtained in the previous step. Simulations were conducted using GROMACS 2025.0. Partial atomic charges of PYs were calculated via Multiwfn [9] and ORCA [10], with their topology files generated by SobTop [11]. Protein topology files and positional restraint files were prepared using GROMACS tools. Each protein-PY complex was centered in a dodecahedron box with a 1.0 Å buffer distance. The system was solvated with the SPC water molecular model and neutralized by adding Na^+ and Cl^- ions. Energy minimization was performed using the steepest descent method (10,000 steps) followed by conjugate gradient optimization (10,000 steps). The system underwent two sequential equilibration phases: 100 ps constant Number, Volume, Temperature (NVT) ensemble and 100 ps constant Number, Pressure, Temperature (NPT) ensemble. The temperature (310 K) and pressure (1 bar) were set. MD simulations were then executed for 100 ns. Trajectory analyses included root mean square fluctuation (RMSF), root mean square deviation (RMSD), radius of gyration (Rg), intramolecular hydrogen bonding, and free energy landscape profiling.

2.7 Calculation of binding free energy

The binding free energy was calculated using the Molecular Mechanics/Poisson-Boltzmann Surface Area (MM/PBSA) method [12]. The core formula is defined as (2)–(4).

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - \mathrm{T}\Delta S \tag{2}$$

$$\Delta E_{\text{MM}} = \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} \tag{3}$$

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB}} + \Delta G_{\text{SA}} \tag{4}$$

where: ΔE_{MM} is the gas-phase molecular mechanical energy, ΔG_{solv} is the solvation free energy, $-T\Delta S$ represents the entropic contribution, ΔE_{vdw} denotes the van der Waals interaction energy, ΔE_{elec} is the electrostatic interaction energy, ΔG_{PB} is the polar solvation energy, and ΔG_{SA} is the nonpolar solvation energy. However, in practical MM/PBSA calculations, the entropic contribution is omitted due to minimal conformational changes between the bound and unbound states of the receptor-ligand complex and the significant computational errors associated with entropy estimation [13]. This study calculated the binding free energy during the stable 60–100 ns phase of the MD simulations using the gmx_MMPBSA tool [14].

2.8 Construction of protein-protein interaction (PPI) networks

Core targets of PY I/II with binding energy < –5 kcal/mol were analyzed via STRING (Homo sapiens; medium confidence = 0.400; disconnected nodes excluded). PPI network TSV files were evaluated in Cytoscape 3.10.1, where cytoNCA calculated degree/betweenness centrality. MCODE identified protein clusters using optimized parameters (degree cutoff = 2, node score = 0.2, k-core = 2).

2.9 Analysis of gene function and pathway enrichment

GO and KEGG enrichment analyses of PY I/II core targets in BC were performed. Gene identifiers were standardized (org.Hs.eg.db) with significance thresholds ($p < 0.05$, FDR $q < 0.05$). Redundant GO terms were filtered via semantic similarity reduction. Integrated with MD simulations and literature, GO analysis covered Biological Process, Molecular Function, and Cellular Component categories, assessed via hypergeometric testing. KEGG prioritized BC-related pathways linked to MD-simulated targets, ranked by gene count.

3 RESULTS

3.1 Shared targets of PYs and BC in public databases

After merging and deduplicating the targets of PY I and II, 558 and 489 potential targets were identified, respectively. BC-related targets were merged and deduplicated, yielding a total of 16,076 targets. We performed differential expression analysis on two gene expression datasets (GSE42568 and GSE15852) and compiled the top 500 upregulated and 500 downregulated genes into potential BC target sets (Figure S1a–S1b). Intersection analysis (Figure S1c–S1h) revealed 491 and 447 common targets between PY I and II and the general BC database, respectively. Additionally, PY I and II exhibited 25 and 38 overlapping genes with GSE42568 and 47 and 51 overlapping genes with GSE15852, respectively.

3.2 Preliminary Screening with DeepPurpose

Using these UniProt IDs of intersection genes, protein sequences were acquired from the RCSB Protein Data Bank. Both the protein sequence information and the SMILES notations of PY I and II were subsequently input into the DeepPurpose model for binding affinity prediction, with results presented as numerical scores (Tables S1 and S2). In the model_MPNN_CNN_DAVIS framework, the output scores directly reflect predicted binding affinity strength, where higher values indicate stronger compound-target interactions. The DTI scores between PY I and the intersecting targets ranged from 5.145 to 8.993, while those between PY II and the targets spanned 4.557 to 8.337 (Figure 1). Subplots (a), (b), and (c) depict binding affinity predictions for PY I against intersecting targets derived from the generic public databases GSE42568 and GSE15852, respectively. Subplots (d), (e), and (f) illustrate analogous predictions for PY II against intersecting targets from the same three datasets. Compounds with scores exceeding 7 (totaling 269) were selected for subsequent molecular docking to refine candidate prioritization.

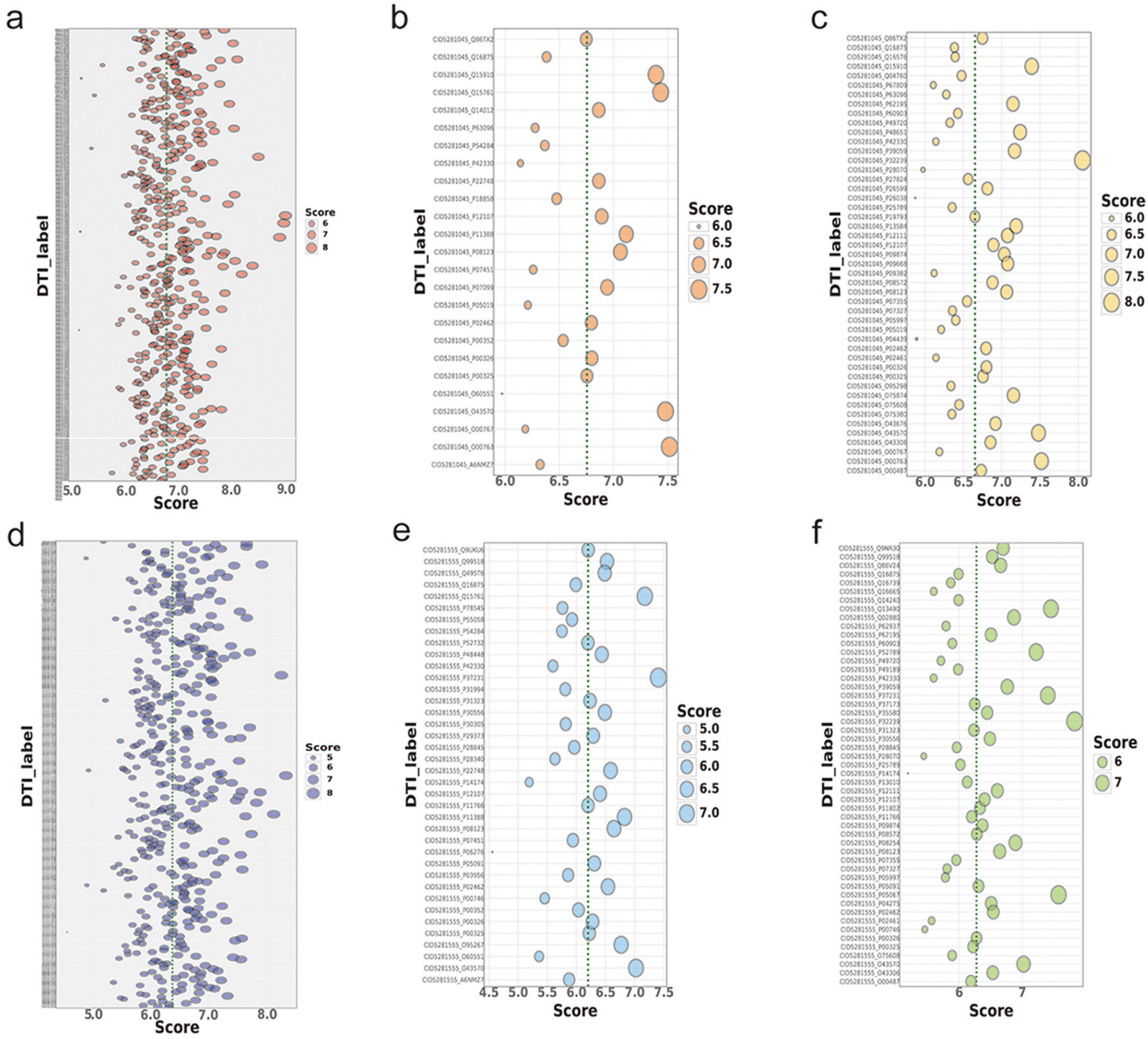

Fig. 1. Bubble plot of DeepPurpose-predicted binding affinity scores between PY I and II and intersecting targets

3.3 Identification of BC-associated targets via MR

From the GWAS data of pQTLs, we yielded 69288 SomaScan-derived and 27146 Olink-derived SNPs as IVs for pQTL analysis (Tables S3 and S4). TSMR analyses were conducted between these protein exposures and various BC subtypes, retaining exposures with consistent effect directions across five MR methods (Figure S2a–S2e). The IVW method served as primary evidence with selection thresholds set at $p < 0.05$ and $FDR < 0.5$, while preserving exposures with horizontal pleiotropy $p > 0.05$. In SomaScan data, we identified NCR3 (OR = 0.9314, 95% CI: 0.9006 – 0.9632, $p = 3.301 \times 10^{-5}$) as protective against ER-negative BC and FBXO3

(OR = 0.9400, 95% CI: 0.9097 – 0.9713, $p = 2.134 \times 10^{-4}$) as protective for ER-positive BC, while RNF146 (OR = 1.0793, 95% CI: 1.0362 – 1.1241, $p = 2.385 \times 10^{-4}$) and EZH2 (OR = 1.0769, 95% CI: 1.0388 – 1.1163, $p = 5.411 \times 10^{-5}$) emerged as risk factors for ER-positive BC. From Olink platform data, TEK (OR = 0.9382, 95% CI: 0.9082 – 0.9692, $p = 1.183 \times 10^{-4}$) showed protective effects against ER-positive BC. Notably, neither platform's pQTL exposures demonstrated significant causal relationships with other BC subtypes in our MR analyses.

Methodological evaluations confirmed the robustness of IVs across analytical parameters. Heterogeneity analyses indicated consistency across IVs (p-values of Cochran's Q test are > 0.05). Pleiotropy assessments through intercept-based (MR-Egger intercept test) and outlier-adjusted (MR-PRESSO global tests) approaches revealed no detectable horizontal bias (all pleiotropy $p > 0.05$). Corresponding SNP effect forest plots for these five genes are shown in Figure S2f–S2j. Sensitivity evaluations employing sequential variant exclusion (leave-one-out analyses) confirmed stable causal estimates unaffected by individual IVs, reinforcing the reliability of observed associations (Figure S2k–S2o). Comprehensive quality control metrics are documented in Tables S5–S7.

3.4 Virtual screening using molecular docking

Molecular docking identified high-affinity BC targets for PY I/II, using a binding energy threshold of < –5 kcal/mol. PY I bound 148 targets, and PY II 67 (Table S8). Strong binding (≤ –7.5 kcal/mol) prompted visualization of docking conformations, revealing hydrogen bonds, hydrophobic contacts, and electrostatic interactions. PY, I exhibited binding energies of 8.42 kcal/mol with MAOB (1OJA), 9.13 kcal/mol with TEK (2OO8), 7.55 kcal/mol with NCR3 (3NOI), 8.77 kcal/mol with EZH2 (4MI5), 7.60 kcal/mol with TNKS2 (5BXO), 8.24 kcal/mol with TOP2A (5NNE), 8.97 kcal/mol with RPS6KB1 (5WBH), 7.80 kcal/mol with MT-CO1 (5Z62), and 7.61 kcal/mol with SCN7A (7TJ9) (Figure 2a–2i). PY II demonstrated binding energies of 9.21 kcal/mol with TEK (2OO8), 8.96 kcal/mol with CAMK2G (2V7O), 10.54 kcal/mol with NCR3 (3NOI), 11.61 kcal/mol with TNKS2 (5BXO), 7.66 kcal/mol with SOS1 (5OVE), 9.25 kcal/mol with CSNK2A1 (7A4Q), and 9.09 kcal/mol with PPARG (8KFC) (Figure 2j–2p).

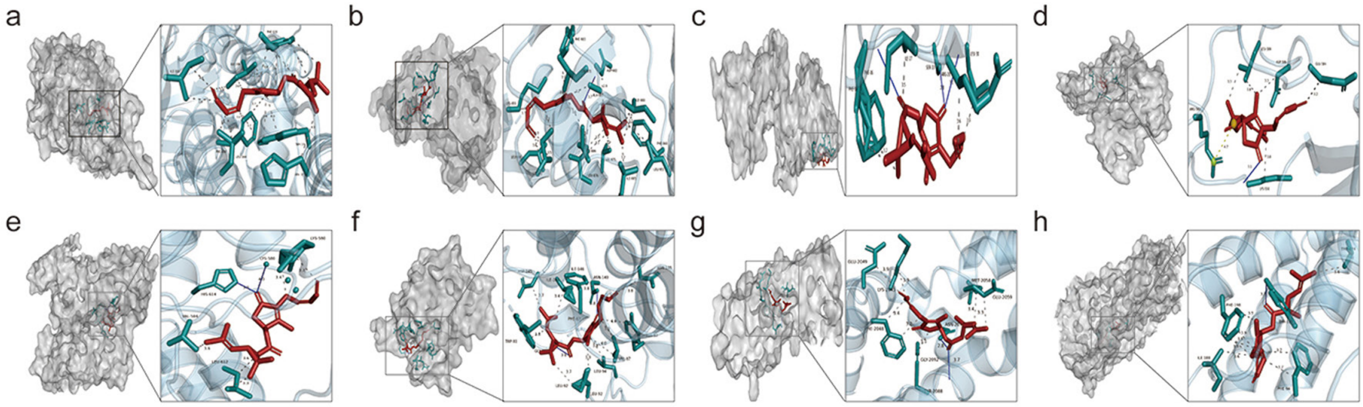

Fig. 2. *(Continued)*

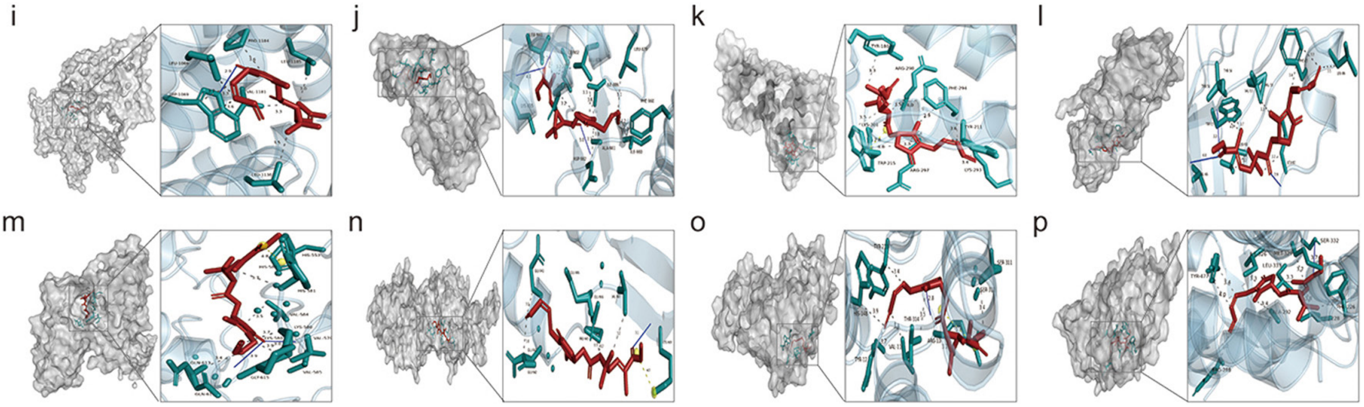

Fig. 2. Molecular docking results of PY I and II with BC targets

3.5 MD simulations and binding stability analyses

The MD simulation analyses encompassed four key metrics: RMSD, RMSF, Rg, and hydrogen bond interactions, as illustrated in Figure 3. Panels (a)–(i) show PY I complexes with MAOB (1OJA), TEK (2OO8), NCR3 (3NOI), EZH2 (4MI5), TNKS2 (5BXO), TOP2A (5NNE), RPS6KB1 (5WBH), MT-CO1 (5Z62), and SCN7A (7TJ9), analyzing RMSD, RMSF, Rg, and hydrogen bond dynamics. Panels (j)–(p) detail PY II bound to TEK (2OO8), CAMK2G (2V7O), NCR3 (3NOI), TNKS2 (5BXO), SOS1 (5OVE), CSNK2A1 (7A4Q), and PPARG (8KFC), with corresponding trajectory profiles. MD simulation results revealed that all investigated systems exhibited satisfactory structural stability and dynamic evolution characteristics during the equilibration phase. Analysis of RMSD demonstrated that most ligand-target complexes displayed characteristic conformational adjustments in the initial simulation phase (0–20 ns), followed by entry into a stable fluctuation regime. Notably, the majority of systems attained thermodynamic equilibrium states after 40 ns. Notably, complexes with MAOB, TEK, MT-CO1, SCN7A, SOS1, CSNK2A1, and PPARG exhibited rapid RMSD convergence, suggesting stable binding modes with minimal backbone rearrangements. In contrast, the PY I-NCR3 complex experienced pronounced conformational fluctuations between 0–50 ns, which progressively stabilized and reached equilibrium after 60 ns.

The RMSF profile exhibited a characteristic biphasic distribution pattern, reflecting the displacement magnitude of atoms or residues relative to their average positions during the simulation. Elevated RMSF values (> 0.2 nm) were indicative of highly flexible or disordered regions, including loop structures, terminal residues, and unfolded segments. These dynamic domains likely participate in conformational rearrangements during complex formation and may contribute to functional regulation following ligand binding, as evidenced by their enhanced structural adaptability. The Rg trajectories demonstrated stable fluctuations throughout most of the simulation period, indicating maintained structural integrity of the complexes. For the TNKS2 and RPS6KB1 complexes with PY I, the Rg values exhibited a reduction compared to initial values, likely reflecting PY I-induced protein compaction phenomena.

In analyses of hydrogen-bonding patterns, the NCR3-PY I complex initially displayed no detectable hydrogen bonds but subsequently developed close molecular contacts (< 0.35 nm) and established hydrogen bonds following stabilization of the RMSD profile, corroborating the attainment of a stable state in the final system. Contrastingly, the MT-CO1-PY I complex exhibited transient hydrogen bond disappearance during intermediate stages, potentially attributable to solvent molecule intercalation at the ligand-receptor interface. This phenomenon suggests a possible solvent-mediated bridging mechanism supplanting direct intermolecular interactions. The SOS1-PY II complex demonstrated limited hydrogen bond formation, implying that the primary driving force for their association likely originates from alternative non-hydrogen bonding intermolecular forces, potentially including hydrophobic interactions or van der Waals forces. We constructed the Gibbs free energy landscape to characterize conformational states of the biomolecular system. From Figure S3, we can intuitively find that in our study, all complexes have a stable lowest-energy conformation state.

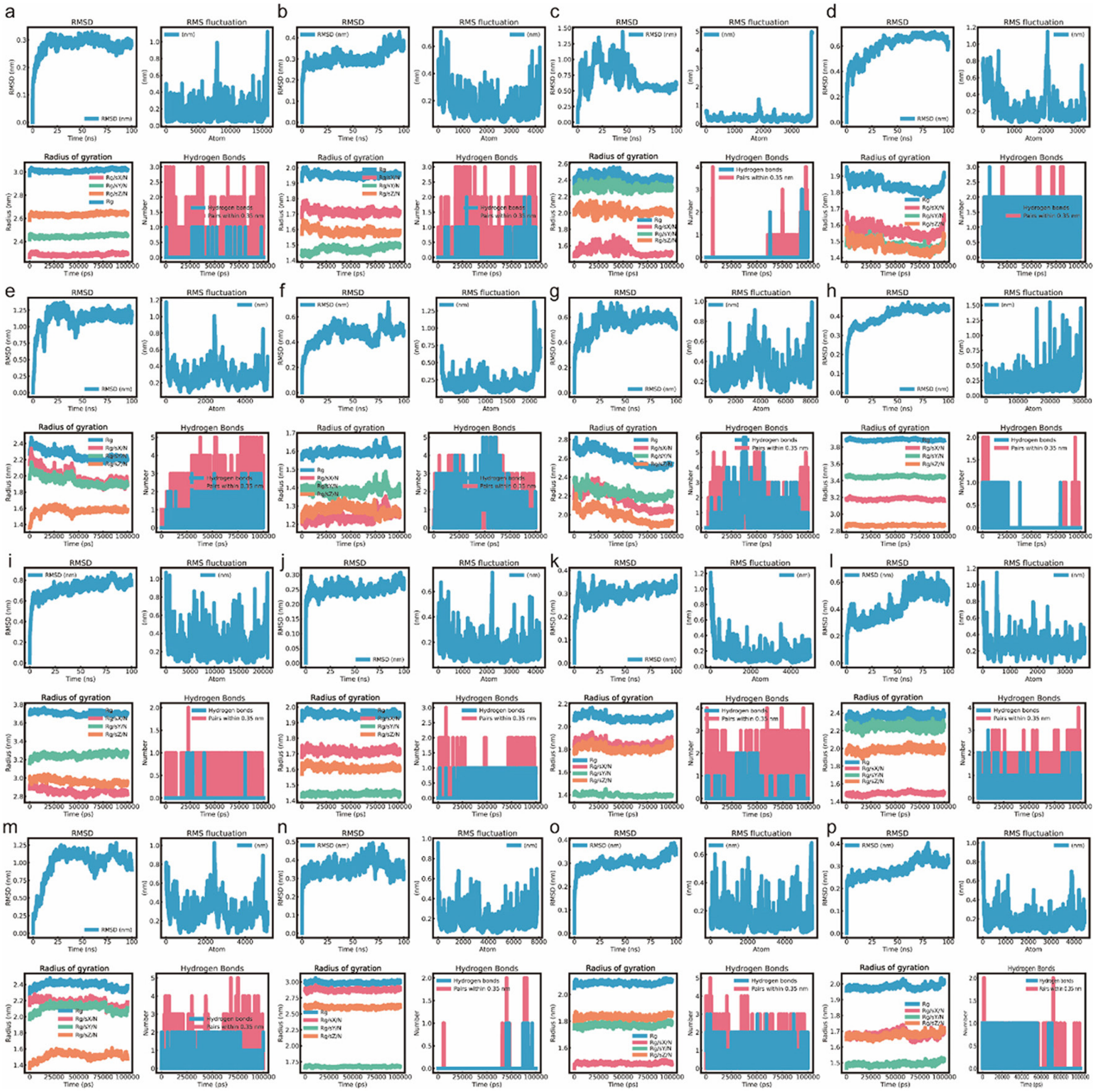

Fig. 3. MD simulation results

3.6 Assessment of average binding free energy

The binding free energies of PY I and II complexes with various target proteins were computationally evaluated using the MM/PBSA method (refer to Table 1). All complexes exhibited negative total binding free energies, indicating thermodynamically spontaneous binding processes. Notably, PY I demonstrated superior binding affinity with RPS6KB1, TOP2A, and TNKS2, while PY II showed preferential binding to CSNK2A1. Energy decomposition analysis revealed gas-phase contributions as the primary binding driver, ranging from –41.94 to –234.45 kcal/mol. Electrostatic interactions dominated in most of the complexes. The exceptional ΔE_{elec} value of –214.72 kcal/mol observed in the PY I-TOP2A complex likely correlates with the high density of charged residues within its binding pocket. Van der Waals interactions displayed moderate contributions, suggesting consistent hydrophobic complementarity across binding interfaces. Solvation effects exerted destabilizing influences, with polar solvation energy accounting for 91.5%–98.6% of this penalty. The PY II-TEK complex exhibited particularly high ΔG_{PB}, significantly counteracting its favorable gas-phase binding. Nonpolar solvation contributions remained minimal, showing positive correlation with ligand hydrophobic surface area exposure. Structural analogs displayed marked binding selectivity differences: PY I demonstrated 84.7% higher affinity for TNKS2 compared to PY II, while PY II showed 36.9% enhanced binding to TEK relative to PY I. These findings suggest distinct electronic and steric complementarity between homologs and respective targets, potentially mediated by variations in binding pocket polarity and residue composition.

Table 1. Binding free energies calculated by the MM/PBSA method (kcal/mol)

Ligand-Targets	ΔE_{vdw}	ΔE_{elec}	ΔG_{PB}	ΔG_{SA}	ΔE_{MM}	ΔG_{solv}	ΔG_{bind}
PY I-EZH2	–24.77	–134.52	145.54	–2.62	–159.29	142.92	–16.37
PY I-TEK	–26.1	–109.35	130.12	–2.91	–135.45	127.21	–8.24
PY I-TOP2A	–19.73	–214.72	213.88	–3.09	–234.45	210.79	–23.66
PY I-RPS6KB1	–31.49	–145.79	153.51	–3.59	–177.29	149.92	–27.37
PY I-MT-CO1	–28.44	–13.5	30.37	–2.81	–41.94	27.56	–14.38
PY I-SCN7A	–34.3	–38.18	56.69	–3.35	–72.48	53.34	–19.14
PY I-TNKS2	–25.62	–111.37	117.71	–2.66	–136.98	115.05	–21.94
PY I-NCR3	–12.84	–76.52	76.51	–1.63	–89.36	74.88	–14.48
PY I-MAOB	–39.55	–122.68	150.42	–3.85	–162.23	146.57	–15.66
PY II-TNKS2	–36.36	–83.97	111.7	–3.24	–120.34	108.46	–11.88
PY II-CAMK2G	–40.32	–128.98	159.67	–3.7	–169.3	155.97	–13.34
PY II-NCR3	–41.05	–51.73	83.5	–3.36	–92.78	80.13	–12.65
PY II-CSNK2A1	–33.82	–79.07	93.4	–2.96	–112.88	90.45	–22.44
PY II-TEK	–38.02	–185.2	216.12	–4.17	–223.22	211.95	–11.28
PY II-PPARG	–25.13	–121.41	140.1	–3.09	–146.53	137.01	–9.53
PY II-SOS1	–15.33	–74.75	82.22	–2.09	–90.08	80.13	–9.95

3.7 Target interaction network

In the constructed PY I-BC PPI network, we identified 137 nodes and 443 edges (Figure 4a). CytoNCA analysis of core target proteins derived from MD simulations revealed the degree centrality values. MCODE plugin analysis further identified functionally significant clusters (Figure 4b), among which the sodium channel-associated cluster (SCN7A, CACNA1C) exhibited the highest score. The cytochrome P450/UGT family cluster (CYP2A1, UGT1A5) ranked second, followed by the adenylate cyclase-related cluster (ADCY1). Additional clusters included collagen-associated proteins (COL18A1), a cell cycle regulatory module (MDM2, TNKS, TOP2A, RNF146), and a mitochondrial respiratory cluster (MT-CO1). These findings suggest that PY I may influence BC pathogenesis through modulation of sodium/calcium ion transport, xenobiotic metabolism, cAMP-mediated signaling, extracellular matrix remodeling, cell cycle regulation, and mitochondrial respiration.

In contrast, the PY II-BC PPI network comprised 52 nodes and 317 edges (Figure 4e). CytoNCA analysis of MD simulation data highlighted key targets including TNKS2, TEK, CSNK2A1, PPARG, BIRC2, and SOS1. MCODE clustering prioritized three functionally distinct modules (Figure 4f): a calcium channel-enriched cluster (CACNA1 family proteins), a ribosylation-associated cluster (TNKS2 interactors), and a cytokine/calcium signaling cluster (CAMK2G, JAK2). These results implicate PY II in calcium homeostasis regulation, post-translational modification processes, and cytokine-calcium signaling crosstalk during BC progression.

3.8 GO and KEGG enrichment analysis

PY I targets yielded 758 significant GO entries (580 biological processes, 67 cellular components, and 111 molecular functions). Key cellular components included transmembrane transporter complexes, synaptic membranes, and plasma membrane rafts. Molecular functions involved ion channel activity, kinase activities (serine/threonine/tyrosine), and glycosyltransferase activity. Biological processes spanned xenobiotic response, MAPK/TOR/ERK/Wnt signaling, immune activation, and vascular regulation (see Figure 4d). KEGG analysis identified 92 pathways, with prioritized BC-relevant pathways: calcium/MAPK signaling, cytochrome P450-mediated metabolism, Rap1/estrogen signaling, chemical carcinogenesis (receptor activation/DNA adducts), and apoptosis (see Figure 4c).

PY II analysis identified 465 biological processes (BPs), 55 cellular components (CCs), and 63 molecular functions (MFs). Key CCs included transporter complexes, membrane rafts, postsynaptic membranes, and nuclear envelopes. MFs involved G protein-coupled receptor activity, ion channel function, kinase activities (tyrosine/serine-threonine), and NAD+ ADP-ribosyltransferase activity. Targeted BPs encompassed hypoxia response, MAPK/Wnt signaling, kinase regulation, inflammation, and hormone secretion (see Figure 4h). KEGG analysis highlighted BC-relevant pathways: MAPK, GnRH, HIF-1, Wnt, NF-κB, and proteoglycan-mediated oncogenesis, linking PY II to apoptosis-efferocytosis imbalance and cancer progression via proteoglycan dysregulation (see Figure 4g).

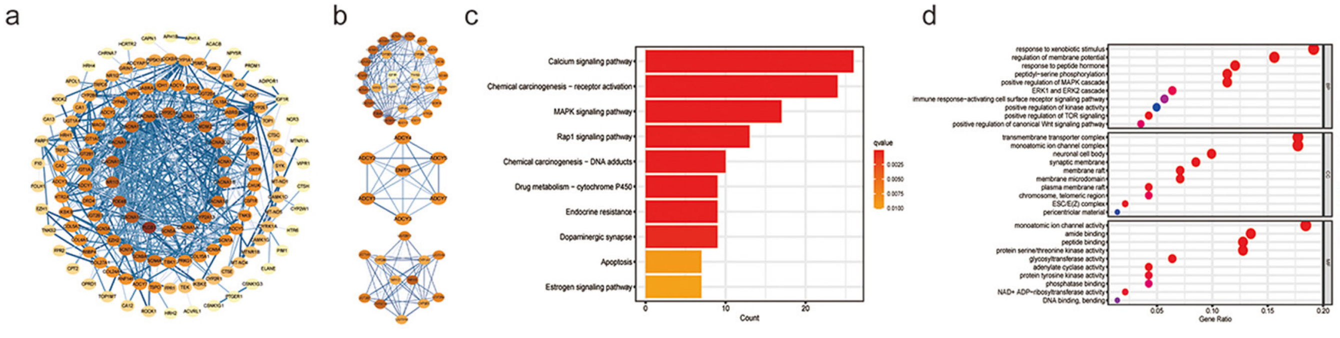

Fig. 4. *(Continued)*

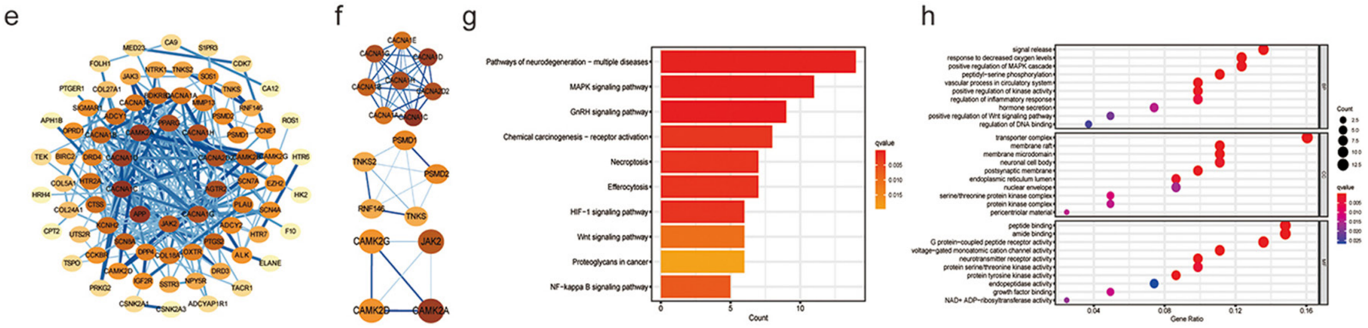

Fig. 4. PPI network diagram and GO/KEGG enrichment analysis plots

4 DISCUSSION

Our *in silico* analyses demonstrated that PY I and II share multiple overlapping molecular targets with breast carcinogenesis pathways, with molecular docking and MD simulations revealing exhibited robust binding affinities. Notably, key targets identified by MR as causally linked to BC risk subtypes, specifically NCR3 (ER-), TEK (ER+), and EZH2 (ER+), also demonstrated high-affinity binding to PY I and/or PY II in molecular docking (≤ –7.5 kcal/mol). This convergence of causal genetic evidence and computational biophysical evidence strengthens the biological relevance of these interactions in the context of PY-associated BC risk. In summary, PY I exhibited high-affinity binding interactions with key molecular targets: MAOB, TEK, NCR3, EZH2, TNKS2, TOP2A, RPS6KB1, MT-CO1, and SCN7A. At the same time, PY II demonstrated strong binding interactions with distinct targets, including TEK, CAMK2G, NCR3, TNKS2, SOS1, CSNK2A1, and PPARG. These findings collectively suggest that PYs may possess intrinsic carcinogenic potential associated with BC development.

Elevated monoamine oxidase B (MAOB) expression correlates with estrogen receptor (ER)-negative BC, particularly in triple-negative BC (TNBC). Estrogen-related receptors (ERRs) transcriptionally activate MAOB, while ER suppresses this regulation via receptor crosstalk. Stromal MAOB expression associates with poor metastatic outcomes, potentially serving as a therapeutic target or prognostic marker. Hypoxic bone metastasis microenvironments upregulate MAOB through HIF-1α-mediated Sp1/Sp3 transcription factor competition at its promoter. Stromal MAOB may drive metastasis via metabolic reprogramming and tumor microenvironment remodeling [15]. PY I could promote BC progression by activating oncogenic pathways through MAOB.

The TEK (TIE2) receptor, central to Angiopoietin signaling, critically regulates BC pathogenesis. PYs I/II showed strong TEK binding in our analyses. TEK promotes metastasis via angiogenesis and cancer cell invasiveness. While TEK overexpression correlates with ER+/PR+ tumors [16], Finngen MR data revealed TEK variants inversely linked to BC risk. Clinically, elevated plasma Tie2 levels predicted better neoadjuvant therapy outcomes (bevacizumab/chemotherapy), with high pre-treatment sTIE2 indicating enhanced efficacy [17]. We propose sTIE2 acts as a decoy receptor by sequestering Ang-1/Ang-2, reducing ligand bioavailability. This suppresses endothelial TIE2 signaling, enhances vascular stability, and inhibits metastasis. However, molecular modeling suggests PYs bind sTIE2, blocking its ligand-scavenging function. This may restore Angiopoietin signaling, exacerbating angiogenesis and tumor spread—paradoxically converting sTIE2's protective role into pro-metastatic activity under PY exposure.

NCR3 (NKp30) and TNKS2 emerge as high-affinity targets for PYs I/II. NCR3 modulates NK cell activity via splice variants: activating isoforms (NKp30a/b) enhance tumor lysis, while inhibitory NKp30c suppresses immunity [18]. Our MR analysis identified NCR3 as protective in ER-negative BC, suggesting its role in NK-mediated tumor suppression. PY binding may favor inhibitory isoform dominance, potentially subverting immune surveillance. TNKS2, upregulated in TNBC, destabilizes β-catenin degradation by degrading AXIN, boosting β-catenin stability and oncogenic Wnt signaling linked to tumor aggression [19]. PYs show high-affinity TNKS2 binding, potentially exacerbating Wnt pathway activation to promote BC progression.

EZH2, an epigenetic driver in aggressive BC subtypes (HER2+, basal-like, TNBC), correlates with tumor invasiveness and poor prognosis. As PRC2's catalytic subunit, it silences tumor suppressors via H3K27me3-mediated repression and drives therapy resistance through ER-Wnt crosstalk [20]. MR analyses identify EZH2 as a genetic risk factor in ER+ BC. TOP2A, upregulated in BC (see Figure S1a), encodes a chemotherapy target (e.g., doxorubicin), yet its overexpression induces anthracycline resistance via target saturation [21]. PY I may block TOP2A's catalytic sites, worsening drug resistance. SOX4 activates CSNK2A1 to enhance TOP2A-driven BC progression, with PY II potentially amplifying oncogenic effects through this axis [22].

RPS6KB1 encodes S6K1, a key PI3K/AKT/mTOR pathway effector hyperactivated in BC via PI3K/PTEN mutations. S6K1 drives tumor progression by (1) enhancing pro-proliferative mRNA translation, (2) suppressing apoptosis, and (3) phosphorylating ERα to sustain endocrine therapy resistance under low-estrogen conditions [23]. MT-CO1, encoding mitochondrial Complex IV's catalytic subunit, shows recurrent mutations in BC linked to electron transport chain dysfunction. These mutations reduce ATP synthesis while elevating ROS production, creating a pro-metastatic microenvironment through redox signaling [24]. Although SCN7A's role remains unclear, its paralog SCN5A correlates with poor prognosis in ER-negative BC [25]. Our findings suggest PY I promotes BC pathogenesis via interactions with these targets through the described mechanisms.

In TNBC, CAMK2G activation via PEAK1 drives malignancy through reciprocal phosphorylation, reinforcing PEAK1/PEAK2 heterodimerization and STAT3 activation [26]. SOS1, a RAS-activating GEF, triggers MEK/ERK and PI3K/AKT pathways to promote BC proliferation and metastasis [27]. PY II binds both CAMK2G and SOS1, potentially activating these oncogenic pathways. PPARG exhibits dual roles: downregulation links to tumor progression, while ER-positive overexpression correlates with better outcomes [28]. PY II may disrupt PPARG's tumor-suppressive transcriptional regulation, exacerbating carcinogenesis, consistent with our differential gene analysis (Figure S1a).

While molecular docking quantifies binding affinity, it does not inherently specify functional consequences (e.g., activation vs. inhibition). Thus, docking energies < -7.5 kcal/mol indicate stable binding but do not demonstrate loss-of-function toxicity. Interpretation of binding data in the context of MR results requires caution: high-affinity binding to MR-identified protective targets (e.g., NCR3, TEK) could disrupt physiological function, but alternative outcomes (e.g., allosteric modulation, functional potentiation) remain plausible without experimental validation. While our multi-omics approach provides mechanistic hypotheses, future work should prioritize: (i) organoid models to test PY-induced TOP2A chemoresistance, (ii) xenograft studies evaluating PY-driven metastasis via TEK/RPS6KB1, and (iii) population biomonitoring correlating PY metabolites with BC incidence.

5 CONCLUSION

This multi-omics computational study suggests that PYs exhibit high-affinity binding to key oncogenic targets (MAOB, TEK, NCR3, TNKS2, EZH2) and may disrupt critical pathways implicated in breast carcinogenesis (e.g., angiogenesis, immune evasion, and Wnt/β-catenin signaling). Our integrative *in silico* evidence positions PYs as a potential risk factor for BC, particularly in ER-negative and triple-negative subtypes. However, these predictions require rigorous experimental validation in biological systems and epidemiological confirmation of exposure-disease associations. Regulatory agencies should consider prioritizing longitudinal biomonitoring studies to assess PY bioaccumulation and its population-level health impacts. This work establishes a framework for preemptive environmental toxicant risk assessment by multi-omics big data, emphasizing chemical carcinogenesis through multi-target network disruption. Future work must validate predicted targets in vivo and elucidate dose-response relationships to conclusively evaluate PYs' carcinogenic hazard.

6 ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (Approval number: 82074508), the Natural Science Foundation of Fujian Province (Approval number: 2023J011627), the Xiamen Key Scientific Research Project in TCM (Approval number: XWZY-2023-0603), the Xiamen Municipal Major Scientific Research Funding Program for High-Quality Development in Health and Science Technology (Approval number: 2024GZL-ZD13), and the Science and Technology Project of the Fujian Provincial Health Commission (Approval number: 2023CXB003).

7 DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in the published article and its supplementary materials (https://tinyurl.com/4yjjeayh).

8 REFERENCES

[1] K. Matsuda, "Understanding pyrethrin biosynthesis: Toward and beyond natural pesticide overproduction," *Biochemical Society Transactions*, vol. 52, no. 4, pp. 1927–1937, 2024. https://doi.org/10.1042/BST20240213

[2] M. M. Colodel, I. Ferreira, V. M. V. Martins, A. A. Almeida, M. D. Lopes, and N. S. Rocha, "Spontaneous mammary carcinomas in female dogs: Association between immunohistochemical degrees of neoplasia aggressiveness and residual pyrethroids," *Open Journal of Veterinary Medicine*, vol. 2, no. 4, pp. 207–215, 2012. https://doi.org/10.4236/ojvm.2012.24033

[3] M. Leemans, S. Couderq, B. Demeneix, and J.-B. Fini, "Pesticides with potential thyroid hormone-disrupting effects: A review of recent data," *Frontiers in Endocrinology*, vol. 10, p. 743, 2019. https://doi.org/10.3389/fendo.2019.00743

[4] E. E. Akselrad, M. de la Cabeza Fernández, P. Moyano, and M. V. Naval, "Unveiling the connections: Chlorpyrifos and its association with breast cancer," *Journal of Clinical Microbiology and Biochemical Technology*, vol. 9, no. 1, pp. 022–029, 2023. https://doi.org/10.17352/jcmbt.000055

[5] I. Young Kim *et al.*, “Pyrethroid insecticides, fenvalerate and permethrin, inhibit progesterone-induced alkaline phosphatase activity in T47D human breast cancer cells,” *Journal of Toxicology and Environmental Health*, vol. 68, nos. 23–24, pp. 2175–2186, 2005. https://doi.org/10.1080/15287390500180523

[6] K. Huang, T. Fu, L. M. Glass, M. Zitnik, C. Xiao, and J. Sun, “DeepPurpose: A deep learning library for drug–target interaction prediction,” *Bioinformatics*, vol. 36, nos. 22–23, pp. 5545–5547, 2020. https://doi.org/10.1093/bioinformatics/btaa1005

[7] M. I. Kurki *et al.*, “FinnGen provides genetic insights from a well-phenotyped isolated population,” *Nature*, vol. 613, no. 7944, pp. 508–518, 2023.

[8] L. Du *et al.*, “Dockey: A modern integrated tool for large-scale molecular docking and virtual screening,” *Briefings in Bioinformatics*, vol. 24, no. 2, 2023. https://doi.org/10.1093/bib/bbad047

[9] T. Lu, “A comprehensive electron wavefunction analysis toolbox for chemists, Multiwfn,” *The Journal of Chemical Physics*, vol. 161, no. 8, 2024. https://doi.org/10.1063/5.0216272

[10] F. Neese, “Software update: The ORCA program system—Version 5.0,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 12, no. 5, p. e1606, 2022. https://doi.org/10.1002/wcms.1606

[11] T. Lu, “Sobtop: A tool of generating forcefield parameters and GROMACS topology file,” Sobtop, vol. 1, 2022. http://sobereva.com/soft/Sobtop

[12] N. Homeyer and H. Gohlke, “Free energy calculations by the molecular mechanics Poisson–Boltzmann surface area method,” *Molecular Informatics*, vol. 31, no. 2, pp. 114–122, 2012. https://doi.org/10.1002/minf.201100135

[13] S. Maier, B. Thapa, J. Erickson, and K. Raghavachari, “Comparative assessment of QM-based and MM-based models for prediction of protein–ligand binding affinity trends,” *Physical Chemistry Chemical Physics*, vol. 24, no. 23, pp. 14525–14537, 2022. https://doi.org/10.1039/D2CP00464J

[14] M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente, and E. Moreno, “gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS,” *Journal of Chemical Theory and Computation*, vol. 17, no. 10, pp. 6281–6291, 2021. https://doi.org/10.1021/acs.jctc.1c00645

[15] Y. J. Cha, W. H. Jung, and J. S. Koo, “Site-specific expression of amine oxidases in breast cancer metastases,” *Tumor Biology*, vol. 40, no. 5, 2018. https://doi.org/10.1177/1010428318776822

[16] M. L. Slattery *et al.*, “Angiogenesis genes, dietary oxidative balance and breast cancer risk and progression: The breast cancer health disparities study,” *International Journal of Cancer*, vol. 134, no. 3, pp. 629–644, 2014. https://doi.org/10.1002/ijc.28377

[17] I. Makhoul *et al.*, “High-circulating Tie2 is associated with pathologic complete response to chemotherapy and antiangiogenic therapy in breast cancer,” *American Journal of Clinical Oncology*, vol. 39, no. 3, pp. 248–254, 2016. https://doi.org/10.1097/COC.0000000000000046

[18] A. Shemesh *et al.*, “NKp44 and NKp30 splice variant profiles in decidua and tumor tissues: A comparative viewpoint,” *Oncotarget*, vol. 7, no. 43, pp. 70912–70923, 2016. https://doi.org/10.18632/oncotarget.12292

[19] Z. Jia *et al.*, “miR-490-3p inhibits the growth and invasiveness in triple-negative breast cancer by repressing the expression of TNKS2,” *Gene*, vol. 593, no. 1, pp. 41–47, 2016. https://doi.org/10.1016/j.gene.2016.08.014

[20] S. Adibfar *et al.*, “The molecular mechanisms and therapeutic potential of EZH2 in breast cancer,” *Life Sciences*, vol. 286, p. 120047, 2021. https://doi.org/10.1016/j.lfs.2021.120047

[21] J.-h. Ye, J. Yu, M.-y. Huang, and Y.-m. Mo, “The correlation study between TOP2A gene expression in circulating tumor cells and chemotherapeutic drug resistance of patients with breast cancer,” *Breast Cancer*, vol. 31, no. 3, pp. 417–425, 2024. https://doi.org/10.1007/s12282-024-01553-x

[22] J. Zou *et al.*, "Novel role of the SOX4/CSNK2A1 axis in regulating TOP2A phosphorylation in breast cancer progression," *The FASEB Journal*, vol. 39, no. 3, p. e70315, 2025. https://doi.org/10.1096/fj.202401907RR

[23] J. Sridhar, R. Komati, and S. Kumar, "Targeting RPS6K1 for refractory breast cancer therapy," in *Breast Cancer*. Brisbane, Australia: Exon Publications, 2022, pp. 163–178. https://doi.org/10.36255/exon-publications-breast-cancer-rps6k1

[24] R. C. de Oliveira, S. P. Dos Reis, and G. C. Cavalcante, "Mutations in structural genes of the mitochondrial complex IV may influence breast cancer," *Genes*, vol. 14, no. 7, p. 1465, 2023. https://doi.org/10.3390/genes14071465

[25] M. A. Erdogan, E. Yuca, A. Ashour, N. Gurbuz, S. Sencan, and B. Ozpolat, "SCN5A promotes the growth and lung metastasis of triple-negative breast cancer through EF2-kinase signaling," *Life Sciences*, vol. 313, p. 121282, 2023. https://doi.org/10.1016/j.lfs.2022.121282

[26] X. Yang *et al.*, "Activation of CAMK2 by pseudokinase PEAK1 represents a targetable pathway in triple negative breast cancer," *Nature Communications*, vol. 16, no. 1, 2025. https://doi.org/10.1038/s41467-025-57046-8

[27] S.-G. Du *et al.*, "Polyphyllin VII promotes apoptosis in breast cancer by inhibiting MAPK/ERK signaling pathway through downregulation of SOS1," *The American Journal of Chinese Medicine*, vol. 52, no. 3, pp. 885–904, 2024. https://doi.org/10.1142/S0192415X24500368

[28] D.-H. Li, X.-K. Liu, X.-T. Tian, F. Liu, X.-J. Yao, and J.-F. Dong, "PPARG: A promising therapeutic target in breast cancer and regulation by natural drugs," *PPAR Research*, vol. 2023, no. 1, p. 4481354, 2023. https://doi.org/10.1155/2023/4481354

9 AUTHORS

Zikang Jiang is currently a Master's candidate at Fujian University of Traditional Chinese Medicine. He is undergoing resident physician standardized training at Xiamen Hospital and Dongzhimen Hospital, Beijing University of Chinese Medicine. His research focus is oncology.

Jinghui Sung is currently a Master's candidate at Fujian University of Traditional Chinese Medicine. She is undergoing resident physician standardized training at Xiamen Hospital and Dongzhimen Hospital, Beijing University of Chinese Medicine. Her research focus is oncology.

Weijie Li is currently a Master's candidate at Fujian University of Traditional Chinese Medicine. He is undergoing resident physician standardized training at Xiamen Hospital and Dongzhimen Hospital, Beijing University of Chinese Medicine. His research focus is oncology.

Yixin Zhuang is currently a student at Xiamen University. Her research focus is oncology.

Yuanpeng Huang, Ph.D. is a Professor and Chief Physician at Beijing University of Chinese Medicine and Fujian University of Traditional Chinese Medicine and serves as Postgraduate Supervisor. He has been engaged in scientific research, clinical practice, and teaching on tumor disease prevention and treatment for over 30 years (E-mail: huangyp998@xmu.edu.cn).

Ting-Yu Chen, Ph.D. is Assistant Professor at the Integrated Research Center for Green Technologies, National Chin-Yi University of Technology. He has long been engaged in research on environmental pollutants.