Sensors and Materials, Vol. 37, No. 6 (2025) 2679–2695
MYU Tokyo

S & M 4081

Predicting Herb Pairs against Triple-negative Breast Cancer by Integrating Graph Neural Network and Multiscale Interactome

Sung-Sam Hong,[1] Sangjin Kim,[2] Yewon Han,[2] Boyun Jang,[3] Youngsoo Kim,[3] Chan Lim Park,[4] Seungho Lee,[5] Sungyoul Choi,[6*] and Won-Yung Lee[7,8**]

[1]Department of Multimedia Contents, Jangan University, Pyeongtaek 17731, Republic of Korea
[2]National Institute for Korean Medicine Development, Seoul 4516, Republic of Korea
[3]IntegroMediLab Co., Ltd., 428-Bonguan, Chungmuro Building of Dongguk University, Republic of Korea
[4]Smart Safety Laboratory, Seoul 05567, Republic of Korea
[5]College of Korean Medicine, Woosuk University, Jeonju 54986, Republic of Korea
[6]College of Korean Medicine, Gachon University, Seungnam 13120, Republic of Korea
[7]School of Korean Medicine, Wonkwang University, Iksan 54538, Republic of Korea
[8]Research Center of Traditional Korean Medicine, Wonkwang University, Iksan 54538, Republic of Korea

(Received April 4, 2025; accepted June 13, 2025)

Keywords: triple-negative breast cancer, graph neural network, GraphSAGE, multiscale interactome, network pharmacology, synergistic herbal therapies

Triple-negative breast cancer (TNBC) is an aggressive subtype lacking targeted therapies, leading to high relapse rates and poor prognoses. Here, we aimed to identify a synergistic herb pair against TNBC by integrating graph neural networks (GNNs) and a multiscale interactome. We curated TNBC-related biomarkers and constructed an herb–compound–target–disease network by integrating multiple data sources. Using this dataset, we trained and evaluated three GNN architectures—graph convolutional network (GCN), graph attention network, and graph sample-and-aggregate (GraphSAGE)—on 6830 herb pairs annotated with compound and target information. We then applied a biased random walk algorithm to estimate the network effect of herb targets and TNBC-related proteins, identifying new herbal candidates with potential synergy. Among the tested GNNs, GraphSAGE showed the highest performance in distinguishing known versus unknown herb pairs, with significant accuracy gains ($p < 0.001$). We subsequently performed diffusion profile analysis on top-ranked herbal combinations, revealing key TNBC targets, such as AKT1 and TP53. This multiscale approach illuminated potential synergistic effects within herbal therapies for TNBC. Our findings demonstrate that integrating GNN-driven deep learning with network pharmacology can systematically uncover multi-target herbal therapies for TNBC. Moreover, the molecular network we present can guide the design of materials for the rapid screening of herb–target interactions, aligning this work with emerging sensing technologies.

*Corresponding author: e-mail: pc1075@gachon.ac.kr
*Corresponding author: e-mail: wonyung21@wku.ac.kr
https://doi.org/10.18494/SAM5650

ISSN 0914-4935
https://myukk.org/

1. Introduction

Triple-negative breast cancer (TNBC) represents an aggressive and heterogeneous subtype of breast cancer defined by the lack of estrogen receptors (ERs), progesterone receptors (PRs), or human epidermal growth factor receptor 2 (HER2).[1,2] Accounting for approximately 10–20% of all breast cancer cases, TNBC is notorious for its rapid progression, high metastatic potential, and early relapse following conventional therapies.[3,4] The absence of specific molecular targets renders conventional hormone therapies and HER2-targeted treatments ineffective, thereby relegating TNBC treatment predominantly to surgery, radiotherapy, and chemotherapy.[1,3] Unfortunately, these standard treatment modalities often fall short in preventing recurrence, and patients frequently experience poor outcomes, underscoring an urgent need for novel therapeutic strategies.

For centuries, traditional herbal medicine has been employed in East Asia as a holistic approach to treating a wide spectrum of diseases, including cancer. Herbal formulations typically consist of multiple bioactive compounds. Each compound contributes to a synergistic effect that modulates various biological processes such as immune regulation, apoptosis induction, and oxidative stress reduction.[5] Emerging evidence suggests that a certain herb pair, namely, ginseng radix–astragali radix, exhibits promising anticancer activities and anti-cancer-related fatigue, including inhibition of tumor cell proliferation and induction of cell death.[6,7] This multitarget potential makes herbal medicine a compelling candidate for complementing conventional cancer therapies, particularly in diseases such as TNBC. Furthermore, combining traditional herbal approaches with modern therapeutic strategies may not only enhance antitumor effects but also mitigate the side effects commonly associated with aggressive chemotherapy regimens.[8,9]

The advent of network pharmacology has advanced the way researchers approach multicomponent therapies. Unlike the traditional "one drug–one target" paradigm, network pharmacology embraces a "multicomponent, multitarget" framework that is especially well suited to analyzing the complex interactions inherent in herbal formulations.[10,11] By constructing integrative networks that map relationships among herbs, their constituent compounds, protein targets, and disease phenotypes, researchers can uncover synergistic interactions and novel therapeutic mechanisms that would otherwise remain obscured using conventional methods.[12,13] Several studies have leveraged this approach to reveal how herbal compounds modulate key signaling pathways in cancer, providing a robust rationale for the clinical application of herbal combinations as adjuvant treatments.[14,15] However, many of these studies have been limited by static network models and relatively simple topological analyses, which may not fully capture the dynamic and multidimensional nature of biological systems.

To elucidate the complex characteristics of traditional herbal medicine, advanced approaches have recently been introduced, such as graph-based machine learning and multiscale interactome approaches. These techniques provide powerful tools to dissect how the myriad bioactive compounds in herbal formulas interact within the biological system. For instance, deep learning models such as graph neural networks (GNNs) are specifically designed to capture intricate relationships in graph-structured data,[16] enabling their application in drug repurposing and

drug–drug interaction analyses.[17,18] On the other hand, the multiscale interactome approach goes beyond merely mapping protein–protein interactions to also incorporate the relationships between proteins and biological functions.[19] By simulating how drugs and diseases affect these layered networks, we can realize this strategy, which can offer more precise insights into therapeutic efficacy and underlying mechanisms. Notably, some studies have employed this approach to accurately predict the effects of herbal formulas or to identify the key mechanisms of specific herbal medicines, such as *Bupleurum* radix.[20, 21]

In this study, we aim to predict promising herb combinations for TNBC treatment by combining a graph convolutional network (GCN) and multiscale interactome analysis (Fig. 1). First, we trained and compared various GCN models to predict optimal herb combinations based on an integrated network. These models will serve as the primary tool to identify synergistic herb pairs. From the predicted combinations, a TNBC-related herb combination was selected on the basis of its simulated impact on a multiscale interactome, ensuring its potential therapeutic effects on TNBC-associated targets. Subsequently, the key mechanisms of the selected herb combination were visualized to elucidate their underlying pathways and interactions. Together, these integrated methodologies not only offer a novel strategy for predicting effective herbal treatments for TNBC but also lay a foundation for subsequent experimental validation and potential clinical applications.

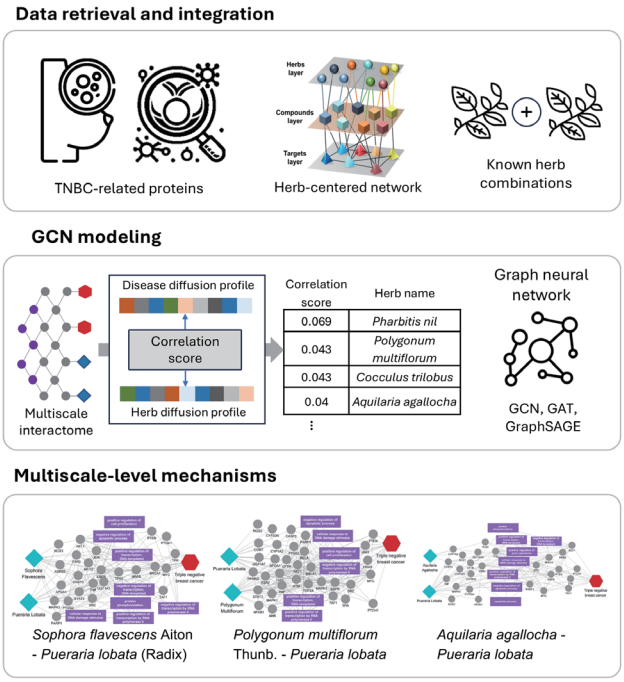

Fig. 1. (Color online) Integrated strategy for identifying herb combinations in TNBC treatment. The top panel integrates TNBC-related proteins, herbs, compounds, and targets, as well as previously known herbal prescriptions. In the middle panel, a correlation score is computed by comparing disease diffusion and herb diffusion profiles within the integrated network. The bottom panel illustrates an example of a network showing how the selected herb combination interacts with TNBC-related proteins.

2. Materials and Methods

2.1 Rule mining associations

Association rule mining was applied to one of the network pharmacology databases, TCMID,[22] to extract herb combination associations from prescription–herb relationships. In this process, it was determined that using only co-occurrence as a criterion will predominantly select frequently appearing herbs (e.g., licorice), whereas relying solely on the lift value can incorporate herb combinations with low co-occurrence. To ensure both reliability and an adequate dataset size, herb pair associations were selected only if they satisfied both criteria. For the co-occurrence-based analysis, the co-occurrence threshold was varied, and the corresponding lift values of known herb pairs were monitored. It was observed that increasing the co-occurrence threshold gradually reduced the total number of herb pairs, whereas the discrimination performance of the herb combinations improved up to a point before eventually declining. A frequency threshold of 5 was ultimately chosen. Subsequently, using the established frequency threshold, we conducted a lift-based analysis by incrementally adjusting the lift score threshold while assessing the predictive performance (measured by MCC) and the distribution of available sample sizes. With an increasing lift threshold, the number of herb pairs continuously decreased, and the discrimination performance initially improved before decreasing again. A lift threshold of 1.6 was selected as optimal, as it yielded the best performance while maintaining an adequate number of samples.

2.2 Herb–compound–target–disease network

Herb–compound interaction data were compiled from multiple herbal medicine databases, including HerDing,[23] TCM-Taiwan,[24] TCMID,[22] TCMSP,[25] TM-MC,[26] and UNPD.[27] From each database, curated lists of herbs and their reported bioactive compounds were retrieved, focusing on those validated on the basis of experimental evidence or widely cited literature. Next, compound names, synonyms, and unique identifiers (e.g., PubChem CID) were harmonized through an ID mapping process to unify compound records across databases. Following standardization and the removal of duplicate entries, a comprehensive herb–compound dataset was generated, forming the basis for subsequent network construction and analyses.

Compound–target interactions were obtained from the STITCH database (version 5.0). In this study, a compound was considered to interact with a target if its combined score exceeded the default threshold of 700, ensuring a high confidence level in the interaction data. In this network analysis, the simple pathway count for each herb was calculated, accounting for instances where multiple components affected a single target. This process enabled for the selection of the top 100 targets, with each target's relative importance assessed accordingly. Disease-related proteins were derived from a curated dataset that included 150 TNBC-associated functional genes manually compiled from over 30 research articles and reviews.[28] This set includes targets from clinical trials, significantly mutated driver genes from large-scale cancer genome studies, and

genes validated through siRNA or gene knockdown experiments. The integration of these datasets enabled the construction of a comprehensive herb–compound–target–disease network that serves as the foundational framework for subsequent analyses.

2.3 Multiscale interactome

The associations among protein–protein, protein–biological function, and biological function–function interactions were retrieved from the methodology described by Ruiz *et al.*[19] Human protein–protein interaction data were obtained from various databases, encompassing 387626 physical interactions among 17660 proteins. Protein–biological function interactions were extracted from the human gene ontology database, assembling 34777 experimentally verified associations between 7993 proteins and 6387 biological functions. Lastly, biological function–function interactions were organized into a highly interconnected hierarchical structure with 22545 associations among 9798 functions.

2.4 Network construction

An integrated network was constructed by mapping each node type—herbs, compounds, targets, and biological functions—to consistent IDs, ensuring alignment across diverse data sources. As a result, 514 herbs and 6351 compounds were linked through 32918 herb–compound relationships, whereas 23684 compound–target interactions were identified, involving the same 6351 compounds and 4062 targets. At the protein level, 17381 proteins were connected via 385659 protein–protein interactions, and 7993 proteins and 6387 biological functions were associated through 34777 experimentally verified links. Lastly, 22545 function–function relationships were captured among 9798 biological functions, forming a highly interconnected functional hierarchy. All these data were consolidated into a unified framework that facilitates subsequent diffusion-based and mechanism-focused analyses.

2.5 Gene set enrichment analysis (GSEA)

Biological processes and signaling pathways associated with the protein targets were identified by GSEA with the GSEApy module in a Python environment, facilitated through the Enrichr platform (http://amp.pharm.mssm.edu/Enrichr/).[29,30] Enrichr performs enrichment analysis by drawing on various gene-set libraries, such as Gene Ontology and the Kyoto Encyclopedia of Genes and Genomes (KEGG). In this study, adjusted p-values, z-scores, and combined scores were calculated to evaluate the signaling pathways and biological functions relevant to herbal ingredient targets. The combined score, obtained by multiplying the logarithm of the p-value with the z-score, provided reliable results, allowing for a systematic evaluation of the effects of herbal components on specific biological pathways. All signaling pathways identified through enrichment analysis were included in the analysis, except for those specifically related to diseases.

2.6 Biased random walker algorithm

Diffusion profiles were subsequently calculated using the multiscale interactome to assess propagation effects between herb targets and proteins associated with disease. A biased random walk with a restart algorithm was employed for this analysis, enabling a quantitative evaluation of the effect exerted by herbal and ingredient targets on disease-related proteins. A correlation score was then computed between herb–ingredient and disease profiles to facilitate the identification of potential candidate herbs and ingredients for treating the disease.

The key mechanisms underlying each ingredient–disease pair were identified by analyzing the diffusion profiles and selecting the top k proteins or biological functions based on their effect from either the herb or the disease. A network was constructed from these selected entities to highlight their significance, whereas targets of ingredients not associated with disease-related proteins or biological functions were excluded. The highest-ranking entity in the diffusion profile was deemed most critical for treatment owing to its substantial effect. In this analysis, the value of k was set to 20 to ensure sufficient exploration of influential nodes. Previous studies have shown that a window size of $k \approx 20$ effectively recovers most disease-relevant nodes.[21,31] For a detailed explanation of the diffusion profile calculation, including mathematical formulas, iterative procedures, and the rationale behind selecting parameter k, please refer to the previous studies.

2.7 GCNs

GCNs, graph attention networks (GATs), and graph sample-and-aggregate (GraphSAGE) were considered as the primary GNN models. Briefly, GCN employs a first-order approximation of spectral graph convolutions to efficiently perform semisupervised learning on graph-structured data, achieving superior accuracy and computational efficiency on citation networks (Cora, Citeseer, Pubmed) and knowledge graphs (NELL).[16] GAT introduces a self-attention mechanism that enables each node to learn dynamic importance weights for its neighbors, improving representational power in both transductive and inductive learning settings.[32] This approach refines node embeddings by emphasizing more relevant neighbors and down-weighting less informative ones. Lastly, GraphSAGE[33] learns an aggregation function to sample and aggregate neighbor information, allowing real-time embedding generation for new or previously unseen nodes. This inductive framework is particularly effective for large-scale or rapidly evolving graphs, such as social networks or biological networks (e.g., protein–protein interactions). Collectively, GCNs, GATs, and GraphSAGE have demonstrated robust performance in node classification, recommendation systems, social network analysis, and various other domains.

2.8 Model training and hyperparameter settings

The GCN models were trained using the following hyperparameters: the hidden layer dimension was set to 128, the batch size to 1024, and the learning rate to 0.001. The Adam

optimizer was employed for model optimization. Training was conducted for 100 epochs, a threshold determined empirically as the point at which the loss function stabilized, ensuring convergence. The dataset was partitioned into training and test sets at a 5:1 ratio to comprehensively evaluate the models' generalization performance.

2.9 Performance evaluation

A total of 6830 herb pairs were used for evaluation, comprising 3415 known herb pairs as positive examples and an equal number of randomly selected unknown herb pairs as negative examples. The model was trained on the training set and subsequently evaluated on the test set. Performance was assessed using the following metrics: accuracy, F1 score, area under the receiver operating characteristic curve (AUC), and area under the precision-recall curve (AUPR). This evaluation framework enabled a comprehensive assessment of the model's ability to differentiate between effective (known) and ineffective (unknown) herb pair combinations. To confirm the statistical significance of differences in model performance, one-way ANOVA was performed on the final metric scores, followed by a post hoc test (e.g., Tukey's HSD) to pinpoint pairwise differences among the evaluated methods.

3. Results

3.1 Performances of GCN-based model for herb combination prediction

We evaluated the predictive performance of three GNN architectures (GCN, GAT, and GraphSAGE) for herb combination prediction. To systematically assess the models, we grouped our metrics into three categories: training-oriented (Train Loss, Test Loss), probability-based (AUC, AUPR), and threshold-based (accuracy, F1 score). Each model was trained under identical conditions, and the results were aggregated over multiple runs to ensure reproducibility. Overall, GraphSAGE demonstrated superior performance across all evaluated metrics (Fig. 2). Specifically, it achieved the lowest train loss and test loss ($p < 0.001$), indicating more stable and generalized learning. Moreover, GraphSAGE achieved the highest mean AUC (0.884) as well as superior accuracy (0.884), F1 score (0.887), and AUPR (0.840), significantly outperforming GCN (0.848, 0.848, 0.853, 0.798) and GAT (0.849, 0.849, 0.855, 0.798) ($p < 0.001$). These findings underscore the advantages of an inductive learning approach for large and evolving networks and highlight GraphSAGE's potential for identifying synergistic herb combinations within complex multicomponent systems.

3.2 Discovering herbal candidate for TNBC via a multiscale interactome

We then applied a multiscale interactome approach to identifying herbal candidates for TNBC by quantifying the impact of herb–disease associations through a diffusion profile derived from a biased random walker algorithm. In this framework, each herb's capability to affect the TNBC-related proteins was measured and subsequently converted into a correlation

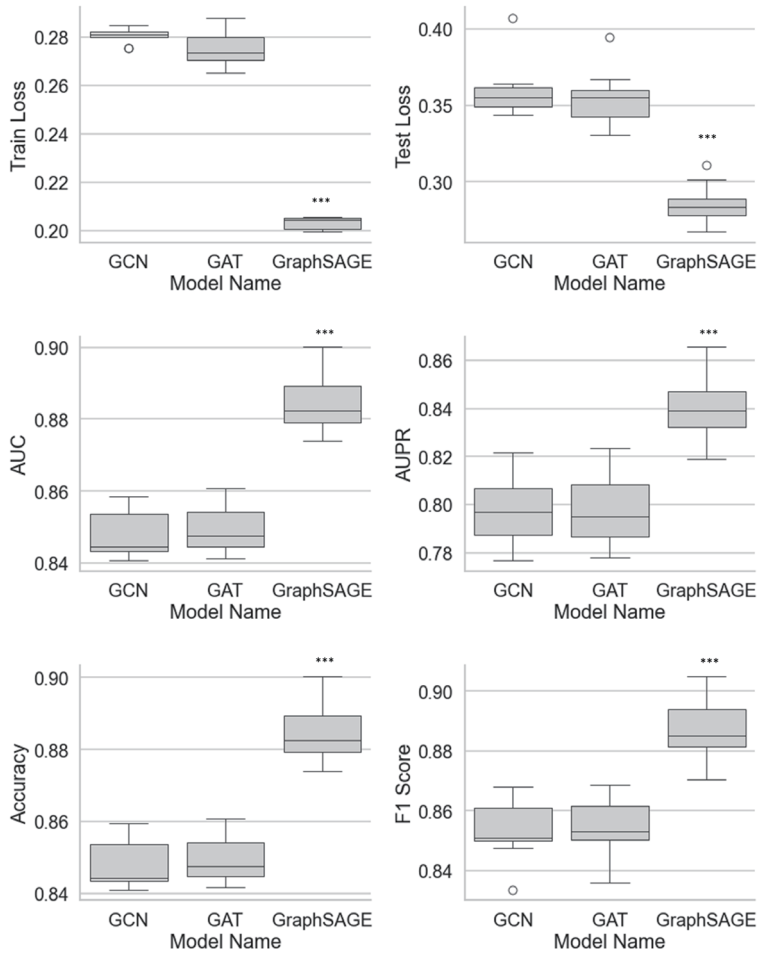

Fig. 2. Performance comparison of GCN, GAT, and GraphSAGE in herb combination prediction. Six boxplots illustrate the distributions of train loss, test loss, accuracy, F1 score, AUC, and AUPR for each model (GCN, GAT, and GraphSAGE). Boxes represent the interquartile range, with whiskers extending to the most extreme data points. Dots above or below the whiskers denote outliers. The asterisks (***) indicate statistically significant differences ($p < 0.001$) according to one-way ANOVA with post hoc tests.

score between the herb and TNBC. Our analysis revealed that ten herbal candidates exhibited relatively high correlation scores with TNBC (Table 1). Among these ten herbs, three—*Pharbitis nil, Sophora flavescens,* and *Sanguisorba officinalis*—have already been experimentally validated to exert therapeutic effects against TNBC.[34–36] This convergence between our computational predictions and the existing experimental evidence reinforces the reliability of our methodology. Moreover, the remaining herbal candidates, including *Polygonum multiflorum*, *Cocculus trilobus*, *Aquilaria agallocha*, *Scapharca broughtonii*, *Dalbergia odorifera*, and both the flos and radix of *Pueraria lobata*, which have not yet been directly associated with TNBC in the literature, emerge as promising novel candidates for use in TNBC treatment and require further investigation.

To investigate the potential mechanisms, we conducted a KEGG signaling pathway enrichment analysis on their predicted protein targets (Table 2). The result showed significant

Table 1
Herbal candidates identified for TNBC using multiscale interactome analysis.

Scientific name	Official name	Correlation score	Reported evidence (PMID)
Pharbitis nil	Pharbitis nil (L.) Choisy (Syn. Ipomoea nil)	0.069	34746014
Polygonum multiflorum	Polygonum multiflorum Thunb. (Syn. Fallopia multiflora, Reynoutria multiflora)	0.043	
Cocculus trilobus	Cocculus trilobus (Thunb.) DC.	0.043	
Aquilaria agallocha	Aquilaria malaccensis Lam.	0.04	
Sophora flavescens	Sophora flavescens Aiton	0.039	32810619
Scapharca broughtonii	Scapharca broughtonii	0.039	
Dalbergia odorifera	Dalbergia odorifera T. Chen	0.038	
Sanguisorba officinalis	Sanguisorba officinalis L.	0.038	33381039
Pueraria lobata	Pueraria lobata (Willd.) Ohwi (Flos)	0.038	
Pueraria lobata	Pueraria lobata (Willd.) Ohwi (Radix)	0.036	

Table 2
Herbal candidates identified for TNBC using multiscale interactome analysis.

Term (Pathway)	Overlap	Adjusted *p*-value	Odds ratio	Combined score
Bladder cancer	18/41	6.74E−26	100.65	6105.20
Prostate cancer	28/97	8.38E−34	55.68	4459.31
AGE-RAGE signaling in diabetic complications	28/100	1.79E−33	53.35	4220.76
Lipid and atherosclerosis	43/215	7.73E−45	38.09	4053.52
Pancreatic cancer	23/76	1.66E−28	57.59	3859.66
General pathways in cancer	68/531	1.63E−59	27.35	3853.21
Non-small cell lung cancer	21/72	8.93E−26	53.93	3253.10
Hepatitis B	33/162	1.98E−34	36.25	2963.79
Chemical carcinogenesis	41/239	7.14E−40	31.03	2934.50
HIF-1 signaling	25/109	1.12E−27	39.97	2599.23

enrichment in several cancer-related pathways, including bladder, prostate, pancreatic, and non-small cell lung cancer pathways, as well as more general pathways in cancer and chemical carcinogenesis. Additionally, pathways such as AGE-RAGE signaling in diabetic complications, lipid and atherosclerosis, and HIF-1 signaling suggest potential links between metabolic or inflammatory regulation and tumorigenesis.

Further enrichment analysis on gene ontology showed that the targets of the herbal candidates are significantly associated with various biological processes, cellular components, and molecular functions (Fig. 3). Notably, there was a strong enrichment in pathways regulating cell proliferation and apoptosis—such as ‘regulation of apoptotic process’ and ‘regulation of cell population proliferation’—indicating potential roles in modulating tumor growth and survival. From a cellular component perspective, the targets were prominently linked to intracellular membrane-bounded organelles, nuclei, and adhesion-related structures, suggesting their involvement in key subcellular signaling and structural processes. In terms of molecular

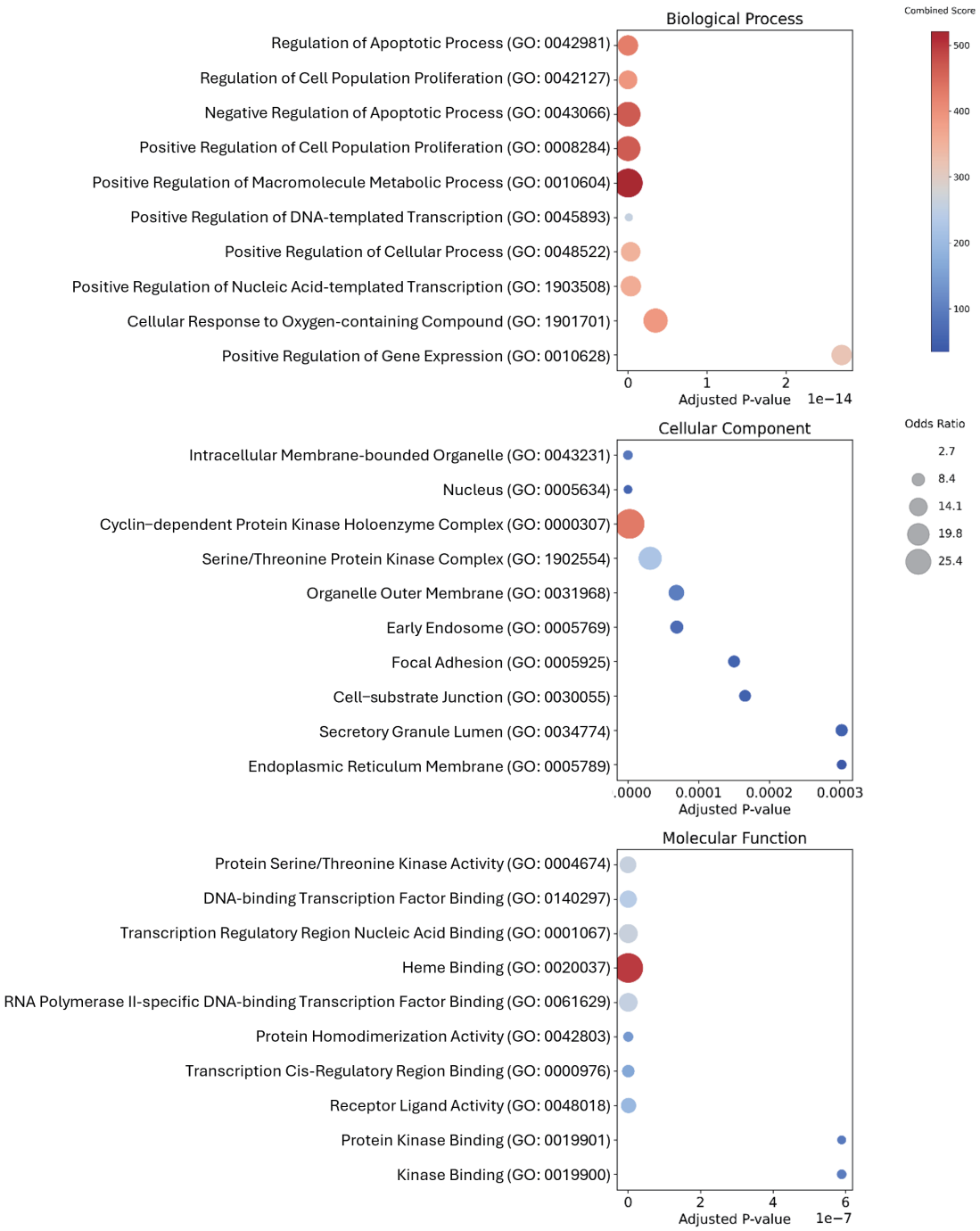

Fig. 3. (Color online) Gene ontology enrichment analysis for core protein targets. Gene ontology enrichment analysis of the 49 core protein targets across three categories: biological processes (top), cellular components (middle), and molecular functions (bottom). The *x*-axis represents the adjusted *p*-value (indicating the association significance), bubble size corresponds to the odds ratio, and bubble color reflects the combined score, which indicates the statistical significance of each term.

function, the enrichment of kinase activity, DNA-binding transcription factor binding, and heme binding underline the importance of both signaling and transcriptional regulation in these predicted targets. These findings collectively suggest that the herbal candidates may exert therapeutic effects through diverse regulatory and signaling mechanisms pertinent to TNBC.

3.3 Predicting herbal combinations for TNBC using graph-based deep learning model

Having identified GraphSAGE as the most effective model for herb combination prediction, we next applied it to the TNBC-related herb candidates. We summarize the top ten predicted herb pairs exhibiting the lowest loss values (Table 3). A low loss value reflects close similarity to embeddings of previously validated herb pairs, indicating a high probability that the candidate pair will exhibit novel synergistic activity. Notably, *S. flavescens* Aiton and *P. lobata* (radix) emerged as the most promising pair (loss = 0.00039), indicating a strong potential for anti-TNBC activity. Moreover, *P. lobata* (radix) consistently appeared in multiple top-ranked pairs, underscoring its potential role as a key ingredient. Interestingly, the combination of *P. lobata* (flos) and *P. lobata* (radix) also ranked within the top 10, suggesting that different parts of the same plant may complement each other's therapeutic actions in TNBC.

3.4 Multiscale interactome-level mechanisms of top-ranked herb pairs

To further investigate the mechanisms of top-ranked herb pairs, we conducted an additional case study at the multiscale interactome level. Among these, the combination of *S. flavescens* Aiton and *P. lobata* (radix) exhibited particularly low loss values, suggesting a high potential for synergistic effects. A multiscale interactome analysis (Fig. 4) highlighted several critical oncogenic and tumor-suppressive proteins. Notably, AKT1, TP53, MYC, and EGFR emerged as key nodes affected by this combination. The network analysis also revealed the involvement in essential cellular processes such as the positive regulation of cell proliferation, the negative regulation of apoptotic processes, and protein phosphorylation.

Another high-ranking herb pair for TNBC was *P. multiflorum* Thunb. and *P. lobata*, both of which demonstrated low loss values in the GraphSAGE-based predictions (Fig. 5). A detailed multiscale interactome analysis revealed several shared and distinct targets compared with the

Table 3
Predicted herbal combinations for TNBC based on GraphSAGE.

Herbal combination (scientific name)	Loss value
Sophora flavescens—Pueraria lobata (Radix)	0.00039
Polygonum multiflorum—Pueraria lobata (Radix)	0.00101
Aquilaria agallocha—Pueraria lobata (Radix)	0.00122
Polygonum multiflorum—Sophora flavescens	0.0015
Sophora flavescens– Cocculus trilobus	0.00164
Pueraria lobata (Radix)—*Sophora flavescens*	0.00168
Sanguisorba officinalis—Pueraria lobata (Radix)	0.0018
Aquilaria agallocha—Sophora flavescens	0.00182
Pueraria lobata (Flos)—*Pueraria lobata* (Radix)	0.00203
Sophora flavescens—Polygonum multiflorum	0.00221

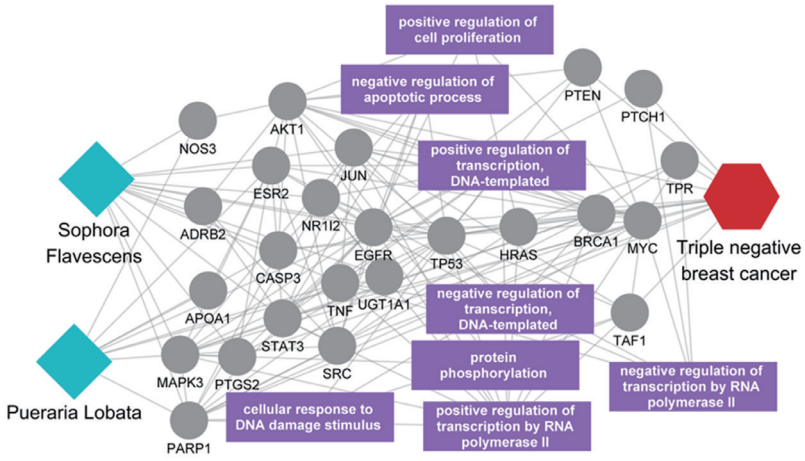

Fig. 4. (Color online) Multiscale interactome analysis of *S. flavescens* Aiton and *P. lobata* (radix) combination. Blue diamonds represent herbs, gray circles denote protein targets, purple boxes indicate biological functions, and the red hexagon denotes TNBC.

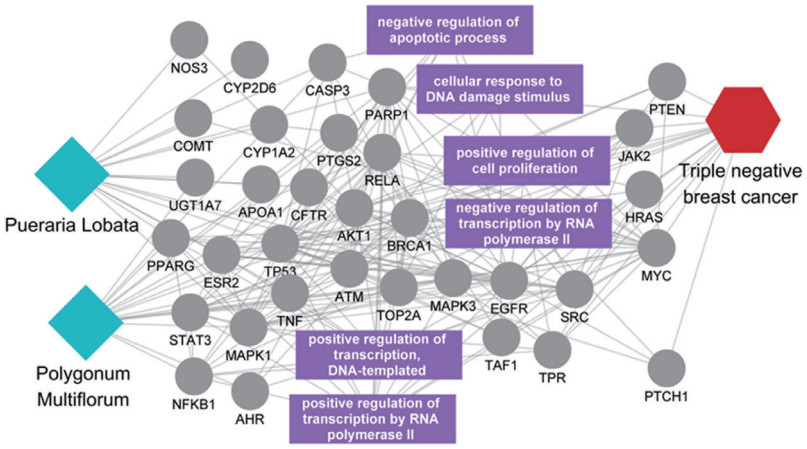

Fig. 5. (Color online) Multiscale interactome analysis for *P. multiflorum* and *P. lobata* combination. Blue diamonds represent herbs, gray circles denote protein targets, purple boxes indicate biological functions, and the red hexagon denotes TNBC.

first pair. We found that multiple oncogenic and regulatory proteins—TP53, EGFR, MYC, MAPK1, and STAT3, among others—were identified as central nodes in this combination. These targets are associated with diverse biological processes, including the negative regulation of apoptosis, the cellular response to a DNA damage stimulus, and the positive regulation of cell proliferation. Notably, the negative regulation of transcription by RNA polymerase II also emerged as a significant term, suggesting an additional layer of transcriptional control that may modulate key cancer pathways in TNBC.

A third noteworthy pair predicted by the GraphSAGE model was *A. agallocha* and *P. lobata*. We visualized their key mechanisms and found that multiple proteins involved in TNBC

progression—such as AKT1, PTEN, EGFR, and MYC—emerged as central nodes in this combination (Fig. 6). The network analysis indicated the potential modulation of pathways related to protein phosphorylation, the negative regulation of the apoptotic process, and the cellular response to the DNA damage stimulus. These functions suggest the capacity of this herb pair to affect key intracellular signaling events.

4. Discussion

TNBC represents one of the most aggressive breast cancer subtypes, characterized by a lack of hormone receptors or HER2, limited treatment options, and a poor prognosis.[37] In this context, leveraging synergistic combinations of herbal medicines offers a promising avenue, as certain multiherb formulations have shown efficacy as adjuvant therapies in TNBC.[38] In our study, we addressed this need by integrating GNNs with multiscale interactome analysis to predict and elucidate effective herb pair combinations against TNBC. First, we constructed a comprehensive herb–compound–target network by compiling known herbal compounds and their corresponding protein targets relevant to breast cancer. This involved mapping herbs to their bioactive ingredients, linking these ingredients to known or predicted target proteins, and integrating TNBC-associated genes to establish a disease-specific network. Next, we applied multiscale interactome analysis to identify promising herb candidates for TNBC. This analysis incorporated not only direct herb–target interactions but also secondary interactions among target proteins (protein–protein interactions) and their roles in broader cellular pathways. Following candidate selection, we utilized graph-based deep learning models to predict potential herb combinations. Using the trained GraphSAGE model, we systematically predicted synergistic herb pairs with high potential for TNBC therapy. Finally, we conducted in-depth case studies on the top-ranked herb pairs. For each pair, we analyzed their combined target profiles, examined the connectivity of these targets within the human interactome, and identified key proteins and biological processes affected by the herb combination. These analyses helped validate the model's predictions and provided mechanistic insights into how these herbal combinations might exert anti-TNBC effects.

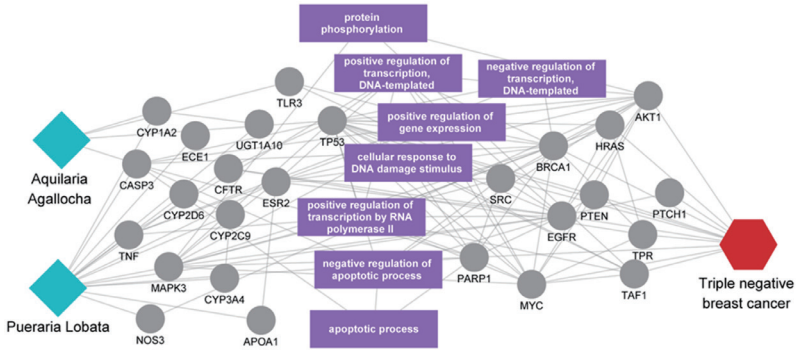

Fig. 6. (Color online) Multiscale interactome analysis for *A. agallocha* and *P. lobata* combination. Blue diamonds represent herbs, gray circles denote protein targets, purple boxes indicate biological functions, and the red hexagon denotes TNBC.

A key strength of our approach was the integration of GraphSAGE-based link prediction with multiscale interactome analysis, combining high predictive accuracy with biological interpretability. GraphSAGE outperformed other models by leveraging its inductive learning capability, allowing it to be generalized to new herbs and interactions beyond the training set.[33] This adaptability is crucial given the evolving knowledge of herbal constituents and targets. Additionally, GraphSAGE effectively captured functional patterns, clustering herbs with similar target profiles while mitigating overfitting, leading to superior predictive performance. Beyond prediction, the multiscale interactome analysis provided deeper mechanistic insights.[19] Instead of a static target-based approach, we mapped herbs to bioactive compounds, linked them to protein targets, and integrated these into the PPI network. This revealed emergent properties such as pathway crosstalk and key intermediary proteins that drive synergistic effects. For instance, one herb's targets might activate a pathway that another herb amplifies, reinforcing their combined therapeutic potential. By bridging computational predictions with biological context, we developed an approach that not only identified synergistic herb pairs with high confidence but also explained their potential mechanisms, moving beyond black-box predictions to interpretable outcomes.

From a clinical standpoint, these findings have several important implications. First, identifying synergistic herb pairs lays the groundwork for developing adjuvant therapies for TNBC. Herbal combinations predicted to synergize can be used alongside standard treatments (such as chemotherapy or targeted drugs) to potentially improve outcomes. From a clinical perspective, our findings highlight a few key points. First, identifying synergistic herb pairs provides a basis for developing herbal adjuvants in TNBC treatment. Testing these predicted combinations alongside standard therapies can be a valuable next step. Second, our network-based approach aligns with the principles of precision medicine. Since TNBC is a heterogeneous disease, the most effective herbal combination may vary depending on the patient. By integrating tumor-specific gene expression or mutation data, we can tailor herb pair recommendations to individual molecular profiles in future studies, supporting a more personalized approach to herbal therapy. Lastly, our study complements existing pharmacological research, which often focuses on single compounds. Herbal medicines act on multiple targets, and our findings embrace this complexity rather than reducing it. By identifying promising herb pairs, our work provides a foundation for further clinical investigation into their safety and efficacy as potential adjuncts to TNBC treatment.

The case studies of top-ranked herb pairs illustrated how these combinations exert their effects through key molecular targets and pathways known to drive TNBC. In analyzing the merged target networks of each herb pair, several central oncogenic or tumor-suppressive nodes emerged consistently. First, AKT1 is a pivotal node in the PI3K/AKT/mTOR signaling pathway. AKT1 activation promotes cell proliferation, survival, and metabolism in TNBC, and its dysregulation is commonly observed in this cancer.[37] In our herb pair networks, AKT1 appeared as a hub connecting multiple herb-targeted proteins, suggesting that the combination may synergistically dampen the PI3K/AKT pathway, thereby inhibiting tumor cell growth. The TP53 gene (encoding p53) is a master tumor suppressor that regulates cell cycle arrest and apoptosis in response to cellular stress. TP53 is frequently mutated or inactivated in TNBC,

contributing to uncontrolled proliferation and genomic instability.[37,39] Our analysis showed that some herb pairs can collectively affect p53 pathways – for instance, one herb in a pair provides compounds that restore p53's apoptotic function while the other herb targets proteins involved in p53 degradation. MYC is a transcription factor oncogene often upregulated in basal-like TNBC, driving robust transcriptional programs for cell growth and division.[40] The herb combinations studied were linked to the MYC regulatory network, either through the direct targeting of MYC itself or through the modulation of upstream regulators and cofactors. This result suggests the following potential key mechanisms: a synergistic herb pair that can collectively dampen MYC signaling may achieve a stronger antiproliferative effect than a single agent alone. In the case studies, we observed that one or both herbs in the pair often targeted the EGFR network (either EGFR directly or its key downstream effectors). By jointly interrupting EGFR-driven signals, the herb pair can synergistically suppress a major growth pathway in TNBC cells.

The predicted targets are closely linked to key TNBC-related processes, including uncontrolled cell proliferation (AKT1, MYC), resistance to cell death (AKT1, TP53, MYC), and dysregulated transcription of growth-related genes (MYC, TP53). The fact that our herb pairs consistently target these well-established cancer regulators supports the reliability of the GraphSAGE model and multiscale analysis. Rather than generating arbitrary predictions, the model identified herb combinations that align with known TNBC vulnerabilities, increasing confidence in their potential therapeutic relevance. These findings also provide testable hypotheses. For instance, if a herb pair is predicted to synergize by co-inhibiting the PI3K/AKT and EGFR pathways, we would expect to see reduced TNBC cell proliferation and increased apoptosis in experimental validation. The multiscale interactome analysis helps explain why these predictions make biological sense by highlighting key target hubs and pathways.

Despite encouraging results, our study is limited by incomplete pharmacology databases, potentially overlooking certain herbs or skewing findings toward well-studied candidates. Moreover, the predicted pairs require experimental validation to confirm real-world synergistic effects on TNBC, including toxicity assessments. Our approach also does not account for patient-specific factors, underscoring the need for personalized models that integrate individual omics data. Finally, extending this methodology to other cancers or diseases and refining it with advanced techniques such as the use of GNNs can further enhance its predictive power. Bridging computational insights with experimental proof remains essential for advancing these findings toward clinical use.

5. Conclusions

In this study, we combined GraphSAGE-based deep learning with a multiscale interactome to systematically uncover herb pairs that target multiple TNBC-related pathways. GraphSAGE outperformed GCN and GAT, and its top-ranked predictions—such as *S. flavescens* with *P. lobata*—converged on key oncogenic nodes (AKT1, TP53, MYC, EGFR), highlighting strong synergistic potential. These results lay a data-driven foundation for developing multi-herb adjuvants in TNBC and illustrate how network-level insights can guide biosensor materials for rapid herb–target screening. Future work should experimentally validate these combinations and extend the framework to patient-specific omics data to support precision herbal therapy.

Acknowledgments

This research was supported by grants from the National Research Foundation of Korea (NRF) funded by the Korean government (RS-2023-00243363 and RS-2023-00218419)

References

1 R. Dent, M. Trudeau, K. I. Pritchard, W. M. Hanna, H. K. Kahn, C. A. Sawka, L. A. Lickley, E. Rawlinson, P. Sun, and S. A. Narod: Clin. Cancer Res. **13** (2007) 4429. https://doi.org/10.1158/1078-0432.CCR-06-3045
2 W. D. Foulkes, I. E. Smith, and J. S. Reis-Filho: New England J. Med. **363** (2010) 1938. https://doi.org/10.1056/NEJMra1001389
3 G. Bianchini, J. M. Balko, I. A. Mayer, M. E. Sanders, and L. Gianni: Nat. Rev. Clin. Oncol. **13** (2016) 674. https://doi.org/10.1038/nrclinonc.2016.66
4 B. D. Lehmann, J. A. Bauer, X. Chen, M. E. Sanders, A. B. Chakravarthy, Y. Shyr, and J. A. Pietenpol: The J. Clin. Invest. **121** (2011) 2750. https://doi.org/10.1172/JCI45014
5 S. Li, B. Zhang, D. Jiang, Y. Wei, and N. Zhang: BMC Bioinf. **11** (2010) 1. https://doi.org/10.1186/1471-2105-11-S11-S6
6 A. Sheik, K. Kim, G. L. Varaprasad, H. Lee, S. Kim, E. Kim, J.-Y. Shin, S. Y. Oh, and Y. S. Huh: Phytomedicine **91** (2021) 153698. https://doi.org/10.1016/j.phymed.2021.153698
7 M. N. Tran, N. S. Kim, and S. Lee: J. Ethnopharmacology **333** (2024) 118447. https://doi.org/10.1016/j.jep.2024.118447
8 B. Fu, N. Wang, H.-Y. Tan, S. Li, F. Cheung, and Y. Feng: Front. Pharmacol. **9** (2018) 1394. https://doi.org/10.3389/fphar.2018.01394
9 Y. Tan, H. Wang, B. Xu, X. Zhang, G. Zhu, Y. Ge, T. Lu, R. Gao, and J. Li: Front. Pharmacol. **13** (2022) 977708. https://doi.org/10.3389/fphar.2022.977708
10 A. L. Hopkins: Nat. Biotechnol. **25** (2007) 1110. https://doi.org/10.1038/nbt1007-1110
11 A.-L. Barabási, N. Gulbahce, and J. Loscalzo: Nat. Rev. Genet. **12** (2011) 56. https://doi.org/10.1038/nrg2918
12 J. Fang, C. Liu, Q. Wang, P. Lin, and F. Cheng: Briefings Bioinf. **19** (2018) 1153. https://doi.org/10.1093/bib/bbx045
13 W.-Y. Lee, C.-Y. Lee, Y.-S. Kim, and C.-E. Kim: Biomolecules **9** (2019) 362. https://doi.org/10.3390/biom9080362
14 Y. W. Kim, S. B. Bak, W.-Y. Lee, S. J. Bae, E. H. Lee, J.-H. Yang, K. Y. Kim, C. H. Song, S. C. Kim, and U.-J. Yun: J. Ginseng Res. **47** (2023) 479. https://doi.org/10.1016/j.jgr.2023.02.002
15 F. Gao, Y. Niu, L. Sun, W. Li, H. Xia, Y. Zhang, S. Geng, Z. Guo, H. Lin, and G. Du: J. Ethnopharmacol. **298** (2022) 115573. https://doi.org/10.1016/j.jep.2022.115573
16 T. N. Kipf and M. Welling: arXiv preprint arXiv:1609.02907 (2016). https://doi.org/10.48550/arXiv.1609.02907
17 Z. Gao, C. Jiang, J. Zhang, X. Jiang, L. Li, P. Zhao, H. Yang, Y. Huang, and J. Li: Nat. Commun. **14** (2023) 1093. https://doi.org/10.1038/s41467-023-36736-1
18 M. Zitnik, M. Agrawal, and J. Leskovec: Bioinformatics **34** (2018) i457. https://doi.org/10.1093/bioinformatics/bty294
19 C. Ruiz, M. Zitnik, and J. Leskovec: Nat. Commun. **12** (2021) 1796. https://doi.org/10.1038/s41467-021-21770-8
20 W.-Y. Lee, K.-I. Park, S.-B. Bak, S. Lee, S.-J. Bae, M.-J. Kim, S.-D. Park, C. O. Kim, J.-H. Kim, and Y. W. Kim: J. Adv. Res. (2024). https://doi.org/10.1016/j.jare.2024.12.040
21 S. B. Bak, Y. R. Song, S. J. Bae, W. Y. Lee, and Y. W. Kim: Free Radical Biol. Med. **199** (2023) 141. https://doi.org/10.1016/j.freeradbiomed.2023.02.016
22 R. Xue, Z. Fang, M. Zhang, Z. Yi, C. Wen, and T. Shi: Nucleic Acids Res. **41** (2012) D1089. https://doi.org/10.1093/nar/gks1100
23 W. Choi, C.-H. Choi, Y. R. Kim, S.-J. Kim, C.-S. Na, and H. Lee: Database **2016** (2016) baw011. https://doi.org/10.1093/database/baw011
24 C. Y.-C. Chen: PloS one **6** (2011) e15939. https://doi.org/10.1371/journal.pone.0015939
25 J. Ru, P. Li, J. Wang, W. Zhou, B. Li, C. Huang, P. Li, Z. Guo, W. Tao, and Y. Yang: J. Cheminformatics **6** (2014) 1. https://doi.org/10.1186/1758-2946-6-13
26 S.-K. Kim, S. Nam, H. Jang, A. Kim, and J.-J. Lee: BMC Complementary Altern. Med. **15** (2015) 1. https://doi.org/10.1186/s12906-015-0758-5

27 A. Naghizadeh, D. Hamzeheian, S. Akbari, F. Mohammadi, T. Otoufat, S. Asgari, A. Zarei, S. Noroozi, N. Nasiri, and M. Salamat: Evidence-Based Complementary Altern. Med. **2020** (2020) 3690781. https://doi.org/10.1155/2020/3690781

28 Y. Huang, J. Fang, W. Lu, Z. Wang, Q. Wang, Y. Hou, X. Jiang, O. Reizes, J. Lathia, and R. Nussinov: Cell Chem. Biol. **26** (2019) 1143. https://doi.org/j.chembiol.2019.05.004

29 Z. Fang, X. Liu, and G. Peltz: Bioinformatics **39** (2023) btac757. https://doi.org/https://doi.org/10.1093/bioinformatics/btac757

30 M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, and A. Lachmann: Nucleic Acids Res. **44** (2016) W90. https://doi.org/10.1093/nar/gkw377

31 B. Jang, Y. Kim, J. Song, Y.-W. Kim, and W.-Y. Lee: Int. J. Mol. Sci. **25** (2024) 12322. https://doi.org/10.3390/ijms252212322

32 P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio: arXiv preprint arXiv:1710.10903 (2017). https://doi.org/10.48550/arXiv.1710.10903

33 W. Hamilton, Z. Ying, and J. Leskovec: Adv. Neural Inf. Process. Syst. **30** (2017) 1. https://dl.acm.org/doi/10.5555/3294771.3294869

34 R. Li, X. Song, Y. Guo, P. Song, D. Duan, and Z.-S. Chen: Front. Oncol. **11** (2021) 772915. https://doi.org/10.3389/fonc.2021.772915

35 Z. Yang, Q. Zhang, L. Yu, J. Zhu, Y. Cao, and X. Gao: J. Ethnopharmacol. **264** (2021) 113249. https://doi.org/10.1016/j.jep.2020.113249

36 N. Wang, G. Muhetaer, X. Zhang, B. Yang, C. Wang, Y. Zhang, X. Wang, J. Zhang, S. Wang, and Y. Zheng: Front. Pharmacol. **11** (2020) 591400. https://doi.org/10.3389/fphar.2020.591400

37 N. Xiong, H. Wu, and Z. Yu: Front. Oncol. **14** (2024) 1405491. https://doi.org/10.3389/fonc.2024.1405491

38 Y.-Z. Zhang, J.-Y. Yang, R.-X. Wu, C. Fang, H. Lu, H.-C. Li, D.-M. Li, H.-L. Zuo, L.-P. Ren, and X.-Y. Liu: Front. Pharmacol. **12** (2021) 714628. https://doi.org/10.3389/fphar.2021.714628

39 R. Sainsbury, I. Proctor, S. Rodriguez, M. Loddo, S. Tudzarova, K. Stoeber, and G. Williams: Breast Cancer Res. **12** (2010) 1. https://doi.org/10.1186/bcr2547

40 D. Horiuchi, L. Kusdra, N. E. Huskey, S. Chandriani, M. E. Lenburg, A. M. Gonzalez-Angulo, K. J. Creasman, A. V. Bazarov, J. W. Smyth, and S. E. Davis: J. Exp. Med. **209** (2012) 679. https://doi.org/10.1084/jem.20111512